MOLECULAR ORGANIZATION, EVOLUTION, AND FUNCTION OF RIBOSOMAL DNA

EDITED BY: Roman A. Volkov, Nikolai Borisjuk, Sonia Garcia, Ales Kovarik and Julio Sáez-Vásquez

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-962-9
DOI 10.3389/978-2-88976-962-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MOLECULAR ORGANIZATION, EVOLUTION, AND FUNCTION OF RIBOSOMAL DNA

Topic Editors:
Roman A. Volkov, Chernivtsi University, Ukraine
Nikolai Borisjuk, Huaiyin Normal University, China
Sonia Garcia, Spanish National Research Council (CSIC), Spain
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia
Julio Sáez-Vásquez, Université de Perpignan Via Domitia, France

Citation: Volkov, R. A., Borisjuk, N., Garcia, S., Kovarik, A., Sáez-Vásquez, J., eds. (2023). Molecular Organization, Evolution, and Function of Ribosomal DNA. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-962-9

Table of Contents

05 ***Editorial: Molecular organization, evolution, and function of ribosomal DNA***
Roman A. Volkov, Nikolai Borisjuk, Sònia Garcia, Aleš Kovařík and Julio Sáez-Vásquez

08 ***G4 Structures in Control of Replication and Transcription of rRNA Genes***
Kateřina Havlová and Jiří Fajkus

14 ***Ancient Origin of Two 5S rDNA Families Dominating in the Genus* Rosa *and Their Behavior in the Canina-Type Meiosis***
Radka Vozárová, Veit Herklotz, Aleš Kovařík, Yuri O. Tynkevich, Roman A. Volkov, Christiane M. Ritz and Jana Lunerová

29 ***Molecular Evolution and Organization of Ribosomal DNA in the Hawkweed Tribe Hieraciinae (Cichorieae, Asteraceae)***
Judith Fehrer, Renáta Slavíková, Ladislava Paštová, Jiřina Josefiová, Patrik Mráz, Jindřich Chrtek and Yann J. K. Bertrand

52 ***Targeted Enrichment of rRNA Gene Tandem Arrays for Ultra-Long Sequencing by Selective Restriction Endonuclease Digestion***
Anastasia McKinlay, Dalen Fultz, Feng Wang and Craig S. Pikaard

60 ***Horizontally Acquired nrDNAs Persist in Low Amounts in Host* Hordeum *Genomes and Evolve Independently of Native nrDNA***
Karol Krak, Petra Caklová, David Kopecký, Frank R. Blattner and Václav Mahelka

73 ***Mosaic Arrangement of the 5S rDNA in the Aquatic Plant* Landoltia punctata *(Lemnaceae)***
Guimin Chen, Anton Stepanenko and Nikolai Borisjuk

83 ***The Arabidopsis 2′-O-Ribose-Methylation and Pseudouridylation Landscape of rRNA in Comparison to Human and Yeast***
Deniz Streit and Enrico Schleiff

102 ***It Is Just a Matter of Time: Balancing Homologous Recombination and Non-homologous End Joining at the rDNA Locus During Meiosis***
Jason Sims, Fernando A. Rabanal, Christiane Elgert, Arndt von Haeseler and Peter Schlögelhofer

113 ***Personal Perspectives on Plant Ribosomal RNA Genes Research: From Precursor-rRNA to Molecular Evolution***
Vera Hemleben, Donald Grierson, Nikolai Borisjuk, Roman A. Volkov and Ales Kovarik

128 ***Integrating Wheat Nucleolus Structure and Function: Variation in the Wheat Ribosomal RNA and Protein Genes***
Rudi Appels, Penghao Wang and Shahidul Islam

147 ***The Nuclear 35S rDNA World in Plant Systematics and Evolution: A Primer of Cautions and Common Misconceptions in Cytogenetic Studies***
Josep A. Rosselló, Alexis J. Maravilla and Marcela Rosato

162 ***The Ribosomal DNA Loci of the Ancient Monocot* Pistia stratiotes *L. (Araceae) Contain Different Variants of the 35S and 5S Ribosomal RNA Gene Units***
Anton Stepanenko, Guimin Chen, Phuong T. N. Hoang, Jörg Fuchs, Ingo Schubert and Nikolai Borisjuk

178 *5S Ribosomal DNA of Genus* **Solanum***: Molecular Organization, Evolution, and Taxonomy*
Yurij O. Tynkevich, Antonina Y. Shelyfist, Liudmyla V. Kozub, Vera Hemleben, Irina I. Panchuk and Roman A. Volkov

207 *To Be or Not to Be Expressed: The First Evidence of a Nucleolar Dominance Tissue-Specificity in* **Brachypodium hybridum**
Natalia Borowska-Zuchowska, Ewa Robaszkiewicz, Serhii Mykhailyk, Joanna Wartini, Artur Pinski, Ales Kovarik and Robert Hasterok

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 04 August 2022
DOI 10.3389/fpls.2022.994380

OPEN ACCESS

EDITED AND REVIEWED BY
Hank W. Bass,
Florida State University, United States

*CORRESPONDENCE
Roman A. Volkov
r.volkov@chnu.edu.ua

SPECIALTY SECTION
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

RECEIVED 14 July 2022
ACCEPTED 22 July 2022
PUBLISHED 04 August 2022

CITATION
Volkov RA, Borisjuk N, Garcia S, Kovařík A and Sáez-Vásquez J (2022) Editorial: Molecular organization, evolution, and function of ribosomal DNA. *Front. Plant Sci.* 13:994380. doi: 10.3389/fpls.2022.994380

Editorial: Molecular organization, evolution, and function of ribosomal DNA

Roman A. Volkov[1]*, Nikolai Borisjuk[2], Sònia Garcia[3], Aleš Kovařík[4] and Julio Sáez-Vásquez[5]

[1]Department of Molecular Genetics and Biotechnology, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, [2]Jiangsu Key Laboratory for Eco-Agricultural Biotechnology Around Hongze Lake and Jiangsu Collaborative Innovation Centre of Regional Modern Agriculture and Environmental Protection, School of Life Sciences, Huaiyin Normal University, Huai'an, China, [3]Institut Botànic de Barcelona - Consejo Superior de Investigaciones Científicas (IBB-CSIC), Barcelona, Spain, [4]Department of Molecular Epigenetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia, [5]CNRS, Laboratoire Génome et Développement des Plantes (LGDP), UMR 5096, Perpignan, France

KEYWORDS

concerted evolution, epigenetics, molecular phylogeny and taxonomy, nucleolus, polyploidy, rRNA processing

Editorial on the Research Topic
Molecular organization, evolution, and function of ribosomal DNA

Introduction

The aim of this Research Topic is to highlight the current status of knowledge and research on plant ribosomal DNA (rDNA). The Topic compiles seven Original Research papers, five Reviews, one Perspective and one Methods articles, viewed more than 26,000 times by the time of this Editorial. The scope covers diverse modern technologies, scientific approaches, and research aimed at achieving a better understanding of the many, complex aspects of rDNA structure, evolution, regulation, and functions in plant development and adaptation.

The rDNA encodes four ribosomal RNA (rRNAs), which are the major components of ribosome and constitute 65–75% of the plant cell's total RNA. Because of its abundance, functional importance and specific organization in evolutionarily conserved rRNA coding sequences, and rapidly evolving intergenic spacer (IGS) regions, the chromosomal and molecular organization, transcription and evolution of the rDNA have been intensively studied since the early days of plant molecular biology.

The history of rDNA research started almost 90 years ago when McClintock (1934) observed that in the interphase nuclei of maize the nucleolus was formed in association with a specific region of a chromosome, which she called the nucleolar organizer region (NOR). Early rDNA research in plants is presented in article of Hemleben et al., which covers topics such as the synthesis of rRNA precursors, processing, the organization and evolution of 5S and 18S-5.8S-26S (or 35-45S) rDNA as well as

epigenetic phenomena and the impact of hybridization and allopolyploidy on rDNA expression and homogenization. This historical view sets the scene for the other articles highlighting the progress in modern rDNA research.

Ribosomal DNA function and expression

The rRNA synthesis involves specialized transcription complexes built around RNA Polymerase I for 35-45S rRNAs and RNA Polymerase III for 5S rRNA. The regulation of rDNA expression in response to multiple internal factors and external stimuli utilizes various epigenetic mechanisms such as DNA methylation, histone modifications or RNA interference. Havlová and Fajkus focus on unusual structural features of DNA, namely shortly spaced oligo-guanine tracts able to form G-quadruplex (G4) structures. They discussed the role of these structures in regulating rDNA activity in two model plants, *Arabidopsis thaliana* (angiosperm) and *Physcomitrella patens* (moss).

Nucleolar dominance (ND) represents the selective silencing of parental 35-45S rDNA loci in the genome of a hybrid or allopolyploid. Borowska-Zuchowska et al. analyzed ND in two genotypes of a model allotetraploid grass, *Brachypodium hybridum* (Poaceae). They found that ND was developmentally stable in one but not the other accession, the latter showing a codominant expression of parental rDNA in adventitious roots.

The current knowledge on the 35-45S pre-rRNA modifications, which is a mandatory step of rRNA processing during ribosome assembly in the nucleolus, is summarized by Streit and Schleiff. Hundreds of ribosome biogenesis factors (RBFs) and small nucleolar RNAs (snoRNA) catalyze the rRNA processing (cleavages and modifications) in *Arabidopsis thaliana*. Small nucleolar ribonucleoprotein particles (snoRNPs) complexes, composed of C/D or H/ACA snoRNAs and RBFs, catalyze, respectively, the two major rRNA modification types, 2'-*O*-ribose-methylation and pseudouridylation, which are required for stability of rRNA structure and for translational accuracy and efficiency.

Coordinated production and integration of rRNA and protein components into cytoplasmic ribosomes and the nucleolar organizer regions (NORs) in response to changes in genetic constitution, biotic and abiotic stresses are reviewed by Appels et al. Using a hexaploid wheat *Triticum aestivum* (Poaceae) as a model, it is argued that unique functionalities of ribosome populations can become central in situations of stress conditions by preferentially translating mRNAs coding for proteins contributing to survival of the cell.

Organization and evolution of rDNA

5S and 35-45S rDNA are composed of numerous copies of tandemly arranged repeated units (repeats), which are usually located on one or few chromosomal loci. In many species, numerous rDNA repeats tend to be very similar due to sequence homogenization, i.e., individual repeats do not evolve independently, but in a concerted manner (Arnheim et al., 1980; Coen et al., 1982; Volkov et al., 1999). The level of intragenomic homogenization may differ in different taxa and for the different rDNAs (5S vs 35-45S), and the reason for this remains enigmatic. Sims et al. provided a hypothetical model under which the genetic landscape of rDNA arrays is driven by a balance between non-homologous end joining (NHEJ) and homologous recombination (HR). While NHEJ increases the array heterogeneity by introducing point mutation and indels, HR acts in the opposite direction homogenizing the units. It is widely believed that the rDNA repeats should be nearly identical within the same chromosomal locus, while the repeats from different loci may show lower similarity. This idea was confirmed by examining diploid and polyploid species of the genus *Rosa* (Rosaceae), for which two highly divergent 5S rDNA families located on different chromosomes were identified (Vozárová et al.). Both gene families arose in the early history of the genus, already 30 myrs ago, and apparently survived numerous speciation events thereafter.

The intragenomic diversity of 5S rDNA was examined for 137 *Solanum* (Solanaceae) species (Tynkevich et al.), possessing one 5S locus per chromosome set. It was shown that many repeat variants coexist within the genome demonstrating incomplete sequence homogenization. The main mechanisms of 5S rDNA molecular evolution in the genus *Solanum* was step-wise accumulation of single base substitution or short insertions/deletions (indels) in the 5S IGS, whereas long indels and multiple base substitutions were mostly not conserved and eliminated.

In this Research Topic, the molecular organization of rDNA is explored for the first time for the aquatic plant duckweed *Landoltia punctata* representing the Lemnaceae. Chen et al. demonstrated the presence of two classes of 5S rDNA repeats, which differ by the composition and distribution of subrepeats in the IGS, and regulatory DNA elements potentially involved in 5S rDNA transcription. The genome of *L. punctata* has one of the lowest copy numbers of rDNA genes among flowering plants and an unusual, mosaic arrangement of 5S rDNA clusters. Stepanenko et al. characterized rDNA of another aquatic species, *Pistia stratiotes*, from the Araceae family. Whereas, the 5S and 35-45S rDNA were localized in a single chromosome locus each, the species' 35-45S rDNA is represented by at least four length variants, distinguished by the number of subrepeats within the IGS. The 5S rDNA locus includes at least six types of functional gene units, intermingled with each other and with pseudogenes.

The data support the idea of a mosaic arrangement of multiple variants of 5S and 35-45S rDNA units in single locus as the rule rather than the exception.

Nuclear rDNA demonstrates extraordinary dynamics during evolution. In diploid *Hordeum* (Poaceae) species, Krak et al. analyzed the fate of alien 35-45S rDNA copies acquired *via* horizontal transfer from panicoid genera. The foreign ribotypes were present in the respective genomes at low copy numbers, likely representing a minor fraction of the total rDNA dedicated to pseudogenization.

It is clear that the knowledge of the intragenomic diversity of 35-45S rDNA is still incomplete, since large NORs are generally missing from existing genome assemblies due to their highly repetitive nature. In the future, organization of these complex areas composed of relatively long (∼9–20 kb) regions can potentially be deciphered by third generation sequencing methods, such as Oxford Nanopore Technology (ONT). This issue is addressed by McKinlay et al., who developed a method providing enrichment of 35-45S rDNA sequences among ultra-long ONT reads.

Taxonomic application of rDNA

The rapidly evolving spacer regions of rDNA provide a convenient tool for phylogenetic studies of lower-ranking taxa. Fehrer et al. point out that the appropriate treatment of intra-individual variation and the investigation of multiple markers allows interesting insights in complex species relationships as well as in the evolution of the markers themselves. In the *Hieracium* (Asteraceae) genus, they found that chromosomal location of the 5S and 35-45S rDNA loci is far more dynamic than the sequences they contain, implying that chromosomal patterns are not suitable to infer species relationships, at least not in *Hieracium*.

The comparison of 5S IGS was successfully applied to reconstruct the phylogeny of the giant genus *Solanum* (Solanaceae) (Tynkevich et al.), allowing clarification of taxonomic position of several species and detection of reticulate evolution, especially in its largest section, *Petota*.

The correct interpretation of rDNA markers in plant taxonomy and evolution is not free of drawbacks. Accordingly, Rosselló et al. aim to discuss the limitations of nuclear 35-45S rDNA markers based on cytological and karyological experimental work to draw sound biological and evolutionary conclusions in a systematic and phylogenetic context. The authors offer clarification for some misconceptions usually found in published work that could help lead to an insightful utilization of the ribosomal DNA world in plant evolution.

Author contributions

The manuscript was written by RV with inputs from NB, SG, JS-V, and AK. All authors approved the submitted version.

Funding

RV and AK thank respectively the Ministry of Education and Science of Ukraine (Grant No. 0122U001335) and the Czech Science Foundation (20-14133J) for support. NB was supported by an individual grant provided by the Huaiyin Normal University (Huai'an, China).

Acknowledgments

We are grateful to all authors and journal editors who contributed to this Research Topic. We appreciate the hard work of peer reviewers who helped to maintain high standards of the publications in this issue.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., and Zimmer, E. (1980). Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proc. Natl. Acad. Sci. U. S. A.* 77, 7323–7327. doi: 10.1073/pnas.77.12.7323

Coen, E. S, Thoday, J. M, and Dover, G. (1982). Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature* 295, 564–568. doi: 10.1038/295564a0

McClintock, B. (1934). The relationship of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Z. Zellforsch Microsk.* 21, 294–398. doi: 10.1007/BF00374060

Volkov, R. A., Borisjuk, N. V., Panchuk, I. I., Schweizer, D., and Hemleben, V. (1999). Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. *Mol. Biol. Evol.* 16, 311–320. doi: 10.1093/oxfordjournals.molbev.a026112

frontiers in Plant Science

PERSPECTIVE
published: 08 October 2020
doi: 10.3389/fpls.2020.593692

G4 Structures in Control of Replication and Transcription of rRNA Genes

Kateřina Havlová[1,2] *and Jiří Fajkus*[1,2,3]*

[1]*Laboratory of Functional Genomics and Proteomics, National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czechia,* [2]*Chromatin Molecular Complexes, Mendel Centre for Plant Genomics and Proteomics, Central European Institute of Technology, Masaryk University, Brno, Czechia,* [3]*Department of Cell Biology and Radiobiology, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czechia*

Genes encoding 45S ribosomal RNA (rDNA) are known for their abundance within eukaryotic genomes and for their unstable copy numbers in response to changes in various genetic and epigenetic factors. Commonly, we understand as epigenetic factors (affecting gene expression without a change in DNA sequence), namely DNA methylation, histone posttranslational modifications, histone variants, RNA interference, nucleosome remodeling and assembly, and chromosome position effect. All these were actually shown to affect activity and stability of rDNA. Here, we focus on another phenomenon – the potential of DNA containing shortly spaced oligo-guanine tracts to form quadruplex structures (G4). Interestingly, sites with a high propensity to form G4 were described in yeast, animal, and plant rDNAs, in addition to G4 at telomeres, some gene promoters, and transposons, suggesting the evolutionary ancient origin of G4 as a regulatory module. Here, we present examples of rDNA promoter regions with extremely high potential to form G4 in two model plants, *Arabidopsis thaliana* and *Physcomitrella patens*. The high G4 potential is balanced by the activity of G4-resolving enzymes. The ability of rDNA to undergo these "structural gymnastics" thus represents another layer of the rich repertoire of epigenetic regulations, which is pronounced in rDNA due to its highly repetitive character.

Keywords: rDNA stability, transcription, replication, quadruplex DNA, G4, ribosomal RNA genes

OPEN ACCESS

Edited by:
Julio Sáez-Vásquez, Université de Perpignan Via Domitia, France

Reviewed by:
Zhong-Hui Zhang, South China Normal University, China
Hao Wang, South China Agricultural University, China

***Correspondence:**
Jiří Fajkus
fajkus@sci.muni.cz;
tttaggg@gmail.com

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *11 August 2020*
Accepted: *22 September 2020*
Published: *08 October 2020*

Citation:
Havlová K and Fajkus J (2020) G4 Structures in Control of Replication and Transcription of rRNA Genes. Front. Plant Sci. 11:593692. doi: 10.3389/fpls.2020.593692

INTRODUCTION

Among many potential reasons to become interested in genes encoding ribosomal RNA (rRNA) is the possibility to study the wide range of regulatory mechanisms used to control their expression and genomic stability. When starting from the genomic level, genes for 45S rRNA (rDNA) usually form the most abundant gene family in most eukaryotes (e.g., 150 copies per haploid genome in *Saccharomyces cerevisiae*, Kobayashi et al., 1998; 300 in human, Schmickel, 1973; Agrawal and Ganley, 2018; and 600 in *Arabidopsis thaliana*, Pruitt and Meyerowitz, 1986; Copenhaver et al., 1995) with a considerable individual variability in a copy number. Variability can also be seen in the lengths and nucleotide sequences of intergenic spacers separating individual transcription units of 18S-5.8S-25S transcribed by RNA Polymerase I, while the nucleotide sequences of genes coding for 18S, 5.8S and 25S rRNAs are highly conserved (reviewed in Dvorackova et al., 2015). rDNAs form one or more tandemly arranged

gene clusters (nucleolus organizing regions, NORs) per haploid genome whose sizes are maintained within a standard range as a result of dynamic balance between the loss and recovery of individual rDNA repeats. rDNA copies are most notably lost by intra-chromatid recombination between distant rDNA copies, leading to excision of the intervening copies in the form of an extrachromosomal rDNA circle. These events can be counteracted by various recombination events, e.g., an unequal sister chromatid recombination or an unequal sister chromatid exchange, which are induced in a response to DNA double strand breaks generated due to arrested replication forks (see Nelson et al., 2019, for recent review).

In some organisms, e.g., *S. cerevisiae* (Bayev et al., 1980) or the moss *Physcomitrella patens* (Goffova et al., 2019), rDNA units also comprise 5S rRNA genes inserted in the intergenic spacers between individual 18S-5.8S-25S transcription units. 5S rRNA is not present in the primary RNA Pol I transcript but is transcribed by RNA Pol III. Besides RNA Pol I and – in some cases – RNA Pol III promoters, intergenic spacers also show the presence of additional promoters (spacer promoters), which may promote transcription by RNA Pol I or II, giving rise to non-coding (nc)RNAs affecting rRNA expression (Doelling et al., 1993; Mayer et al., 2006; Cesarini et al., 2010; Earley et al., 2010; Agrawal and Ganley, 2018).

Indeed, rDNA clusters represent a miniature system of their own where concurrent functions of different kinds of promoters and polymerases can be observed, replication origins are present (and obviously closely spaced), replication and transcription polymerases can meet and occasionally collide, and DNA repair mechanisms must eventually solve problems arising from all this apparent turmoil.

On the other hand, this mini-world has also developed numerous tools of precise regulation which began to be understood in molecular details recently. These include a phenomenon termed nucleolar dominance (see, e.g., Preuss and Pikaard, 2007; Chandrasekhara et al., 2016; Mohannath et al., 2016).

Further, the importance of an appropriate higher order chromatin arrangement for rDNA stability was highlighted in recent studies (Pontvianne et al., 2013, 2016), as well as was the role of histone chaperones in the assembly of the very basic units of chromatin – the nucleosomes (Mozgova et al., 2010; Muchova et al., 2015; Pavlistova et al., 2016). Further, the role of DNA methylation and histone acetylation in the control of rDNA activity has been elucidated (Probst et al., 2004; Grummt, 2007; Mcstay and Grummt, 2008; Pontvianne et al., 2010; Schmitz et al., 2010), as well as the enigmatic importance of keeping a considerable fraction of rDNA units inactive (Kobayashi, 2011).

In addition to all the interesting knowledge accumulated on rDNA/rRNA topics in the last decades, a specific feature of rDNA has been observed – its propensity to form tetraplex (quadruplex) structures (G4), which are based on guanine tetrads. This feature seems to be conserved throughout eukaryotes (Hanakahi et al., 1999; Hershman et al., 2008; Capra et al., 2010; Goffova et al., 2019; Matyasek et al., 2019; Mestre-Fos et al., 2019a) and is thought to contribute significantly to the inherently low stability of rDNA as an obstacle to advancing replication forks. Stalled and collapsed replication forks then induce repair events which may result in rDNA loss or expansion (see above). Effects of the high propensity to form G4 structures become more evident when functions of intrinsic factors (e.g., specific helicases) which are able to dissolve G4 structures are disrupted or compromised, resulting in a hyper-recombinogenic character of rDNA and its instability.

Here, we exemplify the role of G4 structures in rDNA of two model plants, *P. patens* and *A. thaliana*.

DESTABILIZATION OF RDNA DUE TO DYSFUNCTION OF G4-RESOLVING HELICASES AND COLOCALIZATION OF G4 SITES WITH GENE AND SPACER PROMOTERS IN *ARABIDOPSIS THALIANA* RDNA

In *Arabidopsis thaliana*, it was found recently that RecQ-mediated genome instability protein 2 (RMI2) and Regulator of telomere elongation helicase 1 (RTEL1) contribute to the stability of the 45S rDNA copy number (Rohrig et al., 2016). RMI2 in *Arabidopsis*, as well as in yeasts and humans, acts for a proper dissolution of recombination intermediates, thereby suppressing a hyper-recombinogenic phenotype (Wu and Hickson, 2003). Also, RTEL1 (initially described in *Caenorhabditis elegans*) functions as a Fe-S cluster helicase suppressing inappropriate recombination events by promoting disassembly of D-loop recombination. Furthermore, RTEL1 can dissolve quadruplex (G4) DNA structures that otherwise block the extension of telomeres by telomerase (Vannier et al., 2012), and in humans, its dysfunction causes Hoyeraal-Hreidarsson syndrome, a severe form of dyskeratosis congenita, which is characterized by short telomeres and genome instability (Le Guen et al., 2013; Vannier et al., 2013, 2014; Faure et al., 2014). RTEL1 also promotes genome-wide replication through its interaction with PCNA, increasing replication fork stability, extension rates, and origin usage (Vannier et al., 2013).

Both AtRMI2 and AtRTEL1 participate in the maintenance of rDNA stability in parallel pathways. In *atrmi2* plants, 45S rDNA decreased to 80%, in *atrtel1* plants to 40%, and in double *atrmi2 atrtel1* mutants to ca. 30% of their standard copy number (Rohrig et al., 2016). A similar contribution to rDNA stability was also observed in another Fe-S cluster helicase – FANCJ homolog in *Arabidopsis* – AtFANCJB (Dorn et al., 2019).

These results are consistent with the fact that *A. thaliana* rDNA repeat units show the presence of a cluster of sites with a strong potential to form a G4 structure (**Figure 1**). The highest score obtained using the pqsfinder tool (Hon et al., 2017; Labudova et al., 2020) coincides with the gene promoter (GP) site, reaching a value (77) higher than the scores of the best characterized G4-forming DNAs, plant or human telomeric repeats (60 and 64, respectively; Goffova et al., 2019). Presumably, formation of G4 in the plus-strand at the promoter sites may strongly inhibit 45S rDNA transcription and slow down its replication. Two other G4 sites were detected at spacer promoters,

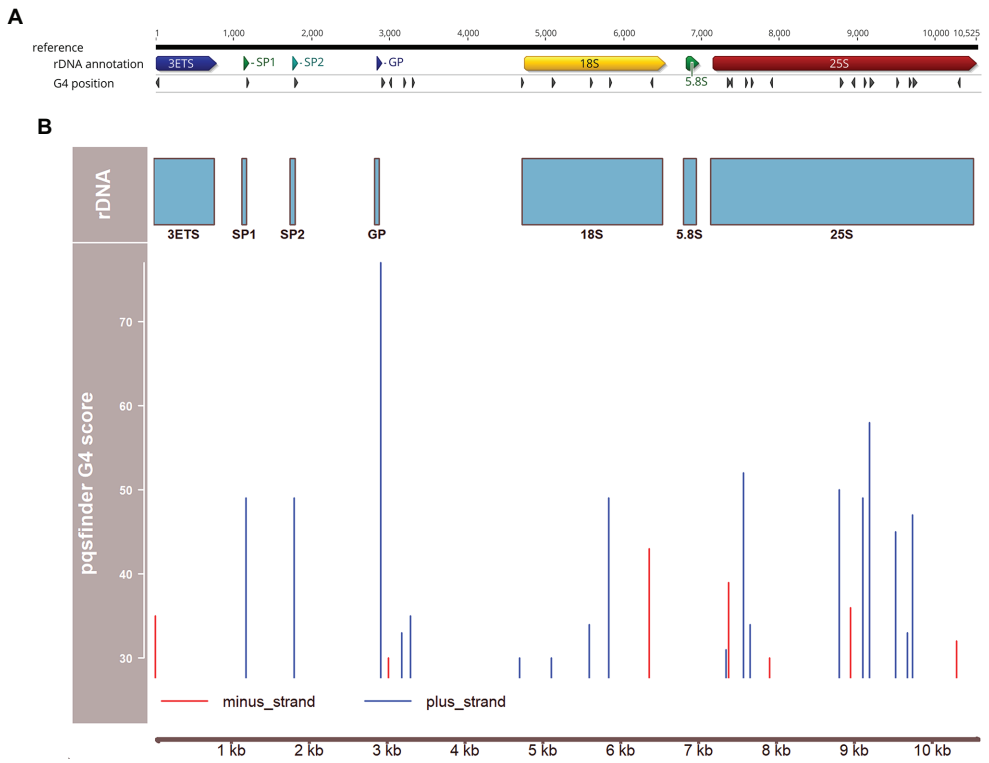


FIGURE 1 | Distribution of potential G4-forming sequences over the 45S ribosomal DNA (rDNA) unit of *Arabidopsis thaliana*. **(A)** Map of the rDNA unit using the data from Chandrasekhara et al. (2016), and the Geneious software platform (Biomatters, Auckland, New Zealand). Positions of 3' external transcribed Spacer (3ETS), spacer promoter 1 and 2 (SP1, SP2, respectively), gene promoter (GP) and 18S, 5.8S, and 25S rRNA genes are indicated. **(B)** Positions and scores of G4 structures predicted using pqsfinder (Hon et al., 2017) and plotted with the Bioconductor package Gviz (Hahne and Ivanek, 2016).

SP1 and SP2. Interestingly, the number of spacer promoters (and, consequently, a number of G4s) varies among rDNA units in *A. thaliana* (Havlova et al., 2016), which may represent a novel layer in regulation of transcription and replication of individual rDNA units. Yet, additional G4 sites were found inside the coding regions for 18S rRNA and 25S rRNA (**Figure 1**). These results are thus consistent with the view that G4 sites play important roles not only in rDNA replication and genome stability (supported by the abovementioned observations on *A. thaliana* helicase mutants) but also in control of rDNA transcription.

IN ADDITION TO THE FEATURES OBSERVED IN *ARABIDOPSIS THALIANA*, A CLUSTER OF G4 SITES SEPARATES 5S AND 18S RRNA GENES TRANSCRIBED WITH POL III AND POL I, RESPECTIVELY, IN *PHYSCOMITRELLA PATENS*

The situation in *P. patens* rDNA is complicated by the linked arrangement between 18S-5.8S-25S units and 5S rRNA genes. This arrangement has been demonstrated recently (Goffova et al., 2019) and is congruent with its earlier description in a liverwort, *Marchantia polymorpha*, and a moss *Funaria hygrometrica* (Sone et al., 1999), as well as with a later systematic study in land plants (Wicke et al., 2011). *P. patens RTEL1* mutants (*pprtel1*), similar to *atrtel1* mutants, also show a marked decrease of 18S rDNA copies (representing 45S rDNA), but, in addition, a comparable decrease of 5S rDNA is observed (Goffova et al., 2019). Interestingly, while reduced relative transcript levels of 18S rRNA roughly correspond to the decrease in their genomic copies in *pprtel1* plants, reduction in 5S rRNA transcripts is more pronounced, without any obvious relation to 5S rDNA copy number. This indicates a relatively independent regulation of 5S and 45S rDNA transcription.

In a search of a mechanistic explanation of our results, we found a noticeable clustering of putative G4 sites in the spacer region between 5S and 18S rRNA genes (Goffova et al., 2019). Prediction of G4 propensity revealed a particularly strong site in the plus-strand (thus with a presumable inhibitory role in transcription) ca. 500 bp upstream of the 18S rRNA gene where the pqsfinder score reached a value of 132, which is twice higher than that of telomere DNA. These results were confirmed by another prediction tool, G4Hunter (Bedrat et al., 2016). An independent indication of the high G4 potential of this region was supported by our observation that PCR amplification was problematic across the linker between the 5S and 18S rRNA genes, requiring addition of dimethyl sulfoxide

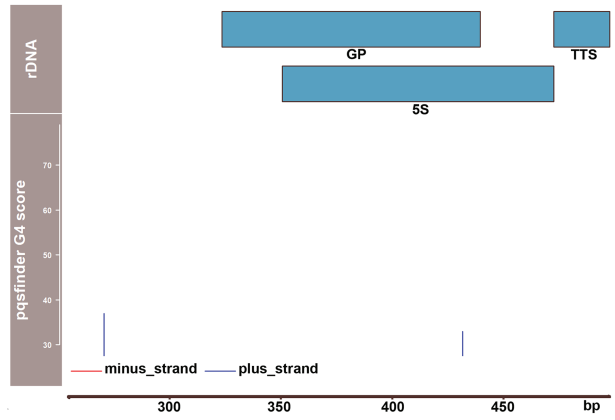


FIGURE 2 | Distribution of potential G4-forming sequences over the 5S rDNA unit of *A. thaliana*. 5S rDNA sequence data (Campell et al., 1992) were used to analyze pqsfinder scores. Two putative G4 sites starting in positions 257 and 414 were identified with pqsfinder showing the scores of 37 and 33, respectively. The GP including both external and internal elements, 5S rRNA gene (5S) and transcription termination site (TTS) are depicted using the data from Cloix et al. (2003).

(DMSO) to the reaction mixture. Furthermore, our repeat clustering analysis indicated a high potential of this region to form non-canonical structures by a dramatically (two orders of magnitude) lower number of NGS reads when compared with the neighboring regions (Goffova et al., 2019). Thus, it is conceivable that in addition to the G4 roles suggested in *A. thaliana* rDNA based on experiments and predictions, yet another putative function is provided by G4 sites in *P. patens* – a protection against collision or interference between advancing RNA Polymerases I and III. This hypothesis is supported by the absence of any sites of a comparable G4 potential in *A. thaliana* 5S rDNA unit, which is located separately from the 45S rDNA locus (**Figure 2**).

CONCLUSION

G4 formation and resolution can be regarded as a dynamic switch whose identity is defined genetically – through its primary DNA sequence – but its "ON" and "OFF" states are controlled by the local availability of G4-targeting proteins or other ligands that affect the G4 stability positively or negatively. As this switch acts in control of transcription and replication without a change in the primary DNA sequence, we suggest that the formation of G4 structures (and possibly also the other relevant non-canonical DNA secondary structures) be included among epigenetic mechanisms.

In rDNA, epigenetic effects of G4 formation can be expected preferentially at active copies (where a lesser nucleosome density or even nucleosome removal can be expected around transcription start sites – thereby facilitating formation of G4) or during replication when DNA strands are temporarily separated and noncanonical intrastrand structures can be formed. In addition to G4s formed by rDNA as discussed above, recent results suggest possible roles of G4s formed by rRNAs. Interestingly, these potential G4s are located on surfaces of both subunits of the human ribosome (Mestre-Fos et al., 2019b). When assuming that rRNA is the most abundant fraction of cellular RNA, then these G4-rRNAs clearly dominate the total population of RNA quadruplexes, thus indicating another perspective topic of future studies.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article.

AUTHOR CONTRIBUTIONS

KH performed annotation of rDNA region and prediction of G4 sites. JF wrote the manuscript. Both authors have read, edited, and approved the final version of the manuscript.

FUNDING

This work was supported by the project SYMBIT, reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed by the ERDF, and by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601) and INTER-COST project LTC20003.

ACKNOWLEDGMENTS

The authors thank JL Mergny for inspiring discussions on G4. We apologize to researchers whose work we have been unable to cite owing to space constraints.

REFERENCES

Agrawal, S., and Ganley, A. R. D. (2018). The conservation landscape of the human ribosomal RNA gene repeats. *PLoS One* 13:e0207531. doi: 10.1371/journal.pone.0207531

Bayev, A. A., Georgiev, O. I., Hadjiolov, A. A., Kermekchiev, M. B., Nikolaev, N., Skryabin, K. G., et al. (1980). The structure of the yeast ribosomal RNA genes. 2. The nucleotide-sequence of the initiation site for ribosomal RNA transcription. *Nucleic Acids Res.* 8, 4919–4926. doi: 10.1093/nar/8.21.4919

Bedrat, A., Lacroix, L., and Mergny, J. L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746–1759. doi: 10.1093/nar/gkw006

Campell, B. R., Song, Y. G., Posch, T. E., Cullis, C. A., and Town, C. D. (1992). Sequence and organization of 5s ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene* 112, 225–228. doi: 10.1016/0378-1119(92)90380-8

Capra, J. A., Paeschke, K., Singh, M., and Zakian, V. A. (2010). G-Quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.* 6:e1000861. doi: 10.1371/journal.pcbi.1000861

Cesarini, E., Mariotti, F. R., Cioci, F., and Camilloni, G. (2010). RNA polymerase I transcription silences noncoding RNAs at the ribosomal DNA locus in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 9, 325–335. doi: 10.1128/EC.00280-09

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., and Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the *Arabidopsis thaliana* 5S rRNA genes. *Plant J.* 35, 251–261. doi: 10.1046/j.1365-313x.2003.01793.x

Copenhaver, G. P., Doelling, J. H., Gens, J. S., and Pikaard, C. S. (1995). Use of RFLPs larger than 100-Kbp to map the position and internal organization of the nucleolus organizer region on chromosome-2 in *Arabidopsis thaliana*. *Plant J.* 7, 273–286. doi: 10.1046/j.1365-313x.1995.7020273.x

Doelling, J. H., Gaudino, R. J., and Pikaard, C. S. (1993). Functional-analysis of *Arabidopsis thaliana* ribosomal RNA gene and spacer promoters in vivo and by transient expression. *Proc. Natl. Acad. Sci. U. S. A.* 90, 7528–7532. doi: 10.1073/pnas.90.16.7528

Dorn, A., Feller, L., Castri, D., Rohrig, S., Enderle, J., Herrmann, N. J., et al. (2019). An *Arabidopsis* FANCJ helicase homologue is required for DNA crosslink repair and rDNA repeat stability. *PLoS Genet.* 15:e1008174. doi: 10.1371/journal.pgen.1008174

Dvorackova, M., Fojtova, M., and Fajkus, J. (2015). Chromatin dynamics of plant telomeres and ribosomal genes. *Plant J.* 83, 18–37. doi: 10.1111/tpj.12822

Earley, K. W., Pontvianne, F., Wierzbicki, A. T., Blevins, T., Tucker, S., Costa-Nunes, P., et al. (2010). Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev.* 24, 1119–1132. doi: 10.1101/gad.1914110

Faure, G., Revy, P., Schertzer, M., Londono-Vallejo, A., and Callebaut, I. (2014). The C-terminal extension of human RTEL1, mutated in Hoyeraal-Hreidarsson syndrome, contains harmonin-N-like domains. *Proteins* 82, 897–903. doi: 10.1002/prot.24438

Goffova, I., Vagnerova, R., Peska, V., Franek, M., Havlova, K., Hola, M., et al. (2019). Roles of RAD51 and RTEL1 in telomere and rDNA stability in *Physcomitrella patens*. *Plant J.* 98, 1090–1105. doi: 10.1111/tpj.14304

Grummt, I. (2007). Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. *Hum. Mol. Genet.* 16, R21–R27. doi: 10.1093/hmg/ddm020

Hahne, F., and Ivanek, R. (2016). "Visualizing genomic data using gviz and bioconductor in *Statistical Genomics*. Methods in Molecular Biology. *Vol. 1418*. eds. E. Mathé and S. Davis (New York, NY: Humana Press), 335–351. doi: 10.1007/978-1-4939-3578-9_16

Hanakahi, L. A., Sun, H., and Maizels, N. (1999). High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* 274, 15908–15912. doi: 10.1074/jbc.274.22.15908

Havlova, K., Dvorackova, M., Peiro, R., Abia, D., Mozgova, I., Vansacova, L., et al. (2016). Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol. Biol.* 92, 457–471. doi: 10.1007/s11103-016-0524-1

Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., et al. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 36, 144–156. doi: 10.1093/nar/gkm986

Hon, J., Martinek, T., Zendulka, J., and Lexa, M. (2017). pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* 33, 3373–3379. doi: 10.1093/bioinformatics/btx413

Kobayashi, T. (2011). Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell. Mol. Life Sci.* 68, 1395–1403. doi: 10.1007/s00018-010-0613-2

Kobayashi, T., Heck, D. J., Nomura, M., and Horiuchi, T. (1998). Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev.* 12, 3821–3830. doi: 10.1101/gad.12.24.3821

Labudova, D., Hon, J., and Lexa, M. (2020). pqsfinder web: G-quadruplex prediction using optimized pqsfinder algorithm. *Bioinformatics* 36, 2584–2586. doi: 10.1093/bioinformatics/btz928

Le Guen, T., Jullien, L., Touzot, F., Schertzer, M., Gaillard, L., Perderiset, M., et al. (2013). Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. *Hum. Mol. Genet.* 22, 3239–3249. doi: 10.1093/hmg/ddt178

Matyasek, R., Kuderova, A., Kutilkova, E., Kucera, M., and Kovarik, A. (2019). Intragenomic heterogeneity of intergenic ribosomal DNA spacers in *Cucurbita moschata* is determined by DNA minisatellites with variable potential to form non-canonical DNA conformations. *DNA Res.* 26, 273–286. doi: 10.1093/dnares/dsz008

Mayer, C., Schmitz, K. M., Li, J. W., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol. Cell* 22, 351–361. doi: 10.1016/j.molcel.2006.03.028

Mcstay, B., and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell Dev. Biol.* 24, 131–157. doi: 10.1146/annurev.cellbio.24.110707.175259

Mestre-Fos, S., Penev, P. I., Richards, J. C., Dean, W. L., Gray, R. D., Chaires, J. B., et al. (2019b). Profusion of G-quadruplexes on both subunits of metazoan ribosomes. *PLoS One* 14:e0226177. doi: 10.1371/journal.pone.0226177

Mestre-Fos, S., Penev, P. I., Suttapitugsakul, S., Hu, M., Ito, C., Petrov, A. S., et al. (2019a). G-quadruplexes in human ribosomal RNA. *J. Mol. Biol.* 431, 1940–1955. doi: 10.1016/j.jmb.2019.03.010

Mohannath, G., Pontvianne, F., and Pikaard, C. S. (2016). Selective nucleolus organizer inactivation in *Arabidopsis* is a chromosome position-effect phenomenon. *Proc. Natl. Acad. Sci. U. S. A.* 113, 13426–13431. doi: 10.1073/pnas.1608140113

Mozgova, I., Mokros, P., and Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in *Arabidopsis thaliana*. *Plant Cell* 22, 2768–2780. doi: 10.1105/tpc.110.076182

Muchova, V., Amiard, S., Mozgova, I., Dvorackova, M., Gallego, M. E., White, C., et al. (2015). Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants. *Plant J.* 81, 198–209. doi: 10.1111/tpj.12718

Nelson, J. O., Watase, G. J., Warsinger-Pepe, N., and Yamashita, Y. M. (2019). Mechanisms of rDNA copy number maintenance. *Trends Genet.* 35, 734–742. doi: 10.1016/j.tig.2019.07.006

Pavlistova, V., Dvorackova, M., Jez, M., Mozgova, I., Mokros, P., and Fajkus, J. (2016). Phenotypic reversion in FAS mutants of *Arabidopsis thaliana* by reintroduction of FAS genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J.* 88, 411–424. doi: 10.1111/tpj.13257

Pontvianne, F., Abou-Ellail, M., Douet, J., Comella, P., Matia, I., Chandrasekhara, C., et al. (2010). Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* 6:e1001225. doi: 10.1371/journal.pgen.1001225

Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgova, I., Hassel, C., Pontes, O. M. F., et al. (2013). Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* 27, 1545–1550. doi: 10.1101/gad.221648.113

Pontvianne, F., Carpentier, M. C., Durut, N., Pavlistova, V., Jaske, K., Schorova, S., et al. (2016). Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the *A. thaliana* genome. *Cell Rep.* 16, 1574–1587. doi: 10.1016/j.celrep.2016.07.016

Preuss, S., and Pikaard, C. S. (2007). RRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. *Biochim. Biophys. Acta* 1769, 383–392. doi: 10.1016/j.bbaexp.2007.02.005

Probst, A. V., Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., et al. (2004). *Arabidopsis* histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* 16, 1021–1034. doi: 10.1105/tpc.018754

Pruitt, R. E., and Meyerowitz, E. M. (1986). Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.* 187, 169–183. doi: 10.1016/0022-2836(86)90226-3

Rohrig, S., Schropfer, S., Knoll, A., and Puchta, H. (2016). The RTR complex partner RMI2 and the DNA helicase RTEL1 are both independently involved in preserving the stability of 45S rDNA repeats in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006394. doi: 10.1371/journal.pgen.1006394

Schmickel, R. D. (1973). Quantitation of human ribosomal DNA-hybridization of human DNA with ribosomal-RNA for quantitation and fractionation. *Pediatr. Res.* 7, 5–12. doi: 10.1203/00006450-197301000-00002

Schmitz, K. M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24, 2264–2269. doi: 10.1101/gad.590910

Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., et al. (1999). Bryophyte 5S rDNA was inserted into 45S rDNA repeat units

after the divergence from higher land plants. *Plant Mol. Biol.* 41, 679–685. doi: 10.1023/a:1006398419556

Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H., and Boulton, S. J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* 149, 795–806. doi: 10.1016/j.cell.2012.03.030

Vannier, J. B., Sandhu, S., Petalcorin, M. I. R., Wu, X. L., Nabi, Z., Ding, H., et al. (2013). RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. *Science* 342, 239–242. doi: 10.1126/science.1241779

Vannier, J. B., Sarek, G., and Boulton, S. J. (2014). RTEL1: functions of a disease-associated helicase. *Trends Cell Biol.* 24, 416–425. doi: 10.1016/j.tcb.2014.01.004

Wicke, S., Costa, A., Munoz, J., and Quandt, D. (2011). Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol. Phylogenet. Evol.* 61, 321–332. doi: 10.1016/j.ympev.2011.06.023

Wu, L., and Hickson, I. D. (2003). The bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426, 870–874. doi: 10.1038/nature02253

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 March 2021
doi: 10.3389/fpls.2021.643548

Ancient Origin of Two 5S rDNA Families Dominating in the Genus *Rosa* and Their Behavior in the Canina-Type Meiosis

*Radka Vozárová[1,2], Veit Herklotz[3], Aleš Kovařík[1], Yuri O. Tynkevich[4], Roman A. Volkov[4], Christiane M. Ritz[3,5] and Jana Lunerová[1]**

[1] Department of Molecular Epigenetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia, [2] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia, [3] Department of Botany, Senckenberg Museum of Natural History Görlitz, Görlitz, Germany, [4] Department of Molecular Genetics and Biotechnology, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, [5] Chair of Biodiversity of Higher Plants, International Institute (IHI) Zittau, Technische Universität Dresden, Zittau, Germany

OPEN ACCESS

Edited by:
Susann Wicke, Humboldt University of Berlin, Germany

Reviewed by:
Hai-Qin Zhang, Sichuan Agricultural University, China
Kyong-Sook Chung, Jungwon University, South Korea
Tae-Soo Jang, Chungnam National University, South Korea

***Correspondence:**
Jana Lunerová
jana.luner@ibp.cz

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *18 December 2020*
Accepted: *15 February 2021*
Published: *08 March 2021*

Citation:
Vozárová R, Herklotz V, Kovařík A, Tynkevich YO, Volkov RA, Ritz CM and Lunerová J (2021) Ancient Origin of Two 5S rDNA Families Dominating in the Genus Rosa and Their Behavior in the Canina-Type Meiosis. Front. Plant Sci. 12:643548. doi: 10.3389/fpls.2021.643548

The genus *Rosa* comprises more than 100 woody species characterized by intensive hybridization, introgression, and an overall complex evolutionary history. Besides many diploid species (2n = 2x = 14) polyploids ranging from 3x to 10x are frequently found. Here we analyzed 5S ribosomal DNA in 19 species covering two subgenera and the major sections within subg. *Rosa*. In addition to diploids and polyploids with regular meiosis, we focused on 5x dogroses (*Rosa* sect. *Caninae*), which exhibit an asymmetric meiosis differentiating between bivalent- and univalent-forming chromosomes. Using genomic resources, we reconstructed 5S rDNA units to reveal their phylogenetic relationships. Additionally, we designed locus-specific probes derived from intergenic spacers (IGSs) and determined the position and number of 5S rDNA families on chromosomes. Two major 5S rDNA families (termed 5S_A and 5S_B, respectively) were found at variable ratios in both diploid and polyploid species including members of the early diverging subgenera, *Rosa persica* and *Rosa minutifolia*. Within subg. *Rosa* species of sect. *Rosa* amplified the 5S_A variant only, while taxa of other sections contained both variants at variable ratios. The 5S_B family was often co-localized with 35S rDNA at the nucleolar organizer regions (NOR) chromosomes, whereas the co-localization of the 5S_A family with NOR was only exceptionally observed. The allo-pentaploid dogroses showed a distinct distribution of 5S rDNA families between bivalent- and univalent-forming chromosomes. In conclusion, two divergent 5S rDNA families dominate rose genomes. Both gene families apparently arose in the early history of the genus, already 30 myrs ago, and apparently survived numerous speciation events thereafter. These observations are consistent with a relatively slow genome turnover in the *Rosa* genus.

Keywords: 5S rDNA, evolution, *Rosa*, genomics, cytogenetics, repeatome, Rosaceae

INTRODUCTION

Ribosomal RNA genes (rDNA) encoding 5S, 5.8S, 18S, and 26S ribosomal RNA are ubiquitous in plants and are organized into arrays containing hundreds to thousands of tandem units at one or more genomic loci (Hemleben et al., 1988; Nieto Feliner and Rossello, 2012; Roa and Guerra, 2012). Each unit consists of an evolutionary conserved coding region of 120 bp and a variable intergenic spacer (IGS) (Long and Dawid, 1980). The units within the 5S arrays retain a high degree of identity due to homogenizing forces referred to as concerted evolution (Dover, 1982; Eickbush and Eickbush, 2007) where unequal crossing-over and gene conversion are major forces driving the process. Regardless of the mechanism, numerous factors such as the number of arrays, their mutation rate, formation of new variants, the intensity of natural selection, or efficient population size can affect homogenization of repeats (Dover, 1982; Ohta, 1984; Nagylaki, 1990). In plant hybrids and allopolyploids, homogenization of 5S rDNA arrays may not always occur as efficiently as that of 35S rDNA (Pedrosa-Harand et al., 2006; Weiss-Schneeweiss et al., 2008; Garcia et al., 2016; Amosova et al., 2019). As a consequence, two or more variants differing in the length and nucleotide sequence may simultaneously exist per genome (Cronn et al., 1996; Volkov et al., 2001, 2017; Fulnecek et al., 2002; Pastova et al., 2019; Benson et al., 2020).

The genus *Rosa* L. (Rosaceae) comprises about 150 species widely distributed across the northern hemisphere. Taxonomy is considered to be challenging because frequent polyploidy in app. 50% of the species (Yokoya et al., 2000; Roberts et al., 2009) and recurrent hybridization events may blur species boundaries (Ritz et al., 2005; Joly and Bruneau, 2006; Koopman et al., 2008). The existence of multiple cytotypes and variable degree of retention of progenitor alleles leading to incomplete lineage sorting complicating taxonomic classifications. In addition, species identification is generally hampered because most species are characterized rather by combinations of morphological characters than by single discriminating traits (Christ, 1873; Wissemann, 2003). Moreover, roses are one of oldest ornamentals (Wang, 2007) and their complex history of cultivation and breeding may generate another uncertainty in phylogenetic studies. Several attempts have been made to reconstruct the phylogeny of the genus (Millan et al., 1996; Matsumoto et al., 1998; Wissemann and Ritz, 2005; Bruneau et al., 2007; Zhang et al., 2013; Fougere-Danezan et al., 2015; Liu et al., 2015). Currently, the system comprises four subgenera: *Hulthemia* (one species), *Hesperhodos* (two species), *Platyrhodon* (one species), and *Rosa*, the latter consisting of 10 sections and comprising the vast majority of species (Wissemann, 2003). The most recent phylogenies detected *Rosa persica* (subg. *Hulthemia*) and *Rosa minutifolia* (subg. *Hesperhodos*) as early diverging lineages, and a major split of the genus into two large clades: the *Synstylae* and allies clade consisting of sect. *Synstylae*, *Indicae, Caninae, Bracteatae, Laevigatae*, and *Gallicanae* and the *Rosa* and allies clade comprising sect. *Rosa* [=*Cinnamomeae*] and *Pimpinellifoliae* (Wissemann and Ritz, 2005; Bruneau et al., 2007; Fougere-Danezan et al., 2015; Zhu et al., 2015; Debray et al., 2019).

The exclusively polyploid members of a large section *Caninae* (DC.) Ser. (dogroses), originated by multiple hybridization events (Ritz et al., 2005; Herklotz et al., 2018) represent a remarkable evolutionary lineage because they exhibit a peculiar unbalanced mode of sexual reproduction also known as Canina meiosis (Täckholm, 1920; Blackburn and Harrison, 1921). Canina meiosis results in a strongly matroclinal inheritance of genetic information since two pairing genomes form bivalents, while the remaining genomes remain unpaired as univalents and are transmitted by the female germ line only. Thus, at least hemisexual reproduction is ensured in the mostly pentaploid (2n = 5x = 35) species but tetraploids, hexaploids, and heptaploids also occur and their meiosis just differs by the number univalents (Wissemann, 2003; Roberts et al., 2009; Pachl, 2011). Amazingly, in plastid phylogenies, sect. *Caninae* appeared to be polyphyletic since species with fragrant glands (subsect. *Rubigineae* and *Vestitae*) were separated from the remaining species (subsect. *Caninae*) by *Rosa gallica* and *Rosa arvensis* which perform regular meiosis (Wissemann and Ritz, 2005; Fougere-Danezan et al., 2015). Thus, Canina meiosis has been probably evolved twice, which is supported by fluorescence *in situ* hybridization (FISH) analyses of meiotic chromosomes (Herklotz et al., 2018; Lunerova et al., 2020).

Ribosomal DNA loci have been studied in several diploid and polyploid species of *Rosa* so far. Ma et al. (1997) found one 35S rDNA locus per genome, located terminally on the short arms of small submetacentric chromosomes in five diploid species and one tetraploid cultivar of *Rosa*. Fernandez-Romero et al. (2001) found one 35S rDNA locus per genome at terminal locations on submetacentric chromosomes in five diploid species. These studies indicate the presence of a single nucleolar organizer regions (NOR) chromosome per haploid set of x = 7. In pentaploid dogroses, four to five 35S loci were reported implying the occasional loss of one locus (Lim et al., 2005; Herklotz et al., 2018). The 5S locus has been less commonly studied, while there is evidence for more than one 5S locus per haploid set. Two loci were found in the diploid *Rosa lucieae* [=*Rosa wichurana*] (Kirov et al., 2016), and some pentaploid dogroses may contain more than five sites (Lim et al., 2005; Herklotz et al., 2018) indicating a variable number of 5S loci per haploid set. The analysis of 5S rDNA clones from diploid *Rosa rugosa* revealed a conserved bipartite polymerase III promoter and non-coding IGS region (Tynkevich and Volkov, 2014b) evidencing that organization at the unit level is similar to most other plants. Analysis of 5S rDNA clones from four distantly related diploid species of *Rosa* (*R. nitida*, *R. rugosa*, *R. sericea*, and *R. lucieae*) showed a high level of intragenomic homogeneity. In contrast, the level of IGS similarity between *R. lucieae* and three other diploid species appeared to be unusually low (less than 58%) arguing for interspecies diversity in *Rosa* (Tynkevich and Volkov, 2014a,b).

In this study, based on genomic and cytogenetic approaches, we aim to map the evolutionary history of 5S rDNA loci across the genus *Rosa.* Based on available phylogenies of the genus, we selected 11 diploid and eight polyploid species representing the genus' diversity (**Table 1**). Bioinformatic methods were used to determine the abundance and homogeneity of 5S rDNA in the genomes. Using locus-specific probes derived from

TABLE 1 | List of Rosaceae species used in this study, ploidy, source, and read archive accessions and the analyses employed.

Taxonomic rank[a]	Species/accession ID[b]	Ploidy	Methods applied[c]	Sequence read archive[d]/clone
Subgenus *Hesperhodos* COCKERELL	*R. minutifolia* ENGELM.	2x	Q, P, R	SRR7077023
Subgenus *Hulthemia* (DUMORT.) FOCKE	*R. persica* JUSS. [=*R. berberifolia* PALL.]	2x	Q, F, P, R	SRR7077021
Subgenus *Rosa*				
Sect. *Caninae* (DC). SER. subsect. *Caninae*	*R. canina* L. (CZ)	5x	Q, F, P, R	SRR8265808
	R. canina L. (DE-S27b)	5x	Q, P, R	ERR1662939
	R. corymbifera BORKH. (DE_2)	5x	Q, F, P, R	SRR8265810
	R. dumalis BECHST. (DE_34)	5x	Q, P, R	ERR1662941
Subsect. *Rubigineae* CHRIST	*R. inodora* FR. (DE_12)	5x	Q, F, P, R	ERR1662940
	R. rubiginosa L.	5x	Q, F, P, R	SRR10402274
Subsect. *Vestitae* CHRIST	*R. sherardii* DAVIES	5x	Q, P, R	SRR10402273
Sect. *Gallicanae* DC.	*R. gallica* L.	4x	Q, P, R	SRR6175524
Sect. *Indicae* THORY	*R. chinensis* JACQ.	2x	Q, P, R	SRR7077020
Sect. *Laevigatae* THORY	*R. laevigata* MICHX.	2x	Q, P, R	SRR7077018
Sect. *Synstylae* DC.	*R. arvensis* HUDS. (DE_8)	2x	Q, F, P, R	SRR8265809
	R. lucieae CRÉP. [=*R. wichurana* Crép.]	2x	Q, P, R	SRR6175519
	R. multiflora THUNB.	2x	Q, F, P, R	DRR059736
Sect. *Pimpinellifoliae* DC.	*R. spinosissima* L.	4x	Q, F, P, R	SRR8422951
Sect. *Rosa* [=*Cinnamomeae* DC.]	*R. majalis* HERRM. (DE_4)	2x	Q, F, P, R	SRR6175513
	R. nitida WILLD.	2x	F	n. d.
	R. pendulina L.	2x	Q, P, R	SRR6175522
	R. rugosa THUNB.	2x	Q, F, P, R	SRR6175514
Outgroups				
Cliffortia curvifolia WEIM.		2x	P	EU931716
Acaena latebrosa (AITON) W.T. AITON		2x	P	EU931698
Geum urbanum L.		2x	P	ERR2187925

[a] *Taxonomy within Rosa is according to Wissemann (2003). Species names are according to Plants of the world online (http://www.plantsoftheworldonline.org/); synonyms used in previous phylogenies of the genus are given in brackets.*
[b] *CZ, Czech Republic; DE, Germany.*
[c] *Q—5S families quantification by reads mapping, P—phylogeny tree construction, R—cluster analysis by RepeatExplorer, and F—fluorescent in situ hybridization.*
[d] *Sequence read archives (ENA/NCBI) submitted as parts of original projects (Herklotz et al., 2018; Nakamura et al., 2018; Saint-Oyant et al., 2018; Lunerova et al., 2020).*

5S IGSs, we identified the two major 5S rDNA loci on the chromosomes by FISH.

MATERIALS AND METHODS

Plant Material

Material of polyploid dogroses was sampled in wild populations in Germany and the Czechia (**Supplementary Table S1**). Diploid species and tetraploid species with regular meiosis were obtained from various Botanical Gardens (**Supplementary Table S1**). In addition, we retrieved sequence information from published work stored in the ENA database for bioinformatics analyses (**Supplementary Table S1**).

Isolation and Cloning of 5S rDNA Sequences From *Rosa canina*

Total genomic DNA of *Rosa canina* was extracted from fresh leaves applying the standard protocol (Rogers and Bendich, 1985). The 5S rDNA repeats were amplified using the primers pr5S-14 and pr5S-15 (Tynkevich and Volkov, 2014b) with the 5'-extensions containing restriction endonuclease *Not*I recognition site. The PCR products were treated by *Not*I, ligated into the *Eco*52I recognition site of the pLitmus 38 plasmid, and used for transformation of *Escherichia coli* XL_blue line by electroporation method. Selected recombinant clones were sequenced using a BigDye Terminator Cycle Sequencing Kit (Thermofisher Scientific, United States). Clones containing inserts of A and B variants of 5S rDNA were identified by sequence analysis. The sequences were submitted to GenBank under the accession numbers MW349696 and MW349697.

The inserts of cloned 5S rDNA sequences contained genic regions and IGS. In order to increase the specificity of probe hybridization, we amplified the IGS sequences using specific primers annealing to 5S_A and 5S_B variants. The oligonucleotide primers' sequences for the 5S_A IGS were as follows: A_for: 5′-CCTCTTTTTTCTGTTTCGGT-3′; A_rev: 5′-ATAAACTCCATTCGCTCAG-3′. Primers for the 5S_B variants were: B_for: 5′-ACCCCTCTTTTTGCCTTT-3′; B_rev: 5′-GCTTCGTCTCACTCCTCT-3'. The 25 μl PCR reaction contained 0.1 ng of plasmid DNA as the template, 4 pmol of each primer, 2.4 nmol of each dNTP, and 0.4 units of Kapa *Taq* DNA polymerase I (Kapa Biosystems). Cycling conditions were as follows: initial denaturation step (94°C, 180 s); 35 cycles (94°C, 20 s; 57°C, 30 s; and 72°C, 30 s). The length of amplified products was 373 nt for the A variant and 394 nt for the B

variant. Purified PCR products were labeled by fluorescent dyes (as below) and used in FISH.

Identification of 5S rDNA Sequences in High-Throughput Reads

For bioinformatic analyses, the whole genomic sequencing data for 19 *Rosa* accessions were used (**Supplementary Table S2**). The genome proportion of 5S rDNA families was determined using the total Illumina reads trimmed for quality (Phred score $\geq$ 30 over $\geq$ 95% read length). Trimmed reads (typically > 7 million) were mapped to corresponding 5S_A and 5S_B reference sequences (IGS subregion between the primers, **Figure 1**) using the following parameters: insertion and deletion costs_3, lengths fraction_0.5, similarity fraction_0.8, and deletion cost_2 (Qiagen, Germany). The distribution of SNPs across the 5S rDNA sequences was recorded when the distribution exceeded a threshold of at least 20 identical SNPs over at least 200 reads that covered the variant position and occurred at $\geq$ 10% frequency. For the alignment, minimal sequence length coverage was 50% and minimum sequence similarity was 90%. For the more distantly related species, *Rosa spinosissima* and *R. persica*, similarity threshold parameter was decreased to 80% (for the 5S_B variant). The genome abundance and copy number was calculated from genome proportions according to the formula stated in (Lunerova et al., 2020).

Generation of Consensus Sequence and Phylogenetic Analyses

For phylogenetic reconstructions, the consensus 5S_A and 5S_B rDNA sequences were extracted from mapped reads using CLC genomic workbench. Additionally, we added the partial sequence of 5S ribosomal RNA genes from *Acaena latebrosa* (EU931698.1)

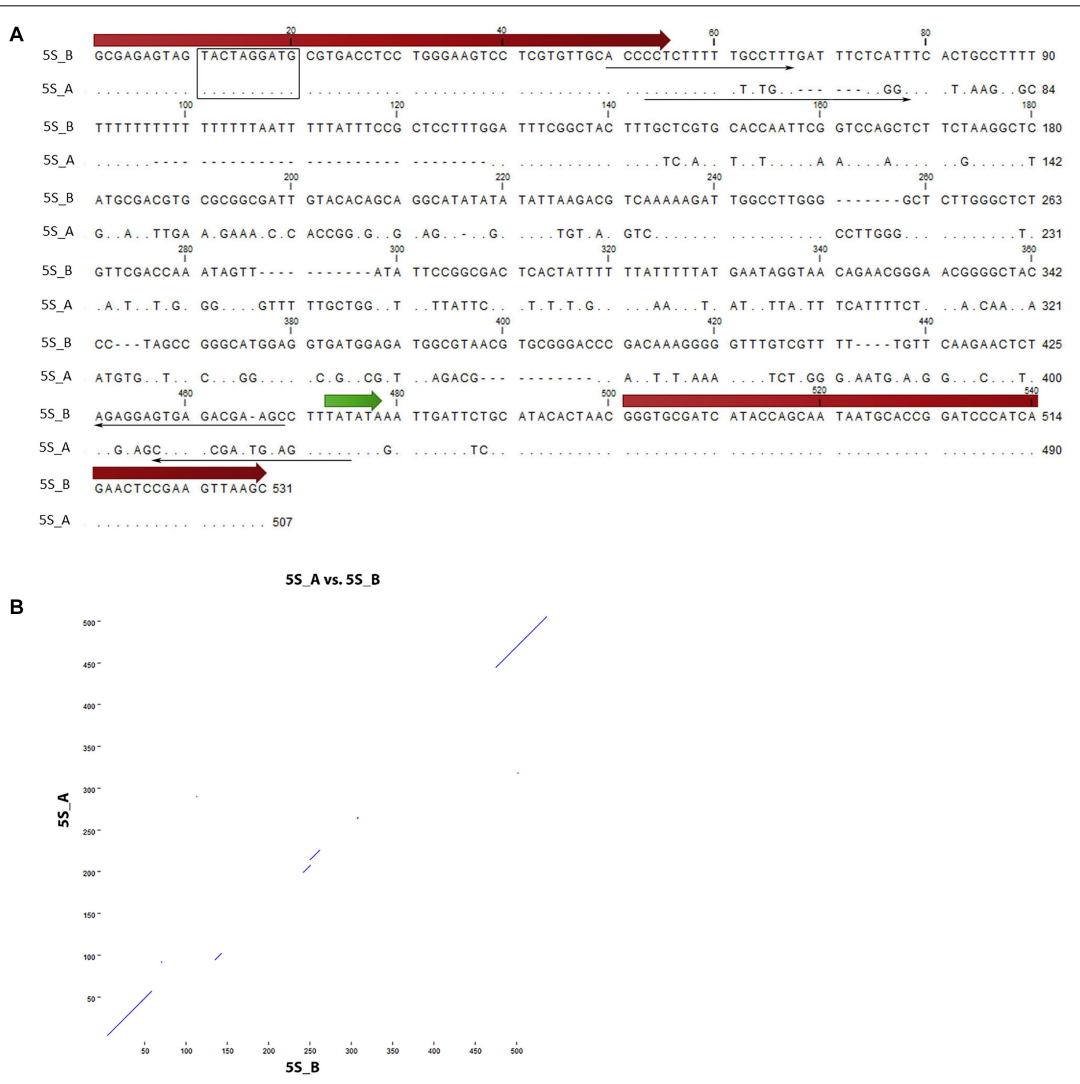


FIGURE 1 | Sequence analysis of 5S rDNA clones from *Rosa canina*. **(A)** Alignment of the 5S_A and 5S_B clones. The position of 5S rRNA coding regions (thick arrows), primers (thin arrows), and regulatory regions (TATA box—in green, Box-C—rectangle) are highlighted. **(B)** Dot plot of pairwise comparisons clones. Note that only coding regions showed significant similarity.

and *Cliffortia curvifolia* (EU931716.1). Paired 250 bp Illumina reads of *Geum urbanum* (SRA accession ERR2187925) were mapped in a first round to the *C. curvifolia* (EU931716.1) sequence. Mapping was done with Bowtie2 (Langmead and Salzberg, 2012) implemented in Geneious® 10.0.9[1] with the lowest sensitivity pre-set. This resulted in two reads out of 22.8 million hitting to a 39 bp conserved region. The two reads were aligned and served in a second round with same parameters as mapping scaffold of 334 bp length. In the second mapping, 50 reads out of 22.8 million were assembled and its consensus was added to the alignment including all rose sequences and the two species from the *Acaena* clade. Alignments were done with MAFFT v7.450 algorithm (Katoh and Standley, 2013) implemented in Geneious using default parameters. Model test implemented in MEGA X v. 10.1.8 (Kumar et al., 2018) revealed the Tamura–parameter substitution model with invariant sites as most appropriate for the data based on Akaike information criterion (Tamura, 1992). Based on this model, we computed a maximum-likelihood tree in MEGA X whose branch support was evaluated by 1000 bootstrap replicates. Rooted with *G. urbanum*, this tree was used for the dating approach conducted with MEGA X. Therefore, we used two calibration points along the tree taken from the respective fossils given in Xiang et al. (2017) *Rosa germerensis*, 48.6 Mya (Edelman, 1975) and *Acaena* sp., 37.2 Mya (Zetter et al., 1999). A timetree inferred by applying the RelTime method (Tamura et al., 2018) was computed using two fixed calibration constraints. All positions containing gaps and missing data were eliminated (complete deletion option). A neighbor joining tree was constructed with the Seaview software (Gouy et al., 2010).

Clustering Analysis of 5S rDNA

The fastq reads were initially filtered for quality and trimmed to uniform length using the pre-processing and QC tools in RepeatExplorer2 (Novak et al., 2013). Read length ranged between 100 and 150 bp, depending on sequencing library and the Illumina sequencing platform. After the fastq > fasta conversion and trimming to uniform length, reads were analyzed with the RepeatExplorer2 clustering program using default parameters. We used 1 million paired-end reads, or 1 million single-end reads as inputs for RepeatExplorer2 clustering. This bioinformatic pipeline runs a graph-based clustering algorithm (Novak et al., 2013) that assembles groups of frequently overlapping reads into clusters of reads, representing a repetitive element or part of a repetitive element with a higher order genome structure. The similarity and structure-based repeat identification tools in RepeatExplorer2 aid in identification of the repeats. RepeatExplorer2 uses a BLAST threshold of 90% similarity across 55% of the read to assign reads to clusters (minimum overlap = 55, cluster threshold = 0.01%, minimum overlap for assembly = 40), and the clusters are identified based on the principle of maximum modularity. We also used the SeqGrapheR (Novak et al., 2010) software in virtual space of Ubuntu 18.04 to visualize the specific reads corresponding to the 5S_A and 5S_B variants.

[1]https://www.geneious.com

Slide Preparation and FISH

For slide preparations, we used young anthers from flower buds of about 0.5 cm in length, which were harvested during spring 2019. Male meiosis was studied at prophase I (diplotene/diakinesis) where the bivalents and univalents could be easily distinguished from each other. Fresh flower buds were fixed using Carnoy solution (acetone:acetic acid, 2:1 or ethanol:acetic acid, 3:1 in some cases), and stored in 70% ethanol at −20°C. Before slide preparation, anthers were pre-treated by 0.5% PVP and 2% Triton-X100 (Sigma–Aldrich, United States) for 2–5 min followed by enzyme digestion overnight at 10°C in 1% cellulase, 0.2% pectolyase Y23, 0.5% hemi-cellulase, and 0.5%, macerozym R10 (Sigma–Aldrich, United States; Duchefa Biochemie, Netherlands) dissolved in citric buffer (0.04 M citric acid and 0.06 M sodium citrate). FISH followed the procedures described in Herklotz et al. (2018). Anthers were separated and squashed on slides in a drop of 70% acetic acid and fixed in liquid nitrogen.

For FISH, we used two probes derived from the 5S_A and 5S_B clones of the IGS region, respectively, and in addition, an 18S rDNA probe that was a 1.7-kb fragment of the 18S rRNA gene of *Solanum lycopersicum* (GenBank # X51576.1). The 5S rDNA genic region originated from *Artemisia tridentata* S4 clone, GenBank # JX101915.1. The probes were labeled by nick translation using Spectrum green dUTPs (Abbott, United States) for 5S_A rDNA, and Cy3-dUTPs (Roche, Switzerland) for 5S_B and 18S rDNA; 5S rDNA was labeled by Atto647N (Jena Bioscience, Germany). Slide preparation and hybridization followed standard protocols (Schwarzacher and Heslop-Harrison, 2000). Chromosomes were counterstained with 1 μg ml^{-1} DAPI (4', 6-diamidine-2'-phenylindole dihydrochloride) diluted in mounting medium for fluorescence (Vectashield, Vector Laboratories, United Kingdom). The slides were scanned using epifluorescence microscopes (Olympus Provis AX70, with cold cube camera, Metasystems, Germany). Imaging software was ISIS (MetaSystems, Germany), and images were optimized for contrast and brightness with Adobe Photoshop CS6 and PS2020.

RESULTS

Cloning and Sequencing of 5S rDNA Variants

Two 5S rDNA clones (5S_A and 5S_B) were isolated from *R. canina* IGS. Sequence analysis revealed some conserved regulatory elements: Box-C within the coding region, the TATA box at –29 (both clones), and T-rich terminators downstream of the coding sequences (**Figure 1A**). Box-A could not be unambiguously determined due to primer overlap. The 5S_B clone had a long (20 nt) T-tract which appears to be missing or was much shorter in clone 5S_A. By analogy with other 5S rRNA transcripts, the putative transcription started at the first G within the GGG motif following the C at −1 (Tynkevich and Volkov, 2014b). Intragenomic homogeneity was high, and no significant SNPs were revealed in mapping experiments (not shown). Pairwise alignment revealed conserved coding

regions, while most of the IGS was dissimilar between both sequences (**Figure 1B**). We took advantage of considerable sequence divergence between both clones and amplified the locus-specific IGS subregions from plasmids. The resulting 373 bp (5S_A family) and 394 bp (5S_B family) PCR products were subsequently used in FISH.

Representation of Individual 5S rDNA Variants in *Rosa* Genomes

To determine the abundance of individual 5S rRNA gene families in *Rosa* genomes, we used available genomic resources (**Table 1**). The genome proportion of the 5S_A family was in average twice of that of the 5S_B family (**Supplementary Tables S2, S3**). The copy number per somatic cell (2C) ranged from 80–8000 (5S_A family) and 0–2400 (5S_B family). Both families appeared to be equally homogeneous containing a relatively low number of SNPs consistent with our previous findings obtained by comparison of individual 5S rDNA clones (Tynkevich and Volkov, 2014a,b). The contribution of each family to total 5S rDNA was expressed for each species by pie charts and visualized in a phylogenetic context (**Figure 2**). The diploid species from sect. *Synstylae*, all polyploid species, and *R. persica* (subg. *Hulthemia*) carried both families. *Rosa laevigata* contained the 5S_A family in low copy (c. 80 copies/2C), while its 5S rDNA was dominated by the 5S_B family (980 copies/2C) (**Supplementary Table S3**). Species from sect. *Rosa* and *R. minutifolia* (subg. *Hesperhodos*) carried the 5S_A rDNA family only. Blast searches failed to reveal significant hits of 5S_A and 5S_B sequences in genomic reads from *Prunus*, *Rubus*, *Fragaria*, *Cliffortia*, *Acaena*, and *Sanguisorba* (all Rosaceae) even at relaxed (e = 0.1) stringencies (not shown).

Repeat Explorer Analysis of 5S rDNA Families in *Rosa* Genomes

Cloning experiments cannot address the question about the distribution of gene families in the genome. Thus, in order to determine the number and genomic representation of individual 5S rRNA gene families, we applied clustering analysis (**Figure 3**). The cluster graph shapes provide information about the number and type of 5S gene families revealing potential hybridization and introgression (Garcia et al., 2020). It visualizes divergent IGS families as loops emanating from the bridge region which contains reads derived from a conserved coding region. Any loop can be considered as a separate gene family. The subregions in the graph can be annotated based on the read alignment against the 5S_A and 5S_B reference clones. Visual exploration indicates that there is no other 5S family amplified except of 5S_A and 5S_B types. The cluster graphs obtained from different *Rosa* genomes were categorized based on their structure into three groups (**Figure 3**). Group 1 comprised a single species *R. laevigata* (sect. *Laevigatae*) with predominant 5S_B type family representation. Group 2 comprised *R. persica* (subg. *Hulthemia*) and the majority of diploid species (sect. *Synstylae* and *Indicae*) showing a typical two-loop structure representing relatively balanced ratios of both families. Group 3 contained the diploid species *R. minutifolia* (subg. *Hesperhodos*) and species of sect. *Rosa* harboring a single 5S rDNA family (A). All polyploid species (both, those with regular meiosis and those with Canina meiosis) showed a Group 2 profile indicating the presence of both A and B families (**Supplementary Figure S1**). In sum, quantitative relationships between both 5S rDNA families were confirmed. Moreover, cluster analysis showed that the maximum number of 5S rDNA families in the *Rosa* genomes is always two, irrespective of the ploidy level.

Phylogenetic Relationships Between 5S rDNA Families

To determine the phylogenetic relationships between 5S rDNA families, we computed phylogenies based on aligned 5S rDNA consensus sequences (obtained from mapping experiments. Both the maximum-likelihood (**Figure 4** and **Supplementary Figure S2**) and neighbor joining (**Supplementary Figure S3**) trees separate the A and B 5S rDNA families clearly into two well-supported clades (A and B). Both, clades A and B contained diploid and polyploid species. Except of sect. *Rosa*, whose members clustered exclusively within the A clade, members of other sections, including *Synstylae*, *Indicae*, *Laevigatae*, *Pimpinellifoliae*, and *Caninae*, partitioned their 5S rDNA between both clades. Sequences from *R. persica* (subg. *Hulthemia*) were consistently positioned on early diverging nodes at both subclades. The major 5S rDNA family of *R. laevigata* (sect. *Laevigatae*) branched off at a rather basal position in clade B. The 5S_B family of 4x *R. spinosissima* positioned as sister to *R. persica*. Five 5S rDNAs of 5x species from sect. *Caninae* clustered together in both clades with negligible resolution between species (**Supplementary Figures S2, S4**). Out of the diploids, 5S sequences of *R. arvensis* (sect. *Synstylae*, B clade) and *R. pendulina* in (sect. *Rosa*, A clade) were most closely related to those of the respective *Caninae* branches. To gauge the length of time these 5S rDNA variants have existed in the *Rosa* genomes, we used two calibration points (48.6 myrs for *R. germerensis* and 37.2 myrs for *Acaena* sp.). We estimated that a common ancestor of both A and B families lived about 32 myrs ago (**Figure 4**) relatively long before separation of modern clades.

FISH Analysis of 5S rDNA Variants on Chromosomes

Fluorescence *in situ* hybridization was conducted to visualize the position and number of 5S rDNA variants on chromosomes in several diploid and polyploid species. The diploids included representatives of subg. *Rosa* sect. *Synstylae* (*R. arvensis* and *R. multiflora*), sect. *Rosa* (*R. rugosa, R. majalis,* and *R. nitida*), and subg. *Hulthemia* (*R. persica*). Meiotic chromosomes from anthers (**Figure 5A**) were hybridized with rDNA probes derived from the 5S rDNA genic (**Figure 5B**, shown in white), 5S_B (red), and 5S_A (green) IGS subregions (**Figure 5C**). Additionally, the same chromosome spreads were re-hybridized with the 18S rDNA probe (shown in cyan, **Figure 5D**). In *R. arvensis* and *R. multiflora*, each 5S_A and 5S_B probe hybridized to one bivalent (one pair of chromosomes). The 5S_B probe was always co-localized on a chromosome bearing also the 18S rDNA signal. In *R. rugosa* and *R. majalis*, the 5S_A probe hybridized to a single bivalent, while we did not detect any hybridization

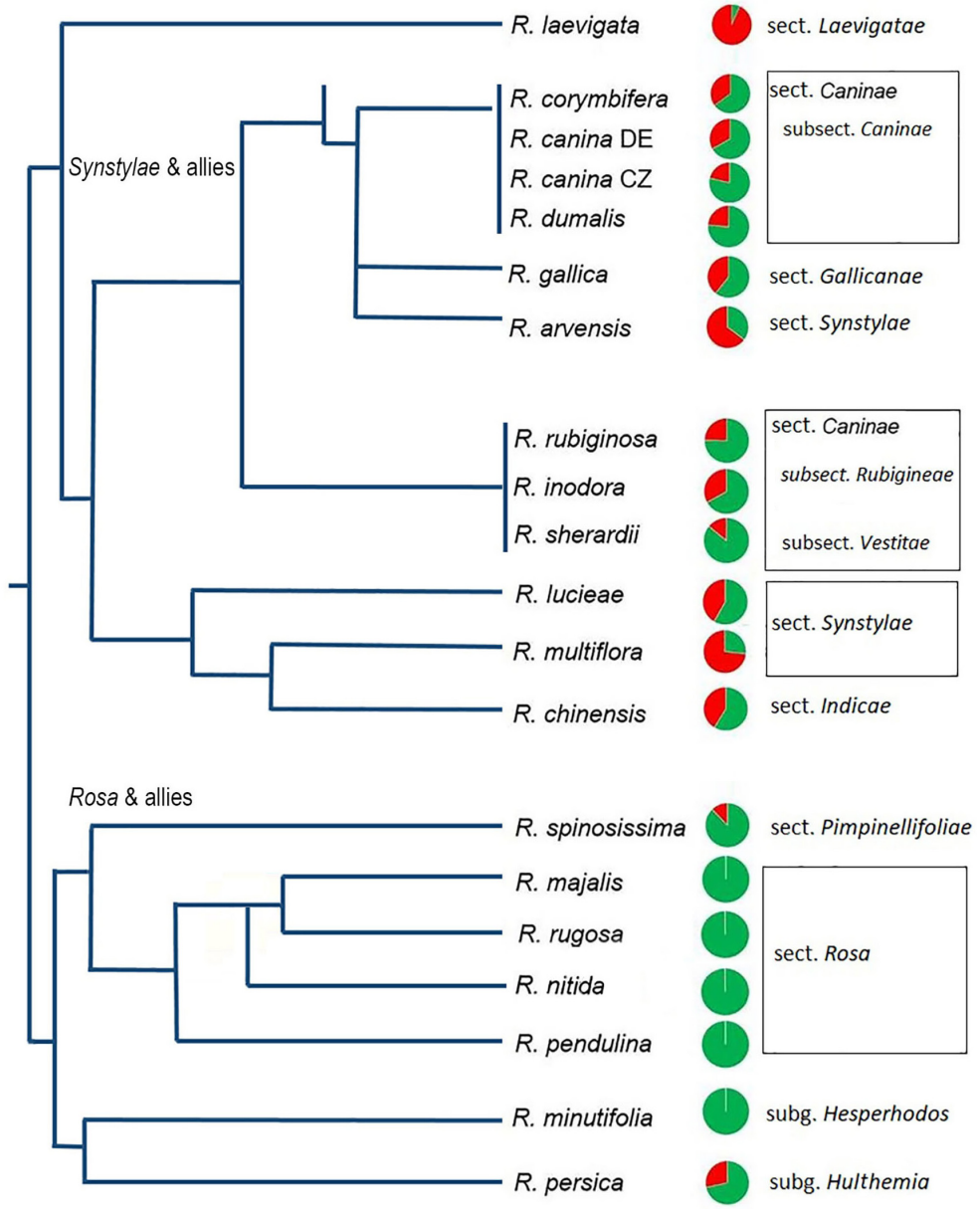


FIGURE 2 | Quantitative relationships between 5S_A and 5S_B rDNA families in *Rosa* genomes. Genomic proportions of families calculated from high-throughput reads are shown as pie charts next to the species names. Green—5S_A family, red—5S_B family. Data are in **Supplementary Table S3**; data for *R. nitida* are taken from FISH (**Supplementary Figure S5**). A simplified phylogenetic tree is redrawn according to published phylogenies based on plastid sequences (Fougere-Danezan et al., 2015).

signals with the 5S_B probe in accordance to genomic analyses (**Figures 2**, **3**). In both species, the 18S and 5S rDNA loci were separate. However, in *R. nitida*, the 5S_A probe hybridized to one pair of chromosomes (mitotic metaphase from root tips, **Supplementary Figure S5**) which carried 18S rDNA signal (NOR). In *R. persica*, both variant-specific 5S rDNA probes hybridized to a single bivalent each, and these bivalents carried also 18S rDNA sites. The number and position of rDNA loci on chromosomes are summarized in **Supplementary Table S4** and are diagrammatically depicted by ideograms (**Figure 6**).

We further analyzed meiotic (**Figure 7**) and mitotic (**Supplementary Figure S5**) chromosomes in polyploid species. As expected meiotic chromosomes of four 5x dogrose species (sect. *Caninae*) were represented by seven bivalents (pairs of chromosomes) and 21 univalents (**Figure 7**). In *R. canina* and *R. corymbifera* (subsect. *Caninae*), the 5S_A probe hybridized to one bivalent and three sites on univalent chromosomes. The 5S_B probe hybridized to one bivalent carrying the 18S (NOR) signal and two sites on univalents. The 5S_A and 5S_B signals occurred on different chromosomes except of one univalent

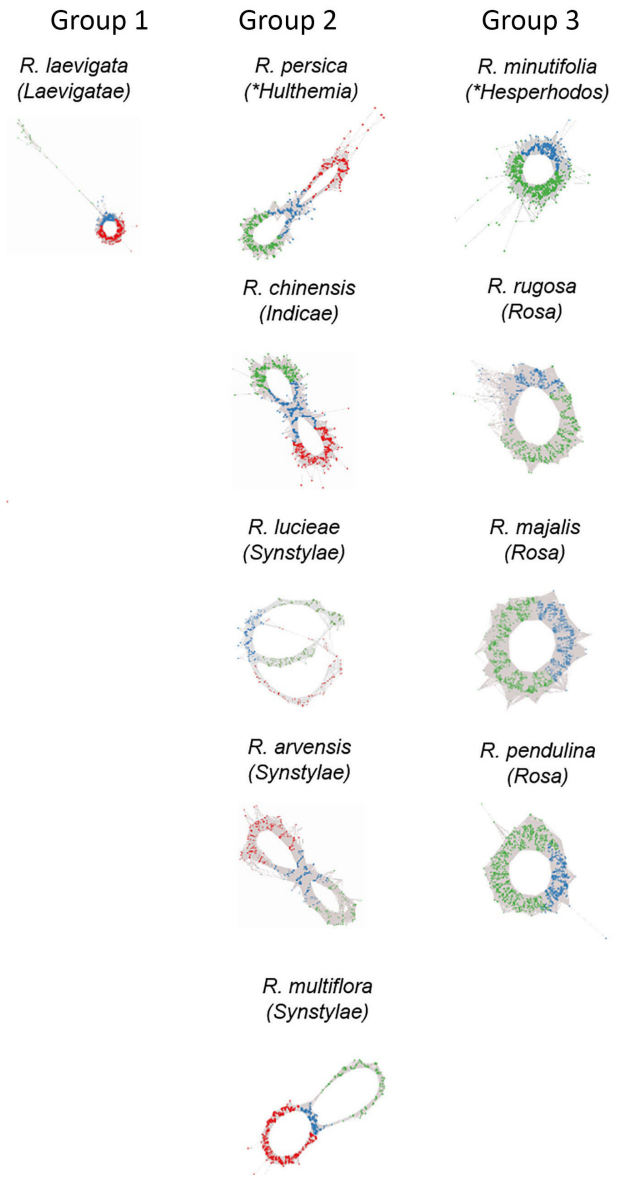


FIGURE 3 | Genomic analysis of 5S rDNA variants in diploid (2n = 2x = 14) species of *Rosa*. Projections of 5S rDNA cluster graphs: Loops represent IGS reads colored according to the 5S_A (green) and 5S_B (red) IGS variants; genic regions are in blue. Group 1 represents species with dominance of 5S_B variant; Group 2 contains diploids with balanced ratio of both variants; and Group 3 shows species with only A variant in the genome. Name of the sections within subg. *Rosa* is given in brackets; other subgenera are marked by asterisks.

chromosome in *R. canina* where both probes were co-localized. In *R. inodora* (subsect. *Rubigineae*), the 5S_A probe hybridized to one bivalent and three univalent chromosomes. The 5S_B probe hybridized to two univalent chromosomes carrying the 18S rDNA signal. *Rosa rubiginosa* (subsect. *Rubigineae*) showed a similar distribution of signals like *R. inodora* except that only one out of two 5S_B univalent chromosomes co-localized with the 18S signal. In addition, there were only two 5S_A sites on univalents. Collectively, these observations indicate that the number of rDNA sites, their chromosome position, and their meiotic behavior differ between subsections *Caninae* and *Rubigineae*. FISH on mitotic chromosomes from 4x *R. spinosissima* is shown in **Supplementary Figure S5**. In this species, the 5S_B probe hybridized to a chromosome pair which also carried the 18S rDNA signal (NOR) (**Supplementary Figure S5**). Two other 5S_B signals were colocalized (but did not overlap) with that of 5S_A on non-NOR chromosomes. 18S rDNA and 5S_A signals were localized on two separate chromosomes. Results are summarized in **Figure 6** and **Supplementary Table S4**.

DISCUSSION

To study chromosome evolution and potential hybridization events in the genus *Rosa*, we analyzed the structure and organization of 5S rDNA in several diploid and polyploid species. We found that the genus is dominated by essentially two 5S rDNA families which markedly differ in IGSs and date back to the genus' base.

Ancient Origin of 5S rDNA Variants in the Genus *Rosa*

The IGSs of rRNA genes are rapidly evolving sequences, and it is common to find variation even between closely related species. It was therefore striking to observe that the genus *Rosa* is dominated essentially by only two 5S rDNA families and that no other family was amplified in any of the species analyzed here. Both families occupy different chromosome loci: the 5S_B family was always co-localized with NOR (35S rDNA), while the 5S_A family was mostly but not exclusively (see below) separate (**Figure 6**). Moreover, *R. persica* (subg. *Hulthemia*) amplified both families at similar ratio in its genome (**Figures 2**, **3**). In contrast to all other diploid species (Ma et al., 1997), *R. persica* is also exceptional in possessing two NORs instead of one per haploid chromosome set. Since *R. persica* was consistently identified as the earliest divergent lineage in most phylogenies (Fougere-Danezan et al., 2015; Debray et al., 2019), we presume that the configuration with NORs co-localizing with distinct 5S rDNA families (**Figure 6**) is an ancient condition, while the NOR chromosome without 5S rDNA locus is derived. This assumption is supported by the following observations: First, the *Rosa* and allies clade contained the 5S_A family which was co-localized with 18S rDNA locus in *R. nitida* but not in *R. majalis* and *R. rugosa*. Second, all members of *Synstylae, Indicae,* and *Pimpinellifoliae* contained two 5S rDNA families albeit at differing ratios. For example, *R. multiflora* (sect. *Synstylae*) had prevalent 5S_B family, while the 5S_A family dominated in *R. chinensis*, a member of the closely related sect. *Indicae*. Similarly, in *R. lucieae* ([*R. wichurana*]; sect. *Synstylae*), both families are likely to be represented by two loci out of which one co-localized with 18S rDNA on the same chromosome (Kirov et al., 2016). Third, *R. laevigata* (sect. *Laevigatae*) was dominated by the 5S_B family (**Figures 2**, **3**). This species is sister to the remaining species

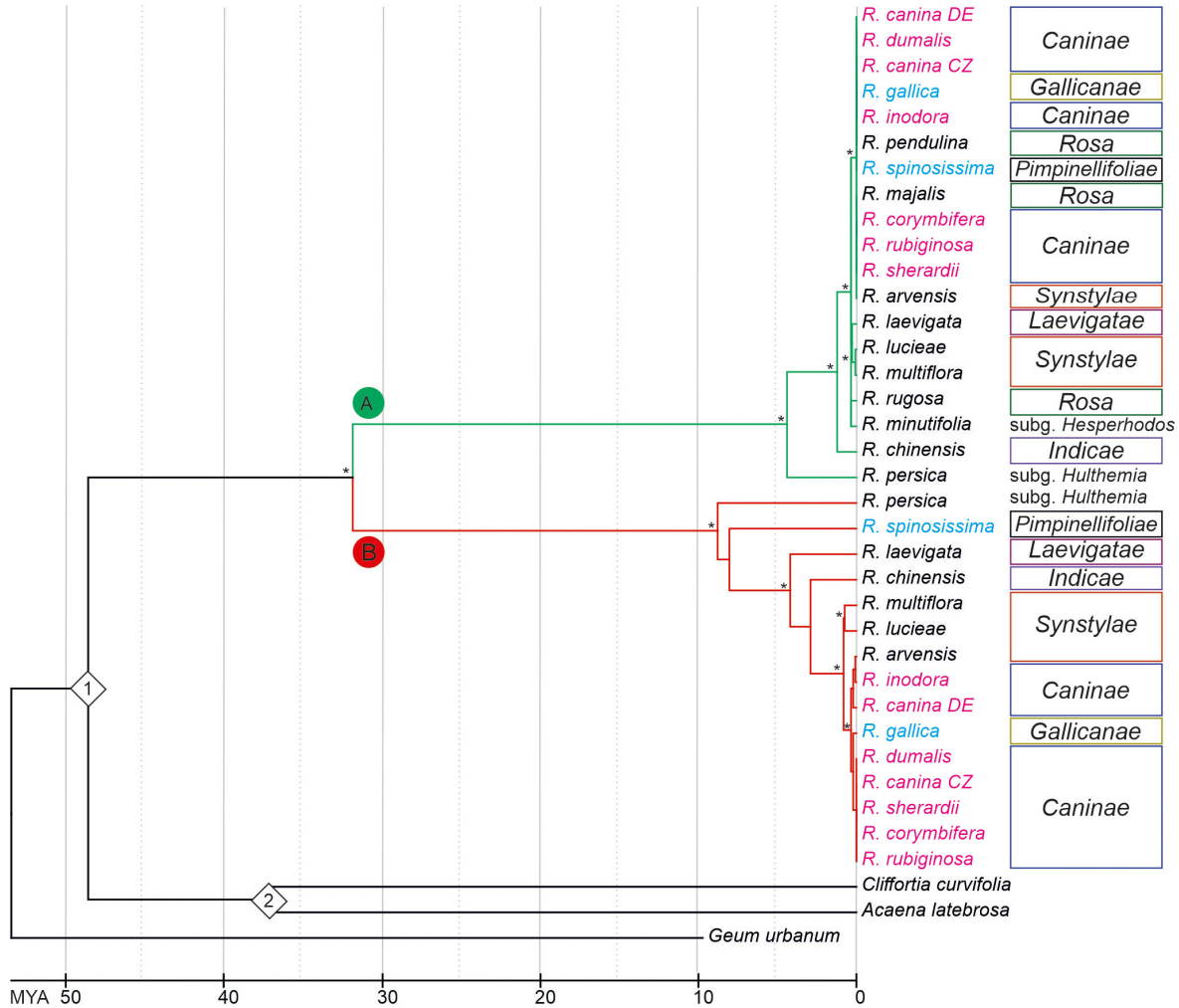


FIGURE 4 | Chronogram based on a phylogeny among *Rosa* species inferred from Maximum-Likelihood analysis of 5S rDNA intergenic spacers. The tree was rooted with *Geum urbanum*. Bootstrap value of ≥ 70% is indicated by an asterisk. Species names are colored according to ploidy level (2x—black, 4x—light blue, 5x—pink). Color of branches represents 5S_A (green) and 5S_B (red) variants of IGS rDNA. Calibration points (1 = *Rosa gemerensis*; 48.6 myrs and 2 = *Acaena* sp.; 37.2 myrs) are indicated by diamonds.

of the *Synstylae* and allies clade (Fougere-Danezan et al., 2015; Debray et al., 2019).

Of note, *R. nitida* differed from *R. majalis* and *R. rugosa* in having the 5S_A family co-localized with NOR. Traditionally, *R. nitida* has been classified into a separate section called *Carolinae* (Crépin, 1889). However, more recent taxonomies based on molecular markers failed to support this distinction and all three species are now placed within the *Rosa* and allies clade (Wissemann and Ritz, 2005; Joly and Bruneau, 2006). Interestingly, members of sect. *Rosa* tend to have much smaller loci of the centromeric satellite repeat CANR4 compared to other species of the genus (Lunerova et al., 2020). However, *R. nitida* is exceptional in having large abundance of the CANR4 satellite (10 out 14 chromosomes carried strong FISH signals, not shown). These features suggest chromosomal rearrangements accompanying speciation events in sect. *Rosa* although the basic chromosome number (x = 7) remained unchanged.

Neither A nor B type sequences were found in 5S rDNA of the genera *Prunus*, *Rubus*, *Fragaria*, *Cliffortia*, *Acaena*, and *Sanguisorba* (all Rosaceae). These observations suggest that both 5S rDNA families have their origin in the early evolution of the genus *Rosa* because the common ancestor of both families was dated at app. 32 myrs ago. Furthermore, molecular dating of diversification within both 5S rDNA clades was dated to app. 5–9 myrs each (**Figure 4**) matching at least roughly the origin of *Synstylae* and allies and *Rosa* and allies (Fougere-Danezan et al., 2015).

The Fate of 5S rDNA in Allopolyploid Dogroses

Despite considerable interest, the composition of pentaploid (2n = 5x = 35) dogrose genome remains enigmatic (Wissemann, 1999; Nybom et al., 2004; Ritz et al., 2005; Herklotz et al., 2018;

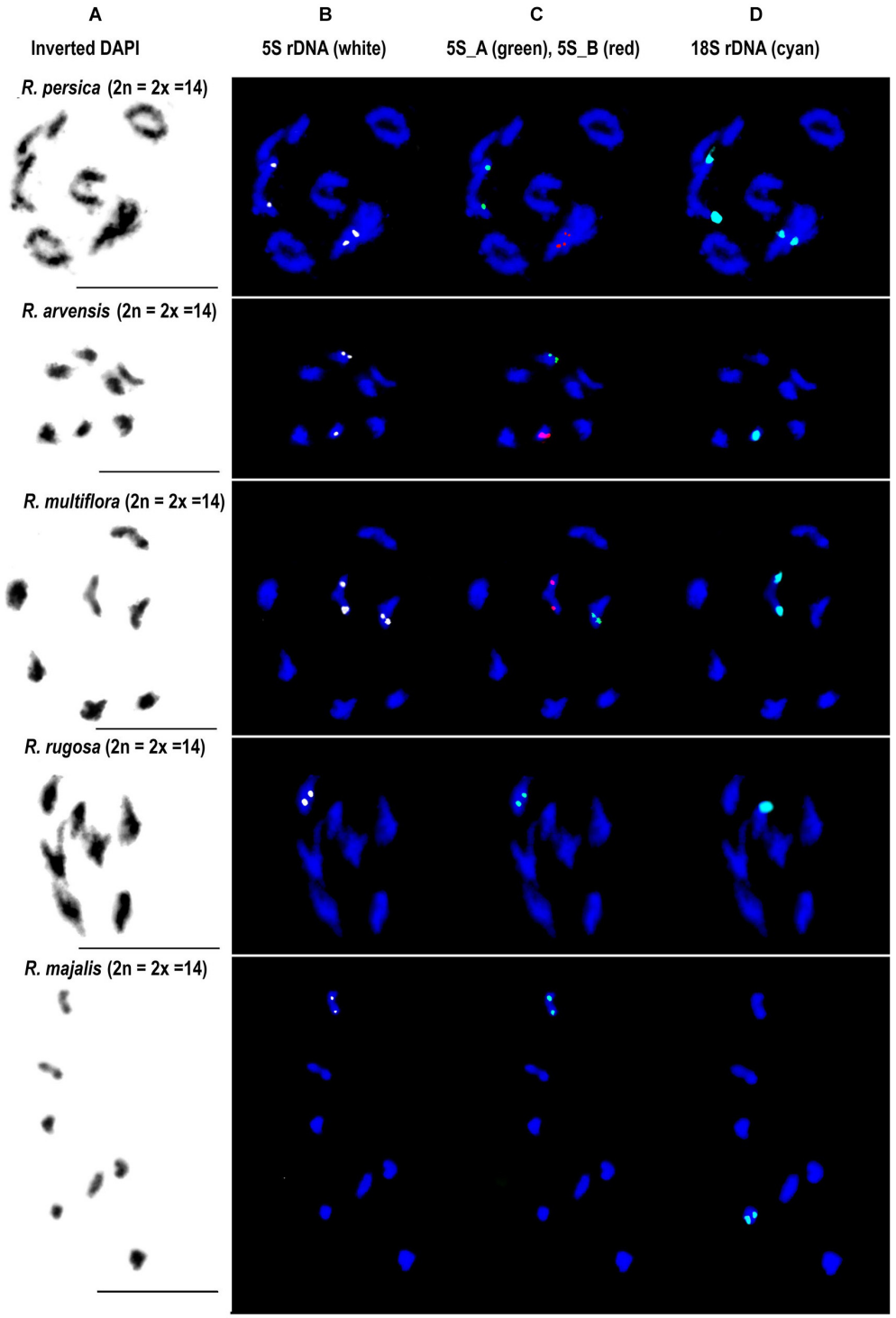


FIGURE 5 | FISH analysis of 5S rDNA in diploid species. Diakinesis of meiotic phase I shown in *R. persica* (subgenera *Hulthemia*), *R. arvensis and R. multiflora* (section *Synstylae*), *R. rugosa* and *R. majalis* (section *Rosa*). In each species, the same metaphase was hybridized with the 5S genic **(B)**, 5S intergenic **(C)**, and 18S rDNA **(D)** probes. **(A)** DAPI staining (gray scale, inverted) of chromatin. Scale bar: 10 μm.

Lunerova et al., 2020). Previous studies based on microsatellite markers revealed genetic distinction between bivalent- and univalent-forming chromosomes (Nybom et al., 2006). The analysis of 35S rDNA markers confirmed these assumptions revealing two highly divergent ITS types (named Canina type and Rubiginosa type) present at variable ratios in subsections

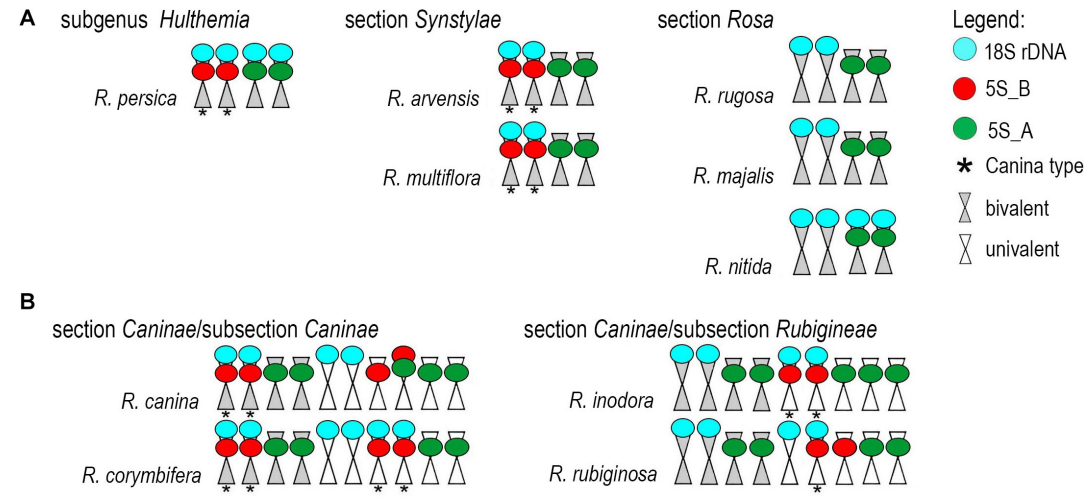


FIGURE 6 | Ideograms of chromosomes carrying the rDNA loci. **(A)** Diploid species. **(B)** Pentaploid species from section *Caninae*.

Caninae and *Rubigineae*, respectively (Ritz et al., 2005; Kovařik et al., 2008). Here we show that the *Canina* type ITS is co-localized with 5S_B rDNA locus (NOR), while the *Rubiginosa* ITS type is not. This Canina type of configuration [equivalent to marker chromosome 1 (Lim et al., 2005) or chromosome 7 (Kirov et al., 2016)] is typical for the bivalent-forming chromosomes in *R. canina* and *R. corymbifera*, both from subsection *Caninae* (**Figure 6**). In contrast, the Canina-type configuration of both rDNAs is found on univalent chromosomes in *R. rubiginosa* and *R. inodora* (subsection *Rubigineae*). The bivalent-forming chromosomes in this subsection carry the 5S_A family on a non-NOR chromosome. These data support the hypothesis that dogroses from both subsections arose by reciprocal hybridization of closely related species (Bruneau et al., 2007; Herklotz and Ritz, 2017). A very similar composition of dogrose genomes is also supported by the shallow nodes among dogroses in the 5S rDNA phylogeny (**Figure 4**). However, within the subsections, we detected some variation in number and distribution of loci especially between the univalent sets. For example, *R. inodora* carried two univalent-forming chromosomes with Canina type configuration, while *R. rubiginosa* contained only one. Additionally, one of the *R. canina* univalent chromosomes carried both 5S_A and 5S_B chromosomes co-localized, while this chromosome was not found in the closely related *R. corymbifera*. The variation between univalent chromosomes is consistent with increased diversity of microsatellite markers on univalent genomes (Nybom et al., 2004) and could be either related to divergence of genome donors and/or to partial degeneration of univalent chromosomes due to their exclusion from recombination in meiosis. Although 5S rDNA pseudogenes seem to be present in *R. rugosa* (Tynkevich and Volkov, 2014b), there is no indication for extensive 5S rDNA pseudogenization in *R. canina* (Lim et al., 2005) and in other dogroses (this work).

Although it might be preliminary to trace potential genome donors of allopolyploid dogroses, it is notable that the tetraploid species with a regular meiosis, *R. gallica* (sect. *Gallicanae*), tends to cluster with dogroses. It is therefore possible that pentaploid dogroses actually arose by pollination of an unreduced non-dogrose tetraploid egg cell by a reduced male gamete from a diploid donor. In support, *R. gallica* ITS types were occasionally found in dogroses (Herklotz et al., 2018) and *R. gallica* belonged to a clade together with dogroses in plastid phylogenies (Fougere-Danezan et al., 2015). Moreover, *R. gallica* shares the distinct morphological feature of partially pinnate sepals with dogroses, which is absent in the remaining species of the genus. However, there are also tetraploid cytotypes within dogroses (e.g., *R. villosa* L., *R. canina*) forming triploid egg cells and haploid sperm cells, which would not fit in the proposed scenario so far but might have arisen by another combination of partially reduced gametes. The also occurring higher ploidy (6x) levels which are less frequently found in dogroses rather originated from hybridizations within dogrose species involving unreduced egg cells (Herklotz and Ritz, 2017).

Potential Factors Influencing Genetic Stability of 5S rDNA Loci in Roses

The maintenance of two 5S rDNA families in the *Rosa* genomes is consistent with the increased stability of 5S loci as compared to 35S loci in allopolyploid genomes documented in several allopolyploid systems (Pedrosa-Harand et al., 2006; Weiss-Schneeweiss et al., 2008; Garcia et al., 2016; Amosova et al., 2019). The reasons for relative stasis of 5S rDNA loci are not well understood, while their position on chromosomes (Garcia et al., 2016) and epigenetic modifications (5S rDNA loci carry mostly heterochromatic landmarks; Simon et al., 2018) have been discussed. One also has to consider the relative scarcity of meiosis driving genetic recombination (and homogenization) in these long-living perennial shrubs. For example, a *R. canina* individual known as "Rose of Hildesheim" (North Germany) is estimated to be more than 700 years old (Peters and Peters, 2013).

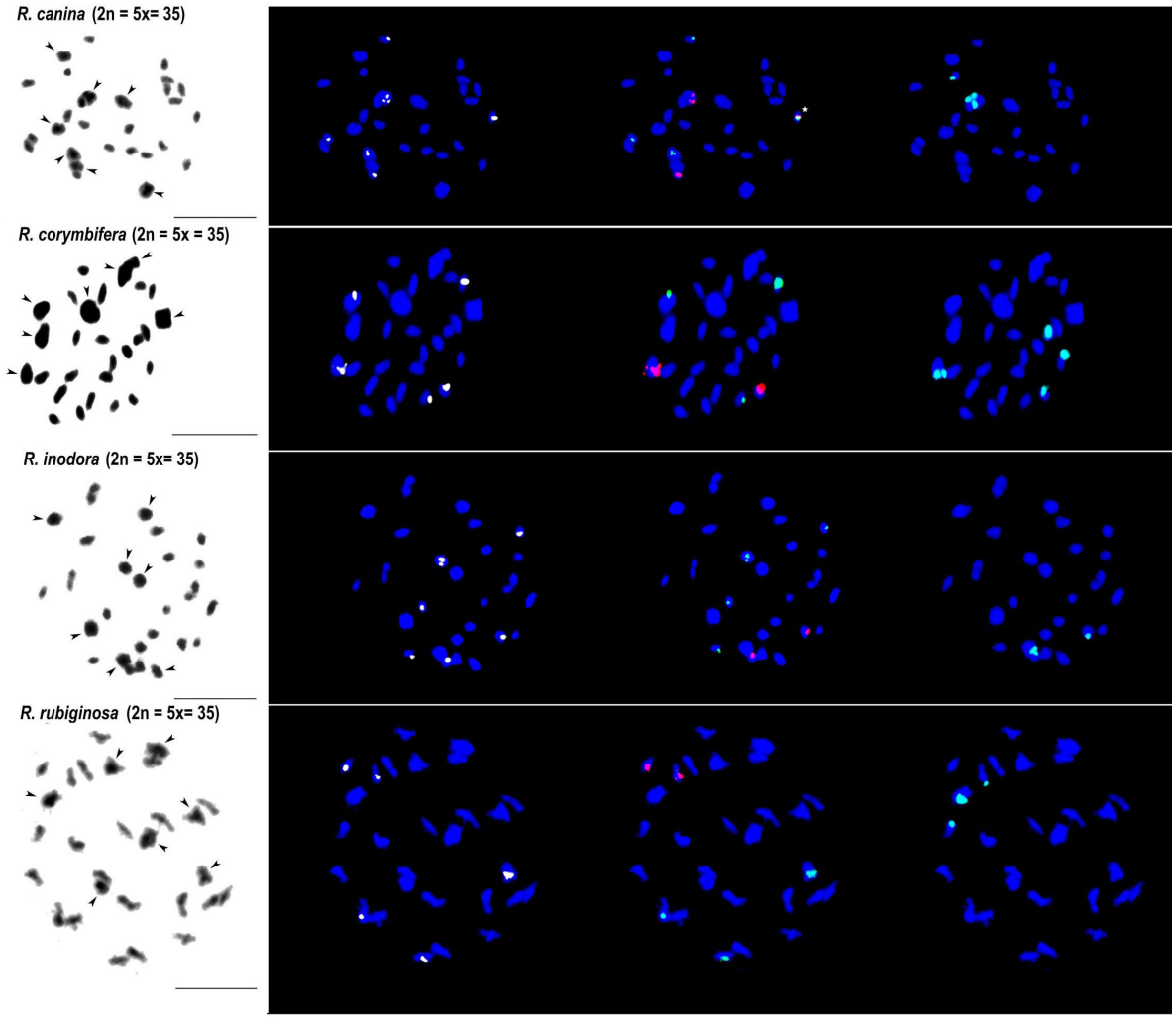


FIGURE 7 | FISH analysis of meiotic chromosomes in pentaploid dogroses section *Caninae*: *R. canina* and *R. corymbifera* (subsection *Caninae*); *R. inodora* and *R. rubiginosa* (subsection *Rubigineae*). In each species, the same diakinesis was hybridized with the 5S genic **(B)**, 5S intergenic **(C)**, and 18S rDNA **(D)** probes. DAPI staining (gray scale, inverted) is to the left margin **(A)**. Arrowheads indicate bivalent chromosomes. In *R. canina*, a chromosome with co-localized 5S_A and 5S_B signals is marked with an asterisk. Scale bar: 10 μm.

Interestingly, *Gossypium* allopolyploids which represent also perennial shrubs tend to maintain 5S rDNA loci relatively intact over millions of years, while they homogenized their 35S rDNA loci (Cronn et al., 1996).

CONCLUSION

We identified two 5S rDNA families which are widespread across the *Rosa* genus. The molecular and cytogenetic observations lead us to propose that both families have their origin deep in the genus history probably close to its base. A remarkably slow tempo of 5S rDNA evolution differs from other systems where these loci show considerable dynamics. The retention of a large number of ancient rDNA sequences in *Rosa* genomes resonates with drastic allelic heterozygosity encountered in previous studies of microsatellites (Nybom et al., 2006), protein coding genes sequences (Joly and Bruneau, 2006), and more recent whole genome sequencing projects (Raymond et al., 2018). In future, it will be interesting to analyze expression of alleles inherited from deep evolutionary times.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and

accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW349696; https://www.ncbi.nlm.nih.gov/genbank/, MW349697.

AUTHOR CONTRIBUTIONS

AK, CR, and RAV conceived and designed the study. CR, JL, RAV, VH, and YT performed the experiments and collected material. JL, RV, RAV, VH, and YT analyzed the data. AK, CR, JL, and RAV wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the German Science Foundation (DFG Ri 2090/3-1), the Czech Science Foundation (GAČR 20-14133J), the Ministry of Education and Science of Ukraine (Grant No. 0106U003620), and DAAD (German Academic Exchange Service) Research Fellowship for Yuri Tynkevich.

ACKNOWLEDGMENTS

We are thankful to M. Schwager, M. Laufer, and J. Lorenz (Senckenberg Museum of Natural History Görlitz) for their excellent help in the laboratory. We thank V. Wissemann (Justus Liebig University Gießen, Germany), G. Vogg (Botanical Garden of University Würzburg, Germany), S. Arndt (Botanical Garden of Friedrich Schiller University Jena, Germany), U. Pietzarka (Forstpark Tharandt, TU Dresden, Germany), and M. I. Vykliuk (Botanical Garden of the Yuriy Fedkovych Chernivtsi National University, Ukraine) for their kind help in obtaining access to plant material. We also thank P. Hotka and P. Ferus (Botanical Garden Mlynany of the Slovak Academy of Sciences) for their kind help and for enabling the collecting of the samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.643548/full#supplementary-material

Supplementary Figure 1 | Genomic analysis of 5S rDNA variants. Projections of 5S rDNA cluster graphs in polyploid *Rosa* species. Loops representing the IGS reads are colored according to the 5S_A (green) and 5S_B (red) IGS sequences. Blue nodes represent the 5S genic regions.

Supplementary Figure 2 | Phylogenic relationships among *Rosa* species inferred from Maximum-Likelihood analyses of 5S rDNA intergenic spacers. The tree was rooted with *Geum urbanum*. Species names are colored according to ploidy level (2x—black, 4x—light blue, 5x—pink). Branch colors represent 5S_A (green) and 5S_B (red) lineages. Bootstrap support >70% is indicated above branches.

Supplementary Figure 3 | A Neighbor-Joining phylogeny tree constructed from the 5S rDNA consensus sequences. The same data set as in **Figure 4** was analyzed using the Juke Cantor model. One thousand repetitions were allowed for the statistical support. Bootstrap support levels >80% are indicated.

Supplementary Figure 4 | Phylogenetic NJ trees constructed from type A and B 5S rDNA sequences, respectively. Note a similar position of *R. persica* on both trees (early separating). Note clustering of *Caninae* species (red asterisks after the names) in an unresolved branch. Note incongruent placement of the *R. spinosissima* variants.

Supplementary Figure 5 | FISH analysis of root tip mitotic chromosomes from *R. spinosissima* and *R. nitida*. Arrows indicate NOR chromosomes with co-localized 5S rDNA loci. Note odd (5) number of 5S rDNA sites in the assumingly *R. spinosissima* allotetraploid.

Supplementary Table 1 | Information about the taxonomy and origin of plants used in this study.

Supplementary Table 2 | Details of bioinformatic procedures leading to 5S rDNA genome proportion and copy number estimation. Sheet 1—5S_A variant. Sheet 2—5S_B variant.

Supplementary Table 3 | Summary of genome proportions and copy number of 5S rDNA families in *Rosa* genomes.

Supplementary Table 4 | Summary of cytogenetic FISH analyses.

REFERENCES

Amosova, A. V., Zoshchuk, S. A., Rodionov, A. V., Ghukasyan, L., Samatadze, T. E., Punina, E. O., et al. (2019). Molecular cytogenetics of valuable Arctic and sub-Arctic pasture grass species from the *Aveneae/Poeae* tribe complex (*Poaceae*). *BMC Genet.* 20:92. doi: 10.1186/s12863-019-0792-2

Benson, C. W., Mao, Q., and Huff, D. R. (2020). Global DNA methylation predicts epigenetic reprogramming and transgenerational plasticity in *Poa annua* L. *Crop Sci.* 2020, 1–12. doi: 10.1002/csc2.20337

Blackburn, K. B., and Harrison, J. W. H. (1921). The status of the British rose forms as determined by their cytological behaviour. *Ann. Bot.* 35, 159–188. doi: 10.1093/oxfordjournals.aob.a089753

Bruneau, A., Starr, J. R., and Joly, S. (2007). Phylogenetic relationships in the genus *Rosa*: New evidence from chloroplast DNA sequences and an appraisal of current knowledge. *Syst. Bot.* 32, 366–378. doi: 10.1600/036364407781179653

Christ, H. (1873). *Die Rosen der Schweiz*. Basel: Verlag H. Georg.

Crépin, F. (1889). Sketch of a new classification of roses. *J. R. Hortic. Soc.* 11, 217–228.

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Debray, K., Marie-Magdelaine, J., Ruttink, T., Clotault, J., Foucher, F., and Malecot, V. (2019). Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). *BMC Evol. Biol.* 19:152. doi: 10.1186/s12862-019-1479-z

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111–117. doi: 10.1038/299111a0

Edelman, D. W. (1975). *The Eocene Germer Basin Flora of South-Central Idaho*. Ph. D. thesis, Moscow: University of Idaho.

Eickbush, T. H., and Eickbush, D. G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175, 477–485. doi: 10.1534/genetics.107.071399

Fernandez-Romero, M. D., Torres, A. M., Millan, T., Cubero, J. I., and Cabrera, A. (2001). Physical mapping of ribosomal DNA on several species of the subgenus *Rosa*. *Theor. Appl. Genet.* 103, 835–838. doi: 10.1007/s001220100709

Fougere-Danezan, M., Joly, S., Bruneau, A., Gao, X. F., and Zhang, L. B. (2015). Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann. Bot.* 115, 275–291. doi: 10.1093/aob/mcu245

Fulnecek, J., Lim, K. Y., Leitch, A. R., Kovařik, A., and Matyasek, R. (2002). Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species. *Heredity* 88, 19–25. doi: 10.1038/sj.hdy.6800001

Garcia, S., Borowska-Zuchowska, N., Wendel, J. F., Ainouche, M., Kuderova, A., and Kovařik, A. (2020). The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids and cryptic introgressants. *Front. Plant Sci.* 11:41. doi: 10.3389/fpls.2020.00041

Garcia, S., Kovařik, A., Leitch, A. R., and Garnatje, T. (2016). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259

Hemleben, V., Ganal, M., Gersnter, J., Schiebel, K., and Torres, R. A. (1988). "Organization and length heterogeneity of plant ribosomal RNA genes," in *The architecture of Eukaryotic Gene*, ed. G. Kahl (Weinheim: VHC), 371–384.

Herklotz, V., and Ritz, C. M. (2017). Multiple and asymmetrical origin of polyploid dog rose hybrids (*Rosa* L. sect. *Caninae* (DC.) Ser.) involving unreduced gametes. *Ann. Bot.* 120, 209–220. doi: 10.1093/aob/mcw217

Herklotz, V., Kovařik, A., Lunerova, J., Lippitsch, S., Groth, M., and Ritz, C. M. (2018). The fate of ribosomal RNA genes in spontaneous polyploid dogrose hybrids [*Rosa* L. sect. *Caninae* (DC.) Ser.] exhibiting non-symmetrical meiosis. *Plant J.* 94, 77–90. doi: 10.1111/tpj.13843

Joly, S., and Bruneau, A. (2006). Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: An example from *Rosa* in north America. *Syst. Biol.* 55, 623–636. doi: 10.1080/10635150600863109

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kirov, I. V., Van Laere, K., Van Roy, N., and Khrustaleva, L. I. (2016). Towards a FISH-based karyotype of *Rosa* L. (Rosaceae). *Comp. Cytogenet.* 10, 543–554. doi: 10.3897/CompCytogen.v10i4.9536

Koopman, W. J. M., Wissemann, V., De Cock, K., Van Huylenbroeck, J., De Riek, J., Sabatlno, G. J. H., et al. (2008). AFLP markers as a tool to reconstruct complex relationships: A case study in *Rosa* (Rosaceae). *Am. J. Bot.* 95, 353–366. doi: 10.3732/ajb.95.3.353

Kovařik, A., Werlemark, G., Leitch, A. R., Souckova-Skalicka, K., Lim, Y. K., Khaitova, L., et al. (2008). The asymmetric meiosis in pentaploid dogroses (*Rosa* sect. *Caninae*) is associated with a skewed distribution of rRNA gene families in the gametes. *Heredity* 101, 359–367. doi: 10.1038/Hdy.2008.63

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/Nmeth.1923

Lim, K. Y., Werlemark, G., Matyasek, R., Bringloe, J. B., Sieber, V., El Mokadem, H., et al. (2005). Evolutionary implications of permanent odd polyploidy in the stable sexual, pentaploid of *Rosa canina* L. *Heredity* 94, 501–506. doi: 10.1038/sj.hdy.6800648

Liu, C. Y., Wang, G. L., Wang, H., Xia, T., Zhang, S. Z., Wang, Q. G., et al. (2015). Phylogenetic relationships in the genus *Rosa* revisited based on *rpl16*, *trnL-F*, and *atpB-rbcL* sequences. *HortScience* 50, 1618–1624. doi: 10.21273/Hortsci.50.11.1618

Long, E. O., and Dawid, I. B. (1980). Repeated genes in eukaryotes. *Annu. Rev. Biochem.* 49, 727–764. doi: 10.1146/annurev.bi.49.070180.003455

Lunerova, J., Herklotz, V., Laudien, M., Vozarova, R., Groth, M., Kovařik, A., et al. (2020). Asymmetrical canina meiosis is accompanied by the expansion of a pericentromeric satellite in non-recombining univalent chromosomes in the genus *Rosa*. *Ann. Bot.* 125, 1025–1038. doi: 10.1093/aob/mcaa028

Ma, Y., Crane, C. F., and Byrne, D. H. (1997). Karyotypic relationships among some *Rosa* species. *Caryologia* 50, 317–326. doi: 10.1080/00087114.1997.10797405

Matsumoto, S., Kouchi, M., Yabuki, J., Kusunoki, M., Ueda, Y., and Fukui, H. (1998). Phylogenetic analyses of the genus *Rosa* using the matK sequence: molecular evidence for the narrow genetic background of modern roses. *Sci. Hortic.* 77, 73–82. doi: 10.1016/S0304-4238(98)00169-1

Millan, T., Osuna, F., Cobos, S., Torres, A. M., and Cubero, J. I. (1996). Using RAPDs to study phylogenetic relationships in *Rosa*. *Theor. Appl. Genet.* 92, 273–277. doi: 10.1007/Bf00223385

Nagylaki, T. (1990). Gene conversion, linkage, and the evolution of repeated genes dispersed among multiple chromosomes. *Genetics* 126, 261–276.

Nakamura, N., Hirakawa, H., Sato, S., Otagaki, S., Matsumoto, S., Tabata, S., et al. (2018). Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* 25, 113–121. doi: 10.1093/dnares/dsx042

Nieto Feliner, G., and Rossello, J. A. (2012). "Concerted evolution of multigene families and homeologous recombination," in *Plant Genome Diversity*, ed. J. F. Wendel (Wien: Springer-Verlag), 171–194.

Novak, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* 11:378. doi: 10.1186/1471-2105-11-378

Novak, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Nybom, H., Esselink, G. D., Werlemark, G., and Vosman, B. (2004). Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog-roses, *Rosa* L. sect. *Caninae* DC. *Heredity* 92, 139–150. doi: 10.1038/sj.hdy.6800332

Nybom, H., Esselink, G. D., Werlemark, G., Leus, L., and Vosman, B. (2006). Unique genomic configuration revealed by microsatellite DNA in polyploid dogroses, *Rosa* sect. *Caninae*. *J. Evol. Biol.* 19, 635–648. doi: 10.1111/j.1420-9101.2005.01010.x

Ohta, T. (1984). Some models of gene conversion for treating the evolution of multigene families. *Genetics* 106, 517–528.

Pachl, Š (2011). *Variablita botanických druhů rodu Rosa L., a možnosti jejich využití v krajinářské tvorbê*. Ph. D.thesis, Nitra, SK: Slovak University of Agriculture.

Pastova, L., Belyayev, A., and Mahelka, V. (2019). Molecular cytogenetic characterisation of *Elytrigia xmucronata*, a natural hybrid of *E. intermedia* and *E. repens* (Triticeae, Poaceae). *BMC Plant. Biol.* 19:230. doi: 10.1186/s12870-019-1806-y

Pedrosa-Harand, A., de Almeida, C. C. S., Mosiolek, M., Blair, M., Schweizer, D., and Guerra, M. (2006). Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. *Theor. Appl. Genet.* 112, 924–933. doi: 10.1007/s00122-005-0196-8

Peters, H., and Peters, H. (2013). *Tausendjähriger Rosenstock from Hildesheim (the thousand-year-oldrosebush)*. Available online at: www.webcitation.org/6PIIQmfN8?url=http://www.worldrose.org/heritage/HeritageNumber10.pdf

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., et al. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nat. Genet.* 50:772. doi: 10.1038/s41588-018-0110-3

Ritz, C. M., Schmuths, H., and Wissemann, V. (2005). Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. *J. Hered.* 96, 4–14. doi: 10.1093/jhered/esi011

Roa, F., and Guerra, M. (2012). Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evol. Biol.* 12:225. doi: 1471-2148-12-225

Roberts, A. V., Gladis, T., and Brumme, H. (2009). DNA amounts of roses (*Rosa* L.) and their use in attributing ploidy levels. *Plant Cell Rep.* 28, 61–71. doi: 10.1007/s00299-008-0615-9

Rogers, S. O., and Bendich, A. J. (1985). Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant-tissues. *Plant Mol. Biol.* 5, 69–76. doi: 10.1007/Bf00020088

Saint-Oyant, L. H., Ruttink, T., Hamama, L., Kirov, I., Lakhwani, D., Zhou, N. N., et al. (2018). A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat. Plants* 4, 473–484. doi: 10.1038/s41477-018-0166-1

Schwarzacher, T., and Heslop-Harrison, P. (2000). *Practical in situ hybridization*. Oxford: BIOS Scientific Publishers.

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., et al. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucl. Acids Res.* 46, 3019–3033. doi: 10.1093/nar/gky163

Täckholm, G. (1920). On the cytology of the genus *Rosa*. *Svensk Bot. Tidskrift* 1920, 300–311.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9, 678–687.

Tamura, K., Tao, Q., and Kumar, S. (2018). Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol. Biol. Evol.* 35, 1770–1782. doi: 10.1093/molbev/msy044

Tynkevich, Y. O., and Volkov, R. A. (2014a). Novel structural class of 5S rDNA of *Rosa wichurana* Crep. *Rep. Natl. Acad. Sci. Ukraine* 5, 143–148. doi: 10.15407/dopovidi2014.05.143

Tynkevich, Y. O., and Volkov, R. A. (2014b). Structural organization of 5Sribosomal DNA in *Rosa rugosa*. *Cytol. Genet.* 48, 1–6. doi: 10.3103/S0095452714010095

Volkov, R. A., Panchuk, I. I., Borisjuk, N. V., Hosiawa-Baranska, M., Maluszynska, J., and Hemleben, V. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol.* 17:6. doi: 10.1186/s12870-017-0978-6

Volkov, R. A., Zanke, C., Panchuk, I. I., and Hemleben, V. (2001). Molecular evolution of 5S rDNA of *Solanum* species (sect. *Petota*): application for molecular phylogeny and breeding. *Theor. Appl. Genet.* 103, 1273–1282. doi: 10.1007/s001220100670

Wang, G. L. (2007). A study on the history of Chinese roses from ancient works and images. *Acta Hortic.* 751, 347–356. doi: 10.17660/ActaHortic.2007.751.44

Weiss-Schneeweiss, H., Tremetsberger, K., Schneeweiss, G. M., Parker, J. S., and Stuessy, T. F. (2008). Karyotype diversification and evolution in diploid and polyploid South American *Hypochaeris* (Asteraceae) inferred from rDNA localization and genetic fingerprint data. *Ann. Bot.* 101, 909–918. doi: 10.1093/aob/mcn023

Wissemann, V. (1999). Genetic constitution of *Rosa* sect. *Caninae* (*R. canina, R. jundzillii*) and sect. *Gallicanae* (*R. gallica*). *Angew. Bot.* 73, 191–196.

Wissemann, V. (2003). "Conventional taxonomy (wild roses)," in *Encyclopedia of Rose Science*, eds A. V. Roberts, T. Debener, and S. Gudin (Oxford: Elsevier Academic Press).

Wissemann, V., and Ritz, C. M. (2005). The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atpB-rbc*L intergenic spacer (IGS) versus conventional taxonomy. *Bot. J. Linn. Soc.* 147, 275–290. doi: 10.1111/j.1095-8339.2005.00368.x

Xiang, Y. Z., Huang, C. H., Hu, Y., Wen, J., Li, S. H., Yi, T. S., et al. (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* 34, 1026–1026. doi: 10.1093/molbev/msx093

Yokoya, K., Roberts, A. V., Mottley, J., Lewis, R., and Brandham, P. E. (2000). Nuclear DNA amounts in roses. *Ann. Bot.* 85, 557–561. doi: 10.1006/ambo.1999

Zetter, R., Hofmann, C., Draxler, I., Durango, de Cabrera, J., Vergel, M., et al. (1999). A rich Middle Eocene Microflora at Arroyo de los Mineros, near Cañadón Beta, NE Tierra del Fuego Province, Argentina. *Abh. Geol. Bundesanst.* 56, 439–460.

Zhang, J., Esselink, G. D., Che, D., Fougere-Danezan, M., Arens, P., and Smulders, M. J. M. (2013). The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. *J. Hortic. Sci. Biotechnol.* 88, 85–92. doi: 10.1080/14620316.2013.11512940

Zhu, Z. M., Gao, X. F., and Fougere-Danezan, M. (2015). Phylogeny of *Rosa* sections *Chinenses* and *Synstylae* (Rosaceae) based on chloroplast and nuclear markers. *Mol. Phylogenet. Evol.* 87, 50–64. doi: 10.1016/j.ympev.2015.03.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 March 2021
doi: 10.3389/fpls.2021.647375

Molecular Evolution and Organization of Ribosomal DNA in the Hawkweed Tribe Hieraciinae (Cichorieae, Asteraceae)

Judith Fehrer[1], Renáta Slavíková[1], Ladislava Paštová[1], Jiřina Josefiová[1], Patrik Mráz[2], Jindřich Chrtek[1,2] and Yann J. K. Bertrand[1]*

[1] Institute of Botany, Czech Academy of Sciences, Průhonice, Czechia, [2] Department of Botany, Charles University, Prague, Czechia

Molecular evolution of ribosomal DNA can be highly dynamic. Hundreds to thousands of copies in the genome are subject to concerted evolution, which homogenizes sequence variants to different degrees. If well homogenized, sequences are suitable for phylogeny reconstruction; if not, sequence polymorphism has to be handled appropriately. Here we investigate non-coding rDNA sequences (ITS/ETS, 5S-NTS) along with the chromosomal organization of their respective loci (45S and 5S rDNA) in diploids of the Hieraciinae. The subtribe consists of genera *Hieracium*, *Pilosella*, *Andryala,* and *Hispidella* and has a complex evolutionary history characterized by ancient intergeneric hybridization, allele sharing among species, and incomplete lineage sorting. Direct or cloned Sanger sequences and phased alleles derived from Illumina genome sequencing were subjected to phylogenetic analyses. Patterns of homogenization and tree topologies based on the three regions were compared. In contrast to most other plant groups, 5S-NTS sequences were generally better homogenized than ITS and ETS sequences. A novel case of ancient intergeneric hybridization between *Hispidella* and *Hieracium* was inferred, and some further incongruences between the trees were found, suggesting independent evolution of these regions. In some species, homogenization of ITS/ETS and 5S-NTS sequences proceeded in different directions although the 5S rDNA locus always occurred on the same chromosome with one 45S rDNA locus. The ancestral rDNA organization in the Hieraciinae comprised 4 loci of 45S rDNA in terminal positions and 2 loci of 5S rDNA in interstitial positions per diploid genome. In *Hieracium*, some deviations from this general pattern were found (3, 6, or 7 loci of 45S rDNA; three loci of 5S rDNA). Some of these deviations concerned intraspecific variation, and most of them occurred at the tips of the tree or independently in different lineages. This indicates that the organization of rDNA loci is more dynamic than the evolution of sequences contained in them and that locus number is therefore largely unsuitable to inform about species relationships in *Hieracium*. No consistent differences in the degree of sequence homogenization and the number of 45S rDNA loci were found, suggesting interlocus concerted evolution.

Keywords: 5S rDNA, 45S rDNA, *Andryala*, concerted evolution, *Hieracium*, *in situ* hybridization, molecular phylogeny, *Pilosella*

OPEN ACCESS

Edited by:
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia

Reviewed by:
Elvira Hörandl, University of Göttingen, Germany
Roman Matyasek, Academy of Sciences of the Czech Republic, Czechia

****Correspondence:***
Judith Fehrer
fehrer@ibot.cas.cz

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *04 January 2021*
Accepted: *19 February 2021*
Published: *12 March 2021*

Citation:
Fehrer J, Slavíková R, Paštová L, Josefiová J, Mráz P, Chrtek J and Bertrand YJK (2021) Molecular Evolution and Organization of Ribosomal DNA in the Hawkweed Tribe Hieraciinae (Cichorieae, Asteraceae). Front. Plant Sci. 12:647375. doi: 10.3389/fpls.2021.647375

INTRODUCTION

The Cichorieae subtribe Hieraciinae is well defined on molecular and morphological grounds (Fehrer et al., 2007a; Krak and Mráz, 2008). Genera of the subtribe are *Hieracium* s.str., *Pilosella* (formerly treated as a subgenus of *Hieracium*, Bräutigam and Greuter, 2007), *Andryala* and monotypic *Hispidella* (Kilian et al., 2009). The basic chromosome number of all Hieraciinae is $x = 9$, with diploid representatives having $2n = 2x = 18$. The mainly European genera *Pilosella* and *Hieracium* comprise many polyploid taxa, most or all of which reproduce apomictically, i.e., they form seeds without fertilization resulting in progeny corresponding to the maternal genotype (Krahulcová et al., 2000; Mráz and Zdvořák, 2019). Distribution ranges of diploids of both genera are usually small and often constrained to glacial refugia (Merxmüller, 1975). *Andryala* is an entirely diploid genus with its main distribution in Macaronesia and the Mediterranean region (Ferreira et al., 2015). *Hispidella hispanica* is also diploid and occurs in the central and western parts of the Iberian Peninsula (Tutin et al., 1976).

Phylogenetic relationships within the Hieraciinae have been previously inferred based on the internal transcribed spacer (ITS) region and the external transcribed spacer (ETS) of nuclear ribosomal DNA (rDNA) as well as on several chloroplast and low-copy nuclear markers (Fehrer et al., 2007a, 2009; Krak et al., 2013; Ferreira et al., 2015; Chrtek et al., 2020). ITS and ETS (the 5′ part of the intergenic spacer) are non-coding parts of the tandemly repeated 18S-5.8S-26S rDNA cistron, whose organization is the same in most organisms (Rogers and Bendich, 1987; Hillis and Dixon, 1991), namely ETS-18S-ITS1-5.8S-ITS2-26S. It is also referred to as 45S rDNA (sometimes 35S or 25S), which is commonly used as a cytogenetic marker (Denduangboripant et al., 2007; Garcia et al., 2010; Lan and Albert, 2011). The tandemly repeated 5S rDNA gene usually occurs separately from the 45S rDNA array in other regions of the genome in plants and animals (Hemleben and Grierson, 1978; Long and Dawid, 1980; Appels and Honeycutt, 1986; Wicke et al., 2011; but see Garcia et al., 2007, 2017 for exceptions), and the non-transcribed spacer (NTS) separates its individual units. The NTS is highly variable in plants (Cronn et al., 1996; Kaplan et al., 2013; Mahelka et al., 2013), but has not yet been used to infer species relationships in the Hieraciinae. The correspondence of cytogenetically employed 45S and 5S rDNA probes with highly variable sequences contained in these regions allows comparing phylogenetic trees of closely related species with the number and localization of the corresponding loci on chromosomes.

Both rDNAs occur in arrays of hundreds to thousands of copies (Long and Dawid, 1980), which are often homogenized by concerted evolution within individuals and species (Arnheim, 1983; Nieto Feliner and Rosselló, 2007). We found previously that ITS is fairly well homogenized in the Hieraciinae (Fehrer et al., 2007a; Ferreira et al., 2015) whereas ETS frequently retained two or more variants in *Hieracium* (Fehrer et al., 2009). This also applied to many diploids investigated and has been attributed to ancient hybridization between lineages or incomplete lineage sorting near the base of the genus. However, some of the ETS variants were found to be homogenized and occasionally shared by other species whereas others were never found as the only variants in any of the species analyzed and were presumed to belong to unknown or extinct lineages. 5S-NTS sequences of two species of *Hieracium* were well homogenized (Zagorski et al., 2020). A few groups of related species were consistently found with different molecular markers (nrDNA, cpDNA, low-copy nuclear genes), but their relationships remained mostly unresolved or were in strong conflict with each other (Krak et al., 2013). So far, each molecular marker applied to *Hieracium* has revealed a particular aspect of the speciation of the genus, but ETS was thought to reflect the evolutionary history best, because it was in concordance with geographic distribution and genome size (Chrtek et al., 2009).

Our initial cytogenetic analyses of *Hieracium* focusing on satellite DNA showed that two species had two 45S rDNA loci and one 5S rDNA locus per haploid genome whereas a third species had three 45S rDNA loci (Belyayev et al., 2018). To assess the variability in the number and position of rDNA loci in *Hieracium*, we extend here the sampling of diploid species and include diploid *Pilosella* and *Andryala* taxa to infer the ancestral pattern in the Hieraciinae. Because, in diploids, the 5S rDNA locus so far always occurred on a chromosome also bearing one of the 45S rDNA loci (Belyayev et al., 2018; Mráz et al., 2019), we ask whether phylogenies based on markers obtained from these regions are congruent or not. We investigate whether the 5S-NTS spacer provides new insights into the diversification of *Hieracium* and related genera, what level of resolution it provides compared to ITS and ETS, how well it is homogenized in the Hieraciinae, if concerted evolution of ITS/ETS and 5S-NTS occurred in the same direction, and how these patterns conform to the number and position of 45S and 5S rDNA loci on chromosomes. We further investigate if cytogenetic patterns are in accordance with the phylogenetic patterns.

MATERIALS AND METHODS

Plant Material

All genera of the Hieraciinae were included in phylogenetic analyses. The little-studied, exclusively American *Hieracium* subgenus *Chionoracium* (Sleumer, 1956) was ignored here because of a lack of material. We also did not include polyploids, because most of the accessions analyzed were found to have allopolyploid origin in *Hieracium* (Krak et al., 2013; Chrtek et al., 2020) and *Pilosella* (Krahulec et al., 2004, 2008; Fehrer et al., 2005, 2007b), and we expected potential confounding effects of reticulation on the organization of the loci (Zagorski et al., 2020).

All major lineages of diploids were represented by 1–3 samples per species, if possible, from different geographic regions. Most diploids of *Hieracium* that had previously shown indications of hybrid origin (showing mixed ETS sequences) were excluded; 18 species included here are representative for all lineages. For *Pilosella* and *Hispidella hispanica*, the same species as in Fehrer et al. (2007a) were sampled (14+1); for some *Pilosella* species, additional accessions were included. *Andryala* was represented by five species, two of which consistently formed long basal branches

and another three belonged to the major radiation of the genus (Ferreira et al., 2015; Zahradníček et al., 2018).

Sampling for cytogenetic investigations was as far as possible based on the same individuals that were sequenced; if this was not feasible (herbarium specimens, lack of good metaphases, plants perished), a sample from the same population was sequenced or a larger geographic range was covered by several accessions. Cytogenetic analyses were carried out for a subset of species representing the major lineages; whenever available, more than one individual per species was included. Altogether, 64 samples of 38 species were sequenced, and 29 samples of 18 species were analyzed by fluorescence *in situ* hybridization (FISH). A list of species is provided in **Table 1**. Details about sample origins and voucher information are included in **Supplementary Table 1**.

Sanger Sequencing

Sequences of ITS and ETS of Hieraciinae from previous studies (Fehrer et al., 2007a, 2009; Ferreira et al., 2015) were complemented by newly generated sequences of the same samples (mainly ITS for *Hieracium* and ETS for *Pilosella*). 5S-NTS sequences, so far available for only two species of *Hieracium* (Zagorski et al., 2020), were newly generated for all other samples.

PCR amplification and sequencing of ITS was done as described in Fehrer et al. (2007a), sequencing of ETS followed Fehrer et al. (2009), and procedures for 5S-NTS were as in Kaplan et al. (2013). *Pilosella* samples show a tandem repeat structure in the ETS region and could only be sequenced with the reverse primer, otherwise all sequencing was done in both directions to account for polymorphic sites and to obtain full-length sequences. Polymorphic sites were represented by the IUPAC ambiguity codes and maintained if they were clearly visible on both strands and if their relative amounts were similar, i.e., small additional peaks were ignored so as not to introduce too much noise in phylogenetic analyses. If direct sequences were unreadable due to longer indels, the respective samples were cloned as described in Fehrer et al. (2009); five clones per sample were sequenced in one direction. Sequences were submitted to GenBank (accession numbers MW325251–MW325296, MW315935–MW315953, MW328890–MW329033, MW587333–MW587351, and MW591759–MW591773), see also **Table 1**.

Genome Skimming Approach

For 20 samples, low-coverage genome sequencing was newly performed. For these, DNA was extracted from fresh or silica-gel dried leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Library preparation and low-coverage Illumina sequencing were performed at GATC Biotech (Konstanz, Germany)/Eurofins Genomics (Ebersberg, Germany) using a standardized protocol that produced 150 bp paired-end reads with an insert size of ∼450 bp. The raw Illumina datasets have been submitted to the European Nucleotide Archive (ENA) under the study no. PRJEB41719. Raw reads were filtered to remove sequences shorter than 120 bp and Illumina adapters using the Trimmomatic v0.39 tool (Bolger et al., 2014) with parameter settings: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:120.

In order to retrieve the sequences corresponding to the 45S rDNA and the 5S rDNA loci, we adopted a reference-guided approach with manual correction based on *de novo* contigs. Our workflow began by creating an unrefined *de novo* assembly from the total low-coverage sequences for a single representative of each of the three genera (*H. kittanae*: 1228/2, *A. laevitomentosa*: Alev18, *P. hoppeana*: H1702) using SPAdes v3.14.0 (Bankevich et al., 2012) with default settings. Contigs corresponding to the 45S rDNA and the 5S/5S-NTS rDNA loci were identified using BLAST+ v2.7.1 (Camacho et al., 2009) (blastn -perc_identity 90 -evalue 1E-50 -max_target_seqs 1) against a database of known sequences (*Helianthus annuus* DQ865267.1 for the 5S, *H. prenanthoides* MN784129.1 for the 5S-NTS, *H. alpinum* EU867634.1 for the 45S rDNA). The contigs provided genus specific reference sequences for the subsequent study.

For each sample, we used BLAST+ to obtain all reads matching the appropriate reference sequences (blastn -word_size 18 -perc_identity 90 -qcov_hsp_perc 55 -max_target_seqs 1). Each sample was thus blasted against one reference sequence for the 45S rDNA and one for the 5S/5S-NTS rDNA locus. Each set of matching reads was corrected for Illumina sequencing errors using the correction algorithm of SPAdes (-only-error-correction -k 21,33,55,77 –careful) followed by correction with Karect (Allam et al., 2015) (correct -matchtype = hamming -celltype = diploid) in order to obtain reads for further mappings and assemblies.

Corrected reads that originated from the two focal markers were mapped on the references using bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012) with stringent settings (-very-fast-local), and reads that failed to align were discarded. For each species, we mapped the reads on the appropriate reference sequence that belonged to the same genus. For each sample/reference combination, we generated a consensus sequence from the mapped reads using Kindel v0.4.2 (Constantinides and Robertson, 2017) (–min-depth 10). Mapped reads were *de novo* assembled using SPAdes (–only-assembler -k 21,33,55,77 –careful). The resulting contigs were aligned together with the consensus sequence with MAFFT v7.471 (Katoh and Standley, 2013) (–adjustdirection –auto –addfragments). The consensus sequences were checked for missing indels by visual comparison with the *de novo* contigs and manually corrected in Bioedit v7.3 (Hall, 1999). A visual sanity check of the bam files was performed in Tablet v1.19.09.03 (Milne et al., 2013). The corrected consensuses were used as references for a final read mapping with bowtie (-very-fast-local) that produced bam formatted files.

Phasing was carried out on the bam files in order to separate allelic variants. Each bam file was analyzed with Samtools v1.10 (Li et al., 2009). The obtained mpileup file was further processed with VarScan v2.3.8 (Koboldt et al., 2012) in order to infer valid single nucleotide polymorphisms (SNPs). Valid SNP positions had to be located in regions with good read coverage (at least eight reads), the minimum number of supporting reads at a position to call a SNP was

TABLE 1 | Samples used in this study, their origin, GenBank accession numbers and results from cytogenetic analyses.

Species	Identifier	Origin[4]	GenBank accession numbers			FISH
			ETS	ITS	5S-NTS	rDNA loci 5S/45S
Hieracium alpinum[1]	alp.Ukr	Ukraine: Polonina Breskulska ridge	EU821408, EU867634	AJ633429	MW328890	
	H63-15-15	Ukraine: Mt. Bliznitsya	MW328990-91	MW325251-52	MW328939	2/4
	H63-30-7	Ukraine: Mt. Bliznitsya	n.d.	n.d.	n.d.	2/4
H. eriophorum	1221/1	France: dépt. Landes	EU821409, EU867639	MW315935	MW328891	
	1222/2	France: dépt. Landes	EU867640-41	MW315936, MW587333	n.d.	
	Bis11b	France: Biscarrosse-Plage	MW328992-93	MW325253-54	MW328940	2/4
H. intybaceum[2]	inb.Kaer	Austria: Turracher Höhe	EU867568, EU821370	AJ633426, KM372113	MW328892	
	1531/8	Austria: Arlbergpass	MW328994-95	MW325255-56	MW328941-42, MN784131	2/4[5]
	6/14/25	France: Col du Petit Saint-Bernard	MW328996-97	MW325257-58	MW328943-44, MN784130	2/4[5]
H. kittanae	1228/2	Bulgaria: central Rhodope Mts	EU821400, EU867622,	MW315937, MW587334	MW328893,	
			MW328998-99	MW325259-60	MW328945-46	
H. laniferum	lanif2	Spain: la Sénia	MK523499, MW591759	MW315938, MW587335	MW328894	
H. lucidum	H. lucidum	Italy: Sicily, distr. Palermo	EU867592-93	MW315939, MW587336	MW328895	
	Hluc_1-1-2	Italy: Sicily, Mt. Gallo	MW329000-01	MW325261-62	MW328947-48	*/6
H. petrovae	1229	Bulgaria: central Rhodope Mts	EU821403, EU867625, MW328989	MW325265, MW587337	MW328949	2/4
H. plumulosum	1218/2	Montenegro: canyon of Mrtvica river	FJ858097, FJ858105, FJ858108, FJ858110, MW329002-03	MW325263-64, MW315940	MW328950	
H. pojoritense	PM2012	Romania: Pojorita	MW328988	MW325266	MW328951-52	2/4
	poi.Rom.1	Romania: Pojorita	EU867635-36, MK523506-07	AJ633412, MW587338	n.d.	
H. porrifolium	1052/9	Austria: Carinthia, Karawanken Mts	EU821407, EU867631	MW315941, MW587339	MW328896	
	Hpor_1-14-2	Slovenia: Podljubelj	MW329004-05	MW325267-68	MW328953-54	
	Hpor_1-14-1	Slovenia: Podljubelj	n.d.	n.d.	n.d.	2/*
	H1463	Slovenia: Julijske Alpe, Spodnja Trenta	n.d.	n.d.	n.d.	*/4
H. prenanthoides	1252	France: La Grave	EU821377, EU867579	MW315942, MW587340	MW328897	
	JC1513-3	France: Villarodin & Modane	n.d.	n.d.	n.d.	*/6[6]
	pre_6/5/5	Italy: Claviere	MW329006-07	MW325269-70	MW328955-56, MN784129	2/6[5]
	pre_6/5/2	Italy: Claviere	n.d.	n.d.	n.d.	2/6[6]
	pre_6/8/5	Italy: Claviere	MW329008-09	MW325271-72	MW328957, MN784128	2/6[5]
	pre_6/4/5	Italy: Claviere	n.d.	n.d.	n.d.	2/6
H. recoderi	1174/4	Spain: Catalunya, prov. Barce	EU821386, EU867603	MW315943, MW587341	MW328898	
H. sparsum	1251/1	Bulgaria: Sofia, Vitoša Mts	EU821404, EU867626	MW315944. MW587342	n.d.	
	spa.sst.2	Bulgaria: Pirin Mts, Vihren	EU867627-28	AJ633431, MW587343	MW328899	
	spa1611/5	Bulgaria: Pirin Mts, Vihren	MW329010-11	MW325273-74	MW328958-59	2/6
	spa1611/6	Bulgaria: Pirin Mts, Vihren	n.d.	n.d.	n.d.	2/6
	PM2099	Bulgaria: Rila Mts, Maljovica	n.d.	n.d.	n.d.	3/6
	PM2102	Bulgaria: Rila Mts, Maljovica	n.d.	n.d.	n.d.	2/6
H. stelligerum	1233/1	France: Vallon Pont d'Arc	EU821383, EU867597	MW315945, MW587344	MW328900	
	Hstel_3-2-1	France: Thueyts	MW329012-13	MW325275-76	MW328960-61	2/7

(Continued)

TABLE 1 | Continued

Species	Identifier	Origin[4]	GenBank accession numbers			FISH
			ETS	ITS	5S-NTS	rDNA loci 5S/45S
H. tomentosum	1066/8	France: valley of la Roya	EU821382, EU867596	MW315946, MW587345	MW328901	
H. transylvanicum[3]	tra.Boa	Romania: Borşa	EU867570-71	MW315947, MW587346	MW328902	
	1077/7	Ukraine: Oblast Zakarpatska	EU821372, EU867572, MW329014-15	MW587347-48 MW325277-78	MW328962	
	Htrans_2-2-1	Romania: Băile Tuşnad	MW328975, MW591760	MW315948	MW328903	2/4
H. umbellatum	1021/1	Poland: Województwo pomorskie	EU821410, EU867642	MW315949, MW587349	MW328904	
	um.AM.1	Germany: Schönau-Berzdorf	EU867643-44	KM372116	MW328905	
	H1617	Czechia: Praha, Troja	MW329016-17	MW325279-80	MW328963-64	2/4
	UMB 8/9/3	Slovakia: Prakovce	MW328976, MW591761	MW315950, MW587350	MW328906	2/3
H. vranceae	Hvran_1-1	Romania: Lepşa	MK523515, MW591762	MW315951, MW587351	MW328907	
	PM2013	Romania: Lepşa	n.d.	n.d.	n.d.	2/4[7]
Pilosella alpicola	pic1141	Slovakia: Vysoké Tatry Mts	MW328977, MW591763	AJ633401	MW328908	
P. angustifolia	ang.Fra	France: dép. Hautes-Alpes	MW328978, MW591764	AJ633407	MW328909-13	
P. argyrocoma	agy.Gra	Spain: Prov. Granada	KM372001	MW315952	MW328914	
P. breviscapa	brc.Bou	France: Lac de Bouillouses	MW328979, MW591765	AJ633393	MW328915	
P. castellana	cas.Nev	Spain: Sierra Nevada	MW328980, MW591766	AJ633392	MW328916	
P. cymosa	cym.12/4	Czechia: Louny	MW328981, MW591767	AJ633398	MW328917	
P. echioides	H1701/2	Czechia: Praha-Čimice	MW329018-19	MW325287-88	MW328965	2/4
P. hoppeana	H1702/1	Austria: Carinthia, Hohe Tauern	MW329020-21	MW325281-82	MW328966-67	2/4
P. lactucella	lac.Jon.1	Germany: Oberlausitz, Jonsdorf	KM372002	AJ633389	MW328918	
	lac.Neu.2	Germany: Erzgebirge, Neuwernsdorf	MW329022-23	MW325283-84	MW328968-69	
	Zebra	Czechia: Světlá nad Sázavou	n.d.	n.d.	n.d.	2/4
P. onegensis	caeb.Jbo.2	Czechia: Krkonoše Mts	MW328982, MW591768	AJ633396	MW328919	
	H1704	Czechia: Krkonoše Mts, distr. Trutnov	MW329024-25	MW325285-86	MW328970-71	2/4
P. pavichii	pav.Oly	Greece: Mt. Olympos	MW328983, MW591769	AJ633400	MW328920	
P. peleteriana	pel.Wal	Switzerland: Kanton Wallis	MW328984, MW591770	AJ633504	MW328921	
P. pseudopilosella	pse.Civ	Spain: prov. Civdad Real	MW328985, MW591771	AJ633390	MW328922	
P. vahlii	vah.Sor	Spain: prov. Soria	MW328986, MW591772	AJ633394	MW328923	
Hispidella hispanica	His.his.2	Spain: Sierra de Guadarrama	EU821365-66	KM372107	MW328924-28	
Andryala agardhii	JC 2011/31/1	Spain: Andalusia, Calar del Desabezedo	KM371905-06	KM372009-10	MW328929	
	A.agaJF	Spain: Sierra Nevada	MW328987, MW591773	MW315953	MW328930	
	PM2390	Spain: garden culture, origin unknown	n.d.	n.d.	n.d.	2/4

(Continued)

TABLE 1 | Continued

Species	Identifier	Origin[4]	GenBank accession numbers			FISH
			ETS	ITS	5S-NTS	rDNA loci 5S/45S
A. glandulosa	A.glan.Mad.1	Portugal: Madeira, Ponta do Pargo	MW329026-27, KM371929-30	MW325289-90, KM372033-34	MW328931	*
	ZF 233	Portugal: Madeira, Seixal	KM371933-34	KM372037-38	MW328932	
A. integrifolia	AZ 4	Algeria: Alger, town distr. Le Caroubier	MW329028-29	MW325291-92	MW328972	*
	AZ 3/1	Algeria: Algiers, Kouba town district	KM371941-42	KM372045-46	MW328933	
	JC 26/1	Spain: Andalusia, province Granada	KM371939-40	KM372043-44	MW328934	
A. laevitomentosa	Alev18	Romania: Pietrosul Bogolin	MW329030-31	MW325295-96	MW328973-74	*
	E8	Romania: Pietrosul Bogolin	KM371945-46	KM372049-50	MW328935-36	
A. pinnatifida	SB T2/1	Spain: Tenerife, Puerto de la Cruz	KM371981-82	KM372086-87	MW328937	
	And.pin.Cer	Spain: La Gomera, El Cercado	KM371971-72, MW329032-33	KM372076-77, MW325293-94	MW328938	

[1]*An additional sample of H. alpinum from a Romanian locality shows the same karyotype (Belyayev et al., 2018), and another sample, also from Romania, has a highly similar ETS sequence (Fehrer et al., 2009) as the Ukrainian material.*
[2]*An additional sample of H. intybaceum from Italy shows the same karyotype (Belyayev et al., 2018).*
[3]*Based on a broad geographic sampling, H. transylvanicum shows two loci of 5S and four loci of 45S rDNA without intraspecific variation (Ilnicki et al., 2010).*
[4]*For details, see* ***Supplementary Table 1****.*
[5]*From Chrtek et al., 2020.*
[6]*From Belyayev et al., 2018.*
[7]*From Mráz et al., 2019.*
n.d., not determined.
**, samples were collected late in the year and root tip quality was insufficient for evaluation or no further material available.*

2, and each read had to show at the position a minimum base quality of 30 in order to be counted (mpileup2cns –min-coverage 8 –min-reads2 2 –min-avg-qual 30). If valid SNPs were present, the phasing was performed with Samtools phase (-A -F -Q 30). The product of Samtools phase consists at most in two bam files that each correspond to a putative allelic variant. When generated, these files were further subjected to a second Samtools and VarScan round of analyses, and in presence of valid SNPs were further phased in order to produce a maximum of four alleles for each sample/marker combination. Putative chimeric alleles identified by Samtools phase were discarded.

ITS and ETS sequences were extracted from contigs of the entire 45S rDNA region. The phased sequences were aligned with Sanger sequenced samples of all sequences of the same species/individual in BioEdit v7.0.9.0 (Hall, 1999), including sequences with all polymorphic sites retained for comparison with the diversity of phased alleles. After inspection of the variation in each alignment, if more than two phased sequences per sample were found, the two most divergent ones accounting for the maximum of alternative character states at variable sites in ITS as well as ETS regions were chosen to represent the sample. Using the same phased sequence allowed to tentatively assign ITS and ETS allelic variants to each other. 5S-NTS sequences were treated in the same way. We use the following terminology for phased alleles: If only a single variant was found in our approach, the allele is referred to as 'single'. If two variants occurred, they are labeled 0 and 1. Four alleles (found in the second round of phasing) are designated as 0.0, 0.1 (phasing of the first main variant), 1.0 and 1.1 (phasing of the second variant). If only one new allele was retrieved in the second round of phasing (i.e., a total of three), their labels are 0, 1.0, 1.1 or 0.0, 0.1, 1.

Phylogenetic Analyses

Total alignments of ITS and ETS were produced in BioEdit and at first subjected to separate phylogenetic analyses to see if topologies were congruent and if phased sequences of both regions corresponded to each other (see Allele matching below). Later, ITS and ETS sequences were concatenated and analyzed together; if one of the regions consisted of a single sequence and the other was represented by two variants, this sequence was concatenated with both sequence variants of the other region. Maximum parsimony (MP), Maximum likelihood (ML), and Bayesian analyses (BA) were carried out using PAUP v4.0b10 (Swofford, 2002), IQ-TREE (Nguyen et al., 2015), and MrBayes v3.2.2 (Ronquist and Huelsenbeck, 2003), respectively. Prior to analysis, gaps were coded as additional characters in FastGap v1.2 (Borchsenius, 2009) using the simple method of Simmons and Ochoterena (2000).

Maximum parsimony analyses were computed as heuristic searches with 100 random addition sequence replicates and TBR branch swapping, saving no more than 100 trees with length greater than or equal to 1 per replicate, automatically increasing the maximum number of trees saved. Bootstrapping was done with the same settings and 1000 replicates, but without branch swapping. For ML and BA, sequence and gap data were treated as separate partitions, applying the GTR2 on the binary partition. Using the ModelFinder (Kalyaanamoorthy et al., 2017) tool of IQ-TREE, the best fitting molecular evolutionary models were determined for ML. The standard non-parametric bootstrap

was performed in IQ-TREE with 1000 replicates. For BA, the models best fitting the presumed molecular evolution of the respective datasets were determined with Modeltest v3.5 (Posada and Crandall, 1998) under the Akaike Information Criterion. Models found were TVM + Γ (ETS) and GTR + Γ (ITS, 5S-NTS, combined ITS + ETS). The basic model parameters, i.e., gamma distribution of rates among sites and six different substitution rates, were set as priors for each analysis; apart from that, the default settings were used. Chains were computed for 2 million generations, sampling every 1000th tree; all indicators suggested that convergence between the different runs was achieved for all datasets. The first 25% of the trees per run were discarded as burn-in and the remaining trees were summarized.

In order to characterize the cause of discordance within and between datasets we carried out a Quartet Sampling (QS) analysis (Pease et al., 2018) with 1000 replicates, implemented in the quartetsampling software[1]. A QS analysis provides for each branch three complementary measures: (1) The Quartet Concordance (QC) score that quantifies the support among the three possible resolutions of four taxa. (2) The Quartet Differential (QD) score that measures the disparity between the sampled proportions of the two discordant topologies. QD is only applicable to branches where resampling produces alternative topologies to the input tree. (3) The Quartet Informativeness (QI) score quantifies the proportion of replicates where the best-likelihood quartet has a likelihood score that exceeds the score of the second best quartet. Therefore, these three measures provide an overview of the structure of the topological conflict distinguishing between uninformative branches (signaled by QI) and the branches characterized by conflicting information (QC and QD).

Allele Matching

We define 'allele phasing' as the process of grouping reads according to their shared polymorphisms in order to reconstruct the sequence variants they originate from. Contrary to haplotype phasing, which aims at separating the alleles of the same gene located on different homologous chromosomes, the phasing process in our case groups reads that derive from the same homogenized variants. As a consequence, each phased sequence (termed 'allele' here for simplicity) does not necessarily match a single genomic unit, but might represent a majority consensus of several units.

As described above, we mapped the reads from each marker to a single reference sequence and separated the alleles during phasing. However, due to the high level of conservation in the intervening 18S region, there were no polymorphisms located between the ITS and ETS domains that would allow connecting the two regions into a single allele using overlapping pair-reads. Consequently, the phasing could lead to *in silico* recombined 45S alleles where the ITS and ETS regions would not share a common history. Furthermore, because of allele loss, unsampled loci and differential rates of homogenization, we observed a frequent unbalance between the number of ITS and ETS alleles retrieved in a given sample. In order to perform cogent comparisons between the pair of trees derived from the two markers, we designed an algorithm that returns for each accession the best allelic combinations between ITS and ETS sequences. The objective function used to assess the combinations is the distance between the two trees after swapping the leaves' labels so that they correspond to the selected combination. During the optimization phase, the algorithm searches for the combinations that produce the most similar trees which correspond to the shortest distance between the two trees. The algorithm pre-processes the trees by pruning them to the accessions they have in common, and in each tree, it collapses sister alleles to a single branch whose length corresponds to one of the alleles. Clades made up exclusively of alleles from the same accession are reduced into a single branch that is set to the length of the most basal allele. As coalescent groups of alleles do not provide information for selecting an optimal solution, their removal reduces the combinatorial load. Because the search space grows exponentially with the number of alleles, the number of possible combinations could prove prohibitive in the case of large trees. As a consequence, we designed a heuristic function, which selects the set of the most favorable solutions among all possible solutions for further optimization. The function firstly performs a local optimization that selects a set of optimal pairings for each accession, thus reducing the search space. The possible pairings are then combined during a global optimization, which completes an exhaustive comparison of all remaining combinations using all accessions. For the local optimization, it compares the pre-processed ITS and ETS trees derived from each combination using the Robinson–Foulds (Robinson and Foulds, 1981) distance (aka symmetric distance), which only takes into account the tree topology. The algorithm proceeds by randomly assigning the alleles for all the accessions. Then for each focal accession, all possible combinations are tested while retaining the random combination for the non-focal accessions. For each focal accession, the combinations that minimize the distance between the two modified trees are retained. The product of the best local combinations is then evaluated using a modified Robinson-Foulds metric on rooted trees: Each possible split is weighted by the length of the corresponding branch and by the support of the child node connected to the branch, a support lesser than 50% leads to the removal of the associated split from the distance calculation. The algorithm has been implemented in a Python 3 based software that relies on the Dendropy library (Sukumaran and Holder, 2010). The novel tool (allele_linker) is available at https://git.sorbus.ibot.cas.cz/allele_linker/allele_linker.

Cytogenetic Experiments and Ancestral State Reconstruction of Locus Numbers

FISH with 45S and 5S rDNA probes was performed as described in Belyayev et al. (2018). The number of 45S and 5S rDNA loci is summarized in **Table 1**. We refer throughout the manuscript to the total number of loci per diploid genome, corresponding to the number of FISH signals.

For the number of 45S rDNA loci, ancestral state reconstruction was performed based on the combined ITS/ETS tree, either omitting taxa for which locus numbers were unknown

[1]https://www.github.com/fephyfofum/quartetsampling

or treating them as missing data. ITS is the molecular marker that reflects species relationships in the Hieraciinae best, in accordance with morphology and other evidence (Fehrer et al., 2007a). ETS is the closest approximation of the species tree in *Hieracium* as relationships are in keeping with geographic distribution and genome size (Chrtek et al., 2009; Krak et al., 2013). The combined tree is therefore, despite a lack of resolution in some parts, the best estimate of the species tree. Besides, it is interesting to reconstruct the evolution of 45S rDNA locus numbers on a tree that is based on sequences contained in this locus. Evolution of 5S rDNA locus numbers was not investigated, because they were uniform except for a single sample (see below).

We performed a maximum likelihood reconstruction of ancestral states as a function of stochastic character mapping (SCM) (Huelsenbeck et al., 2003) in R v4.0.3 (R Core Team, 2020). The {phytools} package v0.7-70 (Revell, 2012) was used to project the number of 45S rDNA loci onto the ML tree. The tree was mid-point rooted. It was time-calibrated using the semiparametric penalized likelihood method implemented in the chronopl function of the {ape} package v5.4-1 with a smoothing parameter of 1 (Sanderson, 2002; Paradis et al., 2004). The three usual transition models (ER – equal rates model; SYM – symmetrical model; ARD – all-rates-different model) were compared by computing their corrected Akaike information criterion (AICc) scores. The best-fitting model for character transformation was the ER model (see **Supplementary Table 2**). Several other custom models that ordered and/or oriented the state transitions were also tested; as they produced identical state reconstructions as the ER model, they will not be further discussed. The character state for specimens that lack a locus count was treated as missing. We reconstructed all changes across the tree based on transitions between the states at each node using the fitDiscrete function of {geiger} package v2.0.7 (Harmon et al., 2008) and mapped them on the ultrametric tree. The magnitude of phylogenetic signal contained in 45S rDNA loci data was evaluated after pruning terminal branches that harbored leaves without locus count. The signal was assessed with Blomberg's *K* statistics (Blomberg et al., 2003) using the phylosignal function from the {picante} package v1.8.2 (Kembel et al., 2010) and with Pagel's λ (Pagel, 1994) using the fitDiscrete function with the ER model. Pagel's λ was computed with 1000 iterations for the pruned tree and for a rescaled tree (no signal model) where all branches were collapsed into a single polytomy. The strength of the phylogenetic signal contained in the locus data was evaluated by comparing the AICc scores for both models.

RESULTS

Comparison of Genome Skimming and Sanger Sequencing

Individual alignments for each species showed that polymorphic sites inferred from direct sequencing corresponded very well to the resolved character states of the phased alleles (not shown). In ITS, ETS, and 5S-NTS trees, the position of phased sequences from genome skimming was compared with Sanger sequenced samples of the same species or individual, the latter represented by either major (usually partly polymorphic) or, in some cases (divergent variants), cloned sequences.

Only in a few cases (*Andryala laevitomentosa*, *Hieracium intybaceum*, *H. porrifolium,* and *H. transylvanicum*), phased alleles and direct sequences of the same species were coalescent. In the ITS tree (**Figure 1**), phased alleles of several samples (*Pilosella echioides*, *A. integrifolia*, *H. lucidum*, *H. stelligerum* and two accessions of *H. prenanthoides*) were more divergent from each other than different species in their respective clades. Direct sequences of the same sample or species either clustered with one of the phased sequences (*P. onegensis*, *H. alpinum*) or occupied intermediate or basal positions (*H. intybaceum*, *H. kittanae*, *H. lucidum*, *A. pinnatifida*). The same was true for the ETS tree (**Figure 2**): Divergent alleles (phased sequences) that were more similar to other species occurred (in *P. echioides*, *P. onegensis*, *P. hoppeana*, *H. alpinum,* and *H. kittanae*), intermediate or basal positions of direct sequences were assumed by some species (*H. intybaceum*, *H. alpinum*, and *H. prenanthoides*), or direct sequences clustered with one of the phased alleles (in *A. pinnatifida*, *H. kittanae,* and *H. lucidum*). Worth mentioning is *H. plumulosum*, which was shown to possess four ETS variants, two occurring in the western European clade and two in the eastern European clade (Fehrer et al., 2009). The eastern variants were found among the phased alleles and clustered with two representative clones (12 and 13, the latter clone was recombinant, therefore, only the unique part of this sequence was included in the analyses) whereas the western types (represented by clones 1 and 7) were not retrieved, but only a phased sequence that was recombinant with western clade sequences was detected (**Figure 2**, inset).

Furthermore, it is likely that phased alleles of ITS and ETS were recombined during the mapping. For example, the ETS sequence of *H. kittanae* 0.0, but the ITS sequence of *H. kittanae* 1.0 clustered with *H. petrovae*. Likewise, *P. echioides* 1 clustered with *P. pavichii* in the ITS tree, but with *P. onegensis* 0.0 in the ETS tree. Therefore, before concatenating ITS and ETS sequences for combined analyses, we tested for *in silico* recombination of the phased alleles using allele_linker (see below).

5S-NTS alleles were generally less divergent than ITS or ETS alleles (**Table 2**); often, only a single variant was found in sequences retrieved from genome skimming, and direct sequences were fairly homogenous as well. Species in which phased alleles and direct sequences formed coalescent groups were *A. laevitomentosa*, *H. alpinum*, *H. sparsum*, *H. kittanae,* and *H. transylvanicum* (**Figure 3**). In *P. hoppeana*, one of the phased alleles was more similar to other *Pilosella* species than to the second allele of the same individual, the same held for *H. lucidum*. One sample of *A. integrifolia* (JC 26/1) did not group with other samples of the same species. *Pilosella angustifolia* and *Hispidella hispanica* showed large indels in direct sequencing. Three of the cloned sequences of *P. angustifolia* grouped together, the other two occurred in unresolved positions among other *Pilosella* species. Four cloned sequences of *Hispidella* formed a well-supported branch, but another cloned sequence inserted near the base of genus *Hieracium*.

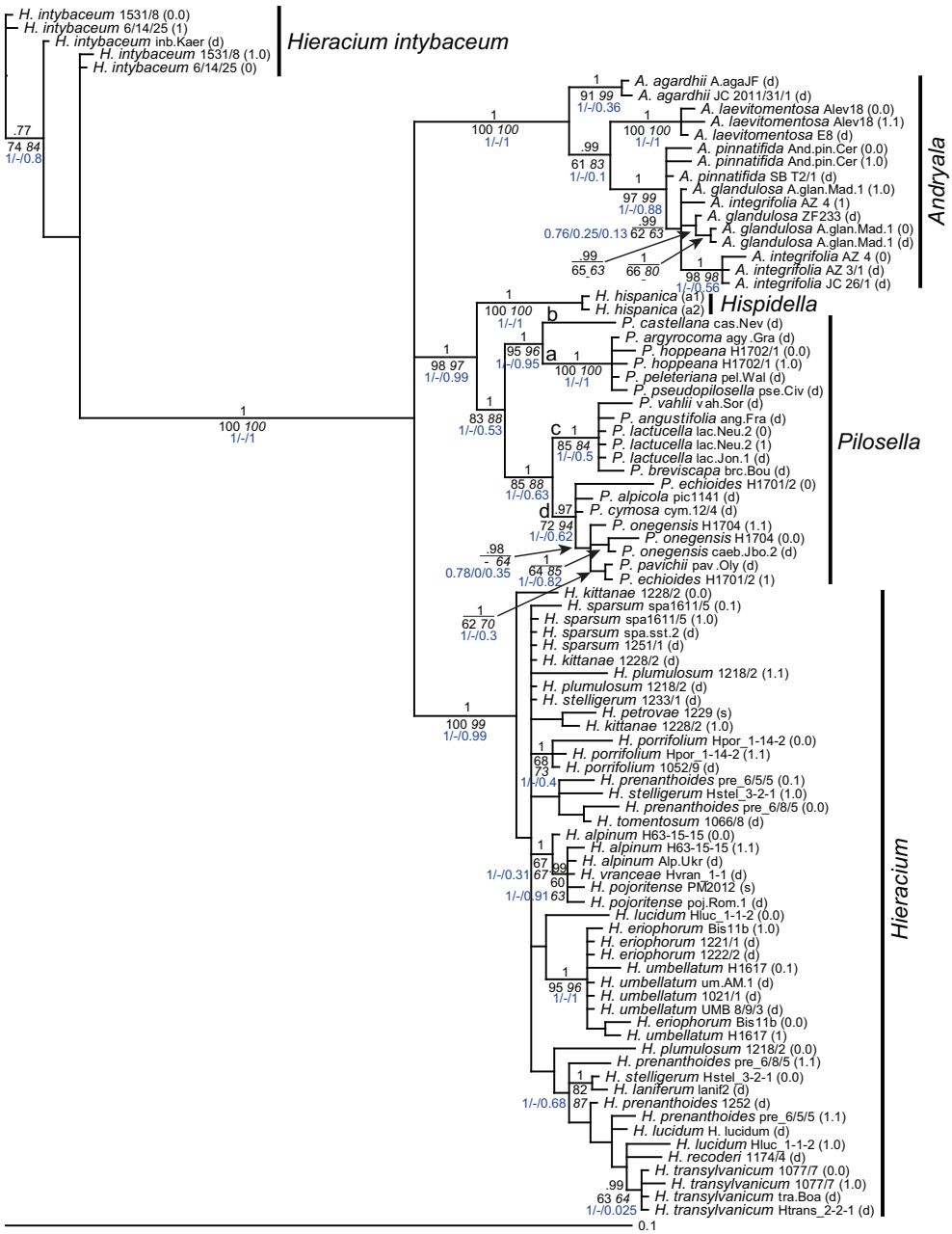


FIGURE 1 | Phylogenetic analysis of the Hieraciinae based on the ITS region. The Bayesian consensus tree is shown with posterior probabilities (pp) above branches and boostrap support (bs) from MP (regular) and ML (italics) analyses below branches. Values are only shown if pp was > 0.94 or bs > 70%. Below the support values, Quartet Concordance/Quartet Differential/Quartet Informativeness scores for 1000 replicates of the full alignment are displayed (in blue). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single); d, direct sequence; a1/a2, two alleles of *Hispidella* occur in ETS sequences, and ITS sequences were duplicated here. a-d, main lineages of *Pilosella*. Accession labels correspond to **Table 1**.

Comparison of Tree Topologies Based on ITS and ETS

Generally, species relationships based on both rDNA regions were fairly similar and in agreement with previous findings (Fehrer et al., 2007a, 2009; Ferreira et al., 2015; Mráz et al., 2019). These concern the outgroup position of *Hieracium intybaceum*, the sister relationship of *Hispidella* and *Pilosella*, a main clade of all *Andryala* species except *A. agardhii* and *A. laevitomentosa*, four lineages within *Pilosella* with identical species compositions, a joint clade of *H. umbellatum* and *H. eriophorum* (with ETS also containing an allele of one sample of *H. pojoritense*). For all these clades, the basal branch displayed a perfect QS score (i.e., 1/-/1), indicative of high data informativeness and no conflict. A clade comprising *H. alpinum*, all other alleles of *H. pojoritense*, and

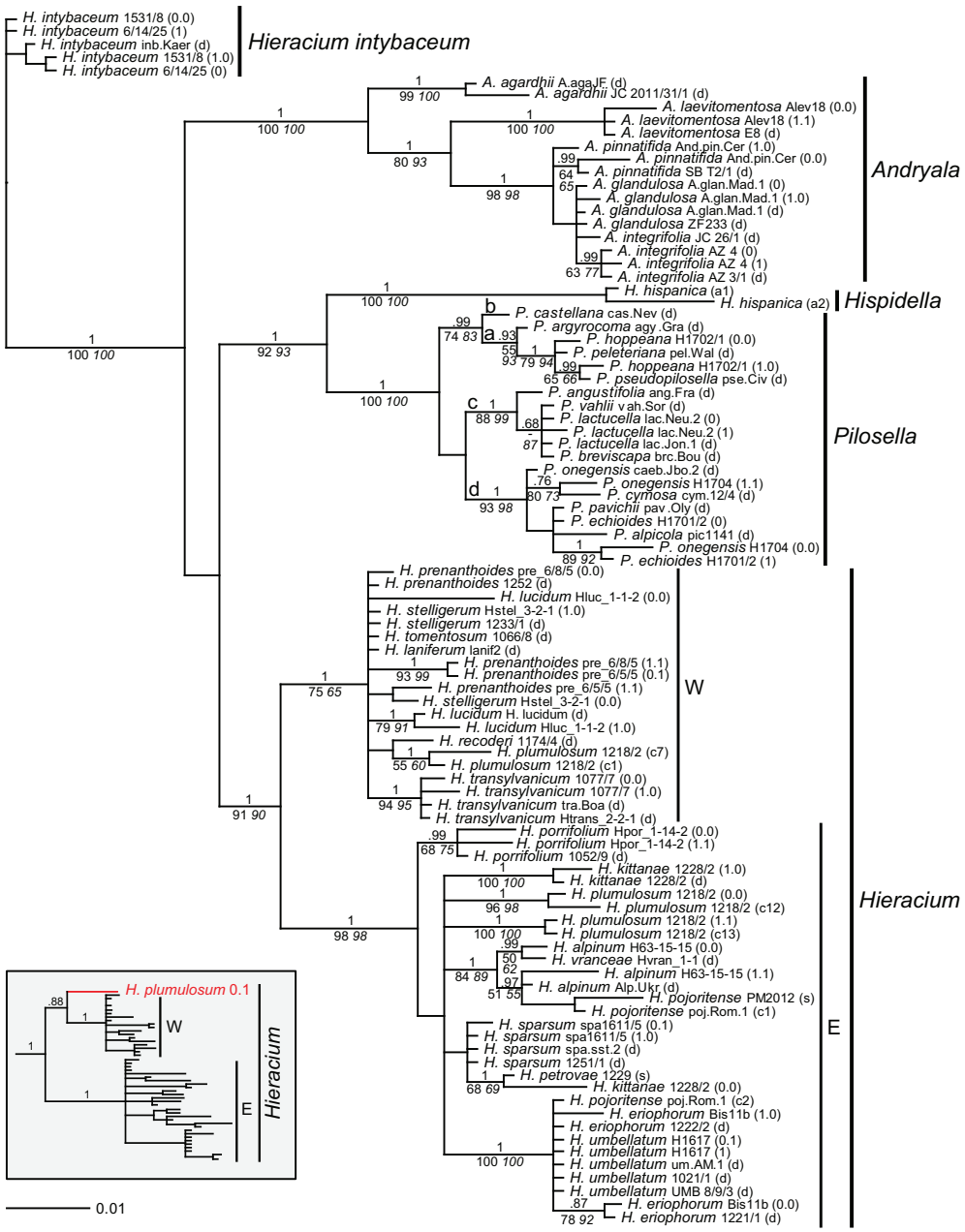


FIGURE 2 | Phylogenetic analysis of the Hieraciinae based on the ETS region. The Bayesian consensus tree is shown with posterior probabilities (pp) above branches and boostrap support (bs) from MP (regular) and ML (italics) analyses below branches. Values are only shown if pp was > 0.94 or bs > 70%. Below the support values, Quartet Concordance/Quartet Differential/Quartet Informativeness scores for 1000 replicates of the full alignment are displayed (in blue). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single); d, direct sequence; a1/a2, two alleles of *Hispidella* (minor and major sequence inferred from direct sequencing); c, cloned sequence. W, E, western and eastern European clades of *Hieracium*. a-d, main lineages of *Pilosella*. Accession labels correspond to **Table 1**. The inset shows the position of a recombinant phased allele of *H. plumulosum*.

H. vranceae was also recovered in both trees, but its basal branch showed slightly less information without conflicting quartets with QS of 1/-/0.91 for the ITS and 1/-/0.69 for the ETS tree. Also, species relationships within *Hieracium* remained largely unresolved with ITS whereas a marked separation into two groups of mainly western or eastern European origin was found with ETS. However, the western clade is poorly supported by MP and ML bootstrap values (75 and 65 respectively) and the QS analysis shows that only a week majority of quartets supports the corresponding branch (QC = 0.062), which can be resolved equally into any of the possible topologies (QD = 0) despite that the data contains a rather high phylogenetic signal (QI = 0.55). According to Pease et al. (2018), this QS configuration is indicative of a rapid radiation or a highly complex conflict.

TABLE 2 | Number of polymorphic sites in direct sequences of Hieraciinae and diversity of phased or cloned sequences.

Species	Identifier	No. of polymorphic sites (d) or substitutions (ph, c)				No. of 45S
		ETS	ITS	ETS + ITS	5S-NTS	rDNA loci
Hieracium alpinum	alp.Ukr	3/4 (d)	3/3 (d)	6/7 (d)	0 (d)	
	H63-15-15	8 (ph)	2 (ph)	10 (ph)	0 (ph)	4
H. eriophorum	1221/1	5/6 (d)	3/3 (d)	8/9 (d)	7 (d)	
	1222/2	7/8 (d)	0/5 (d)	7/13 (d)	n.d.	
	Bis11b	7 (ph)	5 (ph)	12 (ph)	0 (ph)	4
H. intybaceum	inb.Kaer	1/4 (d)	0/8 (d)	1/12 (d)	0 (d)	
	1531/8	1 (ph)	6 (ph)	7 (ph)	2 (ph)	4
	6/14/25	1 (ph)	6 (ph)	7 (ph)	1 (ph)	4
H. kittanae	1228/2	0/15 (d), 17 (ph)	5/9 (d), 9 (ph)	5/24 (d), 26 (ph)	1 (d), 1 (ph)	
H. laniferum	lanif2	1/2 (d)	1/11 (d)	2/13 (d)	8 (d)	
H. lucidum	H. lucidum	1/13 + indel (d)	2/30 (d)	3/43 + indel (d)	3 (d)	
	Hluc_1-1-2	13 + indel (ph)	17 (ph)	30 + indel (ph)	4 (ph)	6
H. petrovae	1229	1/7 (d), 0 (ph)	0/10 (d), 0 (ph)	1/17 (d), 0 (ph)	0 (ph)	4
H. plumulosum	1218/2	35 + indel (ph + c)	21 (d), 19 (ph)	56[2] + indel	0 (ph)	
H. pojoritense	PM2012	0 (ph)	0 (ph)	0 (ph)	2 (ph)	4
	poi.Rom.1	1/26 (d), 11 (c)	2/21 + indel (d)	3/48 + indel (d)	n.d.	
H. porrifolium	1052/9	5/6 (d)	2/3 (d)	7/9 (d)	2 (d)	4*
	Hpor_1-14-2	7 (ph)	7 (ph)	14 (ph)	3 (ph)	4*
H. prenanthoides	1252	3/8 (d)	18/23 (d)	21/31 (d)	4 (d)	
	pre_6/5/5	8 (ph)	13 (ph)	21 (ph)	0 (d), 1 (ph)	6
	pre_6/8/5	6 (ph)	15 (ph)	21 (ph)	0 (d), 0 (ph)	6
H. recoderi	1174/4	2/5 (d)	1/4 (d)	3/9 (d)	4 (d)	
H. sparsum	1251/1	3/12 (d)	8/9 + indel (d)	11/21 + indel (d)	n.d.	
	spa.sst.2	1/12 (d)	4/6 (d)	5/18 (d)	0 (d)	
	spa1611/5	1 (ph)	4 (ph)	5 (ph)	3 (ph)	6
H. stelligerum	1233/1	1/3 (d)	18/21 (d)	19/24 (d)	1 (d)	
	Hstel_3-2-1	2 (ph)	13 (ph)	15 (ph)	1 (ph)	7
H. tomentosum	1066/8	0/6 (d)	1/15 (d)	1/21 (d)	1 (d)	
H. transylvanicum	tra.Boa	0/5 (d)	0/2 (d)	0/7 (d)	0 (d)	
	1077/7	0/7 (d), 3 (ph)	0/3 (d), 2 (ph)	0/10 (d), 5 (ph)	0 (ph)	
	Htrans_2-2-1	0/3 (d)	1/1 (d)	1/4 (d)	0 (d)	4
H. umbellatum	1021/1	2/5 (d)	7/8 (d)	9/13 (d)	0 (d)	
	um.AM.1	0/5 (d)	6/6 (d)	6/11 (d)	1 (d)	
	H1617	0 (ph)	6 (ph)	6 (ph)	2 (ph)	4
	UMB 8/9/3	0/4 (d)	6/8 (d)	6/12 (d)	0 (d)	3
H. vranceae	Hvran_1-1	0/1 (d)	2/4 + indel (d)	2/5 + indel (d)	1 (d)	4*
Pilosella alpicola	pic1141	0/3 (d)	0/n.a. (d)	0/ ≥ 3 (d)	0 (d)	
P. angustifolia	ang.Fra	2/6 + indel (d)	0/n.a. (d)	2/ ≥ 6 + indel (d)	10 + 3 indels (c)	
P. argyrocoma	agy.Gra	6/6 (d)	7/7 (d)	13/13 (d)	0 (d)	
P. breviscapa	brc.Bou	4/5 (d)	0/n.a. (d)	4/ ≥ 5 (d)	1 (d)	
P. castellana	cas.Nev	4/6 (d)	0/n.a. (d)	4/ ≥ 6 (d)	1 (d)	
P. cymosa	cym.12/4	3/10 (d)	0/n.a. (d)	3/ ≥ 10 (d)	6 (d)	
P. echioides	H1701/2	3 (ph)	10 (ph)	13 (ph)	0 (ph)	4
P. hoppeana	H1702/1	3 (ph)	3 (ph)	6 (ph)	4 (ph)	4
P. lactucella	lac.Jon.1	1/1 (d)	0/n.a. (d)	1/ ≥ 1 (d)	2 (d)	4*
	lac.Neu.2	1 (ph)	0 (ph)	1 (ph)	2 (ph)	4*
P. onegensis	caeb.Jbo.2	6/8 (d)	2/n.a. (d)	8/ ≥ 10 (d)	1 (d)	
	H1704	12 (ph)	5 (ph)	17 (ph)	1 (ph)	4
P. pavichii	pav.Oly	0/5 (d)	0/n.a. (d)	0/ ≥ 5 (d)	3 (d)	
P. peleteriana	pel.Wal	0/2 (d)	1/n.a. (d)	1/ ≥ 3 (d)	0 (d)	
P. pseudopilosella	pse.Civ	1/3 (d)	0/n.a. (d)	1/ ≥ 3 (d)	0 (d)	
P. vahlii	vah.Sor	2/4 (d)	0/n.a. (d)	2/ ≥ 4 (d)	0 (d)	
Hispidella hispanica	His.his.2	6 + indel (d)[1]	1/1 (d)	7 + indel (d)	53 + 4 indels (c)	

(Continued)

TABLE 2 | Continued

Species	Identifier	No. of polymorphic sites (d) or substitutions (ph, c)				No. of 45S
		ETS	ITS	ETS + ITS	5S-NTS	rDNA loci
Andryala agardhii	JC 2011/31/1	0/10 (d)	0/13 (d)	0/23 (d)	0 (d)	4*
	A.agaJF	6/9 (d)	11/11 (d)	17/20 (d)	1 (d)	4*
A. glandulosa	A.glan.Mad.1	0/2 (d), 1 (ph)	0/2 (d), 2 (ph)	0/4 (d), 3 (ph)	0 (d)	
	ZF 233	0/1 (d)	1/3 (d)	1/4 (d)	0 (d)	
A. integrifolia	AZ 4	1 (ph)	8 (ph)	9 (ph)	0 (ph)	
	AZ 3/1	1/2 (d)	0/11 (d)	1/13 (d)	0 (d)	
	JC 26/1	0/1 (d)	1/1 (d)	1/2 (d)	0 (d)	
A. laevitomentosa	Alev18	3 (ph)	3 (ph)	6 (ph)	1 (ph)	
	E8	1/4 (d)	2/3 (d)	3/7 (d)	1 (d)	
A. pinnatifida	SB T2/1	0/1 (d)	1/4 (d)	1/5 (d)	0 (d)	
	And.pin.Cer	0/6 (d), 5 (ph)	1/3 (d), 4 (ph)	1/9 (d), 9 (ph)	0 (d)	

For direct ITS and ETS sequences, the number of polymorphic sites is provided for the major sequence/for the sequence including all polymorphisms. Sequence reads of 5S-NTS did in many cases not allow reliable identification of smaller additional peaks due to polynucleotide runs and high GC content. However, very low numbers of substitutions between phased alleles of the same sample indicate that there is indeed not much hidden variation.

d, direct sequence; ph, phased alleles; c, cloned sequences; n.d., not determined; n.a., number of all polymorphic sites not available (sequences provided by a collaborator).

**, locus number determined for another individual and assigned to the species.*

[1]Major and minor variant inferred from direct sequencing.

[2]The maximal number of substitutions is given.

The combined variation of ITS and ETS sequences and the number of corresponding 45S rDNA loci are shown. All sequenced samples have two loci of 5S rDNA.

Combining ITS and ETS Sequences for Phylogenetic Analysis

To account for recombination in the 18S rDNA, the newly designed software run using the ML trees indicated phased ITS and ETS sequences of the following samples should be swapped: *P. echioides*, *H. kittanae*, *H. stelligerum*, *H. eriophorum,* and *H. prenanthoides* pre_6/5/5. The algorithm further suggested to swap alleles of both *H. intybaceum* accessions, which is equivalent to changing none of them and was therefore dismissed, and to swap the alleles of *H. porrifolium*. Because these formed a well-supported branch together with the direct sequence of another accession of the same species in BA of ITS and ETS trees without further resolution of their relationships, swapping of the alleles is irrelevant. Equivocal results (no preference for swapping or not swapping of alleles) were obtained for *H. alpinum*, *H. prenanthoides* pre_6/8/5, *A. integrifolia* and *H. umbellatum*. Alleles of the latter three were not swapped, because nothing indicated that either solution was better for the combined analysis, however, *H. alpinum* allele 0.0 grouped with *H. vranceae* in the ETS tree whereas allele 1.1 grouped with *H. vranceae* in the ITS tree with high support in BA. Therefore, swapping of *H. alpinum* alleles was performed as well.

After swapping of these six allele pairs, combined analyses were performed (omitting *H. plumulosum* and clone 2 of *H. pojoritense* poi.Rom.1). Expectedly, the resolution of the tree was enhanced (**Supplementary Figure 1**) as indicated by higher support values, but no additional species relationships compared to individual analyses of ITS and ETS sequences were found. It is noteworthy that combining the two markers greatly boosts the phylogenetic signal for the western clade in *Hieracium*, which leads to higher MP and ML bootstrap support values (95 and 88 respectively) and a near perfect QS score (1/-/0.67) clear of topological conflict.

Phylogenetic Analysis Based on the 5S-NTS and Comparison With ITS/ETS Tree Topology

The most striking difference of the 5S-NTS tree (**Figure 3**) compared to those based on ITS and ETS was the position of *Hieracium intybaceum*, which did not form an outgroup to the rest of the Hieraciinae, but clustered with several *Hieracium* species of western European origin. The high support values for the western European B clade containing *H. intybaceum* in the former tree and the taxon's outgroup position of *H. intybaceum* in the latter tree point toward a genuine phylogenetic signal as the cause of this topological conflict. QS scores corroborate this conclusion by displaying a perfect score for the branch leading to the ingroup in the ITS/ETS tree (**Supplementary Figure 1**). Although several branches within the B clade possess low QI scores (0.07) in the 5S-NTS tree, no alternative topology is supported (QC of 1 and QD undefined) for any branch between the outgroup and the B clade.

Further differences of 5S-NTS tree topology compared to ITS/ETS topology were as follows: With *Pilosella*, only one of the four lineages found previously was retrieved (a); the other three (b–d) formed a single clade with mostly unresolved species relationships. The majority of cloned *Hispidella* sequences was sister to both *Pilosella* lineages. It is tantalizing to interpret the position of *H. hispanica* (c3) in the 5S-NTS tree as caused by a low phylogenetic signal (QI score of 0.11 for the branch grouping the sequence with the A clade). However, as no alternative topology is supported on any branch leading to this terminal (QC of

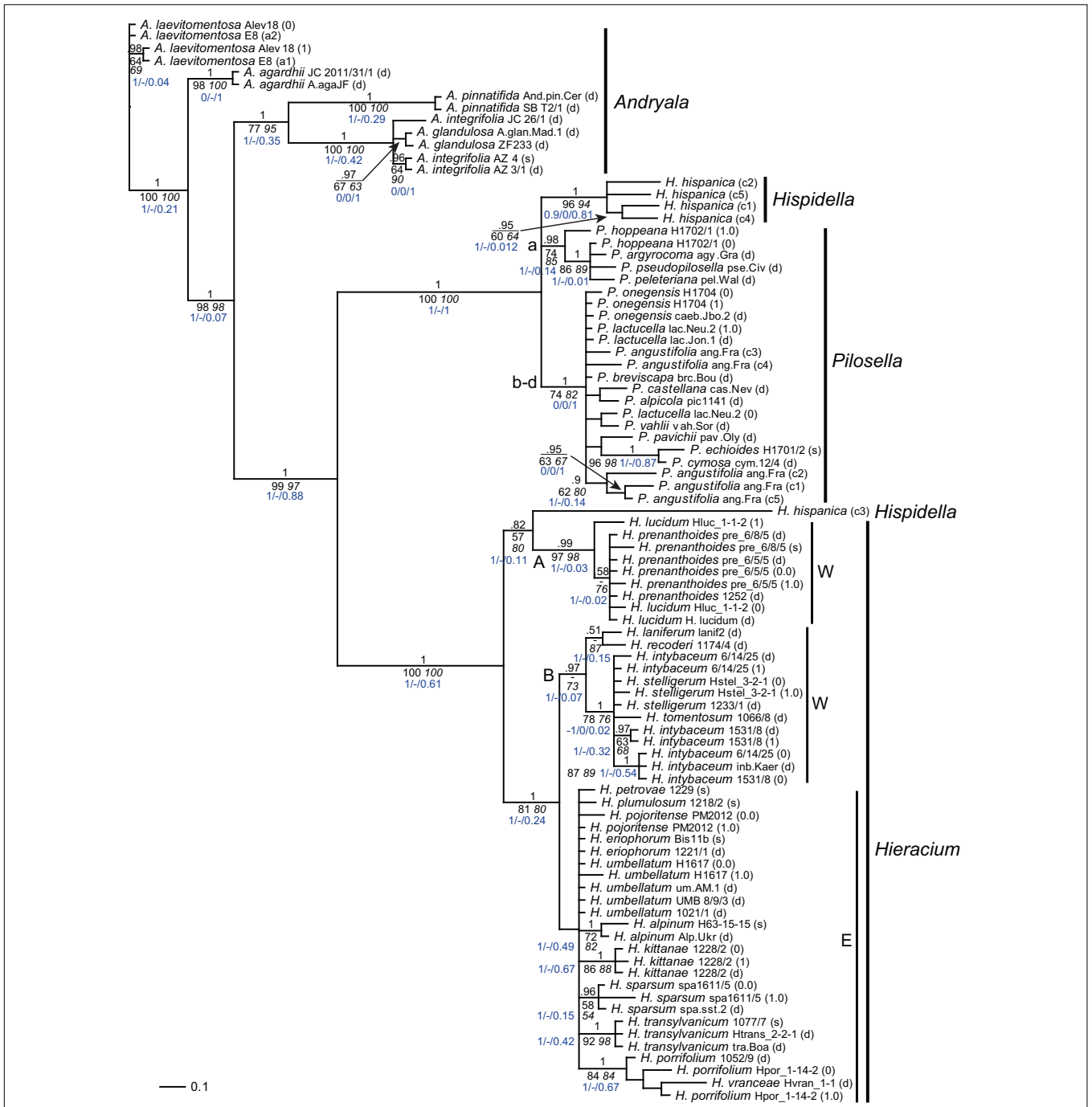


FIGURE 3 | Phylogenetic analysis of the Hieraciinae based on the 5S-NTS region. The Bayesian consensus tree is shown with posterior probabilities (pp) above branches and boostrap support (bs) from MP (regular) and ML (italics) analyses below branches. Values are only shown if pp was > 0.94 or bs > 70%. Below the support values, Quartet Concordance/Quartet Differential/Quartet Informativeness scores for 1000 replicates of the full alignment are displayed (in blue). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single); d, direct sequence; c, cloned sequence; a1/a2, variants of *A. laevitomentosa* E8 differing by a single substitution. W, E, western and eastern European clades of *Hieracium*; A, B, two clades of *Hieracium* with western European origin. a-d, main lineages of *Pilosella*. Accession labels correspond to **Table 1**.

1 and QD undefined), the non-monophyly of *Hispidella* in this tree seems genuine and does not stem from an artifact of the phylogenetic reconstruction, but requires a biological explanation.

With *Andryala*, the only difference concerned a clear separation of *A. pinnatifida* from *A. glandulosa* and *A. integrifolia*. With *Hieracium*, species of western European origin were separated into two clusters, one of them (A,

consisting of *H. lucidum* and *H. prenanthoides*) was sister to one cloned sequence of *Hispidella*, the other cluster (B, consisting of *H. stelligerum*, *H. tomentosum*, *H. intybaceum* and the Pyrenean species *H. laniferum* and *H. recoderi*) formed a well-supported branch together with all species of eastern European origin. The eastern lineage was still recognizable, however, poorly resolved. Finally, *Hieracium vranceae* nested among sequences of *H. porrifolium*, and *H. transylvanicum* belonged to the eastern European species of *Hieracium*. In these two cases, within each tree the branches leading to the sequences show a high level of informativeness (QI above 0.5) and no conflict between the trees and the data used to generate them (QC of 1 and QD undefined), which indicates that no alternative evolutionary history is favored by any of the branches.

Organization of 45S and 5S rDNA in Relation to Phylogeny

45S rDNA loci always occurred in terminal positions and 5S rDNA loci in interstitial positions; the 5S locus was always localized on the same chromosome with one 45S locus. The majority of samples of the three genera showed four loci of 45S and two loci of 5S rDNA per diploid genome (**Table 1** and **Figure 4**). In *Hieracium*, which was investigated in more detail, all analyzed accessions of three species (*H. prenanthoides*, *H. sparsum*, and *H. lucidum*) had six loci of 45S rDNA, *H. stelligerum* (Hstel_3-2-1) had seven loci, and one accession of *H. umbellatum* (UMB 8/9/3) had only three loci. In the latter, the 45S rDNA locus was lost together with a part of the chromosome arm (**Figure 4I**). Differences in the number of 5S loci were only found in one accession of *H. sparsum* (PM2099) whereas another sample of the same population (PM2102) and two samples of another population showed the usual two loci. No indications for translocations or inversions were observed.

For the combined ITS/ETS tree, the magnitude of the phylogenetic signal associated with the 45S rDNA loci count was measured with Blomberg's *K* and Pagel's λ statistics. For both indices, a value close to 0 indicates phylogenetic independence and a value of 1 indicates that species' traits are distributed as expected under a Brownian motion model of trait evolution. Blomberg's *K* revealed a moderate phylogenetic signal that was significantly different from zero (0.3519 *P*-value < 0.005); in contrast, Pagel's λ with a value of 0.7746 indicated an intense signal and strongly rejected the no-signal model (dAICc = 11.0717). This difference in magnitude could stem from the structure of the tree having a differential effect on the ability to accurately measure phylogenetic signal. Our tree contained a number of near zero length internal branches that could hinder Blomberg's *K* performance whereas they should not affect Pagel's λ (Münkemüller et al., 2012). In all cases, they clearly indicate that the trait is not randomly distributed, but follows largely the pattern of the phylogeny.

The ancestral number of 45S rDNA loci in Hieraciinae was four whereas all other numbers were derived as they mainly occurred close to the tips (**Figure 5** and **Supplementary Figure 2**). *Hieracium* species of western European origin showed six or seven loci, the latter appeared to be derived from the former. An exception was the state of *H. transylvanicum*, contained in this clade with four loci, but this was considered as an artifact of the placement of this eastern European species in the 'wrong' clade (see section Discussion) rather than a secondary reduction of locus numbers. Six locus numbers occurred independently in *H. sparsum*, which belongs to the eastern European lineage, and three loci occurred only in one sample of *H. umbellatum*, which is a derived character state.

DISCUSSION

Features and Molecular Evolution of rDNA Sequences

rDNA sequences of many individuals and species of the Hieraciinae were fairly well homogenized for all markers. In these cases, no or only a few polymorphic sites were found in direct sequences, or a single or two fairly similar sequences were retrieved by the genome skimming approach. If more than two alleles were determined for an individual, those from the first round of phasing (0, 1) were, except in *H. plumulosum* (see section "Results"), always more divergent than alleles found in the second round of phasing; these formed pairs of similar alleles (0.0, 0.1 and 1.0, 1.1, respectively). To our knowledge, phasing of rDNA sequences obtained from genome sequencing has not been attempted before, and in fact, even phasing of low-copy genes is rarely being done for phylogenetic inference even though it has been shown that phasing might improve the phylogenetic analysis (Eriksson et al., 2018). We show that results from direct sequencing and cloning are well comparable to the genomic approach in Hieraciinae and provide a new software tool (allele_linker) that allows to assign ITS and ETS alleles for combined analysis. The tool can also be used for pairwise assignment of alleles from two different markers in order to find correspondences between leaves in the trees that are necessary for phylogenetic inferences based on concatenated data, or for genome tree reconstruction where a species tree is built from non-concatenated markers and each allele/paralogue combination represents an evolutionary lineage.

In phylogenetic analyses, inclusion of polymorphic sequences along with phased (or cloned) alleles of the same sample or species resulted in patterns typically observed for ITS and ETS sequences of hybrids or allopolyploids: Polymorphic sequences (analogous to hybrid sequences) end up at or near the base of the clade containing the separated variants, or cluster with one of the variants, or occur in a basal position relative to all other ingroup taxa (Soltis et al., 2008). The latter possibility was not found in our study, probably because polymorphic sequences actually did not belong to hybrids (most diploid samples with mixed sequences that were additive for different lineages were excluded *a priori*), the level of polymorphism was usually low, and the dominant sequence (ignoring small additional peaks) used for phylogeny reconstruction often matched that of one phased allele as should be expected.

Alleles of some species formed coalescent groups, but often, sequences were similar or identical between closely related species indicating very recent divergence of the respective taxa

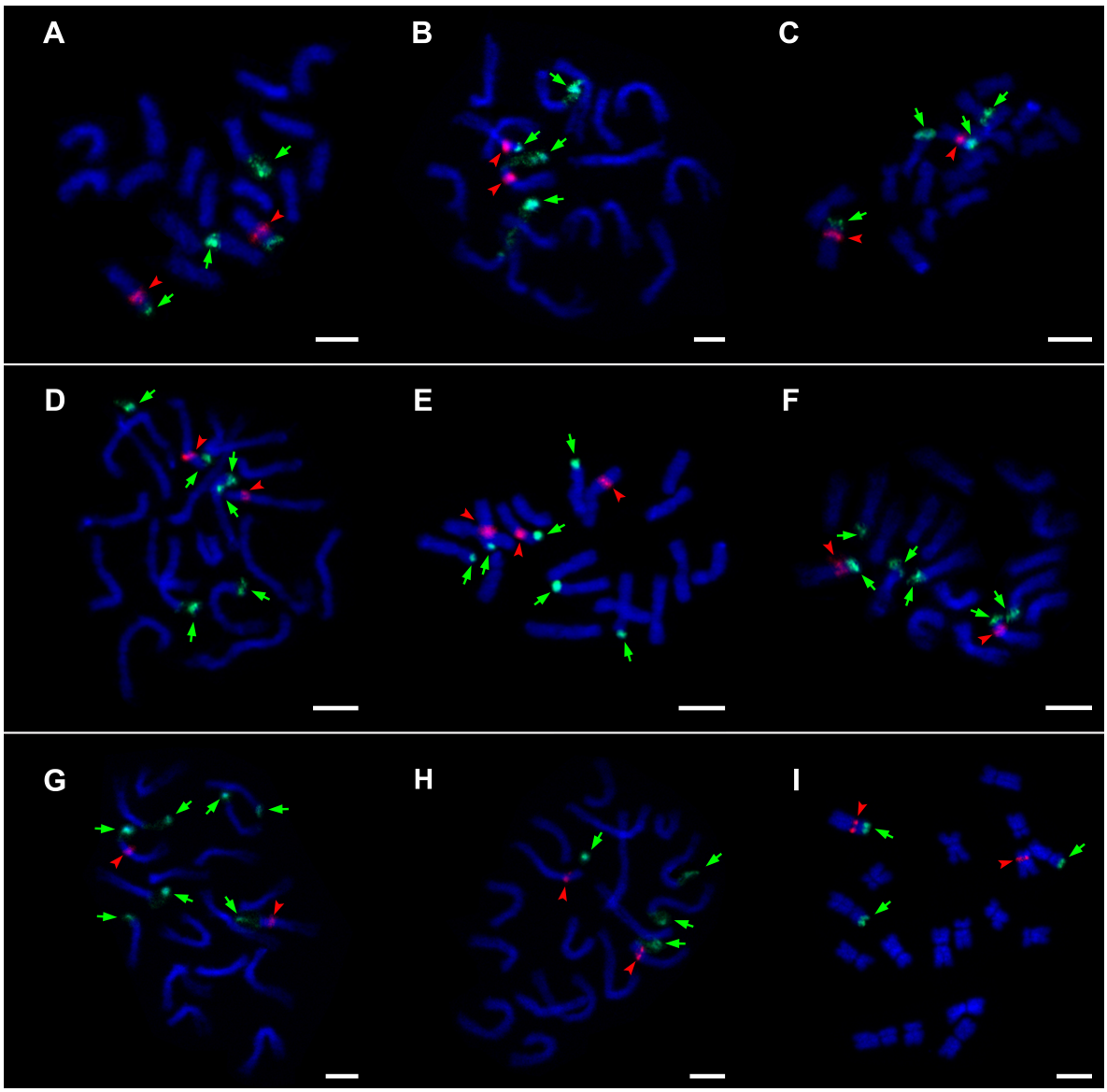


FIGURE 4 | Visualization of 45S rDNA and 5S rDNA loci on metaphase chromosomes. **(A)** *Andryala agardhii* PM2390, **(B)** *Pilosella echioides* H1701/2, **(C)** *P. hoppeana* H1702/1, **(D)** *Hieracium sparsum* PM2102, **(E)** *H. sparsum* PM2099, **(F)** *H. sparsum* spa1611/6, **(G)** *H. stelligerum* Hstel_3-2-1, **(H)** *H. umbellatum* H1617, **(I)** *H. umbellatum* UMB 8/9/3. 45S rDNA (green signal and arrows) and 5S rDNA (red signal and arrowheads). Chromosomes were counterstained with DAPI (blue). Scale bars = 5 μm.

or, maybe less likely, homogenization of rDNA toward the same variant. ITS and ETS sequences were generally more variable within individuals than 5S-NTS sequences (**Table 2**). This might, at least in part, be due to the higher length of aligned ITS (695 bp) and ETS (574 bp) sequences compared to 5S-NTS (296 bp). Nevertheless, compared to its length, the overall variation in the 5S-NTS was larger than that of both ITS and ETS: 5S-NTS showed 114 parsimony informative characters (38.5%) whereas ITS showed 135 (19.4%) and ETS 137 (23.9%). A much higher proportion of parsimony informative characters in 5S-NTS sequences compared to ITS sequences was also found in *Anemone* (Mlinarec et al., 2012). On the other hand, ITS and ETS together provided more than twice as many parsimony informative characters for phylogenetic analysis. ITS/ETS and 5S-NTS showed different resolution in different parts of the tree as well as several highly incongruent patterns. This indicates that while the 5S locus is always located on the same chromosome with one of the 45S rDNA loci, their sequences evolved independently. Independent evolution of both arrays was also reported in other plants and in animals (Rosato et al., 2015; Araya-Jaime et al., 2020).

Very often, ITS sequences of diploids are well homogenized in plants (Baldwin et al., 1995) as evidenced by their widespread use for building phylogenies despite their well-known drawbacks

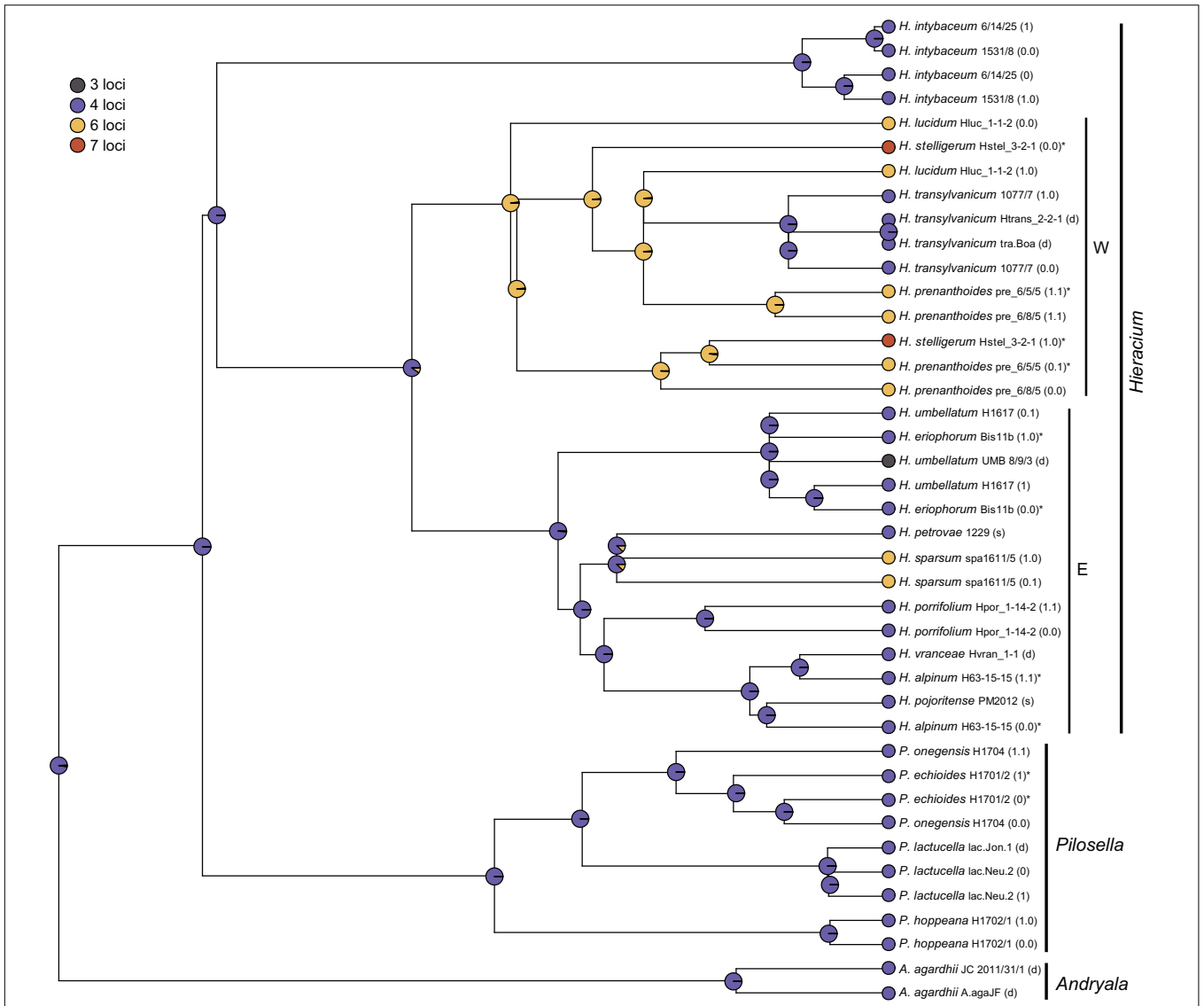


FIGURE 5 | Ancestral character state reconstruction on the maximum likelihood tree based on the combined ITS and ETS sequences using stochastic mapping of 45S rDNA loci. Locus numbers (see **Table 1**) were assigned to sequences (alleles) of the same individual. For *H. transylvanicum*, all further individuals were assigned 4 loci, because this species does not show intraspecific variation (Ilnicki et al., 2010). For *A. agardhii*, *P. lactucella*, *H. porrifolium*, and *H. vranceae*, for which only cytogenetic data from other individuals were available, locus numbers were assigned to the species. Pies at nodes represent the marginal ancestral states (empirical Bayesian posterior probabilities). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single). Labels correspond to those in the ITS tree (**Figure 1**); swapped alleles for ETS are marked by asterisks (*). d, direct sequence; W, E, western and eastern European clades of *Hieracium*.

(Álvarez and Wendel, 2003; Nieto Feliner and Rosselló, 2007). The same holds for the less often used ETS, which is often more variable than the ITS, and the phylogenetic signal from both regions is usually congruent and provides better resolution and higher support in trees (Baldwin and Markos, 1998; Calonje et al., 2009). Both regions are part of the array forming the nucleolar organizer region (Hillis and Dixon, 1991); they associate during interphase, and interlocus homogenization is a common observation where multi-gene families are located in terminal positions on chromosomes (Cronn et al., 1996; Volkov et al., 1999; Rauscher et al., 2004). In contrast, 5S-NTS sequences are usually highly polymorphic within individuals and often exhibit different unit size classes (e.g., Kellogg and Appels, 1995; Besendorfer et al., 2005; Galián et al., 2014; and references therein). For this reason, most research is focused on the molecular patterns of this region, but it is less often used for phylogeny reconstruction (e.g., Lan and Albert, 2011; and references therein). In the Hieraciinae, 5S-NTS sequences are exceptionally well homogenized. As almost all of them possess only one 5S rDNA locus per haploid genome, intralocus homogenization may be more efficient in this case than interlocus homogenization of 45S rDNA. However, Volkov et al. (2007) suggested that concerted evolution generally operates differently in 5S rDNA. Also, other factors such as the number of repeats

in an array, the intensity of natural selection and effective population size can play a role (Smith, 1976; Basten and Ohta, 1992; Schlötterer and Tautz, 1994). Lower copy numbers of 5S arrays compared to 45S arrays are often observed (Sastri et al., 1992; Macas et al., 2011), also in *Hieracium* (Zagorski et al., 2020), although a stoichiometric relationship of mature rRNA copies from genes of both loci is required for ribosome biogenesis (Fromont-Racine et al., 2003). Our findings add to the vast literature on differential behavior of unlinked rDNA arrays in plants and animals.

Phylogenetic Trees Reveal Hybridization and Differential Homogenization of rDNA

In the Hieraciinae, several cases of ancient intergeneric hybridization were found previously based on the discrepancy between chloroplast and nuclear DNA markers (Fehrer et al., 2007a). In all genera, especially within *Hieracium*, massive allele sharing of various molecular markers between species was inferred, and many interspecific relationships remained unresolved (Fehrer et al., 2009; Krak et al., 2013; Ferreira et al., 2015; Mráz et al., 2019; Chrtek et al., 2020). Reasons for these patterns were a lack of divergence in closely related species, incomplete lineage sorting, and hybridization. Phased alleles of ITS and ETS added further complexity at the intra-individual level, and the 5S-NTS, which was investigated for Hieraciinae for the first time, provided novel insights into the intricate evolution of this group. Particular patterns of some species are discussed in the following paragraphs.

In the case of *Hieracium intybaceum*, all nuclear markers employed previously (ITS – Fehrer et al., 2007a; ETS – Fehrer et al., 2009; *sqs* – Krak et al., 2013; *gsh1* – Chrtek et al., 2020) placed the species in an outgroup position. *Hieracium intybaceum* is considered as an ancient intergeneric hybrid involving a parent whose nuclear DNA markers belonged to an extinct taxon (Fehrer et al., 2007a; Krak et al., 2012, 2013; Chrtek et al., 2020). The 5S-NTS is the first nuclear marker that groups *H. intybaceum* with other *Hieracium* species; it occurs in a well-supported cluster with the western European species *H. stelligerum* and *H. tomentosum*, and its sequences are highly similar to those of these species. In contrast, its chloroplast DNA clusters with the eastern European species *Hieracium alpinum*, *H. sparsum*, *H. pojoritense,* and *H. vranceae* (Krak et al., 2013; Mráz et al., 2019). The repetitive part of its genome is highly similar to *Hieracium* species of western as well as eastern European origin (*H. prenanthoides*, *H. umbellatum*; Zagorski et al., 2020), and *H. intybaceum* also shows an abundant, *Hieracium*-specific tandem repeat located in the centromeric regions of all chromosomes (Belyayev et al., 2018). As the species forms many polyploid apomictic hybrids with *Hieracium* species (Zahn, 1921–1923; Chrtek et al., 2020), it was included in this genus despite its markedly different morphology (Zahn, 1921–1923). It shares also the same chromosomal organization of rDNA with the majority of the Hieraciinae, however, concerted evolution of ITS/ETS (45S rDNA loci) and 5S-NTS (5S rDNA loci) operated in opposite directions – toward the extinct parent or toward *Hieracium*, respectively. The fact that it clusters with different lineages of *Hieracium* in chloroplast and 5S-NTS trees may indicate that the hybridization event occurred early, before western and eastern European lineages of *Hieracium* diverged. Alternatively, given the sequence similarity with different contemporary species, *H. intybaceum* may have been introgressed by a second lineage of *Hieracium* after the initial hybridization event. At the intraspecific and intra-individual level, all non-coding rDNA regions are very well homogenized, and indications for its hybrid origin are based on the discrepancies of different molecular markers rather than on mixed sequences of the same markers. Sequences of all other molecular markers as well as AFLP patterns (Zahradníček and Chrtek, 2015) are fairly homogenous across populations indicating a single, however, complex origin of this species.

A further case of homogenization of ITS/ETS and 5S-NTS into different directions concerns *Hieracium transylvanicum*. Intraspecific variation was also very low, and all rDNAs were well homogenized. All alleles of all accessions of this species formed well-supported coalescent groups, but these occurred in the western European clade with ITS/ETS, but among eastern European species with 5S-NTS. The latter placement is in accordance with its geographic origin and genome size (Chrtek et al., 2009). This pattern may be explained by incomplete lineage sorting of western and eastern European alleles of *Hieracium*.

Hieracium vranceae is a recently described species of the Carpathians (Mráz et al., 2019). Based on ITS and ETS (and chloroplast DNA), it is most closely related to *H. alpinum* and *H. pojoritense* occurring in the same area. Surprisingly, it formed a well-supported branch together with the southeastern alpine species *H. porrifolium* with the 5S-NTS. This is not a methodological artifact, because its 5S-NTS sequence was unique and well homogenized, but not identical to any of the *H. porrifolium* sequences. In this case, we may observe either hidden introgression or incomplete lineage sorting. A second sample of *H. vranceae* showed two divergent ETS alleles (Mráz et al., 2019), one of which was nearly identical with the sequence of the individual included here, the other occurred in an unresolved polytomy with other eastern European species (including *H. porrifolium*). Recent hybridization in *Hieracium* despite ample hybridization in the past, leading to thousands of allopolyploid apomictic species is rare, and hybrids are usually female sterile (Mráz et al., 2005, 2011). Besides, diploids are often endemic with very narrow distribution ranges and are sometimes known only from very few populations. Their distribution ranges and ecological requirements rarely overlap, and furthermore, the so-called mentor effect causes a breakdown of self-incompatibility under the influence of foreign pollen, which results in selfing and represents a strong barrier to introgression (Mráz, 2003; Mráz and Paule, 2006). For these reasons, if hybridization is responsible for the different placement of *H. vranceae* with different rDNA markers, it most probably did not occur recently. On the other hand, it is also difficult to explain the pattern by incomplete lineage sorting, because it should not show different relationships of the same individual with different, not closely related species with high support.

Hispidella hispanica is the only annual species of the Hieraciinae, endemic to central and western parts of the Iberian Peninsula. It is a monotypic genus that is, according to all

molecular markers applied so far, sister to *Pilosella*. However, it showed a strongly divergent cloned 5S-NTS sequence that grouped near the base of *Hieracium*. This also is not a methodological artifact, because an indel position that differs between the clones was visible on the direct sequence, and the aberrant clone is very divergent from all other sequences of the Hieraciinae. We assume this represents yet another case of ancient intergeneric hybridization, this time not revealed by a discrepancy between rDNA regions or other molecular markers, but by a mixture of 5S-NTS variants at the intra-individual level. Only a 58 years old herbarium specimen was available for this species, and several attempts by us and a Spanish collaborator to collect *Hispidella* in the field failed, therefore the species can currently not be investigated in more detail. In *Potamogeton*, where ITS sequences are by far better homogenized than 5S-NTS sequences, the latter retained parental copies of hybrids even if the former have lost indications of hybrid origin or nearly so (Kaplan et al., 2018).

In the Mediterranean-Macaronesian genus *Andryala*, *A. integrifolia*, *A. glandulosa,* and *A. pinnatifida* belong to a 'major radiation group' with relatively recent speciation and largely unresolved species relationships (Ferreira et al., 2015). Within this group, three samples of *Andryala integrifolia* formed a well-supported monophyletic branch with ITS, but one phased allele of accession AZ 4 occurred in an unresolved position. In contrast, ETS and 5S-NTS grouped all alleles of two Algerian accessions (AZ 3/1 and AZ 4) whereas an Andalusian sample (JC 26/1) occurred outside this clade. With the low-copy nuclear marker *sqs*, the latter sample showed two rather divergent alleles, and generally, *A. integrifolia* was the most polyphyletic species of *Andryala* according to this marker (Ferreira et al., 2015). It is also known to hybridize with other species (Maire, 1937; García Adá, 1992), however, no indication for a recent hybrid origin of any of the accessions was found. Hybridization of the individuals of *A. integrifolia* studied here with *A. glandulosa* and *A. pinnatifida* can be excluded, because *A. glandulosa* is endemic to Madeira and *A. pinnatifida* to the Canary Islands. *Andryala integrifolia* is the most widespread species of the genus (Ferreira et al., 2015), which implies a larger population size. Here, rDNA patterns can be best interpreted as a consequence of recent divergence with incomplete lineage sorting.

In *Pilosella*, species relationships within clades a, c and d were unresolved, and their sequences were nearly identical within clades. In clade d, phased alleles of *P. echioides* and *P. onegensis* were more different than direct sequences of further species in the same clade with ITS and ETS, but this was not the case with 5S-NTS, where a fully homogenized sequence was found in *P. echioides* and two nearly identical phased alleles in *P. onegensis*. 5S-NTS showed a well-supported relationship of *P. echioides* with *P. cymosa*, but none of its divergent ITS or ETS alleles were grouping with that species. We assume that speciation in this clade was also recent and that incomplete lineage sorting is responsible for the differential behavior of the alleles. In contrast to *Hieracium*, recent hybridization in *Pilosella*, even across ploidy levels, is basically unlimited, and hybrids are usually fertile (Krahulcová et al., 2000; Fehrer et al., 2007b). Nevertheless, the morphologies of species in clade d are very divergent and no indications for introgression were observed in the material studied.

Organization of 45S and 5S rDNA

Previous cytogenetic studies of the Hieraciinae focused on *Pilosella*, where an aposporous-specific meiotic avoidance locus and satellite markers were studied (Okada et al., 2011; Kotani et al., 2014; Belyayev et al., 2018) and on *Hieracium*, where satellite markers and rDNA loci of a few species were investigated (Ilnicki et al., 2010; Belyayev et al., 2018; Mráz et al., 2019; Chrtek et al., 2020; Zagorski et al., 2020). Additional species and populations that cover most of the phylogenetic lineages in both genera were added, and a species of genus *Andryala* was karyotyped here for the first time. *Andryala agardhii*, all *Pilosella* and the majority of *Hieracium* samples showed four loci of 45S rDNA and two loci of 5S rDNA per diploid genome, and their chromosomal organization – 45S rDNA in terminal positions and 5S rDNA in interstitial positions, the latter located on the same chromosome with one of the 45S rDNA loci – was the same. This indicates that this pattern represents the ancestral condition in the Hieraciinae. Furthermore, the same number and position of rDNA loci in diploids was inferred as the ancestral state across plants, except that the Hieraciinae have 18 chromosomes and the general karyotype of plants was inferred to have 16 chromosomes (Garcia et al., 2017).

In three species of *Hieracium* (*H. prenanthoides*, *H. lucidum,* and *H. sparsum*), six loci of 45S rDNA were found. The first two species belong to the western European clade, but *H. sparsum* belongs to the eastern European lineage. *H. prenanthoides* and *H. lucidum* form together a well-supported lineage of western European taxa in the 5S-NTS tree (**Figure 3**, A) and may therefore have acquired the additional 45S rDNA loci prior to their species divergence. In contrast, *H. sparsum* has apparently acquired the additional loci independently. Whether the possession of six loci is species-specific or not cannot be decided, however, without a broader geographic sampling of these species. The distribution areas of diploids (southwestern Alps in *H. prenanthoides*, Sicily in *H. lucidum* and the Balkans in *H. sparsum*) are relatively small so that these species may be actually uniform concerning the number of 45S rDNA loci. In *H. stelligerum*, another western European species, seven loci of 45S rDNA were observed. The ancestral state reconstruction based on the ITS/ETS tree implies a transition from 6 to 7 loci in the western European clade, but another possibility is a direct transition from 4 to 7 loci based on the position of *H. stelligerum* in the 5S-NTS tree (**Figure 3**, B). In this case, there is no indication whether the possession of seven loci is a species-specific pattern or not, because only a single sample was analyzed. The number of major polymorphisms in samples/species with four loci ranged from 0 (in *H. petrovae*, *H. pojoritense* PM2012, and *A. agardhii* JC 2011/31/1) to 17 (in *A. agardhii* A.agaJF and *P. onegensis* H1704); from 5 (in *H. sparsum* spa1611/5) to 31 (in *H. lucidum* Hluc_1-1-2) in samples with six loci (but a second sample of *H. lucidum* from the same population had only three polymorphisms); *H. stelligerum* Hstel_3-2-1 with seven loci showed 15 polymorphisms, and *H. umbellatum* UMB 8/9/3 with only three loci (see below) had six polymorphisms like

another sample of the same species (H1617) that showed the usual four loci. The high variation in intra-individual sequence polymorphisms across samples, largely without any phylogenetic pattern, do not suggest consistent differences in the degree of homogenization of ITS or ETS sequences in relation to 45S rDNA locus number. Generally, interlocus concerted evolution seems to have operated fairly well in most samples, which may have been facilitated by the subtelomeric positions of the 45S rDNA loci (Wendel, 2000; Lan and Albert, 2011).

Within *H. sparsum* and also within *H. umbellatum*, the most widespread diploid *Hieracium* species (Bräutigam, 1992), intraspecific variation was observed. In *H. umbellatum*, one 45S rDNA locus was lost in accession UMB 8/9/3 from Slovakia, but not in accession H1617 from the Czech Republic. Their ITS and ETS sequences (representing the 45S rDNA) are identical. In *H. sparsum*, an additional locus of 5S rDNA was found in one accession from the Rila Mts whereas a second accession of that population and two accessions from the Pirin Mts showed the usual two loci of 5S rDNA. The additional locus of accession PM2099 (**Figure 4E**) occurred in an interstitial position on a chromosome not bearing also a 45S rDNA locus. 5S-NTS sequences from the variable population are not available (the plants perished), but two accessions of *H. sparsum* grouped together, albeit with low support. Intraspecific and even intra-population variation in the number of rDNA loci indicates that locus acquisition and loss can happen very quickly and also that it is not usable as a phylogenetic marker in the Hieraciinae. Also in many other plant groups, variation in the number and distribution pattern of rDNA is commonly observed among closely related species (Lan and Albert, 2011; Garcia et al., 2017; and references therein) and is therefore not informative concerning species relationships. Many studies in plants and animals have shown variation in rDNA locus number (e.g., Raskina et al., 2008; Gouja et al., 2015; Kolano et al., 2015; and references therein), suggesting that rDNA sites are highly dynamic components of the genome (Britton-Davidian et al., 2012).

Interestingly, three hemizygous loci were detected: seven loci and three loci of 45S rDNA in *H. stelligerum* and *H. umbellatum*, and three loci of 5S rDNA in *H. sparsum*. Hemizygous rDNA loci were also observed in other plant groups, for example in diploid and polyploid grasses (Rocha et al., 2018; Chiavegatto et al., 2019) and diploid orchids (Lan and Albert, 2011). Generally, a potential reason for the observation of hemizygous loci is hybridization (Myburg et al., 2003). However, none of the three *Hieracium* species (nor individuals) show any indication of recent or ancient hybridization, neither in their morphology nor with any molecular markers. Therefore, the occurrence of hemizygous loci, which were also frequently observed in other satellite DNAs in *Hieracium* (Belyayev et al., 2018; Zagorski et al., 2020), may have another reason. The genome size of *Hieracium* is approximately twice as high as that in *Pilosella* and *Andryala* (Suda et al., 2007; Chrtek et al., 2009; Zahradníček et al., 2018). This is suggestive of a whole genome duplication (WGD) that may have occurred in the ancestral lineage of *Hieracium*. WGD is widespread in the evolutionary history of the Asteraceae: In addition to a previously suggested paleopolyploidization event at the origin of the core-Asteraceae (Chapman et al., 2008, 2012), phylotranscriptomic analyses have uncovered at least four, possibly seven other events distributed at different levels in the Asteraceae phylogeny (Huang et al., 2016). A detailed genomic investigation of genus *Hieracium* is needed to understand if such an event has actually occurred in this genus as well and, if so, how it has affected its genome organization and the evolution of molecular markers.

CONCLUSION

Molecular evolution of multi-copy sequences such as rDNA arrays poses specific challenges for phylogenetic inference. Appropriate treatment of intra-individual variation and the investigation of multiple markers can provide interesting insights in complex species relationships as well as in the evolution of the markers themselves. Contrary to most other plants, ITS and ETS sequences (45S rDNA locus) are more polymorphic than 5S-NTS sequences (5S rDNA locus) in Hieraciinae even though, generally, concerted evolution homogenized all rDNA arrays fairly well. Several strong discrepancies between ITS/ETS and 5S-NTS phylogenetic trees reveal previously unidentified cases of reticulation, and homogenization of the different arrays sometimes occurs in opposite directions. Comparison with the chromosomal organization of the loci corresponding to the markers shows that their location in the genome is far more dynamic than the sequences they contain, implying that chromosomal patterns are not suitable to infer species relationships, at least not in *Hieracium*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW315935–MW315953, MW325251–MW325296, MW328890–MW329033, MW587333–MW587351, and MW591759–MW591773; https://www.ebi.ac.uk/ena, ERS5458545–ERS5458563.

AUTHOR CONTRIBUTIONS

JF conceived of the study. JF and YB analyzed the data. YB developed the software tool. LP and RS did cytogenetic analyses. JJ did molecular labwork. JC and PM collected and determined the material. JF and YB wrote the manuscript. All the authors contributed to the drafts and gave final approval for publication.

FUNDING

This work was supported by the Czech Science Foundation (17–14620S to JF and PM, 14–02858S to PM) and the long–term

research development project no. RVO 67985939 of the Czech Academy of Sciences.

ACKNOWLEDGMENTS

We thank P. Caklová for supporting the labwork, A. Belyayev for supporting the cytogenetic analyses, M. Hartmann for help with ancestral states reconstructions, and S. Bräutigam, J. Doležal, D. J. Frey, E. di Gristina, K. Kabátová, Z. Szeląg, and A. Zeddam for providing material. Institutional funding was received by the Charles University Prague (to PM). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.647375/full#supplementary-material

Supplementary Figure 1 | Phylogenetic analysis of the Hieraciinae based on the combined ITS and ETS regions. The Bayesian consensus tree is shown with posterior probabilities (pp) above branches and boostrap support (bs) from MP (regular) and ML (italics) analyses below branches. Values are only shown if pp was > 0.94 or bs > 70%. Below the support values, Quartet Concordance/Quartet Differential/Quartet Informativeness scores for 1000 replicates of the full alignment are displayed (in blue). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single). These labels correspond to those in the ITS tree (**Figure 1**); swapped alleles for ETS are marked by asterisks (*). d, direct sequence; a1/a2, two alleles of *Hispidella* (minor and major sequence inferred from direct sequencing); c, cloned sequence. W, E, western and eastern European clades of *Hieracium*. a-d, main lineages of *Pilosella*. Accession labels correspond to **Table 1**.

Supplementary Figure 2 | Ancestral character state reconstruction on the maximum likelihood tree based on the combined ITS and ETS sequences using stochastic mapping of 45S rDNA loci (including taxa with unknown locus numbers). Locus numbers (see **Table 1**) were assigned to sequences (alleles) of the same individual. For *H. transylvanicum*, all further individuals were assigned four loci, because this species does not show intraspecific variation (Ilnicki et al., 2010). For *A. agardhii*, *P. lactucella*, *H. porrifolium*, and *H. vranceae*, for which only cytogenetic data from other individuals were available, locus numbers were assigned to the species. Pies at nodes represent the marginal ancestral states (empirical Bayesian posterior probabilities). Phased alleles are indicated behind accession labels as 0.0, 0.1, 1.0, 1.1., 0, 1, and s (single). Labels correspond to those in the ITS tree (**Figure 1**); swapped alleles for ETS are marked by asterisks (*). For the tree with branch lengths and support values, see **Supplementary Figure 1**. d, direct sequence; a1/a2, two alleles of *Hispidella* (minor and major sequence inferred from direct sequencing); c, cloned sequence. W, E, western and eastern European clades of *Hieracium*.

Supplementary Table 1 | Details of sample origins and voucher information.

Supplementary Table 2 | Evolutionary transition model selection in a likelihood framework for stochastic character mapping using the corrected Akaike information criterion. ER, equal rates model; SYM, symmetrical model; and ARD, all-rates-different model. lnL, log-likelihood of the model; AICc, corrected Akaike information criterion; dAICc, difference between the AICc of the model and the best model; wAICc, AICc weight of the model. Differences in AICc between competing models were considered negligible if they were <3, very strong if >10, and moderately strong between 4 and 7 (Burnham and Anderson, 2002). The AICc weights indicate the probability that the model is the best among the whole set of candidate models.

REFERENCES

Allam, A., Kalnis, P., and Solovyev, V. (2015). Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. *Bioinformatics* 31, 3421–3428. doi: 10.1093/bioinformatics/btv415

Álvarez, I., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/s1055-7903(03)00208-2

Appels, R., and Honeycutt, R. L. (1986). "rDNA evolution over a billion years," in *DNA Systematics: Plants II. Plant DNA*, ed. S. K. Dutta (Boca Raton, FL: CRC Press), 81–125.

Araya-Jaime, C., Claudio Palma-Rojas, C., Von Brand, E., and Silva, A. (2020). Cytogenetic characterization, rDNA mapping and quantification of the nuclear DNA content in *Seriolella violacea* Guichenot, 1848 (Perciformes, Centrolophidae). *Comp. Cytogenetics* 14, 319–328.

Arnheim, N. (1983). "Concerted evolution of multigene families," in *Evolution of Genes and Proteins*, eds M. Nei and R. K. Koehn (Sunderland, MA: Sinauer), 38–61.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., and Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Miss. Bot. Garden* 82, 247–277. doi: 10.2307/2399880

Baldwin, G., and Markos, S. (1998). Phylogenetic utility of the external transcribed spacer (ETS) of 18-26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Mol. Phylogenet. Evol.* 10, 449–463. doi: 10.1006/mpev.1998.0545

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Basten, C. J., and Ohta, T. (1992). Simulation study of a multigene family, with special reference to the evolution of compensatory advantageous mutations. *Genetics* 132, 247–252.

Belyayev, A., Paštová, L., Fehrer, J., Josefiová, J., Chrtek, J., and Mráz, P. (2018). Mapping of *Hieracium* (Asteraceae) chromosomes with genus-specific satDNA elements derived from next-generation sequencing data. *Plant Syst. Evol.* 304, 387–396. doi: 10.1007/s00606-017-1483-y

Besendorfer, V., Krajačić-Sokol, I., Jelenić, S., Puizina, J., Mlinarec, J., Sviben, T., et al. (2005). Two classes of 5S rDNA unit arrays of the silver fir, *Abies alba* Mill.: structure, localization and evolution. *Theor. Appl. Genet.* 110, 730–741. doi: 10.1007/s00122-004-1899-y

Blomberg, S. P., Garland, T. Jr., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borchsenius, F. (2009). *FastGap 1.2. Department of Biosciences, Aarhus University, Denmark*. Available online at: http://www.aubot.dk/FastGap_home.htm. (accessed December 5, 2020)

Bräutigam, S. (1992). "*Hieracium* L," in *Vergleichende Chorologie der Zentraleuropäischen Flora 3*, eds H. Meusel and E. J. Jäger (Stuttgart: Gustav Fischer Verlag), 325–333.

Bräutigam, S., and Greuter, W. (2007). A new treatment of *Pilosella* for the Euro-Mediterranean flora [Notulae ad floram euro-mediterraneam pertinentes 24]. *Willdenowia* 37, 123–137. doi: 10.3372/wi.37.37106

Britton-Davidian, J., Cazaux, B., and Catalan, J. (2012). Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. *J. Heredity* 108, 68–74. doi: 10.1038/hdy.2011.105

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer.

Calonje, M., Martin-Bravo, S., Dobeš, C., Gong, W., Jordon-Thaden, I., Kiefer, C., et al. (2009). Non-coding nuclear DNA markers in phylogenetic reconstruction. *Plant Syst. Evol.* 282, 257–280. doi: 10.1007/s00606-008-0031-1

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Chapman, M. A., Leebens-Mack, J. H., and Burke, J. M. (2008). Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Mol. Biol. Evol.* 25, 1260–1273. doi: 10.1093/molbev/msn001

Chapman, M. A., Tang, S., Draeger, D., Nambeesan, S., Shaffer, H., Barb, J. G., et al. (2012). Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of CYCLOIDEA genes in the Asteraceae. *PLoS Genet.* 8:e1002628. doi: 10.1371/journal.pgen.1002628

Chiavegatto, R. B., Chaves, A. L. A., Rocha, L. C., Benites, F. R. G., Peruzzi, L., and Techio, V. H. (2019). Heterochromatin bands and rDNA sites evolution in polyploidization events in *Cynodon* Rich. (Poaceae). *Plant Mol. Biol. Rep.* 37, 477–487. doi: 10.1007/s11105-019-01173-2

Chrtek, J., Mráz, P., Belyayev, A., Paštová, L., Mrázová, V., Caklová, P., et al. (2020). Evolutionary history and genetic diversity of apomictic allopolyploids in *Hieracium* s.str.: morphological versus genomic features. *Amer. J. Bot.* 107, 66–90. doi: 10.1002/ajb2.1413

Chrtek, J., Zahradníček, J., Krak, K., and Fehrer, J. (2009). Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. *Ann. Bot.* 104, 161–178. doi: 10.1093/aob/mcp107

Constantinides, B., and Robertson, D. L. (2017). Kindel: indel-aware consensus for nucleotide sequence alignments. *J. Open Source Softw.* 2:282. doi: 10.21105/joss.00282

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for statistical computing.

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Denduangboripant, J., Cronk, Q. C. B., Kokubugata, G., and Möller, M. (2007). Variation and inheritance of nuclear ribosomal DNA clusters in *Streptocarpus* (Gesneriaceae) and their biological and phylogenetic implications. *Int. J. Plant Sci.* 168, 455–467. doi: 10.1086/512103

Eriksson, J. S., de Sousa, F., Bertrand, Y. J. K., Antonelli, A., Oxelman, B., and Pfeil, B. E. (2018). Allele phasing is critical to revealing a shared allopolyploid origin of *Medicago arborea* and *M. strasseri* (Fabaceae). *BMC Evol. Biol.* 18:9. doi: 10.1186/s12862-018-1127-z

Fehrer, J., Gemeinholzer, B., Chrtek, J., and Bräutigam, S. (2007a). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Mol. Phylogenet. Evol.* 42, 347–361. doi: 10.1016/j.ympev.2006.07.004

Fehrer, J., Krahulcová, A., Krahulec, F., Chrtek, J. Jr., Rosenbaumová, R., and Bräutigam, S. (2007b). "Evolutionary aspects in *Hieracium* subgenus *Pilosella*," in *Apomixis: Evolution, Mechanisms and Perspectives (Regnum Vegetabile 147)*, eds E. Hörandl, U. Grossniklaus, P. van Dijk, and T. Sharbel (Königstein: Koeltz), 359–390.

Fehrer, J., Krak, K., and Chrtek, J. (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise. *BMC Evol. Biol.* 9:239. doi: 10.1186/1471-2148-9-239

Fehrer, J., Šimek, R., Krahulcová, A., Krahulec, F., Chrtek, J., Bräutigam, E., et al. (2005). "Evolution, hybridisation, and clonal distribution of apo- and amphimictic species of *Hieracium* subgen. *Pilosella* (Asteraceae, Lactuceae) in a Central European mountain range," in *Plant Species-Level Systematics: New Perspectives on Pattern & Process (Regnum Vegetabile 143)*, eds F. T. Bakker, L. W. Chatrou, B. Gravendeel, and P. B. Pelser (Königstein: Koeltz), 175–201.

Ferreira, M. Z., Zahradníček, J., Kadlecová, J., Menezes de Sequeira, M., Chrtek, J. Jr., and Fehrer, J. (2015). Tracing the evolutionary history of the little-known Mediterranean-Macaronesian genus *Andryala* (Asteraceae) by multigene sequencing. *Taxon* 62, 535–551. doi: 10.12705/643.10

Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. *Gene* 313, 17–42. doi: 10.1016/s0378-1119(03)00629-2

Galián, J. A., Rosato, M., and Rosselló, J. A. (2014). Partial sequence homogenization in the 5S multigene families may generate sequence chimeras and spurious results in phylogenetic reconstructions. *Syst. Biol.* 63, 219–230. doi: 10.1093/sysbio/syt101

Garcia, S., Garnatje, T., Hidalgo, O., McArthur, E. D., Siljak-Yakovlev, S., and Valles, J. (2007). Extensive ribosomal DNA (18S-5.8S-26S and 5S) colocalization in the North American endemic sagebrushes (subgenus *Tridentatae*, *Artemisia*, Asteraceae) revealed by FISH. *Plant Syst. Evol.* 267, 79–92. doi: 10.1007/s00606-007-0558-6

Garcia, S., Kovarik, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Garcia, S., Panero, J. L., Siroky, J., and Kovarik, A. (2010). Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC Plant Biol.* 10:176. doi: 10.1186/1471-2229-10-176

García Adá, R. (1992). Un híbrido nuevo en el género *Andryala* (Asteraceae). *Acta Bot. Malac.* 17, 259–260.

Gouja, H., Garnatje, T., Hidalgo, O., Neffati, M., Raies, A., and Garcia, S. (2015). Physical mapping of ribosomal DNA and genome size in diploid and polyploid North African *Calligonum* species (Polygonaceae). *Plant Syst. Evol.* 301, 1569–1579. doi: 10.1007/s00606-014-1183-9

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131. doi: 10.1093/bioinformatics/btm538

Hemleben, V., and Grierson, D. (1978). Evidence that in higher plants the 25S and 18S genes are not interspersed with genes for 5S rRNA. *Chromosoma* 65, 353–358. doi: 10.1007/BF00286414

Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66, 411–453. doi: 10.1086/417338

Huang, C. H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., et al. (2016). Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* 33, 2820–2835. doi: 10.1093/molbev/msw157

Huelsenbeck, J. P., Nielsen, R., and Bollback, J. P. (2003). Stochastic mapping of morphological characters. *Syst. Biol.* 52, 131–158. doi: 10.1080/10635150390192780

Ilnicki, T., Hasterok, R., and Szelkag, Z. (2010). Cytogenetic analysis of *Hieracium transylvanicum* (Asteraceae). *Caryologia* 63, 192–196. doi: 10.1080/00087114.2010.589726

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kaplan, Z., Fehrer, J., Bambasová, V., and Hellquist, C. B. (2018). The endangered Florida pondweed (*Potamogeton floridanus*) is a hybrid. *PLoS One* 13:e0195241. doi: 10.1371/journal.pone.0195241

Kaplan, Z., Jarolímová, V., and Fehrer, J. (2013). Revision of chromosome numbers of Potamogetonaceae: a new basis for taxonomic and evolutionary implications. *Preslia* 85, 421–482.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kellogg, E. A., and Appels, R. (1995). Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* 140, 325–343.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: r tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166

Kilian, N., Gemeinholzer, B., and Lack, H. W. (2009). "24. Cichorieae," in *Systematics, Evolution, and Biogeography of Compositae. International Association for Plant Taxonomy*, eds V. A. Funk, A. Susanna, T. F. Stuessy, and R. J. Bayer (Vienna: International Association for Plant Taxonomy), 343–383.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111

Kolano, B., Siwinska, D., McCann, J., and Weiss-Schneeweiss, H. (2015). The evolution of genome size and rDNA in diploid species of *Chenopodium* s.l. (Amaranthaceae). *Bot. J. Linn. Soc.* 179, 218–235. doi: 10.1111/boj.12321

Kotani, Y., Henderson, S. T., Suzuki, G., Johnson, S. D., Okada, T., Siddons, H., et al. (2014). The LOSS OF APOMEIOSIS (LOA) locus in *Hieracium praealtum* can function independently of the associated large-scale repetitive chromosomal structure. *New Phytol.* 201, 973–981. doi: 10.1111/nph.12574

Krahulcová, A., Krahulec, F., and Chapman, H. M. (2000). Variation in *Hieracium* subgen. *Pilosella* (Asteraceae): what do we know about its sources? *Folia Geobot.* 35, 319–338.

Krahulec, F., Krahulcová, A., Fehrer, J., Bräutigam, S., Plačková, I., and Chrtek, J. (2004). The Sudetic group of *Hieracium* subgen. *Pilosella* from the Krkonoše Mts: a synthetic view. *Preslia* 76, 223–243.

Krahulec, F., Krahulcová, A., Fehrer, J., Bräutigam, S., and Schuhwerk, F. (2008). The structure of the agamic complex of *Hieracium* subgen. *Pilosella* in the Šumava Mts and its comparison with other regions in Central Europe. *Preslia* 80, 1–26.

Krak, K., Álvarez, I., Caklová, P., Costa, A., Chrtek, J., and Fehrer, J. (2012). Development of novel low-copy nuclear markers for Hieraciinae (Asteraceae) and their perspective for other tribes. *Amer. J. Bot.* 99, e74–e77. doi: 10.3732/ajb.1100416

Krak, K., Caklová, P., Chrtek, J., and Fehrer, J. (2013). Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). *Heredity* 110, 138–151. doi: 10.1038/hdy.2012.100

Krak, K., and Mráz, P. (2008). Trichomes in the tribe Lactuceae (Asteraceae) – taxonomic implications. *Biologia* 63, 616–630.

Lan, T., and Albert, V. A. (2011). Dynamic distribution patterns of ribosomal DNA and chromosomal evolution in *Paphiopedilum*, a lady's slipper orchid. *BMC Plant Biol.* 11:126. doi: 10.1186/1471-2229-11-126

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Long, E. O., and Dawid, I. B. (1980). Repeated genes in eukaryotes. *Annu. Rev. Biochem.* 49, 727–764. doi: 10.1146/annurev.bi.49.070180.003455

Macas, J., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A., and Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS One* 6:e27335. doi: 10.1371/journal.pone.0027335

Mahelka, V., Kopecký, D., and Baum, B. R. (2013). Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass *Thinopyrum intermedium* (Triticeae). *Mol. Biol. Evol.* 30, 2065–2086. doi: 10.1093/molbev/mst106

Maire, R. (1937). Contributions à l'etude de la Flore de l'Afrique du Nord. *Fascicule 25. Bull. Soc. Hist. Nat. Afrique N.* 28, 332–420.

Merxmüller, H. (1975). Diploide Hieracien. *Anal. Inst. Botánico A.J. Cavanilles* 32, 189–196.

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* 14, 193–202. doi: 10.1093/bib/bbs012

Mlinarec, J., Šatović, Z., Malenica, N., Ivančić-Baće, I., and Besendorfer, V. (2012). Evolution of the tetraploid *Anemone multifida* (2n = 32) and hexaploid *A. baldensis* (2n = 48) (Ranunculaceae) was accompanied by rDNA loci loss and intergenomic translocation: evidence for their common genome origin. *Ann. Bot.* 110, 703–712. doi: 10.1093/aob/mcs128

Mráz, P. (2003). Mentor effects in the genus *Hieracium* s.str. (Compositae, Lactuceae). *Folia Geobot.* 38, 345–350. doi: 10.1007/BF02803204

Mráz, P., Chrtek, J., and Fehrer, J. (2011). Interspecific hybridization in the genus *Hieracium* (s. str.) – evidence for bidirectional gene flow and spontaneous allopolyploidization. *Plant Syst. Evol.* 293, 237–245. doi: 10.1007/s00606-011-0441-3

Mráz, P., Chrtek, J., Fehrer, J., and Plačková, I. (2005). Rare recent natural hybridization in *Hieracium* s.str. – evidence from morphology, allozymes and chloroplast DNA. *Plant Syst. Evol.* 255, 177–192.

Mráz, P., Filipaş, L., Bărbos, M. I., Kadlecová, J., Paštová, L., Belyayev, A., et al. (2019). An unexpected new diploid *Hieracium* from Europe: integrative taxonomic approach with a phylogeny of diploid *Hieracium* taxa. *Taxon* 68, 1258–1277. doi: 10.1002/tax.12149

Mráz, P., and Paule, J. (2006). Experimental hybridization in the genus *Hieracium* s.str. (Asteraceae): crosses between selected diploid taxa. *Preslia* 78, 1–26.

Mráz, P., and Zdvořák, P. (2019). Reproductive pathways in *Hieracium* s.str. (Asteraceae): strict sexuality in diploids and apomixis in polyploids. *Ann. Bot.* 123, 391–403. doi: 10.1093/aob/mcy137

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., et al. (2012). How to measure and test phylogenetic signal. *Methods Ecol. Evol.* 3, 743–756. doi: 10.1111/j.2041-210X.2012.00196.x

Myburg, A. A., Griffin, A. R., Sederoff, R. R., and Whetten, R. W. (2003). Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theor. Appl. Genet.* 107, 1028–1042. doi: 10.1007/s00122-003-1347-4

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nieto Feliner, G., and Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919. doi: 10.1016/j.ympev.2007.01.013

Okada, T., Ito, K., Johnson, S. D., Oelkers, K., Suzuki, G., Houben, A., et al. (2011). Chromosomes carrying meiotic avoidance loci in three apomictic eudicot *Hieracium* subgenus *Pilosella* species share structural features with two monocot apomicts. *Plant Physiol.* 157, 1327–1341. doi: 10.1104/pp.111.181164

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 255, 37–45. doi: 10.1098/rspb.1994.0006

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. E. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Amer. J. Bot.* 105, 385–403. doi: 10.1002/ajb2.1016

Posada, D., and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Raskina, O., Barber, J. C., Nevo, E., and Belyayev, A. (2008). Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenet. Genome Res.* 120, 351–357. doi: 10.1159/000121084

Rauscher, J. T., Doyle, J. J., and Brown, A. H. D. (2004). Multiple origins and nrDNA internal transcribed spacer homeologue evolution in the *Glycine tomentella* (Leguminosae) allopolyploid complex. *Genetics* 166, 987–998. doi: 10.1534/genetics.166.2.987

Revell, L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147. doi: 10.1016/0025-5564(81)90043-2

Rocha, L. C., Ferreira, M. T. M., Cunha, I. M. F., Mittelmann, A., and Techio, V. H. (2018). 45S rDNA sites in meiosis of *Lolium multiflorum* Lam.: variability, non-homologous associations and lack of fragility. *Protoplasma* 256, 227–235. doi: 10.1007/s00709-018-1292-3

Rogers, S. O., and Bendich, A. J. (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* 9, 509–520. doi: 10.1007/BF00015882

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Rosato, M., Moreno-Saiz, J. C., Galián, J. A., and Rosselló, J. A. (2015). Evolutionary site-number changes of ribosomal DNA loci during speciation: complex scenarios of ancestral and more recent polyploid events. *AoB Plants* 7:lv135. doi: 10.1093/aobpla/plv135

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. doi: 10.1093/oxfordjournals.molbev.a003974

Sastri, D. C., Hilu, K., Appels, R., Lagudah, E. S., Playford, J., and Baum, B. R. (1992). An overview of evolution in plant 5S DNA. *Plant Syst. Evol.* 183, 169–181. doi: 10.1007/BF00940801

Schlötterer, C., and Tautz, D. (1994). Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* 4, 777–783. doi: 10.1016/S0960-9822(00)00175-5

Simmons, M. P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381. doi: 10.1093/sysbio/49.2.369

Sleumer, H. (1956). Die Hieracien Argentiniens unter Berücksichtigung der Nachbarländer. *Bot. Jb. (Stuttgart)* 77, 85–148.

Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191, 528–535. doi: 10.1126/science.1251186

Soltis, D. E., Mavrodiev, E. V., Doyle, J. J., Rauscher, J., and Soltis, P. S. (2008). ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Syst. Bot.* 33, 7–20. doi: 10.1600/036364408783887401

Suda, J., Krahulcová, A., Trávníček, P., Rosenbaumová, R., Peckert, T., and Krahulec, F. (2007). Genome size variation and species relationships in *Hieracium* sub-genus *Pilosella* (Asteraceae) as inferred by flow cytometry. *Ann. Bot.* 100, 1323–1335. doi: 10.1093/aob/mcm218

Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571. doi: 10.1093/bioinformatics/btq228

Swofford, D. (2002). *PAUP*: Phylogenetic Analysis Using Parsimony (*and other Methods), Version 4*. Sunderland, MA: Sinauer.

Tutin, T. G., Heywood, V. H., Burges, N. A., Moore, D. M., Valentine, D. H., Walters, S. M., et al. (1976). *Flora Europaea*, Vol. 4. Cambridge: Cambridge University Press.

Volkov, R. A., Borisjuk, N. V., Panchuk, I. I., Schweizer, D., and Hemleben, V. (1999). Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. *Mol. Biol. Evol.* 16, 311–320. doi: 10.1093/oxfordjournals.molbev.a026112

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Syst. Biodivers.* 5, 261–276. doi: 10.1017/S1477200007002447

Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384

Wicke, S., Costa, A., Muñoz, J., and Quandt, D. (2011). Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol. Phylogenet. Evol.* 61, 321–332. doi: 10.1016/j.ympev.2011.06.023

Zagorski, D., Hartmann, M., Bertrand, Y. J. K., Paštová, L., Slavíková, R., Josefiová, J., et al. (2020). Characterization and dynamics of repeatomes in closely related species of *Hieracium* (Asteraceae) and their synthetic and apomictic hybrids. *Front. Plant Sci.* 11:591053. doi: 10.3389/fpls.2020.591053

Zahn, K. H. (1921–1923). "Hieracium," in *Das Pflanzenreich*, ed. A. Engler (Leipzig: Wilhelm Engelmann).

Zahradníček, J., and Chrtek, J. (2015). Cytotype distribution and phylogeography of *Hieracium intybaceum* (Asteraceae). *Bot. J. Linn. Soc.* 179, 487–498. doi: 10.1111/boj.12335

Zahradníček, J., Chrtek, J., Ferreira, M. Z., Krahulcová, A., and Fehrer, J. (2018). Genome size variation in the genus *Andryala* (Hieraciinae, Asteraceae). *Folia Geobot.* 53, 429–447. doi: 10.1007/s12224-018-9330-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

METHODS
published: 28 April 2021
doi: 10.3389/fpls.2021.656049

Targeted Enrichment of rRNA Gene Tandem Arrays for Ultra-Long Sequencing by Selective Restriction Endonuclease Digestion

Anastasia McKinlay[1†], Dalen Fultz[1,2†], Feng Wang[1,2] and Craig S. Pikaard[1,2]*

[1] Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, United States, [2] Howard Hughes Medical Institute, Indiana University, Bloomington, IN, United States

OPEN ACCESS

Edited by:
Sònia Garcia,
Consejo Superior de Investigaciones Científicas, Spanish National Research Council (CSIC), Spain

Reviewed by:
Martina Dvorackova,
Central European Institute of Technology (CEITEC), Czechia
Dongying Gao,
Small Grains and Potato Germplasm Research Unit (United States Department of Agriculture (USDA)-ARS), United States
Fernando A. Rabanal,
Max Planck Society (MPG), Germany

***Correspondence:**
Craig S. Pikaard
cpikaard@indiana.edu

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: 20 January 2021
Accepted: 06 April 2021
Published: 28 April 2021

Citation:
McKinlay A, Fultz D, Wang F and Pikaard CS (2021) Targeted Enrichment of rRNA Gene Tandem Arrays for Ultra-Long Sequencing by Selective Restriction Endonuclease Digestion. Front. Plant Sci. 12:656049. doi: 10.3389/fpls.2021.656049

Large regions of nearly identical repeats, such as the 45S ribosomal RNA (rRNA) genes of Nucleolus Organizer Regions (NORs), can account for major gaps in sequenced genomes. To assemble these regions, ultra-long sequencing reads that span multiple repeats have the potential to reveal sets of repeats that collectively have sufficient sequence variation to unambiguously define that interval and recognize overlapping reads. Because individual repetitive loci typically represent a small proportion of the genome, methods to enrich for the regions of interest are desirable. Here we describe a simple method that achieves greater than tenfold enrichment of *Arabidopsis thaliana* 45S rRNA gene sequences among ultra-long Oxford Nanopore Technology sequencing reads. This method employs agarose-embedded genomic DNA that is subjected to restriction endonucleases digestion using a cocktail of enzymes predicted to be non-cutters of rRNA genes. Most of the genome is digested into small fragments that diffuse out of the agar plugs, whereas rRNA gene arrays are retained. In principle, the approach can also be adapted for sequencing other repetitive loci for which gaps exist in a reference genome.

Keywords: Oxford Nanopore sequencing, *Arabidopsis thaliana*, ribosomal RNA gene enrichment, Nucleolus Organizer Region, NOR

INTRODUCTION

Many eukaryotic genomes have chromosomal loci that consist of hundreds, if not thousands, of nearly identical repeats, sometimes spanning millions of basepairs. Examples include the AT-rich satellites of pericentromeric regions (Aldrup-Macdonald and Sullivan, 2014), ribosomal RNA (rRNA) gene repeats (Gerbi, 1985; Flavell, 1986) and tandemly repeated transposable element (TE)-derived sequences (Ahmed and Liang, 2012). Distinguishing one repeat from the next can be difficult, precluding the easy determination of how individual repeats are arranged at the locus. As a result, repetitive loci are often miss-assembled or absent from assemblies of eukaryotic genomes (Biscotti et al., 2015).

Two long-read sequencing technologies have greatly improved the ability to close gaps in sequenced genomes, namely Pacific Biosciences (PacBio) SMRT sequencing and Oxford Nanopore MinION sequencing (Besser et al., 2018; Michael et al., 2018). PacBio sequencing can yield reads that are 10–100 kb in length, with the potential to obtain multiple reads of the same DNA

fragment. This allows one to obtain consensus sequence reads whose accuracy rivals that of short-read Illumina or Sanger sequencing. By obtaining highly accurate long reads, single nucleotide polymorphisms (SNPs) can be identified among repeats that are nearly identical in sequence. For instance, PacBio sequencing has been used to identify subtle sequence differences among *Arabidopsis thaliana* rRNA genes (Havlova et al., 2016) that are each ∼10 kb in length. However, PacBio sequencing reads are not long enough for assembly of rRNA genes into long contigs due to the paucity of variation that is unique and thus not shared by numerous genes.

Sequencing using Oxford Nanopore Technology (ONT) yields ultra-long reads that can be hundreds of kilobases in length, but with an accuracy of only 75–95% (Rang et al., 2018). The technology is especially well-suited to identifying chromosomal deletions, insertions, or rearrangements. However, the high error rate of ONT sequencing is problematic for assembly of repetitive regions in which there are few sequence differences to discriminate each repeat (Michael et al., 2018). For successful assembly of these repetitive regions, having multiple overlapping ONT reads is necessary, thus allowing consensus sequences to be deduced to improve the accuracy and confidently identify SNPs and other subtle variation (Ebler et al., 2019).

Targeted enrichment aims to increase sequencing coverage for a region of interest (reviewed in Good, 2011; Kozarewa et al., 2015). Current methods are mostly designed for short-read sequencing, but some are amenable to ultra-long sequencing of large repetitive genomic regions. For instance, the clustered regularly interspaced short palindromic repeats (CRISPR) Cas9 system has been used for targeted sequencing (Bennett-Baker and Mueller, 2017; Gabrieli et al., 2018; Nachmanson et al., 2018). In this approach, megabase-sized genomic regions of interest are cleaved and purified from the rest of the genome by pulsed-field gel electrophoresis. Although compatible with PacBio and ONT sequencing, the strategy poses technical challenges and requires large amounts of starting DNA.

Nucleolus Organizer Regions (NORs) are missing from current genome assemblies of multicellular eukaryotes. The number of NORs in a genome varies between species, and within a species the number of rRNA genes within a NOR can vary between individuals and even among cells of an individual (Stults et al., 2008; Tucker et al., 2010; McStay, 2016). Due to the lack of substantial sequence variation among rRNA genes repeats, NORs of reference genomes are sometimes represented by a single rRNA gene repeat, with actual copy numbers and NOR sizes remaining unknown (Treangen and Salzberg, 2011). ONT sequencing holds promise for the assembly of NORs but has not yet been used to assemble complete NORs (Michael et al., 2018). This poses an obstacle to studies of NOR recombination, replication and repeat homogenization as well as studies of large-scale rRNA gene regulation. For instance, our laboratory is interested in understanding why the two NORs of the *Arabidopsis thaliana* strain Col-0 differ in expression, with one being constitutively active and the other falling silent during development (Chandrasekhara et al., 2016; Mohannath et al., 2016), an epigenetic phenomenon known as nucleolar dominance (McStay, 2006; Tucker et al., 2010). Each *Arabidopsis thaliana* NOR is composed of hundreds of tandemly repeated rRNA genes that are each ∼10 kb in length, such that both span several million basepairs of DNA (Copenhaver and Pikaard, 1996b). Evidence suggests that chromosomal context or position, rather than rRNA gene sequence variation, is responsible for the differential expression of the two NORs (Chandrasekhara et al., 2016; Mohannath et al., 2016), but the chromosomal basis for nucleolar dominance remains unknown. The possibility exists that one of more locus control elements might be embedded within the NORs, thus their complete sequence is desirable.

Here, we describe a simple method for enrichment of ultra-high molecular weight rRNA gene tandem arrays using a cocktail of restriction endonucleases predicted not cut an rRNA gene reference sequence. When used to treat genomic DNA embedded in an agarose plug, the enzymes digest most of the genome into small fragments that passively diffuse out of the agarose plug (Fritz and Musich, 1990). This depletes the plug of unwanted DNA fragments while retaining large DNA fragments that include rRNA gene arrays. Using *A. thaliana* rRNA genes as our example, the strategy yields a tenfold increase in ONT sequencing reads corresponding to rRNA genes. In principle, the method should also be adaptable for the enrichment of other target sequences, simply by altering the choice of restriction endonucleases.

MATERIALS AND METHODS

Plant Material

Arabidopsis thaliana Col-0 plants (Arabidopsis Biological Resource Center stock #CS 70000) were surface-sterilized and grown on agar plates containing 0.5X Murashige and Skoog medium (MS). Plants were harvested after 2 weeks of growth under short-day conditions (8 h light, 16 h dark).

Preparation of Genomic DNA

Ultra-high molecular weight DNA was purified from *Arabidopsis thaliana* Col-0 plants by following the Bionano Prep Plant Tissue DNA Isolation, Liquid Nitrogen Grinding Protocol (Bionano document number 30177)[1] (summarized in **Supplementary Figure 1**). Briefly, 2 g of fresh tissue was placed in a pre-chilled (overnight at −80°C) mortar and ground in liquid nitrogen using a pre-chilled pestle then resuspended in 40 mL of ice-cold Bionano Prep Plant Tissue Homogenization Buffer (Part #20283) supplemented with 2-mercaptoethanol (0.2% final concentration) and 1 mM spermine-spermidine (known as Plant Tissue Homogenization Buffer *plus*). The suspension was passed sequentially through 100 μm (VWR Cat# 21008-950) and 40 μm cell strainers (VWR Cat# 21008-949) into a pre-chilled 50 mL conical tube on ice. Nuclei were then pelleted by centrifugation at 3,500 × g for 20 min at 4°C using a swinging bucket rotor. After discarding the supernatant, the pellet was resuspended in 1 mL of Plant Tissue Homogenization Buffer *plus* buffer with the assistance of a small paint brush that had been presoaked

[1]https://bionanogenomics.com/wp-content/uploads/2018/02/30177-Bionano-Prep-Plant-Tissue-DNA-Isolation-Liquid-Nitrogen-Grinding-Protocol.pdf

in the buffer. The resuspended nuclei were further diluted with 40 mL of ice-cold Plant Tissue Homogenization Buffer *plus* buffer and then subjected to centrifugation at 60 × g for 2 min at 4°C using a swinging bucket rotor to remove cell debris, with no braking during rotor deceleration. The supernatant was then subjected to another 40 μm filtration step (VWR Cat# 21008-949). Nuclei were collected by centrifugation at 3,500 × g for 20 min at 4°C using a swinging bucket rotor and washed three times by resuspension in in 30 mL of ice-cold Plant Tissue Homogenization Buffer *plus* and re-pelleting at 3,500 × g for 20 min at 4°C. The final nuclei pellet was resuspended in 3 mL of ice-cold Plant Tissue Homogenization Buffer *plus* and applied on top of the Density Gradient (Bionano Prep Density Gradient, catalog numbers 20281 and 20280). After centrifugation at 4,500 × g for 40 min at 4°C in a swinging bucket rotor, with no braking during deceleration, nuclei were recovered from the gradient and collected into a pre-chilled 15 mL conical tube on ice. Nuclei were then diluted with 14 mL of ice-cold Plant Tissue Homogenization Buffer *plus* and collected by centrifugation at 2,500 × g for 10 min at 4°C in a swinging bucket rotor, with no braking during deceleration. After carefully decanting the supernatant, nuclei were resuspended in 50 μL of ice-cold Density Gradient Buffer (Bionano Prep Density Gradient Buffer Cat #20280). The nuclei were then equilibrated to 43°C for 3 min and mixed with 30 μL of molten 2% agarose equilibrated at 43°C (CleanCut Low Melting Point, Bio-Rad, Cat# 1703594) using a wide-bore tip, and pipetted into a Bio-Rad CHEF Disposable Plug Mold (Bio-Rad, Cat# 170-3713). The final agarose concentration of the plugs was 0.82%. Plug molds were incubated at 4°C for 15 min to solidify the agarose.

Plugs containing embedded nuclei were then subjected to Proteinase K (20 mg/mL; 0.8 mg/plug; QIAGEN, Cat# 19131) and RNase A (100 μg/mL; 1 μg/plug; QIAGEN, Cat# 19101) digestion and washed according to the Bionano Prep Plant Tissue DNA Isolation, Liquid Nitrogen Grinding Protocol (document #30177). For rRNA gene enrichment, embedded nuclei were treated with a restriction endonuclease cocktail composed of six enzymes predicted to be rRNA gene non-cutters. Briefly, agarose plugs were placed in 50 mL conical tubes and were first incubated in 10 mL of T10E10 buffer (10 mM Tris-HCl, 10 mM EDTA, pH 8.0) supplemented with 2 mM PMSF for 1 h at 4°C. Plugs were then washed four times, 30 min each, at room temperature in 10 mL of T10E10 buffer without PMSF. Next, individual agarose plugs were washed twice, 1 h, at room temperature, with 1 ml of 1× restriction enzyme buffer [1× CutSmart buffer (NEB)]. After a second wash, the plug was incubated with 200 μL of 1× CutSmart buffer (NEB) containing 50 U each of the restriction endonucleases BstZ17I-HF (NEB #R3594), SpeI-HF (NEB #R3133), BclI-HF (NEB #R3160), SnaBI (NEB #R0130), MscI (NEB #R0534), and PvuII-HF (NEB #R3151) at 37°C overnight. The buffer was then removed and replaced with 500 μL of 20 mM Tris-HCl, 50 mM EDTA, pH 8.0 and incubated at 10 min at room to stop further digestion. The agarose plug was then subjected to 5 wash steps, each 15 min at room temperature, with 10 mL of TE buffer in order to deplete the plugs of short DNA digestion products fragments that can diffuse from the plugs, unlike large DNA fragments that are retained.

Plugs were then melted at 70°C for 2 min, equilibrated at 43°C for 5 min, and then incubated with 2 μL of 0.5 U/μL agarase (ThermoFisher Scientific, Cat# EO0461) per plug at 43°C for 45 min to digest the agar and liberate the encapsulated DNA. The resulting solution was then subjected to drop dialysis by applying genomic DNA on a 0.1μm dialysis membrane (Millipore, Cat# VCWP04700) floating on the surface of 15 mL of TE buffer inside a 6 cm petri dish. Dialysis was at room temperature for 45 min. DNA was assessed for quantity and quality using a Qubit dsDNA BR Assay kit and by agarose gel electrophoresis.

Quantitative PCR (qPCR) Assay

Genomic DNA was subjected to electrophoresis using a 0.7% agarose gel in 0.5× TBE buffer for 1 h and 15 min at 100 V. The gel was then stained with GelRed diluted 1:10,000 in water (GoldBio # G-725-100). The intensively stained DNA band >23 kb, consisting of all DNA fragments greater than the resolving limit of the gel, was then excised and the DNA extracted using a QIAEX II Gel Extraction Kit (QIAGEN #20021). The resulting DNA was assessed by qPCR for the presence of 25S rRNA gene sequences and actin genes using the following forward (F) and reverse (R) primers:

25S_F: GAGTGCTTGAAATTGTCGGGAGGGAAG;
25S_R: CGAATCTTAGCGACAAAGGGCTGAATC;
actin_F: GAGAGATTCAGATGCCCAGAAGTC;
actin_R: TGGATTCCAGCAGCTTCCA.

Oxford Nanopore MinION Sequencing and Analysis

DNA sequencing library preparation was performed using the Oxford Nanopore Rapid Library Kit (RAD-004). Sequencing was performed using a MinION sequencer with an R9.4.1 flow cell. Base-calling of raw ONT sequencing data was performed using Albacore v2.3.1. General FASTQ read statistics were calculated by NanoPlot (v 1.13). Length count distribution was analyzed using an R ggplot2 package. Statistical analysis was performed with GraphPad Prizm8 software. Percent sequence identity was calculated using minimap2 and the following Perl script (Li, 2018): <minimap2 -c reference.fasta query.fasta | perl -ane "if(/tp:A:P/&&/NM:i:(\d+)/){$n+ = $1; $m+ = $1 while/(\d+)M/g;$g+ = $1,++$o while/(\d+)[ID]/g} END{print(($n-$g+$o)/($m+$o),"\n")}" >.

RESULTS

Arabidopsis thaliana plants have two NORs that are located on the short arm of chromosomes 2 and 4 (Copenhaver et al., 1995; Copenhaver and Pikaard, 1996a). Each NOR is estimated to span ∼4 Mbp and consist of ∼350 to 400 rRNA genes that are each ∼10 kb in length (Tucker et al., 2010). We used the New England Biolabs (NEB) cutter tool (Vincze et al., 2003) to examine a full-length *Arabidopsis thaliana* (ecotype Col-0) 45S rRNA gene sequence (Chandrasekhara et al., 2016) and identify a list of 24 restriction endonucleases (RE) whose recognition sites are missing within the reference rRNA gene sequence. We then performed virtual *in silico* digestions of the *Arabidopsis thaliana*

Col-0 reference genome (TAIR10) to identify the subset of these enzymes that cut most frequently in the genome (**Supplementary Figures 2**, **3**), selecting six that are each predicted to digest genomic DNA to a median fragment length of 5 kb or smaller (**Figure 1A**) and that display 100% activity in NEB CutSmart buffer. *In silico* digestion using a cocktail of all six enzymes predicted that they would cut genomic *A. thaliana* (ecotype Col-0) into DNA fragments with a mean length of 522 bp (**Figure 1B**).

To test the restriction endonuclease cocktail, Col-0 genomic DNA immobilized in agarose plugs was subjected to in-plug digestion as described in Mohannath et al. (2016). Sizes of genomic DNA fragments were visualized following electrophoresis through a 0.7% agarose gel in TBE buffer. As shown in **Figure 2A**, digestion of genomic DNA with the enzyme cocktail resulted in a significant reduction of high molecular weight DNA (top band) and the appearance of DNA fragments mostly smaller than 5 kb (**Figure 2A**, lane 3). In contrast, treatment of genomic DNA with the rRNA gene-specific endonuclease I-PpoI, which cleaves once per rRNA gene (Muscarella et al., 1990; Copenhaver and Pikaard, 1996a), resulted in a band of ∼10 kb, the expected rRNA gene unit length (**Figure 2A**, lane 2). Quantitative PCR analysis of genomic DNA extracted from the top gel band revealed similar quantities of rRNA gene sequences in undigested genomic DNA and DNA cut by the six-enzyme cocktail. By contrast, digestion with I-PpoI depletes rRNA gene sequences, as expected (**Figure 2B**, top panel). The six-enzyme cocktail reduced the level of control actin gene DNA by 1,000-fold relative to undigested DNA (**Figure 2B**, bottom panel).

Next, we performed ONT sequencing to test the degree to which digestion with the six-enzyme cocktail enriches for reads containing rRNA gene repeats. For this experiment, agarose-embedded nuclei from *Arabidopsis thaliana* Col-0 plants (denoted as whole genome, or WG nuclei in **Figure 3A**) were first subjected to digestion with Proteinase K and RNase A. Half of the sample was then incubated with the cocktail of six restriction endonucleases (RE nuclei) and the other half received only buffer. The resulting DNA was prepared via ONT's rapid library kit (RAD-004) and sequenced on a MinION. Reads were mapped to the Col-0 reference genome (version TAIR10) with the alignment program ngmlr, using the default cutoffs (minimum identity of 65% and at least 25% of the read length aligned to the reference sequence) (Sedlazeck et al., 2018). The resulting FASTQ files (total sequenced DNA) were aligned to a rRNA gene consensus sequence (Chandrasekhara et al., 2016) in order to identify ribosomal gene DNA reads and separate them from remaining TAIR10-mappable sequences (non-ribosomal DNA reads). The sequence identity of the basecalled reads when aligned to the *A. thaliana* nuclear genome (TAIR10) was 85.78%. The sequence identity of the ribosomal gene reads was 86.12% for the 45S rRNA gene region (excluding the variable 3'ETS region). Sequencing statistics are shown in **Figure 3A**.

A **Selected rRNA gene non-cutters**

Enzymes	Sequence specificity	Median DNA fragment length upon digestion of *A. thaliana* TAIR 10 genome
BstZ17I-HF	GTATAC	3863
SpeI-HF	ACTAGT	3149
BclI-HF	TGATCA	1490
SnaBI	TACGTA	3723
MscI	TGGCCA	4632
PvuII-HF	CAGCTG	3552

B **Expected 6-enzyme digestion profile**

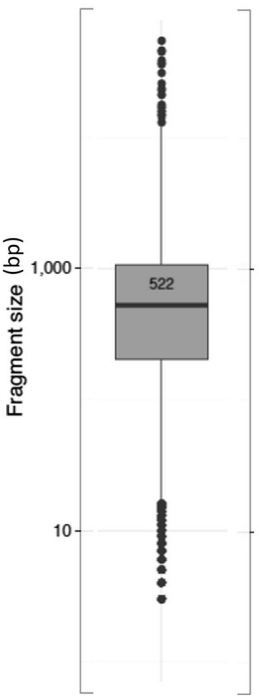


FIGURE 1 | Predicted 45S rRNA gene non-cutting restriction endonucleases used as a cocktail for digestion of genomic DNA. **(A)** Names and sequence specificities of the enzymes and median fragment sizes obtained upon *in silico* digestion of *A. thaliana* Col-0 genomic DNA. **(B)** Expected *in silico* digestion mean size and distribution for Col-0 genomic DNA subjected to digestion by the six-enzyme cocktail.

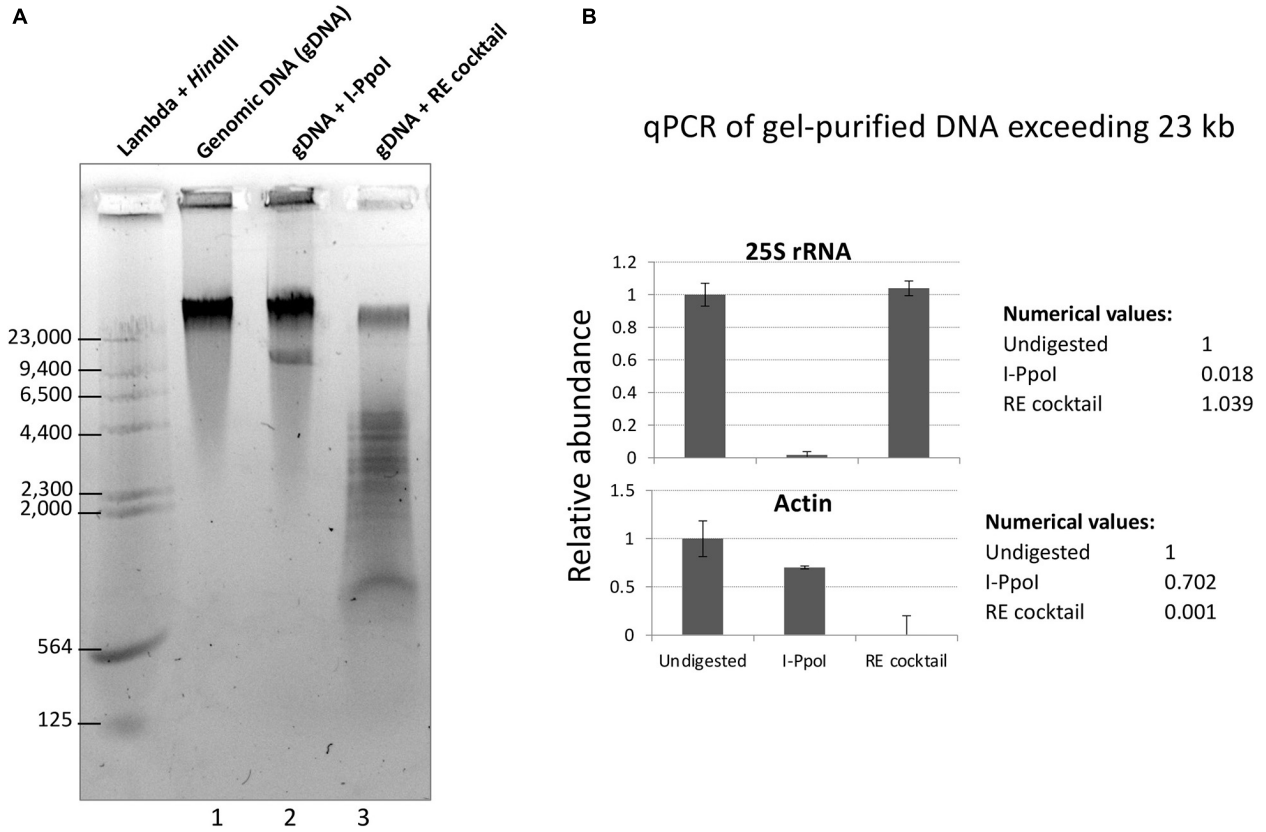


FIGURE 2 | Digestion of genomic DNA with a cocktail of six restriction endonucleases (REs) predicted not to cut rRNA genes. **(A)** Col-0 gDNA that was either undigested (lane 1) or digested with I-PpoI (Lane 2) or the RE cocktail (lane 3) was separated on a 0.7% agarose gel and stained with GelRed dye. Bacteriophage Lambda DNA digested with *Hind*III provided size markers. **(B)** Summary of quantitative PCR (qPCR) results for rRNA gene (top panel) and actin (bottom panel) detection in undigested, I-PpoI, or RE cocktail-digested samples. qPCR was performed on gDNA extracted from the top gel band, containing DNA > 23 kb.

By definition, the N50 value of a sequencing run indicates a read length at which half of the total yield is in read lengths equal to or greater than this value. Consistent with targeted digestion of genomic DNA other than rRNA genes, the six-enzyme cocktail treatment greatly reduced the N50 read length for non-ribosomal DNA reads, whereas the N50 value for ribosomal DNA reads was less affected. Importantly, the percentage of sequenced rRNA gene bases (% ribosomal DNA bases) increased 13.5-fold, from 4.3% in the undigested control sample (Nuclei WG) to 57.9% in the sample digested with the six-enzyme cocktail (Nuclei RE). Additionally, the read length distribution of the restriction enzyme digested sample shows a statistically significant difference (unpaired *t*-test, p-value < 0.0001) between the non-ribosomal DNA reads and ribosomal RNA gene reads (**Figure 3**).

DISCUSSION

Gaps in published reference genomes can be millions of basepairs in size and can consist of repeats with nearly identical sequences, as is the case for NORs and pericentromeric repeats. Assembly of these regions can benefit from ultra-long ONT sequencing that yields reads that span multiple repeat units. However, a high depth of coverage is needed to assure accuracy and continuity of the assembly. Obtaining the needed coverage can be costly when the repeat region represents only a fraction of the genome to be sequenced.

In our proof-of-concept approach described in this brief report, we explored whether targeted enrichment of highly repetitive ribosomal RNA gene arrays can be combined with Oxford Nanopore Technology (ONT) sequencing in order to increase read depth coverage for *A. thaliana* NOR regions. Without enrichment, ribosomal RNA gene sequences are expected to account for ∼4.3% of the sequencing data, based on an estimated size of ∼8 Mbp for the two NORs (Copenhaver and Pikaard, 1996b). Restriction endonuclease-mediated sequence enrichment increased the proportion of rRNA gene reads by ∼13.5-fold. In our test, we used an RE cocktail chosen based on the sequence of a reference consensus gene sequence. A caveat to this approach is that rRNA gene sequence variants that can be cut by one or more of these enzymes may occur at low frequency. Thus, rRNA gene reads obtained by direct sequencing of genomic DNA, without restriction endonuclease digestion, should also be conducted. The latter can provide unbiased "scaffold reads" to which the "enriched" reads can be matched to increase the depth of sequence coverage. Alternatively, two

A **ONT sequencing summary for whole genome (WG) vs. restriction endonuclease (RE)-digested DNA**

Library	Ribosomal DNA reads				Non-ribosomal DNA reads			
	# KBases	Read length N50 (bp)	Max Read (bp)	% ribosomal DNA bases	# KBases	Read length N50 (bp)	Max Read (bp)	% non-ribosomal DNA bases
Nuclei WG	170,914	31,941	327,453	4.3 %	3,786,656	31,851	396,939	95.7 %
Nuclei RE	68,734	27,932	136,682	57.9 %	50,016	3,503	341,013	42.1 %

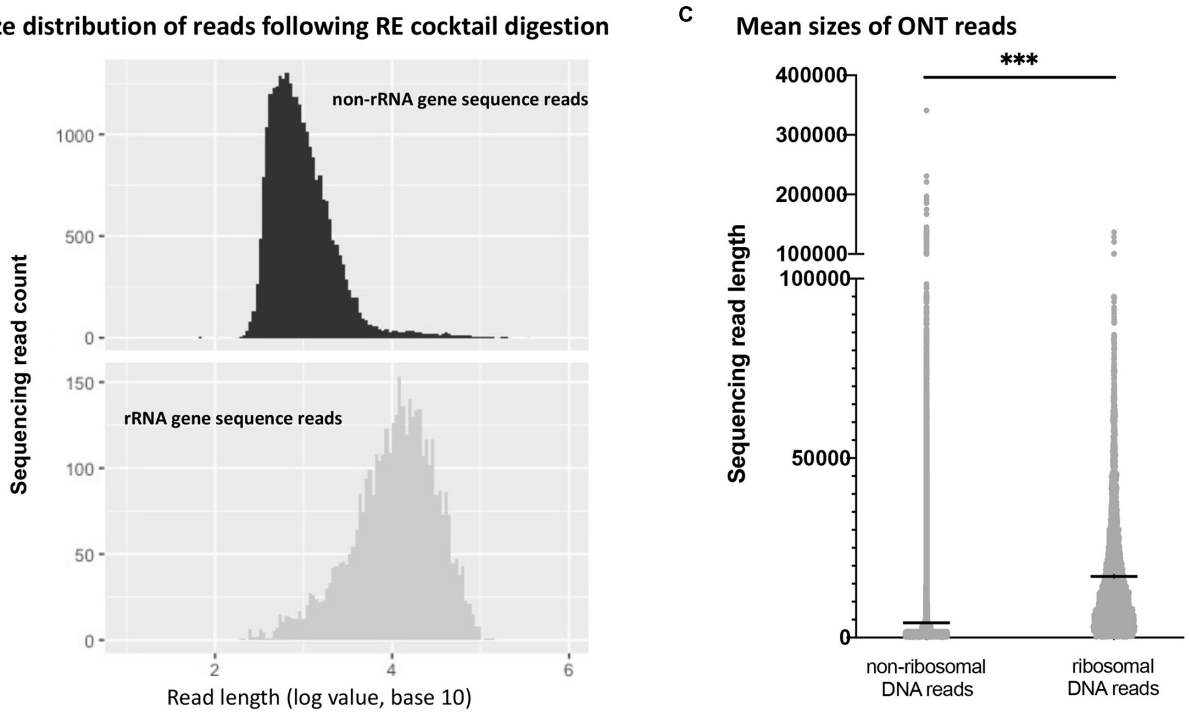


FIGURE 3 | Oxford Nanopore sequencing results for whole genome (WG) vs. restriction endonuclease (RE)-digested DNA. **(A)** Effect of restriction endonuclease cocktail digestion on read distribution for non-ribosomal and ribosomal RNA gene sequences. **(B)** Non-ribosomal DNA (top) and ribosomal RNA gene DNA (bottom) read numbers are plotted as a function of read length. **(C)** Sequencing read mean length (horizontal black lines within the distributions) for non-ribosomal and ribosomal RNA gene sequences are significantly different (*** denotes unpaired *t*-test values of $p < 0.0001$).

or more different restriction endonuclease cocktails could be employed, designed to account for rare variants that might be cut using one cocktail but not another. These and other strategies for improving ultra-long sequencing coverage will likely be needed to achieve complete *de novo* assembly of NORs (Rang et al., 2018).

ONT sequencing of bacterial artificial chromosomes (BACs) is another way to obtain sequences for cloned arrays of tandem repeats, as recently demonstrated for BAC-cloned *Arabidopsis thaliana* rRNA gene arrays (Sims et al., 2021). An advantage of BACs is the ability to achieve high sequence coverage for the cloned insert, allowing high per-base accuracy. However, unlike direct genomic DNA sequencing, BACs tend to be ∼100 kb in size, which may not be long enough to identify sufficient variation for overlapping sequences to be identified and longer contigs assembled. BACs are also known to recombine, especially BACs containing cloned repetitive regions (Mozo et al., 1998). Thus, secondary confirmation of gene arrangements determined by BAC sequencing, obtained by direct genomic DNA sequencing to obtain even longer reads, is desirable and can have synergistic benefits, with ultra-long genomic DNA sequences serving as scaffolds for contig assembly and BAC sequences providing high accuracy at each nucleotide position within the contig.

Despite their overall conservation, 45S rRNA gene sequences are diverse in eukaryotes such that restriction endonuclease enrichment strategies must be adapted on a species-by-species, and even strain-by-strain, basis (Rabanal et al., 2017). However, the large selection of commercially available restriction enzymes makes it likely that the strategy be adapted for most species simply by altering the mix of restriction endonucleases. Exceptionally long read lengths will also likely be required to assemble NORs in species such as humans, in which individual rRNA gene repeats are four times longer (∼42 kb) than the ∼10 kb rRNA

gene repeats of Arabidopsis. Thus, it is noteworthy that some of the longest reported ONT read lengths been obtained using human genomic DNA (Jain et al., 2018), far surpassing the ONT read lengths we have obtained thus far for Arabidopsis. Keeping these considerations in mind, if one has preliminary knowledge of repeat size and sequence variation, and the length of ONT reads possible for the species and cells being studied, enrichment strategies can likely be designed to obtain long reads to help assemble repetitive loci composed of highly similar genes or DNA elements.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CP conceived the experiments. AM designed and performed the experiments. FW performed *in silico* digestion of Col-0 gDNA by restriction enzymes. AM and DF analyzed results of ONT sequencing runs. AM and CP wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funds to CP as an Investigator of the Howard Hughes Medical Institute and by endowment funds of the Carlos O. Miller Professorship at Indiana University.

ACKNOWLEDGMENTS

We thank our fellow Pikaard lab colleagues for valuable discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.656049/full#supplementary-material

REFERENCES

Ahmed, M., and Liang, P. (2012). Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp. Funct. Genomics* 2012:947089. doi: 10.1155/2012/947089

Aldrup-Macdonald, M. E., and Sullivan, B. A. (2014). The past, present, and future of human centromere genomics. *Genes (Basel)* 5, 33–50. doi: 10.3390/genes5010033

Bennett-Baker, P. E., and Mueller, J. L. (2017). CRISPR-mediated isolation of specific megabase segments of genomic DNA. *Nucleic Acids Res.* 45:e165. doi: 10.1093/nar/gkx749

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., and Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin. Microbiol. Infect.* 24, 335–341. doi: 10.1016/j.cmi.2017.10.013

Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Res.* 23, 415–420. doi: 10.1007/s10577-015-9499-z

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Copenhaver, G. P., Doelling, J. H., Gens, S., and Pikaard, C. S. (1995). Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*. *Plant J.* 7, 273–286. doi: 10.1046/j.1365-313x.1995.7020273.x

Copenhaver, G. P., and Pikaard, C. S. (1996a). RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J.* 9, 259–272. doi: 10.1046/j.1365-313x.1996.09020259.x

Copenhaver, G. P., and Pikaard, C. S. (1996b). Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* 9, 273–282. doi: 10.1046/j.1365-313x.1996.09020273.x

Ebler, J., Haukness, M., Pesout, T., Marschall, T., and Paten, B. (2019). Haplotype-aware diplotyping from noisy long reads. *Genome Biol.* 20:116. doi: 10.1186/s13059-019-1709-0

Flavell, R. B. (1986). Repetitive DNA and chromosome evolution in plants. *Philos. Trans. R. Soc. Lond.* 312, 227–242. doi: 10.1098/rstb.1986.0004

Fritz, R. B., and Musich, P. R. (1990). Unexpected loss of genomic DNA from agarose gel plugs. *Biotechniques* 9, 542, 544, 546–550.

Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y., and Ebenstein, Y. (2018). Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.* 46:e87. doi: 10.1093/nar/gky411

Gerbi, S. A. (1985). "Evolution of ribosomal DNA," in *Molecular Evolutionary Genetics*, ed. R. J. McIntyre (New York, NY: Plenum Press), 419–517. doi: 10.1007/978-1-4684-4988-4_7

Good, J. M. (2011). Reduced representation methods for subgenomic enrichment and next-generation sequencing. *Methods Mol. Biol.* 772, 85–103. doi: 10.1007/978-1-61779-228-1_5

Havlova, K., Dvorackova, M., Peiro, R., Abia, D., Mozgova, I., Vansacova, L., et al. (2016). Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol. Biol.* 92, 457–471. doi: 10.1007/s11103-016-0524-1

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060

Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., and Hendrickson, C. L. (2015). Overview of target enrichment strategies. *Curr. Protoc. Mol. Biol.* 112, 7.21.1–7.21.23. doi: 10.1002/0471142727.mb0721s112

Li, H. (2018). On the Definition of Sequence Identity. Available online at: https://lh3.github.io/2018/11/25/onthe-definition-of-sequence-identity

McStay, B. (2006). Nucleolar dominance: a model for rRNA gene silencing. *Genes Dev.* 20, 1207–1214. doi: 10.1101/gad.1436906

McStay, B. (2016). Nucleolar organizer regions: genomic 'dark matter' requiring illumination. *Genes Dev.* 30, 1598–1610. doi: 10.1101/gad.283838.116

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., et al. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9:541. doi: 10.1038/s41467-018-03016-2

Mohannath, G., and Pikaard, C. S. (2016). Analysis of rRNA gene methylation in *Arabidopsis thaliana* by CHEF-conventional 2D gel electrophoresis. *Methods Mol. Biol.* 1455, 183–202. doi: 10.1007/978-1-4939-3792-9_14

Mohannath, G., Pontvianne, F., and Pikaard, C. S. (2016). Selective nucleolus organizer inactivation in *Arabidopsis* is a chromosome position-effect

phenomenon. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13426–13431. doi: 10.1073/pnas.1608140113

Mozo, T., Fischer, S., Shizuya, H., and Altmann, T. (1998). Construction and characterization of the IGF *Arabidopsis* BAC library. *Mol. Gen. Genet.* 258, 562–570. doi: 10.1007/s004380050769

Muscarella, D. E., Ellison, E. L., Ruoff, B. M., and Vogt, V. M. (1990). Characterization of I-Ppo I, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol. Cell. Biol.* 10, 3386–3396. doi: 10.1128/mcb.10.7.3386

Nachmanson, D., Lian, S., Schmidt, E. K., Hipp, M. J., Baker, K. T., Zhang, Y., et al. (2018). Targeted genome fragmentation with CRISPR/Cas9 enables fast and efficient enrichment of small genomic regions and ultra-accurate sequencing with low DNA input (CRISPR-DS). *Genome Res.* 28, 1589–1599. doi: 10.1101/gr.235291.118

Rabanal, F. A., Mandakova, T., Soto-Jimenez, L. M., Greenhalgh, R., Parrott, D. L., Lutzmayer, S., et al. (2017). Epistatic and allelic interactions control expression of ribosomal RNA gene clusters in *Arabidopsis thaliana*. *Genome Biol.* 18:75. doi: 10.1186/s13059-017-1209-z

Rang, F. J., Kloosterman, W. P., and de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19:90. doi: 10.1186/s13059-018-1462-9

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15, 461–468. doi: 10.1038/s41592-018-0001-7

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlogelhofer, P. (2021). Sequencing of the *Arabidopsis* NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12:387. doi: 10.1038/s41467-020-20728-6

Stults, D. M., Killen, M. W., Pierce, H. H., and Pierce, A. J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res.* 18, 13–18. doi: 10.1101/gr.6858507

Treangen, T. J., and Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46. doi: 10.1038/nrg3117

Tucker, S., Vitins, A., and Pikaard, C. S. (2010). Nucleolar dominance and ribosomal RNA gene silencing. *Curr. Opin. Cell Biol.* 22, 351–356. doi: 10.1016/j.ceb.2010.03.009

Vincze, T., Posfai, J., and Roberts, R. J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* 31, 3688–3691. doi: 10.1093/nar/gkg526

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fpls.2021.672879

Horizontally Acquired nrDNAs Persist in Low Amounts in Host *Hordeum* Genomes and Evolve Independently of Native nrDNA

*Karol Krak[1,2], Petra Caklová[1], David Kopecký[3], Frank R. Blattner[4,5] and Václav Mahelka[1]**

[1] Czech Academy of Sciences, Institute of Botany, Prùhonice, Czechia, [2] Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague 6, Czechia, [3] Czech Academy of Sciences, Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czechia, [4] Experimental Taxonomy, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany, [5] German Centre of Integrative Biodiversity Research (iDiv) Halle–Jena–Leipzig, Leipzig, Germany

Nuclear ribosomal DNA (nrDNA) has displayed extraordinary dynamics during the evolution of plant species. However, the patterns and evolutionary significance of nrDNA array expansion or contraction are still relatively unknown. Moreover, only little is known of the fate of minority nrDNA copies acquired between species via horizontal transfer. The barley genus *Hordeum* (Poaceae) represents a good model for such a study, as species of section *Stenostachys* acquired nrDNA via horizontal transfer from at least five different panicoid genera, causing long-term co-existence of native (*Hordeum*-like) and non-native (panicoid) nrDNAs. Using quantitative PCR, we investigated copy number variation (CNV) of nrDNA in the diploid representatives of the genus *Hordeum*. We estimated the copy number of the foreign, as well as of the native ITS types (ribotypes), and followed the pattern of their CNV in relation to the genus' phylogeny, species' genomes size and the number of nrDNA loci. For the native ribotype, we encountered an almost 19-fold variation in the mean copy number among the taxa analysed, ranging from 1689 copies (per 2C content) in *H. patagonicum* subsp. *mustersii* to 31342 copies in *H. murinum* subsp. *glaucum*. The copy numbers did not correlate with any of the genus' phylogeny, the species' genome size or the number of nrDNA loci. The CNV was high within the recognised groups (up to 13.2 × in the American I-genome species) as well as between accessions of the same species (up to 4×). Foreign ribotypes represent only a small fraction of the total number of nrDNA copies. Their copy numbers ranged from single units to tens and rarely hundreds of copies. They amounted, on average, to between 0.1% (*Setaria* ribotype) and 1.9% (*Euclasta* ribotype) of total nrDNA. None of the foreign ribotypes showed significant differences with respect to phylogenetic groups recognised within the sect. *Stenostachys*. Overall, no correlation was found between copy numbers of native and foreign nrDNAs suggesting the sequestration and independent evolution of native and non-native nrDNA arrays. Therefore, foreign nrDNA in *Hordeum* likely poses a dead-end by-product of horizontal gene transfer events.

Keywords: copy number variation (CNV), horizontal gene transfer (HGT), internal transcribed spacer (ITS), nuclear ribosomal DNA (nrDNA), fluorescent *in situ* hybridisation (FISH), phylogeny, qPCR (quantitative PCR), xenolog

OPEN ACCESS

Edited by:
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia

Reviewed by:
Natalia Borowska-Zuchowska, University of Silesia in Katowice, Poland
Anna Dimitrova, Institute of Plant Physiology and Genetics (BAS), Bulgaria

***Correspondence:**
Václav Mahelka
vaclav.mahelka@ibot.cas.cz

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *26 February 2021*
Accepted: *22 April 2021*
Published: *17 May 2021*

Citation:
Krak K, Caklová P, Kopecký D, Blattner FR and Mahelka V (2021) Horizontally Acquired nrDNAs Persist in Low Amounts in Host Hordeum Genomes and Evolve Independently of Native nrDNA. Front. Plant Sci. 12:672879. doi: 10.3389/fpls.2021.672879

INTRODUCTION

Ribosomal RNA (rRNA) is an essential structural component of ribosomes, the sites of protein synthesis. Ribosomes consist of two subunits, each of which is composed of several proteins and rRNA molecules. In eukaryotes, the large subunit consists of three rRNA molecules (25–26S, 5.8S and 5S), whereas the small subunit includes just one rRNA (18S) molecule (Ben-Shem et al., 2011). High demand for ribosomal RNA needed for ribosome assembly is satisfied through the transcription of numerous copies of the nuclear ribosomal (nrDNA) genes. Three of the four rRNA genes in eukaryotes, coding for 18S, 5.8S, and 26S (altogether referred to as 35S), are separated by two internal transcribed spacers (ITS) and together constitute a single transcription unit (Potapova and Gerton, 2019). Transcription units are separated from one another by intergenic spacers (IGS). Plant genomes harbour thousands of nrDNA transcription units, which are organised in large tandem arrays forming the so-called nucleolar organiser regions (NORs) located on a variable number of chromosomes (Srivastava and Schlessinger, 1991; Dubcovsky and Dvořák, 1995).

The number of nrDNA copies has been found to correlate with genome size at large scale in both plants and animals (Prokopowich et al., 2003). Nevertheless, at lower taxonomic levels this relationship is largely unexplored. The extent of the nrDNA copy number variation (CNV) depends on the group studied. It is somewhat conserved among liverworts, mosses and hornworts (Rosato et al., 2016). Conservation is highly variable among species (Cronn et al., 1996) as well as within species of seed plants (Govindaraju and Cullis, 1992; Malinská et al., 2010), fungi (Johnson et al., 2015) and vertebrates (Veiko et al., 2007). Despite the fact that nrDNA occupies a significant portion of the eukaryotic genome, the relationship between species-level phylogeny and nrDNA copy number has barely been investigated (Sproul et al., 2020).

Nuclear ribosomal DNA is an exemplar member of a multigene family and known for its ability to maintain sequence homogeneity. The nrDNA units change their sequences in a highly synchronised manner – described as concerted evolution (Arnheim et al., 1980). Notwithstanding, nrDNAs of different origin have been found to coexist within a single genome as a result of hybridisation and allopolyploidisation events (Álvarez and Wendel, 2003). The patterns of sequence evolution and array contraction and/or expansion of the putative parental copies, co-existing within the hybrid genomes, are well explored (Matyášek et al., 2007; Malinská et al., 2010). The fates of the parental ribotypes range from complete homogenisation (Wendel et al., 1995; Fuertes Aguilar et al., 1999) over independent evolution of parental sequences and their maintenance at different abundances (e.g., Malinská et al., 2010) to the occurrence of newly arising recombinant ribotypes (e.g., Ko and Jung, 2002). Horizontal gene transfer (HGT) represents another mechanism that can contribute to increased diversity of nrDNA within a single genome. However, the evolutionary dynamics of nrDNA sequences acquired via HGT has not been investigated.

Recently, Mahelka et al. (2017) described extensive HGT involving nrDNA. They found that wild barley (*Hordeum*, Pooideae) species possess, in addition to their native nrDNA copies, nrDNA sequences that correspond to grasses from the subfamily Panicoideae. The transferred nrDNAs occur only in the I-genome *Hordeum* taxa (= sect. *Stenostachys*; for infrageneric classification of *Hordeum* see Blattner, 2009; Brassac and Blattner, 2015), and certain wild barley species (and individuals) possess non-native nrDNAs from up to five panicoid donors (namely *Arundinela*, *Euclasta*, *Panicum*, *Paspalum*, and *Setaria*). Phylogenetic patterns suggest the acquisition of the panicoid DNA occurred via at least nine independent horizontal transfers within a timeframe of between 5 and 1 mya. Based on substitution patterns within the ITS region and the absence of mRNA expression, the authors considered the foreign ribotypes as pseudogenes (Mahelka et al., 2017). In a follow-up study, Mahelka et al. (2021) focused specifically on the transfer from a *Panicum*-like donor into *Hordeum* that is likely the oldest of the nine transfers between panicoid grasses and *Hordeum*. It predated the diversification of sect. *Stenostachys*, and occurred between 5 and 1.7 mya. Along with several protein-coding genes and transposable elements, these authors show that the *Panicum*-like ribotypes resided within a specific *Panicum*-derived chromosomal segment, which is located on a NOR-bearing chromosome, although on the opposite chromosome arm in various *Hordeum* species from sect. *Stenostachys*.

Mahelka et al. (2017) carried out a detailed characterisation of the non-native genetic material at the level of sequence variation in a phylogenetic context, but did not focus on quantification of the foreign nrDNAs in the *Hordeum* genomes. The number of particular foreign ribotypes in the *Hordeum* genomes – in terms of copy number, coupled with within – as well as between-species dynamics – remains an unanswered questions.

The objective of our study is to investigate the CNV of nrDNA in diploid representatives of the genus *Hordeum*. In the main, we ask what is the CNV of the foreign ribotypes in the I-genome species? Also, whether, and how, the CNV correlates with the CNV of native ribotypes in *Hordeum*? To answer these questions, we investigated the following: (i) the CNV of the native ribotype in diploid species representing the entire genus *Hordeum*; (ii) the CNV of particular foreign ribotypes in species of section *Stenostachys*; (iii) any correlations in the patterns of CNV between the native and foreign ribotypes; (iv) any correlation between the CNV of both native and foreign ribotypes and species-level phylogeny, genome size and the number of nrDNA loci. We hypothesise that a minor proportion of foreign ribotypes in the total nrDNA, coupled with independent evolution of foreign and native ribotypes, suggest the persistence of foreign nrDNAs within minor nrDNA loci aside from the main and active nrDNA loci (NORs). Together with other lines of evidence, such as non-functionality of the foreign nrDNAs and their confirmed occurrences in chromosomal parts located outside the NORs (Mahelka et al., 2017, 2021), the above scenario would constitute additional support for the status of foreign nrDNAs in *Hordeum* as dead-end by-products of HGT events.

MATERIALS AND METHODS

Plant Material and DNA Quality Check

The plant material and high-quality DNA used in this study was available from a previous study (Mahelka et al., 2017). Details of the plant material is provided in **Table 1** and **Supplementary Table 1**, for further details on the origins of particular accessions see Mahelka et al. (2017). The DNA of all samples was checked for integrity on agarose gel, and quantified using Qubit fluorometer (Thermo Fisher) with the Quibit dsDNA HS Assay kit (Thermo Fisher) according to the manufacturer's instructions.

Preparation of Standards for qPCR Estimation

Estimation of copy number using qPCR, requires the parallel measurement of standards to control qPCR efficiency. We employed results from our previous study (Mahelka et al., 2017), in which bacterial colonies containing all of the different ribotypes (both native and foreign) were stored as deep-frozen glycerol stocks. From this source we retrieved the relevant representative samples, and from these we isolated the plasmids using the Qiagen Plasmid Mini Kit (Qiagen) according to manufacturer's instructions. The plasmids were linearised using an appropriate restriction enzyme with a single recognition site in the plasmid region. The linearised plasmids were purified using the Qiaquick PCR Purification Kit (Qiagen) and their concentrations were measured with a Qubit fluorometer using the Quibit dsDNA HS Assay kit (Thermo Fisher). The plasmids were then diluted to a concentration of 1 ng/μl. Copy numbers of the target ribotypes were calculated for the amplified fragments using the following equation: Copy number (ng^{-1}) DNA = $(6.022 \times 10^{23})/(L \times 10^{-9} \times 660)$, where L is the length in bp of the amplified fragment (e.g., the insert), as implemented in the online tool[1]. Lengths of the inserts were as follows: native (*Hordeum*-like) 644 bp, *Panicum*-like ribotype 538 bp, *Paspalum*-like ribotype 537 bp, *Setaria*-like ribotype 539 bp, and *Euclasta*-like ribotype 536 bp. The plasmids were serially diluted to obtain a range of copies of the target ribotypes between 10^6 and 1 (in a 10-fold dilution series) and these dilution series were used for the qPCR efficiency estimates.

Development and Testing of qPCR Assays

The qPCR assays, targeting the ITS1-5.8S-ITS2 region of nrDNA, were developed based on the sequence variation available from Mahelka et al. (2017). We targeted the primers to the regions showing the highest specificity for each ribotype. At first, the specificity of the primers was tested *in silico* using Geneious 10.2.6 (Biomatters Ltd.), later, specificity was tested using the ribotype-specific plasmid standards. Each standard was amplified by all qPCR assays. Assays showing cross-amplification were further redesigned and tested, until only specific amplification of the target ribotype was obtained. Further, the reaction conditions of each assay were optimised using the serial dilutions of the respective plasmid standards in order to obtain maximal amplification efficiency.

[1]http://cels.uri.edu/gsc/cndna.html

Limitation of the qPCR Assays

Our initial intention was to develop seven qPCR assays: five targeting the foreign ribotypes as found in *Hordeum,* e.g., *Panicum*-, *Paspalum*-, *Euclasta*-, *Setaria*-, and *Arundinella*-like (Mahelka et al., 2017), one assay specific for the native *Hordeum* ribotype and a universal assay that would amplify the native as well as the foreign ribotypes (total nrDNA). Unfortunately, the overall pattern of sequence variation hindered the development of an assay specific for *Arundinella* and the assay targeting the native *Hordeum* ribotype. Hence, we use the universal assay not only to estimate the copy number of total nrDNAs but also to estimate the copy number of the native ribotype. Therefore, in species of sect. *Stenostachys*, an overestimate of native ribotype is likely. However, given the low proportion of foreign ribotypes among the total nrDNAs, the bias is considered unlikely to be serious. In the species outside this section (*H. vulgare, H. marinum, H. murinum, H. gussoneanum*) that lack the foreign ribotypes, the estimate is likely to be unbiased.

Estimation of nrDNA Copy Number Using qPCR

All qPCRs were carried out using the LightCycler 480 II Real-Time PCR instrument (Roche). To quantify the native ribotype, a TaqMan probe-based assay with the Light Cycler 480 Probes Master kit (Roche) was used, under the following cycling conditions: 10 min at 95°C, followed by 45 cycles of denaturation (95°C, 10 s), annealing (60°C, 30 s), and extension (72°C, 1 s).

The foreign ribotypes were quantified using SYBR Green I-based assays and the LightCycler 480 SYBR green I master kit (Roche), under the following cycling conditions: 5 min at 95°C, followed by 40 cycles of denaturation (95°C, 10 s), annealing (assay-specific Ta, 10 s; for specific conditions see **Supplementary Table 2**) and extension (72°C, 15 s). To detect the possible formation of primer dimers or unspecific amplification, the cycling was concluded with a standard melting curve analysis.

The efficiencies of the qPCR assays were estimated from standard calibration curves based on serial 10-fold dilutions of plasmid standards with specific ribotype sequence inserts, ranging from 10^6 to 1 copies of target nrDNA. The absolute quantification of the target sequences was carried out based on the standard calibration curves using the LightCycler 480 software, version 1.5 (Roche). The resulting concentrations of amplicon DNA, expressed as DNA copy number (ng^{-1}), were further normalised to the genome sizes of the *Hordeum* species analysed (Jakob et al., 2004). For each quantification standard, three technical replicates were used, whereas for each analysed individual two technical replicates were included in the qPCR reactions.

Estimation of nrDNA Copy Number Using Illumina Data

In addition to the qPCR-based estimation of CNV, we estimated CNV of native ribotypes from whole-genome sequencing data.

TABLE 1 | Copy numbers of native and foreign ribotypes in *Hordeum* as estimated using qPCR.

Taxon	Geography	Clade (Subclade)	Genome size (pg/2C)	No. of loci (pairs)	Mean ribotype copy number (s.d.) per 2C				
					Native	*Panicum*	*Paspalum*	*Setaria*	*Euclasta*
H. bogdanii	Central Asia	I (Eurasian)	9.48	1*	9320 (1989)	17 (5)	n.d.	n.d.	n.d.
H. californicum	North America	I (American, californicum)	8.19	1*	8478 (2516)	12 (8)	n.d.	n.d.	19 (8)
H. chilense	Central Argentina	I (American, "core" species)	8.77	2	21665	36	98	n.d.	n.d.
H. comosum	Patagonia	I (American, "core" species)	8.97	n.d.	3929	10	12	4	60
H. cordobense	Central Argentina	I (American, muticum/cordobense)	9.19	2*	4122	14	9	n.d.	14
H. erectifolium	Central Argentina	I (American, "core" species)	9.49	2*	22282	n.d.	45	4	n.d.
H. euclaston	Central Argentina	I (American, "core" species)	6.85	2*	6416 (1798)	n.d.	134 (177)	n.d.	n.d.
H. flexuosum	Central Argentina	I (American, "core" species)	8.51	2*	13980	16	59	n.d.	n.d.
H. intercedens	North America	I (American, "core" species)	7.01	2*	12779 (10846)	n.d.	9 (9)	n.d.	n.d.
H. muticum	Central Argentina	I (American, muticum/cordobense)	9.57	1*	3498 (1933)	15 (7)	n.d.	n.d.	28 (31)
H. patagonicum subsp. *magellanicum*	Patagonia	I (American, "core" species)	9.33	2	13212	25	39	n.d.	73
H. patagonicum subsp. *mustersii*	Patagonia	I (American, "core" species)	8.77	2	1689	15	18	n.d.	n.d.
H. patagonicum subsp. *patagonicum*	Patagonia	I (American, "core" species)	9.46	2	2337 (291)	36 (26)	59 (27)	n.d.	132
H. patagonicum subsp. *setifolium*	Patagonia	I (American, "core" species)	9.61	2*	10996 (6986)	23 (3)	10 (8)	13	134 (20)
H. pubiflorum	Patagonia	I (American, "core" species)	8.70	2*	3830 (2727)	23 (8)	6 (3)	n.d.	26 (27)
H. pusillum	North America	I (American, "core" species)	7.16	2*	12488 (2802)	n.d.	401	n.d.	n.d.
H. roshevitzii	Central Asia	I (Eurasian)	9.69	1*	11138 (880)	18 (9)	n.d.	n.d.	n.d.
H. stenostachys	Central Argentina	I (American, "core" species)	9.38	2*	7257 (4683)	26	50 (38)	16 (7)	474
H. gussoneanum	Western Eurasia	Xa	10.41	1	5031 (3101)	n.d.	n.d.	n.d.	n.d.
H. marinum	Western Eurasia	Xa	9.10	1*	14026 (8769)	n.d.	n.d.	n.d.	n.d.
H. murinum subsp. *glaucum*	Western Eurasia	Xu	9.11	2*	31342 (13353)	n.d.	n.d.	n.d.	n.d.
H. vulgare subsp. *spontaneum*	Western Eurasia	H	10.59	2	24415	n.d.	n.d.	n.d.	n.d.
H. vulgare subsp. vulgare	Western Eurasia	H	10.59	2	26556	n.d.	n.d.	n.d.	n.d.

*Standard deviation (brackets) is given for taxa with two accession analysed. Only diploid taxa were analysed. Geography – geographic distribution; Clade – phylogenetic position within Hordeum; the I-genome species are further divided into monophyletic subclades based on Brassac and Blattner (2015); Genome size values are from Jakob et al. (2004); * numbers of loci identified in this study (for details see* ***Supplementary Table 1*** *and* ***Supplementary Figures 3, 4****); n.d. = ribotype not detected from the sample.*

This analysis provides an alternative method to the qPCR-based estimation, enabling estimation of relative sensitivity of both methods. Therefore, only a subset of *Hordeum* samples was included in this analysis (specified below). In principal, the CNV was calculated from the number of Illumina reads mapped to a reference out of the total number of reads used for the mapping. The CN estimation was calculated as described in Wang et al. (2019), with two modifications. First, to avoid potential bias caused by uneven coverage within the rDNA region (genes vs. spacers), we used a complete repeat unit (18S-ITS1-5.8S-ITS2-26S) as a reference. Second, we used 2C DNA content as input data in the formula to conform with the qPCR-based CN estimation. A unique reference was used for each species: Each reference consisted of a species-specific ITS region (retrieved from Mahelka et al., 2017), which was surrounded by universal 18S and 26S rDNA genes. The genes were common to all species and have been derived from *Hordeum bogdanii*. The transcription unit of *H. bogdanii* was completed by mapping genomic reads of *H. bogdanii* against a *Secale cereale* reference (JF489233). The genes' and spacers' boundaries were then adjusted with the aid of an alignment consisting of multiple rDNA units of grasses including the newly assembled *H. bogdanii*. Lengths of the reference sequences varied slightly among the *Hordeum* samples (see **Table 2**). Mapping of reads was done using Bowtie2 (Langmead and Salzberg, 2012) implemented in Geneious version R10 (Biomatters Ltd.), using local alignment type and lowest sensitivity.

We used for the analysis samples representing all major *Hordeum* lineages (genomes H, Xu, Xa, and I; Blattner, 2009): *H. vulgare* subsp. *vulgare* (cultivars Morex, BW457, AAC Synergy, Igri; genome H), *H. vulgare* subsp. *spontaneum* (H), *H. murinum* subsp. *glaucum* (Xu), *H. marinum* (Xa), *H. gussoneanum* (Xa), *H. bogdanii* (I), and *H. pubiflorum* (I). For *H. murinum* subsp. *glaucum*, *H. marinum*, *H. gussoneanum*, *H. bogdanii*, and *H. pubiflorum*, we used the same accessions (but not individuals), as we used for the qPCR-based estimation of CN (compare **Table 2** and **Supplementary Table 1**). Short read archives were either downloaded from NCBI database, or we used our unpublished data (**Table 2**). For each species, the reads were subsampled to ca 1 × genomic coverage. Details on the accessions analysed and short read archives used for the analysis are provided in **Table 2**.

Identification of nrDNA Loci Number

The number of nrDNA loci was determined by fluorescent *in situ* hybridisation on metaphase chromosomes using the pTa71 probe as described in Mahelka and Kopecký (2010). The experiments were done under conditions of ∼77% stringency.

Data Analysis

We carried out a number of basic exploratory statistics to evaluate the variation of estimated nrDNA copy numbers. To determine the effects of variables, we used linear regression, ANOVA, Tukey's HSD test and two-way ANOVA in R version 3.6.3 (R Core Team, 2020).

TABLE 2 | Copy numbers of native ribotype in *Hordeum* as estimated using Illumina data.

***Hordeum* sample**	**BioProject**	**Experiment accession**	**Run**	**Genome size (Mb/2C)**	**Number of reads (subsample)**	**Reads mapped**	**Genome proportion (%)[1]**	**Genome space (Mb)[2]**	**Genome space (Kb)**	**Copy number[3,4] (rounded)**
H. vulgare cv. Morex	PRJEB31444[a]	ERX3211415	ERR3183564	10357.02	35000000	178318	0.51	52.781	52781	9100
H. vulgare cv. BW457	PRJEB3038	ERX103249	ERR127100	10357.02	34190711	251901	0.74	76.306	76306	13150
H. vulgare cv. AAC Synergy	PRJNA665698[b]	SRX9220769	SRR12748507	10357.02	35000000	124428	0.36	36.820	36820	6350
H. vulgare cv. Igri	PRJEB36576[c]	ERX4041629	ERR4040343	10357.02	35000000	116486	0.33	34.470	34470	5940
H. vulgare subsp. *spontaneum*	PRJEB25923[d]	ERX2779022	ERR2766181	10435.26	35000000	206923	0.59	61.694	61694	10640
H. murinum subsp. *glaucum* BCC2017	PRJNA491526[e]	SRX4789710	SRR7956029	8909.58	31000000	158437	0.51	45.536	45536	7850
H. marinum BCC2001	PRJNA720259[f]		SRR14162036	8899.8	48000000	112020	0.23	20.770	20770	3580
H. gussoneanum BCC2012	PRJNA720259[f]		SRR14162035	10180.98	54086101	71061	0.13	13.376	13376	2310
H. bogdanii BCC2063	PRJNA720259[f]		SRR14162034	9271.44	47328373	73152	0.15	14.330	14330	2470
H. pubiflorum BCC2028	PRJEB18129	ERX246085	ERR271809	8508.60	44000000	89665	0.20	17.339	17339	2990

Superscript letters in BioProjects refer to studies in which the data were generated (if available). Genome size values are from Jakob et al. (2004).

[1]Calculated as [(number of mapped reads/total number of reads) × 100].

[2]Calculated as [(genome proportion × genome size)/100].

[3]Calculated as [(genome space in kb/length of the reference rDNA unit in kb].

[4]Reference rDNA unit lengths: H. vulgare (incl. ssp. spontaneum), H. bogdanii 5.801 kbp, H. marinum, murinum, gussoneanum 5.803 kbp, H. pubiflorum 5.799 kbp.

[a]Monat et al., 2019, [b]Xu et al., 2021, [c]Jayakodi et al., 2020, [d]Liu et al., 2019, [e]Kono et al., 2019, [f]this study, [g]Mascher et al., 2013.

RESULTS

We carried out a qPCR-based estimation of nrDNA copy number in the diploid representatives of the genus *Hordeum*. The main objective was to investigate the dynamics and evolution of the foreign ribotypes, in particular by estimating the CNV of the foreign ribotypes and comparing these with the CNV of native *Hordeum* ribotypes. For a subset of taxa, representing all major *Hordeum* clades (genomes), we performed an alternative estimation of CNV of native ribotype using low coverage Illumina sequencing.

Copy Number Estimates Based on qPCR and Illumina Read Mapping Are Correlated

The copy numbers of the native ribotype estimated using qPCR and Illumina read mapping differed in all analysed taxa. In all samples, mean qPCR-based values were higher than the values obtained from read mapping. The biggest difference was observed for *H. murinum* (4×), whereas in *H. pubiflorum* the difference was much smaller (1.3×, **Supplementary Table 3**). Overall, the copy numbers estimated using the two methods were correlated (adjusted $R^2 = 0.743$, $p = 0.0078$; **Supplementary Figure 1**). Since we focus primarily on relative copy number differences, and a detailed sample including all diploid *Hordeum* species was used only to estimate CN based on qPCR, we consider qPCR as the default method. Therefore, CN estimates are hereafter qPCR-based, unless otherwise stated.

Copy Number of Native Ribotype Varies Between Species and Individuals

For the native *Hordeum* ribotype, we found an up to 19-fold difference in the mean copy number among the taxa analysed. The lowest mean value, 1689 copies per 2C content, was estimated for *H. patagonicum* subsp. *mustersii* and the highest, 31,342 copies, for *H. murinum* subsp. *glaucum* (**Table 1**). A high variation in copy number (up to fourfold) was also observed between different accessions of the same taxa (**Supplementary Table 1**). To determine whether the observed variation in nrDNA copy number was affected by variations in the DNA extraction efficiency (and therefore represents a methodological artefact), we carried out a regression analysis. We identified no correlation between DNA extract concentration (as a measure of DNA extraction efficiency) and nrDNA copy number (adjusted $R^2 = -0.025$, $p = 0.71$; **Supplementary Figure 2**). This indicates that there is no significant effect of DNA extraction efficiency on the measured CN value. For *H. vulgare* subsp. *vulgare*, intraspecific CNV could be estimated also from Illumina read mapping. Here, the four accessions (different cultivars) showed up to a 2.21-fold variation in the copy number (**Table 2**).

Phylogeny, Genome Size, and Number of nrDNA Loci Have No Effect on CNV of the Native Ribotype

We assessed the effect of genome size, number of nrDNA loci and phylogeny on the observed patterns of CNV. There were either one or two pairs of nrDNA loci in the *Hordeum* samples analysed (**Supplementary Table 1** and **Supplementary Figures 3**, **4**). We found no significant relationship between the copy number of the native ribotype and genome size or the number of nrDNA loci (**Supplementary Figure 5**). Conversely, the nrDNA copy number was significantly affected by a group's phylogeny (ANOVA, $F = 3.396$, $p < 0.05$). However, the *post hoc* comparisons (Tukey's HSD test) revealed that the only significant difference was that between the Xu-genome and the American I-genome species (**Figure 1**).

TABLE 3 | Copy number variation of native *Hordeum* ribotype within the major phylogenetic lineages.

Lineage	Mean copy number (s.d.)	Copy number variation
H	25485 (1513)	1.1
I – American	8682 (6392)	13.2
I – Eurasian	10229 (1636)	1.5
Xa	9528 (7470)	7.1
Xu	31342 (13353)	1.9

Phylogenetic lineages follow the phylogeny of Brassac and Blattner (2015); standard deviation (brackets); copy number variation – ratio of maximal and minimal values estimated within each lineage.

The CNV was high within the recognised phylogenetic groups (**Table 3**). The highest variation (in the ratio of maximal and minimal values, estimated within a lineage) was observed among the American I-genome species (13.2-fold). On the other hand, the H genome, represented by *H. vulgare*, was the most homogeneous group with respect to nrDNA copy number. Within this group, the two subspecies differed only by a factor of 1.1 (**Table 3**).

Foreign Ribotypes Represent a Minor Proportion of Total nrDNA

Apart from the native nrDNA, the I-genome *Hordeum* species contain foreign ribotypes acquired from panicoid grasses. The sequence divergence of particular ribotypes allowed us to design specific qPCR assays targeting four out of the five foreign ribotypes (namely *Panicum*-, *Setaria*-, *Paspalum*-, and *Euclasta*-like) and hence to estimate their CNV. The estimated copy number of the foreign ribotypes was remarkably low compared with that of the native ribotypes. The estimates ranged from a few copies to tens, solely hundreds of copies (**Table 1**). Individual foreign ribotypes represent only a small fraction of the estimated total number of nrDNA copies (0.01 – 12.02% depending on ribotype; **Supplementary Table 1**). Overall, the *Euclasta*-like ribotype was the most abundant and the *Setaria*-like ribotype was the least abundant (**Figure 2**). The two remaining ribotypes showed intermediate values. The differences among the abundances of the foreign ribotypes were not significant (ANOVA, $F = 2.572$, $p = 0.06$).

We carried out a regression analysis to test whether the copy numbers of the foreign ribotypes were affected by the copy number of the native ribotype. We found the estimated abundances of foreign ribotypes were independent of the abundance of the native ribotype (**Supplementary Figure 6**).

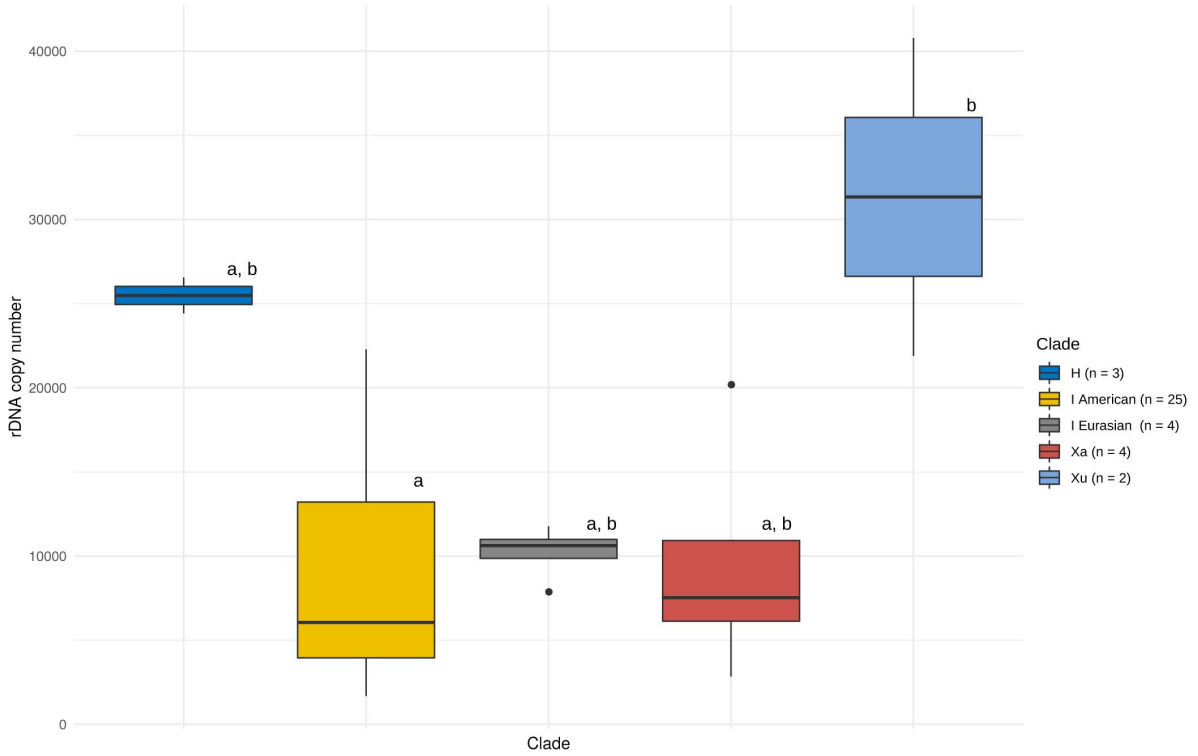


FIGURE 1 | Copy number variation of the native ribotype in *Hordeum*. Variation within and between major phylogenetic groups is shown. The boxes represent the range between 25th and 75th percentil. The whiskers represent the distance between hinges to the 1.5 × IQR (IQR is the inter-quartile range, or distance between the first and third quartiles) to both sides (toward the maximum as well as minimum values). Vertical bars represent the median values and outlayers are dispalyed as individual points. Phylogenetic grouping is based on Brassac and Blattner (2015). Differences among means (solid lines in boxes) at $P = 0.05$ (as analysed by ANOVA) are indicated by letters above boxes (boxes with the same letters are not significantly different).

This indicates the high specificity of the qPCR assays (indicating the absence of significant cross-amplification of the non-target template by the qPCR assays). It points to the independent evolution of the foreign and native ribotypes within the genomes studied.

Phylogeny and Geography Have No Effect on the Copy Numbers of Foreign Ribotypes

Next, we asked whether the abundances of the foreign ribotypes correlated with the phylogenetic relationships of the I-genome *Hordeum* species. For this purpose, we subdivided the I-genome group into subclades, which reflected the group's phylogeny (**Table 1**). We found no significant relationship between the abundance of the foreign ribotypes and the phylogeny (ANOVA, $F = 0.758$, $p = 0.52$ for the differences between the subclades, $F = 0.968$, $p = 0.43$ for the interaction of the subclade and ribotype, **Figure 3**). We further investigated whether the CNV of the foreign ribotypes was affected by the geographic origins of the species analysed. Again, the copy number of the foreign ribotypes did not differ significantly among samples of different geographic origin (ANOVA, $F = 1.077$, $p = 0.37$) and the interactions between the foreign ribotype and geographic origin of the sample were not significant either ($F = 0.551$, $p = 0.74$).

DISCUSSION

Apart from their native nrDNA copies, species of section *Stenostachys* from the barley genus *Hordeum* harbour foreign nrDNA copies, presumably acquired after a series of HGT events from panicoid grasses (Mahelka et al., 2017). To investigate the evolution and dynamics of the foreign ribotypes, we estimated their CNV using specific qPCR assays, and related the values to the CNV of native *Hordeum* ribotypes. We hypothesised that the HGTs resulted in insertions of foreign nrDNA arrays at random places across *Hordeum* genomes, so that the foreign ribotypes represent non-functional entities, which reside apart from the NORs in the *Hordeum* species. Such a pattern would imply independent evolution of native and foreign ribotypes, which can be tested using relationships between CNV of foreign and native ribotypes.

The Usefulness of qPCR and Mapping of Short Reads for Quantification of nrDNA

To investigate the differences in CNV of individual ribotypes, we targeted our qPCR assays to ITS of nrDNA. These regions showed sufficient variation to discriminate between the foreign ribotypes, as well as between the foreign ribotypes and the native one. One of the assays was used to estimate the CNV

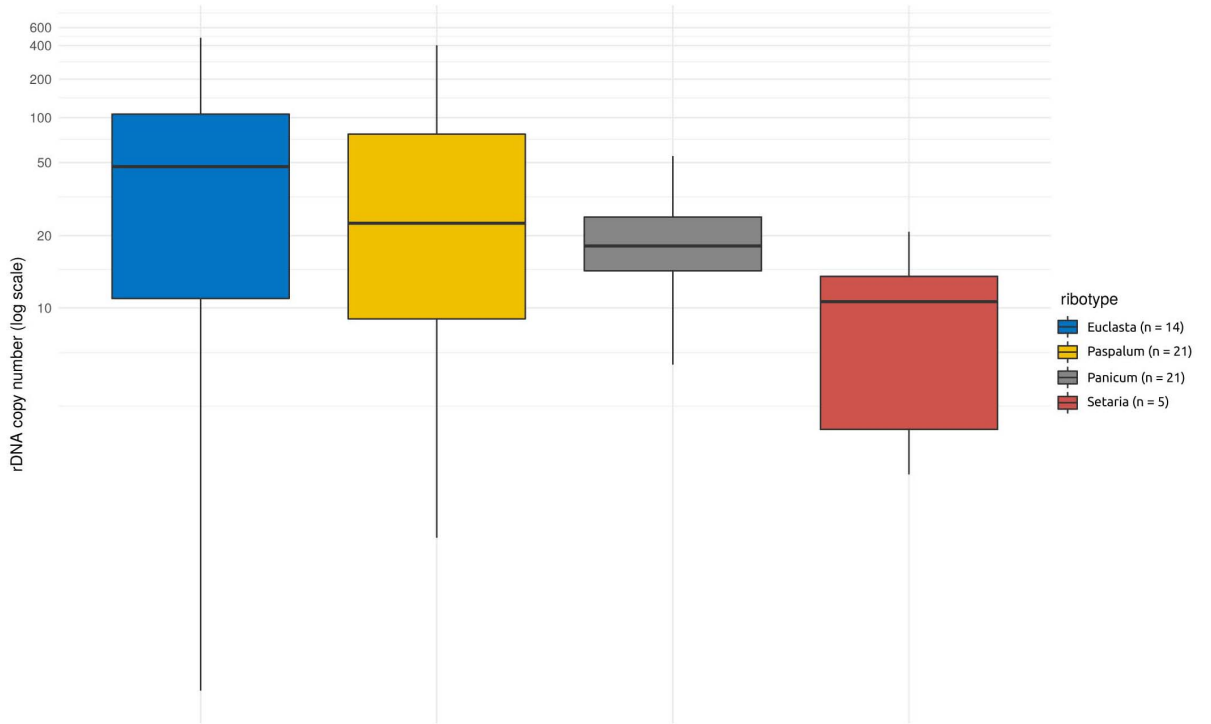


FIGURE 2 | Copy number variation within and among the foreign ribotypes in *Hordeum* species of sect. *Stenostachys*. No significant differences at $P = 0.05$ were detected using ANOVA.

of the native (*Hordeum*) ribotype across the genus *Hordeum*. We used this method mainly to seek for relative differences in CNV of particular ribotypes in the different species, or in the major phylognetic groups. For precise estimates of absolute copy numbers, one should target the rRNA genes (Long et al., 2013; Wang et al., 2019), ideally coupled with a normalisation using a well-defined and confirmed single-copy gene(s) (Long et al., 2013; Rosato et al., 2016; Rabanal et al., 2017). Conversely, specific qPCR assays targeting the ITS region, capture all target sequences irrespective of their genomic locations, and thus represent an efficient tool for detection of potentially pseudogenised, interspersed, non-native nrDNA copies. Thus, this approach is suitable for detection and quantification of minority acquired copies, for example via HGT.

To get insight into the sensitivity of the qPCR-based CN estimation, we implemented an alternative method based on Illumina read mapping. We thus obtained an independent estimate of CNV for a subset of taxa, representing all the major phylogenetic groups included in the study. We found considerable differences between the two methods, with qPCR-based CN estimates showing 1.3–4 × higher values (**Supplementary Table 3**). Despite the differences between the absolute values, the overall CN estimates were correlated. Hence, we believe that relative differences (between taxa and ribotypes), as inferred from qPCR here, are valid.

The consistently higher CN values estimated using qPCR, than those estimated by read mapping, may partly stem from different methodology of both approaches. While the qPCR targeted ITS region and therefore measured CN thereof, using Illumina data we estimated the number of complete rDNA units. It is likely that the ITS region outnumbers complete rDNA repeat units in plant genomes, because additive ITS copies are scattered across genomic locations outside the main NORs. We found that there are tens of ITS1 copies dispersed across all chromosomes except chr3 within the Morex V2 assembly of *H. vulgare* (Monat et al., 2019) (**Supplementary File 1**). In *Hordeum* samples analysed by us, a visual inspection of mapped reads suggests that there is no obvious increase in coverage within the ITS region. On the other hand, there is a decrease of reads, which mapped to the 26S gene in *H. bogdanii*, *H. marinum*, and *H. vulgare* cultivars BW457 and AAC Synergy. Thus, Illumina-based CNs can be underestimated in these cases. In any case, uneven coverage of a region of interest by reads is a potential source of error in CN estimation based on read mapping.

Potential discrepancies of relative values (e.g., the opposite ratios of CNs obtained for *H. bogdanii* accession BCC2063 and *H. pubiflorum* BCC2028; **Table 2** and **Supplementary Table 1**), can be attributed to intraspecific variation of CN, when differences among individuals are commonplace (discussed below). In this particular case (*H. bogdanii* vs. *H. pubiflorum*), identical accessions were used, but not the same individuals. Even the use of plant material from a seed bank is not a guarantee of genotype homogeneity. For example, Jakob et al. (2014) studied genetic diversity in wild barley *Hordeum vulgare* subsp. *spontaneum*. Interestingly, *ex situ* samples of this taxon showed genetic admixture, as 22% of gene bank samples did

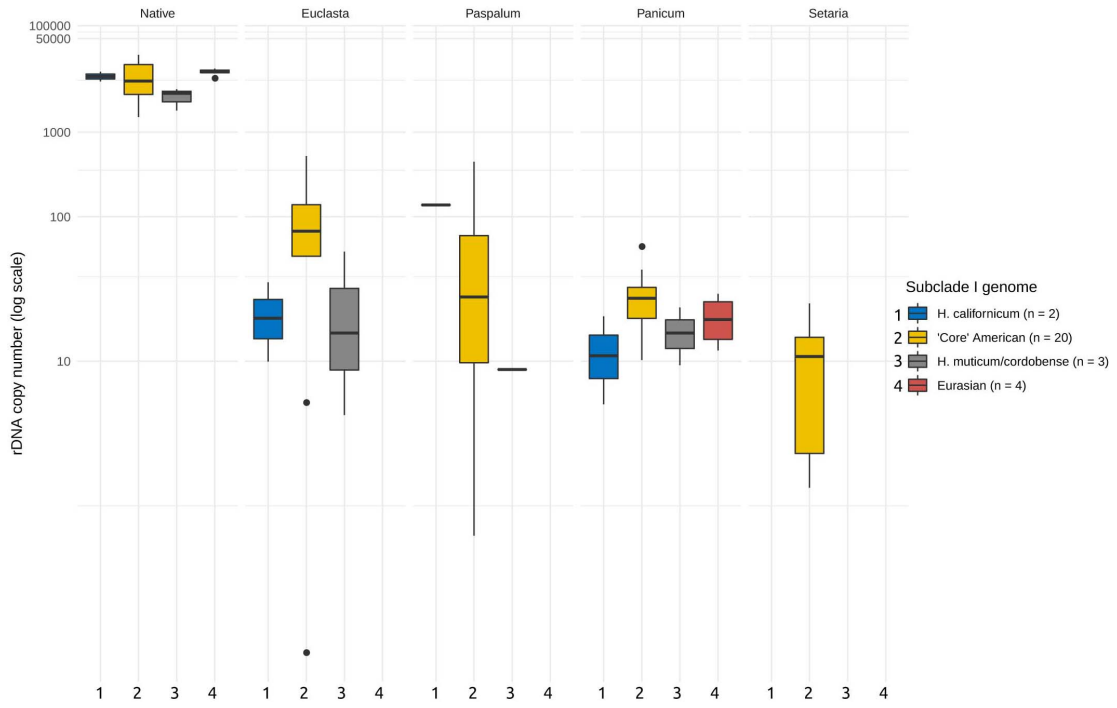


FIGURE 3 | Copy number variation of the foreign ribotypes in relation to major phylogenetic groups. Phylogenetic grouping is based on Brassac and Blattner (2015). Foreign ribotypes only occur in *Hordeum* species of sect. *Stenostachys*. No significant differences at $P = 0.05$ were revealed by ANOVA. Copy numbers of native ribotype are given for comparison.

not correspond with the geographic pattern found in wild barley accessions collected in the field (in which the incongruence was only found in 5% of samples). The authors speculated, that although *H. vulgare* is classified as an inbreeder, the admixture was likely caused by hybridisation during propagation and maintenance, and, possibly, handling of the *ex situ* gene bank samples. Thus, different individuals may actually represent diverse genotypes even if they are maintained under the same accession number. In this respect, revealing the net effect of particular methods investigated in the same individual plants (or even better in the same DNA extract), would be necessary.

CNV Among Ribotypes

One result of the multiple HGTs from panicoid grasses into *Hordeum* is that the I-genome species (and even some individual plants) harbour up to five foreign ribotypes. Moreover, these correspond to distinct panicoid genera, namely *Panicum*, *Paspalum*, *Setaria*, *Euclasta*, and *Arundinella* (Mahelka et al., 2017). In most cases, multiple foreign ribotypes were found in single individual plants. One individual of *H. stenostachys* even combined foreign ribotypes from all of the five panicoid genera. Although the distribution of foreign ribotypes among the I-genome species has been well described, information on their quantity, dynamics (variation among individuals) and evolution (variation among species) is largely lacking.

The CNV of individual foreign ribotypes provides valuable information, as it may indicate potential functionality and thus evolutionary potential in *Hordeum* genomes through their quantity and inferred chromosomal localisation. In another study, Mahelka et al. (2021) provide detailed characterisation of the *Panicum*-derived chromosomal segment in *Hordeum* species. The authors reconstructed the foreign DNA segment using sequencing of BAC clones in *H. bogdanii* (specimen BCC2063) and *H. pubiflorum* (BCC2028). There were two *Panicum*-like copies of the ITS region identified within the segment, and the copies occurred at two distinct sites. Thus, given the confirmed presence of a *Panicum*-derived segment on both chromosomal homologs, there are at least four *Panicum*-like copies in *H. bogdanii* and *H. pubiflorum*. The copy number of *Panicum*-like ribotypes as inferred from qPCR shows higher values (21 copies in *H. bogdanii* BCC2023 and 29 copies in *H. pubiflorum* BCC2028). This higher count can be explained by the presence of additional copies beyond the boundaries of the *Panicum*-derived segment, which would not have been detected by Mahelka et al. (2021). Furthermore, we still cannot rule out the possibility that some of the foreign ribotypes reside within the native nrDNA loci. The coexistence of different nrDNA variants within a single genome has been documented in allopolyploid and homoploid hybrids (e.g., Kovařík et al., 2008; Fehrer et al., 2009; Zozomová-Lihová et al., 2014) or even in presumed pure diploid species with different paralogous rDNA loci (Blattner, 2004). For these, several scenarios have been described, including maintenance of the different ribotypes within their original loci, formation of new recombinant ribotypes or gradual replacement of one ribotype by another due to interlocus recombination (reviewed in Álvarez and Wendel, 2003). We found only a

small proportion of nrDNA is of foreign origin, supporting their presence solely in minor arrays, co-localised with other genetic material of foreign origin as expected under the HGT scenario (Mahelka et al., 2017, 2021). Due to their localisation, out of reach of homogenisation mechanisms, such minor arrays may remain conserved in plant genomes in relatively stable amounts over millions of years (Mahelka et al., 2021).

CNV of nrDNA Is High Both Within and Between Species, and Is Independent of Phylogeny, Genome Size, and Number of nrDNA Loci in *Hordeum* Species

A high level of CNV of the native ribotype detected between individuals of the same species (e.g., a 4-fold difference in *H. intercedens*) as well as between species within phylogenetic groups (e.g., a 13.2-fold difference within the American I-genome species) is not unusual. In other studies, high variation in nrDNA copies has been reported not only among closely related species and among individuals within a species but even between different tissues of the same individual. For example, Govindaraju and Cullis (1992) detected a 12-fold variation in nrDNA copy numbers among individuals within populations of *Pinus rigida*, whereas variation between populations reached values up to 21-fold. Populations of *Arabidopsis thaliana* from northern and southern Sweden differed in their genome size by more than 10%, mainly due to the variability in nrDNA copy number (Long et al., 2013). In *Vicia faba*, nrDNA copy number was shown to differ between individuals within a population (up to 95-fold differences) and between different tissues of the same individual, showing up to a 12-fold difference (Rogers and Bendich, 1987). Mechanisms underlying such differences still remain to be revealed. In *Hordeum*, the highest variation within phylogenetic groups was observed for the American I-genome species. In this case, the results may have been partly skewed by the number of samples, since the I-genome group is the most species-rich within *Hordeum*. Since we aimed to analyse two individuals per each species, some of the species-poor clades (H, Xu) were represented by only two individuals. Multiple samples from within other than I-genome group were only included for *H. vulgare* (H genome), in which CN of four cultivars was estimated using read mapping (**Table 2**). These four samples showed variation that was twice that of the two samples from the H group estimated using qPCR (*H. vulgare* subsp. *vulgare* and *spontaneum*). Since intra-specific variation of rDNA was described for both (Saghai-Maroof et al., 1984), it is likely that the inclusion of more wild barley individuals could further increase the CNV within the H-genome group.

In *Hordeum*, genome size is correlated with the genus' phylogeny (Jakob et al., 2004). Our data on the CNV of nrDNA suggest that this correlation is not due to quantity of nrDNA. Despite the high level of CNV detected between *Hordeum* species, the only significant difference was that between the Xu- and the American I-genome species. We did not find a significant correlation between the nrDNA copy number and the genome size even when individual plants were examined (**Supplementary Figure 5A**). We further questioned whether the observed CNV could be an effect of the number of nrDNA loci. Since the number of nrDNA loci (particularly of the minor ones) varies in *Hordeum* (Taketa et al., 1999, 2001), using nrDNA-FISH we identified the number of loci in the same accessions, as used for the CNV analysis. The *Hordeum* species involved in this study carried one or two pairs of loci (**Table 1**). This is in accord with other studies, reporting for diploid species one or two pairs of major loci plus up to four pairs of minor loci (Leitch and Heslop-Harrison, 1992; Taketa et al., 1999, 2001). The absence of a significant relationship between CNV and the number of nrDNA loci suggests that the number of loci is not responsible for the variation in nrDNA copy number in *Hordeum*. Instead, within-loci dynamics (causing contraction or expansion of nrDNA arrays) seems to be the dominant mechanism underlying CNV in this group. A striking example fitting this hypothesis is *H. murinum* subsp. *glaucum* analysed here. The hybridisation signal after nrDNA-FISH in accession BCC2017 (**Supplementary Figure 4B**) is particularly stronger than that observed in the other accession of the same species (BCC2002, **Supplementary Figure 4A**). This pattern is consistent with copy number of nrDNA, when accession BCC2017 contains almost twice the number of copies. Although nrDNA-FISH is not a precise method to quantify copy number of rDNA, a correlation between copy number of nrDNA and intensity of hybridisation signal presumably exists.

At inter-specific level, the relationship between nrDNA copy number and phylogeny is still poorly explored. Inter-specific genome size evolution has been addressed for almost two decades, especially for plants (e.g., Albach and Greilhuber, 2004; Jakob et al., 2004; Weiss-Schneeweiss et al., 2006; Chrtek et al., 2009; Kang et al., 2014; Mandák et al., 2016). Besides investigating the patterns of genome size variation in a phylogenetic context, recent studies have focused on identification of particular genomic components as the main drivers of these changes (Zedek et al., 2010; Talla et al., 2017; Blommaert et al., 2019; Hloušková et al., 2019; Wong et al., 2019; McCann et al., 2020; Vitales et al., 2020). To date, nrDNA has been confirmed as being a substantial driver of interspecific genome size evolution only in a group of ground beetles (Sproul et al., 2020). Otherwise, no obvious relationship has been observed to enable generalisation of nrDNA evolution.

CONCLUSION

Here, we reveal the patterns of copy number variation of nrDNA in diploid species of the genus *Hordeum*. For this purpose, we categorised the nrDNA types as native or foreign ribotypes. While the native, *Hordeum*-like, ribotype is present in all *Hordeum* species, the occurrence of foreign ribotypes is restricted to the I-genome *Hordeum* species (*Hordeum* sect. *Stenostachys*), which acquired these ribotypes via a series of horizontal transfers. The foreign ribotypes were present in the respective genomes only at low copy numbers (a few copies to hundreds of copies) and represent a relatively minor fraction of the total nrDNA. We detected a high level of variation in the copy numbers of particular ribotypes at all hierarchical levels examined – between species groups, between species, and between individuals within

a species. This variation did not correspond to any of the genus' phylogeny, the species' genome size, or the number of nrDNA loci. Overall, we can consider nrDNA copy number as a dynamic trait in *Hordeum*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

KK, VM, and PC: conceived and designed the study. PC: conducted lab experiments. DK: provided *in situ* hybridisation experiments. FB: provided the Illumina data. KK and VM: analysed the data. KK: drafted the manuscript with contribution from VM. All co-authors contributed to the final version of the manuscript.

FUNDING

The research was financially supported by the Czech Science Foundation (Grant No. 17-06548S) and the long-term research development project RVO 67985939 of the Czech Academy of Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.672879/full#supplementary-material

Supplementary Figure 1 | Correlation between copy numbers of native ribotype estimated using qPCR and Illumina read mapping.

Supplementary Figure 2 | Correlation between the DNA extraction efficiency (expressed as DNA extract concentration) and the copy number of nrDNA, for both native **(A)** and foreign **(B)** ribotypes.

Supplementary Figure 3 | Molecular cytogenetic analysis of *Hordeum* species. Fluorescent signal of 45S-rDNA (labeled with biotin, arrows) has been observed after FISH on one or two pairs of chromosomes of *H. bogdanii* BCC2063 **(A)**, *H. bulbosum* GRA970 **(B)**, *H. californicum* BCC2057 **(C)**, *H. cordobense* BCC2067 **(D)**, *H. erectofolium* BCC2026 **(E)**, *H. euclaston* BCC2022 **(F)**, *H. flexuosum* BCC2023 **(G)**, *H. intercedens* BCC2044 **(H)**, and *H. marinum* BCC2001 **(I)**. In *H. cordobense* **(D)** and *H. erectifolium* **(E)**, minor loci of 45S-rDNA on one pair of chromosomes have been observed (arrowheads). *Hordeum bulbosum* GRA970 is tetraploid (2*n* = 28), hence it was excluded from qPCR analyses. Scale bar 10 μ m.

Supplementary Figure 4 | Molecular cytogenetic analysis of *Hordeum* species. Fluorescent signal of 45S-rDNA (labeled with biotin, arrows) has been observed after FISH on one or two pairs of chromosomes of *H. murinum* BCC2002 **(A)**, *H. murinum* BCC2017 **(B)**, *H. muticum* BCC2042 **(C)**, *H. patagonicum* subsp. *setifolium* BCC2032 **(D)**, *H. pubiflorum* BCC2028 **(E)**, *H. pusillum* BCC2049 **(F)**, *H. roshevitzii* BCC2015 **(G)**, *H. stenostachys* BCC2041 **(H)**, and *H. vulgare* cv. Morex **(I)**. In *H. murinum* BCC2017 **(B)** and *H. vulgare* cv. Morex **(I)**, minor loci of 45S-rDNA on one pair of chromosomes have been observed (arrowheads). Scale bar 10 μ m.

Supplementary Figure 5 | (A) Correlation between the genome size (2C value) and the copy number of native ribotype, **(B)** correlation between the number of nrDNA loci and the copy number of native ribotype.

Supplementary Figure 6 | Correlations between the copy numbers of native and foreign ribotypes. **(A)** *Setaria*-like ribotype; **(B)** *Paspalum*-like ribotype; **(C)** *Panicum*-like ribotype; **(D)** *Euclasta*-like ribotype.

Supplementary Table 1 | Characteristics of *Hordeum* samples used in the study.

Supplementary Table 2 | Detailed description of the qPCR assays used to quantify nrDNA in *Hordeum*.

Supplementary Table 3 | Comparison of copy numbers of native ribotype estimated using qPCR and Illumina read mapping.

Supplementary File 1 | The distribution of internal transcribed spacer 1 (ITS1) of nuclear ribosomal DNA in the genome of *Hordeum vulgare*.

REFERENCES

Albach, D. C., and Greilhuber, J. (2004). Genome size variation and evolution in Veronica. *Ann. Bot.* 94, 897–911. doi: 10.1093/aob/mch219

Álvarez, I., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/S1055-7903(03)00208-2

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., and Zimmer, E. (1980). Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc. Natl. Acad. Sci. U.S.A.* 77, 7323–7327. doi: 10.1073/pnas.77.12.7323

Ben-Shem, A., de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529. doi: 10.1126/science.1212642

Blattner, F. R. (2004). Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. *Mol. Phylogenet. Evol.* 33, 289–299. doi: 10.1016/j.ympev.2004.05.012

Blattner, F. R. (2009). Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). *Breed. Sci.* 59, 471–480. doi: 10.1270/jsbbs.59.471

Blommaert, J., Riss, S., Hecox-Lea, B., Mark Welch, D. B., and Stelzer, C. P. (2019). Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* 20:466. doi: 10.1186/s12864-019-5859-y

Brassac, J., and Blattner, F. R. (2015). Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and in silico cloning of multiple nuclear loci. *Syst. Biol.* 64, 792–808. doi: 10.1093/sysbio/syv035

Chrtek, J., Zahradníček, J., Krak, K., and Fehrer, J. (2009). Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. *Ann. Bot.* 104, 161–178. doi: 10.1093/aob/mcp107

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Dubcovsky, J., and Dvořák, J. (1995). Ribosomal RNA multigene loci: nomads of the Triticeae genomes. *Genetics* 140, 1367–1377. doi: 10.1093/genetics/140.4.1367

Fehrer, J., Krak, K., and Chrtek, J. (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise. *BMC Evol. Biol.* 9:239. doi: 10.1186/1471-2148-9-239

Fuertes Aguilar, J., Rosselló, J. A., and Nieto Feliner, G. (1999). Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria*

(Plumbaginaceae). *Mol. Ecol.* 8, 1341–1346. doi: 10.1046/j.1365-294X.1999.00690.x

Govindaraju, D. R., and Cullis, C. (1992). Ribosomal DNA variation among populations of a *Pinus rigida* Mill. (pitch pine) ecosystem: I. distribution of copy numbers. *Heredity (Edinb)* 69, 133–140. doi: 10.1038/hdy.1992.106

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., and Lysák, M. A. (2019). The large genome size variation in the *Hesperis clade* was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Ann. Bot.* 124, 103–120. doi: 10.1093/aob/mcz036

Jakob, S. S., Meister, A., and Blattner, F. R. (2004). The considerable genome size variation of *Hordeum* Species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Mol. Biol. Evol.* 21, 860–869. doi: 10.1093/molbev/msh092

Jakob, S. S., Rödder, D., Engler, J. O., Shaaf, S., Özkan, H., Blattner, F. R., et al. (2014). Evolutionary history of wild barley (*Hordeum vulgare* subsp. *spontaneum*) analyzed using multilocus sequence data and paleodistribution modeling. *Genome Biol. Evol.* 6, 685–702. doi: 10.1093/gbe/evu047

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., et al. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289. doi: 10.1038/s41586-020-2947-8

Johnson, S. M., Carlson, E. L., and Pappagianis, D. (2015). Determination of ribosomal DNA copy number and comparison among strains of *Coccidioides*. *Mycopathologia* 179, 45–51. doi: 10.1007/s11046-014-9820-y

Kang, M., Tao, J., Wang, J., Ren, C., Qi, Q., Xiang, Q. Y., et al. (2014). Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. *New Phytol.* 202, 1371–1381. doi: 10.1111/nph.12726

Ko, K. S., and Jung, H. S. (2002). Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum*. *Mol. Phylogenet. Evol.* 23, 112–122. doi: 10.1016/S1055-7903(02)00009-X

Kono, T. J. Y., Liu, C., Vonderharr, E. E., Koenig, D., Fay, J. C., Smith, K. P., et al. (2019). The fate of deleterious variants in a barley genomic prediction population. *Genetics* 213, 1531–1544. doi: 10.1534/genetics.119.302733

Kovařík, A., Dadejová, M., Lim, Y. K., Chase, M. W., Clarkson, J. J., Knapp, S., et al. (2008). Evolution of rDNA in *Nicotiana* allopolyploids: A potential link between rDNA homogenization and epigenetics. *Ann. Bot.* 101, 815–823. doi: 10.1093/aob/mcn019

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leitch, I. J., and Heslop-Harrison, J. S. (1992). Physical mapping of the 18S-5.8S-26S rRNA genes in barley by in situ hybridization. *Genome* 35, 1013–1018. doi: 10.1139/g92-155

Liu, M., Li, Y., Ma, Y., Zhao, Q., Stiller, J., Feng, Q., et al. (2019). The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. *Plant Biotechnol. J.* 18, 443–456. doi: 10.1111/pbi.13210

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884–890. doi: 10.1038/ng.2678

Mahelka, V., and Kopecký, D. (2010). Gene capture from across the grass family in the allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triticeae) as evidenced by ITS, GBSSI, and molecular cytogenetics. *Mol. Biol. Evol.* 27, 1370–1390. doi: 10.1093/molbev/msq021

Mahelka, V., Krak, K., Fehrer, J., Caklová, P., Nagy Nejedlá, M., Čegan, R., et al. (2021). A *Panicum*-derived chromosomal segment captured by *Hordeum* a few million years ago preserves a set of stress-related genes. *Plant J.* 105, 1141–1164. doi: 10.1111/tpj.15167

Mahelka, V., Krak, K., Kopecký, D., Fehrer, J., Šafář, J., Bartoš, J., et al. (2017). Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1726–1731. doi: 10.1073/pnas.1613375114

Malinská, H., Tate, J. A., Matyášek, R., Leitch, A. R., Soltis, D. E., Soltis, P. S., et al. (2010). Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon* (Asteraceae) allotetraploids. *BMC Evol. Biol.* 10:291. doi: 10.1186/1471-2148-10-291

Mandák, B., Krak, K., Vít, P., Pavlíková, Z., Lomonosova, M. N., Habibi, F., et al. (2016). How genome size variation is linked with evolution within *Chenopodium* sensu lato. *Perspect. Plant Ecol. Evol. Syst.* 23, 18–32. doi: 10.1016/j.ppees.2016.09.004

Mascher, M., Richmond, T. A., Gerhardt, D. J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 76, 494–505. doi: 10.1111/tpj.12294

Matyášek, R., Tate, J. A., Lim, Y. K., Šrubařová, H., Koh, J., Leitch, A. R., et al. (2007). Concerted evolution of rDNA in recently formed *Tragopogon* allotetraploids is typically associated with an inverse correlation between gene copy number and expression. *Genetics* 176, 2509–2519. doi: 10.1534/genetics.107.072751

McCann, J., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., and Weiss-Schneeweiss, H. (2020). Differential genome size and repetitive DNA evolution in diploid species of *Melampodium* sect. *Melampodium* (Asteraceae). *Front. Plant Sci.* 11:362. doi: 10.3389/fpls.2020.00362

Monat, C., Padmarasu, S., Lux, T., Wicker, T., Gundlach, H., Himmelbach, A., et al. (2019). TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* 20, 284. doi: 10.1186/s13059-019-1899-5

Potapova, T. A., and Gerton, J. L. (2019). Ribosomal DNA and the nucleolus in the context of genome organization. *Chromosom. Res.* 27, 109–127. doi: 10.1007/s10577-018-9600-5

Prokopowich, C. D., Gregory, T. R., and Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46, 48–50. doi: 10.1139/g02-103

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rabanal, F. A., Nizhynska, V., Mandáková, T., Novikova, P. Y., Lysák, M. A., Mott, R., et al. (2017). Unstable inheritance of 45S rRNA genes in *Arabidopsis thaliana*. *G3 Genes Genom. Genet.* 7, 1201–1209. doi: 10.1534/g3.117.040204

Rogers, S. O., and Bendich, A. J. (1987). Heritability and variability in ribosomal RNA genes of *Vicia faba*. *Genetics* 117, 285–295.

Rosato, M., Kovařík, A., Garilleti, R., and Rosselló, J. A. (2016). Conserved organisation of 45S rDNA sites and rDNA gene copy number among major clades of early land plants. *PLoS One* 11:e0162544. doi: 10.1371/journal.pone.0162544

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 81, 8014–8018. doi: 10.1073/pnas.81.24.8014

Sproul, J. S., Barton, L. M., and Maddison, D. R. (2020). Repetitive DNA profiles reveal evidence of rapid genome evolution and reflect species boundaries in ground beetles. *Syst. Biol.* 69, 1137–1148. doi: 10.1093/sysbio/syaa030

Srivastava, A. K., and Schlessinger, D. (1991). Structure and organization of ribosomal DNA. *Biochimie* 73, 631–638. doi: 10.1016/0300-9084(91)90042-Y

Taketa, S., Ando, H., Takeda, K., and von Bothmer, R. (2001). Physical locations of 5S and 18S-25S rDNA in Asian and American diploid *Hordeum* species with the I-genome. *Heredity (Edinb)* 86, 522–530. doi: 10.1046/j.1365-2540.2001.00768.x

Taketa, S., Harrison, G. E., and Heslop-Harrison, J. S. (1999). Comparative physical mapping of the 5S and 18S-25S rDNA in nine wild *Hordeum* species and cytotypes. *Theor. Appl. Genet.* 98, 1–9. doi: 10.1007/s001220051033

Talla, V., Suh, A., Kalsoom, F., Dinca, V., Vila, R., Friberg, M., et al. (2017). Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (Leptidea) butterflies. *Genome Biol. Evol.* 9, 2491–2505. doi: 10.1093/gbe/evx163

Veiko, N. N., Shubaeva, N. O., Malashenko, A. M., Beskova, T. B., Agapova, R. K., and Lyapunova, N. A. (2007). Ribosomal genes in inbred mouse strains: interstrain and intrastrain variation of copy number and extent of methylation. *Russ. J. Genet.* 43, 1021–1031. doi: 10.1134/S1022795407090086

Vitales, D., Álvarez, I., Garcia, S., Hidalgo, O., Nieto Feliner, G., Pellicer, J., et al. (2020). Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Ann. Bot.* 125, 611–623. doi: 10.1093/aob/mcz183

Wang, W., Wan, T., Becher, H., Kuderová, A., Leitch, I. J., Garcia, S., et al. (2019). Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Ann. Bot.* 123, 767–781. doi: 10.1093/aob/mcy172

Weiss-Schneeweiss, H., Greilhuber, J., and Schneeweiss, G. M. (2006). Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. *Am. J. Bot.* 93, 148–156. doi: 10.3732/ajb.93.1.148

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. U.S.A.* 92, 280–284. doi: 10.1073/pnas.92.1.280

Wong, W. Y., Simakov, O., Bridge, D. M., Cartwright, P., Bellantuono, A. J., Kuhn, A., et al. (2019). Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22915–22917. doi: 10.1073/pnas.1910106116

Xu, W., Tucker, J. R., Bekele, W. A., You, F. M., Fu, Y.-B., Khanal, R., et al. (2021). Genome assembly of the Canadian two-row malting barley cultivar AAC Synergy. *G3 Genes Genom. Genet.* 11:jkab031. doi: 10.1093/g3journal/jkab031

Zedek, F., Šmerda, J., Šmarda, P., and Bureš, P. (2010). Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. *BMC Plant Biol.* 10:265. doi: 10.1186/1471-2229-10-265

Zozomová-Lihová, J., Mandáková, T., Kovaříková, A., Muhlhausen, A., Mummenhoff, K., Lysák, M. A., et al. (2014). When fathers are instant losers: homogenization of rDNA loci in recently formed *Cardamine schulzii* trigenomic allopolyploid. *N. Phytol.* 203, 1096–1108. doi: 10.1111/nph.12873

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2021
doi: 10.3389/fpls.2021.678689

Mosaic Arrangement of the 5S rDNA in the Aquatic Plant *Landoltia punctata* (Lemnaceae)

Guimin Chen[1,2†], *Anton Stepanenko*[1,2†‡] *and Nikolai Borisjuk*[1,2*‡]

[1]*Jiangsu Key Laboratory for Eco-Agricultural Biotechnology Around Hongze Lake, School of Life Sciences, Huaiyin Normal University, Huai'an, China,* [2]*Jiangsu Collaborative Innovation Centre of Regional Modern Agriculture & Environmental Protection, Huaiyin Normal University, Huai'an, China*

OPEN ACCESS

Edited by:
Ana F. Miranda,
RMIT University, Australia

Reviewed by:
Aidyn Mouradov,
RMIT University, Australia
Eduardo Cires,
University of Oviedo, Spain
Gennady I. Karlov,
All-Russia Research Institute of Agricultural Biotechnology, Russia

***Correspondence:**
Nikolai Borisjuk
nborisjuk@hytc.edu.cn

[†]*These authors have contributed equally to this work*

[‡]**ORCID:**
Anton Stepanenko
orcid.org/0000-0003-2326-6613
Nikolai Borisjuk
orcid.org/0000-0001-5250-9771

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *15 March 2021*
Accepted: *31 May 2021*
Published: *24 June 2021*

Citation:
Chen G, Stepanenko A and Borisjuk N (2021) Mosaic Arrangement of the 5S rDNA in the Aquatic Plant Landoltia punctata (Lemnaceae). Front. Plant Sci. 12:678689. doi: 10.3389/fpls.2021.678689

Duckweeds are a group of monocotyledonous aquatic plants in the Araceae superfamily, represented by 37 species divided into five genera. Duckweeds are the fastest growing flowering plants and are distributed around the globe; moreover, these plants have multiple applications, including biomass production, wastewater remediation, and making pharmaceutical proteins. Dotted duckweed (*Landoltia punctata*), the sole species in genus Landoltia, is one of the most resilient duckweed species. The ribosomal DNA (rDNA) encodes the RNA components of ribosomes and represents a significant part of plant genomes but has not been comprehensively studied in duckweeds. Here, we characterized the 5S rDNA genes in *L. punctata* by cloning and sequencing 25 PCR fragments containing the 5S rDNA repeats. No length variation was detected in the 5S rDNA gene sequence, whereas the nontranscribed spacer (NTS) varied from 151 to 524 bp. The NTS variants were grouped into two major classes, which differed both in nucleotide sequence and the type and arrangement of the spacer subrepeats. The dominant class I NTS, with a characteristic 12-bp TC-rich sequence present in 3–18 copies, was classified into four subclasses, whereas the minor class II NTS, with shorter, 9-bp nucleotide repeats, was represented by two identical sequences. In addition to these diverse subrepeats, class I and class II NTSs differed in their representation of cis-elements and the patterns of predicted G-quadruplex structures, which may influence the transcription of the 5S rDNA. Similar to related duckweed species in the genus Spirodela, *L. punctata* has a relatively low rDNA copy number, but in contrast to Spirodela and the majority of other plants, the arrangement of the 5S rDNA units demonstrated an unusual, heterogeneous pattern in *L. punctata*, as revealed by analyzing clones containing double 5S rDNA neighboring units. Our findings may further stimulate the research on the evolution of the plant rDNA and discussion of the molecular forces driving homogenization of rDNA repeats in concerted evolution.

Keywords: duckweed, 5S rRNA genes, gene organization, molecular evolution, *Landoltia punctata*

INTRODUCTION

The plant ribosomal DNA (rDNA) consists of highly conserved regions coding for 18S–5.8S–25S ribosomal RNAs (rRNAs), 45S rDNA, and 5S rRNAs, 5S rDNA, intertwined with more rapidly evolving nontranscribed spacer (NTS) sequences. Based on their high copy number in plant genomes, and structural features, the 45S rDNA and 5S rDNA loci have been broadly used in studies of plant systematics, evolution, and biodiversity and as molecular markers of ancestral genomes in polyploids and various hybrids (Borisjuk et al., 1988; Stadler et al., 1995; Mahelka et al., 2017). The 5S rDNA is especially well suited for such studies, due to the smaller size of its repeat units, making the sequences technically easier to handle compared to the much larger 45S rDNA, and to the higher variability exhibited by 5S rDNA NTSs relative to those of the 45S rDNA. The 5S rDNA loci have been characterized for representatives of numerous plant taxa to reveal phylogenetic relationships (Baker et al., 2000; Pan et al., 2000; Röser et al., 2001), genome evolution (Kellogg and Appels, 1995; Allaby and Brown, 2001; Simon et al., 2018), and the subgenome composition of polyploids (Kovarik et al., 2008; Baum and Feldman, 2010; Sergeeva et al., 2017; Volkov et al., 2017), natural hybrids, and artificial hybrids (Zanke et al., 1995; Fulnecek et al., 2002; Matyácek et al., 2002; Volkov et al., 2007; Mahelka et al., 2013).

Duckweeds are a group of floating plants present in local aquatic ecosystems worldwide, where they often cover large areas of the water surface (Landolt, 1986; Tippery and Les, 2020). Duckweeds were a favorite model for plant biochemistry studies from the 1950s to the 1980s before being supplanted by model plants, such as Arabidopsis (*Arabidopsis thaliana*). However, these aquatic plants came back into the spotlight in the 2010s, primarily because of their potential as a promising feedstock for the production of biofuels and other valuable biochemicals (Cao et al., 2018; Zhou and Borisjuk, 2019; Liu et al., 2021). Different duckweed species are also widely used for wastewater treatment (Zhao et al., 2015; Zhou et al., 2018) and biosensing (Ziegler et al., 2019). The establishment of living *in vitro* collections hosting ~2,000 duckweed ecotypes (Sree and Appenroth, 2020), primarily at the world duckweed depository hosted by Prof. E. Lam at the Rutgers Duckweed Stock Cooperative at Rutgers University, New Brunswick, NJ, United States,[1] and a number of local collections in Canada, China, Germany, Hungary, India, Ireland, and Switzerland (Lam et al., 2020) have supported and helped promote modern duckweed research.

Duckweeds are an ancient group of monocot plants with extremely reduced morphology. Their exact taxonomic status, as a distinct family (Lemnaceae) or as a subfamily that belongs to the Araceae (Les and Tippery, 2013), is still debated (Tippery and Les, 2020). The 37 known species of duckweeds are currently classified into five genera: Spirodela, Landoltia, Lemna, Wolffia, and Wolffiella (Appenroth et al., 2013). The genus Landoltia, represented by the single species dotted duckweed or duckmeat (*Landoltia punctata*), is believed to have separated relatively recently from Spirodela, based on the morphological and new molecular data (Les and Crawford, 1999). In addition to the benefits commonly provided by duckweed species, such as fast, dense growth on the water surface, easy harvesting, convenient enzymatic saccharification of biomass, and efficient phytoremediation of wastewater (Xu and Shen, 2011; Zhao et al., 2015; Ziegler et al., 2015; Zhou et al., 2018), *L. punctata* has the added advantage of being one of the most resilient and stress-resistant among all duckweeds. For example, in the subtropical climate of Eastern China, *L. punctata* is the first duckweed species to colonize water reservoirs in the Spring and the last remaining in the Fall, thus exhibiting the longest vegetative growth period compared with other duckweeds. Based on these qualities, the species has attracted much attention as a promising, inexpensive, and sustainable source of valuable biomass for the production of biofuels, such as ethanol, butanol, biogas, and hydrogen (Tao et al., 2017; Toyama et al., 2018; Miranda et al., 2020), and high-value biochemicals, such as succinic acid (Shen et al., 2018).

In addition to these applications, the genetic diversity seen in duckweeds has stimulated a recent burst of studies examining duckweed genomics, molecular evolution, ecology, and biodiversity (Appenroth et al., 2015; Laird and Barks, 2018; Ho et al., 2019). Despite the fact that the duckweed 5S rDNA was one of the first sequenced genes in plants (Vandenberghe et al., 1984), the rDNA remains relatively poorly studied in duckweeds and in the Araceae. Our current knowledge of the molecular organization of the rDNA in the duckweed relates to the partial sequencing of 35S rDNA repeats from representative duckweed species in a study aiming to investigate the phylogenetic relationships and evolutionary history of Lemnaceae by Tippery et al. (2015) and to whole genome sequences of *Spirodela polyrhiza* (Michael et al., 2017) and *Spirodela intermedia* (Hoang et al., 2020), which revealed important characteristics of the 35S and 5S rDNA loci.

In this study, we present the molecular organization of the 5S rDNA locus in one *L. punctata* ecotype originating from Eastern China, based on the characterization of 25 independent sequences derived from cloned PCR products. Our results provide new information on the diversity and arrangement of the rDNA in this species and shed new light on general principles of evolution and arrangement of the 5S rDNA in plants.

MATERIALS AND METHODS

Plant Materials

The duckweed ecotype used in this study was collected in summer of 2017 from a lake (GPS location: N 33″618817, E 119″001941) in one of the parks in the East China city of Huai'an. The collected fronds were surface sterilized in 0.5% sodium hypochlorite and 0.1% benzalkonium bromide in order to establish an aseptically grown strain. The NB0014 strain, developed from a single frond, is maintained as an *in vitro* culture on 0.8% agar containing 0.5× Schenk and Hildebrandt (SH) salts (Sigma-Aldrich, St. Louis, MO, United States) and 0.5% sucrose, pH 5.7–6.0, under axenic conditions. The identity of the NB0014 as the species of *Landoltia punctata* was confirmed by DNA barcoding using primers specific for chloroplast

[1]http://www.ruduckweed.org

intergenic spacers atpF-atpH (ATP) and psbK-psbL (PSB), recommended by Consortium for the Barcode of Life (CBOL), as previously described (Borisjuk et al., 2015).

Cloning and Sequence Characterization of *L. punctata* 5S rDNA Genes

For analysis of 5S rDNA genes, total DNA was isolated from the *in vitro* propagated biomass of *L. punctata* NB0014 using a cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson, 1980) modified according to Borisjuk et al. (2015). The 5S rDNA genes were amplified from genomic DNA by PCR using the 5S rDNA gene-specific primers DW-5S-F: CTTGGGCGAGAGTAGTACTAGG and DW-5S-R: CACGCTTA ACTTCGGAGTTCTG. The generated DNA fragments were purified by gel electrophoresis, cloned into the vector pMD19 (TaKaRa, Dalian, China), and custom sequenced by the Sangon Biotech (Shanghai, China). The obtained sequences were primarily analyzed using the "Online Analysis Tools" package.[2] The subrepeats were characterized using the advanced hidden Markov model with the CLC Main Workbench (Version 6.9.2, QIAGEN Digital Insights, Redwood City, CA, United States) software. For the detection of the DNA regions likely to fold into G-quadruplex structures, we have primarily used the pqsfinder prediction tool (Labudová et al., 2020) available at the website: https://pqsfinder.fi.muni.cz/, with further verification by the G4Hunter algorithm (Brázda et al., 2019), freely available at the DNA Analyzer server: https://bioinformatics.ibp.cz.

Estimation of 5S and 35S rDNA Copy Number

The estimation of 5S and 35S rDNA gene copies was carried out by quantitative PCR (qPCR), relating the rates of the DNA amplification of samples to the standard curve. The standard curves were established based on the amplification reads of independent dilution series of two specially constructed reference plasmids, pAS-Lp1 and pAS-Lp2. The plasmids were assembled using a backbone of pAS-Sp1, previously constructed to calculate rDNA copy number in *S. intermedia* (Hoang et al., 2020) by replacing a single-copy Actin gene specific for Spirodela with *L. punctata* sequences encoding nitrate reductase, NR (pAS-Lp1) and nitrite reductase, NiR (pAS-Lp2). The NR and NiR sequences were PCR amplified from genomic DNA of *L. punctata* using primers designed according to the gene sequences kindly shared by Todd Michael (J. Craig Venter Institute, San Diego, CA, United States). The arrangements of reference genes in pAS-Lp1 and pAS-Lp2 plasmids with primers used in qPCR are represented in **Supplementary Figure S1**. The number of rDNA gene copies was determined in qPCRs prepared with the UltraSybr Mixture (CWBio, Taizhou, China), run on the CFX Connect Real-Time Detection System (Bio-Rad, Hercules, CA, United States). The samples and 10-fold dilution series of the reference plasmids were assayed in the same run. The quality of products was checked by the thermal denaturation cycle. Only the experiments providing a single peak were considered. Three technical replicates were performed for each sample. The obtained data were analyzed using the program BIO-RAD CFX Manager 3.1 (Hercules, CA, United States) and Microsoft Excel 2016 software (Microsoft Corp., Redmond, WA, United States).

[2]https://molbiol-tools.ca

RESULTS

Characterization of the 5S rDNA in *L. punctata*

Landoltia punctata, a duckweed species inhabiting mostly tropical and subtropical regions (Tippery and Les, 2020), is represented in this study by ecotype NB0014, which was isolated in the Jiangsu province of Eastern China. To detect possible intragenomic variation, we analyzed 5S rDNA repeats by cloning PCR products amplified with primers designed to cover the two halves of neighboring 5S rDNA genes with the NTS in the center, then sequencing individual clones. In total, 25 clones containing inserts ranging from 260 to 653 bp were sequenced and analyzed, including five clones containing two sequential 5S rDNA units (**Figure 1**).

Conserved 5S rDNA Gene Sequence and rDNA Copy Number

All sequenced clones representing building blocks of the 5S rDNA locus consisted of a common unit of 119 bp coding for 5S rDNA and an adjacent NTS. Across the 30 5S rDNA sequences (from 20 clones with one copy and five clones with two copies of the rDNA), we detected six nucleotide substitutions but no variation in the 5S rDNA gene length (**Supplementary Figure S2**). Five of these variants were T/C or A/G transitions, with the final substitution being a T/G transversion.

We identified all regulatory sequences in the 5S rDNA locus from *L. punctata*, such as the A-box, intermediate element (IE), and C-box, which are characteristic of plant genes (Hemleben and Werts, 1988; Cloix et al., 2003). The 5S rDNA transcribed from the locus was predicted to form a secondary structure similar to that seen in other plant species (**Figure 2**). The specific A/G transitions at nucleotide +50 in clones NB0014-22 and NB0014-23, which had the two shortest rDNA units with an NTS of 151 bp, mapped to a loop in the predicted 5S rDNA secondary structure, where it is unlikely to interfere with rDNA folding (**Figure 2**).

We next estimated the 5S rDNA and 25S rDNA gene copy numbers by qPCR using the approach previously developed to estimate rDNA copy number in *S. intermedia* (Hoang et al., 2020), with NR and NiR as two single-copy genes in *L. punctata* for normalization. We determined that the 5S rDNA locus was represented by 168 ± 25 gene copies, whereas the 35S rDNA had 176 ± 37 copies in the genome of the *L. punctata* ecotype NB0014.

The NTS Shows Variant Subrepeat Structures

The 5S rDNA NTS region showed a significant variation in sequence length, ranging from 151 to 524 bp. A full alignment of the 30 NTS sequences represented in the 25 clones clearly

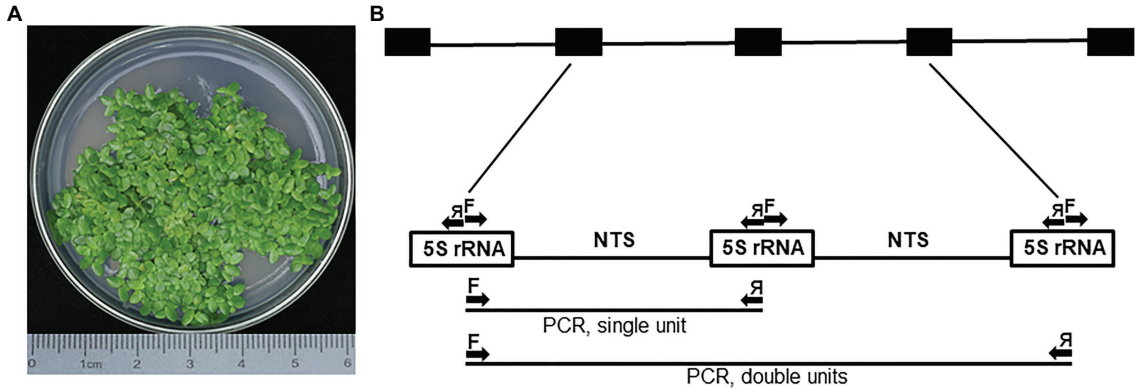


FIGURE 1 | The PCR amplification of 5S rDNA units in *Landoltia punctata* NB0014. **(A)** Representative image of the *Landoltia punctata* ecotype NB0014 during *in vitro* culture; the species identity was confirmed by the chloroplast DNA barcoding. **(B)** Schematic representation of the complete 5S rDNA locus and the PCR amplification scheme. Arrows represent forward (F) and reverse (я) primers used during PCR and may amplify single or double 5S rRNA units. The array of black boxes at the top represents the 5S rDNA locus, and the white boxes represent the DNA sequences transcribed to produce the 5S rRNA. NTS, nontranscribed spacer.

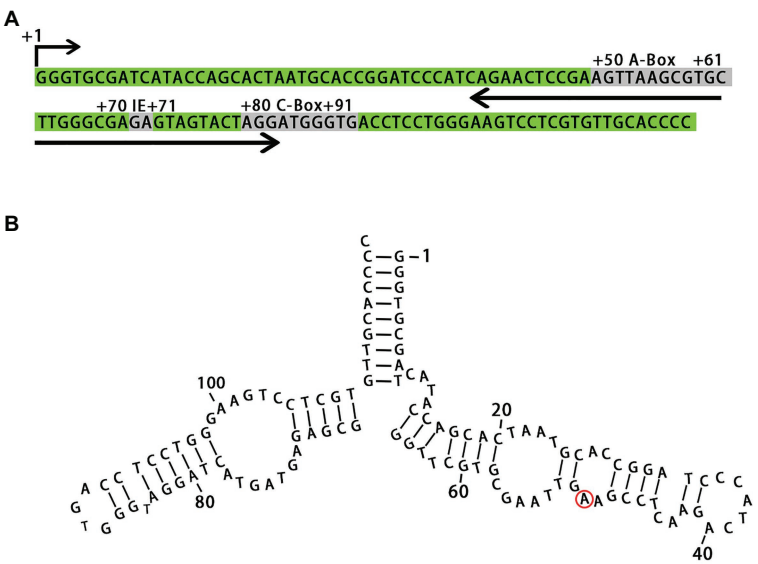


FIGURE 2 | Primary nucleotide sequence of the 5S rDNA gene and secondary structure of the 5S rRNA. **(A)** Conserved motifs involved in transcriptional regulation are marked as A-box, IE, and C-box. **(B)** The position of nucleotide substitution specific for clones with short NTS (NB0014-22 and NB0014-23) is marked with a red ring. IE, intermediate element.

separated them into two groups: one with 28 NTS sequences, and the other represented by two identical sequences from clones NB0014-16A and NB0014-25 (**Supplementary Figure S3**). Both classes were characterized by specific conserved sequences at their 5' ends (with lengths of 95 bp for NTS class I and 169 bp for NTS class II) and started with the transcription termination sequence TTTT (Hemleben and Werts, 1988). Both classes also shared a more variable TC-rich region in their center. In 26 out of the 28 clones with a class I NTS, the length of the NTS was over 400 bp, with a range between 432 and 524 bp (**Supplementary Figure S4**). The two remaining clones, NB0014-22 and NB0014-23, had the shortest variants, as they lacked a large portion of the 3' end, with the exception of the 12 nucleotides directly adjacent to the 5S rDNA gene (**Supplementary Figure S5**).

An in-depth analysis of all NTS sequences using CLC software (Version 6.9.2, QIAGEN Digital Insights, Redwood City, CA, United States) revealed that the variable central region of both NTS classes is composed of repeated units (subrepeats) of 12 or 9 bp for class I and class II NTS, respectively, in various arrangements. Notably, both NTS classes shared a 6-bp core (T-C-T/C-T/C-T/C-T/C; **Figure 3**). The 9-bp repeats detected in the two clones with class II NTS were organized as a single copy followed by a 12-bp sequence resembling an almost perfectly duplicated 6-bp core, tCTTCT-cCTTCC, followed by seven copies of the 9-bp element (**Figure 3B**).

The arrangements of the 12-bp units among clones with class I NTS were more complex and were divided into four subclasses (**Figure 3A**). Subclasses A1, A2, and B, represented by 26 sequences of class I NTS, all started with two copies of the 12-bp unit, followed by the sequence TCC, another 12-bp repeat, the sequence CC, one more 12-bp repeat and the sequence TCC. The three subclasses then diverged in their arrangement, with five (A1), four (A2), or six (B) 12-bp repeats. The predominant subclass A1 with 22 NTS clones was further extended by a block of four to six consecutive 12-bp units with no spacer sequence. Subclass A2, which was represented by three individual clones, showed a slight, distinct extension of its repeat sequence by a block of five consecutive 12-bp units. Finally, subclass B, represented by a single NTS variant, contained a single 12-bp unit following the basic unit described above. Subclasses A1, A2, and B were characterized by a highly conserved 194-bp sequence between the subrepeats and the 5S rDNA gene. We named the fourth subclass ABbr (for abbreviated), as it's NTS comprised only three 12-bp repeats identical to subclasses A and B, but lacked the rest of the NTS sequences, with the exception of the final 12 bp upstream of the 5S rDNA gene (**Figure 3B**).

Possible Alternative Regulation of 5S rDNA Variants in *L. punctata*

The basic regulatory elements for transcription by Pol III are located within the 5S rDNA gene sequence (**Figure 1**), but upstream cis-elements, such as TATA-like motifs and GC dinucleotides, may also significantly contribute to the modulation of transcription. The upstream regions of representative class I and class II NTSs of the 5S rDNA locus in *L. punctata* showed some divergence from the previously published arrangement in plants (Venkateswarlu et al., 1991; Cloix et al., 2003). For example, in clone NB0014-15, the GC dinucleotide was preserved at conserved position −12 and −11 from the transcription start in the class I 5S rDNA; however, the −28 to −23 location normally occupied by the TATATA-box in plants had the sequence ACATGA instead (**Figure 4A**). In the class II NTS 5S rDNA locus, represented by clone NB0014-25, the TATA-box was replaced by the related sequence ATATGT, but a TG dinucleotide occupied the −12 to −11 position, instead of the conserved GC dinucleotide.

We then used the pqsfinder algorithm (Labudová et al., 2020) to determine the potential for forming regulatory G-quadruplex structures (G4), which consist of four guanine-rich regions held together *via* unconventional base pairing (Lipps and Rhodes, 2009; Havlová and Fajkus, 2020). We discovered that the TG dinucleotide is a part of the 3' end of the sequence TGGGA on the reverse strand within the −37 to −8 sequence predicted to form a G4 structure, just upstream of the 5S rDNA gene (**Figure 4B**).

In addition to the differences in nucleotide organization noted above in the immediate 5S rDNA gene upstream regions, the general patterns of the revealed G-quadruplex structure were also distinct between class I and II NTSs. The pqsfinder algorithm predicted three G4-forming regions on the forward DNA strand of class I NTS, between the subrepeats and the 5S rDNA gene, with the highest score at position −142 to −106. Class II NTS had six potential G4-forming regions, three each on the forward and reverse DNA strand, scattered over the entire length of the NTS, with the highest scores obtained for the element located on the reverse strand within the subrepeats region. We validated these specific patterns with the G4 prediction tool G4Hunter (Brázda et al., 2019).

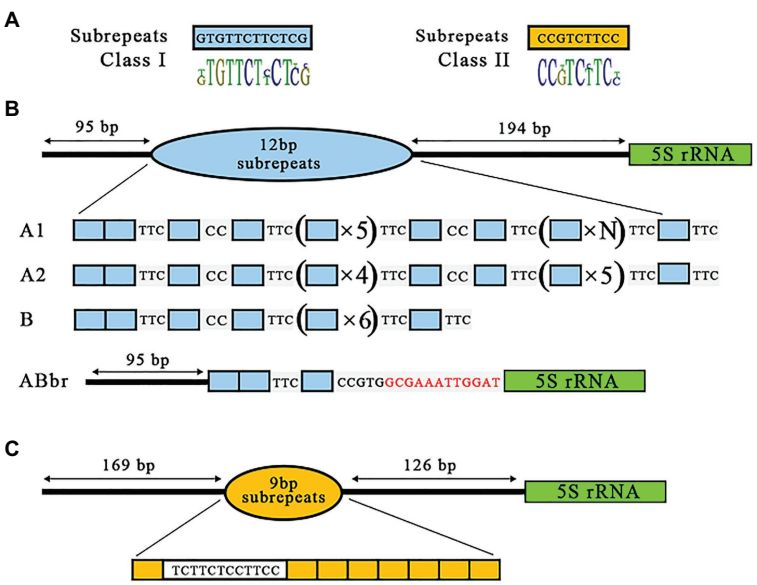


FIGURE 3 | Diversity of 5S rDNA NTS repeats in the gene of *Landoltia punctata* NB0014. **(A)** The NB0014 5S rDNA NTS repeats are classified into class I and class II, based on the type, number, and arrangement of their 12-bp (blue blocks) or 9-bp (yellow blocks) subrepeats in relation to the 5S rDNA gene (marked by green; **B**). Class I is further divided into four subclasses: A1, A2, B, and ABbr. Red letters in the ABbr variant indicate the 12 nucleotides upstream of the 5S rDNA gene that are shared by all NTS class I sequences **(C)**. Class II is represented by a single NTS type of 9-bp units.

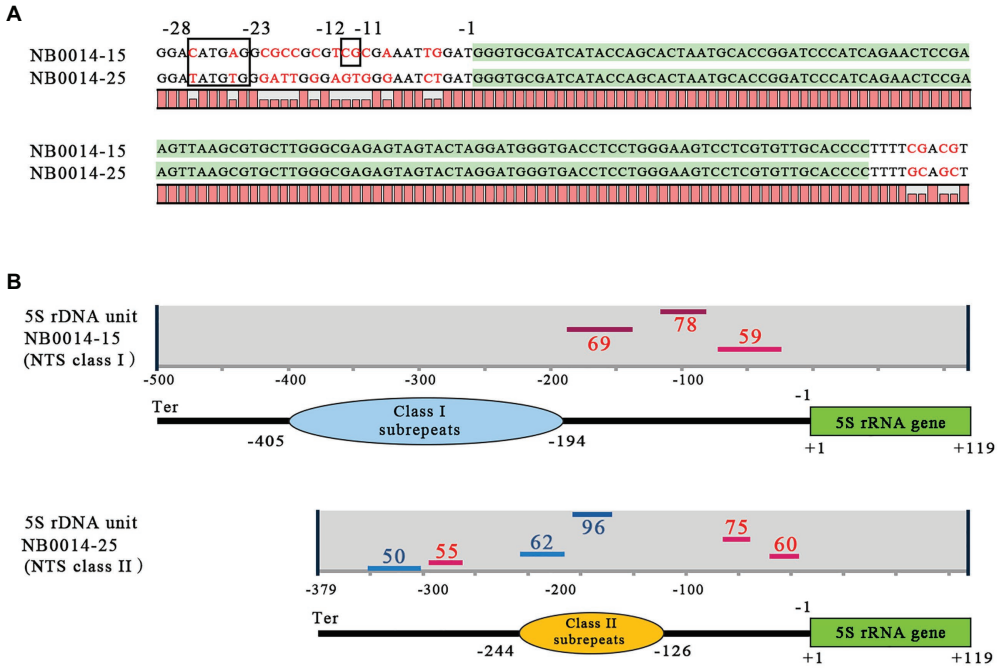


FIGURE 4 | Features of class I and class II NTS sequences with the potential to modulate 5S rDNA transcription. **(A)** Regulatory DNA elements upstream and downstream of the 5S rDNA gene. The coding sequence for the 5S rRNA is highlighted in green; nucleotide positions from −28 to −23 mark the TATA-like motif; nucleotide positions −12 to −11 mark the GC dinucleotide; position −1 marks the first nucleotide upstream of the 5S rDNA transcription start. **(B)** Patterns of G-quadruplex structures predicted for class I and class II NTS sequences. Positions of G-quadruplex structures are indicated by horizontal lines (red for forward DNA strand, blue for reverse strand); numbers next to bars indicate the relative strength of each G-quadruplex structure, and the negative (–) numbers indicate their positions relative to the transcription start of the 5S rDNA gene; Ter marks the position of the terminator with sequence TTTT. NTS, nontranscribed spacer.

Clones With Two 5S rDNA Units Hint at Low Repeat Homogenization Within the Locus

Tandemly repeated rDNA units are thought to undergo high levels of homogenization within the array, due to concerted evolution (Coen et al., 1982; Cronn et al., 1996). Of the 25 sequenced clones, five harbored two repeat units, as illustrated in **Figure 1**. These paired repeats offered us a glimpse into the arrangement of individual units along the 5S rDNA locus. Two clones with two 5S rDNA copies, NB0014-17 and NB0014-20, contained identical A1 variants of the NTS: two NTSs of 524 bp for NB0014-17 and two NTSs of 500 bp for NB0014-20. Three other clones harbored pairs of NTS each belonging to different classes, with NB0014-18 and NB0014-19 having class A1 and class A2 units. Even more surprising was the NTS pair in clone NB0014-16, with a class I type A1 unit followed by a class II NTS (**Figure 5**). The obtained results suggest a rather random mosaic arrangement of 5S rDNA units in *L. punctata*.

DISCUSSION

The duckweed 5S rDNA locus was among the first plant loci whose sequence was determined in the 1980s (Vandenberghe et al., 1984). However, rDNA loci of duckweeds did not attract substantial attention until recently. Tippery et al. (2015) investigated the phylogenetic relationships and evolutionary history of Lemnaceae by sequencing different regions of duckweed 35S rDNA repeats. Genome surveys revealed an unusually low 35S rDNA copy number in the great duckweed *S. polyrhiza* relative to other plants (Michael et al., 2017). Later studies of duckweed genomes showed that 5S rDNA genes are present as two loci in *S. polyrhiza*, *S. intermedia*, and *L. punctata* (Hoang et al., 2018, 2019). The recent sequencing of two *S. intermedia* genomes (Hoang et al., 2020) was consistent with these observations, and indicated that each 5S rDNA locus was populated by distinct units with locus-specific NTS. The total number of 5S rDNA copies was estimated to be around 70 per genome in *S. intermedia*, which is the lowest number reported to date for plants, as rDNA copy numbers typically reach the thousands.

The estimated rDNA copy number in the *L. punctata* ecotype NB0014 was 168 ± 25 copies for 5S rDNA and 176 ± 37 copies for 25S rDNA. These numbers are consistent with those seen in *S. polyrhiza* and *S. intermedia* when normalized to genome size, as the *L. punctata* genome is roughly three times bigger than that of Spirodela (Hoang et al., 2019). It is worth noting that the almost equal copy number for 5S rDNA and 25S rDNA units is atypical for plants, where the evolution of 5S and 25S rDNA loci appears to follow different patterns (Mahelka et al., 2013; Volkov et al., 2017); rather, the *L. punctata*

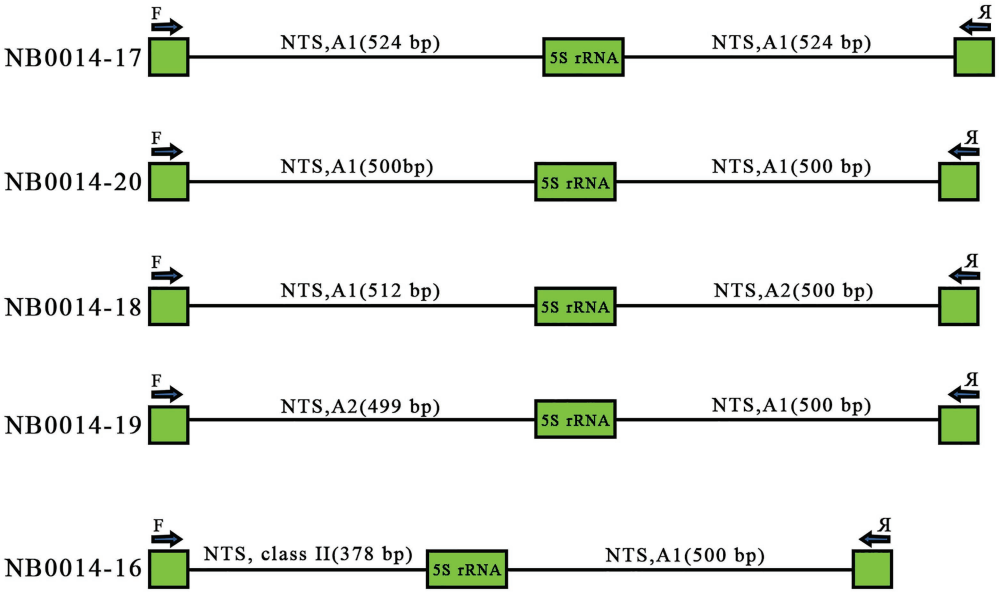


FIGURE 5 | Various modes of 5S rDNA unit arrangement in clones containing pairs of consecutive repeated units. Clones NB0014-17 and NB0014-20 each contain identical A1 type NTS sequences of 524 bp and 500 bp, respectively. Clones NB0014-18 and NB0014-19 have interspersed type A1 and A2 units, whereas the clone NB0014-16 has interspersed class II and class I, type A1 units. NTS, nontranscribed spacer.

pattern is more reminiscent of animal genomes, including human (Gibbons et al., 2015).

The two major classes of 5S rDNA repeats detected in *L. punctata* are in agreement with the genomic data produced for *S. polyrhiza* and *S. intermedia*, both of which have two types of 5S rDNA units composed of a conserved 119-bp gene coding sequence interspersed with two types of NTS, with slight but significant sequence differences in *S. intermedia* (Hoang et al., 2020) and much more profound differences both in length (~400 vs. ~1,070 bp) and nucleotide composition in *S. polyrhiza*, as can be seen in online databases for strain 7498[3] and strain 9509.[4] However, there were significant differences between the arrangements of the 5S rDNA locus in the two Spirodela species and *L. punctata*. First, almost no intragenomic or intergenomic variation was observed within each NTS type in Spirodela, which is in agreement with our own analysis of a smaller NTS variant in four *S. polyrhiza* ecotypes (Borisjuk et al., 2018). By contrast, we identified extensive variation in class I NTS in *L. punctata*, as illustrated in **Figure 3**. Second, both class I and class II NTSs in *L. punctata* contained multiple subrepeats, which is typical for intergenic spacers of 25S rDNA (Borisjuk and Hemleben, 1993; Borisjuk et al., 1997; Hemleben et al., 2021) but generally not observed for plant 5S rDNA spacers, including Spirodela.

In addition, because variation in NTS sequences directly adjacent to the 5S rDNA gene may modulate transcription (**Figure 4A**), NTS subrepeats may also contribute to regulating rDNA activity in a manner similar to 25S rDNA (Zentgraf and Hemleben, 1992; Schlögelhofer et al., 2002). This assumption is further strengthened by a prediction of strong 4G structures, which participate in the regulation of gene expression in many eukaryotic organisms (Lipps and Rhodes, 2009; Yadav et al., 2017), in the subrepeat region of class II *L. punctata* NTS (**Figure 4B**).

Even more intriguing was the finding that in three out of five clones with PCR amplicons composed of double 5S rDNA units, each neighboring repeat was represented by a distinct type of NTS (**Figure 5**). This result contradicts the basic concept of extended repeat homogenization along rDNA arrays (Eickbush and Eickbush, 2007) and contrasts with the arrangement of the 5S rDNA locus in the related Spirodela species, where each locus contained a single type of 5S rDNA unit, as shown by extra-long OxfordNano sequencing.[5]

Thus our finding, coupled with a recent discovery of variation and clustering of 35S rDNA repeats within the nucleolus organizing region (NOR2) locus in Arabidopsis (Sims et al., 2021), raises a question of the extent of homogeneity for rDNA repeats within their loci. Is the level of repeat homogenization species-specific, and if so, what are the mechanisms responsible for the differential manifestation of concerted evolution?

CONCLUSION

Our data provide the first comprehensive report on the arrangement of 5S rDNA in a representative of the Araceae plant family. In particular, the study reveals two major classes of repeated units, which differ by the composition and distribution of

[3]https://genomevolution.org/coge/GenomeInfo.pl?gid=55812
[4]https://genomevolution.org/coge/GenomeInfo.pl?gid=51364
[5]https://genomevolution.org

subrepeats in the nontranscribed intergenic spacer of the 5S rDNA, and representations of DNA elements potentially involved in the regulation of 5S rDNA transcription. The genome of *L. punctata* has one of the lowest copy numbers of rDNA genes among flowering plants and an unusual, mosaic arrangement of 5S rDNA clusters. Overall, the findings of our study shed a new light on the organization of plant rDNA and may stimulate further discussion of plant genome evolution and the molecular forces that drive the homogenization of rDNA repeats.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

GC generated the sequencing data and prepared the manuscript's figures. AS conducted genes quantifications by qPCR, organized and analyzed the data. NB conceived the idea and prepared the manuscript with contributions from GC and AS. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by an individual grant provided by Huaiyin Normal University (Huai'an, China) to NB.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Todd Michael (J. Craig Venter Institute, San Diego, CA, United States) for providing information on *Landoltia punctata* sequences for NR and NiR genes, and to Dr. Olena Kishchenko for the photograph of *in vitro* growing duckweed.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.678689/full#supplementary-material

Supplementary Figure S1 | Schematic representation of pAS-Lp1 and pAS-Lp-2 plasmids containing sequences of 25S rDNA (yellow), 5S rDNA (green), and the sequences of single copy genes for nitrate reductase (NR; dark blue) and nitrite reductase (NiR; violet). The specific primer pairs to amplify each reference gene are as follows: 5'-TCCCACTGTCCCTGTCTACT/5'-CCCACTTATCCTACACCTCT for 25S rDNA; 5'-GGGTGCGATCATACCAGCAC/5'-GGGTGCAACACGAGGACTTC and 5'-GGGTGCGATCATACCAGCAC/5'-GGGTGCAACACGAGGACTTC for 5S rDNA; 5'-ATTTTCCTTTCCGCCACC/5'-CCGATGTACTCAATGTGCC for NR; and 5'-CGCCAAAAGACGCGCATGA/5'-AAGACCGATGAAGCAGAACCC for NiR. The GenBank accession numbers are: pAS-Lp1, MW803142; pAS-Lp2, MW841295.

Supplementary Figure S2 | Nucleotide alignment of 30 sequences of the 5S rDNA genes cloned from *Landoltia punctata*. Most of the aligned sequences are compilations of the halves of neighboring gene sequences, the sequences marked with M are the 5S genes in the middle of clones containing two consecutive 5S rDNA units, as depicted in **Figure 1B**. The randomly varied nucleotides are highlighted by blue; the nucleotide substitution specific for the clones with short NTS (NB0014-22 and NB0014-23) at position 50, are highlighted by violet. NTS, nontranscribed spacer.

Supplementary Figure S3 | Bulk alignment of *Landoltia punctata* 5S rDNA NTS sequences. The sequences embrace the whole intergenic region between two 5S rDNA genes. The sequence with their names followed by A and B represent the two NTS from clones containing two consecutive 5S rDNA repeats according to **Figure 1B**. Sequence areas which differ from the consensus are highlighted in blue.

Supplementary Figure S4 | Nucleotide alignment of the *Landoltia punctata* 5S rDNA NTS class I sequences. 12 bp subrepeats are marked by black rectangles. Sequence areas that differ from the consensus are highlighted in blue.

Supplementary Figure S5 | Nucleotide alignment of the *Landoltia punctata* 5S rDNA NTS class II sequences. 9 bp subrepeats are marked by black rectangles.

REFERENCES

Allaby, R. G., and Brown, T. A. (2001). Network analysis provides insights into evolution of 5S rDNA arrays in Triticum and Aegilops. *Genetics* 157, 1331–1341. doi: 10.1093/genetics/157.3.1331

Appenroth, K. J., Borisjuk, N., and Lam, E. (2013). Telling duckweed apart: genotyping technologies for the Lemnaceae. *Chin. J. Appl. Environ. Biol.* 19, 1–10. doi: 10.3724/SP.J.1145.2013.00001

Appenroth, K. J., Crawford, D. J., and Les, D. H. (2015). After the genome sequencing of duckweed - how to proceed with research on the fastest growing angiosperm? *Plant Biol.* 17, 1–4. doi: 10.1111/plb.12248

Baker, W. J., Hedderson, T. A., and Dransfield, J. (2000). Molecular phylogenetics of Calamus (Palmae) and related rattan genera based on 5S nrDNA spacer sequence data. *Mol. Phylogenet. Evol.* 14, 218–231. doi: 10.1006/mpev.1999.0697

Baum, B. R., and Feldman, M. (2010). Elimination of 5S DNA unit classes in newly formed allopolyploids of the genera Aegilops and Triticum. *Genome* 53, 430–438. doi: 10.1139/G10-017

Borisjuk, N., Chu, P., Gutierrez, R., Zhang, H., Acosta, K., Friesen, N., et al. (2015). Assessment, validation and deployment strategy of a two-barcode protocol for facile genotyping of duckweed species. *Plant Biol.* 17, 42–49. doi: 10.1111/plb.12229

Borisjuk, N., Davidjuk, I., Kostishin, S., Miroshnichenko, G., Velasko, R., Hemleben, V., et al. (1997). Structural analysis of rDNA in the genus Nicotiana. *Plant Mol. Biol.* 35, 655–660. doi: 10.1023/A:1005856618898

Borisjuk, N., and Hemleben, V. (1993). Nucleotide sequence of the potato rDNA intergenic spacer. *Plant Mol. Biol.* 21, 381–384.

Borisjuk, N. V., Momot, V. P., and Gleba, Y. (1988). Novel class of rDNA repeat units in somatic hybrids between Nicotiana and Atropa. *Theor. Appl. Genet.* 76, 108–112. doi: 10.1007/BF00288839

Borisjuk, N., Peterson, A., Lv, J., Qu, G., Luo, Q., Chen, G., et al. (2018). Structural and biochemical properties of duckweed surface cuticle. *Front. Chem.* 6:317. doi: 10.3389/fchem.2018.00317

Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., et al. (2019). G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495. doi: 10.1093/bioinformatics/btz087

Cao, H. X., Fourounjian, P., and Wang, W. (2018). "The importance and potential of duckweeds as a model and crop plant for biomass-based applications and beyond," in *Handbook of Environmental Materials Management*. ed. C. Hussain (Cham: Springer), 1–16.

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., and Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the *Arabidopsis thaliana* 5S rRNA genes. *Plant J.* 35, 251–261. doi: 10.1046/j.1365-313X.2003.01793.x

Coen, E., Strachan, T., and Dover, G. (1982). Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of drosophila. *J. Mol. Biol.* 158, 17–35. doi: 10.1016/0022-2836(82)90448-X

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Eickbush, T. H., and Eickbush, D. G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175, 477–485. doi: 10.1534/genetics.107.071399

Fulnecek, J., Lim, K. Y., Leitch, A. R., Kovarik, A., and Matyasek, R. (2002). Evolution and structure of 5S rDNA loci in allotetraploid Nicotiana tabacum and its putative parental species. *Heredity* 88, 19–25. doi: 10.1038/sj.hdy.6800001

Gibbons, J. G., Branco, A. T., Godinho, S. A., Yu, S., and Lemos, B. (2015). Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112, 2485–2490. doi: 10.1073/pnas.1416878112

Havlová, K., and Fajkus, J. (2020). G4 structures in control of replication and transcription of rRNA genes. *Front. Plant Sci.* 11:593692. doi: 10.3389/fpls.2020.593692

Hemleben, V., Grierson, D., Borisjuk, N., Volkov, R. A., and Kovarik, A. (2021). Molecular organization, structure, cytogenetics and function of plant ribosomal RNA genes: from precursor-rRNA to molecular evolution (personal perspectives). *Front. Plant Sci.* (in press).

Hemleben, V., and Werts, D. (1988). Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (Vignaradiata and Matthiolaincana). *Gene* 62, 165–169. doi: 10.1016/0378-1119(88)90591-4

Ho, E. K. H., Bartkowska, M., Wright, S. I., and Agrawal, A. F. (2019). Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. *New Phytol.* 224, 1361–1371. doi: 10.1111/nph.16056

Hoang, P. T. N., Fiebig, A., Novák, P., Macas, J., Cao, H. X., Stepanenko, A., et al. (2020). Chromosome-scale genome assembly for the duckweed *Spirodela intermedia*, integrating cytogenetic maps, PacBio and Oxford Nanopore libraries. *Sci. Rep.* 10:19230. doi: 10.1038/s41598-020-75728-9

Hoang, P. N. T., Michael, T. P., Gilbert, S., Chu, P., Motley, S. T., Appenroth, K. J., et al. (2018). Generating a high-confidence reference genome map of the greater duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *Plant J.* 96, 670–684. doi: 10.1111/tpj.14049

Hoang, P. T. N., Schubert, V., Meister, A., Fuchs, J., and Schubert, I. (2019). Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci among duckweeds. *Sci. Rep.* 9:3234. doi: 10.1038/s41598-019-39332-w

Kellogg, E. A., and Appels, R. (1995). Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* 140, 325–343. doi: 10.1093/genetics/140.1.325

Kovarik, A., Dadejova, M., Lim, Y. K., Chase, M. W., Clarkson, J. J., Knapp, S., et al. (2008). Evolution of rDNA in Nicotiana allopolyploids: a potential link between rDNA homogenization and epigenetics. *Ann. Bot.* 101, 815–823. doi: 10.1093/aob/mcn019

Labudová, D., Hon, J., and Lexa, M. (2020). pqsfinder web: G-quadruplex prediction using optimized pqsfinder algorithm. *Bioinformatics* 36, 2584–2586. doi: 10.1093/bioinformatics/btz928

Laird, R. A., and Barks, P. M. (2018). Skimming the surface: duckweed as a model system in ecology and evolution. *Am. J. Bot.* 105, 1962–1966. doi: 10.1002/ajb2.1194

Lam, E., Appenroth, K. J., Ma, Y., Shoham, T., and Sree, K. S. (2020). Registration of duckweed clones/strains-future approach. Duckweed Forum. 8, 35–37. Available at: http://www.ruduckweed.org

Landolt, E. (1986). The family of Lemnaceae - a monographic study. *Ver. Geobot. Inst. ETH StiftRübel.* 71, 1–563.

Les, D. H., and Crawford, D. J. (1999). Landoltia (Lemnaceae), a new genus of duckweeds. *Novon* 9, 530–533. doi: 10.2307/3392157

Les, D. H., and Tippery, N. P. (2013). "In time and with water… the systematics of alismatid monocotyledons," in *Early Events in Monocot Evolution*. eds. P. Wilkin and S. J. Mayo (Cambridge: Cambridge University Press), 118–164.

Lipps, H. J., and Rhodes, D. (2009). G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* 19, 414–422. doi: 10.1016/j.tcb.2009.05.002

Liu, Y., Xu, H., Yu, C., and Zhou, G. (2021). Multifaceted roles of duckweed in aquatic phytoremediation and bioproducts synthesis. *GCB Bioenergy* 13, 70–82. doi: 10.1111/gcbb.12747

Mahelka, V., Kopecky, D., and Baum, B. R. (2013). Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the argonomically important grass *Thinopyrum intermedium* (Triticeae). *Mol. Biol. Evol.* 30, 2065–2086. doi: 10.1093/molbev/mst106

Mahelka, V., Krak, K., Kopecký, D., Fehrer, J., Šafář, J., Bartoš, J., et al. (2017). Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1726–1731. doi: 10.1073/pnas.1613375114

Matyácek, R., Fulnecek, J., Lim, K. Y., Leitch, A. R., and Kovarík, A. (2002). Evolution of 5S rDNA unit arrays in the plant genus Nicotiana (Solanaceae). *Genome* 45, 556–562. doi: 10.1139/g02-017

Michael, T., Bryant, D., Gutierrez, R., Borisjuk, N., Chu, P., Zhang, H., et al. (2017). Comprehensive definitions of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* 89, 617–637. doi: 10.1111/tpj.13400

Miranda, A. F., Kumar, N. R., Spangenberg, G., Subudhi, S., Lal, B., and Mouradov, A. (2020). Aquatic plants, Landoltia punctata, and *Azolla filiculoides* as bio- converters of wastewater to biofuel. *Plants* 9:437. doi: 10.3390/plants9040437

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325.

Pan, Y. B., Burner, D. M., and Legendre, B. L. (2000). An assessment of the phylogenetic relationship among sugarcane and related taxa based on the nucleotide sequence of 5S rRNA intergenic spacers. *Genetica* 108, 285–295. doi: 10.1023/A:1004191625603

Röser, M., Winterfeld, G., Grebenstein, B., and Hemleben, V. (2001). Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). *Mol. Phylogenet. Evol.* 21, 198–217. doi: 10.1006/mpev.2001.1003

Schlögelhofer, P., Nizhynska, V., Feik, N., Chambon, C., Potuschak, T., Wanzenböck, E. M., et al. (2002). The upstream Sal repeat-containing segment of *Arabidopsis thaliana* ribosomal DNA intergenic region (IGR) enhances the activity of adjacent protein-coding genes. *Plant Mol. Biol.* 49, 655–667. doi: 10.1023/a:1015556531074

Sergeeva, E. M., Shcherban, A. B., Adonina, I. G., Nesterov, M. A., Beletsky, A. V., Rakitin, A. L., et al. (2017). Fine organization of genomic regions tagged to the 5S rDNA locus of the bread wheat 5B chromosome. *BMC Plant Biol.* 17(Suppl. 1):183. doi:10.1186/s12870-017-1120-5.

Shen, N., Zhang, H., Qin, Y., Wang, Q., Zhu, J., Li, Y., et al. (2018). Efficient production of succinic acid from duckweed (*Landoltia punctata*) hydrolysate by Actinobacillus succinogenes GXAS137. *Bioresour. Technol.* 250, 35–42. doi: 10.1016/j.biortech.2017.09.208

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., et al. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Res.* 46, 3019–3033. doi: 10.1093/nar/gky163

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the Arabidopsis NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12:387. doi: 10.1038/s41467-020-20728-6

Sree, K. S., and Appenroth, K. J. (2020). "Worldwide genetic resources of duckweed: stock collections," in *The Duckweed Genomes*, *Compendium of plant genomes*. eds. X. Cao, P. Fourounjian and W. Wang (Cham: Springer), 39–46.

Stadler, M., Stelzer, T., Borisjuk, N., Zanke, C., Schilde-Rentschler, L., and Hemleben, V. (1995). Distribution of novel and known repeated elements of Solanum and application for the identification of somatic hybrids among Solanum species. *Theor. Appl. Genet.* 91, 1271–1278. doi: 10.1007/BF00220940

Tao, X., Fang, Y., Huang, M. J., Xiao, Y., Liu, Y., Ma, X. R., et al. (2017). High flavonoid accompanied with high starch accumulation triggered by nutrient starvation in bioenergy crop duckweed (*Landoltia punctata*). *BMC Genomics* 18:166. doi: 10.1186/s12864-017-3559-z

Tippery, N. P., and Les, D. H. (2020). "Tiny plants with enormous potential: phylogeny and evolution of duckweeds," in *The Duckweed Genomes*, *Compendium of plant genomes*. eds. X. H. Cao, P. Fourounjian and W. Wang (Switzerland, AG: Springer Nature), 19–38.

Tippery, N. P., Les, D. H., and Crawford, D. J. (2015). Evaluation of phylogenetic relationships in Lemnaceae using nuclear ribosomal data. *Plant Biol.* 17(Suppl. 1), 50–58. doi: 10.1111/plb.12203

Toyama, T., Hanaoka, T., Tanaka, Y., Morikawa, M., and Mori, K. (2018). Comprehensive evaluation of nitrogen removal rate and biomass, ethanol, and methane production yields by combination of four major duckweeds and three types of wastewater effluent. *Bioresour. Technol.* 250, 464–473. doi: 10.1016/j.biortech.2017.11.054

Vandenberghe, A., Chen, M. W., Dams, E., de Baere, R., de Roeck, E., Huysmans, E., et al. (1984). The corrected nucleotide sequences of 5 S RNAs from six angiosperms. *FEBS Lett.* 171, 17–22. doi: 10.1016/0014-5793(84)80452-4

Venkateswarlu, K., Lee, S. W., and Nazar, R. N. (1991). Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. *Gene* 105, 249–254. doi: 10.1016/0378-1119(91)90158-8

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Syst. Biodivers.* 5, 261–276. doi: 10.1017/S1477200007002447

Volkov, R. A., Panchuk, I. I., Borisjuk, N., Hosiawa-Baranska, M., Maluszynska, J., and Hemleben, V. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol.* 17:21. doi: 10.1186/s12870-017-0978-6

Xu, J., and Shen, G. (2011). Growing duckweed in swine wastewater for nutrient recovery and biomass production. *Bioresour. Technol.* 102, 848–853. doi: 10.1016/j.biortech.2010.09.003

Yadav, V., Hemansi, K. N., Tuteja, N., and Yadav, P. (2017). G Quadruplex in plants: a ubiquitous regulatory element and its biological relevance. *Front. Plant Sci.* 8:1163. doi: 10.3389/fpls.2017.01163

Zanke, C., Borisjuk, N., Ruoss, B., Schilde-Rentschler, L., Ninnemann, H., and Hemleben, V. (1995). A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between Solanum tuberosum and *S. pinnatisectum*. *Theor. Appl. Genet.* 90, 720–726. doi: 10.1007/BF00222139

Zentgraf, U., and Hemleben, V. (1992). Complex-formation of nuclear proteins with the RNA polymerase-I promoter and repeated elements in the external transcribed spacer of Cucumissativus ribosomal DNA. *Nucleic Acids Res.* 20, 3685–3691. doi: 10.1093/nar/20.14.3685

Zhao, Y., Fang, Y., Jin, Y., Huang, J., Bao, S., Fu, T., et al. (2015). Pilot-scale comparison of four duckweed strains from different genera for potential application in nutrient recovery from wastewater and valuable biomass production. *Plant Biol.* 17(Suppl. 1), 82–90. doi:10.1111/plb.12204

Zhou, Y., and Borisjuk, N. (2019). Small aquatic duckweed plants with big potential for production of valuable biomass and wastewater remediation. *Int. J. Environ. Sci. Nat. Res.* 16:555942. doi: 10.19080/IJESNR.2019.16.555942

Zhou, Y., Chen, G., Peterson, A., Zha, X., Cheng, J., Li, S., et al. (2018). Biodiversity of duckweeds in eastern China and their potential for bioremediation of municipal and industrial wastewater. *J. Geosci. Environ. Protect.* 6, 108–116. doi: 10.4236/gep.2018.63010

Ziegler, P., Adelmann, K., Zimmer, S., Schmidt, C., and Appenroth, K. J. (2015) Relative in vitro growth rates of duckweeds (Lemnaceae) - the most rapidly growing higher plants, *Plant Biol.* 17(Suppl. 1), 33–41. doi:10.1111/plb.12184.

Ziegler, P., Sree, K. S., and Appenroth, K. J. (2019). Duckweed biomarkers for identifying toxic water contaminants? *Environ. Sci. Pollut. Res. Int.* 26, 14797–14822. doi: 10.1007/s11356-018-3427-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

REVIEW
published: 26 July 2021
doi: 10.3389/fpls.2021.684626

The Arabidopsis 2′-O-Ribose-Methylation and Pseudouridylation Landscape of rRNA in Comparison to Human and Yeast

Deniz Streit[1] *and Enrico Schleiff*[1,2]*

[1] Department of Biosciences, Molecular Cell Biology of Plants, Goethe University, Frankfurt, Germany, [2] Frankfurt Institute for Advanced Studies (FIAS), Frankfurt, Germany

OPEN ACCESS

Edited by:
Roman A. Volkov, Chernivtsi National University, Ukraine

Reviewed by:
Jonas Barandun, Umeå University, Sweden
Douglas Muench, University of Calgary, Canada
Brigitte Pertschy, University of Graz, Austria

***Correspondence:**
Enrico Schleiff
schleiff@bio.uni-frankfurt.de

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *23 March 2021*
Accepted: *16 June 2021*
Published: *26 July 2021*

Citation:
Streit D and Schleiff E (2021) The Arabidopsis 2′-O-Ribose-Methylation and Pseudouridylation Landscape of rRNA in Comparison to Human and Yeast. Front. Plant Sci. 12:684626. doi: 10.3389/fpls.2021.684626

Eukaryotic ribosome assembly starts in the nucleolus, where the ribosomal DNA (rDNA) is transcribed into the 35S pre-ribosomal RNA (pre-rRNA). More than two-hundred ribosome biogenesis factors (RBFs) and more than two-hundred small nucleolar RNAs (snoRNA) catalyze the processing, folding and modification of the rRNA in *Arabidopsis thaliana*. The initial pre-ribosomal 90S complex is formed already during transcription by association of ribosomal proteins (RPs) and RBFs. In addition, small nucleolar ribonucleoprotein particles (snoRNPs) composed of snoRNAs and RBFs catalyze the two major rRNA modification types, 2′-O-ribose-methylation and pseudouridylation. Besides these two modifications, rRNAs can also undergo base methylations and acetylation. However, the latter two modifications have not yet been systematically explored in plants. The snoRNAs of these snoRNPs serve as targeting factors to direct modifications to specific rRNA regions by antisense elements. Today, hundreds of different sites of modifications in the rRNA have been described for eukaryotic ribosomes in general. While our understanding of the general process of ribosome biogenesis has advanced rapidly, the diversities appearing during plant ribosome biogenesis is beginning to emerge. Today, more than two-hundred RBFs were identified by bioinformatics or biochemical approaches, including several plant specific factors. Similarly, more than two hundred snoRNA were predicted based on RNA sequencing experiments. Here, we discuss the predicted and verified rRNA modification sites and the corresponding identified snoRNAs on the example of the model plant *Arabidopsis thaliana*. Our summary uncovers the plant modification sites in comparison to the human and yeast modification sites.

Keywords: snoRNA, *A. thaliana*, rRNA, ribosome, 2′-O-ribose-methylation, pseudouridylation

INTRODUCTION

Ribosome biogenesis is an essential biochemical process in all existing organisms. The formation of functional ribosomes involves a huge number of different RNAs and proteins, which have to act in a defined order. These factors catalyze various steps during the maturation of ribosomal RNA (rRNA) from the initial precursor including, their folding, modifications of the rRNA and the assembly of ribosomal proteins. For model systems like yeast, the understanding of molecular events during ribosome biogenesis are already well described. For example, comprehensive number of ribosome assembly factors and their functions, in addition to availability of high resolution ribosome structure, paved the way for in-depth analysis of ribosome maturation in yeast (Woolford and Baserga, 2013; Klinge and Woolford, 2019). While for the same processes in plant systems, many aspects are yet to be given a detailed account. For the analysis of the processes in plants, *Arabidopsis thaliana* has become the model plant for the examination of ribosome biogenesis next to crop plants like wheat and rice (Armache et al., 2010b; Hang et al., 2018).

The maturation of 80S ribosomes is coordinated between three different compartments of the cell. It begins with the transcription of the 35S pre-rRNA by RNA-polymerase I in the nucleolus (Tsang et al., 2003; Henras et al., 2008; Missbach et al., 2013; Hellmann, 2020). The 35S pre-rRNA consists of the three rRNAs 18S, 5.8S, and 25S. The 18S and 5.8S rRNA are separated by the internal transcribed spacer 1 (ITS1), 5.8S and 25S rRNA by the internal transcribed spacer 2 (ITS2) and the three maturing rRNAs are additionally flanked by the 5′- and 3′-external transcribed spacers (ETSs); (Tollervey and Kiss, 1997; Lafontaine, 2015; Weis et al., 2015b; **Figure 1**). This precursor is subsequently processed and modified. The maturation of the rRNA is assisted by ribosome biogenesis factors (RBFs) and small nucleolar RNAs (snoRNAs) (**Figure 2A**). Initially, the 90S particle formation is followed by subsequent splitting into pre-40S and pre-60S particles. The maturation of these particles occurs in the nucleolus, nucleoplasm and in the cytosol (**Figure 1**). During the maturation of ribosomal subunits, the precursors of 18S, 5.8S, and 25S rRNA are processed, folded and modified, and the final steps occur in the cytoplasm before final assembly of the 80S ribosomes (Henras et al., 2008; Palm et al., 2019; Sáez-Vásquez and Delseny, 2019). In addition to the rRNAs transcribed on the 35S transcript, the large ribosomal subunit (LSU) contains a 5S rRNA. This rRNA is transcribed independently in the nucleus by RNA-polymerase III (Henras et al., 2008; Missbach et al., 2013; Bassham and MacIntosh, 2017). The 5S rRNA forms the 5S RNP together with the ribosomal proteins L5 and L18, which associates with the 60S pre-ribosomal particle in the nucleoplasm (Leidig et al., 2014; Lafontaine, 2015).

To date, more than 200 RBFs and more than 200 snoRNAs were described to regulate ribosome maturation. In plants, the inventory for both has been established by a combination of experimental evidence and bioinformatics prediction (Barneche et al., 2001; Brown et al., 2001; Qu et al., 2001; Chen C. L. et al., 2003; Chen and Wu, 2009; Kim et al., 2010; Liu et al., 2013; Palm et al., 2016; Azevedo-Favory et al., 2020). Considering the importance of rRNA modifications for proper processing and maturation of rRNA but also for the function of ribosomes, we discuss in the following the current knowledge on the rRNA modifications and snoRNAs in Arabidopsis and compare these with the human and yeast modifications sites.

THE EUKARYOTIC snoRNAs

Alike messenger RNA (mRNA) and transfer RNA (tRNA), rRNA is highly post-transcriptional modified (Sloan et al., 2017). For yeast, it could be shown that the loss of individual modifications within the rRNA is non-essential, while the lack of more than one modification site, especially in important regions of the ribosome has led to alterations in ribosomal processing but also rRNA processing can be affected (Liang et al., 2009; Demirci et al., 2010; Polikanov et al., 2015). Furthermore, different distributions of modification sites can be related to different cell type as reported for human ribosomes, where cancer cells carry a different subset of modifications (Natchiar et al., 2017). Nonetheless, in contrast it could be demonstrated that the loss of a single modification site in zebrafish can have harmful effects during the early development (Higa-Nakamine et al., 2012). However, only a small number of rRNA modification types are known. The two major modifications are 2′-O-ribose-methylation (2′-O-ribose-me), where a methyl group is attached to the 2′ hydroxyl-group of the ribose within nucleosides, and pseudouridylation involving the conversion of uridine to pseudouridine (Zhao and Yu, 2004; Ito et al., 2014). Recent studies in yeast showed, that the acetyltransferase Kre33 acetylates the sites ac^4C1773 and ac^4C1280 of the 18S rRNAs, which are guided by the two orphan C/D box snoRNAs snR4 and snR45 (Ito et al., 2014; Sharma et al., 2017b). Some snoRNAs like the abundant C/D box snoRNAs U3 and U14 are rather involved in pre-rRNA cleavage at the 5′-ETS site and are therefore involved in 18S rRNA production (Brown and Shaw, 1998; Venema and Tollervey, 1999; Brown et al., 2003).

The 2′-O-ribose-me is the most frequently occurring modification within RNA and can be important for RNA degradation. For example, it was observed that miRNA and siRNAs lacking 2′-O-ribose-me on the 3′ terminal ribose are exposed to degradation (Zhao et al., 2012). In addition, 2′-O-ribose-me defines local secondary structures (Filippova et al., 2017). Likewise, the isomerization of uridine to pseudouridine confers stability of hairpins by base stacking (Desaulniers et al., 2008; Filippova et al., 2017). Modifications often occur in functionally relevant areas of the ribosomes such as A, P, and E sites, the peptidyl transferase center (PTC) and the intersubunit bridge (Decatur and Fournier, 2002; Watkins and Bohnsack, 2012). Remarkably, RNA and especially rRNA modifications appear to be altered during development in addition to environmental changes, which could indicate ribosome heterogeneity (Sloan et al., 2017).

Small nucleolar RNAs are small RNA molecules essential for the regulation and guidance of the post-transcriptional modifications of rRNA, tRNA, and snRNAs (Kiss, 2001; Kruszka et al., 2003; Chow et al., 2007). SnoRNAs exist in eukaryotes and archaea, but not in bacteria (Terns and Terns, 2002;

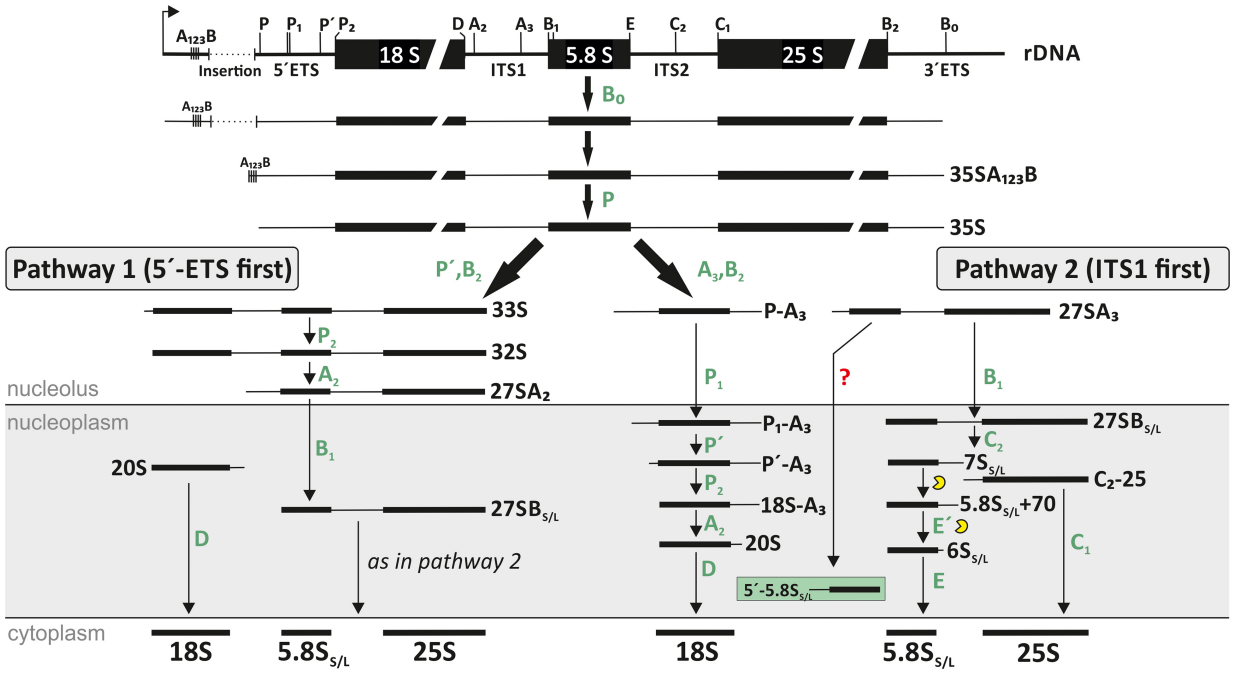


FIGURE 1 | Ribosomal RNA (rRNA) processing pathway in *A. thaliana*. After transcription the 35S pre-rRNA is formed. After cleavage at site P, the further processing depends on the next cleavage site. While for pathway 1 the next cleavage occurs in the 5′-ETS (similar to the main pathway in yeast), the second and majorly used pathway is characterized by a first cleavage in the ITS1 region (similar to the human pathway). The location of each existing precursor of Arabidopsis is shown. However, the exact localization of 27S and 18S precursors are not fully analyzed yet. The figure was modified from Weis et al. (2015a).

Bhattacharya et al., 2016). Because of their importance for rRNA folding and modification, they are often localized in the nucleus where processing and modification of rRNA takes place (Kiss-László et al., 1998). They are re-localized to the cytoplasm in response to stress, which has so far only been described in yeast (Holley et al., 2015). Whether this holds true for eukaryotes in general needs to be elucidated.

The sizes of snoRNAs vary between 60 and 300 nucleotides (nt; Falaleeva et al., 2017) and they are mostly transcribed by RNA-polymerase II (Qu et al., 2015). Nevertheless, in cases of the U3 gene in plants and in dicistronic tRNA-snoRNA genes, RNA-Polymerase III is responsible for the transcription (Dieci et al., 2009). The snoRNA gene organization varies between organisms. Most snoRNAs in yeast are independently encoded, and only a minority is localized in intronic regions or in cluster with other snoRNA genes (Brown et al., 2003; **Figure 2B**). The majority of snoRNAs in humans are organized in intronic regions, and only few are encoded as independent genes (**Figure 2B**). In plant genomes snoRNAs are encoded either independently, in intronic regions or in intronic gene clusters as shown in rice (**Figure 2B**; Leader et al., 1997; Brown et al., 2003). In addition, snoRNAs are also organized in dicistronic tRNA-snoRNAs or snoRNA-miRNA clusters as described in *A. thaliana* and rice (Kruszka et al., 2003; Qu et al., 2015).

The snoRNAs are classified by the existence of conserved sequence motifs (Bachellerie and Cavaillé, 1997; Weinstein and Steitz, 1999). The so called C/D box and H/ACA box snoRNAs form the two major classes, while some minor classes have been identified as well. The C/D box snoRNAs are characterized by a "C box" with a consensus sequence RUGAUGA (R stands for any purine) and a "D box" (consensus sequence: CUGA) (**Figure 2A**). Frequently, these snoRNAs contain additional, less conserved boxes annotated as C' and D'. The conserved C and D boxes fulfill a multitude of functions and are amongst necessary for the snoRNA import into the nucleolus (Samarsky et al., 1998; Newman et al., 2000; Bertrand and Fournier, 2013). In contrast, binding to the rRNA target region is accomplished by one or two antisense elements of about 10–21 nt positioned upstream of the D or D'-boxes. In most cases, the complement to the fifth nucleotide of this element is modified in the rRNA (Barneche et al., 2001; Kiss, 2001; Kruszka et al., 2003). The secondary structure of C/D box snoRNA is characterized by a K-turn motif that brings the C and D box (C' and D') in proximity through stem loop formation, and by guide elements carrying the antisense sequences (Matera et al., 2007). The C/D box snoRNAs are components of small nucleolar ribonucleoprotein particles (snoRNPs) that in addition consists of described four core proteins fibrillarin (methyltransferase; Nop1p in yeast), NOP58 (Nop58p in yeast), NOP56 (Nop56p in yeast), and 15.5K (Snu13p in yeast) (Rodor et al., 2011; Huang et al., 2016; **Figure 2A**).

The hinge box (H-box: sequence ANANNA, N stands for any nucleotide) and the 3′ terminal located ACA box characterize the H/ACA box snoRNA family (Brown and Shaw, 1998; Kiss, 2001). The H and ACA boxes are required for nucleolar import, for example (Bertrand and Fournier, 2013). H/ACA snoRNPs form a hairpin-hinge-hairpin-tail structure with the

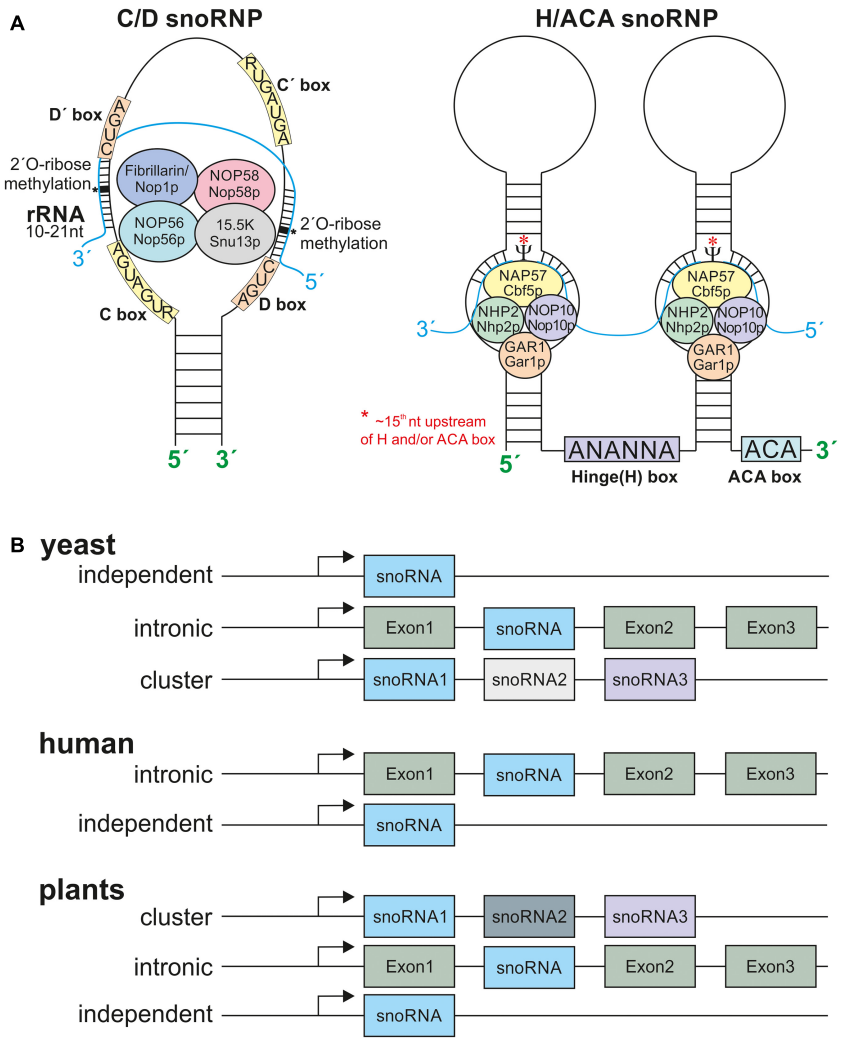


FIGURE 2 | The snoRNPs and snoRNA localization. **(A)** Depicted are the C/D box snoRNPs and H/ACA snoRNPs. The C/D box snoRNPs contain the conserved C (RUGAUGA) and D (CUGA) boxes and the two less conserved C' and D'-boxes. The C/D box snoRNPs are composed of the methyltransferase fibrillarin/Nop1p, NOP56/Nop56p, Nop58/Nop58p, and 15.5K/Snu13p. The rRNA has a 10–21 nt long complementary site to the according snoRNA. The 2′-O-ribose-methylation takes place 5 nt downstream of the D or D' box (asterisks). The H/ACA box snoRNPs are composed of at least one stem loop. The Hinge Box (ANANNA) is located between two stem loops and the ACA box at the 3′ end of the snoRNA. The pseudouridine synthase NAP57/Cbf5p is modifying approx. the 15th nucleotide (asterisks) of the rRNA upstream of the H and or ACA box, further the proteins NHP2/Nhp2p, Nop10/Nop10p, and GAR1/Gar1p are required. **(B)** The snoRNA gene organization in different eukaryotes. Yeast snoRNAs are mainly localized in independent regions and lesser in intronic and polycistronic (cluster) regions. The human snoRNA gene organization is mostly intronic with few independent genes and plant snoRNAs are mostly located within clusters of many snoRNAs. Only very few examples for intronic and independent gene organization are known. For references see main text.

tail and the hinge region being single stranded (Dragon et al., 2006). The nucleotide to be modified is positioned about 15 nt upstream of the ACA or hinge motif, respectively (Lindsay et al., 2013; **Figure 2A**). Alike the C/D box snoRNAs, H/ACA box snoRNAs are components of snoRNPs. However, the known snoRNPs containing an H/ACA box snoRNA consist of the proteins dyskerin/NAP57 (pseudouridine synthase; Cbf5p in yeast), NHP2 (Nhp2p), NOP10 (Nop10p), and GAR1 (Gar1p) (Rodor et al., 2011; **Figure 2A**).

Another minor class of snoRNAs unifies the mitochondrial RNA processing (MRP)-RNAs, a snoRNA family which lacks conserved boxes, but harbors rRNA processing activity (Bertrand and Fournier, 2013). Additional snoRNAs without typical motifs are deposited in plant snoRNA databases as well (Yoshihama et al., 2013).

THE snoRNAs IN PLANTS

Since the first discoveries of snoRNAs in 1970's (Reddy et al., 1974, 1979) different studies targeted the identification of plant snoRNAs by experimental approaches. Early on, the snoRNAs U3, U14, and U49 were identified in plants based on similarity to the snoRNAs of yeast and vertebrates (Kiss et al., 1991;

Leader et al., 1994, 1997). Remarkably, U3 in plants is transcribed by the RNA polymerase III and possess a different capping than found for U3 in yeast or human (Kiss et al., 1991). By dot-matrix analysis of Fib1 and Fib2, the plant-specific snoRNAs U60.1f and U60.2f were discovered (Barneche et al., 2000).

After the release of the *A. thaliana* genome (Kaul et al., 2000) snoRNAs were identified by computational strategies searching for C/D box characteristics, rRNA complementarities or other structural attributes (Zhou et al., 2000; Barneche et al., 2001; Brown et al., 2001; Qu et al., 2001). The next boost for the discovery of plant snoRNAs came by RNomics on either total RNA from different tissues or from the nucleolar RNA of *A. thaliana* (Marker et al., 2002; Kim et al., 2010; Streit et al., 2020) and by re-analysis of existing small RNA datasets of different *A. thaliana* tissues and growth stages (Chen and Wu, 2009). This analysis was initially focused on Arabidopsis and was then extended to *Oryza sativa* (Chen C. L. et al., 2003; Liu et al., 2013). Today, 10,654 different H/ACA box snoRNA genes and 6064 different C/D box snoRNA genes are deposited in the database snOPY (Yoshihama et al., 2013).

In contrast to globally discovered snoRNAs, only a single plant snoRNA is functionally characterized. The C/D box type snoRNA HIDDEN TREASURE 2 (HID2) associates with 45S pre-rRNA but is not relevant for 2′-ribose methylation at position G2620 as this modification was not altered in an according mutant (Zhu et al., 2016). It was speculated that other snoRNAs might complement for HID2 function (Zhu et al., 2016), which needs to be verified. Thus, the analysis of the snoRNA function and the complementarity of the different snoRNA genes of the different families will be a major target of future research.

THE rRNA MODIFICATION IN PLANTS

There are two major types of rRNA modifications, namely 2′-O-ribose-methylation and pseudouridylation. However, for yeast and human rRNAs, base methylations were additionally described (e.g., for yeast dimethylase Dim1p) (Lafontaine et al., 1995). Furthermore, yeast rRNA was also found to be acetylated (Ito et al., 2014; Sharma et al., 2017b). However, the latter two modification types have not been described so far in plants (Piekna-Przybylska et al., 2007).

Initially, the analysis of the individual snoRNAs was accompanied by the analysis of the rRNA modification sites, e.g., by primer extension analysis (Barneche et al., 2000). Recently, genome-wide pseudouridine sequencing verified predicted pseudouridine modifications in cytosolic and plastidic ribosomes (Sun et al., 2019). At the same time, this approach led to the discovery of yet unknown modification sites as well (Sun et al., 2019). Remarkably, the ITS1 separating the 18S rRNA from 5.8S is modified as well (Sun et al., 2019). However, future studies are required to explore whether this is a unique modification or whether ITS1, ITS2 and the ETS regions are generally modified, and to understand the role of modifications of the pre-rRNA. A complementary analysis using RiboMethSeq for detection of 2′-O-ribose-me modifications yielded novel modification sites as well (Azevedo-Favory et al., 2020).

Different approaches exploring the modifications of the rRNA in Arabidopsis yielded a total of 321 rRNA modification sites (Barneche et al., 2001; Qu et al., 2001; Chen and Wu, 2009; Kim et al., 2010; Sun et al., 2019; Azevedo-Favory et al., 2020; Streit et al., 2020). A total of 79 2′-O-ribose-me and 43 pseudouridylation sites were assigned for the 18S rRNA, of which 44 2′-O-ribose-me and 28 pseudouridylation sites were experimentally confirmed. For 25S rRNA, 132 2′-O-ribose-me and 64 pseudouridylation sites are proposed, of which 86 and 51, respectively, are experimentally confirmed. For 5.8S rRNA, three sites carrying 2′-O-ribose-me were predicted due to antisense elements found in three snoRNAs of which two are experimentally confirmed. Accordingly, a recent study confirmed the predicted U22 pseudouridylation and mapped a new site at U78 (Sun et al., 2019).

It has to be considered that the existing discrepancy between detected and predicted modification sites might result from a variability of the modification pattern in ribosomes of one cell, in different tissues, at different developmental stages or in response to environmental changes as discovered for other species named ribosome heterogeneity (Sloan et al., 2017). However, as the ribosome turn-over is comparatively slow, alterations in rRNA modifications are considered to be more meaningful for long-term changes (Ferretti and Karbstein, 2019). Although the final annotation and confirmation of predicted sites requires further research, in here the predicted sites are discussed as well. In the following sections, the positioning of the modifications in the rRNA, for selected modifications the function and the required snoRNAs are discussed.

A VIEW ON THE rRNA MODIFICATION SITES AND snoRNAs IN *Arabidopsis thaliana*

A View on the Modifications in Plant 5.8S rRNA

For the 5.8S rRNA, three 2′-O-ribose-me sites were predicted (**Figure 3** and **Supplementary Table 1**), of which two sites were mapped by primer extension (Barneche et al., 2001; Brown et al., 2001; Qu et al., 2001). Additionally, two pseudouridylation sites in 5.8S could be mapped as well (Sun et al., 2019). Thus, the modification of the 5.8S rRNA in *A. thaliana* is more similar to human with four modifications (two 2′-O-ribose-me and two pseudouridylation sites) than to yeast, where only a single Ψ-site at U73 is known so far (Piekna-Przybylska et al., 2007). However, the Ψ-site in yeast exists in *A. thaliana* at position Ψ78, although this Ψ-site was found at the adjacent uracil in Arabidopsis (**Figure 3**). The H/ACA box snoRNA snR43 targeting this site in yeast (Piekna-Przybylska et al., 2007) could not be identified in *A. thaliana*.

The sites Am47, Gm79, and Ψ78 are localized in the 5.8S secondary structure, which is formed by three bulges between the helices 5, 6, and 7 (**Figure 3** and **Supplementary Table 1**). Worth mentioning, the 2′-O-ribose-me at Gm79 in *A. thaliana* (**Figure 3**) represents the Gm75 modification site in

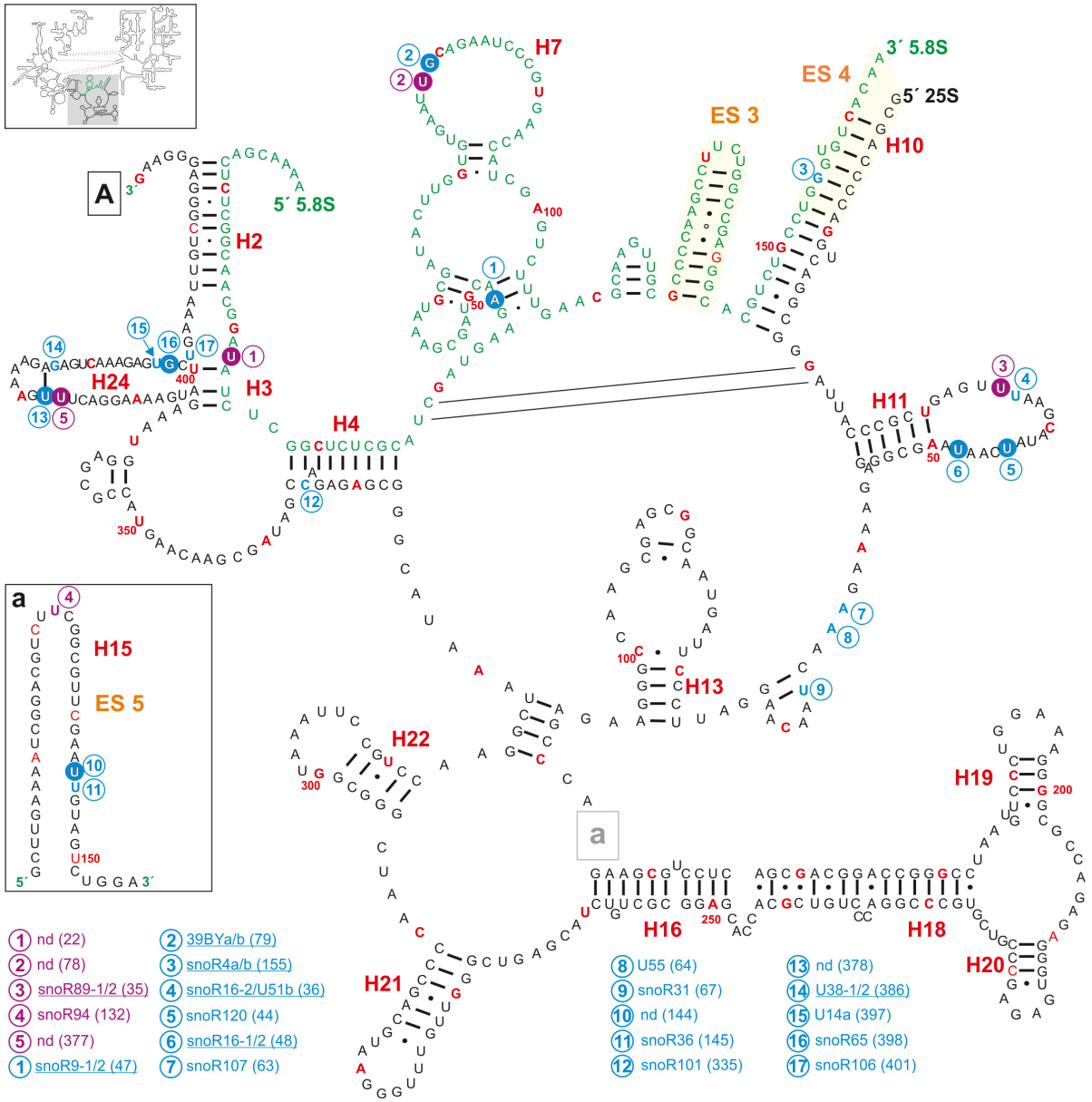


FIGURE 3 | Secondary structure diagram of the 5.8S rRNA and domain I of 25S rRNA of *A. thaliana*. There is no experimental determined secondary structure map of the rRNA of *A. thaliana*. Hence, the secondary map of 5.8S (letters in green) and 25S (letters in black) was created based on the RNAcentral database (The RNAcentral Consortium, 2019) and http://www.rna.icmb.utexas.edu. The positions for predicted (blue letter) and verified 2′-O-ribose-me (white letter in blue circle) sites and the predicted (violet letter) and verified positions for pseudouridylation (white letter in violet circle) are shown. If several snoRNAs are annotated to target the same site, the name is underlined. Analyses were conducted by using the snoRNA databases snOPY (Yoshihama et al., 2013) and the plant snoRNA DB (Brown et al., 2003). Predicted and verified positions for 2′-O-ribose-methylations and pseudouridylations were obtained from Barneche et al. (2001), Brown et al. (2001), Qu et al. (2001), Sun et al. (2019), Azevedo-Favory et al. (2020). Every tenth nucleotide is marked in red and every 50th nucleotide is labeled with the according number. Framed small letters indicate the position of the structures shown separately and framed large letters indicate the position of the connections in subsequent images (for A see **Figure 4**). The sequence used for the secondary structure map refers to the sequence of 5.8S and 25S of chromosome 2. The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 60S rRNA secondary structure highlighting the according region is enclosed.

humans (Piekna-Przybylska et al., 2007). Remarkably, in human ribosomes the region which includes helices 5, 6, and 7 is sandwiched between uL26, L35/uL29, L37, and eL39 (Khatter et al., 2015) and the structure changes between the mRNA free and the mRNA bound state (Graifer et al., 2005). A similar structure was obtained in plant ribosomes, where the helices 5, 6, and 7 are sandwiched by L24, L29, L37e, and L39e (Armache et al., 2010a). Moreover, L29 was identified as one of the ribosomal

proteins with diurnal alteration of the phosphorylation state in Arabidopsis (Turkina et al., 2011). This suggest that the modification in this region of the 5.8S rRNA might be important for the ribosomal activity in translation (Gulay et al., 2017) or for the ribosomal translation elongation (translocation), which was found in cell-free extracts to be under the regulation of the 5.8S rRNA as well (Elela and Nazar, 1997).

In addition, one modification is found in the bulge between helix 2 and helix 3, and one in helix 10. All three helices are formed by base pairing between 5.8S and 25S rRNA (**Figure 3**) and are deeply buried in the ribosomal structure in the human ribosomes (Khatter et al., 2015). Thus, it is likely that the modifications are required for stabilizing the structure of the ribosomes. Interestingly, the predicted 2′-ribose-O-me site at position Gm155, which hypothetically is targeted by snoR4a/4b, could not be confirmed by radiographic labeling of modified nucleotides in wheat-embryo (Lau et al., 1974), suggesting that this snoRNA is probably not involved in the modification but rather in rRNA processing in the ITS2 region (Brown et al., 2001). However, it is known that certain modifications of the eukaryotic 5.8S are tissue specific (Nazar et al., 1975). Hence, it remains possible that modifications like Gm155 are only present in selected tissues or in developmental manner.

Furthermore, modifications such as 2′-O-ribose-methylations at Um14 in rat liver appeared to be present in a higher degree in the cytoplasmic fraction than in nuclear fractions (Nazar et al., 1980). However, the cellular distribution of the rRNA modifications in plant cells was not experimentally approached so far.

A View on the Modifications in Arabidopsis 25S rRNA

The 25S rRNA is the largest RNA within ribosomes and thus it contains numerous modifications. For better discussion, the 25S rRNA is dissected in here into five domains along the rRNA sequence. The 5′ region (bp 1–660) is assigned as domain I (**Figures 3**, **4**), bp 660–1440 as domain II (**Figure 5**), bp 1440–1870 as domain III (**Figure 4**), bp 1870–2370 as domain IV (**Figure 6**) and the 3′ region (bp 2370–3375) is assigned as domain V/VI (**Figure 7**; Paci and Fox, 2015). The helical domains (H), the expansion segments (ES), and pivoting regions (PR) are in part numbered according to previous annotations (Taylor et al., 2009; Paci and Fox, 2015). Expansion segments are additional rRNA parts in eukaryotic rRNA compared to the prokaryotic rRNA. Though expansion segments can vary in their sequence and length but are rather conserved in their overall secondary structure (Ramesh and Woolford, 2016).

Domain I of the 25S rRNA contains five mapped and additionally nine predicted 2′-O-ribose-me sites (**Figures 3**, **4** and **Supplementary Table 2**), as well as two mapped and one predicted pseudouridylation site(s). The methylations can be found in helix 11, 15, and 24. Among others, helices 7, 18, 19, 20, and 24 surround the exit tunnel for the nascent polypeptide chain (Spahn et al., 2001). However, only in helix 24 two 2′-O-ribose-me sites and one pseudouridylation site were mapped (**Figure 3**). Accordingly, in human 28S rRNA, helix 24 carries a modification at site Am389 and Am391 (Sharma et al., 2017a), while no modification was found in domain I in yeast (Yang et al., 2016). In fact, signal recognition particles (SRPs) (Halic et al., 2004), which recognize specific sequences of nascent polypeptide chains from the translating ribosomes interact with the tip of helix 24 of the 25/28S rRNA (Beckmann et al., 2001). In turn one could conclude that human and plant ribosomes have evolved a similar mode of binding to such molecular mechanisms. More intriguingly, the presence or absence of modified nucleotides in this region could be used as a complex regulator for the interaction of such particles with the ribosomes of humans and plants. Furthermore, snoRNAs for helix 24 modification were identified for humans, but not for Arabidopsis (**Supplementary Table 2**). Hence, a stand-alone enzyme might be responsible for these modifications that could act in the cytoplasm of Arabidopsis.

Helix 11 of the 25S rRNA in Arabidopsis contains three positions with modifications (two 2′-O-ribose-me and one pseudouridylation site; **Figure 3**). However, this helix does not carry modifications in yeast or human (Piekna-Przybylska et al., 2007). Thus, it is highly likely that this helix has in plants or at least in Arabidopsis a special function within the 60S ribosomal subunits or even in the 80S ribosomes, which requires such modification.

The region annotated as expansion segment 7 (ES7) contains only one described modification site at Am660 which is targeted by the snoRNAs U18-1 and 2 (**Figure 4** and **Supplementary Table 2**). The ES7 is known be localized at the ribosome surface and belongs to the largest expansion segments with the highest variability in eukaryotes (Ramos et al., 2016). It was found that proteins binding to ES7 were relevant for regulations upon environmental changes, 60S subunit biogenesis and transcription elongation (Ramos et al., 2016). In contrast, yeast and human ES7 of the 25/28S is substantially greater than in plants (Parker et al., 2018). Together with the fact that Arabidopsis ES7 carries a 2′-O-ribose-me it can be concluded that plants evolved a special way of regulating those important features during stress conditions as well as in other regulatory functions.

Domain II of Arabidopsis 25S rRNA carries 29 putative 2′-O-ribose-me sites of which 14 could be successfully mapped. For pseudouridylation, 16 sites were predicted of which 13 were mapped (**Figure 5** and **Supplementary Table 2**). Domain II contains the GTPase center mainly composed of helix 43 and helix 44 (**Figure 5**), which is highly conserved in all ribosomes (Ryan and Draper, 1991). In *Escherichia coli*, this region including the ribosomal proteins L10 and L11 is involved in the regulation of GTP hydrolysis by the elongation factor G and TU (Egebjerg et al., 1990; Briones et al., 1998). The rRNA of the GTPase center in *A. thaliana* contains three mapped and two predicted modification sites, for which associated snoRNAs are assigned (**Figure 5** and **Supplementary Table 2**). Interestingly the modifications seem to be unique for *A. thaliana* since this segment is not modified in human or yeast (Piekna-Przybylska et al., 2007).

In general, many of the mapped modification sites in domain II are localized in stem structures (**Figure 3**). The helices 27, 31, 32, 35, and 38 carry many modifications. In yeast and human, helix 38 is exceedingly pseudouridylated but

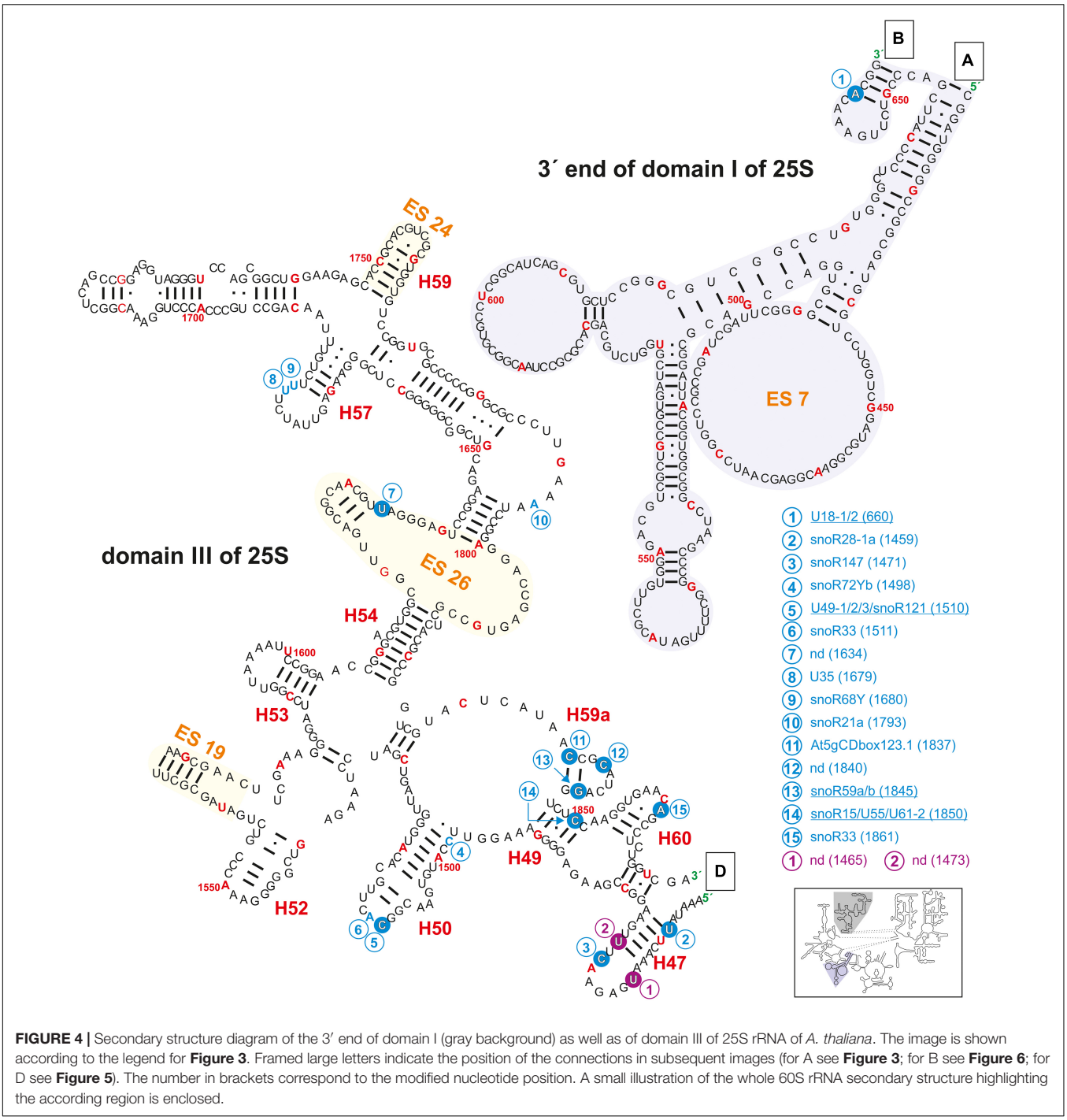


FIGURE 4 | Secondary structure diagram of the 3′ end of domain I (gray background) as well as of domain III of 25S rRNA of *A. thaliana*. The image is shown according to the legend for **Figure 3**. Framed large letters indicate the position of the connections in subsequent images (for A see **Figure 3**; for B see **Figure 6**; for D see **Figure 5**). The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 60S rRNA secondary structure highlighting the according region is enclosed.

does not contain any 2′-O-ribose-me modification (Piekna-Przybylska et al., 2007). In Arabidopsis, pseudouridylations and 2′-O-ribose-me modifications were predicted in this particular helix based on the detection of according snoRNAs. Moreover, three pseudouridylation sites and one 2′-O-ribose-me were experimentally confirmed in helix 38 of the rRNA of Arabidopsis. Nevertheless, for the pseudouridylation site at ψ1060 an according snoRNA could not be identified so far. Intriguingly, helix 38 is involved in the formation of the intersubunit bridge between the 60S and 40S subunit by interacting with S19p of the 40S particle, and it is contacting the A-site bound tRNA in yeast (Spahn et al., 2001). However, in comparison to the yeast and human ribosomes, it can be proposed that the modifications in this helix are involved in the structural stabilization of this important subunit-subunit interaction site in Arabidopsis (Karijolich et al., 2010; Sloan et al., 2017).

Further, helix 35 carries one mapped 2′-O-ribose-me site and two pseudouridylation sites at U892 and U899. For the latter

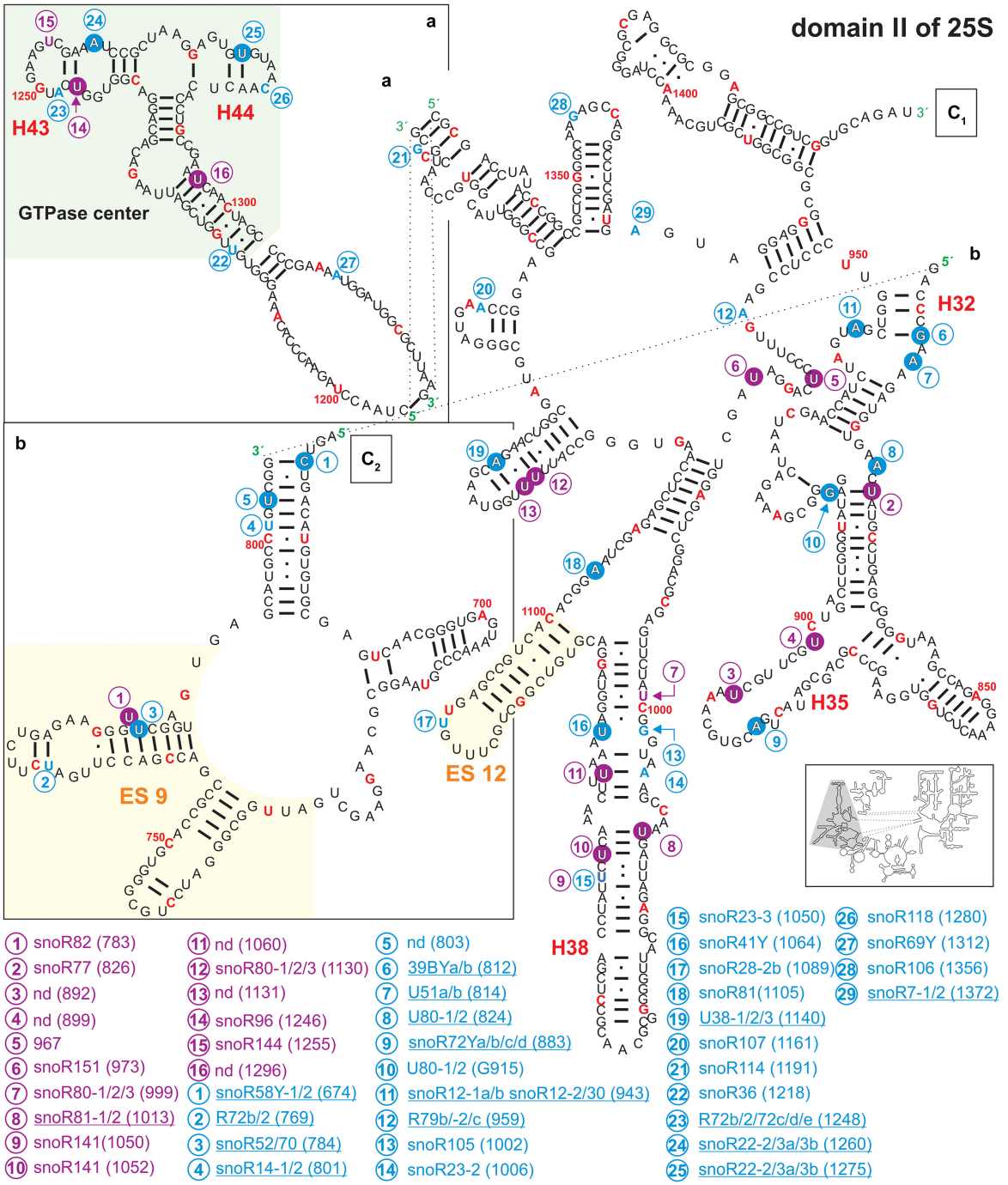


FIGURE 5 | Secondary structure diagram of domain III of 25S rRNA of *A. thaliana*. The image is shown according to the legend for **Figure 3**. Framed large letters indicate the position of the connections in subsequent images (for C see **Figure 6**). PE annotates a pivoting element previously identified (Paci and Fox, 2015). The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 60S rRNA secondary structure highlighting the according region is enclosed.

two sites guiding snoRNAs were not discovered (**Figure 4** and **Supplementary Table 2**), leading to the assumption that stand-alone enzymes may be responsible. In yeast, helix 35 carries two 2′-O-ribose-me sites and in humans one pseudouridylation site (Piekna-Przybylska et al., 2007). A second intersubunit bridge is formed by helix 34 with the 40S subunit. Thus, also in this case the modifications of helix 35 in plants are likely involved in the stabilization of the neighboring structural element.

In domain III, a high density of modifications is present in the region of helix 47, 50, 59a, and 60 (**Figure 4**). In yeast, it is

assumed that Nop4 is binding to helices 47, 32, 26, 33 but also to helix 60 bringing domain II and III in proximity (Granneman et al., 2011). However, helix 47 in yeast carries no modifications, while Arabidopsis helix 47 is highly modified. It can be speculated that these modifications are required for proper processing of 25S precursors like 27SB or 27S-A_2/27S-A_3.

Domains IV (**Figure 6**) and V (**Figure 7**) of the 25S rRNA contain the highest degree of modifications. In both domains a total of 66 sites for 2′-O-ribose-me sites are annotated from which 53 could be successfully confirmed. However, for 13 sites no snoRNA could be identified (**Figures 6**, **7** and **Supplementary Table 2**). In contrast, 31 pseudouridylation sites were mapped, while seven sites could not be verified yet. For 13 of the mapped sites, the associated snoRNA is not known (**Figures 6**, **7** and **Supplementary Table 2**).

The secondary structure map of the core region of 25S rRNA of *A. thaliana* points to a high density of modifications surrounding the PTC (**Figure 7**), which parallels findings for other organisms (Decatur and Fournier, 2002). The PTC is required for the peptide bond formation and peptide release (Lilley, 2001; Polacek and Mankin, 2005; Torres de Farias et al., 2017). In yeast, defective rRNA modifications in this region lead to increased sensitivity to translational inhibitors or changes in translational fidelity (Baxter-Roshek et al., 2007). In Arabidopsis, especially the helices H73, H74, H75, H88, H89, H90, H91, H92, and H93 contain the highest density of modifications (**Figure 7**). Interestingly, Arabidopsis contains the highest number of modification sites (34) in these particular helices in contrast to yeast (16) and human (22). This leads to the conclusion that these modifications are of prime importance for the stability of the PTC structure. Nevertheless, Arabidopsis 60S subunit seems to be closer related to the human 60S regarding the high density of modifications. In human ribosomes, helix 74 is important for the accurate structure of the nascent polypeptide exit tunnel (NPET) (Wilson et al., 2020) and helix 93 is a contact site for hydroxylated uL2, which induces structural rearrangements in the PTC of the mature ribosomes (Yanshina et al., 2015). The same could hold true for plants as well. The tip of helix 89 interacts with the GTPase-associated center which might depend on the modifications (**Figure 5**; Sergiev et al., 2005; Baxter-Roshek et al., 2007) and the modifications in helix 92 were found to be necessary for the correct folding of helix 90–92 in yeast (Baxter-Roshek et al., 2007).

In domain IV, Helix 68, 69, and 71 (**Figure 6**), are involved in the inter-subunit bridge formation between the 40S and 60S (Spahn et al., 2001; Gigova et al., 2014). Helix 68 contains three mapped methylation and one pseudouridylation sites (**Figure 6**). One 2′-O-ribose-me site (Am 2210) is conserved in yeast (Am2220) and human (Am3703; Piekna-Przybylska et al., 2007; **Supplementary Table 2**). The yeast helix 68 contains two E-sites (exit sites), which most probably exist in *A. thaliana* as well (Xie et al., 2012).

Helix 69 is highly modified with two mapped pseudouridylation sites and one mapped 2′-O-ribose-me site in *A. thaliana* (**Figure 6**). This helix interacts with the tRNAs located in the A and P-site, respectively (Ge and Yu, 2013). Similarly, a cluster of modifications is localized in helix 69 in the yeast rRNA, and their deletions led to e.g., severe growth phenotypes and a lower translational rate (Liang et al., 2007). Helix 71 contains two mapped 2′-O-ribose-me sites at Cm2283 and Gm2278 (**Figure 6** and **Supplementary Table 2**). The site Cm2283 was newly identified but the snoRNA targeting this region was not found (Azevedo-Favory et al., 2020). However, while the human rRNA is lacking this modification, it is conserved between yeast and plants (Piekna-Przybylska et al., 2007).

Interestingly, the enigmatic exceptionally large expansion segment 27 (ES27, **Figure 6**) was recently unveiled as essential for translational fidelity, in which it seems to regulate amino acid incorporation and by that prevents frameshift errors (Fujii et al., 2018). Furthermore, it was found that this very flexible region of the eukaryotic ribosomes serves as a scaffold for the conserved enzyme methionine amino peptidase (MetAP), which is required to remove co-translationally the first methionine from the nascent polypeptide chain (Fujii et al., 2018; Knorr et al., 2019). Just recently one pseudouridylation site (Ψ2028) without known snoRNA was found in ES27 of Arabidopsis (Sun et al., 2019).

The P-loop in helix 80 and the A-loop in helix 92 are direct pairing sites for A- and P-site tRNA (Kim and Green, 1999). While the P-loop contains one confirmed 2′-O-ribose-methylation site, the A-loop contains two mapped 2′-O-ribose-me sites with the site Gm2912 having a known snoRNA targeting this region. Moreover, the A-loop contains a pseudouridylation site (**Figure 7**).

Domain VI containing the 3′-end of 25S rRNA from nucleotide 2986 to 3375 contains the conserved sarcin/ricin loop (S/R-Loop; **Figure 7**). This loop is the site of attack of the two toxins α-sarcin, which is a ribonuclease produced by a fungus and ricin, which is an RNA N-glycosylase synthesized by plants (Endo et al., 1988; Macbeth and Wool, 1999). The attack inhibits proper binding of the elongation factors, and thus, translation is blocked (Szewczak and Moore, 1995). In human 60S subunits the S/L-Loop shows a high degree of modifications in comparison to plants or yeast (**Figure 7**; Piekna-Przybylska et al., 2007).

Domain VI harbors the lowest degree of modifications with two mapped 2′-O-ribose-me sites with associated snoRNAs and two mapped pseudouridylations sites in Arabidopsis. The two 2′-O-me sites are within helix 100, one pseudouridylation sites in H97 and one in H98 of ES39 (**Figure 7**). ES39 is exposed to the ribosome surface, the exact function remains elusive, however due to its presence in all eukaryotes it is obvious that eukaryotic ribosomes require this segment (Nygård et al., 2006). For the two pseudouridylation sites U3177 and U3100 a snoRNA is not known so far. Interestingly, the Arabidopsis U3100 is conserved in the human 28S rRNA (U4659), while yeast has not even one modification regarding this specific region (Piekna-Przybylska et al., 2007; **Supplementary Table 2**).

A View on the Modifications in Plant 18S rRNA

The *A. thaliana* genome encodes for two different 18S rRNA variants. While the 18S gene on chromosome 3 has a size of 1808 nt, the copies on chromosomes 2 and 4 contain 1804 nt.

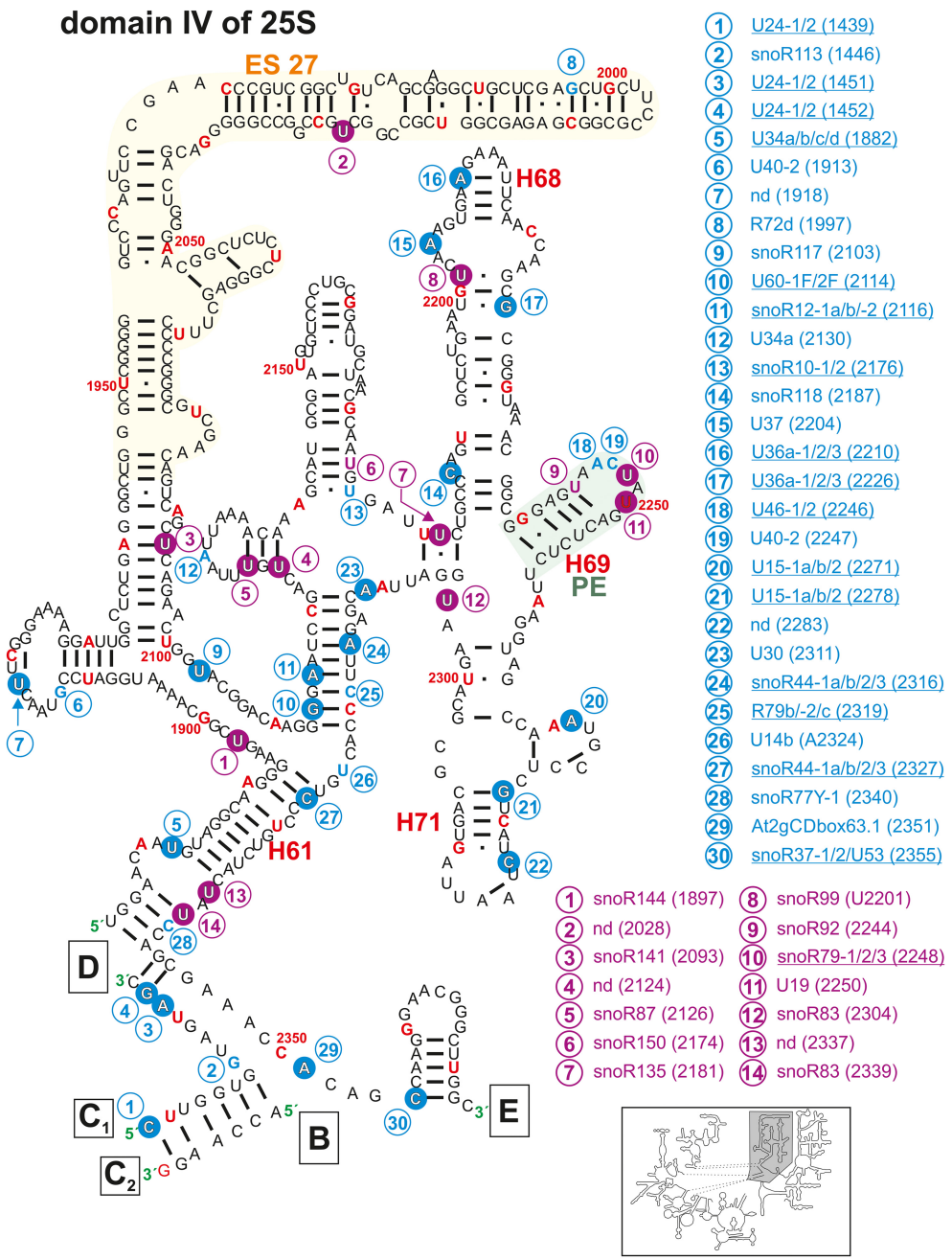


FIGURE 6 | Secondary structure diagram of domain IV of 25S rRNA of *A. thaliana*. The image is shown according to the legend for **Figure 3**. Framed large letters indicate the position of the connections in subsequent images (for B see **Figure 4**; for C see **Figure 5**; for D see **Figure 4**; for E see **Figure 7**). The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 60S rRNA secondary structure highlighting the according region is enclosed.

The secondary structure model of 18S in here refers to the gene in chromosomes 2 and 4, respectively (**Figures 8**, **9**). The SSU binds the mRNA to decode the genetic information in the "decoding center" (Schluenzen et al., 2000). For the 18S rRNA in total 79 sites are predicted to be 2′-O-ribose methylated, of which 44 are experimentally verified (**Supplementary Table 3**). Similarly, from 64 predicted pseudouridylation sites 28 were experimentally confirmed.

It is proposed that the decoding center within the SSU consists of the helices 18, 24, 31, 34, and 44 (Liang et al., 2009), which harbor many modifications in yeast. Helix 24 carries the same Ψ-modification in yeast, human and *A. thaliana*. The modifications in helix 18 vary between two 2′-O-ribose-me and one Ψ-site in human, one 2′-O-ribose-me site in yeast and just one Ψ-site in *A. thaliana* (**Figure 8**), while no methylation site exists in *A. thaliana*.

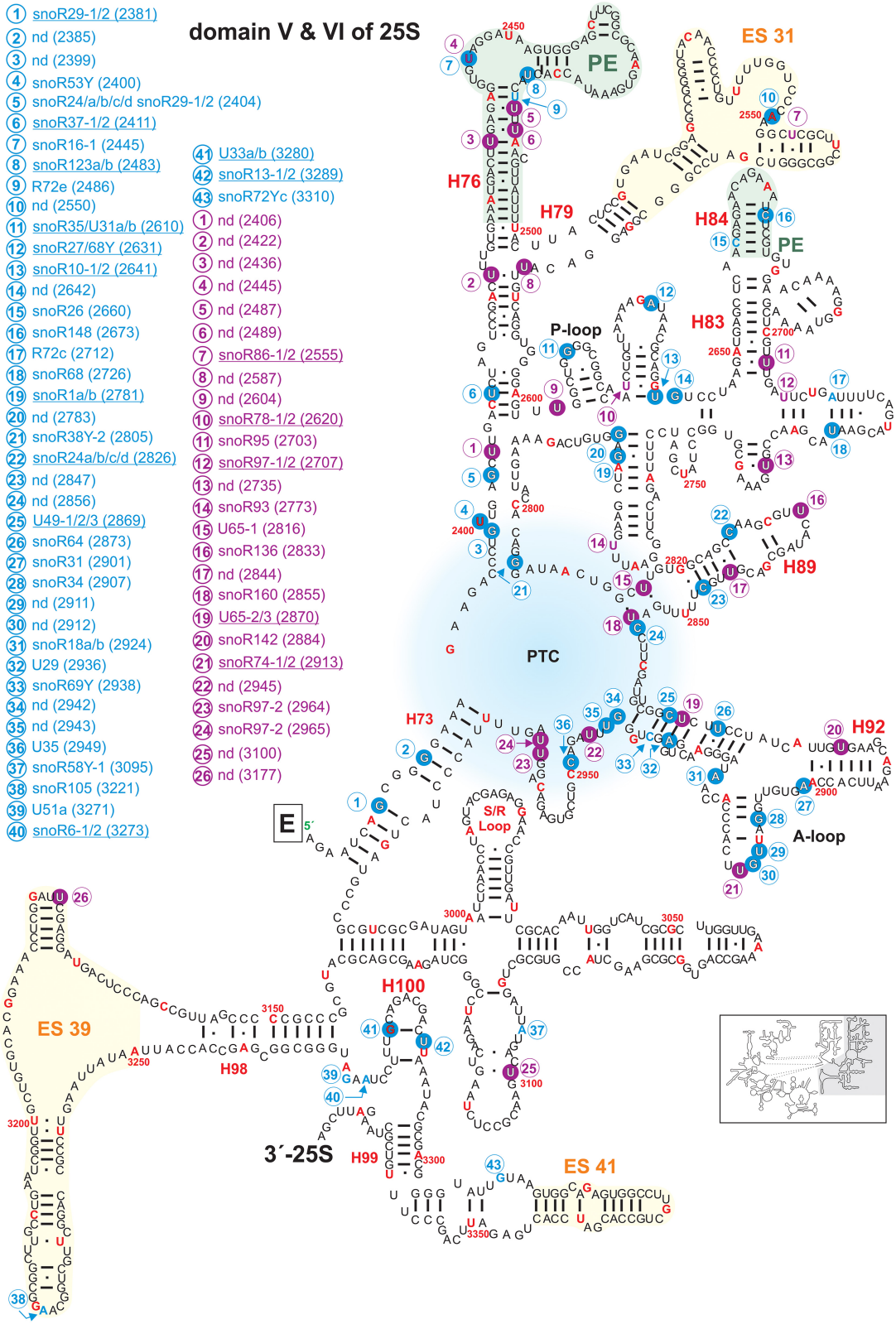


FIGURE 7 | Secondary structure diagram of domain V and VI of 25S rRNA of *A. thaliana*. The image is shown according to the legend for **Figure 3**. Framed large letters indicate the position of the connections in subsequent images (for E see **Figure 6**). The region of the PTC is indicated in blue. The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 60S rRNA secondary structure highlighting the according region is enclosed.

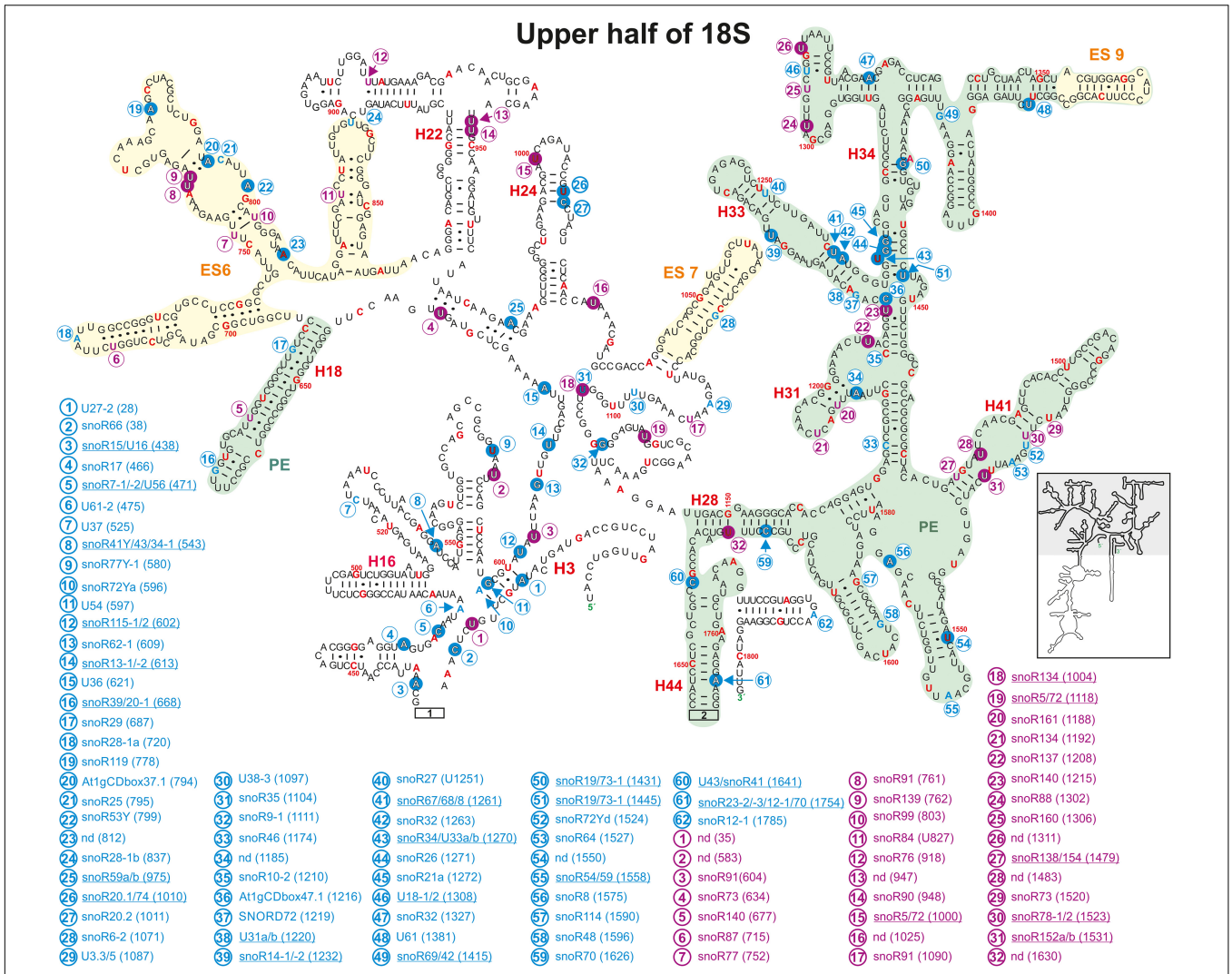


FIGURE 8 | Secondary structure diagram of the 18S rRNA (At2g01010) of *A thaliana* according to The RNAcentral Consortium (2019). Depicted are the positions from 1 to 41, 436 to 1655, and 1750 to 1804. Predicted and verified positions for 2′-O-ribose-methylations and pseudouridylations, as well as of snoRNAs were obtained from Brown et al. (2001, 2003), Qu et al. (2001), Yoshihama et al. (2013), Sun et al. (2019), Streit et al. (2020). The coloring is according to **Figure 3**. Framed large numbers indicate the position of the connections in **Figure 9**. PE annotates a pivoting element and ES expansion elements previously identified (Paci and Fox, 2015). The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 40S rRNA secondary structure highlighting the according region is enclosed.

Helix 34 contains four mapped 2′-O-ribose-me sites in *A. thaliana*, while human and yeast contain three 2′-O-ribose-me sites. Both yeast and human 18S rRNA contain three pseudouridine sites in helix 31 with one site having the hypermodification *m1acp3*Ψ (N1-methyl-N3-aminocarboxypropyl-pseudouridine) at U1191 and U1248, respectively. *A. thaliana* contains two predicted Ψ-sites in this helix, while one of these sites U1192 is the yeast and human equivalent of *m1acp3*Ψ (Piekna-Przybylska et al., 2007). However, it needs to be analyzed whether the plant rRNA also carries hypermodifications as their vertebrate and yeast counterparts.

In helix 44 of the *A. thaliana*, two 2′-O-ribose-me are mapped and additional three are predicted, while only a single Ψ-site is annotated (**Figures 8**, **9**). In contrast, human 18S rRNA contain two and yeast only one 2′-O-ribose-me sites in H44 (Piekna-Przybylska et al., 2007). Interestingly, *A. thaliana* harbors a novel and unique Ψ-site at U1702 in the helix 44 region of ES12 (**Figure 9**). In yeast, H44 just close to the 3′ end is one of the binding sites for the helicase Prp43, which is required for final maturation of 20S and 27S pre-rRNA, respectively (Bohnsack et al., 2009; Pertschy et al., 2009). For example, in the case of 20S maturation in yeast, Prp43 is assumed to be involved in unwinding of the pre-rRNA enabling endonuclease Nob1 for cleavage at site D (**Figure 1**; Bohnsack et al., 2009). It is likely, the modification at Am1754 could be necessary for binding of the Prp43 helicase. Yeast, of all things lacks this modification in H44 (Piekna-Przybylska et al., 2007).

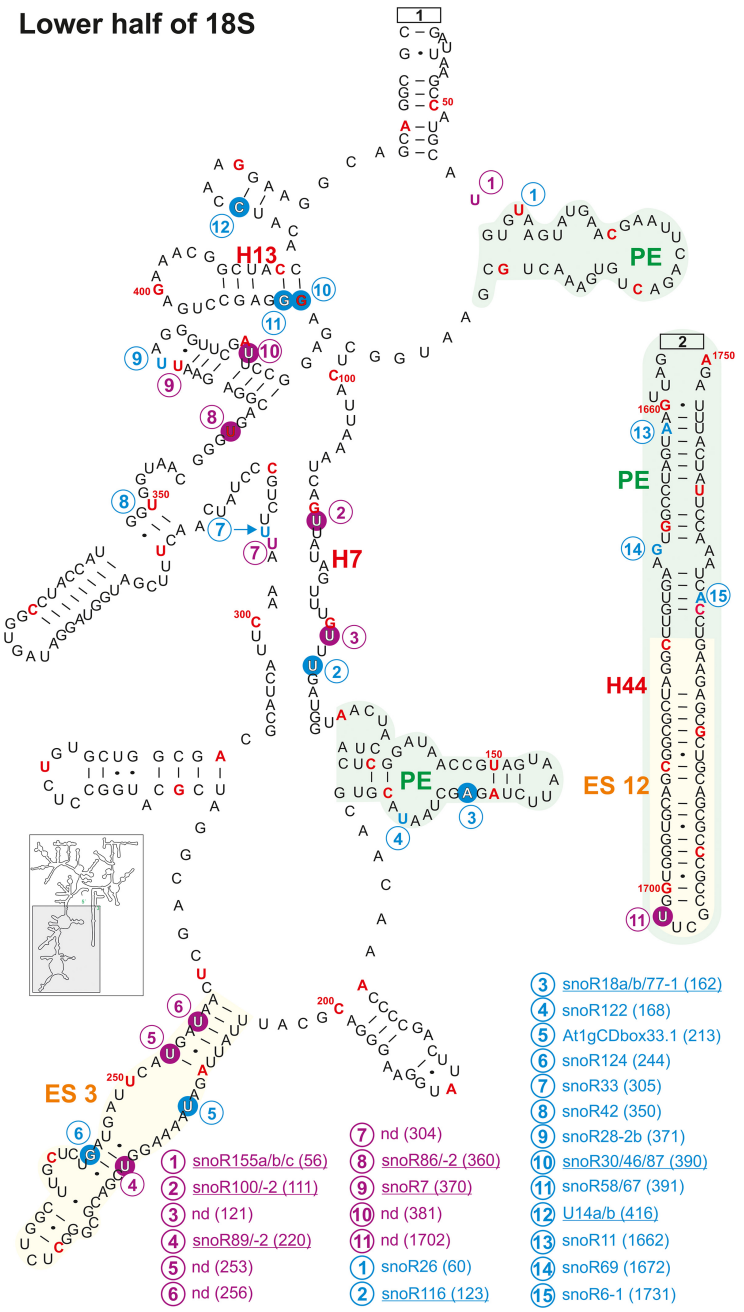


FIGURE 9 | Continuation of the secondary structure diagram of the 18S rRNA (At2g01010) of *A. thaliana* according to The RNAcentral Consortium (2019). All informations are listed in the legend of **Figure 8**. The number in brackets correspond to the modified nucleotide position. A small illustration of the whole 40S rRNA secondary structure highlighting the according region is enclosed.

Expansion segment 6 (ES6) is located at the surface of the small subunit and highly conserved in plants, vertebrates, and yeast (Alkemar and Nygård, 2003, 2006). As longest expansion segment it contains two mapped Ψ-sites, two mapped 2′-O-ribose-me sites and five predicted sites in *A. thaliana* (**Figure 8**). In contrast, the rRNA of human ES6 has six Ψ-sites and two 2′-O-ribose-me sites, while the same element in yeast contains two Ψ-sites and one 2′-O-ribose-me site (Piekna-Przybylska et al., 2007). Remarkably, Arabidopsis ES6 shows similarities to both human and yeast. The two Ψ-sites found in *A. thaliana* (Ψ761 and Ψ762) are conserved with the human sites (Ψ814 and Ψ815) and that one of the Arabidopsis 2′-O-ribose-me site (Am799) is conserved to the yeast site (Am796; **Figure 8**, Piekna-Przybylska et al., 2007; **Supplementary Table 3**). However, the function of this segment and thus, their modifications remain elusive. *In vivo* crosslinking

in yeast suggested that snR30/U17 snoRNAs can bind to two conserved sites in the ES6 to permit proper 18S rRNA processing (Fayet-Lebaron et al., 2009). As snR30 and U17 could not be identified so far in *A. thaliana*, independent pathway for modification might have evolved for *A. thaliana*.

Ribosomal proteins and RBFs bind to specific regions within the rRNA. It could be shown that Enp1 (essential nuclear protein 1) binds to an AUU sequence in helix 33 in yeast, where it is required for pre-rRNA processing of 18S (Chen W. et al., 2003; Granneman et al., 2010). Helix 33 in *Arabidopsis* contains two 2′-O-ribose-methylations (Um1261 and Am1263) in the downstream adjacent region of the AUU site (**Figure 8** and **Supplementary Table 3**). This seems to be unique for plants as yeast and human 18S rRNA lack these modifications (Piekna-Przybylska et al., 2007). Besides, a new modification site at Cm1219 in helix 33 was predicted based on the identification of the snoRNA SNORD72 (Streit et al., 2020).

Helix 41 of the 18S rRNA in *A. thaliana* contains two mapped Ψ-sites at U1483 and U1531 located within a region, which is a binding site for rpS5 in yeast (**Figure 8**; **Supplementary Table 3**; Granneman et al., 2010). However, human and yeast helix 41 of the 18S rRNA is not modified (Piekna-Przybylska et al., 2007).

Helix 28 of *A. thaliana* 18S rRNA contains a single mapped 2′-O-ribose-me site at Cm1626 targeted by snoR70 and a single novel Ψ-site at U1630 targeted from an unknown snoRNA (**Figure 8**). Human 18S rRNA contains a Ψ-site at the same position (U1692) in helix 28, whereas yeast lacks any modification is this region (Piekna-Przybylska et al., 2007; **Supplementary Table 3**). The helices 36 and 37 of *A. thaliana* contain each one mapped Ψ-site at U1302 (H36, snoR88) and U1311 (H37, unknown snoRNA; **Figure 8**). Although the human helix 37 contains at least the counterpart of U1311 of Arabidopsis at U1367, yeast helices are absent of these modification sites (Piekna-Przybylska et al., 2007; **Supplementary Table 3**).

CONCLUSION

Prediction and experimental verification suggest that the rRNA of *A. thaliana* is extensively decorated with different varieties of modifications, where pseudouridylations and 2′-O-ribose-methylations represent most of the modifications (Charette and Gray, 2000; Dimitrova et al., 2019). For Arabidopsis 18S rRNA almost 55% of all predicted 2′-O-ribose-methylation sites and 65% of all predicted pseudouridylation sites were successfully experimentally verified (**Supplementary Table 3**). For the 25S rRNA even 65% of all predicted 2′-O-ribose-methylation sites and almost 58% of predicted pseudouridylation sites were experimentally confirmed (**Supplementary Table 2**). In contrast, the 5.8S rRNA carries only a low number of modifications (**Figure 3** and **Supplementary Table 1**). Two new modification sites at Ψ22, which is plant specific and Ψ78 (yeast Ψ73) were found in the past (Sun et al., 2019; **Supplementary Table 1**). In contrast, based on the antisense element of the snoRNA snoR4a/4b (Brown et al., 2001) the modification at the 3′-end of 5.8S rRNA (Gm155) was predicted, but could not be experimentally confirmed (**Supplementary Table 1**).

The absence of experimental confirmation of predicted sites can have three different reasons. (i) Although nowadays a huge repertoire of techniques is used for mapping of modification sites, a certain limitation in detection sensitivity still exists. An interesting technique would be the use of mung bean nuclease protection assay coupled to RP-HPLC (Yang et al., 2016). (ii) It is discussed those modifications of the rRNA can be tissue or development specific (Chen and Wu, 2009; Sloan et al., 2017; Streit et al., 2020). Thus, the absence of detection can be the result of the analysis of a specific type of ribosomal systems. (iii) It is known that snoRNAs are also involved in the folding of rRNA elements (Bertrand and Fournier, 2013). Hence, it cannot be excluded at stage that some of the modification sites predicted by the detection of snoRNAs might not exist, as the snoRNA is required for guiding a snoRNP involved in rRNA processing or folding.

The modifications in 5.8S varies between the three model species. While yeast 5.8S rRNA contains only one modification site in H7, human and plant 5.8S rRNAs carry two 2′-O-ribose-me and two pseudouridylations (**Figure 3**; **Supplementary Table 1**; Piekna-Przybylska et al., 2007). Modifications like pseudouridylations in rRNA are required for the stability of the RNA structure in the ribosomes and 2′-O-ribose-me are necessary for translational accuracy and efficiency (Wu et al., 2015; Erales et al., 2017). Consistent with this finding, the 5.8S rRNA plays a crucial role in translation elongation (translocation; Elela and Nazar, 1997), for which the modifications of 5.8S might be important. In turn, it appears that the 5.8S rRNA of human and plants share high similarities while yeast seems to have evolved a unique way for keeping the structural and translational balance.

Remarkably, human and *A. thaliana* rRNAs share many conserved sites, which are not present in yeast. This elucidates that plant and human rRNA, despite the different sizes, are closer related than plant to yeast rRNA. Moreover, a subset of modifications is clearly unique to Arabidopsis like modifications in the GTPase center (**Figure 5**), in ES27 (**Figure 6**) or ES3 in the 18S rRNA (**Figure 9**). All these regions are targets of a subset of RBFs and RPs. Hence, the plant specific modification pattern stands in relation to the observed plant specificities of the rRNA processing (Weis et al., 2015b; Sáez-Vásquez and Delseny, 2019; Palm et al., 2019) and the modifications might be required for stabilizing the binding of the rRNA to proteins.

Furthermore, alternative functions of snoRNAs were proposed for the plant system. It was suggested that snoRNAs may regulate the modification level of rRNAs and snRNAs under stress as found for drought stress (Zheng et al., 2019). Thus, snoRNAs might have additional functions in plants, which must be discovered.

In future, it will be important to identify the snoRNAs responsible for certain newly discovered modifications sites, and in turn to map rRNA modifications in ribosomes isolated form different tissues, from plants at different developmental stages and after various stress treatments to complete the picture of the Arabidopsis rRNA modification landscape. The latter would perhaps show whether the high number of predicted modifications sites argues for a high ribosome heterogeneity. This concept might even be valid for ribosomes within a

single cell but required for the translation of different mRNA pools, which begs the analysis of the rRNA modification profile associated with different mRNAs. Moreover, it will be important to establish a complete profile of rRNA modifications of other plants to allow conclusions on globally conserved and species-specific modifications. The latter is of particular importance to transfer the knowledge based on model systems into agricultural applications.

AUTHOR CONTRIBUTIONS

DS and ES: conceptualization, data curation, writing – original draft preparation, and visualization. ES: supervision and funding acquisition. Both authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Deutsche Forschungsgemeinschaft, SFB 902, B9 to ES.

ACKNOWLEDGMENTS

We would like to thank Sotirios Fragkostefanakis and Thiruvenkadam Shanmugam for reading the manuscript and for the helpful and critical discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.684626/full#supplementary-material

Supplementary Table 1 | List of all mapped and predicted modification sites within 5.8S rRNA in *Arabidopsis thaliana* and corresponding human and yeast sites. Not determined (nd).

Supplementary Table 2 | List of all mapped or predicted modification sites within 25S rRNA in *Arabidopsis thaliana* and corresponding human and yeast sites. Not determined (nd).

Supplementary Table 3 | List of all predicted and mapped modification sites within 18S rRNA in *Arabidopsis thaliana* and corresponding human and yeast sites. Not determined (nd).

REFERENCES

Alkemar, G., and Nygård, O. (2006). Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry* 45, 8067–8078. doi: 10.1021/bi052149z

Alkemar, G., and Nygård, O. D. D. (2003). A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits. *RNA* 9, 20–24. doi: 10.1261/rna.2108203

Armache, J. P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Bhushan, S., et al. (2010a). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *PNAS* 107, 19748–19753. doi: 10.1073/pnas.1009999107

Armache, J. P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Bhushan, S., et al. (2010b). Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *PNAS* 107, 19754–19759. doi: 10.1073/pnas.1010005107

Azevedo-Favory, J., Gaspin, C., Ayadi, L., Montacié, C., Marchand, V., Jobet, E., et al. (2020). Mapping rRNA 2′-O-methylations and identification of C/D snoRNAs in Arabidopsis thaliana plants. *RNA Biol.* 1–18. doi: 10.1080/15476286.2020.1869892

Bachellerie, J. P., and Cavaillé, J. (1997). Guiding ribose methylation of rRNA. *Trends Biochem. Sci.* 22, 257–261. doi: 10.1016/s0968-0004(97)01057-8

Barneche, F., Gaspin, C., Guyot, R., and Echeverria, M. (2001). Identification of 66 box C/D snoRNAs in Arabidopsis thaliana: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2′-O-methylation sites. *J. Mol. Biol.* 311, 57–73. doi: 10.1006/jmbi.2001.4851

Barneche, F., Steinmetz, F., and Echeverrìa, M. (2000). Fibrillarin genes encode both a conserved nucleolar protein and a novel small nucleolar RNA involved in ribosomal RNA methylation in arabidopsis thaliana. *J. Biol. Chem.* 275, 27212–27220. doi: 10.1016/s0021-9258(19)61499-7

Bassham, D. C., and MacIntosh, G. C. (2017). Degradation of cytosolic ribosomes by autophagy-related pathways. *Plant Sci.* 262, 169–174. doi: 10.1016/j.plantsci.2017.05.008

Baxter-Roshek, J. L., Petrov, A. N., and Dinman, J. D. (2007). Optimization of ribosome structure and function by rRNA base modification. *PLoS One* 2:e174. doi: 10.1371/journal.pone.0000174

Beckmann, R., Spahn, C. M., Eswar, N., Helmers, J., Penczek, P. A., Sali, A., et al. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107, 361–372. doi: 10.1016/s0092-8674(01)00541-4

Bertrand, E., and Fournier, M. J. (2013). *The snoRNPs and related machines: ancient devices that mediate maturation of rRNA and other RNAs. In Madame Curie Bioscience Database [Internet]*. Austin, TX: Landes Bioscience.

Bhattacharya, D. P., Canzler, S., Kehr, S., Hertel, J., Grosse, I., and Stadler, P. F. (2016). Phylogenetic distribution of plant snoRNA families. *BMC Genom.* 17:969. doi: 10.1186/s12864-016-3301-2

Bohnsack, M. T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E., and Tollervey, D. (2009). Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol. Cell* 36, 583–592. doi: 10.1016/j.molcel.2009.09.039

Briones, E., Briones, C., Remacha, M., and Ballesta, J. P. (1998). The GTPase Center Protein L12 is required for correct ribosomal stalk assembly but Not for *Saccharomyces* cerevisiae Viability. *J. Biol. Chem.* 273, 31956–31961. doi: 10.1074/jbc.273.48.31956

Brown, J. W., Clark, G. P., Leader, D. J., Simpson, C. G., and Lowe, T. O. D. D. (2001). Multiple snoRNA gene clusters from Arabidopsis. *RNA* 7, 1817–1832.

Brown, J. W., Echeverria, M., and Qu, L. H. (2003). Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.* 8, 42–49. doi: 10.1016/s1360-1385(02)00007-9

Brown, J. W., and Shaw, P. J. (1998). Small nucleolar RNAs and pre-rRNA processing in plants. *Plant Cell* 10, 649–657. doi: 10.2307/3870654

Charette, M., and Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* 49, 341–351. doi: 10.1080/152165400410182

Chen, C. L., Liang, D., Zhou, H., Zhuo, M., Chen, Y. Q., and Qu, L. H. (2003). The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from Oryza sativa. *Nucleic Acids Res.* 31, 2601–2613. doi: 10.1093/nar/gkg373

Chen, H. M., and Wu, S. H. (2009). Mining small RNA sequencing data: a new approach to identify small nucleolar RNAs in Arabidopsis. *Nucleic Acids Res.* 37:e69. doi: 10.1093/nar/gkp225

Chen, W., Bucaria, J., Band, D. A., Sutton, A., and Sternglanz, R. (2003). Enp1, a yeast protein associated with U3 and U14 snoRNAs, is required for pre-rRNA processing and 40S subunit synthesis. *Nucleic Acids Res.* 31, 690–699. doi: 10.1093/nar/gkg145

Chow, C. S., Lamichhane, T. N., and Mahto, S. K. (2007). Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem. Biol.* 2, 610–619. doi: 10.1021/cb7001494

Decatur, W. A., and Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends Biochem. Sci.* 27, 344–351. doi: 10.1016/s0968-0004(02)02109-6

Demirci, H., Murphy, F., Belardinelli, R., Kelley, A. C., Ramakrishnan, V., Gregory, S. T., et al. (2010). Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function. *RNA* 16, 2319–2324. doi: 10.1261/rna.2357210

Desaulniers, J. P., Chang, Y. C., Aduri, R., Abeysirigunawardena, S. C., SantaLucia, J. Jr., and Chow, C. S. (2008). Pseudouridines in rRNA helix 69 play a role in loop stacking interactions. *Org. Biomol. Chem.* 6, 3892–3895. doi: 10.1039/b812731j

Dieci, G., Preti, M., and Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94, 83–88. doi: 10.1016/j.ygeno.2009.05.002

Dimitrova, D. G., Teysset, L., and Carré, C. (2019). RNA 2′-O-methylation (Nm) modification in human diseases. *Genes* 10:117. doi: 10.3390/genes10020117

Dragon, F., Lemay, V., and Trahan, C. (2006). snoRNAs: biogenesis, structure and function. *Encycl. Life Sci.* 2006, 1–7. doi: 10.1002/9780470015902.a0001377

Egebjerg, J., Douthwaite, S. R., Liljas, A., and Garrett, R. A. (1990). Characterization of the binding sites of protein L11 and the L10.(L12) 4 pentameric complex in the GTPase domain of 23 S ribosomal RNA from *Escherichia coli*. *J. Mol. Biol.* 213, 275–288. doi: 10.1016/s0022-2836(05)80190-1

Elela, S. A., and Nazar, R. N. (1997). Role of the 5.8 S rRNA in ribosome translocation. *Nucleic Acids Res.* 25, 1788–1794. doi: 10.1093/nar/25.9.1788

Endo, Y., Chan, Y. L., Lin, A., Tsurugi, K., and Wool, I. G. (1988). The cytotoxins alpha-sarcin and ricin retain their specificity when tested on a synthetic oligoribonucleotide (35-mer) that mimics a region of 28 S ribosomal ribonucleic acid. *J. Biol. Chem.* 263, 7917–7920. doi: 10.1016/s0021-9258(18)68418-2

Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S. E., et al. (2017). Evidence for rRNA 2′-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *PNAS* 114, 12934–12939. doi: 10.1073/pnas.1707674114

Falaleeva, M., Welden, J. R., Duncan, M. J., and Stamm, S. (2017). C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *Bioessays* 39:1600264. doi: 10.1002/bies.201600264

Fayet-Lebaron, E., Atzorn, V., Henry, Y., and Kiss, T. (2009). 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO J.* 28, 1260–1270. doi: 10.1038/emboj.2009.79

Ferretti, M. B., and Karbstein, K. (2019). Does functional specialization of ribosomes really exist? *RNA* 25, 521–538. doi: 10.1261/rna.069823.118

Filippova, J. A., Semenov, D. V., Juravlev, E. S., Komissarov, A. B., Richter, V. A., and Stepanov, G. A. (2017). Modern approaches for identification of modified nucleotides in RNA. *Biochem Mosc.* 82, 1217–1233. doi: 10.1134/s0006297917110013

Fujii, K., Susanto, T. T., Saurabh, S., and Barna, M. (2018). Decoding the function of expansion segments in ribosomes. *Mol. Cell* 72, 1013–1020. doi: 10.1016/j.molcel.2018.11.023

Ge, J., and Yu, Y. T. (2013). RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci.* 38, 210–218. doi: 10.1016/j.tibs.2013.01.002

Gigova, A., Duggimpudi, S., Pollex, T. I. M., Schaefer, M., and Koš, M. (2014). A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability. *RNA* 20, 1632–1644. doi: 10.1261/rna.043398.113

Graifer, D., Molotkov, M., Eremina, A., Ven'Yaminova, A., Repkova, M., and Karpova, G. (2005). The central part of the 5.8 S rRNA is differently arranged in programmed and free human ribosomes. *Biochem. J.* 387, 139–145. doi: 10.1042/bj20041450

Granneman, S., Petfalski, E., Swiatkowska, A., and Tollervey, D. (2010). Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA–protein cross-linking. *EMBO J.* 29, 2026–2036. doi: 10.1038/emboj.2010.86

Granneman, S., Petfalski, E., and Tollervey, D. (2011). A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8 S rRNA maturation by the Rat1 exonuclease. *EMBO J.* 30, 4006–4019. doi: 10.1038/emboj.2011.256

Gulay, S. P., Bista, S., Varshney, A., Kirmizialtin, S., Sanbonmatsu, K. Y., and Dinman, J. D. (2017). Tracking fluctuation hotspots on the yeast ribosome through the elongation cycle. *Nucleic Acids Res.* 45, 4958–4971. doi: 10.1093/nar/gkx112

Halic, M., Becker, T., Pool, M. R., Spahn, C. M., Grassucci, R. A., Frank, J., et al. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* 427, 808–814. doi: 10.1038/nature02342

Hang, R., Wang, Z., Deng, X., Liu, C., Yan, B., Yang, C., et al. (2018). Ribosomal RNA biogenesis and its response to chilling stress in Oryza sativa. *Plant physiol.* 177, 381–397. doi: 10.1104/pp.17.01714

Hellmann, E. (2020). How to make an extraordinary machine: SMALL ORGAN4 regulates ribosome biogenesis in plants. *Plant Physiol.* 184, 1627–1629. doi: 10.1104/pp.20.01456

Henras, A. K., Soudet, J., Gerus, M., Lebaron, S., Caizergues-Ferrer, M., Mougin, A., et al. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci.* 65, 2334–2359. doi: 10.1007/s00018-008-8027-0

Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., et al. (2012). Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* 40, 391–398. doi: 10.1093/nar/gkr700

Holley, C. L., Li, M. W., Scruggs, B. S., Matkovich, S. J., Ory, D. S., and Schaffer, J. E. (2015). Cytosolic accumulation of small nucleolar RNAs (snoRNAs) is dynamically regulated by NADPH oxidase. *J. Biol. Chem.* 290, 11741–11748. doi: 10.1074/jbc.m115.637413

Huang, C., Karijolich, J., and Yu, Y. T. (2016). Detection and quantification of RNA 2′-O-methylation and pseudouridylation. *Methods* 103, 68–76. doi: 10.1016/j.ymeth.2016.02.003

Ito, S., Akamatsu, Y., Noma, A., Kimura, S., Miyauchi, K., Ikeuchi, Y., et al. (2014). A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal subunit in *Saccharomyces* cerevisiae. *J. Biol. Chem.* 289, 26201–26212. doi: 10.1074/jbc.m114.593996

Karijolich, J., Kantartzis, A., and Yu, Y. T. (2010). RNA modifications: a mechanism that modulates gene expression. *Methods Mol. Biol.* 629, 1–19. doi: 10.1007/978-1-60761-657-3_1

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., et al. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. doi: 10.1038/35048692

Khatter, H., Myasnikov, A. G., Natchiar, S. K., and Klaholz, B. P. (2015). Structure of the human 80S ribosome. *Nature* 520, 640–645.

Kim, D. F., and Green, R. (1999). Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol. Cell.* 4, 859–864. doi: 10.1016/s1097-2765(00)80395-0

Kim, S. H., Spensley, M., Choi, S. K., Calixto, C. P., Pendle, A. F., Koroleva, O., et al. (2010). Plant U13 orthologues and orphan snoRNAs identified by RNomics of RNA from Arabidopsis nucleoli. *Nucleic Acids Res.* 38, 3054–3067. doi: 10.1093/nar/gkp1241

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 20, 3617–3622. doi: 10.1093/emboj/20.14.3617

Kiss, T., Marshallsay, C., and Filipowicz, W. (1991). Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro. *Cell* 65, 517–526. doi: 10.1016/0092-8674(91)90469-f

Kiss-László, Z., Henry, Y., and Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* 17, 797–807. doi: 10.1093/emboj/17.3.797

Klinge, S., and Woolford, J. L. (2019). Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* 20, 116–131. doi: 10.1038/s41580-018-0078-y

Knorr, A. G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., et al. (2019). Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol* 26, 35–39. doi: 10.1038/s41594-018-0165-y

Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., et al. (2003). Plant dicistronic tRNA–snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.* 22, 621–632. doi: 10.1093/emboj/cdg040

Lafontaine, D., Vandenhaute, J., and Tollervey, D. (1995). The 18S rRNA dimethylase Dim1pis required for pre-ribosomal RNA process-ing in yeast. *Genes Dev.* 9, 2470–2481. doi: 10.1101/gad.9.20.2470

Lafontaine, D. L. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat. Struct. Mol. Biol.* 22, 11–19. doi: 10.1038/nsmb.2939

Lau, R. Y., Kennedy, T. D., and Lane, B. G. (1974). Wheat-embryo ribonucleates. III. Modified nucleotide constituents in each of the 5.8 S, 18S and 26S ribonucleates. *Can. J. Biochem.* 52, 1110–1123. doi: 10.1139/o74-155

Leader, D. J., Clark, G. P., Watters, J., Beven, A. F., Shaw, P. J., and Brown, J. W. (1997). Clusters of multiple different small nucleolar RNA genes in plants

are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J.* 16, 5742–5751. doi: 10.1093/emboj/16.18.5742

Leader, D. J., Sanders, J. F., Waugh, R., Shaw, P., and Brown, J. W. (1994). Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucleic Acids Res.* 22, 5196–5203. doi: 10.1093/nar/22.24.5196

Leidig, C., Thoms, M., Holdermann, I., Bradatsch, B., Berninghausen, O., Bange, G., et al. (2014). 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nat. Commun.* 5, 1–8.

Liang, X. H., Liu, Q., and Fournier, M. J. (2007). rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell.* 28, 965–977. doi: 10.1016/j.molcel.2007.10.012

Liang, X. H., Liu, Q., and Fournier, M. J. (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 15, 1716–1728. doi: 10.1261/rna.1724409

Lilley, D. M. (2001). The ribosome functions as a ribozyme. *Chembiochem.* 2, 31–35. doi: 10.1002/1439-7633(20010105)2:1<31::aid-cbic31>3.0.co;2-p

Lindsay, M. A., Griffiths-Jones, S., Lui, L., and Lowe, T. (2013). Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem.* 54, 53–77. doi: 10.1042/bse0540053

Liu, T. T., Zhu, D., Chen, W., Deng, W., He, H., He, G., et al. (2013). A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in Oryza sativa. *Mol. Plant.* 6, 830–846. doi: 10.1093/mp/sss087

Macbeth, M. R., and Wool, I. G. (1999). The phenotype of mutations of G2655 in the sarcin/ricin domain of 23 S ribosomal RNA. *J. Mol. Bio.* 285, 965–975. doi: 10.1006/jmbi.1998.2388

Marker, C., Zemann, A., Terhörst, T., Kiefmann, M., Kastenmayer, J. P., Green, P., et al. (2002). Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. *Curr. Biol.* 12, 2002–2013.

Matera, A. G., Terns, R. M., and Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* 8, 209–220. doi: 10.1038/nrm2124

Missbach, S., Weis, B. L., Martin, R., Simm, S., Bohnsack, M. T., and Schleiff, E. (2013). 40S ribosome biogenesis co-factors are essential for gametophyte and embryo development. *PloS One* 8:e54084. doi: 10.1371/journal.pone.0054084

Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I., and Klaholz, B. P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. *Nature* 551, 472–477. doi: 10.1038/nature24482

Nazar, R. N., Sitz, T. O., and Busch, H. (1975). Tissue specific differences in the 2′-O-methylation of eukaryotic 5.8 S ribosomal RNA. *FEBS Lett.* 59, 83–87. doi: 10.1016/0014-5793(75)80346-2

Nazar, R. N., Sitz, T. O., and Somers, K. D. (1980). Cytoplasmic methylation of mature 5.8 S ribosomal RNA. *J. Mol. Biol.* 142, 117–121. doi: 10.1016/0022-2836(80)90209-0

Newman, D. R., Kuhn, J. F., Shanab, G. M., and Maxwell, E. S. (2000). Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* 6, 861–879. doi: 10.1017/s1355838200992446

Nygård, O., Alkemar, G., and Larsson, S. L. (2006). Analysis of the secondary structure of expansion segment 39 in ribosomes from fungi, plants and mammals. *J. Mol. Biol.* 357, 904–916. doi: 10.1016/j.jmb.2006.01.043

Paci, M., and Fox, G. E. (2015). Major centers of motion in the large ribosomal RNAs. *Nucleic Acids Res.* 43, 4640–4649. doi: 10.1093/nar/gkv289

Palm, D., Simm, S., Darm, K., Weis, B. L., Ruprecht, M., Schleiff, E., et al. (2016). Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in Arabidopsis thaliana. *RNA Biol.* 13, 441–454. doi: 10.1080/15476286.2016.1154252

Palm, D., Streit, D., Shanmugam, T., Weis, B. L., Ruprecht, M., Simm, S., et al. (2019). Plant-specific ribosome biogenesis factors in Arabidopsis thaliana with essential function in rRNA processing. *Nucleic Acids Res.* 47, 1880–1895. doi: 10.1093/nar/gky1261

Parker, M. S., Balasubramaniam, A., Sallee, F. R., and Parker, S. L. (2018). The expansion segments of 28S Ribosomal RNA extensively match human messenger RNAs. *Front. Genet.* 9:66. doi: 10.3389/fgene.2018.00066

Pertschy, B., Schneider, C., Gnädig, M., Schäfer, T., Tollervey, D., and Hurt, E. (2009). RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *J. Biol. Chem.* 284, 35079–35091. doi: 10.1074/jbc.m109.040774

Piekna-Przybylska, D., Decatur, W. A., and Fournier, M. J. (2007). The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.* 36, D178–D183.

Polacek, N., and Mankin, A. S. (2005). The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit. Rev. Biochem. Mol. Biol.* 40, 285–311. doi: 10.1080/10409230500326334

Polikanov, Y. S., Melnikov, S. V., Söll, D., and Steitz, T. A. (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* 22, 342–344. doi: 10.1038/nsmb.2992

Qu, G., Kruszka, K., Plewka, P., Yang, S. Y., Chiou, T. J., Jarmolowski, A., et al. (2015). Promoter-based identification of novel non-coding RNAs reveals the presence of dicistronic snoRNA-miRNA genes in Arabidopsis thaliana. *BMC Genom.* 16:1009. doi: 10.1186/s12864-015-2221-x

Qu, L. H., Meng, Q., Zhou, H., and Chen, Y. Q. (2001). Identification of 10 novel snoRNA gene clusters from Arabidopsis thaliana. *Nucleic Acids Res.* 29, 1623–1630. doi: 10.1093/nar/29.7.1623

Ramesh, M., and Woolford, J. L. (2016). Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA* 22, 1153–1162. doi: 10.1261/rna.056705.116

Ramos, L. M. G., Smeekens, J. M., Kovacs, N. A., Bowman, J. C., Wartell, R. M., Wu, R., et al. (2016). Yeast rRNA expansion segments: folding and function. *J. Mol. Biol.* 428, 4048–4059. doi: 10.1016/j.jmb.2016.08.008

Reddy, R., Henning, D., and Busch, H. (1979). Nucleotide sequence of nucleolar U3B RNA. *J. Biol. Chem.* 254, 11097–11105. doi: 10.1016/s0021-9258(19)86635-8

Reddy, R., Sitz, T. O., Ro-Choi, T. S., and Busch, H. (1974). Two-dimensional polyacrylamide gel electrophoresis separation of low molecular weight nuclear RNA. *Biochem. Biophys. Res. Commun.* 56, 1017–1022. doi: 10.1016/s0006-291x(74)80290-1

Rodor, J., Jobet, E., Bizarro, J., Vignols, F., Carles, C., Suzuki, T., et al. (2011). AtNUFIP, an essential protein for plant development, reveals the impact of snoRNA gene organisation on the assembly of snoRNPs and rRNA methylation in Arabidopsis thaliana. *Plant J.* 65, 807–819. doi: 10.1111/j.1365-313x.2010.04468.x

Ryan, P. C., and Draper, D. E. (1991). Detection of a key tertiary interaction in the highly conserved GTPase center of large subunit ribosomal RNA. *PNAS* 88, 6308–6312. doi: 10.1073/pnas.88.14.6308

Sáez-Vásquez, J., and Delseny, M. (2019). Ribosome biogenesis in plants: from functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell.* 31, 1945–1967. doi: 10.1105/tpc.18.00874

Samarsky, D. A., Fournier, M. J., Singer, R. H., and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.* 17, 3747–3757. doi: 10.1093/emboj/17.13.3747

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., et al. (2000). Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* 102, 615–623. doi: 10.1016/s0092-8674(00)00084-2

Sergiev, P. V., Lesnyak, D. V., Burakovsky, D. E., Kiparisov, S. V., Leonov, A. A., Bogdanov, A. A., et al. (2005). Alteration in location of a conserved GTPase-associated center of the ribosome induced by mutagenesis influences the structure of peptidyltransferase center and activity of elongation factor G. *J. Biol. Chem.* 280, 31882–31889. doi: 10.1074/jbc.m505670200

Sharma, S., Marchand, V., Motorin, Y., and Lafontaine, D. L. (2017a). Identification of sites of 2′-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* 7, 1–15.

Sharma, S., Yang, J., van Nues, R., Watzinger, P., Kötter, P., Lafontaine, D. L., et al. (2017b). Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet.* 13:e1006804. doi: 10.1371/journal.pgen.1006804

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L., and Bohnsack, M. T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G., et al. (2001). Structure of the 80S ribosome from Saccharomyces cerevisiae—tRNA-ribosome and subunit-subunit interactions. *Cell* 107, 373–386. doi: 10.1016/s0092-8674(01)00539-6

Streit, D., Shanmugam, T., Garbelyanski, A., Simm, S., and Schleiff, E. (2020). The existence and localization of nuclear snoRNAs in Arabidopsis thaliana revisited. *Plants* 9:1016. doi: 10.3390/plants9081016

Sun, L., Xu, Y., Bai, S., Bai, X., Zhu, H., Dong, H., et al. (2019). Transcriptome-wide analysis of pseudouridylation of mRNA and non-coding RNAs in Arabidopsis. *J. Exp. Bot.* 70, 5089–5600. doi: 10.1093/jxb/erz273

Szewczak, A. A., and Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* 247, 81–98. doi: 10.1006/jmbi.1994.0124

Taylor, D. J., Devkota, B., Huang, A. D., Topf, M., Narayanan, E., Sali, A., et al. (2009). Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 17, 1591–1604. doi: 10.1016/j.str.2009.09.015

Terns, M. P., and Terns, R. M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.* 10, 17–39.

The RNAcentral Consortium (2019). RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.* 47, D221–D229.

Tollervey, D., and Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* 9, 337–342. doi: 10.1016/s0955-0674(97)80005-1

Torres de Farias, S., Gaudêncio Rêgo, T., and José, M. V. (2017). Peptidyl transferase center and the emergence of the translation system. *Life* 7:21. doi: 10.3390/life7020021

Tsang, C. K., Bertram, P. G., Ai, W., Drenan, R., and Zheng, X. S. (2003). Chromatin-mediated regulation of nucleolar structure and RNA Pol I localization by TOR. *EMBO J.* 22, 6045–6056. doi: 10.1093/emboj/cdg578

Turkina, M. V., Årstrand, H. K., and Vener, A. V. (2011). Differential phosphorylation of ribosomal proteins in Arabidopsis thaliana plants during day and night. *PloS One* 6:e29307. doi: 10.1371/journal.pone.0029307

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in *Saccharomyces* cerevisiae. *Annu. Rev. Genet.* 33, 261–311.

Watkins, N. J., and Bohnsack, M. T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA.* 3, 397–414. doi: 10.1002/wrna.117

Weinstein, L. B., and Steitz, J. A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.* 11, 378–384. doi: 10.1016/S0955-0674(99)80053-2

Weis, B. L., Kovacevic, J., Missbach, S., and Schleiff, E. (2015b). Plant-specific features of ribosome biogenesis. *Trends Plant Sci.* 20, 729–740. doi: 10.1016/j.tplants.2015.07.003

Weis, B. L., Palm, D., Missbach, S., Bohnsack, M. T., and Schleiff, E. (2015a). atBRX1-1 and atBRX1-2 are involved in an alternative rRNA processing pathway in *Arabidopsis thaliana*. *RNA* 21, 415–425. doi: 10.1261/rna.047563.114

Wilson, D. M., Li, Y., LaPeruta, A., Gamalinda, M., Gao, N., and Woolford, J. L. (2020). Structural insights into assembly of the ribosomal nascent polypeptide exit tunnel. *Nat. commun.* 11, 1–15.

Woolford, J. L. Jr., and Baserga, S. J. (2013). Ribosome biogenesis in the yeast *Saccharomyces* cerevisiae. *Genetics* 195, 643–681. doi: 10.1534/genetics.113.153197

Wu, G., Huang, C., and Yu, Y. T. (2015). Pseudouridine in mRNA: incorporation, detection, and recoding. *Meth. Enzymol.* 560, 187–217.

Xie, Q., Wang, Y., Lin, J., Qin, Y., Wang, Y., and Bu, W. (2012). Potential key bases of ribosomal RNA to kingdom-specific spectra of antibiotic susceptibility and the possible archaeal origin of eukaryotes. *PloS One* 7:e29468. doi: 10.1371/journal.pone.0029468

Yang, J., Sharma, S., Watzinger, P., Hartmann, J. D., Kötter, P., and Entian, K. D. (2016). Mapping of complete set of ribose and base modifications of yeast rRNA by RP-HPLC and mung bean nuclease assay. *PloS One* 11:e0168873. doi: 10.1371/journal.pone.0168873

Yanshina, D. D., Bulygin, K. N., Malygin, A. A., and Karpova, G. G. (2015). Hydroxylated histidine of human ribosomal protein uL2 is involved in maintaining the local structure of 28S rRNA in the ribosomal peptidyl transferase center. *FEBS J.* 282, 1554–1566. doi: 10.1111/febs.13241

Yoshihama, M., Nakao, A., and Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. *BMC Res. Notes.* 6:426. doi: 10.1186/1756-0500-6-426

Zhao, X., and Yu, Y. T. (2004). Detection and quantitation of RNA base modifications. *RNA* 10, 996–1002. doi: 10.1261/rna.7110804

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T. T., Meyers, B. C., et al. (2012). The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.* 22, 689–694. doi: 10.1016/j.cub.2012.02.051

Zheng, J., Zeng, E., Du, Y., He, C., Hu, Y., Jiao, Z., et al. (2019). Temporal small RNA expression profiling under drought reveals a potential regulatory role of small nucleolar RNAs in the drought responses of Maize. *Plant Genome* 12, 1–15.

Zhou, H., Meng, Q., and Qu, L. (2000). Identification and structural analysis of a novel snoRNA gene cluster from Arabidopsis thaliana. *Sci. China C. Life Sci.* 43, 449–453.

Zhu, P., Wang, Y., Qin, N., Wang, F., Wang, J., Deng, X. W., et al. (2016). Arabidopsis small nucleolar RNA monitors the efficient pre-rRNA processing during ribosome biogenesis. *Proc. Natl. Acad. Sci.* 113, 11967–11972. doi: 10.1073/pnas.1614852113

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

REVIEW
published: 28 October 2021
doi: 10.3389/fpls.2021.773052

It Is Just a Matter of Time: Balancing Homologous Recombination and Non-homologous End Joining at the rDNA Locus During Meiosis

Jason Sims[1], Fernando A. Rabanal[2], Christiane Elgert[3], Arndt von Haeseler[3,4] and Peter Schlögelhofer[1]**

[1] Department of Chromosome Biology, Max Perutz Labs, University of Vienna, Vienna BioCenter, Vienna, Austria, [2] Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany, [3] Center for Integrative Bioinformatics Vienna (CIBIV), Max Perutz Labs, University of Vienna and Medical University of Vienna, Vienna BioCenter, Vienna, Austria, [4] Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:
Roman A. Volkov,
Chernivtsi University, Ukraine

Reviewed by:
Frédéric Pontvianne,
UMR5096 Laboratoire Génome et développement des plantes, France
Yingxiang Wang,
Fudan University, China

***Correspondence:**
Jason Sims
jason.sims@univie.ac.at
Peter Schlögelhofer
peter.schloegelhofer@univie.ac.at

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: 09 September 2021
Accepted: 04 October 2021
Published: 28 October 2021

Citation:
Sims J, Rabanal FA, Elgert C, von Haeseler A and Schlögelhofer P (2021) It Is Just a Matter of Time: Balancing Homologous Recombination and Non-homologous End Joining at the rDNA Locus During Meiosis. Front. Plant Sci. 12:773052. doi: 10.3389/fpls.2021.773052

Ribosomal RNA genes (rDNAs) are located in large domains of hundreds of rDNA units organized in a head-to-tail manner. The proper and stable inheritance of rDNA clusters is of paramount importance for survival. Yet, these highly repetitive elements pose a potential risk to the genome since they can undergo non-allelic exchanges. Here, we review the current knowledge of the organization of the rDNA clusters in *Arabidopsis thaliana* and their stability during meiosis. Recent findings suggest that during meiosis, all rDNA loci are embedded within the nucleolus favoring non-homologous end joining (NHEJ) as a repair mechanism, while DNA repair *via* homologous recombination (HR) appears to be a rare event. We propose a model where (1) frequent meiotic NHEJ events generate abundant single nucleotide polymorphisms and insertions/deletions within the rDNA, resulting in a heterogeneous population of rDNA units and (2) rare HR events dynamically change rDNA unit numbers, only to be observed in large populations over many generations. Based on the latest efforts to delineate the entire rDNA sequence in *A. thaliana,* we discuss evidence supporting this model. The results compiled so far draw a surprising picture of rDNA sequence heterogeneity between individual units. Furthermore, rDNA cluster sizes have been recognized as relatively stable when observing less than 10 generations, yet emerged as major determinant of genome size variation between different *A. thaliana* ecotypes. The sequencing efforts also revealed that transcripts from the diverse rDNA units yield heterogenous ribosome populations with potential functional implications. These findings strongly motivate further research to understand the mechanisms that maintain the metastable state of rDNA loci.

Keywords: rDNA, meiosis, homologous recombination, non-homologous end joining, evolution

INTRODUCTION

The central importance of the rRNA genes for the biology of any organism is evident, as they are essential for survival and for all cellular processes. They are among the evolutionary oldest and also most highly transcribed genomic regions forming the RNA building blocks of ribosomes. Most eukaryotic genomes contain clusters with hundreds to thousands of rRNA gene copies arranged in tandem which are transcribed and processed within the nucleolus.

In eukaryotes, the 18S, 5.8S, and 25S rRNAs form the scaffold for the small and large ribosomal subunits and all three are encoded together in functional units and transcribed as a single polycistronic 45S precursor transcript by RNA polymerase I (Wallace and Birnstiel, 1966; Moss et al., 2007; Layat et al., 2012). The 45S rRNA gene units (also termed rDNA units) are arranged in a head-to-tail manner in large clusters known as nucleolus organizing regions (NORs; Ritossa and Spiegelman, 1965; Wallace and Birnstiel, 1966). In the *Arabidopsis thaliana* reference ecotype Col-0, the 45S rDNA units are approximately 10 kb long and arranged in two clusters, each with ~400 repeats, at the top of chromosomes 2 and 4 (Copenhaver and Pikaard, 1996; Sims et al., 2021). A further component of the large ribosomal subunit, the 5S rRNA, is located on chromosomes 3, 4, and 5 in the *A. thaliana* Col-0 ecotype, also arranged in clusters and transcribed by RNA polymerase III (Murata et al., 1997; Layat et al., 2012). One 45S rDNA unit is also found in proximity of the 5S rDNA located on chromosome 3 (Abou-Ellail et al., 2011). Although rRNA transcripts account for approximately 50% of all transcribed RNAs in a cell, only a fraction of the rRNA genes is transcribed at a given time (Warner, 1999; Grummt and Pikaard, 2003; Pontvianne et al., 2010, 2012).

Recent studies have shown that individual 45S and also 5S rDNA units are not identical. Instead, they display a substantial amount of variability, not only within the intergenic regions but also in the genic regions transcribing the conserved ribosomal RNA subunits (Chandrasekhara et al., 2016; Havlová et al., 2016; Rabanal et al., 2017b). These variants have been exploited as molecular markers to study rDNA cluster-specific expression (see below).

The high level of transcriptional activity leads to torsional stress in rDNA and requires the activity of topoisomerases to relieve the positive and negative torsions (French et al., 2011). During replication, highly transcribed regions of the genome, such as the rDNA loci, may encounter frequent collisions between transcription and replication machineries, which need to be resolved (Castel et al., 2014; García-Muse and Aguilera, 2016; Sims et al., 2021). Both processes mentioned above are sources of DNA damage and genome instability in general. The unique nature of the rDNA loci not only makes them especially vulnerable to various types of DNA damage, it also demands special attention during DNA repair. The highly repetitive rDNA loci, with their hundreds of nearly identical rDNA units arranged head-to-tail, may undergo dramatic re-arrangements during homologous recombination (HR) DNA repair. HR may ultimately lead to lengthening or shortening of the rDNA arrays and in general to copy number instability (Warmerdam et al., 2016). Furthermore, the presence of rDNA clusters on multiple chromosomes adds the additional risk of inter-chromosomal recombination. As outlined in more detail below, during meiosis, a developmental program essential for the recombination of genetic traits, numerous DNA double-strand breaks (DSBs) are introduced. In this context, the rDNA loci are sequestered away from the canonical HR pathway and a different DNA repair pathway is employed, termed non-homologous end joining (NHEJ; Sims et al., 2019). NHEJ will less likely lead to genome re-arrangements and rDNA copy number loss, but may lead to the introduction of single nucleotide polymorphisms (SNPs) and short-range insertions/deletions (InDels; Chang et al., 2017; Wright et al., 2018; Xu and Xu, 2020).

In this review, we summarize the recent findings concerning the stability of the rDNA loci and their inheritance from a perspective of meiosis. We also provide a model, in agreement with the current data, that defines HR and NHEJ as the major determinants of rDNA cluster size and rDNA unit sequence variability.

DNA DOUBLE-STRAND BREAK FORMATION AND REPAIR

DNA damage, if not appropriately repaired, leads to loss of genetic material, genome re-arrangements, and cell cycle arrest. One of the most deleterious DNA insults are DNA DSBs which can for instance be generated by genotoxic agents or molecular tools like homing endo-nucleases, TALE nucleases, and CRISPR/Cas9 (Wu et al., 2014; Lopez et al., 2021) by endogenous processes like the re-establishment of collapsed replication forks or by a dedicated machinery during meiosis. As mentioned above, DSBs can be repaired by different repair pathways, the most prominent being NHEJ and HR. NHEJ has been found to be active during all cell cycle stages, whereas HR is the dominant repair pathway during S and G2 and is briefly introduced below.

NON-HOMOLOGOUS END JOINING

Non-homologous end joining was first described in mammals, where it is the predominant mechanism for DSB repair in non-cycling, somatic cells. It is differentiated in c-NHEJ (canonical) and a-NHEJ (alternative) pathways, the latter including microhomology-mediated end joining (MMEJ), all with the direct ligation of processed DNA ends as common denominator. Re-joining of blunt ends or ends with a few overlapping bases occurs without regard for preserving the sequence or context integrity (Hays, 2002; McVey and Lee, 2008; Chang et al., 2017).

c-NHEJ is initiated with the recognition and the juxtaposition of the broken ends. In mammals, this step is promoted by the DNA-dependent protein kinase (DNA-PK), a complex composed of the KU heterodimer (Xrcc5/6) and

the kinase DNA-PK catalytic subunit (DNA-PKcs; Blackford and Jackson, 2017). It is important to note that DNA-PKcs have not been found to be encoded in plant genomes, indicating that in plants NHEJ is orchestrated differently (Templeton and Moorhead, 2005; Yoshiyama et al., 2013). The Artemis protein and the Xrcc4/DNA ligase IV heterodimer are subsequently recruited, with Artemis involved in the maturation of the DSB ends and the Xrcc4/DNA ligase IV complex catalyzing the resealing of the ends (Lees-Miller and Meek, 2003; Meek et al., 2004; Bleuyard et al., 2006). The KU heterodimer is composed of Ku70 and Ku80 and is involved in recognition, protection, and juxtaposition of the ends of a DSB. DNA-PKcs proteins are recruited to the DSB sites *via* interactions with the Ku/DNA complex and by phosphorylating various substrates (e.g.: Ku70, Ku80, Artemis, Xrcc4; Fell and Schild-Poulter, 2015). Artemis possesses both exo- and endo-nuclease activities and performs phospho-regulated maturation of the DSB ends as it cleaves DNA hairpins and other DNA structures (Lobrich and Jeggo, 2017). The final step, consisting of the ligation of broken ends, is carried out by the Xrcc4/DNA ligase IV heterodimer, which is recruited by DNA-PK. The MRN complex, composed of the proteins Mre11, Rad50, and Nbs1, stimulates this ligase activity *in vitro* and is also implicated in the juxtaposition of the ends of the break (Grawunder et al., 1997; Durdikova and Chovanec, 2017).

The alternative NHEJ pathway MMEJ is promoted in the absence of c-NHEJ factors and involves the alignment of microhomologies at the DSB site (Seol et al., 2018). DNA ends are bound by PARP1 (potentially competing with Ku proteins; Wang et al., 2006; Cheng et al., 2011). Following DNA binding, PARP1 gets activated and poly-ADP-ribosylates itself and various targets in the vicinity leading to more accessible chromatin (Polo and Jackson, 2011; Beck et al., 2014). Subsequently, the MRE11-complex is recruited to process the DNA and prepares them for ligation *via* Ligase I or Ligase III.

The counterparts of most NHEJ proteins have been identified and characterized in plants (Bleuyard et al., 2006; Charbonnel et al., 2011). For instance, LIGASE4 is a well-conserved hallmark factor, also in plants, in the c-NHEJ DNA repair pathway (Friesner and Britt, 2003). MRE11 and its complex partners have also been identified and characterized in plants, and they together are required for both HR and MMEJ. Importantly, no homologs of DNA-PK and Ligase III and some further factors have been identified in plant genomes (Manova and Gruszka, 2015; Yoshiyama, 2016), highlighting some fundamental differences in the DNA damage response in plants and other organisms.

HOMOLOGOUS RECOMBINATION

In contrast, DNA DSB repair *via* the HR pathway preserves sequence integrity. Following DSB formation (see above), initiation of HR depends on the localization of the Mre11-Rad50-Nbs1/Xrs2 (MRN/X) complex and its partner CtIP/Sae2/Com1 to the DSB sites (Wright et al., 2018). The MRN complex bridges the two ends, is involved in DNA end processing, and recruits further processing proteins (e.g., a 5' to 3' exo-nuclease). The nucleolytic activities yield a 3' ssDNA overhang, competent to invade dsDNA to probe for a homologous repair template. In addition, the MRN/X complex recruits the DNA damage kinase ATM/Tel1 which phosphorylates a large number of downstream targets (including Rad9, Rad17, Rad53, Rpa1, Xrs1 Com1/Sae2, and Exo1) involved in DNA repair and checkpoint control (Clerici et al., 2005; Roitinger et al., 2015). The ssDNA ends are coated with the replication protein A (RPA), thereby stimulating the recruitment of recombinases [in yeast *via* Rad52; in higher eukaryotes *via* BRCA2 (Krogh and Symington, 2004)]. The recombinase Rad51 (and in meiosis its relative Dmc1; see below) mediates subsequent strand invasion to probe for homologue sequences, assisted and stimulated by a battery of accessory proteins (Sung et al., 2003; Chan et al., 2019). In S/G2, the cell cycle stage during which HR is promoted, the sister chromatid and the chromatids of the homologous chromosome are available as repair templates. Following invasion and successful homology check, the invading strand is elongated and the displaced strand captured by the ssDNA overhang at the DSB site. Subsequently, the elongated strands are ligated yielding a double holiday junction (dHJ) that can lead, after resolution, to restoration of the original chromosome or to a cross-over and therefore a mutual exchange of chromosome arms. In case the sister chromatid has been used as a repair template, such an exchange is genetically neutral; in case a chromatid of the homologous chromosome has been used, such an exchange yields a chimeric chromosome. The latter is the desired repair product during meiosis to support meiotic chromosome disjunction and increase genetic diversity (Ohkura, 2015).

Alternatively, prior to second-end capture, the recombination intermediate can be dismantled by helicases and the invading, now elongated, strand anneals to the DSB site it originated from (also known as SDSA – synthesis dependent strand annealing). Subsequent DNA synthesis and ligation repairs the lesion, with the potential of some genetic information transfer (gene conversion) in case the template strand contained sequence polymorphisms, but without exchange of chromosome arms.

Different pathways have been identified to dismantle dHJs, utilizing structure-specific resolvases like GEN1, MUS81-EM1, or SLX1-SLX4 (or MLH1/3-EXO1 in meiosis; see also below; San-Segundo and Clemente-Blanco, 2020). Alternatively, dHJs can also be dissolved by a complex containing a helicase (BLM, bloom helicase) and a topoisomerase (TOP3-RMI1), to yield intact, but non-recombined chromosomes (Bizard and Hickson, 2014). HR is a conserved process and plants encode all of the important mediators (Knoll et al., 2014).

In this sense, in canonical non-repetitive regions of the genome, HR delivers a more faithful repair outcome with a high likelihood to re-establish the original DNA sequence, while NHEJ leads mostly to short-range deletions and to some extent to insertions and SNPs (Betermier et al., 2014; Liu and Huang, 2014; Ceccaldi et al., 2016).

MEIOSIS

Meiosis is a specific developmental process required for the formation of gametes, carrying the genetic information for the next generation. Meiosis is characterized by two consecutive cell divisions that reduce the genome size by half and by recombination of the paternal and maternal genomes. Novel allelic combinations are created by the mutual exchange of genetic information between parental chromosomes. This depends on meiotic DNA DSBs which are enzymatically induced by the conserved SPO11 protein (together with less conserved partners; Hunter, 2015; Mercier et al., 2015; Robert et al., 2016). About 250–300 DSBs are introduced in each individual meiocyte in Arabidopsis (Edlinger et al., 2011), and they all have to be reliably repaired for successful completion of meiosis. As mentioned, meiotic DSBs are introduced following DNA replication; therefore, cells are in G2-phase with HR being the predominant DNA repair pathway. Meiotic HR is specifically tuned to generate genetic diversity, preferentially using a chromatid of the homologue, and not the sister chromatid, as a repair template [inter-homolog (IH) bias]. Multiple such events along a chromosome ensure that homologous chromosomes recognize each other. At least one IH interaction per chromosome pair has to mature into a cross-over to ensure correct segregation of homologs during the first meiotic division (Gray and Cohen, 2016). In non-repetitive regions, recognition of the homologous partners works very reliably and non-allelic recombination events are not observed. This process is also aided by a meiosis-specific chromosome organization ("bouquet"), clustering telomers (and often also centromeres) to reduce the search space for the ssDNA nucleoprotein filaments (Harper et al., 2004). Genomic loci that are comprised of repetitive sequences, like the rDNA clusters, create a liability during recombination since they can undergo non-allelic exchanges and are a potential source of deletions, duplications, inversions, or translocations (Sasaki et al., 2010).

DSB FORMATION AND REPAIR AT THE rDNA LOCUS

Most of the studies concerning DSB repair at the rDNA region involve the use of induced DSBs by exogenous factors or the use of mutants that perturb the stability of the rDNA (Harding et al., 2015; Sluis et al., 2015; Warmerdam et al., 2016). In plants, a recent study employed CRISPR-Cas9 to induce DSBs at the rDNA locus. This led to a large population of plants each containing a varying number of rDNA repeats ranging from about 20 to 200% of the wild-type copy number (Lopez et al., 2021). While these plants represent a powerful resource to study rDNA dynamics in the future, the actual response to the Cas9-mediated DNA lesions has not yet been studied. In mammalian cells, it has been established that the DNA damage response at the rDNA and within the nucleolus depends on a critical threshold: low levels of DSB formation activate NHEJ, excessive DSB formation within the rDNA is repaired *via* HR, concomitant with transcriptional downregulation and nucleolus re-organization (van Sluis and McStay, 2017).

Studying rDNA repair in a meiotic environment is advantageous since a relatively defined number of endogenous DSBs are formed in a tightly regulated fashion. This allows monitoring DSB repair at the rDNA loci under physiological conditions (Sims et al., 2019). In plants, only a handful of factors are known to be involved in the repair process and stability of the rDNA in somatic and meiotic tissues after DSB formation. The RECQ/TOP3/RMI1 complex partner RMI2, the DNA helicases RTEL1, and FANCJ have been shown to be independently needed for maintaining the stability of the 45S rDNA loci in somatic tissues of some plants (Rohrig et al., 2016; Dorn et al., 2019). Furthermore, several additional studies have shown the importance of the chromatin assembly complex CAF-1in preventing DSB formation at the rDNA loci and maintaining rDNA copy numbers (Mozgová et al., 2010; Pavlistova et al., 2016). In addition, low amounts of 45S rDNA copies have shown to promote genomic instability in a genome-wide manner by generating large genomic re-arrangements (Picart-Picolo et al., 2020; Lopez et al., 2021). In meiosis, c/a-NHEJ factors, such as LIG4 and MRE11, have been shown to be important for DNA repair within the rDNA region, whereas HDA6 and NUC2, which are involved in regulating rDNA transcription and nucleolus integrity, are essential for limiting HR at the rDNA (Sims et al., 2019).

A BALANCE BETWEEN HR AND NHEJ

Studies in human cells, employing artificially induced DSBs, have described a re-organization of the nucleolus and a shift from NHEJ repair to HR upon reaching a certain threshold of DNA damage (van Sluis and McStay, 2017). This is concomitant with the formation of the nucleolar caps (Reynolds et al., 1964) and a shutdown of rRNA transcription while breaks in the rDNA persist. Nucleolar caps have not yet been described in other organisms other than humans and mice. In yeast, sites of DSBs within the rDNA re-localize to an extra nucleolar site for repair by HR (Horigome et al., 2019).

Work performed in *A. thaliana* shows that in physiological conditions, such as meiosis, the DNA lesions in the rDNA are preferentially repaired by NHEJ. The nucleolus creates a HR-refractory zone with strongly reduced numbers of HR events at the NORs (Sims et al., 2019). It is anticipated that sporadic events of HR can still occur, and they may leave noticeable traces, like rDNA unit duplication/amplification/loss (copy number variation) and variable numbers of sequence repeats within the 45S rDNA units. Maintaining this unique HR-refractory domain depends on specific chromatin modifications which are distinct from the meiotic nucleus (Sims et al., 2019). In general, an increase in HR at the rDNA locus leads to the loss of units and reduced cell fitness. This is, for instance, well described in the *FAS1* mutant background, in which an HR-dependent shortening of the NORs has been reported (Mozgová et al., 2010; Muchova et al., 2015). Nevertheless, HR events likely

also occur in wild-type plants, since a dramatic divergence in rDNA copy numbers is detectable within tens of generations, suggesting the presence of low recurring HR events that lead to a change in NOR length (Rabanal et al., 2017a). Furthermore, the presence of long segments of identical rDNA units is an indication of homogenization events at the rDNA locus, likely mediated by HR repair (Copenhaver and Pikaard, 1996; Sims et al., 2021).

We suggest that DSBs within the rDNA, occurring at physiological levels (e.g., as a results of transcription/replication collisions or generated during the meiotic program), are repaired *via* NHEJ, preserving rDNA unit numbers and unit-internal repeat structures, but at the cost of producing errors. Currently, it is unclear whether there is a preference toward canonical or alternative NHEJ pathways (Sims et al., 2019). However, there are indications that both pathways are necessary for repairing lesions within the rDNA since LIG4 and MRE11 have an equal impact on rDNA stability. It is important to mention that DSBs generated within the rDNA by transcription/replication conflicts or during meiosis are still rare events (Sims et al., 2019).

In general, the errors produced by the NHEJ pathways have the potential to generate sequence diversity between the rDNA units, and one would expect for them to accumulate at the transcriptional start and termination sites of each unit, due to selection against mutations in the portions of the rDNA that yield rRNA integrated into ribosomes. In fact, the highest number of SNPs and InDels is found in the external transcribed sequences (ETS), particularly close to the promoter and terminator regions (Chandrasekhara et al., 2016; Rabanal et al., 2017b). In contrast, very few SNPs/InDels are found within the portions transcribing the ribosomal RNA subunits (18S, 5.8S and 25S). This correlates well with the suggested high levels of transcriptional stress in the rDNA, with purifying selection acting on the regions transcribing rRNA subunits and with DNA lesions being repaired *via* NHEJ.

Imbalanced accumulation of polymorphisms is also apparent between the two NORs of *A. thaliana*. A recent study combining long- and short-read sequencing technologies to define the nucleotide composition and organization of 405 individual rDNA units of NOR2 of ecotype Col-0 identified less SNPs/InDels on the transcriptionally less active NOR2, than on NOR4 (Sims et al., 2021). To display the sequence diversity of these 405 rDNA units, we utilized their data and generated a phylogenetic network. For the analysis, we excluded the highly repetitive region of the *Sal*I repeat boxes from each unit. The TCS network was inferred (**Supplementary Figure S1**) using the integrated method of the TCS approach (Templeton et al., 1992; Clement et al., 2002), which is based on the concept of statistical parsimony in PopArt (Leigh and Bryant, 2015). The network shows a lack of phylogenetic structure in the data indicating that a lot of parallel and reverse mutations obscure the relations between the units and that the conservative nature of the rDNA units in general may possibly mask local phylogenetic information. To address this latter point, we repeated the TCS analysis for short stretches of rDNA units (represented in the 59 BACs as published in Sims et al., 2021). Indeed, the majority of the BACs show a clear tree-like structure, with only very little reticulation. Thus, locally, the evolutionary process follows a classical tree-like pattern. Moreover, directly adjacent units on a BAC tend to be next to each other in the tree (data and visualization available upon request). The contigs identified in (Sims et al., 2021) provide additional evidence of tree-like evolution of the NOR2 region (**Figure 1**). Though the tree-like relation breaks down, the more rDNA units are analyzed due to multiple identical units occurring along the NOR2 region.

A plausible explanation for the higher abundance of SNPs/InDels on NOR4 could be derived from the fact that in the ecotype Col-0, NOR4 is transcriptionally active in all analyzed tissues, while NOR2 is selectively silenced during development, and it is only active in certain tissue types (Chandrasekhara et al., 2016; Rabanal et al., 2017a). Transcriptional stress *per se* is a prime source of DSBs, and the rDNA is considered a hotspot of transcription and replication stress (Takeuchi et al., 2003). Since the pattern of NOR expression varies greatly among Arabidopsis ecotypes with some expressing predominantly one and some the other NOR (and some both), it would be interesting to analyze whether rDNA polymorphisms are positively correlated with transcriptional activity in different ecotypes.

It is interesting to speculate that the nucleolus represents an HR-refractory sub-compartment within the nucleus during meiosis (and after pre-meiotic DNA replication). As stated above, both NORs are transcriptionally activated in order to be recruited to the nucleolus and embedded in its HR-refractory zone (Sims et al., 2019). Perturbing the rDNA transcriptional activity or the nucleolar architecture generates an imbalance in the rDNA protective mechanism. In this sense, rDNA transcriptional activation, and subsequent recruitment into the nucleolus, could be a key regulatory mechanism to determine the mode of rDNA repair after DNA damage. The recruitment into the nucleolus following transcription is a conserved feature of rDNA (Pontvianne et al., 2013; Sims et al., 2019).

The protective mechanisms surrounding the 45 rDNA regions could not be limited to the nucleolus itself, since in certain tissues, the majority of 45S rDNA genes are not transcribed and excluded from the nucleolus. Inactive NOR4 rDNA genes are generally located at the nucleolar periphery, whereas NOR2 rDNA genes are completely excluded from the nucleolus area (Pontvianne et al., 2013).

It remains unknown whether the nucleolus plays a protective role in other plant tissues or in other organisms. In human cells, massive DNA damage of the rDNA leads to the formation of nucleolar caps. It has been shown that these caps contain broken rDNA which then becomes available to the HR machinery of the nucleus (Sluis et al., 2015), lending support to the idea that the nucleolus represents a general and conserved HR-refractory sub-compartment. Hence, the nucleolus might have the intrinsic property of excluding HR-related proteins. In line with this idea, the nucleolar proteomes of *Arabidopsis* and of humans showed no evidence of the presence of HR proteins (Andersen et al., 2005; Montacié et al., 2017).

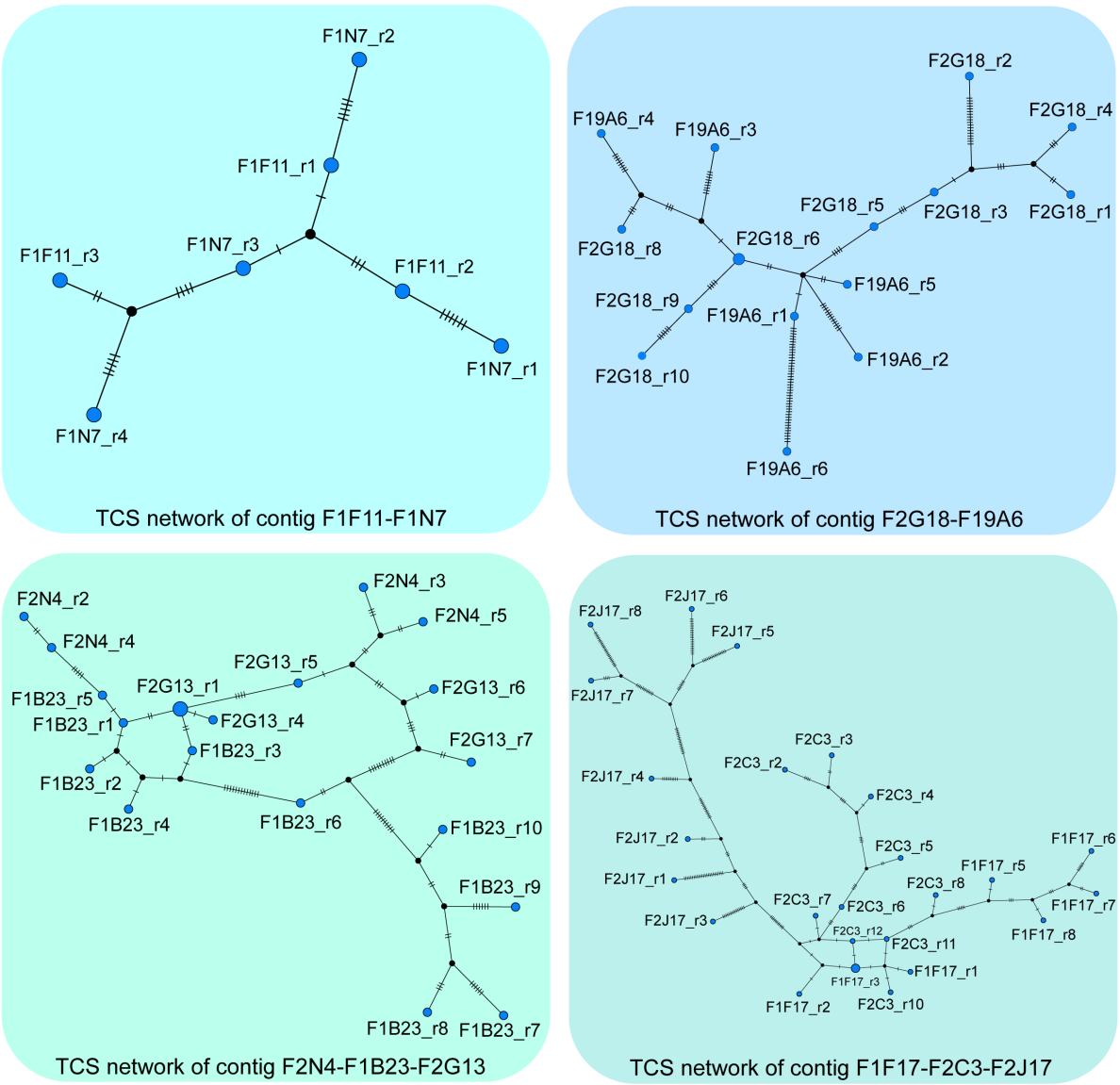


FIGURE 1 | The TCS networks were inferred from rDNA units of the contigs F2N4-F1B23-F2G13, F1F17-F2C3-F2J17, F1F11-F1N27 and F2G18-F19A6 identified in (Sims et al., 2021). The first unit of BAC F2N4 and the fourth unit of BAC F1F11 were excluded from the analysis. Furthermore, the highly repetitive *SalI* boxes were not taken into account for this data analysis. In the network, each node represents a unique sequence and its size is proportional to its frequency within the data. Short vertical bars on the lines connecting similar sequences represent the number of variations between them. Visualizations of the analyses of the 59 individual BACs are available upon request.

CONTROLLING SEQUENCE HOMOGENEITY AND HETEROGENEITY

The repetitive rDNA loci are considered intrinsically unstable genomic regions since they are prone to various types of DNA damage and repair events. The sequence variations identified in individual rDNA units (Chandrasekhara et al., 2016; Havlová et al., 2016; Rabanal et al., 2017b; Sims et al., 2021) may represent past DNA repair events following an error-prone pathway (NHEJ). Taking into consideration rDNA copy numbers, it is possible to evaluate the history of DNA repair events following an error-free pathway (HR). While sequence variations of rDNA units can readily be analyzed in individual plants, the evaluation of rDNA unit copy number variations demands the analysis of large populations or multiple successive generations (Rabanal et al., 2017b; Sims et al., 2021).

The rDNA copy number can also be considered as a genetic trait and studied in pedigrees. Indeed, when analyzing the trait of "rDNA copy number" over a few generations (two generations in F2s, about eight in recombinant inbred lines – RILs), it appears stable enough that it can be mapped

to either NOR in segregating populations. Moreover, the apparent lack of F1-like rDNA copy number phenotypes after several generations of inbreeding in a RIL population further strengthens the notion that the NORs of homologous chromosomes rarely recombine, in agreement with the idea that the nucleolus is a HR-refractory sub-compartment of the nucleus. Importantly, analyzing a wider generational time window, a progressive divergence in the number of rDNA units in single seed descent *A. thaliana* plants was apparent within tens of generations. As a consequence of this unstable inheritance, and in spite of the fact that rDNA unit numbers vary considerably in natural *A. thaliana* populations (Davison et al., 2007; Long et al., 2013), genome-wide association studies failed to map the source of the variation to either of the NORs (Long et al., 2013). This means that rare events of HR might take place, only evident in large populations or when observing multiple successive generations, which lead to dramatic rDNA unit number variations.

In contrast, within plants containing a small amount of 45S rDNA units, the rDNA gene copy numbers can be quickly restored and amplified to wild-type levels. This indicated that there is mechanism in place to restore the 45S rDNA copy numbers within individuals with low amount of rRNA genes (Pavlistova et al., 2016).

FUNCTIONAL AND EVOLUTIONARY IMPACT OF rDNA HETEROGENEITY

Different studies on different organisms (including humans, flies, worms, and plants) have shown that the rDNA genes are not identical either within or among individuals of the same species (Gonzalez et al., 1985; Keller et al., 2006; Stage and Eickbush, 2007; Pillet et al., 2012; Bik et al., 2013). Nevertheless, certain SNPs/Indels are stable and abundant enough in either of the two NORs in *A. thaliana* that they qualify to serve as reporters of NOR-specific expression (Chandrasekhara et al., 2016; Rabanal et al., 2017a). There is unequivocal evidence of selective silencing of one of the two NORs during vegetative development in *A. thaliana*, with the majority of all rRNAs being generated just from one locus. Nevertheless, there is also compelling evidence that (1) there is selective transcriptional activation of certain rDNA units from the otherwise silenced NOR locus in some tissues and (2) that not all rDNA units at the active NOR locus are transcribed at the same time (Pontvianne et al., 2013). rDNA unit variants are not randomly distributed along the NORs [at least established for NOR2 (Sims et al., 2021)], but rather in variant sub-clusters that share certain SNPs/Indels combinations. In some instances, these blocks of corresponding rDNA units are disrupted by rDNA units of a different subtype, but still are transcriptionally co-regulated (Sims et al., 2021). These findings provide a solid base for the future dissection of the fine-tuned regulation of expression of rDNA variant units within a NOR.

It is tempting to speculate that the heterogeneous population of rDNA units and their regulated expression has an important impact on protein translation. The presence of expressed rRNA variants has been shown in various organisms by analyzing total RNA (Kuo et al., 1996; Carranza et al., 1999; Tseng et al., 2008; Rabanal, Mandáková, et al., 2017; Simon et al., 2018). Some of the identified SNPs/InDels were located within the genic regions that encode the 25S and 18S rRNA subunits which are integrated into ribosomes. Furthermore, several studies demonstrated that variant rRNAs are incorporated into polysomes, the ribosomal fraction actively committed to protein translation (Gonzalez et al., 1988; Cloix et al., 2002; Mentewab et al., 2011; Dimarco et al., 2012; Kurylo et al., 2018; Parks et al., 2018; Sims et al., 2021). Interestingly, various of these rRNA gene variants are differentially expressed in a tissue-specific manner. Furthermore, some sequence variations are located in regions that could have a functional impact on the biology of ribosomes. Most of the genic rRNA sequence variations are located in ribosomal expansion segments, that vary greatly between species, but could have an important impact on interacting proteins. A few SNPs/InDels occur in the rRNA core domains. For instance, one G to T transition present in *A. thaliana* is located between the H74 and H88 ribosomal domains at the peptidyl transferase site and thus has the potential to impact ribosomal translation directly. In the parasite *Plasmodium,* two structurally distinct 18S rRNAs are differentially expressed during its life cycle (Gunderson et al., 1987; Waters et al., 1989). And more recently, the expansion segment 9S has been shown to selectively recruit *Hox9* mRNA *via* its 5' UTR stem-loop (Leppek et al., 2020).

In addition, it has been shown that in the bacteria *Vibrio vulnificus*, from a heterogeneous population of ribosomes, it primarily uses ribosomes containing a particular ribosomal RNA variant to translate stress-related mRNA (Song et al., 2019; Leppek et al., 2020). Similarly, in *Escherichia coli,* a specific branch of the stress response utilizes a truncated rRNA to selectively bias translation of stress response proteins (Vesper et al., 2011).

CONCLUSIONS/PERSPECTIVES

The most current sequencing technologies, in combination with detailed and large-scale population studies and in-depth analyses of ribosomal RNA variants, have generated novel and exciting insights.

Without any doubt, the NORs cannot be regarded as stable, rigid domains comprised of – nearly – identical rDNA units anymore, but rather as dynamic chromosomal loci with high variation in rDNA unit copy numbers and sequences. We consider a delicate balance of the HR and NHEJ DNA repair mechanisms to be responsible for the dynamic nature of the NORs. We suggest that frequent (meiotic) NHEJ events generate abundant SNPs and InDels within the rDNA, resulting in a heterogeneous population of rDNA units. We also propose that rare HR events dynamically change rDNA unit numbers. The latter may only be observed in large populations and/or over many generations (**Figure 2**).

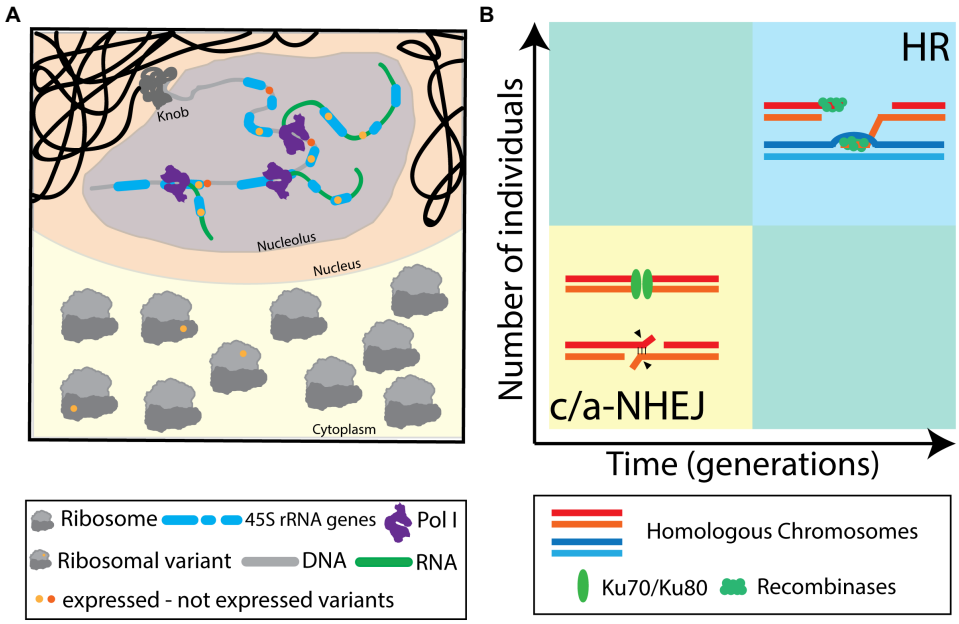


FIGURE 2 | (A) Illustration of transcription of variant rRNAs from non-identical 45 rDNA units and their integration into translating ribosomes. The concept of heterogeneous ribosomes has been introduced considering different protein compositions. Here, this concept is extended, also considering different rRNA variants. **(B)** Diagram illustrating the occurrence of non-homologous end joining (NHEJ) and homologous recombination (HR) as DNA repair modes in the highly repetitive nucleolus organizing regions (NORs) during meiosis. NHEJ is considered to be the commonly deployed repair pathway, leading to short-range repair scars in the affected rDNA units, contributing to sequence heterogeneity and preserving the integrity of the NOR. Meiotic DNA repair events *via* HR are considered rare events and will only be evident in large populations, over multiple generations. HR may contribute to NOR size variability and rDNA unit homogenization.

Furthermore, the ribosomes are no longer seen as invariant machines that translate proteins from available mRNAs but rather as a heterogeneous population of ribonuclear complexes, differing in rRNA and protein composition, with defined functions controlling protein translation (**Figure 2**).

In the future, it will be interesting to generate the detailed sequence information of NORs from various organisms, ecotypes, and individuals. Knowledge of the precise rDNA unit sequences will allow detailed analyses of the dynamic changes of the NORs, their (potentially context dependent) differential transcriptional regulation, and the integration of rRNA variants into actively translating ribosomes (with the potential to impact protein translation).

AUTHOR CONTRIBUTIONS

JS, FR, and PS: conceptualization. JS and CE: data analysis and visualization. PS, AH, and FR: funding acquisition. All authors contributed to the writing of the article and approved the submitted version.

FUNDING

We thank the Austrian Science Fund FWF (SFB F34, DK W1238-B20, I 3685-B25 to PS; W1207-B09 to AH), the European Union (FP7-ITN 606956 to PS), the Human Frontiers Science Program (HFSP Long-Term Fellowship LT000819/2018-L to FR), and the Max Planck Society (FR) for funding.

ACKNOWLEDGMENTS

We thank the members of the PS and AH labs for critical discussions and constructive feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.773052/full#supplementary-material

Supplementary Figure S1 | TCS network was inferred from all sequences of the rDNA units of 59 BACs without the region of the highly repetitive *SalI* boxes. To build the network, the integrated method of the TCS approach, which is based on the concept of statistical parsimony, in PopArt was used. In the network, each node represents a unique sequence, its size is proportional to its frequency within the data, and its color indicates from which BAC a rDNA originated from. In other words, a big, multicolored node is a collection of different rDNA units which are identical and come from different BACs (and therefore different locations within the NOR2). Short vertical bars on the lines connecting similar sequences represent the number of variations between them. Furthermore, sequences, not present in the data, were inferred and represented as small black dots. The TCS network contains 39 nodes which comprise more than one rDNA unit. In total, 238 rDNA units occur in the 39 nodes. The biggest node includes 62 rDNA units.

REFERENCES

Abou-Ellail, M., Cooke, R., and Sáez-Vásquez, J. (2011). Variations in a team: major and minor variants of *Arabidopsis thaliana* rDNA genes. *Nucleus* 2, 294–299. doi: 10.4161/nucl.2.4.16561

Andersen, J. S., Lam, Y. W., Leung, A. K. L., Ong, S.-E., Lyon, C. E., Lamond, A. I., et al. (2005). Nucleolar proteome dynamics. *Nature* 433, 77–83. doi: 10.1038/nature03207

Beck, C., Robert, I., Reina-San-Martin, B., Schreiber, V., and Dantzer, F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: focus on PARP1, PARP2 and PARP3. *Exp. Cell Res.* 329, 18–25. doi: 10.1016/j.yexcr.2014.07.003

Betermier, M., Bertrand, P., and Lopez, B. S. (2014). Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet.* 10:e1004086. doi: 10.1371/journal.pgen.1004086

Bik, H. M., Fournier, D., Sung, W., Bergeron, R. D., and Thomas, W. K. (2013). Intra-genomic variation in the ribosomal repeats of nematodes. *PLoS One* 8:e78230. doi: 10.1371/journal.pone.0078230

Bizard, A. H., and Hickson, I. D. (2014). The dissolution of double Holliday junctions. *Cold Spring Harb. Perspect. Biol.* 6:a016477. doi: 10.1101/cshperspect.a016477

Blackford, A. N., and Jackson, S. P. (2017). ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol. Cell* 66, 801–817. doi: 10.1016/j.molcel.2017.05.015

Bleuyard, J.-Y. Y., Gallego, M. E., and White, C. I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair* 5, 1–12. doi: 10.1016/j.dnarep.2005.08.017

Carranza, S., Baguñà, J., and Riutort, M. (1999). Origin and evolution of paralogous rRNA gene clusters within the flatworm family Dugesiidae (Platyhelminthes, Tricladida). *J. Mol. Evol.* 49, 250–259. doi: 10.1007/PL00006547

Castel, S. E., Ren, J., Bhattacharjee, S., Chang, A. Y., Sánchez, M., Valbuena, A., et al. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell* 159, 572–583. doi: 10.1016/j.cell.2014.09.031

Ceccaldi, R., Rondinelli, B., and D'Andrea, A. D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* 26, 52–64. doi: 10.1016/j.tcb.2015.07.009

Chan, Y. L., Zhang, A., Weissman, B. P., and Bishop, D. K. (2019). RPA resolves conflicting activities of accessory proteins during reconstitution of Dmc1-mediated meiotic recombination. *Nucleic Acids Res.* 47, 747–761. doi: 10.1093/nar/gky1160

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506. doi: 10.1038/nrm.2017.48

Charbonnel, C., Allain, E., Gallego, M. E., and White, C. I. (2011). Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. *DNA Repair* 10, 611–619. doi: 10.1016/j.dnarep.2011.04.002

Cheng, Q., Barboule, N., Frit, P., Gomez, D., Bombarde, O., Couderc, B., et al. (2011). Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks. *Nucleic Acids Res.* 39, 9605–9619. doi: 10.1093/nar/gkr656

Clement, M., Snell, Q., Walke, P., Posada, D., and Crandall, K. (2002). "TCS: estimating gene genealogies." in *Proceedings of the 16th International Parallel Distributed Processing Symposium*; April 15-19, 2002; Lauderdale, FL, USA, 184.

Clerici, M., Mantiero, D., Lucchini, G., and Longhese, M. P. (2005). The *Saccharomyces cerevisiae* Sae2 protein promotes resection and bridging of double strand break ends. *J. Biol. Chem.* 280, 38631–38638. doi: 10.1074/jbc.M508339200

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.-C., et al. (2002). Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Res.* 12, 132–144. doi: 10.1101/gr.181301

Copenhaver, G. P., and Pikaard, C. S. (1996). Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* 9, 273–282. doi: 10.1046/j.1365-313X.1996.09020273.x

Davison, J., Tyagi, A., and Comai, L. (2007). Large-scale polymorphism of heterochromatic repeats in the DNA of *Arabidopsis thaliana*. *BMC Plant Biol.* 7:44. doi: 10.1186/1471-2229-7-44

Dimarco, E., Cascone, E., Bellavia, D., and Caradonna, F. (2012). Functional variants of 5S rRNA in the ribosomes of common sea urchin *Paracentrotus lividus*. *Gene* 508, 21–25. doi: 10.1016/j.gene.2012.07.067

Dorn, A., Feller, L., Castri, D., Röhrig, S., Enderle, J., Herrmann, N. J., et al. (2019). An Arabidopsis FANCJ helicase homologue is required for DNA crosslink repair and rDNA repeat stability. *PLoS Genet.* 15:e1008174. doi: 10.1371/journal.pgen.1008174

Durdikova, K., and Chovanec, M. (2017). Regulation of non-homologous end joining via post-translational modifications of components of the ligation step. *Curr. Genet.* 63, 591–605. doi: 10.1007/s00294-016-0670-7

Edlinger, B., Schlogelhofer, P., Schlögelhofer, P., Schlogelhofer, P., and Schlögelhofer, P. (2011). Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants. *J. Exp. Bot.* 62, 1545–1563. doi: 10.1093/jxb/erq421

Fell, V. L., and Schild-Poulter, C. (2015). The Ku heterodimer: function in DNA repair and beyond. *Mutat. Res. Rev. Mutat. Res.* 763, 15–29. doi: 10.1016/j.mrrev.2014.06.002

French, S. L., Sikes, M. L., Hontz, R. D., Osheim, Y. N., Lambert, T. E., El Hage, A., et al. (2011). Distinguishing the roles of topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell. Biol.* 31, 482–494. doi: 10.1128/MCB.00589-10

Friesner, J., and Britt, A. B. (2003). Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *Plant J.* 34, 427–440. doi: 10.1046/j.1365-313X.2003.01738.x

García-Muse, T., and Aguilera, A. (2016). Transcription-replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* 17, 553–563. doi: 10.1038/nrm.2016.88

Gonzalez, I. L., Gorski, J. L., Campen, T. J., Dorney, D. J., Erickson, J. M., Sylvester, J. E., et al. (1985). Variation among human 28S ribosomal RNA genes. *Proc. Natl. Acad. Sci. U. S. A.* 82, 7666–7670. doi: 10.1073/pnas.82.22.7666

Gonzalez, I. L., Sylvester, J. E., and Schmickel, R. D. (1988). Human 28S ribosomal RNA sequence heterogeneity. *Nucleic Acids Res.* 16, 10213–10224. doi: 10.1093/nar/16.21.10213

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., et al. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388, 492–495. doi: 10.1038/41358

Gray, S., and Cohen, P. E. (2016). Control of meiotic crossovers: from double-strand break formation to designation. *Annu. Rev. Genet.* 50, 175–210. doi: 10.1146/annurev-genet-120215-035111

Grummt, I., and Pikaard, C. S. (2003). Epigenetic silencing of RNA polymerase I transcription. *Nat. Rev. Mol. Cell Biol.* 4, 641–649. doi: 10.1038/nrm1171

Gunderson, J. H., Sogin, M. L., Wollett, G., Hollingdale, M., de la Cruz, V. F., Waters, A. P., et al. (1987). Structurally distinct, stage-specific ribosomes occur in Plasmodium. *Science* 238, 933–937. doi: 10.1126/science.3672135

Harding, S. M., Boiarsky, J. A., and Greenberg, R. A. (2015). ATM dependent silencing links nucleolar chromatin reorganization to DNA damage recognition. *Cell Rep.* 13, 251–259. doi: 10.1016/j.celrep.2015.08.085

Harper, L., Golubovskaya, I., and Cande, W. Z. (2004). A bouquet of chromosomes. *J. Cell Sci.* 117, 4025–4032. doi: 10.1242/jcs.01363

Havlová, K., Dvořáčková, M., Peiro, R., Abia, D., Mozgová, I., Vansáčová, L., et al. (2016). Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol. Biol.* 92, 457–471. doi: 10.1007/s11103-016-0524-1

Hays, J. B. (2002). *Arabidopsis thaliana*, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair* 1, 579–600. doi: 10.1016/S1568-7864(02)00093-9

Horigome, C., Unozawa, E., Ooki, T., and Kobayashi, T. (2019). Ribosomal RNA gene repeats associate with the nuclear pore complex for maintenance after DNA damage. *PLoS Genet.* 15:e1008103. doi: 10.1371/journal.pgen.1008103

Hunter, N. (2015). Meiotic recombination: the essence of heredity. *Cold Spring Harb. Perspect. Biol.* 7:a016618. doi: 10.1101/cshperspect.a016618

Keller, I., Chintauan-Marquier, I. C., Veltsos, P., and Nichols, R. A. (2006). Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics* 174, 863–874. doi: 10.1534/genetics.106.061341

Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: mechanisms and evolutionary consequences. *Chromosom. Res.* 22, 191–201. doi: 10.1007/s10577-014-9415-y

Krogh, B. O., and Symington, L. S. (2004). Recombination proteins in yeast. *Annu. Rev. Genet.* 38, 233–271. doi: 10.1146/annurev.genet.38.072902.091500

Kuo, B. A., Gonzalez, I. L., Gillespie, D. A., and Sylvester, J. E. (1996). Human ribosomal RNA variants from a single individual and their expression in different tissues. *Nucleic Acids Res.* 24, 4817–4824. doi: 10.1093/nar/24.23.4817

Kurylo, C. M., Parks, M. M., Juette, M. F., Thibado, J. K., Vincent, C. T., Blanchard, S. C., et al. (2018). Endogenous rRNA sequence variation can regulate stress response gene expression and phenotype article endogenous rRNA sequence variation can regulate stress response gene expression and phenotype. *Cell Rep.* 25, 236.e6–248.e6. doi: 10.1016/j.celrep.2018.08.093

Layat, E., Sáez-Vásquez, J., and Tourmente, S. (2012). Regulation of pol I-transcribed 45S rDNA and pol III-transcribed 5S rDNA in arabidopsis. *Plant Cell Physiol.* 53, 267–276. doi: 10.1093/pcp/pcr177

Lees-Miller, S. P., and Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* 85, 1161–1173. doi: 10.1016/j.biochi.2003.10.011

Leigh, J. W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Leppek, K., Fujii, K., Quade, N., Susanto, T. T., Boehringer, D., Lenarčič, T., et al. (2020). Gene- and species-specific hox mRNA translation by ribosome expansion segments. *Mol. Cell* 80, 980.e13–995.e13. doi: 10.1016/j.molcel.2020.10.023

Liu, T., and Huang, J. (2014). Quality control of homologous recombination. *Cell. Mol. Life Sci.* 71, 3779–3797. doi: 10.1007/s00018-014-1649-5

Lobrich, M., and Jeggo, P. (2017). A process of resection-dependent nonhomologous end joining involving the goddess Artemis. *Trends Biochem. Sci.* 42, 690–701. doi: 10.1016/j.tibs.2017.06.011

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884–890. doi: 10.1038/ng.2678

Lopez, F. B., Fort, A., Tadini, L., Probst, A. V., McHale, M., Friel, J., et al. (2021). Gene dosage compensation of rRNA transcript levels in *Arabidopsis thaliana* lines with reduced ribosomal gene copy number. *Plant Cell* 33, 1135–1150. doi: 10.1093/plcell/koab020

Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants – from models to crops. *Front. Plant Sci.* 6:885. doi: 10.3389/fpls.2015.00885

McVey, M., and Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* 24, 529–538. doi: 10.1016/j.tig.2008.08.007

Meek, K., Gupta, S., Ramsden, D. A., and Lees-Miller, S. P. (2004). The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* 200, 132–141. doi: 10.1111/j.0105-2896.2004.00162.x

Mentewab, A. B., Jacobsen, M. J., and Flowers, R. A. (2011). Incomplete homogenization of 18 S ribosomal DNA coding regions in *Arabidopsis thaliana*. *BMC. Res. Notes* 4:93. doi: 10.1186/1756-0500-4-93

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., Grelon, M., Mezard, C., et al. (2015). The molecular biology of meiosis in plants. *Annu. Rev. Plant Biol.* 66, 297–327. doi: 10.1146/annurev-arplant-050213-035923

Montacié, C., Durut, N., Opsomer, A., Palm, D., Comella, P., Picart, C., et al. (2017). Nucleolar proteome analysis and proteasomal activity assays reveal a link between nucleolus and 26S proteasome in *A. thaliana*. *Front. Plant Sci.* 8:1815. doi: 10.3389/fpls.2017.01815

Moss, T., Langlois, F., Gagnon-Kugler, T., and Stefanovsky, V. (2007). A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* 64, 29–49. doi: 10.1007/s00018-006-6278-1

Mozgová, I., Mokroš, P., and Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45s rDNA in *Arabidopsis thaliana*. *Plant Cell* 22, 2768–2780. doi: 10.1105/tpc.110.076182

Muchova, V., Amiard, S., Mozgova, I., Dvorackova, M., Gallego, M. E., White, C., et al. (2015). Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants. *Plant J.* 81, 198–209. doi: 10.1111/tpj.12718

Murata, M., Heslop-Harrison, J. S., and Motoyoshi, F. (1997). Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence in situ hybridization with cosmid clones. *Plant J.* 12, 31–37. doi: 10.1046/j.1365-313X.1997.12010031.x

Ohkura, H. (2015). Meiosis: an overview of key differences from mitosis. *Cold Spring Harb. Perspect. Biol.* 7:a015859. doi: 10.1101/cshperspect.a015859

Parks, M. M., Kurylo, C. M., Dass, R. A., Bojmar, L., Lyden, D., Vincent, C. T., et al. (2018). Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci. Adv.* 4:eaao0665. doi: 10.1126/sciadv.aao0665

Pavlistova, V., Dvorackova, M., Jez, M., Mozgova, I., Mokros, P., and Fajkus, J. (2016). Phenotypic reversion in fas mutants of *Arabidopsis thaliana* by reintroduction of FAS genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J.* 88, 411–424. doi: 10.1111/tpj.13257

Picart-Picolo, A., Grob, S., Picault, N., Franek, M., Llauro, C., Halter, T., et al. (2020). Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. *Genome Res.* 30, 1583–1592. doi: 10.1101/gr.261586.120

Pillet, L., Fontaine, D., and Pawlowski, J. (2012). Intra-genomic ribosomal RNA polymorphism and morphological variation in elphidium macellum suggests inter-specific hybridization in Foraminifera. *PLoS One* 7:e32373. doi: 10.1371/journal.pone.0032373

Polo, S. E., and Jackson, S. P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* 25, 409–433. doi: 10.1101/gad.2021311

Pontvianne, F., Abou-Ellail, M., Douet, J., Comella, P., Matia, I., Chandrasekhara, C., et al. (2010). Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* 6:e1001225. doi: 10.1371/journal.pgen.1001225

Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S. E., et al. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. *Genes Dev.* 26, 945–957. doi: 10.1101/gad.182865.111

Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgova, I., Hassel, C., Pontes, O. M. F., et al. (2013). Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* 27, 1545–1550. doi: 10.1101/gad.221648.113

Rabanal, F. A., Mandáková, T., Soto-Jiménez, L. M., Greenhalgh, R., Parrott, D. L., Lutzmayer, S., et al. (2017a). Epistatic and allelic interactions control expression of ribosomal RNA gene clusters in *Arabidopsis thaliana*. *Genome Biol.* 18:75. doi: 10.1186/s13059-017-1209-z

Rabanal, F. A., Nizhynska, V., Mandáková, T., Novikova, P. Y., Martin, A., Lysak, M. A., et al. (2017b). Unstable inheritance of 45S rRNA genes in *Arabidopsis thaliana*. *G3* 7, 1201–1209. doi: 10.1534/g3.117.040204

Reynolds, R. C., Montgomery, P. O. B., and Hughes, B. (1964). Nucleolar "caps" produced by actinomycin D. *Cancer Res.* 24, 1269–1277.

Ritossa, F. M., and Spiegelman, S. (1965). Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 53, 737–745. doi: 10.1073/pnas.53.4.737

Robert, T., Vrielynck, N., Mézard, C., de Massy, B., Grelon, M., Mezard, C., et al. (2016). A new light on the meiotic DSB catalytic complex. *Semin. Cell Dev. Biol.* 54, 165–176. doi: 10.1016/j.semcdb.2016.02.025

Rohrig, S., Schropfer, S., Knoll, A., Puchta, H., Röhrig, S., Schröpfer, S., et al. (2016). The RTR complex partner RMI2 and the DNA helicase RTEL1 are both independently involved in preserving the stability of 45S rDNA repeats in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006394. doi: 10.1371/journal.pgen.1006394

Roitinger, E., Hofer, M., Kocher, T., Pichler, P., Novatchkova, M., Yang, J., et al. (2015). Quantitative phosphoproteomics of the ATM and ATR dependent DNA damage response in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 14, 556–571. doi: 10.1074/mcp.M114.040352

San-Segundo, P. A., and Clemente-Blanco, A. (2020). Resolvases, dissolvases, and helicases in homologous recombination: clearing the road for chromosome segregation. *Genes* 11:71. doi: 10.3390/genes11010071

Sasaki, M., Lange, J., and Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. *Nat. Rev. Mol. Cell Biol.* 11, 182–195. doi: 10.1038/nrm2849

Seol, J. H., Shim, E. Y., and Lee, S. E. (2018). Microhomology-mediated end joining: good, bad and ugly. *Mutat. Res.* 809, 81–87. doi: 10.1016/j.mrfmmm.2017.07.002

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., et al. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Res.* 46, 3019–3033. doi: 10.1093/nar/gky163

Sims, J., Copenhaver, G. P., and Schlögelhofer, P. (2019). Meiotic DNA repair in the nucleolus employs a nonhomologous end-joining mechanism. *Plant Cell* 31, 2259–2275. doi: 10.1105/tpc.19.00367

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the Arabidopsis NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12:387. doi: 10.1038/s41467-020-20728-6

Sluis, M. V., McStay, B., van Sluis, M., and McStay, B. (2015). A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev.* 29, 1151–1163. doi: 10.1101/gad.260703.115

Song, W., Joo, M., Yeom, J. H., Shin, E., Lee, M., Choi, H. K., et al. (2019). Divergent rRNAs as regulators of gene expression at the ribosome level. *Nat. Microbiol.* 4, 515–526. doi: 10.1038/s41564-018-0341-1

Stage, D. E., and Eickbush, T. H. (2007). Sequence variation within the rRNA gene loci of 12 Drosophila species. *Genome Res.* 17, 1888–1897. doi: 10.1101/gr.6376807

Sung, P., Krejci, L., Van Komen, S., and Sehorn, M. G. (2003). Rad51 recombinase and recombination mediators. *J. Biol. Chem.* 278, 42729–42732. doi: 10.1074/jbc.R300027200

Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev.* 17, 1497–1506. doi: 10.1101/gad.1085403

Templeton, A. R., Crandall, K. A., and Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics* 132, 619–633. doi: 10.1093/genetics/132.2.619

Templeton, G. W., and Moorhead, G. B. (2005). The phosphoinositide-3-OH-kinase-related kinases of *Arabidopsis thaliana*. *EMBO Rep.* 6, 723–728. doi: 10.1038/sj.embor.7400479

Tseng, H., Chou, W., Wang, J., Zhang, X., Zhang, S., and Schultz, R. M. (2008). Mouse ribosomal RNA genes contain multiple differentially regulated variants. *PLoS One* 3:e1843. doi: 10.1371/journal.pone.0001843

van Sluis, M., and McStay, B. (2017). Nucleolar reorganization in response to rDNA damage. *Curr. Opin. Cell Biol.* 46, 81–86. doi: 10.1016/j.ceb.2017.03.004

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., et al. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* 147, 147–157. doi: 10.1016/j.cell.2011.07.047

Wallace, H., and Birnstiel, M. L. (1966). Ribosomal cistrons and the nucleolar organizer. *Biochim. Biophys. Acta* 114, 296–310. doi: 10.1016/0005-2787(66)90311-X

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., et al. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res.* 34, 6170–6182. doi: 10.1093/nar/gkl840

Warmerdam, D. O. O., van den Berg, J., Medema, R. H. H., van den Berg, J., Medema, R. H. H., van den Berg, J., et al. (2016). Breaks in the 45S rDNA lead to recombination-mediated loss of repeats. *Cell Rep.* 14, 2519–2527. doi: 10.1016/j.celrep.2016.02.048

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* 24, 437–440. doi: 10.1016/S0968-0004(99)01460-7

Waters, A. P., Syin, C., and McCutchan, T. F. (1989). Developmental regulation of stage-specific ribosome populations in Plasmodium. *Nature* 342, 438–440. doi: 10.1038/342438a0

Wright, W. D., Shah, S. S., and Heyer, W. D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10524–10535. doi: 10.1074/jbc.TM118.000372

Wu, Y., Gao, T., Wang, X., Hu, Y., Hu, X., Hu, Z., et al. (2014). TALE nickase mediates high efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. *Biochem. Biophys. Res. Commun.* 446, 261–266. doi: 10.1016/j.bbrc.2014.02.099

Xu, Y., and Xu, D. (2020). Repair pathway choice for double-strand breaks. *Essays Biochem.* 64, 765–777. doi: 10.1042/EBC20200007

Yoshiyama, K. O. (2016). Recent progress in research on DNA damage responsesin animals and plants. *Genes Genet. Syst.* 90, 185–186. doi: 10.1266/ggs.15-10001

Yoshiyama, K. O., Sakaguchi, K., and Kimura, S. (2013). DNA damage response in plants: conserved and variable response compared to animals. *Biology* 2, 1338–1356. doi: 10.3390/biology2041338

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

REVIEW
published: 21 December 2021
doi: 10.3389/fpls.2021.797348

Personal Perspectives on Plant Ribosomal RNA Genes Research: From Precursor-rRNA to Molecular Evolution

Vera Hemleben[1†], Donald Grierson[2†], Nikolai Borisjuk[3†], Roman A. Volkov[4†] and Ales Kovarik[5*†]*

[1]Center of Plant Molecular Biology (ZMBP), University of Tübingen, Tübingen, Germany, [2]Plant and Crop Sciences Division, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, United Kingdom, [3]School of Life Sciences, Huaiyin Normal University, Huai'an, China, [4]Department of Molecular Genetics and Biotechnology, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, [5]Laboratory of Molecular Epigenetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia

OPEN ACCESS

Edited by:
Wellington Ronildo Clarindo, Universidade Federal de Viçosa, Brazil

Reviewed by:
Andrea Pedrosa-Harand, Federal University of Pernambuco, Brazil
Milene Miranda Praça-Fontes, Universidade Federal do Espírito Santo, Brazil

***Correspondence:**
Vera Hemleben
vera.hemleben@uni-tuebingen.de
Ales Kovarik
kovarik@ibp.cz

†ORCID:
Vera Hemleben
orcid.org/0000-0002-5171-472
Donald Grierson
orcid.org/0000-0002-2238-8072
Nikolai Borisjuk
orcid.org/0000-0001-5250-9771
Roman A. Volkov
orcid.org/0000-0003-0673-2598
Ales Kovarik
orcid.org/0000-0003-2896-0698

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 18 October 2021
Accepted: 26 November 2021
Published: 21 December 2021

Citation:
Hemleben V, Grierson D, Borisjuk N, Volkov RA and Kovarik A (2021) Personal Perspectives on Plant Ribosomal RNA Genes Research: From Precursor-rRNA to Molecular Evolution. Front. Plant Sci. 12:797348. doi: 10.3389/fpls.2021.797348

The history of rDNA research started almost 90 years ago when the geneticist, Barbara McClintock observed that in interphase nuclei of maize the nucleolus was formed in association with a specific region normally located near the end of a chromosome, which she called the nucleolar organizer region (NOR). Cytologists in the twentieth century recognized the nucleolus as a common structure in all eukaryotic cells, using both light and electron microscopy and biochemical and genetic studies identified ribosomes as the subcellular sites of protein synthesis. In the mid- to late 1960s, the synthesis of nuclear-encoded rRNA was the only system in multicellular organisms where transcripts of known function could be isolated, and their synthesis and processing could be studied. Cytogenetic observations of NOR regions with altered structure in plant interspecific hybrids and detailed knowledge of structure and function of rDNA were prerequisites for studies of nucleolar dominance, epistatic interactions of rDNA loci, and epigenetic silencing. In this article, we focus on the early rDNA research in plants, performed mainly at the dawn of molecular biology in the 60 to 80-ties of the last century which presented a prequel to the modern genomic era. We discuss – from a personal view – the topics such as synthesis of rRNA precursor (35S pre-rRNA in plants), processing, and the organization of 35S and 5S rDNA. Cloning and sequencing led to the observation that the transcribed and processed regions of the rRNA genes vary enormously, even between populations and species, in comparison with the more conserved regions coding for the mature rRNAs. Epigenetic phenomena and the impact of hybridization and allopolyploidy on rDNA expression and homogenization are discussed. This historical view of scientific progress and achievements sets the scene for the other articles highlighting the immense progress in rDNA research published in this special issue of Frontiers in Plant Science on "Molecular organization, evolution, and function of ribosomal DNA."

Keywords: rDNA research history, rRNA precursor, rRNA processing, molecular evolution, epigenetics, polyploidy, hybridization, nucleolar dominance

INTRODUCTION

Cytologists in the twentieth century recognized the nucleolus as a common structure in all eukaryotic cells, using both light and electron microscopy. During the winter of 1931, Barbara McClintock observed that in interphase nuclei of maize the nucleolus was formed in association with a specific chromosomal region, normally located at the end of chromosome 6, which she called the nucleolar organizer region (NOR; McClintock, 1934). It took, however, almost 40 years before the composition of the NOR was deciphered (i.e., rRNA genes) with the aid of chromosome *in situ* hybridization techniques in the seventies (Gall, 1981). The molecular structure and function of the ribosomal RNA genes, which are located in or around the nucleolus, were analyzed after the structure of DNA, the triplet code and the central dogma "DNA makes RNA makes protein" were established. In early days of rDNA research, general molecular biological principles were being established rapidly. For example, experiments with cultured cells showed that radioactive amino acids were first polymerized in the cytosol, in association with ribosomes, which identified them as the subcellular sites of protein synthesis. The frog *Xenopus* became a crucial model system for studying the ribosomal RNA genes (rDNA) when Brown and Gurdon (1964) showed that the arrested development and eventual death of *Xenopus* anucleolate mutant embryos was due to their inability to make new rRNA. Birnstiel et al. (1966) demonstrated that these mutants lacked the rDNA, which was subsequently isolated from the wild type and shown to be composed of multiple copies of alternating 18S and 28S rDNA cistrons (Birnstiel et al., 1968). Distinct plant ribosomal DNA satellites had been noted around the same time by Matsuda and Siegel (1967) and characterized subsequently by Goldberg et al. (1972) and Bendich and Anderson (1974). In the mid to late 1960s the synthesis of nuclear-encoded rRNA was the only system in multicellular organisms where transcripts of known function could be isolated, and their synthesis and processing could be studied.

In this retrospective article, we review research carried out in plants from the 60 to 80-ties of the last century focusing on nuclear encoded ribosomal RNA genes as a prequel to recent genomic activities (reviewed in Volkov et al., 2004, 2007; Layat et al., 2012; Sáez-Vásquez and Delseny, 2019; Appels et al., 2021 – this issue). The continuing rise in publications on rDNA in plants witnessed over the past 60 years (**Figure 1**) is testimony to its continuing importance.

GENERAL ORGANIZATION OF NUCLEAR-ENCODED 35S rDNA, rRNA PRECURSOR, AND 5S rDNA: THE BEGINNING OF PLANT MOLECULAR RESEARCH

Many scientists in the 1960s felt that the existence of the tough plant cell wall surrounding a small amount of cytoplasm and a large vacuole containing many secondary products made it very difficult to carry out molecular research on plants. This was certainly felt by some people in the Max Planck Society in Germany and also in the United Kingdom and United States, but plant molecular research began to flourish as methods to overcome these difficulties were developed. Research had been restarted rapidly at the University of Tübingen after the end of World War II. At the Botanical Institute of the University, most people were engaged in studies of the "Biological Clock" and circadian rhythms with Prof. Erwin Bünning (a leading plant physiologist and one of the founders of plant research in Tübingen, (Chandrashekaran, 2006), when a new assistant and later lecturer, Gerhard Richter, arrived. He had spent a research stay performing molecular research in the lab of the biochemist James F. Bonner in Pasadena, the United States. At that time in Tübingen, people from the Max Planck Institute were involved in codon studies and how messenger RNA transported the "Bauanleitung" ("contruction manual") of proteins to the ribosomes, the sites for protein synthesis. So, the environment was prepared for molecular biology, and Gerhard Richter had no problem in convincing Prof. Bünning to establish new laboratories for this kind of research and, most importantly, for working with radioactive substances. Vera Hemleben (VH) thought it would be interesting to work with higher plants and to study nucleic acid synthesis in dark-grown seedlings of beans which could be cultivated under semi-sterile conditions. Other PhD students got involved, and ^{32}P-phosphate radioactively labeled nucleic acids were isolated and separated on MAK [methylated albumin on Kieselgur ("silica")] columns; we isolated the ribosomal RNA fractions and determined the GC-content of the 18S and 25S rRNA (Hemleben-Vielhaben, 1966). Other researchers, e.g., Joe Key now in Athens, Georgia, did similar work (Leaver and Key, 1970). VH decided later to work with *Lemna perpusilla*, the small aquatic monocot plant, which could be cultivated under completely sterile conditions. This was necessary for radioactive pulse-labeling and pulse-chase experiments to follow the fate of the newly synthesized RNA and to characterize the nuclear encoded rRNA precursor, which was 2.3×10^6 Da in size (Hemleben, 1972). From the late 1960s the polycistronic transcription unit, the rRNA precursor, was studied in several different eukaryotic systems. Studies with animals established that the rRNA genes were transcribed as a polycistronic precursor, of variable size in different organisms (45S in humans and 35S in plants), which subsequently underwent endonucleolytic cleavage and methylation before being incorporated, together with ribosomal proteins, into nascent ribosome subunits and transported from the nucleolus to the cytosol (**Figure 2**). Cytosolic ribosomes, formed by 18S (in the small 40S ribosome subunit), 5.8S and 25S plus 5S rRNA (in the large 60S subunit) and ribosomal proteins, were isolated and were, of course, essential constituents in the popular wheat germ *in vitro* protein synthesis system.

In September 1968, Don Grierson joined Ulrich Loening's laboratory in Edinburgh to study the synthesis of rRNA in primary leaves of mung bean seedlings (Grierson et al., 1970; Grierson and Loening, 1972, 1974). Ulrich had developed

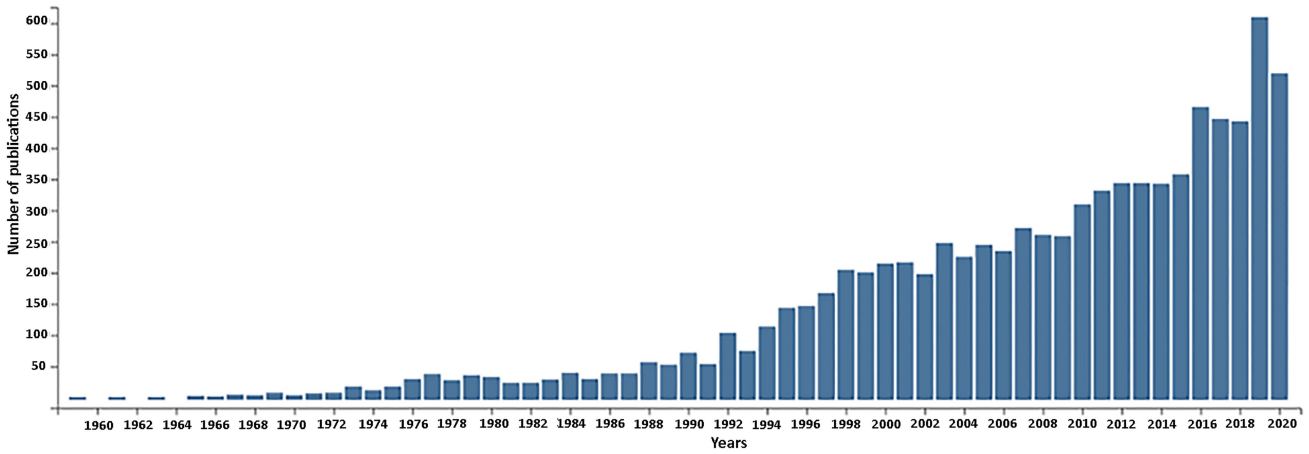


FIGURE 1 | Publications related to plant nuclear rDNA research over the timespan of 1960–2020. The number of retrieved publications is shown in 2-year increments. The total number of publications returned by the Web of Science database was 8,948. Key words used for searched fields: rRNA *or* rDNA *and* plant with following filters: *no* animal, *no* fungal, *no* chloroplast (plastom), *no* mitochondrion.

methods for extracting undegraded RNA, fractionating it on the basis of size in polyacrylamide gels and determining the molecular weights with great precision (Loening, 1969). Ulrich built electrophoresis tanks from Perspex sheets, with platinum wire electrodes, insisted on redistilling phenol for use in RNA extraction, recrystallizing acrylamide and using deoxygenated monomer solutions to get consistent polymerization and electrophoresis results. Disposable plastic ware and automatic microsyringes were the subject of dreams. Ulrich also devised a novel gel-scanner and apparatus for slicing gels and automated detection of radioactivity in the slices. These methods generated worldwide interest and attracted collaborators and many visitors to the laboratory. The synthesis and processing of a polycistronic precursor (pre-rRNA) was studied in several plants, including mung bean leaves and roots (Grierson et al., 1970; Grierson and Loening, 1974), pea roots and cultured artichoke cells (Fraser and Loening, 1974), carrot (Leaver and Key, 1970), cultured sycamore cells (Cox and Turnock, 1973). In general, the observations were similar: when seedlings, plant tissues, or organs were incubated with ^{3}H-uridine or ^{32}P-phosphate for short periods of time the radioactivity was incorporated into distinct macromolecular transcripts. On polyacrylamide gels these could be seen above a polydisperse array of RNA molecules, presumably mRNAs, nascent molecules and processing products. Molecular weight estimates for the largest molecules ranged from 2.3 to 2.8×10^6 Da. At slightly later times, radioactivity was also detected in molecules of around 1.4 and 0.75×10^6 Da and pulse-chase experiments, kinetics of accumulations and comparison of the nucleotide composition of these molecules all supported the conclusion that the initial transcript was a large polycistronic molecule that included one 18S, 5.8S and 25S transcript, together with "transcribed spacer" RNA. Subsequent ribonucleolytic cleavage gave rise to the immediate rRNA precursors, each slightly larger than the mature rRNAs. Aggregates and breakdown products of rRNAs could also be distinguished (Grierson and Loening, 1974), and electrophoresis under denaturing conditions in formamide gels gave lower molecular weight estimates of the size of the initial transcripts in carrot, parsley, and sycamore (Seitz and Seitz, 1979). The size characterization of individual pre-RNA molecules and the complexities of the rRNA maturation pathway stimulated further research. The processing of plant pre-rRNA, updated to show recent findings, is illustrated in **Figure 2** and the enzymes catalyzing individual RNA cleavage steps (A0, A1, A2 and C2 sites) have now been identified (Tomecki et al., 2017). Moreover, plant small nucleolar RNA (snoRNAs) that are thought to take part in pre-rRNA cleavage events were identified (Brown and Shaw, 1998). It became clear that the cleavage events are compartmentalized, some occurring in the nucleolus (A0, and A2) with others take place in the nucleoplasm (C2; **Figure 2**). Finally, although key steps of plant pre-rRNA processing seem to be similar to that of other eukaryotes (Grierson, 1984) notable differences exist between yeast, plant and animal pre-RNA pathways. For example, the analysis of ribosome biogenesis in plants revealed two alternative processing pathways coexisting in plants (Weis et al., 2015; Sáez-Vásquez and Delseny, 2019). A major pathway 1 is initiated by ITS1 cleavage (A2 site, **Figure 2**) and subsequent removal of the 5'-ETS, which is comparable to the human processing pathway. Pathway 2 starts with the 5'-ETS removal (cleavage at the A0 site, **Figure 2**) followed by the ITS1 cleavage which leads to the separated assembly of the pre-40S and pre-60S ribosomal subunits. This pathway is reminiscent of rRNA processing in yeast. In addition, 5'-ETS processing is initiated by exoribonucleolytic trimming of the 5'-end by XRNs in *Arabidopsis thaliana* (Zakrzewska-Placzek et al., 2010).

Other rRNA transcription units were found transcribed from the DNA in chloroplasts, containing their smaller 16S and 23S rRNAs, components of the distinct 70S ribosomes, and a further class of ribosomes in the mitochondria, although their synthesis was not studied in such detail. Double labeling experiments showed a stable polycistronic precursor of rRNA in leaves,

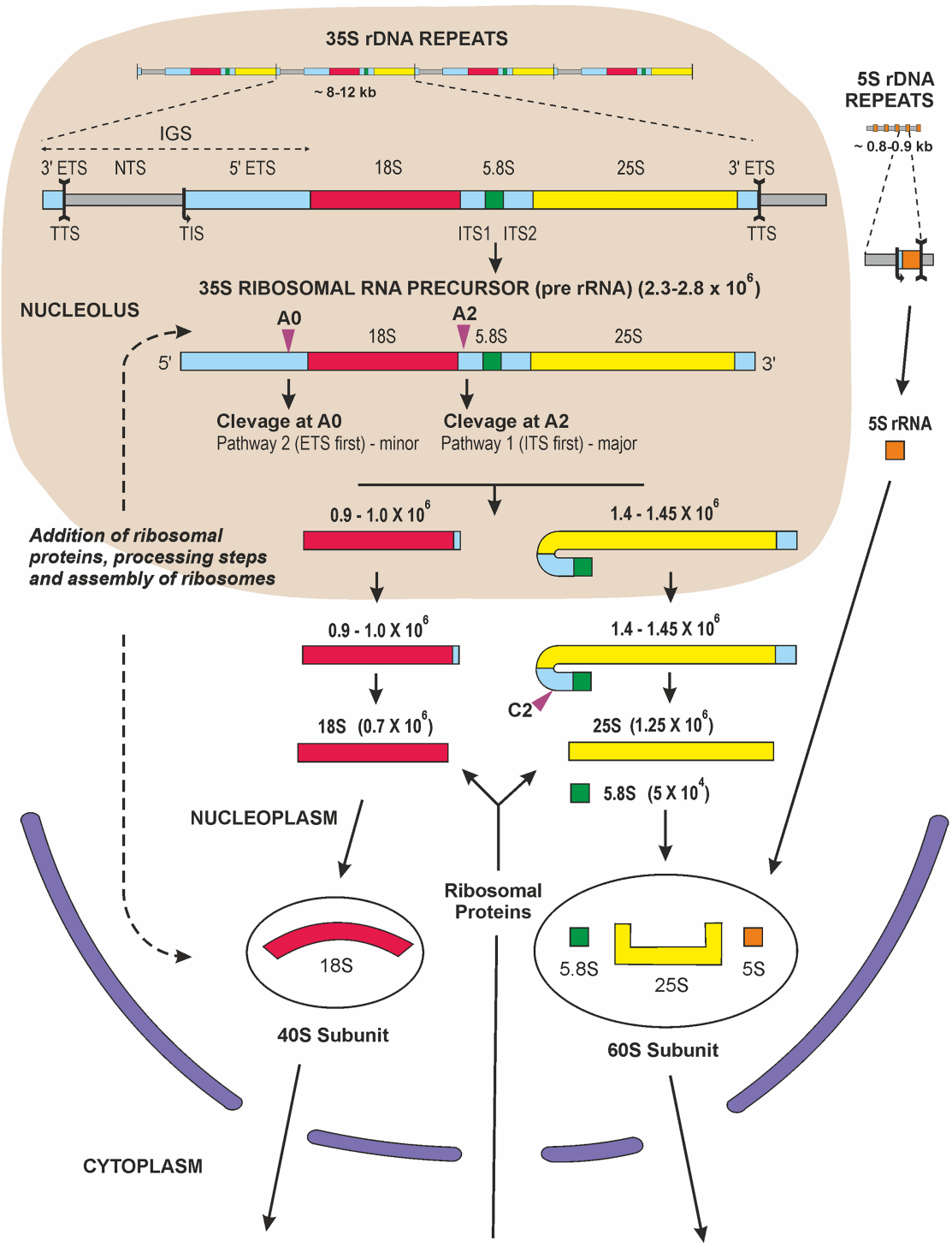


FIGURE 2 | Graphic representation of ribosome biogenesis in eukaryotic cells (adapted from (Grierson, 1984; Sáez-Vásquez and Delseny, 2019). Transcription of rDNA requires RNA Pol I activity and a subset of general transcription factors. The primary transcript, precursor-rRNA (35S in plants and yeast or 45S in mammals), encodes three rRNAs and is first co-transcriptionally processed into the mature 18S, 5.8S, and 25S/28S rRNA. This processing steps involve multiple endonucleolytic and exonucleolytic cleavages (violet arrowheads) occurring in the nucleolus [A0 site or P site according to nomenclature of (Sáez-Vásquez and Delseny, 2019)], A2 site and the nucleoplasm (C2). The 18S rRNAs assemble with ribosomal proteins of the small 40S ribosomal subunit, RPSs, while 5.8S, 25S/28S and 5S rRNA assemble with ribosomal proteins, RPLs, forming the large 60S ribosomal subunit. The 5S rRNA is transcribed in the nucleoplasm by RNA Pol III and imported into the nucleolus. Assembly and transport of ribosomal particles from nucleolus to cytoplasm requires hundreds of specific 40S and 60S RBFs (Ribosome Biogenesis Factors). The 40S and 60S ribosomal subunits finally join to form translationally competent ribosomes in the cytoplasm. Sizes of individual rRNA molecules are in daltons.

which was distinct from, and larger than that in roots (Grierson and Loening, 1972). This RNA species was subsequently shown to be synthesized by chloroplasts, although the conclusion that it represented a polycistronic chloroplast rRNA precursor was not unanimous (Hartley and Ellis, 1973; Grierson and Loening, 1974). This may have been because chloroplasts were believed to have been derived from blue-green bacteria during the course of evolution and prokaryotic rRNAs had been found to be monocistronic (Adesnik and Levinthal, 1969; Dahlberg and Peacock, 1971; Grierson and Smith, 1973; Seitz and Seitz, 1973). Of course, the similarity between the rRNAs of bacteria and chloroplasts had evolutionary significance, as is now widely recognized (Gray, 2017). Similarities between bacteria and these cellular organelles had been noted in the 19th and early 20th centuries but this idea was not given much credence until Lynn Margulis (then Lynn Sagan) published her account of what was described as "perhaps the first unified theory of eukaryogenesis," proposing, as is now widely accepted, that "mitochondria and plastids might have originated endosymbiotically from prokaryotic progenitors" (Gray, 2017).

The discovery and application of restriction enzymes by Werner Arber, Daniel Nathans, and Hamilton O. Smith, who were jointly awarded the Nobel Prize in Physiology or Medicine in 1978 and the work of Sir Kenneth Murray (department of Molecular Biology, University of Edinburgh), in combination with development of other new technologies for transformation of bacteria with foreign DNA, paved the way for gene cloning and experimental gene transfer between organisms. In Tübingen, now at the Genetics department, the VH group planned to study gene transfer in higher plants with the genetically well-defined *Matthiola incana*. Therefore, highly ^{3}H-labeled total DNA was supplied to plant seedlings, and we found integration of this foreign DNA into the nuclei (Hemleben et al., 1975; Leber and Hemleben, 1979). We reported these results at the first Plant Molecular Biology Conference in Liege in 1974. Here, we met Don Grierson, and we organized a 1-year project in 1975/76, supported by an EMBO Long-Term Fellowship to DG. During a meeting in Edinburgh (a summer school run by Ulrich Loening and the late Max Birnstiel in Edinburgh in 1975), we had learned how to separate DNA on Actinomycin D-CsCl gradients and to separate the often highly repeated rRNA genes (rDNA) from the main-band DNA in animal systems (see Birnstiel et al., 1968). Don had also done this with mung beans for his PhD (Grierson, 1972), and this opened up the possibility to purify plant rDNA (Grierson and Hemleben, 1977; Hemleben et al., 1977) and later to characterize it by restriction enzyme analysis, which we carried out firstly during a research stay at Joe Key's lab in Athens/Georgia in 1978 (Friedrich et al., 1979).

The DNA-content of animals and plants (for plants see: Nagl et al., 1979; Wenzel and Hemleben, 1982a; Michael and Van Buren, 2020) appeared to vary enormously, and the question was: What is the explanation for repetitive genome components and which sequences are redundant? The mature, purified and radioactively labeled rRNAs were used as probes in classical liquid phase hybridization experiments with genomic DNA, showing often enormously high numbers of the tandemly repeated genes for the 18S, 5.8S and 25S (Matsuda and Siegel, 1967; Bendich and Anderson, 1974; Ingle et al., 1975; Wenzel and Hemleben, 1982a). Matsuda and Siegel (1967) further showed that the amount of rDNA cistrons (units) in the nuclear DNA varied among tobacco, pumpkin, pinto beans and Chinese cabbage plants over a 10-fold range. Therefore, it became clear that, besides the highly repetitive satellite DNA, the tandemly arranged and highly redundant rRNA genes contribute to this variability (Marazia et al., 1980; Rogers and Bendich, 1987). Our first chromatin and methylation studies with *Matthiola incana* and *Brassica pekinensis* showed that most of the rRNA genes were not transcriptionally active. The rDNA-containing chromatin was not accessible to DNase I digestion, and most of the rRNA genes appeared highly methylated (Leber and Hemleben, 1979; Leweke and Hemleben, 1982; Wenzel and Hemleben, 1982b). At that time, this gene silencing phenomenon was not yet widely called "epigenetics," although there was a department of Epigenetics established at the Edinburgh University. Transcriptional regulation of nuclear encoded rRNA genes by methylation and demethylation, respectively, was observed also by other researchers (Flavell et al., 1988; Thompson and Flavell, 1988). In Tübingen, we were working with *Cucurbitaceae* species, which were known for an enormously high number of ribosomal RNA genes (Bendich and Anderson, 1974) and in the 1990s a PhD student Ramon Torres-Ruiz was able to identify the pattern and degree of methylation in the rDNA of *Cucurbita pepo* (Torres-Ruiz and Hemleben, 1994).

The nuclear ribosomal DNA could, as described, be separated from the main nuclear DNA by Act D-CsCl gradient ultracentrifugation followed by restriction enzyme mapping. This enabled in the early 1980s cloning of plant rDNA using specific gene probes for the 18S, 5.8S, and 25S rRNA and for the internal transcribed (ITS1 and IT2), external transcribed (ETS) and the non-transcribed (NTS) regions of the intergenic spacer (IGS; **Figure 2**), which allowed subsequent rDNA sequencing. Early on, large DNA fragments, especially those containing internal repeated DNA elements (subrepeats) identified later in the 35S rDNA IGS (35S IGS), were difficult to clone in plasmid vectors, but they could be analyzed by restriction endonucleases and often showed length heterogeneity even in single individuals (Ganal and Hemleben, 1986; Rogers et al., 1986; Yokota et al., 1989). Later, we and others were able to clone the complete large IGS and to characterize this region by DNA sequencing using at that time the radioactive sequencing methods (Sanger et al., 1977). Of course, the polymerase chain reaction (PCR) developed by the biochemist Karl Mullis, who in 1993 shared the Nobel Prize in Chemistry, facilitated enormously this procedure. Length heterogeneity of the rDNA repeats was due mostly to different numbers of subrepeats in the 35S IGS upstream or downstream of the transcription initiation site, TIS (Appels et al., 1986; Ganal et al., 1988; Rathgeber and Capesius, 1990; Borisjuk and Hemleben, 1993; King et al., 1993; Zentgraf and Hemleben, 1993; **Figure 2**, upper part). Interestingly, it appeared that in some plants (e.g., some *Vigna* sp.) these repeated elements of the 35S IGS formed independent highly amplified satellite DNA genome components (Unfried et al., 1991; Macas et al., 2003; Lim et al., 2004b; Jo et al., 2009; Kirov et al., 2018). In contrast, 35S DNA amplification of satellites within the 5S rDNA loci is rare. Nevertheless, a 170-bp

satellite sequence (termed jumper) apparently invaded 5S rDNA in the evolutionary history of the *Phaseolus* genus (Ribeiro et al., 2017). Of note, only a single satellite monomer is found in the 5S IGS while there may be thousands of copies outside of 5S rDNA loci, forming pericentromeric and subtelomeric domains. This may suggest that the rDNA intergenic spacers have relatively frequently hosted non-coding satellites but their expansion is limited to one or a few copies, probably due to selection constrains imposed on spacer lengths.

In chromosomes, the 5S rRNA genes occur either as long tandem repeats of regularly spaced units (S-type arrangement) or as solitary insertions within the 35S rDNA intergenic spacer (L-type arrangement). The S-type arrangement is the most frequent organization of 5S rDNA in angiosperms (Ellis et al., 1988; Hemleben and Werts, 1988; Scoles et al., 1988; Röser et al., 2001) while, so far, the L-type arrangement has been detected only in some member of the Asteraceae family (Garcia et al., 2009, 2010; Souza et al., 2019). The L-type arrangement is more typical for plants that diverged early during angiosperm evolution (Capesius, 1997; Wicke et al., 2011) and some groups of gymnosperms (Galian et al., 2012; Garcia and Kovarik, 2013). The 5S rDNA was apparently invaded by an LTR transposon in the early evolutionary history of angiosperms, giving rise to Cassandra retrotransposons (Kalendar et al., 2008), which are now widespread in modern species. Truncated incomplete copies of 35S rDNA seem to be scattered in genomes of both plants (Tulpová et al., 2020) and animals (Robicheau et al., 2017) and likely represent remnants of former NORs.

TRANSCRIPTIONAL REGULATION OF rRNA GENES

The nuclear encoded 18S, 5.8S, and 25S ribosomal RNA genes were already known in yeast and animals to be transcribed by RNA polymerase I, and this was confirmed in plants in Joe Key's lab (Lin et al., 1976) and in Tübingen (Grossmann et al., 1979, 1980). Functional studies identified the putative transcription initiation (TIS) and transcription termination sites (TTS) for plant rDNA (Delcasso-Tremousaygue et al., 1988; Gerstner et al., 1988; Gruendler et al., 1989; Schiebel and Hemleben, 1989; Zentgraf and Hemleben, 1992, 1993; **Figure 2**, upper part). In addition, the repeated elements upstream or downstream of the TIS obviously had an enhancer function (reviewed in Hemleben and Zentgraf, 1994). Under the current view, the nucleoprotein complex responsible for transcription initiation of 45S (35S) rRNA is composed of numerous protein components (Nucleolar Remodelling Complex (NoRC), UBF, histone acetyltransferases, helicases, and RNA polymerase I, among others) and several species of noncoding RNA (Bersaglieri and Santoro, 2019; Yan et al., 2019). The RNA polymerase I holoenzyme has been purified to apparent homogeneity by biochemical approaches, maintaining the capacity to initiate rDNA transcription (Sáez-Vásquez and Pikaard, 2000). Nucleolin, a relatively abundant nucleolar structural protein, seems to be involved in selection of rDNA variants for transcription in Arabidopsis (Pontvianne et al., 2010).

The 5S rRNA genes are transcribed by RNA polymerase III and their promoter elements and termination sites were putatively described for plants (Hemleben and Werts, 1988). Transcriptional regulation of the multigenic 5S rDNA in *Arabidopsis* has been extensively studied in Sylvette Tourmentes's laboratory (University of Clermont-Ferrand, France), confirming the originally described 5S rRNA internal and external elements involved in regulation of the gene's transcription (reviewed by Layat et al. (2012). Later, in Roman Volkov's laboratory, conservation of the putative external promoter elements in the 5S rDNA intergenic spacer (5S IGS) was demonstrated for several families of angiosperms (Volkov et al., 2001, 2017; Tynkevich and Volkov, 2014, 2019; Ishchenko et al., 2018, 2020).

While the major advances in our knowledge of rDNA regulation were achieved in yeast (Moss, 2004) and animals (rat, mouse and *Xenopus*; (Grummt et al., 1985; Pikaard and Reeder, 1988), the research on plant rDNA also made significant progress over the years (Weis et al., 2015; Sáez-Vásquez and Delseny, 2019). Now it is widely accepted that in addition to transcriptional regulation of individual rDNA repeat units, the entire rDNA arrays (NOR) are targets of regulation as exemplified by studies in *Arabidopsis* (Mohannath et al., 2016) and wheat (Handa et al., 2018). The findings described recently by (Sims et al., 2021), sequencing entire rDNA arrays and deciphering their higher structure promise further deeper insights into the functional organization and molecular regulation of plants rDNA loci. By applying a combination of long- and short-read sequencing the authors revealed clustering of rDNA domains in *Arabidopsis* NOR2 and expression of several variants of rRNAs with their tissue-specific integration into active ribosomes (Sims et al., 2021). In leaf tissue of the ecotype Columbia-0 (reference genome), the rDNA of NOR4 on chromosome 4 is usually more active than that on chromosome 2 (NOR2; Chandrasekhara et al., 2016). However, NOR4 is not always dominant and many natural populations of *Arabidopsis thaliana* show considerable epigenetic variability, i.e., dominant expression of NOR4, NOR2, or codominant expression of both loci (Rabanal et al., 2017).

The duckweed species *Spirodela polyrhiza* and *S. intermedia* might emerge as a promising new model to study rDNA regulation (in addition to *Arabidopsis*) because of their compact rDNA loci, composed of no more than a hundred 35S repeated units (Michael et al., 2017; Hoang et al., 2020), even fewer than the 100–200 rDNA copies in yeast (Salim et al., 2017), and an order of magnitude lower than rRNA gene copies in other plants (Wang et al., 2019). The copy number of 5S rDNA is estimated to be about 170 in *Landoltia punctuata* (Chen et al., 2021) – this issue. This exceptionally low copy number makes the duckweed rDNA locus relatively simple to access applying the third-generation sequencing platforms for ultra-long sequencing reads (Jung et al., 2019).

MOLECULAR EVOLUTION OF rDNA LOCI

Although it was initially believed that the rRNA genes are quite conserved, it turned out in the 1980s that the regions

of the 5' and 3' ETS, NTS, and ITS1 and 2 are characterized by a huge intra- and interspecies variability. This made them highly suitable as markers for our phylogenetic and molecular evolution studies at population or interspecies levels (Torres et al., 1989; Grebenstein et al., 1998; Jobst et al., 1998; Volkov et al., 2003, 2010; Denk et al., 2005; Grimm et al., 2005; Komarova et al., 2008; Schlee et al., 2011). **Table 1** summarizes individual 35S rDNA subregions used for phylogenetic markers.

Similarly, the 5S IGS appeared very useful for clarifying phylogenetic relationships at low taxonomic levels (Röser et al., 2001; Volkov et al., 2001; Denk and Grimm, 2010; Tynkevich and Volkov, 2019; Ishchenko et al., 2021) and for identification of interspecific hybrids (Garcia et al., 2020). This contrasts with the really strongly conserved sequences of the mature 5S, 5.8 S, 18S, and 25S rRNA coding regions. However, variable segments of 18S and 25S coding regions ("expansion segments") evolve faster than the conserved stems. These features of rRNA coding regions can be used for phylogenetic studies particularly at higher taxonomic levels (Poczai and Hyvonen, 2010; Soltis and Soltis, 2016). The secondary structure of rRNA transcripts is another layer of phylogenetic information in addition to primary sequence (Noller et al., 1981; Selig et al., 2008). Currently, a public database of ITS2 secondary structure models comprise more than 80 thousand sequences (Selig et al., 2008). Consequently, various rDNA regions have been studied worldwide in the 1990s until the present day as phylogenetic markers in population, species, genus, and higher systematic order studies and helped to solve the phylogenetic relationships between organisms. Public databases storing biological information about rDNA loci (Szymanski et al., 1998; Selig et al., 2008; Cantara et al., 2011; Garcia et al., 2012; Quast et al., 2012) represent a valuable source for structural, functional and phylogenetic studies.

In the 1990s Nikolai (NB) and Ljudmilla Borisjuk and later Roman Volkov (RV) and Irina Panchuk from the Ukraine joined the VH laboratory, NB and RV as Alexander v. Humboldt fellows. They had already started to study plant rDNA, and we had a very successful cooperation over the years working mostly on several genera of Solanaceae (Borisjuk et al., 1994, 1997; Volkov et al., 1999a,b; Komarova et al., 2004). In 1983–1985, RV and NB worked in the group of Andrey S. Antonov at Moscow University (Russia). NB worked on the characterization of genomes of Solanaceae and Brassicaceae somatic hybrids generated by protoplast fusion in the Lab of Yuri Gleba in Kyiv (Ukraine). RV was interested in describing rearrangements of repeated sequences in natural allopolyploids, particularly in the genus *Nicotiana*, which includes several allopolyploids and aneuploids. These young researchers decided to perform the 35S rDNA restriction mapping for several artificial and natural allopolyploids. They used as probes for Southern hybridization 18S and 25S rRNA from maize and a fragment of 25S coding sequence of lemon isolated by Volodymyr Kolosha (Pushchino, Russia) under the supervision of Tengiz Beridze (Georgia). Mapping experiments revealed that generally the interspecies and inter-tribal hybrids obtained by protoplast fusion or sexual crossing inherited a combination of parental rDNA (Gleba et al., 1988; Borisjuk and Miroshnichenko, 1989; Miroshnichenko et al., 1989). Additionally, a novel class of rDNA repeats was found in somatic hybrids between distantly related *Nicotiana* and *Atropa* (Borisjuk et al., 1988). Later on, this observation let to the discovery of an "amplification promoting sequence" (APS) within the tobacco 35S IGS. The cloned APS element apparently increased the copy number of linked reporter genes in transgenic experiments resembling the origin of DNA replication (Borisjuk et al., 2000). It transpired that the structure of the *Nicotiana* 35S IGS is highly complex, bearing repetitive subregions which apparently account for species-specific differences in rDNA structure (Volkov et al., 1996; Borisjuk et al., 1997). Genomic analysis showed numerous SNPs (single nucleotide polymorphisms) in the tobacco 35S IGS, which indicated that the mutation rate in that region may be faster than that of coding regions, arguing for variable selection pressures acting on different parts of the rDNA unit (Lunerova et al., 2017).

THE FATE OF rDNA IN SOLANACEAE HYBRIDS AND ALLOPOLYPOIDS

The multigene families in both plants and animals reveal high levels of intra-species homogeneity and inter-species diversity. These features underlie the concept of concerted evolution put forward by geneticists in the second half of the last century (Brown et al., 1972; Zimmer et al., 1980; Dover, 1982). It has become clear that concerted evolution (i.e., homogenization)

TABLE 1 | Characteristics of individual subregions of plant rDNA units and their relevance for phylogenetic analysis.

Feature	Coding regions (5S, 5.8S, 18S, 25S rRNA)	35S-NTS[a]	5' ETS[b]	ITS1/ITS2	5S-NTS[c]
Tempo of evolution	Slow	Extremely fast	Fast	Fast to moderate	Fast to moderate
Subrepeated structure	No	Frequent	Occasionally	Exceptional	Rare
Resolution power in phylogenetic studies	Order/family	Species/subspecies/ cultivars/populations	Genus/species	Genus/species	Genus/species
Utility for interspecific hybrids identification	No	Intermediate	Good	Good	Excellent

[a]*A part of the 35S rDNA intergenic spacer located between the transcription termination site (TTS) and the transcription initiation site (TIS - see **Figure 2**).*
[b]*A part of the 35S rDNA intergenic spacer located between the TIS and the 18S rRNA gene.*
[c]*Non-transcribed intergenic spacer of variable size between tandemly arranged 5S rRNA genes.*

processes affect nearly all repeated families including non-coding satellites and rDNA. The tandemly arranged rDNA represent a textbook example of concerted evolution since their hundreds of units show little or no intragenomic variation (reviewed by (Eickbush and Eickbush, 2007; Nieto Feliner and Rossello, 2012). The presence of rDNA arrays that are homogeneous for different variants in interbreeding populations of *Drosophila melanogaster* indicated that there is little recombination between the arrays while there might be intensive recombination within the arrays, leading to their overall homogeneity (Schlotterer and Tautz, 1994).

Towards the end of the last century the VH group explored the Solanaceae family (nightshades) which includes many economically important crops such as tomato, potato and tobacco. Within the family, the *Nicotiana* genus, whose center of diversity is Latin America, contains at least 50 allopolyploids of different ages and genome compositions. *Nicotiana tabacum* (tobacco) is, perhaps, the most well-known allotetraploid (2n = 4x = 48, genome composition SSTT) and has long been a favorite model for plant genetic studies (Goodspeed, 1954) including transgenosis and chromosome evolution (Manoharlal et al., 2019; Dodsworth et al., 2020). It is a relatively recent (*ca.* 0.1 myrs old; Lim et al., 2007) allopolyploid originating from hybridization of progenitor species close to *Nicotiana sylvestris* (2n = 2x = 24, S genome) and *Nicotiana tomentosiformis* (2n = 2x = 24, T genome). Its parental S- and T-genomes are relatively intact with few intergenomic translocations (Lim et al., 2004a). The question was: What is the fate of parental 35S rDNAs in tobacco? Are they intact or have they been modified by allopolyploidy?

In order to clarify the fate of parental 35S rDNA in the genome of *N. tabacum*, NB and RV decided to sequence the 35S IGS regions. From 1990 to 1992, with the support of the Alexander v. Humboldt Foundation, NB worked in the group of VH, who by then was a well-known leader in rDNA research. Here he performed restriction mapping of numerous *Solanum* species and other Solanaceae (Borisjuk et al., 1994) and cloned and sequenced the 35S IGS of *S. tuberosum* (Borisjuk and Hemleben, 1993) and *N. tabacum* (Borisjuk et al., 1997). During this time VH's lab used RFLP (restriction fragment length polymorphism) for characterizing artificial somatic hybrids of *Solanum tuberosum* and various wild *Solanum* species produced by protoplast fusion in the lab of Prof. Helga Ninnemann by Dr. Lieselotte Schilde-Rentschler, mainly to introduce pathogen resistant characters into the cultivated potato (Schweizer et al., 1993). Therefore, the *Solanum* research of NB and RV was very complementary to the hybrid identification research. Interestingly, the rDNA of one fusion partner disappeared very quickly (see below for *Nicotiana*).

RV obtained an Austrian exchange service Research Fellowship, and in 1993 he went to the lab of Prof. Dieter Schweizer (Department of Cytology and Genetics, University of Vienna), where the 35S IGS of *Arabidopsis thaliana* had recently been sequenced and characterized (Gruendler et al., 1989). In Dieter Schweizer's laboratory, RV cloned and sequenced 35S IGS of *N. sylvestris* and *N. tomentosiformis* (Volkov et al., 1996, 1999b).

In 1996, RV received the AvH Research Fellowship, moved to Tübingen and joined the VH group in order to investigate further the rDNA in Solanaceae. Comparative analysis of the 35S IGS sequences of N. *tabacum*, *N. tomentosiformis* and *N. sylvestris* allowed the molecular evolution of parental rDNA in the genome of *N. tabacum* to be deduced. Strikingly, only units similar to the paternal *N. tomentosiformis* genome (T-genome) were cloned from the tobacco genome (Volkov et al., 1999b), while clones from the maternal parent were not recovered, indicating elimination of the S-genome rDNA. It was found that the rDNA repeats of *N. tabacum* originated from *N. tomentosiformis*, which involved reconstruction of subrepeated regions in the 35S IGS upstream and downstream of the transcription initiation site. These cloning results well-resonated with the non-additivity of tobacco 35S rDNA restriction fragments observed in previous Southern hybridization experiments (Borisjuk et al., 1989; Miroshnichenko et al., 1989; Volkov et al., 1991; Kovarik et al., 1996). Clearly, thousands of parental rDNA units were overwritten by novel hybrid-specific units in relatively short evolutionary time (<100 thousand years; Borisjuk et al., 1997; Volkov et al., 1999a,b). Molecular cytogenetics approaches carried out by Andrew Leitch's group (University of London) revealed that the number of tobacco rDNA loci is additive, i.e., there is a single locus in the T-genome and three loci in the S-genome (Lim et al., 2000). Only a small number of unconverted and highly methylated S-genome units were detected in the tobacco genome by molecular methods. In 2017 Jana Lunerova (AK group) using locus-specific FISH probes addressed the question of chromosomal localization of these transcriptionally inactive rRNA genes (Lunerova et al., 2017). It appeared that the residual (about 8% of total rDNA) S-genome units are located on a small acrocentric chromosome S12 while active homogenized genes are located on chromosomes T3, S10, and S11. Of note, the S-genome rDNA loci were found in the variety SR-1 and wild tobacco collected by Sandy Knapp (Natural History Museum London) in Bolivia, but not in the variety 09555 (Kovarik et al., 2004). Thus, the process of cultivation and high inbreeding may potentially influence the behavior of rDNA in allopolyploids. The Leitch's and AK's groups further confirmed partial and complete homogenization of parental rDNAs in another two *Nicotiana* allotetraploids, *Nicotiana rustica* (Indian tobacco, 2n = 4x = 48; Matyasek et al., 2003) and *Nicotiana arentsii* (2n = 4x = 48), respectively (Kovarik et al., 2004). Hence genetic interactions of rDNA loci seem to be a rather general feature of rDNA evolution in *Nicotiana* allopolyploids.

Fulnecek et al. (2002) investigated the structure of 5S rDNA loci which occurs separately from 35S rDNA loci in most plant genomes (Hemleben and Grierson, 1978; Garcia et al., 2016). Locus-specific FISH together with pulsed-field gel electrophoresis mapping showed that parental 5S rDNA arrays remained relatively intact and were inherited at expected ratios in tobacco allotetraploid. Therefore, in contrast to 35S rDNA, the 5S rDNA loci do not genetically interact in tobacco allotetraploids. The reason for higher genetic stability of 5S rDNA compared to 35S rDNA is not fully understood. However, 5S rDNA is highly methylated (more than the genome average; Fulnecek et al.,

1998), while 35S rDNA units contain many undermethylated sites particularly in intergenic spacers (Torres-Ruiz and Hemleben, 1994). Hypomethylated sites in 35S IGS were also observed in other species including *Arabidopsis* (Earley et al., 2006), cucumber (Torres-Ruiz and Hemleben, 1994), potato (Komarova et al., 2004) and wheat (Sardana et al., 1993). It is possible that apart from transcription regulation these undermethylated sites might be important for adopting chromatin conformation favorable to recombination processes (Kovarik et al., 2008).

Volkov et al. (2017) applied a combination of karyological and molecular methods to investigate chromosomal localization, molecular organization and evolution of 5S and 35S rDNA in *Atropa belladonna* (Solanaceae), one of the oldest known flowering plant allohexaploids. Intensive sequence homogenization between three pairs of 35S rDNA loci on separate chromosomes was found, presumably inherited from tetraploid and diploid ancestor species. Only four out of six 35S rDNA sites appeared transcriptionally active, demonstrating nucleolar dominance. For 5S rDNA, three size variants of repeats were detected, with the major class represented by repeats containing all functional 5S IGS elements required for transcription, whereas intermediate and short length repeats contained defects both in the spacer and coding sequences. The functional 5S rDNA variants are nearly identical at the sequence level, pointing to their origin from a single parental species. Localization of the 5S rRNA genes on two chromosome pairs further supports uniparental inheritance from the tetraploid progenitor. The data demonstrate complex evolutionary dynamics of rDNA loci in allohexaploid species of *Atropa belladonna*. The high level of sequence unification revealed in 5S and 35S rDNA loci of this ancient hybrid species have been seemingly achieved by different molecular mechanisms.

NUCLEOLAR DOMINANCE IN SOLANACEAE HYBRIDS AND ALLOPOLYPLOIDS

Nucleolar dominance is an epigenetic phenomenon in which one parental array is inactivated in interspecific hybrids and allopolyploids. It was first described at the cytological level by Navashin (Navashin, 1934), who observed that in interspecific hybrids of *Crepis* (Asteraceae) only chromosomes of one crossing partner carried secondary constrictions at metaphase. This chromosomal region was not lost in hybrids but could be reactivated to produce normal nucleoli in hybrids with a different crossing partner. Experiments with epigenetic inhibitors performed in plants towards the end of the last century established that histone deacetylation and DNA methylation pathways interact in a self-reinforcing mechanism, maintaining silencing of partner rDNA units in hybrids (Chen and Pikaard, 1997a; Chen et al., 1988). The VH and AK groups investigated nucleolar dominance from different perspectives, asking questions about the influence of structural features of the 35S IGS, cytosine methylation of rDNA units and developmental stability of nucleolar dominance. To address these questions, they used well-defined natural and synthetic *Nicotiana* and *Solanum* (Solanaceae) allotetraploids. A comparison of 35S rDNA organization in several *Solanum* species revealed (Borisjuk and Hemleben, 1993; Borisjuk et al., 1994) that *S. lycopersicum* (tomato), *S. tuberosum* (potato) and wild species *S. bulbocastanum* possess 35S IGS of nearly identical length but contain different number of subrepeats up- and downstream of the TIS. Accordingly, VH and RV suggested using these species to elucidate the presumptive role of subrepeated elements in nucleolar dominance. In 1998, Nataliya Komarova from RV's group moved from Ukraine to VH's lab, where she studied expression of parental 35S rDNA in *Solanum lycopersicum* x *S. tuberosum* and *S. tuberosum* x *S. bulbocastanum* artificial somatic alloploids produced by protoplast fusion and back-crossed lines, which were kindly provided by E. Jacobsen and H.J. de Jong (Wageningen University, The Netherlands) and by L. Schilde-Rentschler and H. Ninnemann (University of Tübingen, Germany). It appeared that an expression hierarchy exists: In leaves, roots, and petals of the respective allopolyploids, rDNA of *S. lycopersicum* dominates over rDNA of *S. tuberosum*, whereas rDNA of *S. tuberosum* dominates over that of the wild species *S. bulbocastanum*. Also, in a monosomic addition line carrying only one NOR-bearing chromosome of tomato in a potato background, the dominance effect was maintained. These results demonstrated that there is possible correlation between transcriptional dominance and number of conserved elements downstream of the transcription start in the *Solanum* rDNA (Komarova et al., 2004). The authors proposed that this sequence motif could be a recognition site for DNA-interacting proteins involved in modulation of rDNA transcription (Borisjuk et al., 1997; Volkov et al., 2003). Remarkably, no correlation between the number of upstream subrepeats and differential transcription/silencing of 35S rDNA in *Solanum* allopolyploids was found. The latter contrasts with observations made in the allohexaploid wheat (AABBDD), where longer B-genome units containing more upstream subrepeats are active while the shorter units located in the D genome are usually inactive (Sardana et al., 1993). Units bearing longer upstream elements also seem to be dominant in recently (<100 years) formed *Tragopogon allotetraploids* (Matyasek et al., 2016). In contrast, sexual F1 hybrids resulting from crossing of *N. sylvestris* x *N. tomentosiformis* plants and a synthetic tobacco line show codominance (Dadejova et al., 2007), despite apparent differences in the 35S IGS structure of progenitor units. Thus, it seems that the role of 35S IGS repeat elements in regulation of rDNA expression varies from system to system. Certainly, a direct proof for an enhancer/silencing role of these elements, as shown in *Xenopus laevis* 35S IGS (Caudy and Pikaard, 2002), is missing in plants.

Developmental stability of nucleolar dominance was investigated in hybrids of *Brassica* and *Solanum*. Classical experiments in *Brassica napus* allotetraploids showed developmental lability of nucleolar dominance and partial reactivation of under-dominant genes in floral organs (Chen and Pikaard, 1997b). In the early 1990s RV in the VH group employed, perhaps for the first time, quantitative RT-PCR for the analysis of nucleolar dominance in plants. Using this method, they determined the levels of homoelogous ETS transcripts in different organs of tomato x potato hybrids showing a strong

nucleolar dominance of tomato genes in leaf but not in anthers and calli (Komarova et al., 2004). Activation of partner units was apparently linked to changes in DNA methylation and chromatin organization. Indeed, profound changes in condensation of rDNA chromatin were observed between tobacco leaf and root (Koukalova et al., 2005).

Earlier cytogenetic data from *N. tabacum* indicated that unconverted parental units of *N. sylvestris*-origin were highly methylated, perhaps located at a locus on chromosome S12 (Lim et al., 2000) that does not show secondary constrictions at metaphase (a hallmark of genetic inactivity). More recently (Dadejova et al., 2007) used RT-PCR to investigate expression of rRNA genes in a number of synthetic *Nicotiana* hybrids (including reciprocal crosses) with a genomic composition similar to natural *N. tabacum* (SSTT), *N. rustica* (PPUU) and *N. arentsii* (UUWW) allotetraploids differing in age and genome donors. They found strong uniparental rDNA silencing of *N. paniculata* genes in *N. paniculata* × *N. undulata* F1 hybrids (genome composition corresponding to natural *N. rustica*), whereas *N. sylvestris* × *N. tomentosiformis* (*N. tabacum*) and *N. undulata* × *N. wigandioides* (*N. arentsii*) F1 hybrids showed little or no silencing (i.e., co-dominance). Based on these observations, Kovarik et al. (2008) proposed that nucleolar dominance, established early in allopolyploid formation including F1 hybrids, plays a significant role in further molecular evolution of rDNA. It has been suggested that epigenetic silencing of rDNA loci makes them less vulnerable to homogenization and more likely to be lost, perhaps thousands or millions of years later.

In 2003, AK visited the VH lab in Tübingen. At that time, both groups were fascinated by the dynamics of repetitive sequences, especially their species- and sometimes even population-specific features. As a result of fruitful discussions during a stroll around the old castle (whose walls remember the discovery of DNA by Friedrich Miescher) we wrote a review paper on the behavior of satellite DNA repeats in plant hybrids (Hemleben et al., 2007). The outcome was a productive collaborative research on allopolyploidy carried out in labs at the University of Tübingen, Queen Mary College of the University of London and the Czech Academy of Science. The findings are significant for our understanding of evolution of plant species since the world of angiosperms is largely dominated by allopolyploids.

CONCLUDING REMARKS

Looking back over 50 years of research on rDNA, it is amazing to see how many complex factors stimulated or supported this field of study. To start with: In the early 1960s, the interest in molecular biology was rising: Basic functions of cell organelles were clarified. Electron microscopy, ultracentrifugation, radioactive labelling, and gel electrophoresis and hybridization assays were the key experimental methods, revealing a complex and highly organized ribosome construction and assembly process. As in bacteria, but with variation in size in all eukaryotes, the 40S subunit were shown to contain 18S rRNA while the 60S subunit associates with 25/28S (plants/animals), 5.8S and a smaller 5S rRNA. Both subunits form a ribonucleoprotein complex which assembles into a functional 80S ribosome. Transcription of the 18S, 5.8S and 25S was found to occur by RNA polymerase I as a large polycistronic precursor (pre-rRNA) containing tandem repeat sequences of the 18S-5.8S-25S rDNA multigene family, which is subsequently processed into the mature rRNAs, whereas the 5S rRNA genes (5S rDNA) are transcribed by RNA polymerase III. Gene technologies, gene cloning and DNA sequencing showed that strongly conserved parts alternate with more variable regions, especially in the intergenic regions of these multigene families. Cooperation of various experts delivered further valuable information and opened up the broad field of plant molecular phylogeny, molecular evolution and molecular systematics, supported by the rapidly growing field of whole genome sequencing coupled with more and more sophisticated computer evaluation of the data obtained. New techniques of cytogenetics made genome evolution and species formation visible. Especially for plants, the process of homo- and allopolyploidy by natural hybridization could be followed. The phenomenon of nucleolar dominance helped us to understand the mechanisms of silencing of rDNA from one partner probably leading to elimination of rRNA genes in allopolyploids.

AUTHOR CONTRIBUTIONS

VH conceived and designed the study. AK, DG, NB, RV, and VH contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AK: The work was supported by the Czech Science Foundation (grant 19-03442S). VH: Our work was always supported by the German Science Foundation (DFG), some parts by the Federal Ministry of Research and Technology (BMFT), today called BMBF; the Alexander v. Humboldt Foundation (Bonn, Germany) greatly supported the research stay of NB and RV at Tübingen University which is gratefully acknowledged. DG is grateful to former UK governments for financial support during BSc studies at University of East Anglia and PhD (Science & Industry Award) at the University of Edinburgh and to the University of Nottingham. RA is grateful to the Ministry of Education and Science of Ukraine, the Alexander von Humboldt Foundation, German Academic Exchange Service (DAAD) and Austrian Exchange Service (OeAD) for the financial support provided to him and members of his research group. NB is grateful for the earlier support from the Alexander von Humboldt Foundation; he is currently supported by a personal grant from the Huaiyin Normal University (Huai'an, China).

ACKNOWLEDGMENTS

Many Diploma and PhD students and Postdocs contributed to the rDNA research in VH's lab, especially Birgit Lewecke, Harald Friedrich (†), Katrin Schiebel, Jutta Gerstner, Martin Ganal,

Ramon Torres-Ruiz, Ulrike Zentgraf, Klaus King, Jürgen Jobst, Klaus Unfried, Günther Schweizer, Christine Zanke, Guido Grimm, Matthias Schlee, Thomas Grabe, which is highly acknowledged. DG is grateful to Ulrich Loening for his guidance and encouragement during the course of my PhD studies at the University of Edinburgh. Thanks also to the late Max Birnstiel for shining the light on nucleolar rDNA. VH and RV: Many Diploma and PhD students and Postdocs were involved in the study of rDNA in our labs. Especially, we would like to express our appreciation for the great contribution of Irina Panchuk, Nataliya Komarova, Yuri Tynkevich, Antonina Shelyfist, Olga Ishchenko and Kristina Bushyla. NB is grateful for the contribution made by Ljudmilla Borisjuk, Volodymyr Kolosha, Nikolai Friesen, Anton Stepanenko, Guimin Chen and Gregoriy Petjuch. NB and RV are grateful to Galina Miroshnichenko (Moscow, Russia) for her support/advice at early stage of our research careers. We are grateful to the members of The International Research Network (IRN) Polyploidy and Biodiversity (PolyDiv). AK wish to remember Yoong Lim (Queen Mary, University of London) and Ann Kenton (Kew Garden, Richmond, UK), both excellent cytogeneticists, who pioneered molecular cytogenetics of *Nicotiana* in early 90s, both of whom have sadly passed.

REFERENCES

Adesnik, M., and Levinthal, C. (1969). Synthesis and maturation of ribosomal RNA in *Escherichia coli*. *J. Mol. Biol.* 46, 281–303. doi: 10.1016/0022-2836(69)90422-7

Appels, R., Moran, L. B., and Gustafson, J. P. (1986). Rye heterochromatin. 1. Studies on clusters of the major repeating sequence and the identification of a new dispersed repetitive sequence element. *Can. J. Genet. Cytol.* 28, 645–657. doi: 10.1139/g86-094

Appels, R., Wang, P., and Islam, S. (2021). Integrating wheat nucleolus structure and function: variation in the wheat ribosomal RNA and protein genes. *Front. Plant Sci.*

Bendich, A. J., and Anderson, R. S. (1974). Novel properties of satellite DNA from muskmelon. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1511–1515. doi: 10.1073/pnas.71.4.1511

Bersaglieri, C., and Santoro, R. (2019). Genome organization in and around the nucleolus. *Cell* 8:579. doi: 10.3390/cells8060579

Birnstiel, M., Speirs, J., Purdom, I., Jones, K., and Loening, U. E. (1968). Properties and composition of isolated ribosomal DNA satellite of *Xenopus laevis*. *Nature* 219, 454–463. doi: 10.1038/219454a0

Birnstiel, M. L., Wallace, H., Stirlin, L., and Fischberger, M. (1966). Localization of the ribosomal DNA complements in the nucleolar organizer regions in *Xenopus laevis*. *Natl. Cancer Inst. Monogr.* 23, 431–447.

Borisjuk, N., Borisjuk, L., Komarnytsky, S., Timeva, S., Hemleben, V., Gleba, Y., et al. (2000). Tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes. *Nat. Biotechnol.* 18, 1303–1306. doi: 10.1038/82430

Borisjuk, N., Borisjuk, L., Petjuch, G., and Hemleben, V. (1994). Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae. *Genome* 37, 271–279. doi: 10.1139/g94-038

Borisjuk, N. V., Davidjuk, Y. M., Kostishin, S. S., Miroshnichenco, G. P., Velasco, R., Hemleben, V., et al. (1997). Structural analysis of rDNA in the genus *Nicotiana*. *Plant Mol. Biol.* 35, 655–660. doi: 10.1023/A:1005856618898

Borisjuk, N., and Hemleben, V. (1993). Nucleotide sequence of the potato rDNA intergenic spacer. *Plant Mol. Biol.* 21, 381–384. doi: 10.1007/Bf00019953

Borisjuk, N. V., Kostyshin, S. S., Volkov, R. A., and Miroshnichenko, G. P. (1989). Ribosomal RNA gene organization in higher plants from *Nicotiana* genus. *Mol. Biol.* 23, 1067–1074.

Borisjuk, N. V., and Miroshnichenko, G. P. (1989). Organization of ribosomal RNA genes in *Brassica oleracea*, *B. campestris* and their natural allotetraploid hybrid *B. napus*. *Genetika* 25, 417–424.

Borisjuk, N. V., Momot, V. P., and Gleba, Y. (1988). Novel class of rDNA repeat units in somatic hybrids between *Nicotiana* and *Atropa*. *Theor. Appl. Genet.* 76, 108–112. doi: 10.1007/Bf00288839

Brown, D. D., and Gurdon, J. B. (1964). Absence of ribosomal RNA synthesis in anucleolate mutant of *Xenopus laevis*. *Proc. Natl. Acad. Sci. U. S. A.* 51, 139–146. doi: 10.1073/pnas.51.1.139

Brown, J. W. S., and Shaw, P. J. (1998). Small nucleolar RNAs and pre-rRNA processing in plants. *Plant Cell* 10, 649–657. doi: 10.1105/tpc.10.5.649

Brown, D. D., Wensink, P. C., and Jordan, E. (1972). A comparison of ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: evolution of tandem genes. *J. Mol. Biol.* 63, 57–73. doi: 10.1016/0022-2836(72)90521-9

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X. N., et al. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucl. Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028

Capesius, I. (1997). Analysis of the ribosomal RNA gene repeat from the moss *Funaria hygrometrica*. *Plant Mol. Biol.* 33, 559–564. doi: 10.1023/a:1005740031313

Caudy, A. A., and Pikaard, C. S. (2002). *Xenopus* ribosomal RNA gene intergenic spacer elements conferring transcriptional enhancement and nucleolar dominance-like competition in oocytes. *J. Biol. Chem.* 277, 31577–31584. doi: 10.1074/jbc.M202737200

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Chandrashekaran, M. (2006). Erwin Bünning (1906–1990): a centennial homage. *J. Biosci.* 31, 5–12. doi: 10.1007/BF02705230

Chen, Z. J., Comai, L., and Pikaard, C. S. (1988). Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis allopolyploids*. *Proc. Natl. Acad. Sci. USA* 95, 14891–14896. doi: 10.1073/pnas.95.25.14891

Chen, Z. J., and Pikaard, C. S. (1997a). Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev.* 11, 2124–2136. doi: 10.1101/gad.11.16.2124

Chen, Z. J., and Pikaard, C. S. (1997b). Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3442–3447. doi: 10.1073/pnas.94.7.3442

Chen, G., Stepanenko, A., and Borisjuk, N. (2021). Mosaic arrangement of the 5S rDNA in the aquatic plant *Landoltia punctata* (Lemnaceae). *Front. Plant Sci.* 12:678689. doi: 10.3389/fpls.2021.678689

Cox, B. J., and Turnock, G. (1973). Synthesis and processing of ribosomal RNA in cultured plant cells. *Eur. J. Biochem.* 37, 367–376. doi: 10.1111/j.1432-1033.1973

Dadejova, M., Lim, K. Y., Souckova-Skalicka, K., Matyasek, R., Grandbastien, M. A., Leitch, A., et al. (2007). Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana* allotetraploids. *New Phytol.* 174, 658–668. doi: 10.1111/j.1469-8137.2007.02034.x

Dahlberg, A. E., and Peacock, A. C. (1971). Studies of 16 and 23 S ribosomal RNA of *Escherichia coli* using composite gel electrophoresis. *J. Mol. Biol.* 55, 61–74. doi: 10.1016/0022-2836(71)90281-6

Delcasso-Tremousaygue, D., Grellet, F., Panabieres, F., Ananiev, E. D., and Delseny, M. (1988). Structural and transcriptional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant. *Eur. J. Biochem.* 172, 767–776. doi: 10.1111/j.1432-1033.1988.tb13956.x

Denk, T., and Grimm, G. (2010). The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. *Taxon* 59, 351–366. doi: 10.1002/tax.592002

Denk, T., Grimm, G. W., and Hemleben, V. (2005). Patterns of molecular and morphological differentiation in *Fagus* (Fagaceae): phylogenetic implications. *Am. J. Bot.* 92, 1006–1016. doi: 10.3732/ajb.92.6.1006

Dodsworth, S., Kovarik, A., Grandbastien, M.-A., Leitch, I. J., and Leitch, A. R. (2020). "Repetitive DNA dynamics and polyploidization in the genus *Nicotiana*

(Solanaceae)," in *The Tobacco Plant Genome. Compendium of Plant Genomes*. eds. N. Ivanov, N. Sierro and M. Peitsch (Cham: Springer).

Dover, G. A. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111–117. doi: 10.1038/299111a0

Earley, K., Lawrence, R. J., Pontes, O., Reuther, R., Enciso, A. J., Silva, M., et al. (2006). Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev.* 20, 1283–1293. doi: 10.1101/gad.1417706

Eickbush, T. H., and Eickbush, D. G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175, 477–485. doi: 10.1534/genetics.107.071399

Ellis, T. H., Lee, D., Thomas, C. M., Simpson, P. R., Cleary, W. G., Newman, M. A., et al. (1988). 5S rRNA genes in *Pisum*: sequence, long range and chromosomal organization. *Mol. Gen. Genet.* 214, 333–342. doi: 10.1007/BF00337732

Flavell, R. B., O'Dell, M., and Thompson, W. F. (1988). Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. *J. Mol. Biol.* 204, 523–534. doi: 10.1016/0022-2836(88)90352-X

Fraser, R., and Loening, U. (1974). RNA synthesis during synchronous cell division in cultured explants of Jerusalem artichoke tuber. *J. Exp. Bot.* 25, 847–859. doi: 10.1093/jxb/25.5.847

Friedrich, H., Hemleben, V., Meagher, R. B., and Key, J. L. (1979). Purification and restriction endonuclease mapping of soybean 18 S and 25 S ribosomal RNA genes. *Planta* 146, 467–473. doi: 10.1007/Bf00380862

Fulnecek, J., Lim, K. Y., Leitch, A. R., Kovarik, A., and Matyasek, R. (2002). Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species. *Heredity* 88, 19–25. doi: 10.1038/sj.hdy.6800001

Fulnecek, J., Matyasek, R., Kovarik, A., and Bezdek, M. (1998). Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. *Mol. Gen. Genet.* 259, 133–141. doi: 10.1007/s004380050798

Galian, J. A., Rosato, M., and Rossello, J. A. (2012). Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba*. *Heredity* 108, 640–646. doi: 10.1038/hdy.2012.2

Gall, J. G. (1981). Chromosome structure and the C-value paradox. *J. Cell Biol.* 91, 3s–14s. doi: 10.1083/jcb.91.3.3s

Ganal, M., and Hemleben, V. (1986). Comparison of the ribosomal-RNA genes in 4 closely related Cucurbitaceae. *Plant Syst. Evol.* 154, 63–77. doi: 10.1007/Bf00984868

Ganal, M., Torres, R., and Hemleben, V. (1988). Complex structure of the DNA ribosomal spacer of *Cucumis sativus* (cucumber). *Mol. Gen. Genet.* 212, 548–554. doi: 10.1007/BF00330863

Garcia, S., Borowska-Zuchowska, N., Wendel, J. F., Ainouche, M., Kuderova, A., and Kovarik, A. (2020). The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids and cryptic introgressants. *Front. Plant Sci.* 11:41. doi: 10.3389/fpls.2020.00041

Garcia, S., Garnatje, T., and Kovarik, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121, 389–394. doi: 10.1007/s00412-012-0368-7

Garcia, S., and Kovarik, A. (2013). Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity* 111, 23–33. doi: 10.1038/hdy.2013.11

Garcia, S., Kovarik, A., Leitch, A. R., and Garnatje, T. (2016). Cytogenetic features of rRNA genes across land plants: analysis of the plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Garcia, S., Lim, K. Y., Chester, M., Garnatje, T., Pellicer, J., Valles, J., et al. (2009). Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Chromosoma* 118, 85–97. doi: 10.1007/s00412-008-0179-z

Garcia, S., Panero, J. L., Siroky, J., and Kovarik, A. (2010). Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC Plant Biol.* 10:176. doi: 10.1186/1471-2229-10-176

Gerstner, J., Schiebel, K., von Waldburg, G., and Hemleben, V. (1988). Complex organization of the length heterogeneous 5' external spacer of mung bean (*Vigna radiata*) ribosomal DNA. *Genome* 30, 723–733. doi: 10.1139/g88-120

Gleba, Y. Y., Hinnisdaels, S., Sidorov, V. A., Kaleda, V. A., Parokonny, A. S., Boryshuk, N. V., et al. (1988). Intergeneric asymmetric hybrids between *Nicotiana plumbaginifolia* and *Atropa belladonna* obtained by "gamma-fusion". *Theor. Appl. Genet.* 76, 760–766. doi: 10.1007/BF00303523

Goldberg, R. B., Bemis, W. P., and Siegel, A. (1972). Nucleic acid hybridization studies within the genus Cucurbitaceae. *Genetics* 72, 253–266. doi: 10.1093/genetics/72.2.253

Goodspeed, T.H. (1954). *The Genus Nicotiana*. Massachusetts, USA: Waltham

Gray, M. W. (2017). Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Mol. Biol. Cell* 28, 1285–1287. doi: 10.1091/mbc.E16-07-0509

Grebenstein, B., Röser, M., Sauer, W., and Hemleben, V. (1998). Molecular phylogenetic relationships in Aveneae (Poaceae) species and other grasses as inferred from ITS1 and ITS2 rDNA sequences. *Plant Syst. Evol.* 213, 233–250. doi: 10.1007/Bf00985203

Grierson, D. (1972). The Synthesis of rRNA in Developing Primary Leaves of *Phaseolus aureus*. PhD, Edinburgh, UK.

Grierson, D. (1984). "Structure and expression of nuclear genes," in *Plant Molecular Biology*. eds. D. Grierson and S. N. Covey (Glasgow: Blackie).

Grierson, D., and Hemleben, V. (1977). Ribonucleic acid from higher plant *Matthiola incana*. Molecular weight measurements and DNA RNA hybridization studies. *Biochim. Biophys. Acta* 475, 424–436. doi: 10.1016/0005-2787(77)90058-2

Grierson, D., and Loening, U. E. (1972). Distinct transcription products of ribosomal genes in 2 different tissues. *Nature New Biol.* 235, 80–82. doi: 10.1038/newbio235080a0

Grierson, D., and Loening, U. (1974). Ribosomal RNA precursors and synthesis of chloroplast and cytoplasmic ribosomal acid in leaves of *Phaseolus aureus*. *Eur. J. Biochem.* 44, 501–507. doi: 10.1111/j.1432-1033.1974.tb03508.x

Grierson, D., Rogers, M. E., Sartiran, M., and Loening, U. E. (1970). The synthesis of ribosomal RNA in different organisms: structure and evolution of rRNA precursor. *Cold Spring Harb. Symp. Quant. Biol.* 35, 589–598. doi: 10.1101/Sqb.1970.035.01.074

Grierson, D., and Smith, H. (1973). The synthesis and stability of ribosomal RNA in blue-green algae. *Eur. J. Biochem.* 36, 280–285. doi: 10.1111/j.1432-1033.1973.tb02911.x

Grimm, G. W., Schlee, M., Komarova, N. Y., Volkov, R. A., and Hemleben, V. (2005). "Low-level taxonomy and intrageneric evolutionary trends in higher plants," in *From Plant Taxonomy to Evolutionary Biology. Nova Acta Leopoldina NF, Vol. 92, No. 342*. eds. P. K. Endress, U. Lüttge and B. Parthier (Stuttgart: Wissenschaftl. Verlagsges. mbH), 129–145.

Grossmann, K., Friedrich, H., and Seitz, U. (1980). Purification and characterization of chromatin-bound DNA-dependent RNA polymerase-I from parsley (*Petroselinum crispum*) - influence of nucleoside triphosphates. *Biochem. J.* 191, 165–171. doi: 10.1042/bj1910165

Grossmann, K., Seitz, U., and Seitz, H. U. (1979). Transcription and release of RNA in isolated nuclei from parsley cells. *Z. Naturforsch. C. J. Biosci.* 34, 431–435. doi: 10.1515/znc-1979-5-619

Gruendler, P., Unfried, I., Pointner, R., and Schweizer, D. (1989). Nucleotide sequence of the 25S-18S ribosomal gene spacer from *Arabidopsis thaliana*. *Nucl. Acids Res.* 17, 6395–6396. doi: 10.1093/nar/17.15.6395

Grummt, I., Sorbaz, H., Hofmann, A., and Roth, E. (1985). Spacer sequences downstream of the 28S RNA coding region are part of the mouse rDNA transcription unit. *Nucl. Acids Res.* 13, 2293–2304. doi: 10.1093/nar/13.7.2293

Handa, H., Kanamori, H., Tanaka, T., Murata, K., Kobayashi, F., Robinson, S. J., et al. (2018). Structural features of two major nucleolar organizer regions (NORs), nor-B1 and nor-B2, and chromosome-specific rRNA gene expression in wheat. *Plant J.* 96, 1148–1159. doi: 10.1111/tpj.14094

Hartley, M. R., and Ellis, R. J. (1973). Ribonucleic acid synthesis in chloroplasts. *Biochem. J.* 134, 249–262. doi: 10.1042/bj1340249

Hemleben, V. (1972). Untersuchungen zur Biosynthese und Funktion von Nucleinsäuren in höheren Pflanzen. Habilitation, University of Tübingen.

Hemleben, V., Ermisch, N., Kimmich, D., Leber, B., and Peter, G. (1975). Studies on fate of homologous DNA applied to seedlings of *Matthiola incana*. *Eur. J. Biochem.* 56, 403–411. doi: 10.1111/j.1432-1033.1975.tb02246.x

Hemleben, V., and Grierson, D. (1978). Evidence that in higher plants the 25S and 18S rRNA genes are not interspersed with genes for 5S rRNA. *Chromosoma* 65, 353–358. doi: 10.1007/Bf00286414

Hemleben, V., Grierson, D., and Dertmann, H. (1977). Use of equilibrium centrifugation in actinomycin cesium chloride for purification of ribosomal DNA. *Plant Sci. Lett.* 9, 129–135. doi: 10.1016/0304-4211(77)90090-6

Hemleben, V., Kovarik, A., Torres, R. A., Volkov, R. A., and Beridze, T. (2007). Plant highly repeated satellite DNA: molecular evolution, distribution and

use for identification of hybrids. *Syst. Biodivers.* 5, 277–289. doi: 10.1017/S147720000700240X

Hemleben, V., and Werts, D. (1988). Sequence organization and putative regulatory elements in the 5S rRNA genes of 2 higher plants (*Vigna radiata* and *Matthiola incana*). *Gene* 62, 165–169. doi: 10.1016/0378-1119(88)90591-4

Hemleben, V., and Zentgraf, U. (1994). "Structural organisation and regulation of transcription by RNA polymerase I of plant nuclear ribosomal genes," in *Plant Promoters and Transcription Factors*. ed. L. Nover (Berlin/Heidelberg: Springer-Verlag), 3–24.

Hemleben-Vielhaben, V. (1966). Characterization of rapidly labelled nucleic acids in tissues of plant seedlings. *Z. Naturforsch. Pt. B.* 21, 983–992. doi: 10.1515/znb-1966-1016

Hoang, P. T. N., Fiebig, A., Novák, P., Macas, J., Cao, H. X., Stepanenko, A., et al. (2020). Chromosome-scale genome assembly for the duckweed *Spirodela intermedia*, integrating cytogenetic maps, PacBio and Oxford Nanopore libraries. *Sci. Rep.* 10:19230. doi: 10.1038/s41598-020-75728-9

Ingle, J., Timmis, J. N., and Sinclair, J. (1975). The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. *Plant Physiol.* 55, 496–501. doi: 10.1104/pp.55.3.496

Ishchenko, O. O., Bednarska, O. I., and Panchuk, I. I. (2021). Application of 5S ribosomal DNA for molecular taxonomy of subtribe Loliinae (Poaceae). *Cytol. Genet.* 55, 10–18. doi: 10.3103/S0095452721010096

Ishchenko, O. O., Mel'nyk, V. M., Parnikoza, I. Y., Budzhak, V. V., Panchuk, I. I., Kunakh, V. A., et al. (2020). Molecular organization of 5S ribosomal DNA and taxonomic status of *Avenella flexuosa* (L.) Drejer (Poaceae). *Cytol. Genet.* 54, 505–513. doi: 10.3103/S0095452720060055

Ishchenko, O. O., Panchuk, I. I., Andreev, I. O., Kunakh, V. A., and Volkov, R. A. (2018). Molecular organization of 5S ribosomal DNA of *Deschampsia antarctica*. *Cytol. Genet.* 52, 416–421. doi: 10.3103/S0095452718060105

Jo, S. H., Koo, D. H., Kim, J. F., Hu, C.-G., Lee, S., Yang, T. J., et al. (2009). Evolution of ribosomal DNA-derived satellite repeat in tomato genome. *BMC Plant Biol.* 9:42. doi: 10.1186/1471-2229-9-42

Jobst, J., King, K., and Hemleben, V. (1998). Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family Cucurbitaceae. *Mol. Phylogenetics Evol.* 9, 204–219. doi: 10.1006/mpev.1997.0465

Jung, H., Winefield, C., Bombarelym, A., Prentis, P., and Waterhouse, P. (2019). Tools and strategies for long-read sequencing and de novo assembly of plant genomes. *Trends Plant Sci.* 24, 700–724. doi: 10.1016/j.tplants.2019.05.003

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., et al. (2008). Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5833–5838. doi: 10.1073/pnas.0709698105

King, K., Torres, R. A., Zentgraf, U., and Hemleben, V. (1993). Molecular evolution of the intergenic spacer in the nuclear ribosoma RNA genes of Cucurbitaceae. *J. Mol. Evol.* 36, 144–152. doi: 10.1007/Bf00166250

Kirov, I., Gilyok, M., Knyazev, A., and Fesenko, I. (2018). Pilot satellitome analysis of the model plant, *Physcomitrella patens*, revealed a transcribed and high-copy IGS related tandem repeat. *Comp. Cytogenet.* 12, 493–513. doi: 10.3897/CompCytogen.v12i4.31015

Komarova, N. Y., Grabe, T., Huigen, D. J., Hemleben, V., and Volkov, R. A. (2004). Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. *Plant Mol. Biol.* 56, 439–463. doi: 10.1007/s11103-004-4678-x

Komarova, N. Y., Grimm, G. W., Hemleben, V., and Volkov, R. A. (2008). Molecular evolution of 35S rDNA and taxonomic status of *Lycopersicon* within *Solanum* sect. Petota. *Plant Syst. Evol.* 276, 59–71. doi: 10.1007/s00606-008-0091-2

Koukalova, B., Fojtova, M., Lim, K. Y., Fulnecek, J., Leitch, A. R., and Kovarik, A. (2005). Dedifferentiation of tobacco cells is associated with ribosomal RNA gene hypomethylation, increased transcription, and chromatin alterations. *Plant Physiol.* 139, 275–286. doi: 10.1104/pp.105.061788

Kovarik, A., Dadejova, M., Lim, Y. K., Chase, M. W., Clarkson, J. J., Knapp, S., et al. (2008). Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. *Ann. Bot.* 101, 815–823. doi: 10.1093/aob/mcn019

Kovarik, A., Fajkus, J., Koukalova, B., and Bezdek, M. (1996). Species-specific evolution of telomeric and rDNA repeats in the tobacco composite genome. *Theor. Appl. Genet.* 92, 1108–1111. doi: 10.1007/BF00224057

Kovarik, A., Matyasek, R., Lim, K. Y., Skalicka, K., Koukalova, B., Knapp, S., et al. (2004). Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. *Biol. J. Linn. Soc.* 82, 615–625. doi: 10.1111/j.1095-8312.2004.00345.x

Layat, E., Sáez-Vásquez, J., and Tourmente, S. (2012). Regulation of pol I-transcribed 45S rDNA and pol III-transcribed 5S rDNA in *Arabidopsis*. *Plant Cell Physiol.* 53, 267–276. doi: 10.1093/pcp/pcr177

Leaver, C. J., and Key, J. L. (1970). Ribosomal RNA synthesis in plants. *J. Mol. Biol.* 49, 671–680. doi: 10.1016/0022-2836(70)90290-1

Leber, B., and Hemleben, V. (1979). Structure of plant nuclear and ribosomal DNA containing chromatin. *Nucl. Acids Res.* 7, 1263–1282. doi: 10.1093/nar/7.5.1263

Leweke, B., and Hemleben, V. (1982). Organization of rDNA in chromatin: plants. *Cell Nucleus* 11, 225–253.

Lim, K. Y., Kovarik, A., Matyasek, R., Bezdek, M., Lichtenstein, C. P., and Leitch, A. R. (2000). Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. *Chromosoma* 109, 161–172. doi: 10.1007/s004120050424

Lim, K. Y., Kovarik, A., Matyasek, R., Chase, M. W., Clarkson, J. J., Grandbastien, M. A., et al. (2007). Sequence of events leading to near complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol.* 175, 756–763. doi: 10.1111/j.1469-8137.2007.02121.x

Lim, K. Y., Matyasek, R., Kovarik, A., and Leitch, A. (2004a). Genome evolution in allotetraploid *Nicotiana*. *Biol. J. Linn. Soc.* 82, 599–606. doi: 10.1111/j.1095-8312.2004.00344.x

Lim, K. Y., Skalicka, K., Koukalova, B., Volkov, R. A., Matyasek, R., Hemleben, V., et al. (2004b). Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. *Genetics* 166, 1935–1946. doi: 10.1534/genetics.166.4.1935

Lin, C. Y., Chen, Y. M., Guilfoyle, T. J., and Key, J. L. (1976). Selective modulation of RNA polymerase I activity during growth transitions in the soybean seedling. *Plant Physiol.* 58, 614–617. doi: 10.1104/pp.58.5.614

Loening, U. E. (1969). The determination of molecular weight of ribonucleic acid by polyacrylamide-gel electrophoresis - effects of changes in conformation. *Biochem. J.* 113, 131–138. doi: 10.1042/bj1130131

Lunerova, J., Renny-Byfield, S., Matyasek, R., Leitch, A., and Kovarik, A. (2017). Concerted evolution rapidly eliminates sequence variation in rDNA coding regions but not in intergenic spacers in *Nicotiana tabacum* allotetraploid. *Plant Syst. Evol.* 303, 1043–1060. doi: 10.1007/s00606-017-1442-7

Macas, J., Navratilova, A., and Meszaros, T. (2003). Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma* 112, 152–158. doi: 10.1007/s00412-003-0255-3

Manoharlal, R., Saiprasad, G. V. S., and Kovarik, A. (2019). "Smoking and tobacco use," in *New Research on Tobacco*. ed. U. Dreher (New York: Nova Science Publishers Inc), 23–82.

Marazia, T., Barsanti, P., and Maggini, F. (1980). Individual quantitative rDNA variation in 3 species of the Cucurbitaceae family. *Biochem. Genet.* 18, 509–517. doi: 10.1007/Bf00484398

Matsuda, K., and Siegel, A. (1967). Hybridization of plant ribosomal RNA to DNA - isolation of a DNA component rich in ribosomal RNA cistrons. *Proc. Natl. Acad. Sci. U. S. A.* 58, 673–680. doi: 10.1073/pnas.58.2.673

Matyasek, R., Dobesova, E., Huska, D., Jezkova, I., Soltis, P. S., Soltis, D. E., et al. (2016). Interpopulation hybridization generates meiotically stable rDNA epigenetic variants in allotetraploid *Tragopogon mirus*. *Plant J.* 85, 362–377. doi: 10.1111/tpj.13110

Matyasek, R., Lim, K. Y., Kovarik, A., and Leitch, A. R. (2003). Ribosomal DNA evolution and gene conversion in *Nicotiana rustica*. *Heredity* 91, 268–275. doi: 10.1038/sj.hdy.6800333

McClintock, B. (1934). The relationship of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Z. Zellforsch Microsk.* 21, 294–398. doi: 10.1007/BF00374060

Michael, T., Bryant, D., Gutierrez, R., Borisjuk, N., Chu, P., Zhang, H., et al. (2017). Comprehensive definitions of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* 89, 617–635. doi: 10.1111/tpj.13400

Michael, T. P., and Van Buren, R. (2020). Building near-complete plant genomes. *Curr. Opin. Plant Biol.* 54, 26–33. doi: 10.1016/j.pbi.2019.12.009

Miroshnichenko, G. P., Borisjuk, N. V., and Volkov, R. A. (1989). Organization of rDNA repeat units in the Solanaceae sexual and parasexual hybrids. *Biochemistry* 54, 669–675.

Mohannath, G., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific selective nucleolus organizer inactivation in *Arabisopsis* is a chromosome position-effect phenomenon. *Proc. Natl. Acad. Sci. U. S. A.* 113, 13426–13431. doi: 10.1073/pnas.1608140113

Moss, T. (2004). At the crossroads of growth control: making ribosomal RNA. *Curr. Opin. Genet. Dev.* 14, 210–217. doi: 10.1016/j.gde.2004.02.005

Nagl, W., Ehrendorfer, F., and Hemleben, V. (1979). "Genome and Chromatin: Organization, Evolution and Function," in *Plant Systematics and Evolution. 2nd Edn.* eds. W. Nagl, F. Ehrendorfer and V. Hemleben (Wien: Springer Verlag).

Navashin, M. (1934). Chromosomal alterations caused by hybridisation and their bearing upon certain genetic problems. *Cytologia* 5, 169–203.

Nieto Feliner, G., and Rossello, J. A. (2012). "Concerted evolution of multigene families and homeologous recombination," in *Plant Genome Diversity*. ed. J. F. Wendel (Wien: Springer-Verlag), 171–194.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., et al. (1981). Secondary structure model for 23S ribosomal RNA. *Nucl. Acids Res.* 9, 6167–6189. doi: 10.1093/nar/9.22.6167

Pikaard, C. S., and Reeder, R. H. (1988). Sequence elements essential for function of the *Xenopus laevis* ribosomal DNA enhancers. *Mol. Cell. Biol.* 8, 4282–4288. doi: 10.1128/Mcb.8.10.4282

Poczai, P., and Hyvonen, J. (2010). Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol. Biol. Rep.* 37, 1897–1912. doi: 10.1007/s11033-009-9630-3

Pontvianne, F., Abou-Ellail, M., Douet, J., Comella, P., Matia, I., Chandrasekhara, C., et al. (2010). Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* 6:e1001225. doi: 10.1371/journal.pgen.1001225

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rabanal, F. A., Mandáková, T., Soto-Jiménez, L. M., Greenhalgh, R., Parrott, D. L., Lutzmayer, S., et al. (2017). Epistatic and allelic interactions control expression of ribosomal RNA gene clusters in *Arabidopsis thaliana*. *Genome Biol.* 18:75. doi: 10.1186/s13059-017-1209-z

Rathgeber, J., and Capesius, I. (1990). Nucleotide sequence of the intergenic spacer and the 18S ribosomal-RNA gene from mustard (*Sinapis alba*). *Nucl. Acids Res.* 18, 1288–1288. doi: 10.1093/nar/18.5.1288

Ribeiro, T., Dos Santos, K. G., Richard, M. M., Sévignac, M., Thareau, V., Geffroy, V., et al. (2017). Evolutionary dynamics of satellite DNA repeats from *Phaseolus* beans. *Protoplasma* 254, 791–801. doi: 10.1007/s00709-016-0993-8

Robicheau, B. M., Susko, E., Harrigan, A. M., and Snyder, M. (2017). Ribosomal RNA genes contribute to the formation of pseudogenes and junk DNA in the human genome. *Genome Biol. Evol.* 9, 380–397. doi: 10.1093/gbe/evw307

Rogers, S. O., and Bendich, A. J. (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* 9, 509–520. doi: 10.1007/BF00015882

Rogers, S. O., Honda, S., and Bendich, A. J. (1986). Variation in the ribosomal RNA genes among individuals of *Vicia faba*. *Plant Mol. Biol.* 6, 339–345. doi: 10.1007/Bf00034941

Röser, M., Winterfeld, G., Grebenstein, B., and Hemleben, V. (2001). Molecular diversity and physical mapping of 5S rDNA in wild, and cultivated oat grasses (Poaceae: Aveneae). *Mol. Phylogenet. Evol.* 21, 198–217. doi: 10.1006/mpev.2001.1003

Sáez-Vásquez, J., and Delseny, M. (2019). Ribosome biogenesis in plants: from functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell* 31, 1945–1967. doi: 10.1105/tpc.18.00874

Sáez-Vásquez, J., and Pikaard, C. S. (2000). RNA polymerase I holoenzyme-promoter interactions. *J. Biol. Chem.* 275, 37173–37180. doi: 10.1074/jbc.M006057200

Salim, D., Bradford, W. D., Freeland, A., Cady, G., Wang, J., Pruitt, S. C., et al. (2017). DNA replication stress restricts ribosomal DNA copy number. *PLoS Genet.* 13:e1007006. doi: 10.1371/journal.pgen.1007006

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Sardana, R., Odell, M., and Flavell, R. (1993). Correlation between the size of the intergenic regulatory region, the status of cytosine methylation of ribosomal-RNA genes and nucleolar expression in wheat. *Mol. Gen. Genet.* 236, 155–162. doi: 10.1007/Bf00277107

Schiebel, K., and Hemleben, V. (1989). Nucleotide sequence of the 18S-25S spacer region from rDNA of mung bean. *Nucl. Acids Res.* 17, 2852–2852. doi: 10.1093/nar/17.7.2852

Schlee, M., Göker, M., Grimm, G. W., and Hemleben, V. (2011). Genetic patterns in the *Lathyrus pannonicus* complex (Fabaceae) reflect ecological differentiation rather than biogeography and traditional subspecific division. *Biol. J. Linn. Soc.* 165, 402–421. doi: 10.1111/j.1095-8339.2011.01125.x

Schlotterer, C., and Tautz, D. (1994). Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* 4, 777–783. doi: 10.1016/S0960-9822(00)00175-5

Schweizer, G., Borisjuk, N., Borisjuk, L., Stadler, M., Stelzer, T., Schilde, L., et al. (1993). Molecular analysis of highly repeated genome fractions in *Solanum* and their use as markers for the characterization of species and cultivars. *Theor. Appl. Genet.* 85-85, 801–808. doi: 10.1007/Bf00225022

Scoles, C. J., Gill, B. S. Z., Xin, Y., Clarke, B. C., McIntyre, C. L., Chapman, C., et al. (1988). Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the Triticeae. *Plant Syst. Evol.* 160, 105–122.

Seitz, U., and Seitz, U. (1973). Biosynthetic pathway of ribosomal-RNA in blue-green algae (*Anacystis nidulans*). *Archiv fur Mikrobiologie.* 90, 213–222. doi: 10.1007/Bf00424973

Seitz, U., and Seitz, U. (1979). Molecular weight of rRNA precursor molecules and their processing in higher plant cells. *Z. Naturforsch. C. J. Biosci.* 34, 253–258. doi: 10.1515/znc-1979-3-416

Selig, C., Wolf, M., Müller, T., Dandekar, T., and Schultz, J. (2008). The ITS2 database II: homology modelling RNA structure for molecular systematics. *Nucl. Acids Res.* 36, D377–D380. doi: 10.1093/nar/gkm827

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the *Arabidopsis* NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12, 387. doi: 10.1038/s41467-020-20728-6

Soltis, D. E., and Soltis, P. S. (2016). Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Divers.* 38, 264–270. doi: 10.1016/0022-2836(88)90353-1

Souza, G., Marques, A., Ribeiro, T., Dantas, L. G., Speranza, P., Guerra, M., et al. (2019). Allopolyploidy and extensive rDNA site variation underlie rapid karyotype evolution in *Nothoscordum* section *Nothoscordum* (Amaryllidaceae). *Bot. J. Linn. Soc.* 190, 215–228. doi: 10.1093/botlinnean/boz008

Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J., and Erdmann, V. A. (1998). 5S rRNA Data Bank. *Nucl. Acids Res.* 26, 156–159. doi: 10.1093/nar/26.1.156

Thompson, W. F., and Flavell, R. B. (1988). DNAse-I sensitivity of ribosomal RNA genes in chromatin and nucleolar dominance in wheat. *J. Mol. Biol.* 204, 535–548. doi: 10.1016/0022-2836(88)90353-1

Tomecki, R., Sikorski, P. J., and Zakrzewska-Placzek, M. (2017). Comparison of preribosomal RNA processing pathways in yeast, plant and human cells - focus on coordinated action of endo- and exoribonucleases. *FEBS Lett.* 591, 1801–1850. doi: 10.1002/1873-3468.12682

Torres, R. A., Zentgraf, U., and Hemleben, V. (1989). Species and genus specificity of the intergenic spacer (IGS) in the ribosoma RNA genes of Cucurbitaceae. *Z. Naturforsch. C. J. Biosci.* 44, 1029–1034. doi: 10.1515/znc-1989-11-1224

Torres-Ruiz, R. A., and Hemleben, V. (1994). Pattern and degree of methylation in ribosomal RNA genes of *Cucurbita pepo* L. *Plant Mol. Biol.* 26, 1167–1179. doi: 10.1007/Bf00040697

Tulpová, Y.O., Kovařík, A., Toegelová, H., Navrátilová, P., Kapustová, V., Hřibová, E., et al. (2020). Anatomy, transcription dynamics and evolution of wheat ribosomal RNA loci deciphered by a multi-omics approach. BioRxriv [Preprint]. doi: 10.1101/2020.08.29.273623

Tynkevich, Y. O., and Volkov, R. A. (2014). Structural organization of 5S ribosomal DNA in *Rosa rugosa*. *Cytol. Genet.* 48, 1–6. doi: 10.3103/S0095452714010095

Tynkevich, Y. O., and Volkov, R. A. (2019). 5S ribosomal DNA of distantly related *Quercus* species: molecular organization and taxonomic application. *Cytol. Genet.* 53, 459–466. doi: 10.3103/S0095452719060100

Unfried, K., Schiebel, K., and Hemleben, V. (1991). Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not in *Vigna angularis*. *Gene* 99, 63–68. doi: 10.1016/0378-1119(91)90034-9

Volkov, R. A., Bachmair, A., Panchuk, I. I., Kostyshyn, S. S., and Schweizer, D. (1999a). 25S-18S rDNA intergenic spacer of *Nicotiana sylvestris* (Solanaceae): primary and secondary structure analysis. *Plant Syst. Evol.* 218, 89–97. doi: 10.1007/bf01087037

Volkov, R. A., Borisjuk, N. V., Kostishin, S. S., and Panchuk, I. I. (1991). Variability of rRNA genes in *Nicotiana* correlates with the chromosome reconstruction. *Mol. Biol.* 25, 442–450.

Volkov, R. A., Borisjuk, N. V., Panchuk, I. I., Schweizer, D., and Hemleben, V. (1999b). Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. *Mol. Biol. Evol.* 16, 311–320. doi: 10.1093/oxfordjournals.molbev.a026112

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Syst. Biodivers.* 5, 261–276. doi: 10.1017/S1477200007002447

Volkov, R. A., Komarova, N. Y., Panchuk, I. I., and Hemleben, V. (2003). Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. Petota (genus *Solanum*). *Mol. Phylogenetics Evol.* 29, 187–202. doi: 10.1016/s1055-7903(03)00092-7

Volkov, R., Kostishin, S., Ehrendorfer, E., and Schweizer, D. (1996). Molecular organization and evolution of the external transcribed rDNA spacer region in two diploid relatives of *Nicotiana tabacum* (Solanaceae). *Plant Syst. Evol.* 201, 117–129. doi: 10.1007/Bf00989055

Volkov, R. A., Kozeretska, I. A., Kyryachenko, S. S., Andreev, I. O., Maidanyuk, D. N., Parnikoza, I. Y., et al. (2010). Molecular evolution and variability of ITS1–ITS2 in populations of *Deschampsia antarctica* from two regions of the maritime Antarctic. *Pol. Sci.* 4, 469–478. doi: 10.1016/j.polar.2010.04.011

Volkov, R. A., Medina, F. J., Zentgraf, U., and Hemleben, V. (2004). "Molecular cell biology: organization and molecular evolution of rDNA, nucleolar dominance and nucleolus structure," in *Progress in Botany*. *Vol. 65*. eds. K. Esser, U. Lüttge, W. Beyschlag and J. Murata (Berlin, Heidelberg, New York: Springer Verlag), 106–146.

Volkov, R. A., Panchuk, I. I., Borisjuk, N. V., Maluszynska, J., and Hemleben, V. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol.* 17, 21–24. doi: 10.1186/s12870-017-0978-6

Volkov, R. A., Zanke, C., Panchuk, I. I., and Hemleben, V. (2001). Molecular evolution of 5S rDNA of *Solanum* species (sect. Petota): application for molecular phylogeny and breeding. *Theor. Appl. Genet.* 103, 1273–1282. doi: 10.1007/s001220100670

Wang, W., Wan, T., Becher, H., Kuderova, A., Leitch, I. J., Garcia, S., et al. (2019). Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Ann. Bot.* 123, 767–781. doi: 10.1093/aob/mcy172

Weis, B. L., Kovacevic, J., Missbach, S., and Schleiff, E. (2015). Plant-specific features of ribosome biogenesis. *Trends Plant Sci.* 20, 729–740. doi: 10.1016/j.tplants.2015.07.003

Wenzel, W., and Hemleben, V. (1982a). A comparative study of genomes in angiosperms. *Plant Syst. Evol.* 139, 209–227. doi: 10.1007/Bf00989326

Wenzel, W., and Hemleben, V. (1982b). DNA-sequence organization and RNA complexity in *Matthiola incana* (Brassicaceae). *Plant Syst. Evol.* 140, 75–86. doi: 10.1007/Bf02409898

Wicke, S., Costa, A., Munoz, J., and Quandt, D. (2011). Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol. Phylogenet. Evol.* 61, 321–332. doi: 10.1016/j.ympev.2011.06.023

Yan, Q., Zhu, C. M., Guang, S. H., and Feng, X. Z. (2019). The functions of non-coding RNAs in rRNA regulation. *Front. Genet.* 10:290. doi: 10.3389/fgene.2019.00290

Yokota, Y., Kawata, T., Iida, Y., Kato, A., and Tanifuji, S. (1989). Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. *J. Mol. Evol.* 29, 294–301. doi: 10.1007/Bf02103617

Zakrzewska-Placzek, M., Souret, F. F., Sobczyk, G. J., Green, P. J., and Kufel, J. (2010). *Arabidopsis thaliana* XRN2 is required for primary cleavage in the pre-ribosomal RNA. *Nucl. Acids Res.* 38, 4487–4502. doi: 10.1093/nar/gkq172

Zentgraf, U., and Hemleben, V. (1992). Complex-formation of nuclear proteins with the RNA polymerase-I promoter and repeated elements in the external transcribed spacer of *Cucumis sativus* ribosomal DNA. *Nucl. Acids Res.* 20, 3685–3691. doi: 10.1093/nar/20.14.3685

Zentgraf, U., and Hemleben, V. (1993). Nuclear proteins interact with RNA polymerase-I promoter and repeated elements of the 5' external transcribed spacer of the rDNA of cucumber in a single-stranded stage. *Plant Mol. Biol.* 22, 1153–1156. doi: 10.1007/Bf00028984

Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W., and Wilson, A. C. (1980). Rapid duplication and loss of genes-coding for the alpha-chains of hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 77, 2158–2162. doi: 10.1073/pnas.77.4.2158

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

REVIEW
published: 24 December 2021
doi: 10.3389/fpls.2021.686586

Integrating Wheat Nucleolus Structure and Function: Variation in the Wheat Ribosomal RNA and Protein Genes

Rudi Appels[1,2], Penghao Wang[3] and Shahidul Islam[4]*

[1] AgriBio, Centre for AgriBioscience, La Trobe University, Bundoora, VIC, Australia, [2] Faculty of Veterinary and Agricultural Science, Melbourne, VIC, Australia, [3] School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia, [4] Centre for Crop Innovation, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

We review the coordinated production and integration of the RNA (ribosomal RNA, rRNA) and protein (ribosomal protein, RP) components of wheat cytoplasmic ribosomes in response to changes in genetic constitution, biotic and abiotic stresses. The components examined are highly conserved and identified with reference to model systems such as human, Arabidopsis, and rice, but have sufficient levels of differences in their DNA and amino acid sequences to form fingerprints or gene haplotypes that provide new markers to associate with phenotype variation. Specifically, it is argued that populations of ribosomes within a cell can comprise distinct complements of rRNA and RPs to form units with unique functionalities. The unique functionalities of ribosome populations within a cell can become central in situations of stress where they may preferentially translate mRNAs coding for proteins better suited to contributing to survival of the cell. In model systems where this concept has been developed, the engagement of initiation factors and elongation factors to account for variation in the translation machinery of the cell in response to stresses provided the precedents. The polyploid nature of wheat adds extra variation at each step of the synthesis and assembly of the rRNAs and RPs which can, as a result, potentially enhance its response to changing environments and disease threats.

Keywords: nucleolar dominance, rRNA structure, ribosomal protein (RP), sequence variation, associated phenotypes

OPEN ACCESS

Edited by:
Nikolai Borisjuk,
Huaiyin Normal University, China

Reviewed by:
Hirokazu Handa,
Kyoto Prefectural University, Japan
Jason Sims,
Max F. Perutz Laboratories GmbH, Austria

****Correspondence:***
Rudi Appels
rudiappels5@gmail.com

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *27 March 2021*
Accepted: *08 November 2021*
Published: *24 December 2021*

Citation:
Appels R, Wang P and Islam S (2021) Integrating Wheat Nucleolus Structure and Function: Variation in the Wheat Ribosomal RNA and Protein Genes. Front. Plant Sci. 12:686586. doi: 10.3389/fpls.2021.686586

INTRODUCTION

The wheat seed, like all plant seeds, is a "special living state" retaining only 10–15% moisture in the "dry" state to which tissue in the grain has adapted for long-term storage (Bonner and Varner, 1965; Swift and O'Brien, 1972). The dry scutellum/embryo (="wheat germ") is a source of viable ribosomes for the translation of stored messenger RNA when the seed rehydrates, and historically

the wheat germ was an early source of ribosomes for the *in vitro* translation of isolated messenger RNA (Armache et al., 2010). In this review, the overall 3D structure of wheat ribosomes determined by Armache et al. (2010) is used as a basis for reviewing the component RNA and proteins in wheat in order to understand changes in ribosome structure and the adaptation of the translational process in cells to biotic and abiotic stresses.

The RNA components of wheat cytoplasmic ribosomes (rRNA) are encoded by very large tandem arrays of gene units that are transcribed by a dedicated RNA polymerase (RNApol1) within a compartment of the nucleus called the nucleolus (Appels and Dvorak, 1982; Barker et al., 1988; Thompson and Flavell, 1988; Handa et al., 2018; Tulpová et al., 2020). The hexaploid nature of wheat, comprising seven chromosome pairs in the A, B, and D genomes, means that there exists a complex set of interactions between the genes coding for the rRNA and ribosomal proteins (RPs) to generate flexibility in the composition of ribosomes. An estimated 42–100 Mb of the genome is devoted to coding for rRNA and, at this mega-level, structural differences between the major rDNA loci on chromosomes 1B and 6B are argued to be important in the autoregulation of rDNA expression and the silencing of minor rDNA arrays. Accompanying the production of rRNA, a total of 170 proteins have been assigned (high confidence, HC) to the cytoplasmic ribosome subunits 40S and 60S, and organelle subunits 30S and 50S in wheat proteome studies (Ford et al., 2011; Duncan et al., 2017; Islam et al., 2020). The coordinated production and integration of both RNA and protein components into the wheat cytoplasmic ribosome assembly processes are considered in the present review in the context of adjustment to selection pressures and response to biotic and abiotic stresses. The compilation of protein sequences in this review focused on wheat *per se* and entries from Arabidopsis and rice. UniProt identifiers were used to recover amino acid sequences for searches against the *Triticum aestivum* L. reference genome using BLASTP in Ensembl[1] and manual recovery of low-confidence (LC) gene models in a genome viewer Apollo instance[2] for the Chinese Spring wheat genome assembly ver1 to curate their status. The predicted gene models were confirmed using the Phyre2[3] 3D structure predictions for domains and inspecting protein domains in Pfam[4] and InterPro[5] databases. Coordinates for the Traes gene IDs from Ensembl Plants were used to locate the gene models in the CS wheat genome ver1 assembly using the Apollo instance for the wheat CS genome assembly ver1 noted above with the gene models curated by the alignments to RNA-seq from the standard tissues, grain, leaf, roots, stem, and spike (**Supplementary Table 1**). During the course of producing this review, version 2.1 of the wheat Chinese Spring genome was published (Zhu et al., 2021) and although version 1 gene identifiers are used in the present manuscript, **Supplementary Table 2** provides a cross-reference of the version 1 gene identifiers to the version 2.1 gene identifiers that will appear in Ensembl Plants updates in due course.

[1]http://plants.ensembl.org/index.html

[2]http://apollo.tgac.ac.uk/Wheat_IWGSC_WGA_v1_0_browser

[3]http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

[4]http://pfam.xfam.org/search

[5]http://www.ebi.ac.uk/interpro/search/sequence-search

THE rDNA GENOME REGIONS IN WHEAT NUCLEI

The nucleolus organizer region (NOR) is a classical feature of metaphase chromosomes where major rDNA loci are visible as so-called secondary constrictions in the respective chromosomes (**Figure 1A**; Silva et al., 2008) in addition to the primary constriction (centromere) visible in all chromosomes. In **Figure 1**, silver (Ag) staining was used to visualize the location of the NOR because it detects concentrations of nucleic acid molecules and argyrophilic Ag-binding proteins and provides a sensitive assay for the NOR and associated rDNA transcription (Silva et al., 2008) in both plants and animals (**Figure 1**). The Ag staining assays a set of argyrophilic protein markers associated with active ribosomal genes (AgNOR proteins) that reduce and thus deposit the Ag under cytochemically acidic conditions in NORs of metaphase chromosomes and nucleoli of interphase nuclei in fixed tissue (Sirri et al., 2000; Caperta et al., 2002). Studies in model systems have identified nucleolin and nucleophosmin as major AgNOR proteins in the nucleolus with smaller contributions coming from RNA polymerase 1 (RNA pol I) subunits and the transcription factor UBF/UAF30 (Sirri et al., 2000; see also **Figure 2** and **Supplementary Table 1** for wheat homologs). The wheat models for nucleophosmin show a very clear acidic amino acid cluster feature considered to be characteristic for Ag binding, DDLMKNNFGVEGDEDDEDDDEDED, in the C-terminal region (Sirri et al., 2000).

Nucleolus organizer regions are generally considered to feature three subcompartments, namely pale staining structures fibrillar centers (FCs) comprised of fine fibrils, a surrounding densely stained fibrillar component (DFC), and a granular component (GC) in which the FC and DFC are embedded. Given the fundamental nature of ribosome production, it seems reasonable to adopt this model of the nucleolus established in human studies, for wheat, and to envision transcription of the rDNA occurring at the interface between the FCs and the DFC with nascent transcripts and pre-ribosomes progressively migrating from the DFC to the GC (Thiry and Lafontaine, 2005). The high concentration of a limited set of nucleic acid molecules plus associated proteins and rDNA chromatin are considered to form the nucleolus as a phase separated feature within the nucleus not bound by membranes (concept developed in model systems, Berry et al., 2015). Within the nucleolus in wheat, the ribosomal chromatin is located in interphase nuclei as condensed perinucleolar chromatin knobs varying between 3 and 4 μm, and intranucleolar condensed dots ranging from 1 to 2 μm in diameter (**Figure 1B**; Silva et al., 2008). Variation involving the relative distribution of rDNA in the intranucleolar dots and perinucleolar knobs is found. The long arrays of rDNA units that are not transcribed into rRNA are generally located in condensed chromatin (**Figure 1B**). The long arrays of rDNA active chromatin are distributed across most of the nucleolus

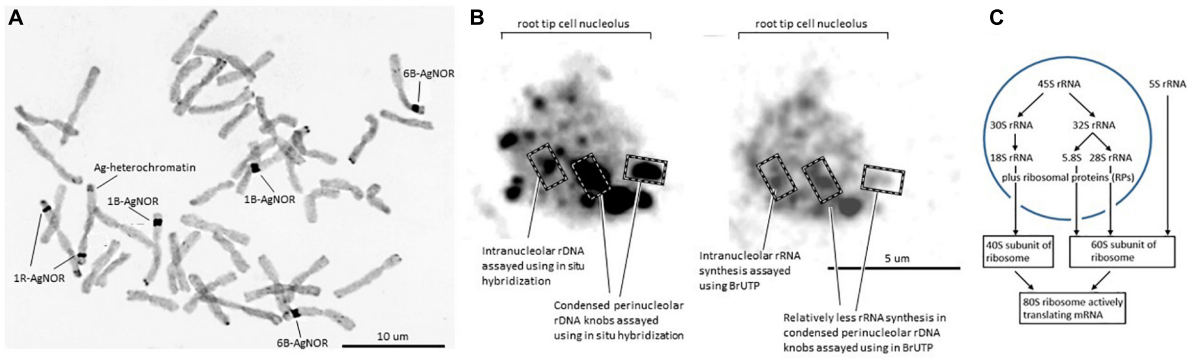


FIGURE 1 | (A) Metaphase chromosomes from a wheat line with a 1R disomic addition (modified from Silva et al., 2008, doi.org/10.1371/journal.pone.0003824.g002). Dark staining regions corresponding to the nucleolar organizers are indicated as AgNOR; other dark staining regions of the rye chromosome correspond to heterochromatin/C-banding regions (marked as Ag-heterochromatin). **(B)** Wheat nucleolus, modified from Silva et al.(2008, doi.org/10.1371/journal.pone.0003824.g002) and Correll et al.(2019, doi.org/10.3390/cells8080869). In the left panel, dark regions are the condensed chromatin regions with rDNA as assayed by the wheat clone pTa71 (Gerlach and Bedbrook, 1979) as a probe for *in situ* hybridization. Silva et al. (2008) defined intranuclear dots housing the more dispersed rDNA that is more active in transcription (examples indicated) and perinucleolar knobs housing the condensed rDNA that was relatively inactive in transcription. The distribution of newly synthesized RNA in the right panel is distributed throughout the nucleolus as measured by the incorporation of labeled UTP (BrUTP). **(C)** Summary of the flow of RNA processing, and assembly into the mature cytoplasmic small ribosomal subunit (40S) and the large subunit (60S), to form the active 80S ribosome, including the independent production of 5S rRNA.

based on the dispersed distribution of newly synthesized RNA (**Figure 1B**). The 45S rRNA precursor that is initially formed interacts with RPs and assembly factors (AFs) in the nucleolar space for processing to the 18S, 5.8S, and 26S rRNAs, the foundations for the ribosome subunits (**Figure 1C**).

The 5SrRNA indicated in **Figure 1C** is produced from independent loci composed of tandem arrays of short units (120-bp gene sequence, intergenic spacer ca. 280 bp, **Supplementary Figure 2**) predominantly in the satellite region of chromosome 1BS and smaller numbers on 1DS, and tandem arrays of long units (intergenic spacer ca. 380 bp) predominantly on 5BS with smaller numbers on 5AS and 5DS (Dvorak et al., 1989; Reddy and Appels, 1989; Baum and Bailey, 2001; Sergeeva et al., 2017). Genome arrays for the 5SrRNA gene units on 5BS have been shown to be in uninterrupted tandem arrays of the long units, and lower numbers of these long units in clusters interrupted by inserts of mobile elements (Sergeeva et al., 2017).

The genome regions encoding rRNA components at the NORs are a challenge in genome assemblies because of the extreme length of the repetitive arrays of near identical gene units and only recently optical mapping has provided a clear view of the tandem arrays in wheat (Tulpová et al., 2020). In the human genome studies, some full and partial arrays of 18S-5.8S-28S rDNA units are assembled on each of the five acrocentric p-arms, but the centers of these arrays are currently represented by a total of 11.5 Mbp of unknown sequence (Ns) in the assemblies on chromosomes 13, 14, 15, 21, and 22. The arrays are near-identical tandem repeats, and so while the content of these arrays is known, the variation in substructure within the arrays remains to be determined. A detailed study of the 18S-5.8S-28S rDNA units on human chromosome 21 succeeded in assembling long arrays (Kim et al., 2018) and revealed heterogeneity at the single locus for rDNA units on this chromosome suggesting the possibility that this variation may relate to the dynamics of ribosome function. In wheat, the application of optical mapping on high molecular weight DNA isolated from flow sorted chromosome arms physically identified the arrays of 2,813 units on 6BS (26.87 Mb), 1,378 (12.96 Mb) on 1BS, and 170 (1.63 Mb) on 5D. A small number of complete units were identified on 1AS (29 units, 0.43 Mb). The optical mapping thus provided a minimum estimate of 4,390 units (42 Mb) in tandem arrays, within the wheat cv Chinese Spring (CS) genome, but does not allow for rDNA fragments external to the tandem arrays and dispersed in the genome. The latter most likely contribute to the higher numbers of total rDNA units reported by Handa et al. (2018). Similar to some human rDNA arrays, the rDNA arrays on chromosome 6BS of wheat showed heterogeneity due to interspersion of non-rDNA sequences in contrast to the relatively homogeneous arrays on 1BS and 5DS (Handa et al., 2018; Tulpová et al., 2020). Studies in wheat are generally agreed on the relative proportions of rDNA units on 1BS (31%), 6BS (61%), and 5DS (8%) (Flavell and O'Dell, 1976; Handa et al., 2018; Tulpová et al., 2020).

Assembly factors initiate the formation of small and large "preribosomal" subunits that accumulate in the GC. The AgNOR proteins, such as nucleolin and nucleophosmin (**Figure 1**), appear to be within the network of AFs active in the ribosome assembly process based on their capacity to bind RNA. Nucleolin is an RNA-binding protein in the nucleoli of all eukaryotes, and in plants it has been well studied at the structure-function level in Arabidopsis, rice (Pontvianne et al., 2007), and pea (Nasirudin et al., 2004). In wheat, the nucleolin gene models are located on chromosomes 2A, 2D, 5D, 7A, 7B, and 7D and display RNA-binding domains in Pfam (RMM domains, **Supplementary Figure 1**) and Phyre2 analyses (Kelley et al., 2015; predicted fold c6r5kH). The C-terminal glycine-rich region is homologous to the region assigned to have helicase attributes as reported by Nasirudin et al. (2004) and has a number of

FIGURE 2 | Termination of rDNA transcription and the engagement of RNA polymerase II to ensure efficient rRNA production. **(A)** Summary of the 9,000 bp rDNA unit in wheat. The 26S gene-3′ downstream is followed by the repetitive region of ca. 130 bp units leading to the promoter region and start of transcription (see **Supplementary Figure 3**) in the 18S gene 5′ upstream region. **(B)** The 26S gene-3′ downstream ca. 650 bp that distinguishes the S1–S4 subtypes of rDNA in wheat. The *Taq*1, *Hae*III, and *Hha*1 that could be related to the structure defined for the clone pTa250 by Appels and Dvorak (1982) are boxed. The (*Hae*III) sites were not identified in pTa250. The TATA box regions are emphasized in red and are postulated to engage the TATA box-binding protein (TBP) for contributing to engaging RNA polymerase II in forming an R-loop shield. **(C)** Model for the protein complexes formed in the 26S gene-3′ downstream region to ensure RNA pol I transcription is properly terminated, based on the studies on human rDNA structure and function (Fomproix et al., 1998; Abraham et al., 2020; doi.org/10.1038/s41586-020-2497-0). The prominent nucleolar proteins, nucleolin and fibrillarin, are possibly involved as discussed in the earlier text (Nasirudin et al., 2004; Pontvianne et al., 2007; Rodriguez-Corona et al., 2015). The inset (boxed) indicates the 3′-end of the rice 26S RNA gene to confirm the 3′-end of the wheat 26S RNA gene, according to Armache et al. (2010).

differences between the wheat gene models (**Supplementary Figure 1**). Based on RNA-seq alignments, all the homologs are highly expressed in the standard, root, leaf, stem, spike, and grain tissues assayed except for the 5D homolog, which has a low-to-moderate expression. The N- and C-terminal regions of the nucleolins are the main source of single amino acid differences between the gene models, with the 5D nucleolin having an additional major difference in missing the entire C-terminal glycine-rich domain and seem to be a pseudogene. The 2A and 2D wheat gene models have 19 exons, and the 7A, 7B, and 7D gene models have 13 exons, similar to the distinction between Arabidopsis AtNuc-L2 (18 exons, At3g18610) and AtNuc-L1 (15 exons, At1g48920) nucleolins. Comparisons between the genomic regions of wheat varieties available in the genome viewer DAWN (Watson-Haigh et al., 2018) for the wheat nucleolins indicated 50 positions in the coding regions of the genes (CDS) were captured as variable positions assayed as single nucleotide polymorphisms (SNPs) at the DNA sequence level. In the wheat varieties inspected for variation in the nucleolin gene region, a total of nine of the variable positions in the genome sequence also showed differences at the amino acid sequence level between the homologous gene models of the reference genome *per se* (**Supplementary Figure 1**). An additional set of nine amino acid sequence differences between the 2A and 2D nucleolins and the 7A, 7B, and 7D nucleolins of the reference genome were not captured as SNPs at the genome level in the wheat varieties set available in DAWN.

Mutation studies in Arabidopsis (Pontvianne et al., 2007) emphasize the fundamental importance of nucleolin in plant growth and development, and for wheat the alignment of the 2A and 2D nucleolin gene models indicates the conserved domains (RMM, RNA-binding domains, **Supplementary Figure 1**) that actually show relatively high levels of polymorphisms at the genome level in terms of the distribution of SNPs. The glycine-rich region at the C-terminal end of the protein shows variation between the 2A and 2D nucleolin gene models, this has not been captured by variation in breeding programs as judged from inspecting varieties available in the DAWN viewer (**Supplementary Figure 1**). The combination of changes at the amino acid and SNP levels can be considered as gene haplotypes and can be viewed as defining fingerprints that could provide markers for associating rRNA loci that are preferentially expressed in terms of the phenomenon of nucleolar dominance described below or in response to environmental stress. The concept of gene haplotypes (broad sense as per Wilhelm et al., 2013) providing function-based markers for associating particular ribosome-related protein (RPs) variants with responses in translation activity of the ribosomes to stress is considered further (later) as we establish a haplotype dictionary for the RPs in wheat.

The fibrillarin protein is another major component of the nucleolus (Rodriguez-Corona et al., 2015) that co-localizes with AgNOR proteins located in the DFC and FC in model systems. The fibrillarin gene is conserved in model plants and animals and thus allows the wheat homologs to be identified on chromosomes 6A, 6B, 6D, 7A, 7B, and 7D. As was the case for the nucleolin gene homologs, RNA-seq alignments indicate that all the fibrillarin homologs are highly expressed in the standard, root, leaf, stem, spike, and grain tissues assayed.

The alignments for fibrillarin shown in **Figure 3A** complement the observations for nucleolin in identifying gene haplotypes that can provide functional markers for associating protein variants with plant phenotypes. The N-terminal glycine-rich parts of the predicted fibrillarins are particularly variable and are actually absent from the 6A and 7A gene models. Since this part of the molecule is responsible for targeting the protein to the nucleolus as suggested by mutation studies in Arabidopsis (review, Rodriguez-Corona et al., 2015), the structural variation within wheat suggests that a considerable flexibility exists for delivering the methyltransferase activity required for modifying rRNA. As is evident in **Figure 3B**, the fibrillarin gene models in wheat show clear haplotype differences between wheat varieties based on the inspection of varieties available in the genome viewer DAWN (Watson-Haigh et al., 2018). The haplotypes are defined by differential SNP distributions. The representative example in **Figure 3B** shows exon 1 in particular to have a clear haplotype difference between the wheat varieties Mace and Lancer, and since this exon encodes the nucleolar-targeting region for fibrillarin, it raises the possibility of associating this genome difference with phenotypes that differentiate wheat cultivars. The association would be considered in the context that fibrillarin is biologically an essential protein (Loza-Muller et al., 2015), well known as a molecular marker of transcriptionally active RNA polymerase. Fibrillarin methyltransferase activity is argued to be the primary methyltransferase for methylated sites early in preribosomal processing and subsequent structural ribosome stability (Rodriguez-Corona et al., 2015). Consistent with this role for fibrillarin in rRNA synthesis, Tessarz et al. (2014) demonstrated that methylation of Q105 or a substitution Q105A in histone 2A by fibrillarin in human and yeast cells specifically increased rRNA synthesis. In wheat, Q105 is substituted Q105H in the same H2A sequence segment and has a Q98 in the same sequence segment where human H2A has a D genome so although the fibrillarins are identical in wheat, human, and yeast, a direct parallel for the effects of H2A methylation needs further experimental work.

VARIATION IN THE NUCLEOLAR DOMINANCE PHENOMENON IN WHEAT

The diversity in major nucleolar proteins, such as nucleolin and fibrillarin, may inform the well-studied phenomenon of nucleolar dominance. Observations on rDNA transcription in the nucleoli of wheat lines with and without a rye chromosome (1R, the major source of rDNA in rye, Silva et al., 2008) or 1U from *Aegilops umbellulata* (Flavell et al., 1988) have indicated that different sources of rDNA regions moderate the expression of each other. Similarly within wheat *per se*, Handa et al. (2018) identified four rDNA unit subtypes (S1–S4) based on differences within the 3′ transcribed spacer regions in Nor-B1 (on 1BS) and Nor-B2 (on 6BS), **Figure 2**, and demonstrated by quantitative PCR that S1 subunits were predominantly expressed. The S2 subunits were relatively more abundant, but only weakly expressed. Overall, 31.4 and 64.1% of the rDNA units have been assigned to the major NORs in 1BS and 6BS, respectively (Tulpová et al., 2020). The minor loci in 5DS and 1AS have 3.9 and 0.7% of the rDNA units, respectively. The expression of S3 subunits on 5DS increased in the ditelosomic genetic stocks Dt1BL (1BS missing) and Dt6BL (6BS missing), suggesting that S3 is subjected to the chromosome-mediated silencing. In the context of the differential distribution of rDNA in the condensed chromatin (**Figure 1B**), Handa et al. (2018) found genome regions adjacent and distal to the major NORs were expanded compared to homologous regions on 1A, 1D, 6A, and 6D, where rDNA loci are no longer present. Handa et al. (2018) suggested that these regions flanking the rDNA loci on chromosomes 1B and 6B could be a potential source for distinguishing the respective rDNA regions for macro (chromatin)-level condensation and render their rDNA units transcriptionally inactive. Similar models based on modifying the chromosome structure around the rDNA units on the *Drosophila* X chromosome have been developed to account for heterochromatin modifying the relative expression of rDNA on the X and Y chromosomes (Hilliker and Appels, 1982).

In wheat, the "non-syntenic" regions of 5–12 Mb of DNA flanking the 1B and 6B rDNA regions are mainly distinguished by a relatively higher transposable element content (Handa et al., 2018) that is often associated with inactive chromatin. At this macro-level and complementing, the cytological observations in **Figure 1B**, the degree of condensation of rDNA chromatin has been assayed using the sensitivity of rDNA to the enzyme

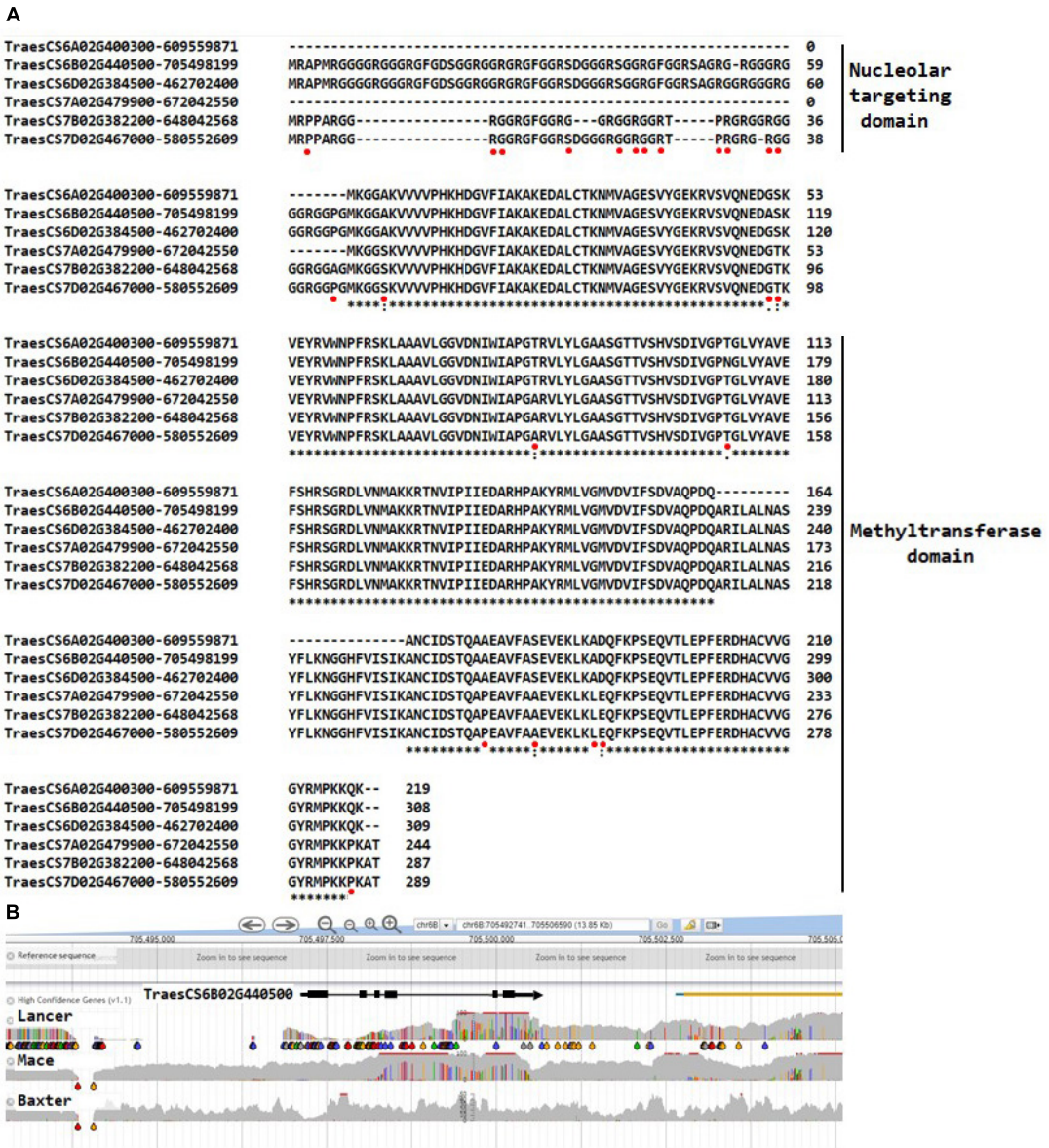


FIGURE 3 | (A) The wheat fibrillarin genes on chromosomes 6A, 6B 6D, 7A, 7B, and 7D. Sequences from Arabidopsis and rice were downloaded, and UniProt identifiers were used to recover amino acid sequences for searches against the *Triticum aestivum* L. genomes using BLASTP in Ensembl (http://plants.ensembl.org/index.html). Alignment at the amino acid level to validate the identification on the wheat gene models against the well-characterized Arabidopsis AAF00542 gene; the fibrillarin TraesCS6D02G462702400 had an identical amino acid sequence to the fibrillarins from human and yeast. The methyltransferase domain is well conserved in contrast to the nucleolar-targeting domain, which shows relatively more diversity. The * indicates the same amino acid is at the respective positions, and spaces and dots indicate amino acid change. The red spots indicate the positions of the single nucleotide polymorphisms (SNPs) in an assessment of wheat varieties available in the viewer DAWN. **(B)** A representative view of the SNP diversity at the DNA sequence level identifying gene haplotypes for TraesCS6B02G440500 using wheat varieties Lancer, Mace, and Baxter as examples. Lancer and Mace have been sequenced in the 10 genome project (Walkowiak et al., 2020). The gray areas indicate the variable genome coverage of the available sequence data, and the colored "drops" identify positions in the sequence that are uniformly changed from that of the reference Chinese Spring genome sequence (orange = change to G; red = change to T; green = change to A; blue = change to C). The SNP analysis was possible using DAWN (Watson-Haigh et al., 2018; http://crobiad.agwine.adelaide.edu.au/dawn/jbrowse/). The DAWN viewer uses standard genome format and can show the location of SNP at the genome sequence level.

DNAase I in isolated nuclei (Thompson and Flavell, 1988). The DNAase I sensitivity assays led to the conclusion that the promoter regions of some wheat rRNA genes possess a more accessible chromatin structure, with the proportion of hypersensitive genes in a NOR argued to be related to observed activity. The genes that displayed hypersensitive DNase I sites were preferentially non-methylated at CCGG sites in the intergenic spacer immediately preceding the promoter. Thus,

the chromatin structure around the promoter of active rRNA genes was differentiated from that in transcriptionally inactive genes and correlated with changes in cytosine methylation. In the case of the wheat-1U addition line studied by Thompson and Flavell (1988), the affinity for (predicted) factors within the DFC/FC interface to assign rDNA units to active transcription is 1U > 1B > 6B. Similarly within wheat *per se*, the additional identification of the S1–S4 rDNA subtypes, and SNP-based haplo-subtypes (Tulpová et al., 2020), is consistent with the preferential recruitment of rDNA units into an active state having a structural basis. Quantitative differences in transposable element levels both within the 6B NOR rDNA arrays (Tulpová et al., 2020) and the regions flanking the 6B NOR could in principle account for the lower transcription of the 6B rDNA genes if the 1B and 6B NORs were competing for limited sites for condensation within the FC, leaving relatively more 1B rDNA for transcription at the DFC/FC interface. A competition model would account for changes in the source of NORs utilized for producing rRNA, depending on the different NORs present in the genetic makeup of the wheat analyzed. Consistent with "non-syntenic" 6B regions as drivers for differentiating the 6B NOR from the 1B NOR is the finding by Handa et al. (2018) that this region is characterized by higher levels of the histone methylation mark, H3K27me3, a chromatin feature that is generally associated with a condensed/gene repressed state of chromatin. The finding that hypersensitive DNase sites included CCGG sites in the intergenic spacer immediately preceding the promoter were preferentially non-methylated in the rDNA units from the 1B NOR compared to the 6B NOR (Thompson and Flavell, 1988) is also consistent with a chromatin structure-based differentiation of the 1B and 6B NOR loci.

Although the structure of the S1–S4 subtypes of rDNA units within the rDNA arrays in wheat NORs could not be extended beyond the 26S gene-3′downstream region shown in **Figure 2** due to ambiguities in recovering experimental chimeric sequences (artifacts) in genome assemblies (H. Handa, personal communication), it is evident that more sequence variation is observed in the so-called non-transcribed spacer (NTS) regions further from the 26S gene-3′downstream region. Extensive polymorphism in the repetitive sequence region comprised of 120–130 bp units within the NTS is well established (Appels and Dvorak, 1982; Lassner et al., 1987; Lagudah et al., 1991) at the level of the number of repetitive units within the NTS and at the DNA sequence level. Duplicated sequences dominate NTS variation and Lassner et al. (1987) identified a consensus sequence of CACGTACACGGA as a signature and basis for the range of variation found, suggesting that the sequence possibly provides sites for within locus recombination or DNA replication slippage events.

In terms of sequence variation within the genes coding for the 18S, 5.8S, and 26S rRNA genes, Tulpová et al. (2020) identified pairs of SNPs for consensus 26S gene sequences from each of the 1BS, 6BS, and 5DS NOR loci that defined unique 26S gene haplotypes. The fact that the 26S gene haplotypes could be defined indicated that there was little if any genetic exchange between the NOR loci even though the long tandem arrays were similar in sequence. This finding was consistent with the conclusions by Lassner et al. (1987) from their sequence comparisons of rDNA clones from the B and D genomes of wheat. Interestingly, the 26S gene haplotypes also allowed the source of rDNA transcripts to be identified in different tissues, and this showed that RNA from mature leaf had the lowest proportion of 6BS transcripts relative to root tip and coleoptile samples. In seeds some unassigned transcripts were found. The possibility that the lower amount of 6B 26S rRNA in the leaf tissue was due to RNA undergoing a faster turnover in this tissue was raised by Tulpová et al. (2020). Consistent with the observations in wheat, the detailed analysis of the structure of tandem rDNA units at Arabidopsis NOR loci (Sims et al., 2021) indicated a clustering of variants that could be traced by SNP haplotypes for the respective 26S rRNA genes. The Arabidopsis study also indicated that ribosomal variants showed tissue-specific expression as well responses to certain stress conditions.

TRANSCRIPTION OF 18S-5.8S-26SrDNA AND 5SrDNA

The RNA polymerases essential for rRNA synthesis are RNA polymerase I (for 18S, 5.8S, and 26S RNAs) and RNA polymerase III (for 5S RNA, see **Supplementary Figure 2**). Unlike some of the RNA pol II subunits, none of the subunits defined in Arabidopsis by Ream et al. (2015) are sufficiently conserved to allow homologs to be identified in the wheat or rice genomes. The subunits shared between RNA pol I, pol II, and pol III (Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12) for RNA pol I do not give clear homologs even though the respective subunits for RNA pol II (from Arabidopsis) can identify homologs in wheat and rice, and this suggests that the boundaries of conservation are not as constrained for RNA pol I and III as they are for RNA pol II. At the genome sequence, there are features of the wheat intergenic rDNA region that can relate to aspects of the controls operating on RNA pol I found in studies of human rDNA transcription (Abraham et al., 2020). In human rDNA transcription, the 3′ region downstream from the 26S rDNA unit has been identified as a point of engagement of RNA pol II in controlling rRNAs expression. The authors argued that RNA pol II generated structures known as R-loops in the intergenic spacers flanking nucleolar rRNA genes (**Figure 2**) and prevented RNA pol I from producing sense intergenic non-coding RNAs (sincRNA) that could disrupt nucleolar organization and rRNA expression. In this context, it is possible that the finding by Handa et al. (2018) of sequence differences in the 26S gene 3′ downstream region in wheat (**Figures 2A,B**) may relate more directly to influencing rDNA transcription activity depending on the efficiency of R-loop formation and associated variation in the activity of RNA pol I. In **Figure 2C**, the structural features of the S1–S4 subtype 26S gene-3′downstream region defined by Handa et al. (2018) are interpreted in the context of RNA pol II engagement in highlighting the prominence of the TATA motifs (classically core elements of RNA pol II promoters) in these regions. Major nucleolus proteins, such as the nucleolin and fibrillarin gene models in wheat (**Supplementary Figure 1**), have RNA-interacting domains that could also contribute to facilitating the

establishment of a proposed R-loop shield. Rodriguez-Corona et al. (2015) noted that the rDNA transcription instability in permeabilized mammalian tissue culture cells infused with fibrillarin antibodies (Fomproix et al., 1998) may be due to the antibodies preventing fibrillarin from contributing to the R-loop shield, which normally blocks sense sincRNA formation by RNA pol I (**Figure 2**).

At the end of the NTS, downstream from the 26S gene 3′-downstream region is the promoter region for forming the correct RNA pol I initiation complex at − 1 to approximately − 200 bp upstream from the start of transcription (Lassner et al., 1987; **Supplementary Figure 3**). The region shows the sequence features of AT and GC clusters that are well characterized for the core promoter region in model systems such as AATGGGGG^{-20}CTAAAACCTC^{-10}GGGTATAGT^{-1} (TATA box underlined), and further upstream for the binding site for the upstream activity factor (UAF), referred to as the upstream-control-element region (G^{-200}GTCCGGGAGA^{-190}AAAAAAGGCC^{-180}; Pisl and Engel, 2020). The TBP and Rrn3 (TF-A1) factors are two significant components that direct RNA pol I into the initiation complex, and they are sufficiently well conserved to allow wheat homologs to be identified using the bioinformatics identification pipeline documented earlier in this review. The wheat homologs include TraesCS1B02G151700 (TBP-1), TraesCS5A02G022000 (TBP-2), TraesCS5B02G018500 (TBP-2), TraesCS5D02G027800 (TBP-2), TraesCS6A02G171400 (Rrn3), and TraesCS6D02G161100 (Rrn3); see also **Supplementary Table 1**. Another significant component for directing RNA pol I to the promoter is the upstream activity factor (UAF, UBF), which belongs to a large family of transcription control proteins characterized by the SWIB/MDM2 domain associated with proteins involved in chromatin remodeling. The wheat homologs could be identified using the yeast UAF30 as a reference (Iida and Kobayashi, 2019) and had the predicted fold of d1v31a in Phyre2 (Kelley et al., 2015) for UAF30. The wheat homologs were found on chromosomes 2A and 2B, TraesCS2A02G488300, TraesCS2B02G515900, and chromosome 7D, TraesCS7D02G242300; see also **Supplementary Table 1**. Inspection of the genome sequence indicated homologous gene models also existed on 2D, 7A, and 7B, but gaps in the genome sequence prevented unambiguous identification. One of the AFs in the spatial and temporal coordination of rRNA production in model systems is the factor Rrp5 (Khoshnevis et al., 2019), and HC gene models were located on chromosomes 1A, 1B, and 1D. The wheat Rrp5 gene models were confirmed based on the presence of a domain covering 14–15% of the CDS sharing a predicted 3D structure, c5c9sB in Phyre2, and were assigned to the gene models TraesCS1A02G06730, TraesCS1B02G085800, and TraesCS1D02G068300; see also **Supplementary Table 1**. A striking feature for all gene models likely to represent factors involved in the different levels of rRNA production is the variation that exists for the gene models in the wheat genome, as was found for nucleolin and fibrillarin. The variation exists at the amino acid sequence level between homologous members of a locus and at a broader genome sequence level where gene haplotypes for SNPs are clear between wheat varieties analyzed for SNPs relative to the reference genome of Chinese Spring ver 1.0 in DAWN (see **Supplementary Figure 4** for the Rrp5 example).

The 5S rRNA component of the rRNA is synthesized independent from the 18S, 5.8S, and 26S rRNA (see **Figure 1C** and **Supplementary Figure 2**), and in model systems, the ribosomal L5 protein (RPL5) has been shown to be an important factor for the correct assembly of 5S rRNA into the 60S subunit (**Figure 4**) together with RPL11, into a feature of the 60S subunits called the central protuberance (CP). The CP feature is close to the peptidyl-transferase center (PTC) and GTPase-associating center sites (Armache et al., 2010). Although the exact function of 5S rRNA is not well defined, Kang et al. (2011) have shown that in wheat the effects of salt, drought, and/or freezing stress caused a rapid accumulation of the RPL5 (TaL5) transcript in seedling leaves. It is thus possible that the variation in the formation of the 5S rRNA–RPL5 complex as a result of quantitative changes and qualitative variation in the RPL5 amino acid sequence (gene-level haplotypes, **Figure 4**) could modify the translation properties of the ribosome to be more suited to the stress conditions.

THE RIBOSOMAL PROTEINS ASSEMBLED INTO RIBOSOMES

The structure of the wheat ribosome has been determined at the 5.5-Å angstrom level of resolution (Armache et al., 2010), and we have used the respective accession numbers of the 40S and 60S RPs in this structure to identify the gene models in the wheat genome. The cryo-EM technology combined with modeling utilizing yeast and other microbe ribosome structures allowed Armache et al. (2010) to compile a consensus structure of a translating wheat ribosome in which RP α-helices were observed as rod-like densities and β-sheets were assigned by smooth surfaces. The authors noted that α-helix pitch and β-sheet strand separation could not be determined. The wheat RPs identified as gene models using the information from Armache et al. (2010) mostly identified homologous gene models at loci on each of the A, B, and D genomes (IWGSCrefSeqver1, IWGSC, 2018). A key criterion for gene models was that they were all highly transcribed in the standard tissues, grain, leaf, roots, stem, and spike (**Supplementary Table 1**). The translated gene models were run in Phyre2 (Kelley et al., 2015) to identify RP domains and were then cross-referenced to InterPro (Blum et al., 2020) for annotating the RP-encoding genes in the wheat genome followed by matching them to Traes ID codes for gene models in the reference wheat genome. Intron– exon structures were checked for consistency with the aligned RNA-seq available in the Apollo genome viewer. The annotations in **Supplementary Table 1** identified 25 groups of Traes IDs for the 40S subunit RPs and 37 groups of Traes IDs for 60S subunit RPs with the respective reference gene (plus UniProt ID), usually from rice, also indicated for each group. The Traes gene model alignments from a given group or subgroup showed high levels of conservation as a foundation for assessing the low levels of variation that defined gene haplotypes at the amino acid level (concept developed further below). Proteome level

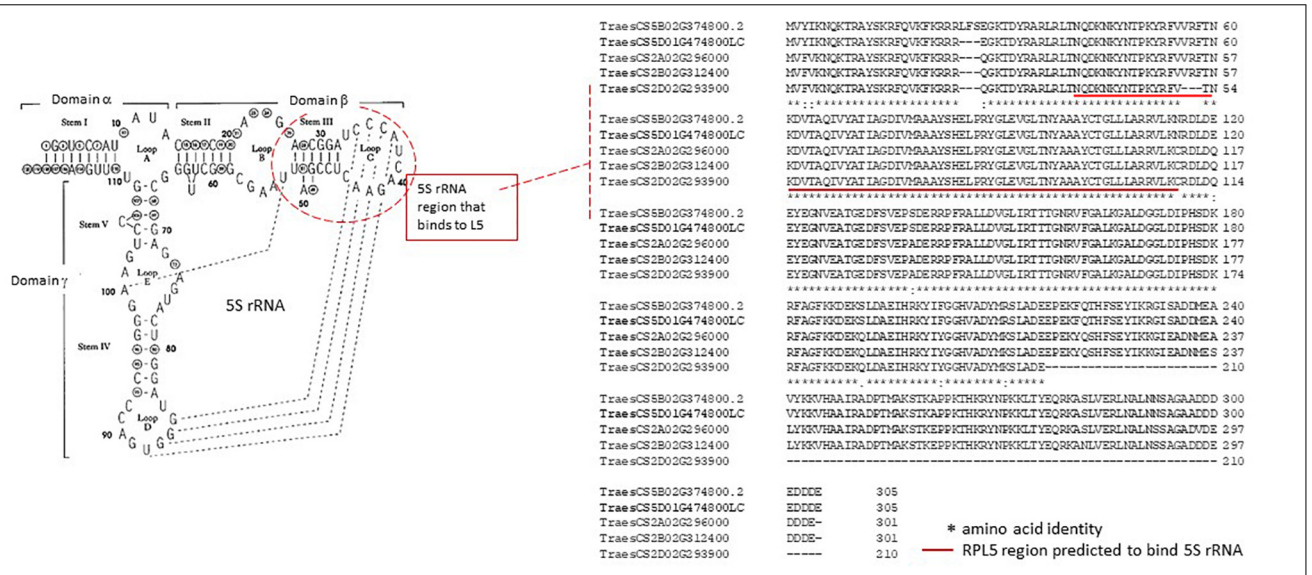


FIGURE 4 | Structure of wheat 5S rRNA and the ribosomal L5 protein (RPL5) from wheat. The consensus 5S rRNA sequence and secondary structure modified from Lee et al. (2006) is in the left panel. The dashed lines link bases that can form H-bonds in a tertiary folding. The structure of wheat 5SRNA indicates the C loop in domain β is important in binding to the RPL5 as discussed in the text. The right panel (dotted line) indicates the wheat RPL5 protein gene models (reference sequence from Kang et al., 2011) and illustrates the gene-level haplotypes discussed in the text for TraesCS2A02G296000, TraesCS2B02G312400, TraesCS2D02G293900, TraesCS5B02G374800.2, and TraesCS5D1G474800LC (see main text for details of the identification process). The TraesCS2D02G293900 gene model has the C-terminal 84 amino acids deleted and is most likely a pseudogene. The significance of the inserted three amino acids at positions 25–27 is unclear. The alignment of the five wheat RPL5 gene models indicates that they are largely separated into either chromosome 2A/2B or chromosome 5B/5D protein haplotypes with the chromosome 2D entry showing a large deletion. At the genome level, the single nucleotide polymorphism (SNP)-based haplotypes were limited to introns and thus useful for tracking the gene region but none of the within-genome variation between homologous genes has been captured in the wheat varieties accessible in the genome viewer DAWN. The red line highlights the region predicted to bind the section of 5SRNA highlighted in a red dashed line in the left panel. The breaks in the rows of *s indicate amino acid changes in the alignments, and these generally define the chromosome 2A/2B or chromosome 5B/5D protein haplotypes.

confirmation that the gene models coded for wheat proteins was obtained for eight of the 25 groups assigned to 40S subunits and for 13 of the 37 groups assigned to 60S subunits. In the case of the RPL6 group, these entries were checked in detail because of an interest in the change in quantity of the protein designated as RPL6, in response to water stress (Islam et al., 2020), and it was found that the two amino acid positions that differentiated TraesCS6B02G225600 and TraesCS6D02G190100 could be identified in the respective peptide maps. The grouping and naming of RPs in **Supplementary Table 1** was ambiguous in some cases due to the presence of shared RNA-binding domains and the historical aspects surrounding RP nomenclature (Ban et al., 2014); the assembly of RPs in **Supplementary Table 1** is intended to provide a sequence-based point of reference for the wheat RPs. Only 10 of the wheat gene models in **Supplementary Table 1** have "LC" added in the IWGSC reference, indicating "low confidence"; for these gene models, the intron– exon structures and reading frames were curated manually to ensure that the models included in **Supplementary Table 1** were in fact HC. For the 40S subunit RPs, Armache et al. (2010) included the RP, RACK1 protein C kinase as one of the proteins in their 3D wheat ribosome assembly, and the respective wheat genome Traes IDs are provided. It is possible that in light of the wheat germination study by Smailov et al. (2020) of the phosphorylation of RPS6 protein by RPS6 kinase, TaS6K1 (AK451448), that the RPS6 kinase may be a more appropriate model than RACK1 protein C kinase with respect to identifying a relevant 40S subunit RP-Traes IDs in the wheat genome. For completeness, both sets of Traes IDs are indicated in **Supplementary Table 1**, in addition to the broader regulator of translation, TOR, which is responsible for activating the RPS6 kinase (Smailov et al., 2020).

The framework for defining the wheat RPs in the context of the detailed 3D compilation of the translating ribosome is provided in **Figure 5**. In **Figures 5A,B**, the maps of the wheat rRNA molecules are shown with the ES annotations, indicating the extension segments to RNA molecules relative to bacterial rRNA reference sequences (Armache et al., 2010). **Figures 5A,B** also indicates the codes for some of the helical structures of the rRNA molecules since these are sites for binding RPs.

The wheat 5.8S rRNA (**Figure 5B**) is assembled into the 60S subunit near the PTC, whereas in model systems, it has been suggested that the highly conserved GAACG in the anti-T loop (see **Figure 5B**) contributes to engaging incoming tRNA-amino acid entities at the PTC through the equally highly conserved (complementary) sequence GTΨC in tRNAs (Nishikawa and Takemura, 1974; Mackay et al., 1980). The 5.8S rRNA has also been argued, in model systems, to contribute to the inclusion of the translation elongation factors involved in the peptide translocation process for peptide synthesis (Elela et al., 1994).

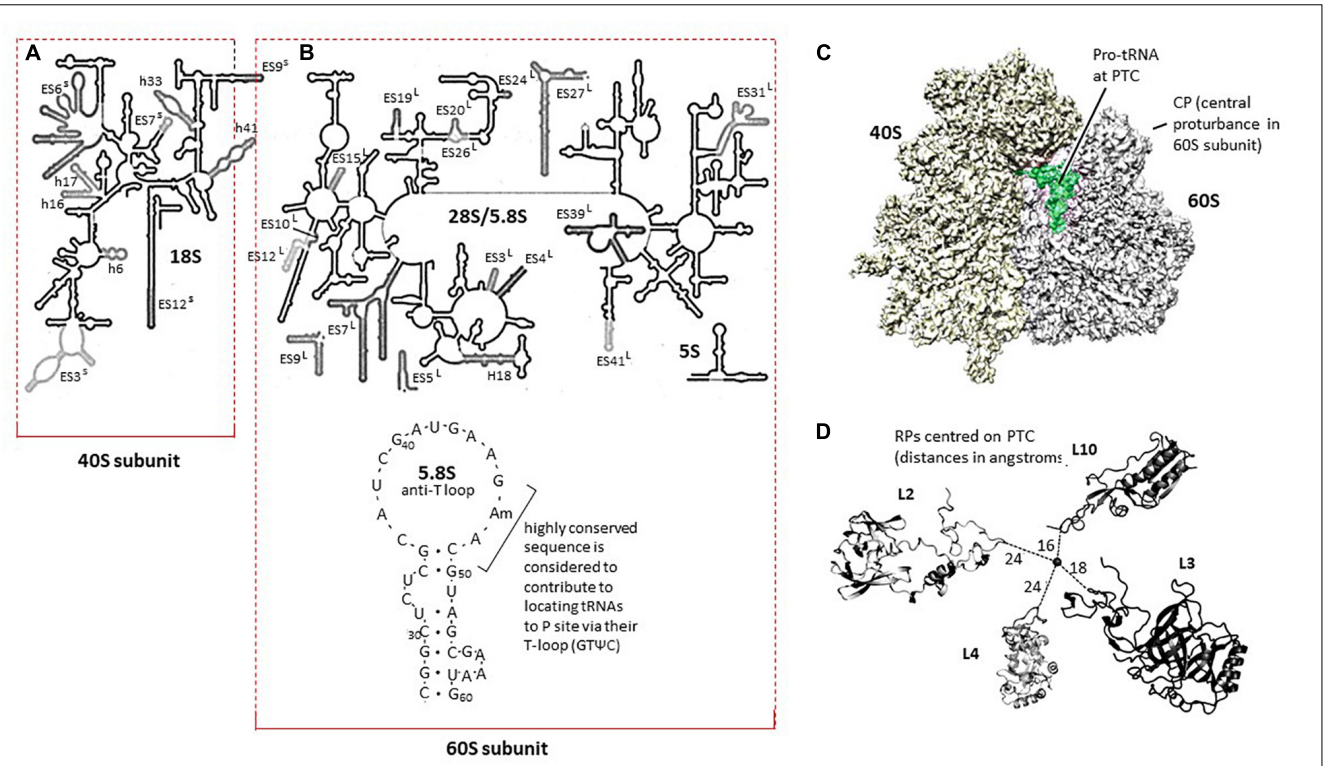


FIGURE 5 | (A) Map of the 18S rRNA in the 40S ribosome subunit modified from Armache et al.(2010; doi.org/10.1073/pnas.1009999107). The h and ES loops in the rRNA molecule are discussed in the text. **(B)** Map of the 26S rRNA in the 60S ribosome subunit modified from Armache et al.(2010, doi.org/10.1073/pnas.1009999107). The wheat 5.8S rRNA molecule shown was based on Mackay et al. (1980). **(C)**. The 3D representation of the translating wheat ribosome with the Proline-tRNA at the peptidyl-transferase center (PTC) in place Armache et al.(2010, doi.org/10.1073/pnas.1009999107). The central protuberance in the 60S subunit is a standard landmark for the 60S subunit. **(D)** Relationship between RPL2, RPL3, RPL4, and RPL10 modeled in the wheat ribosome PTC from Armache et al. (2010) **Supplementary data** in doi.org/10.1073/pnas.100999910.

Although model systems have shown that rRNA segments configured at the PTC provide ribozyme activity for catalyzing peptide bond formation, and that RPs are not strictly required for this chemical reaction (reviewed in De la Cruz et al., 2015), it is evident that, within the context of the biology of the cell, RPs are critical (reviewed in De la Cruz et al., 2015). The RPs are required for the many steps in forming and stabilizing the ribosome complex in order to ensure the efficient translation of mRNA. Utilizing the gene models documented in **Supplementary Table 1**, variation in wheat RPs can now be compiled to provide gene haplotypes (as more broadly considered by Wilhelm et al., 2013) that document variation between proteins from homologous loci on the A, B, and D genomes and SNP variation at the DNA sequence level, to indicate the potential functional markers for associating particular RP variants with phenotypic attributes. The alignments of the RPs within the groups and subgroups in **Supplementary Table 1** indicate six groups of RPs in the 40S subunit and 11 in the 60S subunit and show no variation in their amino acid sequence within the IWGSCrefseqver1 genome sequence, whereas the remainder provides the amino acid variation that can be considered as gene haplotypes. The variation complements the variation presented earlier for nucleolin (**Supplementary Figure 1**) and fibrillarin (**Figure 3**), two abundant proteins important for the infrastructure of the nucleolus and rRNA production and the TBP, Rrn3, UAF30, and Rrp5 gene models for establishing the initiation complex to start rRNA synthesis.

Specific examples of the variation between proteins from homologous loci on the A, B, and D genomes are provided below for RPS6 because it is historically significant, RPL6 as an example of an RP that was responsive to water stress in wheat (Islam et al., 2020), and RPL2, RPL3, RPL4, and RPL10 because of their particular importance in forming the PTC (**Figure 5D**).

RPS6

One of the earliest examples of phosphorylation of an RP was for RPS6 (reviewed in Biever et al., 2015) and, based on studies in model systems, it is one of the RPs interacting with rRNA transcripts during their processing in the nucleolus (Bernstein et al., 2004) to form mature rRNAs. Although the functions of RPS6 have not been clearly defined, its phosphorylation has been used as a marker for the coordinated phosphorylation and activation of RPS6 kinase, S6K1, and activation of the translation initiation factor eIF4B (Holz et al., 2005). Importantly, the activation of RPS6 kinase also reflects the activity of a central regulator of cell proliferation and growth in eukaryotic cells, target of rapamycin (TOR). In wheat, the TOR–S6K1 signaling

pathway has been shown to be a key step in GA-induced digestion of starch in the germinating wheat grain for seedling growth (Smailov et al., 2020).

The alignments shown in **Figure 6** indicate that only eight amino acid differences are found between the predicted proteins. However, at the genome level, wheat varieties analyzed in DAWN show a clear haplotype difference at the genome level that is evident in the example shown for TraesCS2B02G189500 at location 165111139 on chromosome 2B. Among the nine SNPs at the genome level in the CDS, only one caused an amino acid change, F45L, which was not represented in the differences found between the A, B, and D homologs within the reference wheat genome *per se*. None of the other Traes-RPS6 models in **Figure 6** showed SNPs in the CDS and in light of the polyploid nature of wheat, indicates that a major change, with an unknown phenotype, such as F45L in one RPS6 gene is extensively buffered by no changes in the other gene models. A similar situation is indicated for all the amino acid variations identified in **Figure 6A** where the amino acid changes for homologous genes on the long arm of 2A, 2B, and 2D (lower three entries in the alignment) for example are not found in the short arm loci. The missing sequences from TraesCS2A02G066100 and TraesCS2D02G064500 are due to gaps in the respective genome assemblies of chromosomes 2A and 2D, based on the inspection of the published reference genome of Chinese Spring. It is of interest that the phosphorylation of the serine at position 237 (S237) in Arabidopsis is closely linked to the light–dark cycle in the environment and the internal circadian rhythm of the plant (Enganti et al., 2018), and since the respective amino acid sequence section of RPS6 can be clearly identified in the wheat gene model (see insert **Figure 6A**), it adds interest to the phosphorylation of RPS6 as a marker in wheat as discussed above. The increase in S237 phosphorylation in the dark to light transition correlates with the increased loading of ribosomes onto mRNA to form polysomes in Arabidopsis (Enganti et al., 2018).

Considering the well-established nature of RPS6 as a marker for the status of cell proliferation and growth of the plant, in many model systems, the clear haplotype at the genome level differentiating two successful varieties, Mace and Lancer (**Figure 6B**) suggest that the genome level variation provides a useful fingerprint for associating genome changes with phenotypic variation. The details of the SNP scoring in **Figure 6B** using DAWN (Watson-Haigh et al., 2018) relative to the reference wheat genome assembly are provided in the legend of the figure.

RPL6

In an extensive iTRAQ proteomic analysis of water stress administered at the reproductive stage in wheat, Islam et al. (2020) found that ribosomal protein L6 (RPL6) was one of the few RPs that responded to the stress and showed a 4.2-fold increase (p-value of 0.008). The RPL6 identified in the proteome study was TraesCS6D02G190100, which is one of a pair of RPL6 proteins on chromosome 6D (**Figure 7**); pairs of RPL6 proteins also exist at homologous loci on chromosomes 6A and 6B (**Figure 7**). The response to water stress was considered rapid because the expressional change was detected before head emergence, immediately after the drought stress was imposed. RPL6s are among key proteins for controlling and enhancing protein synthesis and have been studied in several plant species in response to different environmental stresses including high and low temperatures, salinity, and water deficit. In rice, the overexpression of RPL6 resulted in tolerance to moderate (150 mM) to high (200 mM) levels of NaCl (Sahi et al., 2006; Moin et al., 2020). In addition, 50S RPL6 was upregulated after 48 h of drought stress in maize (Pei et al., 2019). Salt stress can result in modification of protein synthesis, and it has been observed that in RPL6 transgenic rice plants the upregulation of genes encoding RPs in plants under stressed conditions can lead to efficient reconstruction of protein-synthesizing machinery in cells under stress without compromising the growth and development. The RPL6 protein family members are also highly upregulated in heat-primed wheat plants compared with the non-heat-primed plants (Wang et al., 2016).

Although RPL6 is highly conserved and identification through homology to model systems was unambiguous, we note that 24 variable positions in the amino acid sequence are evident within the hexaploid wheat entries (**Figure 7**). These variable positions provided sufficient resolution between the wheat gene models to allow the pairs of genes on chromosome 6D to be distinguished (different haplotypes at the amino acid level) and specifically assign TraesCS6D02G190100 to be the gene that was upregulated at the protein level as a result of water stress early in the head development (Islam et al., 2020). The example shows that one mechanism for selecting RP sequence variants to make up the pool of ribosomes in wheat includes the possibility of quantitative changes in the level of expression.

Inspection of wheat varieties available in the DAWN genome viewer for SNPs in the RPL6 CDS detected a G to T change leading to an altered amino acid (V45K) relative to the reference sequence. The arrow in **Figure 7** indicates that at this position a V45I variation is detected from including a non-reference wheat sequence (from the UniProt database) in the alignment. The biological consequence of this polymorphism remains to be determined.

RPL2

Relatively few studies are available for the biological effects of modifying RPL2 levels of expression or mutational changes since most studies focus on the plastid RPL2 protein. For the 60S subunit RPL2, Ludwig and Tenhaken (2001) found that the transient loss of RPL2 expression in soybean challenged with a fungal infection correlated with a loss of protein synthesis activity but improved tolerance to the infection. The authors argued that the reduced protein synthesis activity in infected cells provided a short-term response to reduce the capacity of the infecting fungus to utilize host cell resources and allowed the host cell to produce its own defense molecules. Interestingly, the RPL2 gene models in wheat are one of the few examples of absolute conservation of amino acid sequence (**Supplementary Figure 5**); hence, there exists no

FIGURE 6 | Alignments for wheat RPS6 gene models at the amino acid level. **(A)** The first six entries in the alignments are on the short arms of chromosomes 2A, 2B, and 2D, and the lower three entries are on the respective long arms. CLUSTAL omega (Sievers et al., 2011; https://www.ebi.ac.uk/Tools/services/web/) was used to carry out the alignments using standard parameters. The Traes IDs are for IWGSCrefseqver1 and a hyphen separates the coordinate position of the gene model; the gaps in the entries TraesCS2A02G066100 and TraesCS2DG0264500 are the result of gaps in the IWGSCrefseq-ver1 assembly. The insert is the alignment of a small section of amino acid sequence from TraesCS2A02G066200 and an Arabidopsis RPS6 (O48549) referred to in the text in relation to the Serine237. This section is also underlined in the main sequence and is shown in location relative to the highly conserved RPS6 domain (black box). The single nucleotide polymorphisms (SNPs) indicated in panel **(B)** (below) at the genome level are indicated with red dots where they occur in the CDS and in only one case the SNP changed the code for the amino acid (F45L, indicated in red). **(B)** Comparison of the SNP profiles from two representative wheat varieties showing strikingly different haplotypes in the region of the wheat chromosome 2B locus for RPS6 as defined in Pfam (Orengo et al., 2020) at positions 1–128. The gray areas indicate the variable genome coverage of the available sequence data and the colored "drops" identify positions in the sequence that are uniformly changed from that of the reference Chinese Spring genome sequence (orange = change to G; red = change to T; green = change to A; blue = change to C). The SNP analysis was possible using DAWN (Watson-Haigh et al., 2018; http://crobiad.agwine.adelaide.edu.au/dawn/jbrowse/). * means identical amino acid in that position across the genes.

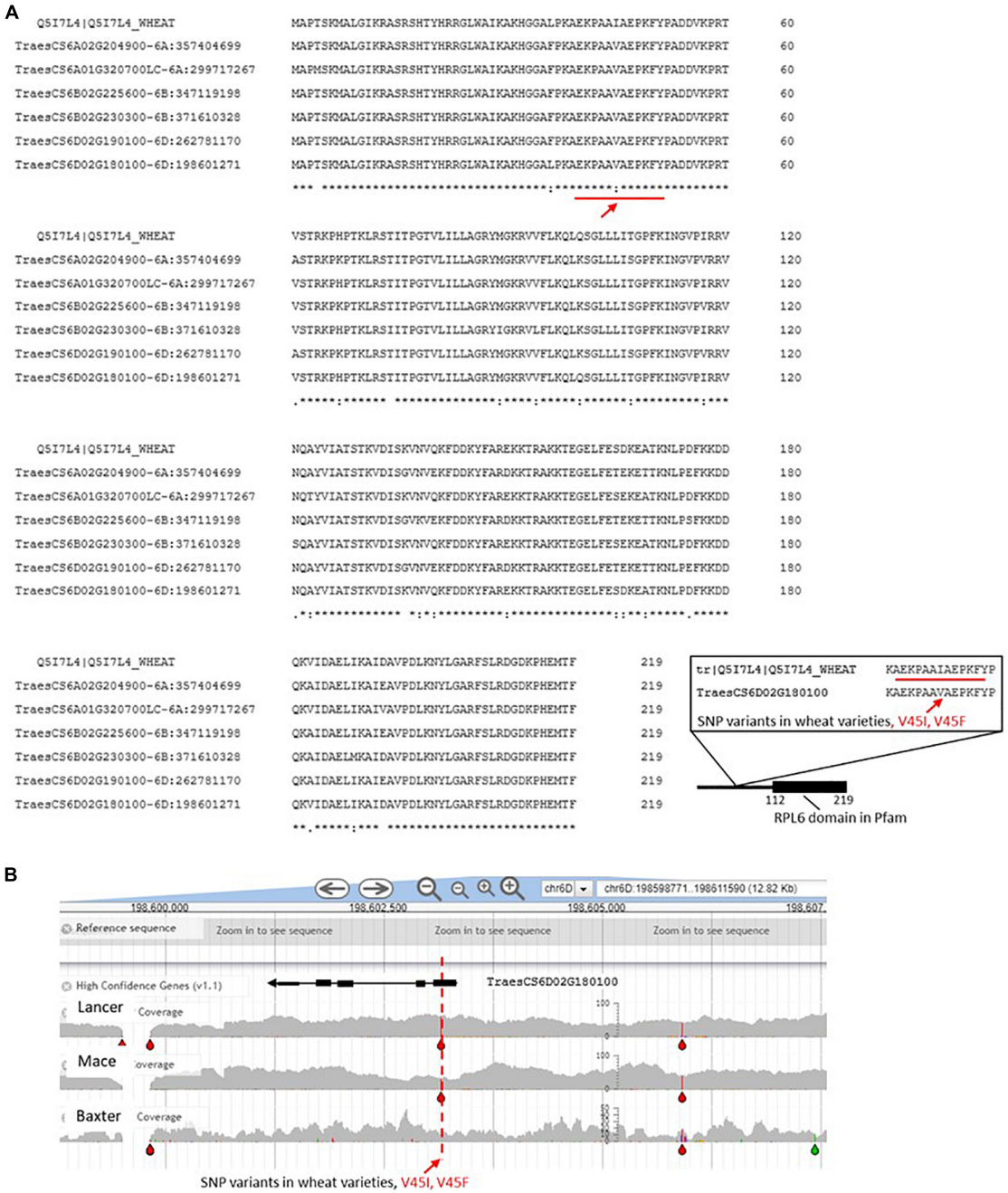


FIGURE 7 | (A) RPL6 wheat gene models. See legend **Figure 3** for details of the identification process. The red line refers to a section of the amino acid sequence that is also represented in the inset to emphasize that the V45L and V45F changes have been captured in the wheat varieties examined in DAWN [see panel **(B)**]. **(B)** Screen for single nucleotide polymorphism (SNPs) in the RPL6 region of chromosome 6D. The gray areas indicate the variable genome coverage of the available sequence data and the colored "drops" identify positions in the sequence that are uniformly changed from that of the reference Chinese Spring genome sequence (orange = change to G; red = change to T; green = change to A; blue = change to C). The SNP analysis was possible using DAWN (Watson-Haigh et al., 2018; http://crobiad.agwine.adelaide.edu.au/dawn/jbrowse/). The red dashed lines emphasizes the V45L and V45F changes in the amino acid sequence shown in **Figure 6A** within an otherwise conserved CDS.

option to recruit suitable, preexisting sequence variants of RPL2 from the genes in wheat. This leaves only a quantitative change in the expression as a mechanism for a change in ribosome translation attributes driven by RPL2. At the genomic level, no SNPs are found within the CDS when wheat varieties were inspected.

FIGURE 8 | RPL3 gene models in wheat. **(A)** See legend **Figure 3** for details of the identification process. Locations of differences in amino acid sequence within a single genome are indicated by the absence of a *. The red dots indicate the locations of mutations associated with medical conditions in humans and are thus predicted to have a phenotype in wheat if they were found to occur. The two red arrows highlight mutations at positions W258R and H259Y in the tomato RPL3 that provided an improved tolerance to *F. graminearum* in transgenic tobacco. **(B)** Single nucleotide polymorphism (SNP) variation in the genome region around the RPL3 gene on chromosome 5B, even though variation in the CDS has not been captured in the set of wheat varieties available in the DAWN viewer used to generate the image. The gray areas indicate the variable genome coverage of the available sequence data and the colored "drops" identify positions in the sequence that are uniformly changed from that of the reference Chinese Spring genome sequence (orange = change to G; red = change to T; green = change to A; blue = change to C). The SNP analysis was possible using DAWN (Watson-Haigh et al., 2018; http://crobiad.agwine.adelaide.edu.au/dawn/jbrowse/).

RPL3

In contrast to RPL2, the RPL3 gene models do show some sequence variation (**Figure 8**) although a screen for SNP variation in the DAWN genome viewer for wheat varieties indicated that no variation in the CDS has been captured in the set of varieties examined (**Supplementary Figure 6**). The RPL3 is located in

FIGURE 9 | (A) RPL10 wheat gene models. See legend **Figure 3** for details of the identification process. The red dots indicate the locations of the amino acid changes found in human, wheat, rice tobacco, and yeast shown in the respective boxes to the left of the alignment, discussed in the text. In some positions, breaks in the *s indicating differences within the homologous wheat genes align correspond to red dots (for example Q123L) indicating the change is also found in humans and other eukaryotes as defined in the boxes to the left of the alignments. The red line indicates the P-site loop region that is important in the catalytic site (PTC, see **Figure 5**). **(B)** Single nucleotide polymorphism (SNP) variation in the genome region around the RPL10 gene on chromosome 1B, even though variation in the CDS has not been captured in the wheat varieties examined (exemplar shown). The gray areas indicate the variable genome coverage of the available sequence data and the colored "drops" identify positions in the sequence that are uniformly changed from that of the reference Chinese Spring genome sequence (orange = change to G; red = change to T; green = change to A; blue = change to C). The SNP analysis was possible using DAWN (Watson-Haigh et al., 2018; http://crobiad.agwine.adelaide.edu.au/dawn/jbrowse/).

the PTC of wheat ribosomes (see **Figure 5D**) and reduced levels of RPL3 in transgenic *Nicotiana* (Popescu and Tumer, 2004) correlated with reductions in cell number, stunting, and inhibition of lateral root growth. Precursor rRNA levels (32S, see **Figure 1C**) were elevated in the transgenic plants consistent with an early engagement of RPL3 as part of the generation of rRNA. Mutations identified in human RPL3 (G27N, A75V, R161W, T189M, D308N(V), and R343W) are associated with medical conditions (Thorolfsdottir et al., 2017; Ganapathi et al., 2020) but have not been identified as variable positions in wheat RPL3 to date. The findings in human RPL3 do serve to indicate that variation in this gene can be explored in terms of fine-tuning the translation process to particular phenotypic requirements such as tolerance to stress conditions. The mutations at positions W258R and H259Y in the tomato RPL3 provided an improved tolerance to *Fusarium graminearum* in transgenic tobacco (Safipoor-Afshar et al., 2007) due to a reduced sensitivity to the trichothecen mycotoxin DON in the host plants, and even though these are not variable positions in the survey of wheat genome SNPs, they provide interesting targets. Di and Tumer (2005) reported that

a transgenic construct carrying a truncated RPL3 missing the C-terminal region also conferred increased tolerance to DON in transgenic tobacco plants.

RPL4

The wheat RPL4 gene models show a low level of sequence divergence, which is spread across the proteins domains that define the gene model (**Supplementary Figure 7**). In Arabidopsis, the mutation G73R (numbering follows **Supplementary Figure 7**) modifies the phenotype extensively to generate a plant with narrow leaves, abnormal numbers of cotyledons, short roots, and short hypocotyls. Although the G73R mutation has undesirable phenotypic consequences, it is able to suppress deleterious mutations (Horiguchi et al., 2011; Kakehi et al., 2015), which may be an advantage in certain situations of a genetic modification pipeline to suppress deleterious mutations transfers to wheat.

RPL10

The location of RPL10 is also near the PTC (**Figure 5D**) in microorganism, plants, and animals and extends from the CP (**Figure 5C**) to the PTC/GTP-ase center *via* the P-site loop region of RPL10. The P-site loop is in the middle of RPL10 at positions 102–112 and is a conserved amino acid sequence. The P-site is argued to be required for the conformational changes within the ribosome that are associated with the elongation cycle of the translation process for synthesizing peptide chains (reviewed in Pollutri and Penzo, 2020) in model systems. In the P-site loop region of the wheat RPL10, no differences exist among the gene models (**Figure 9**). Consistent with its fundamental importance in biology, variation is limited to 41 positions (including the very C-terminal end) across humans to plants in the 224 amino acid sequence; among 33 induced mutations in yeast that could be located in the wheat sequence, 17 were lethal, whereas others such as the change in E (glutamic acid) at position 180 significantly reduced yeast growth. In a mutation study in Arabidopsis (Falcone-Ferreyra et al., 2013), the altering of three copies of the RPL10 gene resident in the genome indicated that compensation between copies occurred and expression differed between tissues. These Arabidopsis studies and studies in human (Klauck et al., 2006; reviewed in Pollutri and Penzo, 2020) and *Nicotiana benthamiana* (Ramu et al., 2020) are consistent with the more general model that variation in essential RPs, such as RPL10, can moderate the translation activity of ribosomes to preferentially accept certain mRNAs and/or undertake non-ribosomal level regulation of transcription and signal transduction.

DISCUSSION

The cytoplasmic ribosomes constitute the central RNA–protein complex responsible for synthesizing new proteins in wheat cells, and studies in model organisms have developed the concept that diversity in the composition of ribosomes can be linked to phenotypic diversity (reviewed in Pollutri and Penzo, 2020). In plants, the observations relating to clusters of rRNA unit types that can be co-regulated in a tissue-specific manner support the concept of tissue-specific ribosome subpopulations differing in their functional attributes and contributing to responses to environmental challenges (Tulpová et al., 2020; Sims et al., 2021). The ribosomes from wheat-germ extracts were among the early eukaryotic cell-free systems for translating mRNA, and the analyses summarized in the present study provide a foundation for characterizing populations of ribosomes with a range of functional attributes. Specifically, the model argues that certain ribosomes can become predominant in situations of stress for preferentially translating mRNAs to generate proteins better suited to contributing to the survival of the cell in situations of, for example, water stress (Martinez-Seidel et al., 2020). This level of structural variation in the population of ribosomes would interact with the well-established variation in the translation machinery of the cell in response to stresses with respect to the engagement of initiation factors and elongation factors involving a translational regulator TOR in wheat (Smailov et al., 2020), and more broadly across plants and animals. The outputs from integrating the networks that determine the translation of populations of mRNA in a cell can be visualized experimentally using ribosome profiling which is based on sequencing of ribosome protected mRNA fragments, combined with total RNA-seq data, to provide an estimate of the efficiency of the utilization of particular mRNAs (Lei et al., 2015). Maize seedlings under water stress (Lei et al., 2015) provided evidence for changes in the sequence profile of translated mRNA and changes in transcription *per se* that correlated with a water stress response.

In wheat, an additional variable relates to the discovery of ribosome-inactivating proteins (tritins) that have different specificities and cofactor requirements depending on whether they are from seed or other tissues (Massiah and Hartley, 1995). The tritins in wheat comprise a family of 17 genes and inspection of IWGSC ver2.1 (Zhu et al., 2021), using the UniProtKB—Q07810 sequence, indicated six gene models expressed predominantly in grain tissue and at a lower level in root tissue in a cluster on chromosome 5B. An additional eight gene models on 5B expressed only in root tissue but at a lower level. Three tritin gene models were identified on chromosome 5A expressed at a relatively level in grain and root tissue, and no significant hits could be identified on chromosome 5D. Two gene models were closely linked and moderately expressed in grain and root and on a genome segment not assigned to a chromosome. The rRNA N-glycosidase activity associated with tritins can preferentially depurinate highly conserved 26S rRNA SRL sequences (sarcin–ricin loop, AGUACGAGAGGA) required for elongation factor engagement at the PTC, in the ribosomes from invading pathogens (Fernandez-Puentes and Vazquez, 1977). This activity can therefore provide a novel approach for developing disease resistance in wheat. Similarly, the RPL3 protein is also located at the PTC and mutations at positions W258R and H259Y, or deletions of the C-terminal region, in the RPL3 gene provided an improved tolerance to *F. graminearum* in transgenic tobacco (Di and Tumer, 2005; Safipoor-Afshar et al., 2007). The improved resistance was due to a reduced sensitivity to DON produced by *F. graminearum* and thus provides a target for breeding tolerance to this challenging

pathogen in wheat. This approach could complement the success in detoxifying trichothecene mycotoxins such as DON using the glutathione-*S*-transferase gene encoded by the Fhb7 locus (Wang et al., 2020).

Mutations and variation, in model systems, in RPL2, RPL4, and RPL10 proteins that are also located at the PTC support the concept of the rRNA–protein complex as possible targets for modifying the phenotype of the wheat plant. In the case of RPL10, some of the variable positions between the wheat homologous loci corresponded to the position of mutations in the human RPL10 that associated with disease phenotypes (ribosomopathies; Pollutri and Penzo, 2020). The example of RPL6 discussed in this manuscript also brings into play the ribosome independent functions of many RPs where it has been argued in model systems that stress can initiate the translocation of RPL6 from the nucleolus to the nucleoplasm where it can interact with chromatin histone H2A and alter cell biology (Yang et al., 2019).

Inspection of variation among the wheat RP gene models in **Supplementary Table 1** generates a picture of single amino acid changes (gene haplotypes) that provide wheat with a flexible pool of RPs and hence pools of ribosomes with unique functional attributes for changing the balance of mRNAs translated in particular environmental conditions. The variation in RPs interface with the variation documented in the rRNA that is exemplified by nucleolar dominance, to generate populations of ribosomes with unique compositions suitable in certain physiological conditions of the cell. An extensive literature exists to describe the effects of ABA on ribosome attributes (Grill and Himmelbach, 1998; Guo et al., 2011; Zhang et al., 2012; Li et al., 2021) and would be expected to reflect the engagement of the protein synthesis machinery in the higher level control networks within the cell.

AUTHOR CONTRIBUTIONS

RA conceived, analyzed/interpreted the data, and drafted the manuscript. SI and PW carried out the proteome studies and contributed to writing of the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.686586/full#supplementary-material

REFERENCES

Abraham, K. J., Khosraviani, N., Chan, J. N. Y., Gorthi, A., Samman, A., Zhao, D. Y., et al. (2020). Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature* 585, 298–302. doi: 10.1038/s41586-020-2497-0

Appels, R., and Dvorak, J. (1982). The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. *Theor. Appl. Genet.* 63, 337–348.

Armache, J.-P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Bhushan, S., et al. (2010). Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19754–19759. doi: 10.1073/pnas.1010005107

Ban, N., Beckmann, R., Cate, J. H. D., Dinman, J. D., Dragon, F., Ellis, S. R., et al. (2014). A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* 24, 165–169.

Barker, R. F., Harberd, N. P., Jarvist, M. G., and Flavell, R. B. (1988). Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. *J. Mol. Biol.* 201, 1–17. doi: 10.1016/0022-2836(88)90434-2

Baum, B. R., and Bailey, L. G. (2001). The 5S rRNA gene sequence variation in wheats and some polyploid wheat progenitors (Poaceae: Triticeae). *Genet. Resour. Crop Evol.* 48, 35–51.

Bernstein, K. A., Gallagher, J. E. G., Mitchell, B. M., Granneman, S., and Baserga, S. J. (2004). The small-subunit processome is a ribosome assembly intermediate. *Eukaryot. Cell* 3, 1619–1626. doi: 10.1128/EC.3.6.1619-1626.2004

Berry, J., Weber, S. C., Vaidya, N., Haataja, M., and Brangwynne, C. P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5237–E5245. doi: 10.1073/pnas.1509317112

Biever, A., Valjent, E., and Puighermanal, E. (2015). Ribosomal protein S6 phosphorylation in the nervous system: from regulation to function. *Front. Mol. Neurosci.* 8:75. doi: 10.3389/fnmol.2015.00075

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977

Bonner, J., and Varner, J. E. (1965). "Seed development and germination," in *Plant Biochemistry*, eds J. Bonner and J. E. Varner (New York, NY: Academic Press).

Caperta, A. D., Neves, N., Morais-Cecílio, L., Malhó, R., and Viegas, W. (2002). Genome restructuring in rye affects the expression, organization and disposition of homologous rDNA loci. *J. Cell Sci.* 115, 2839–2846.

Correll, C. C., Jiri Bartek, J., and Dundr, M. (2019). The nucleolus: a multiphase condensate balancing ribosome synthesis and translational capacity in health, aging and ribosomopathies. *Cells* 8:869. doi: 10.3390/cells8080869

De la Cruz, J., Karbstein, K., and Woolford, J. L. Jr. (2015). Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu. Rev. Biochem.* 84, 93–129. doi: 10.1146/annurev-biochem-060614-033917

Di, R., and Tumer, N. E. (2005). Expression of a truncated form of ribosomal protein L3 confers resistance to pokeweed antiviral protein and the *Fusarium mycotoxin* deoxynivalenol. *Mol. Plant Microbe Interact.* 18, 762–770. doi: 10.1094/MPMI-18-0762

Duncan, O., Trösch, J., Fenske, R., Taylor, N. L., and Millar, A. H. (2017). Resource: mapping the *Triticum a estivum* proteome. *Plant J.* 89, 601–616. doi: 10.1111/tpj.13402

Dvorak, J., Zhang, H.-B., Kota, R. S., and Lassner, M. (1989). Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. *Genome* 32, 1003–1016. doi: 10.1093/molbev/mst106

Elela, S. A., Good, L., Melekhovets, Y. F., and Nazar, R. N. (1994). Inhibition of protein synthesis by an efficiently expressed mutation in the yeast 5.8S ribosomal RNA. *Nucleic Acids Res.* 22, 686–693. doi: 10.1093/nar/22.4.686

Enganti, R., Cho, S. K., Toperzer, J. D., Urquidi-Camacho, R. A., Cakir, O. S., Ray, A. P., et al. (2018). Phosphorylation of ribosomal protein RPS6 integrates light signals and circadian clock signals. *Front. Plant Sci.* 8:2210. doi: 10.3389/fpls.2017.02210

Falcone-Ferreyra, M. L., Casadevall, R., Luciani, M. D., Pezza, A., and Casati, P. (2013). New evidence for differential roles of L10 ribosomal proteins from *Arabidopsis*. *Plant Physiol.* 163, 378–391. doi: 10.1104/pp.113.223222

Flavell, R., and O'Dell, M. (1976). Ribosomal RNA genes on homoeologous chromosomes of groups 5 and 6 in hexaploid wheat. *Heredity* 37, 377–385. doi: 10.1038/hdy.1976.102

Flavell, R. B., O'Dell, M., and Thompson, W. F. (1988). Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. *J. Mol. Biol.* 204, 523–534.

Fernandez-Puentes, C., and Vazquez, D. (1977). Effects of some proteins that inactivate the eukaryotic ribosome. *FEBS Lett.* 78, 143–146.

Fomproix, N., Gébrane-Younès, J., and Hernandez-Verdun, D. (1998). Effects of anti-fibrillarin antibodies on building of functional nucleoli at the end of mitosis. *J. Cell Sci.* 111, 359–372. doi: 10.1242/jcs.111.3.359

Ford, K., Cassin, A., and Bacic, A. (2011). Quantitative proteomic analysis of wheat cultivars with differing drought stress tolerance. *Front. Plant Sci.* 2:44. doi: 10.3389/fpls.2011.00044

Ganapathi, M., Argyriou, L., Martínez-Azorín, F., Morlot, S., Yigit, G., Lee, T. M., et al. (2020). Bi-allelic missense disease-causing variants in RPL3L associate neonatal dilated cardiomyopathy with muscle-specific ribosome biogenesis. *Hum. Genet.* 139, 1443–1454. doi: 10.1007/s00439-020-02188-6

Gerlach, W. L., and Bedbrook, J. R. (1979). Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl. Acids Res.* 7, 1869–1885.

Grill, E., and Himmelbach, A. (1998). ABA signal transduction. *Curr. Opin. Plant Biol.* 1, 412–418. doi: 10.1016/s1369-5266(98)80265-3

Guo, J., Wang, S., Valerius, O., Hall, H., Zeng, Q., Li, J.-F., et al. (2011). Involvement of *Arabidopsis* RACK1 in protein translation and its regulation by Abscisic Acid. *Plant Physiol.* 155, 370–383. doi: 10.1104/pp.110.160663

Handa, H., Kanamori, H., Tanaka, T., Murata, K., Kobayashi, F., Robinson, S. J., et al. (2018). Structural features of two major nucleolar organizer regions (NORs), Nor-B1 and Nor-B2, and chromosome-specific rRNA gene expression in wheat. *Plant J.* 96, 1148–1159. doi: 10.1111/tpj.14094

Hilliker, A. J., and Appels, R. (1982). Pleiotropic effects associated with the deletion of heterochromatin surrounding rDNA on the X chromosome of *Drosophila*. *Chromosoma (Berl.)* 86, 469–490. doi: 10.1007/BF00330122

Holz, M. K., Ballif, B. A., Gygi, S. P., and Blenis, J. (2005). mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* 123, 569–580. doi: 10.1016/j.cell.2005.10.024

Horiguchi, G., Molla-Morales, A., Perez-Perez, J. M., Kojima, K., Robles, P., Ponce, M. R., et al. (2011). Differential contributions of ribosomal protein genes to *Arabidopsis thaliana* leaf development. *Plant J.* 65, 724–736. doi: 10.1111/j.1365-313X.2010.04457.x

Iida, T., and Kobayashi, T. (2019). RNA polymerase I activators count and adjust ribosomal RNA gene copy number. *Mol. Cell* 73, 645–654. doi: 10.1016/j.molcel.2018.11.029

Islam, S., Wang, P., Vincent, D., Khan, J. M., Juhasz, A., Diepeveen, D., et al. (2020). Proteomic profiling of developing wheat heads under water-stress. *Funct. Integr. Genom.* 20, 695–710. doi: 10.1007/s10142-020-00746-9

IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Kakehi, J.-I., Kawano, E., Yoshimoto, K., Cai, Q., Imai, A., and Takahashi, T. (2015). Mutations in ribosomal proteins, RPL4 and RACK1, suppress the phenotype of a thermospermine-deficient mutant of *Arabidopsis thaliana*. *PLoS One* 10:e0117309. doi: 10.1371/journal.pone.0117309

Kang, G.-Z., Peng, H.-F., Han, Q.-X., Wang, Y.-H., and Guo, T.-C. (2011). Identification and expression pattern of ribosomal L5 gene in common wheat (*Triticum aestivum* L.). *Gene* 493, 62–68. doi: 10.1016/j.gene.2011.11.023

Kelley, L., Mezulis, S., Yates, C., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Khoshnevis, S., Liu, X., Dattolo, M. D., and Karbstein, K. (2019). Rrp5 establishes a checkpoint for 60S assemblyduring 40S maturation. *RNA* 25, 1164–1176. doi: 10.1261/rna.071225.119

Kim, J.-H., Dilthey, A. T., Nagaraja, R., Lee, H.-S., Koren, S., Dudekula, D., et al. (2018). Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. *Nucleic Acids Res.* 46, 6712–6725. doi: 10.1093/nar/gky442

Klauck, S. M., Felder, B., Kolb-Kokocinski, A., Schuster, C., Chiocchetti, A., Schupp, I., et al. (2006). Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. *Mol. Psychiatry* 11, 1073–1084. doi: 10.1038/sj.mp.4001883

Lagudah, E. S., Appels, R., and McNeil, D. (1991). The Nor-D3 locus *Triticum tauschii*: natural variation and genetic linkage to markers in chromosome 5. *Genome* 34, 387–395. doi: 10.1139/g91-060

Lassner, M., Anderson, O., and Dvorak, J. (1987). Hypervariation associated with a 12-nucleotide direct repeat and inferences on intergenomic homogenization of ribosomal RNA gene spacers based on the DNA sequence of a clone from the wheat Nor-D3 locus. *Genome* 29, 770–781. doi: 10.1139/g87-130

Lee, B. M., Xu, J., Clarkson, B. K., Martinez-Yamout, M. A., Dyson, H. J., Case, D. A., et al. (2006). Induced fit and "lock and key" recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J. Mol. Biol.* 357, 275–291. doi: 10.1016/j.jmb.2005.12.010

Lei, L., Shi, J., Chen, J., Zhang, M., Sun, S., Xie, S., et al. (2015). Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* 84, 1206–1218. doi: 10.1111/tpj.13073

Li, L., Zhu, T., Yun Song, Y., Li Feng, L., Farag, E. A. H., and Ren, M. (2021). ABSCISIC ACID INSENSITIVE-5 interacts with ribosomal S6 kinase2 to mediate ABA responses during seedling growth in *Arabidopsis*. *Front. Plant Sci.* 11:598654. doi: 10.3389/fpls.2020.598654

Loza-Muller, L., Rodríguez-Corona, U., Sobol, M., Rodríguez-Zapata, L. C., Hozak, P., and Castano, E. (2015). Fibrillarin methylates H2A in RNA polymerase I trans-active promoters in Brassica oleracea. *Front. Plant Sci.* 6:976. doi: 10.3389/fpls.2015.00976

Ludwig, A., and Tenhaken, R. (2001). Suppression of the ribosomal L2 gene reveals a novel mechanism for stress adaptation in soybean. *Planta* 212, 792–798.

Mackay, R. M., Spencer, D. F., Doolittle, W. F., and Gray, M. W. (1980). Nucleotide sequences of wheat-embryo cytosol 5-S and 5.8-S ribosomal ribonucleic acids. *Eur. J. Biochem.* 112, 561–576. doi: 10.1111/j.1432-1033.1980.tb06122.x

Martinez-Seidel, F., Beine-Golovchuk, O., Hsieh, Y.-C., and Kopka, J. (2020). Systematic review of plant ribosome heterogeneity and specialization. *Front. Plant Sci.* 11:948. doi: 10.3389/fpls.2020.00948

Massiah, A. J., and Hartley, M. R. (1995). Wheat ribosome-inactivating proteins: seed and leaf forms with different specificities and cofactor requirements. *Planta* 197, 633–640. doi: 10.1007/BF00191571

Moin, M., Saha, A., Bakshi, A., Madhav, M. S., and Kirti, P. B. (2020). Ribosomal protein large subunit RPL6 modulates salt tolerance in rice. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.31.126102

Nasirudin, K. M., Ehtesham, N. Z., Tuteja, R., Sopory, S. K., and Tuteja, N. (2004). The Gly-Arg-rich C-terminal domain of pea nucleolin is a DNA helicase that catalytically translocates in the 5′ to 3′ direction. *Arch. Biochem. Biophys.* 434, 306–315. doi: 10.1016/j.abb.2004.11.016

Nishikawa, K., and Takemura, S. (1974). Nucleotide sequence of 5SRNA from *Torulopsis utilis*. *FEBS Lett.* 40, 106–109. doi: 10.1016/0014-5793(74)80904-x

Orengo, C., Velankar, S., Wodak, S., Zoete, V., Bonvin, A. M. J. J., Elofsson, A., et al. (2020). A community proposal to integrate structural bioinformatics activities in ELIXIR (3D-Bioinfo Community). *F1000Research* 9:278. doi: 10.12688/f1000research.20559.1

Pei, Y., Bai, J., Guo, X., Zhao, M., Ma, Q., and Song, X. (2019). Comparative proteome analysis of drought-sensitive and drought-tolerant maize leaves under osmotic stress. *Can. J. Plant Sci.* 99, 467–479. doi: 10.1139/cjps-2018-0115

Pisl, M., and Engel, C. (2020). Structural basis of RNA polymerase I pre-initiation complex formation and promoter melting. *Nat. Commun.* 11:1206. doi: 10.1038/s41467-020-15052-y

Pollutri, D., and Penzo, M. (2020). Ribosomal protein L10: from function to dysfunction. *Cells* 2020:2503. doi: 10.3390/cells9112503

Pontvianne, F., Matia, I., Douet, J., Tourmente, S., Medina, F. J., Echeverria, M., et al. (2007). Characterization of AtNUC-L1 reveals a central role of nucleolin in nucleolus organization and silencing of AtNUC-L2 gene in *Arabidopsis*. *Mol. Biol. Cell* 18, 369–379. doi: 10.1091/mbc.e06-08-0751

Popescu, S. C., and Tumer, N. E. (2004). Silencing of ribosomal protein L3 genes in *N. tabacum* reveals coordinate expression and significant alterations in plant growth, development and ribosome biogenesis. *Plant J.* 39, 29–44. doi: 10.1111/j.1365-313X.2004.02109.x

Ramu, V. S., Dawane, A., Lee, S., Oh, S., Lee, H.-K., Sun, L., et al. (2020). Ribosomal protein QM/RPL10 positively regulates defence and protein translation mechanisms during nonhost disease resistance. *Mol. Plant Pathol.* 21, 1481–1494. doi: 10.1111/mpp.12991

Ream, T. S., Haag, J. R., Pontvianne, F., Nicora, C. D., Norbeck, A. D., Pasa-Tolic, L., et al. (2015). Subunit compositions of *Arabidopsis* RNA polymerases

I and III reveal Pol I- and Pol III-specific forms of the AC40 subunit and alternative forms of the C53 subunit. *Nucleic Acids Res.* 43, 4163–4178. doi: 10.1093/nar/gkv247

Reddy, P. R., and Appels, R. (1989). A second locus for the 5S rRNA multigene family in *Secale* L.: sequence divergence in two lineages of the family. *Genome* 32, 456–467. doi: 10.1139/g89-469

Rodriguez-Corona, U., Sobol, M., Rodriguez-Zapata, L. C., Hozak, P., and Castano, E. (2015). Fibrillarin from *Archaea* to human Biol. *Cell* 107, 159–174. doi: 10.1111/boc.201400077

Safipoor-Afshar, A., Mousavi, A., Renu, A. M., and Adam, G. (2007). Double mutation in tomato ribosomal protein L3 cDNA confers tolerance to deoxynivalenol (DON) in transgenic Tobacco. *Pak. J. Biol. Sci.* 10, 2327–2333. doi: 10.3923/pjbs.2007.2327.2333

Sahi, C., Singh, A., Kumar, K., Blumwald, E., and Grover, A. (2006). Salt stress response in rice: genetics, molecular biology, and comparative genomics. *Funct. Integr. Genom.* 6, 263–284. doi: 10.1007/s10142-006-0032-5

Sergeeva, E. M., Shcherban, A. B., Adonina, I. G., Nesterov, M. A., Beletsky, A. V., Rakitin, A. L., et al. (2017). Fine organization of genomic regions tagged to the 5S rDNA locus of the bread wheat 5B chromosome. *BMC Plant Biol.* 17(Suppl. 1):183. doi: 10.1186/s12870-017-1120-5

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Silva, M., Pereira, S., Bento, M., Santos, A. P., Shaw, P., Delgado, M., et al. (2008). Interplay of ribosomal DNA loci in nucleolar dominance: dominant NORs are up-regulated by chromatin dynamics in the wheat-rye system. *PLoS One* 3:e3824. doi: 10.1371/journal.pone.0003824

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the *Arabidopsis* NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12:387. doi: 10.1038/s41467-020-20728-6

Sirri, V., Roussel, P., and Hernandez-Verdun, D. (2000). The AgNOR proteins: qualitative and quantitative changes during the cell cycle. *Micron* 31, 121–126. doi: 10.1016/s0968-4328(99)00068-2

Smailov, B., Alybayev, S., Smekenov, I., Mursalimov, A., Saparbaev, M., Sarbassov, D., et al. (2020). Wheat germination is dependent on plant target of Rapamycin signaling. *Front. Cell Dev. Biol.* 8:60668. doi: 10.3389/fcell.2020.606685

Swift, J. G., and O'Brien, T. P. O. (1972). The fine structure of the wheat scutellum before germination. *Aust. J. Biol. Sci.* 25, 9–22. doi: 10.1071/bi9720009

Tessarz, P., Santos-Rosa, H., Robson, S. C., Sylvestersen, K. B., Nelson, C. J., Nielsen, M. L., et al. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* 505, 564–567. doi: 10.1038/nature12819

Thiry, M., and Lafontaine, D. L. J. (2005). Birth of a nucleolus: the evolution of nucleolar compartments. *Trends Cell Biol.* 15, 194–199. doi: 10.1016/j.tcb.2005.02.007

Thompson, W. F., and Flavell, R. B. (1988). DNase I sensitivity of ribosomal RNA genes in chromatin and nucleolar dominance in wheat. *J. Mol Biol.* 204, 535–548. doi: 10.1016/0022-2836(88)90353-1

Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Jonsson, S., Halldorsson, G. H., Melsted, P., et al. (2017). Mutations in RPL3 and MYZAP increase risk of atrial fibrillation. *bioRxiv* [Preprint]. doi: 10.1101/223578

Tong, C. G., Reichler, S., Sonal Blumenthal, S., Balk, J., Hsieh, H.-L., and Roux, S. J. (1997). Light regulation of the abundance of mRNA encoding a nucleolin-like protein localized in the nucleoli of pea nuclei. *Plant Physiol.* 114, 643–652. doi: 10.1104/pp.114.2.643

Tulpová, Z., Aleš Kovařík, A., Toegelová, H., Kapustová, V., Navrátilová, P., Hřibová, E., et al. (2020). Anatomy and transcription dynamics of wheat ribosomal RNA loci revealed by optical mapping and RNA sequencing. *bioRxiv* [Preprint]. doi: 10.1101/2020.08.29.273623

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283. doi: 10.1038/s41586-020-2961-x

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., et al. (2020). Horizontal gene transfer of Fhb7 from fungus underlies *Fusarium* head blight resistance in wheat. *Science* 368:eaba5435. doi: 10.1126/science.aba5435

Wang, J., Ji, C., Li, Q., Zhou, Y., and Wu, Y. (2018). Genome-wide analysis of the plant-specific PLATZ proteins in maize and identification of their general role in interaction with RNA polymerase III complex. *BMC Plant Biol.* 18:221. doi: 10.1186/s12870-018-1443-x

Wang, X., Xin, C., Cai, J., Zhou, Q., Dai, T., Cao, W., et al. (2016). Heat Priming Induces Trans-generational tolerance to high temperature stress in wheat. *Front. Plant Sci.* 7:501. doi: 10.3389/fpls.2016.00501

Watson-Haigh, N. S., Suchecki, R., Kalashyan, E., Garcia, M., and Baumann, U. (2018). DAWN: a resource for yielding insights into the diversity among wheat genomes. *BMC Genomics* 19:941. doi: 10.1186/s12864-018-5228-2

Wilhelm, E. P., Mackay, I. J. M., Saville, R. J., Korolev, A. V., Balfourier, F., Greenland, A. J., et al. (2013). Haplotype dictionary for the Rht-1 loci in wheat. *Theor. Appl. Genet.* 126, 1733–1747. doi: 10.1007/s00122-013-2088-2097

Yang, C., Zhang, W., Ji, Y., Li, T., Yang, Y., and Zheng, X. X. (2019). Ribosomal protein L6 (RPL6) is recruited to DNA damage sites in a poly(ADP-ribose) polymerase–dependent manner and regulates the DNA damage response. *J. Biol. Chem.* 294, 2827–2838. doi: 10.1074/jbc.RA118.007009

Zhang, L., Hu, Y., Yan, S., Li, H., He, S., Huang, M., et al. (2012). ABA-mediated inhibition of seed germination is associated with ribosomal DNA chromatin condensation, decreased transcription, and ribosomal RNA gene hypoacetylation. *Plant Mol. Biol.* 79, 285–293. doi: 10.1007/s11103-012-9912-3

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant J.* 107, 303–314. doi: 10.1111/tpj.15289

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

REVIEW
published: 24 February 2022
doi: 10.3389/fpls.2022.788911

The Nuclear 35S rDNA World in Plant Systematics and Evolution: A Primer of Cautions and Common Misconceptions in Cytogenetic Studies

*Josep A. Rosselló, Alexis J. Maravilla and Marcela Rosato**

Jardín Botánico, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universitat de València, Valencia, Spain

The ubiquitous presence of rRNA genes in nuclear, plastid, and mitochondrial genomes has provided an opportunity to use genomic markers to infer patterns of molecular and organismic evolution as well as to assess systematic issues throughout the tree of life. The number, size, location, and activity of the 35S rDNA cistrons in plant karyotypes have been used as conventional cytogenetic landmarks. Their scrutiny has been useful to infer patterns of chromosomal evolution and the data have been used as a proxy for assessing species discrimination, population differentiation and evolutionary relationships. The correct interpretation of rDNA markers in plant taxonomy and evolution is not free of drawbacks given the complexities derived from the lability of the genetic architecture, the diverse patterns of molecular change, and the fate and evolutionary dynamics of the rDNA units in hybrids and polyploid species. In addition, the terminology used by independent authors is somewhat vague, which often complicates comparisons. To date, no efforts have been reported addressing the potential problems and limitations involved in generating, utilizing, and interpreting the data from the 35S rDNA in cytogenetics. This review discusses the main technical and conceptual limitations of these rDNA markers obtained by cytological and karyological experimental work, in order to clarify biological and evolutionary inferences postulated in a systematic and phylogenetic context. Also, we provide clarification for some ambiguity and misconceptions in terminology usually found in published work that may help to improve the usage of the 35S ribosomal world in plant evolution.

Keywords: 35S rDNA, secondary constriction, satellite chromosome, NOR, nucleolus, amphiplasty, cytogenetic markers, rRNA genes

OPEN ACCESS

Edited by:
Sonia Garcia,
Botanical Institute of Barcelona (CSIC), Spain

Reviewed by:
Teresa Garnatje,
Spanish National Research Council (CSIC), Spain
Martina Dvorackova,
Central European Institute of Technology (CEITEC), Czechia

***Correspondence:**
Marcela Rosato
marcela.rosato@uv.es

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *03 October 2021*
Accepted: *27 January 2022*
Published: *24 February 2022*

Citation:
Rosselló JA, Maravilla AJ and Rosato M (2022) The Nuclear 35S rDNA World in Plant Systematics and Evolution: A Primer of Cautions and Common Misconceptions in Cytogenetic Studies. Front. Plant Sci. 13:788911. doi: 10.3389/fpls.2022.788911

INTRODUCTION

Ribosomes are universal cellular components found across all domains of life. Research suggests they represent the most critical macromolecular machine in living organisms, as they are trusted with carrying out protein synthesis in cells by converting information encoded within mRNA into peptides (Ojha et al., 2020). Ribosome biogenesis in eukaryotes is a process of extraordinary complexity (Thomson et al., 2013). Four rRNA species are transcribed by two RNA polymerases, RNA Pol I (18S, 5.8S, 26S rRNA) and RNA Pol III (5S rRNA) being extensively modified during

their subsequent maturation in the macromolecular complex of the nucleolus, the nucleus and the cytoplasm (Sloan et al., 2017).

The ubiquitous presence of rDNA genes in nuclear, plastid and mitochondrial genomes has provided an opportunity to use ribosomal sequences as homologous markers to infer evolutionary processes and to assess systematic issues at the three basic domains of life, Bacteria, Archaea, Eukarya, including plants (Álvarez and Wendel, 2003; Nieto Feliner and Rosselló, 2007). In fact, the 5S intergenic spacers and the internal transcribed spacers (ITS) from the nuclear ribosomal 35S repeat have been proposed as nuclear standards for species identification (Gemeinholzer et al., 2006; Doveri and Lee, 2007; Gao et al., 2010), species delimitation (Müller et al., 2007) and DNA barcoding (Kress et al., 2005; Chase et al., 2007; CBOL Plant Working Group, 2009) in land plant lineages.

Ribosomal markers are not only limited to the use of rDNA sequences. Since the pioneering studies on the topic (Heitz, 1931), there has been extensive karyological work characterizing rDNA (García et al., 2017). The study of the development of the nucleolus, the assessment of karyological landmarks related to the linear differentiation of chromosomes, like non-centromeric constrictions and associated (satellite) regions, and the physical mapping and linkage between 35S rDNA and 5S rDNA families, have provided anchor points for comparative studies (Weiss-Schneeweiss and Schneeweiss, 2013). These include assessing the molecular evolution of rDNA units, gene silencing, evolutionary trends on karyotype differentiation, producing genetic maps and identifying ancestors in hybrids and polyploid species (Weiss-Schneeweiss et al., 2013). These karyo-evolutionary trends have complemented the knowledge on nuclear rDNA using DNA sequencing and are clearly relevant for providing cytogenetic markers in species that are routinely used for plant systematics and evolution purposes, and to postulate phylogenetic hypotheses (Totta et al., 2017).

The number, size, location and transcriptional activity of the 35S rDNA cistrons in plant karyotypes are some of the traditional cytogenetic landmarks reported in conventional karyotype descriptions (Battaglia, 1955). Their scrutiny has been useful to infer patterns of chromosomal evolution and the data have been used as a proxy for species discrimination, population differentiation and relationships at several evolutionary levels (Guerra, 2012). In agreement with their continued use in the field of cytotaxonomy, no major concerns have been reported for their application as cytogenetic markers in plant taxonomy and evolution (García et al., 2017).

The correct interpretation of 35S rDNA markers in plant taxonomy and phylogenetics is not free of drawbacks. This is not surprising given the complexities derived from the lability of the genetic architecture, the patterns of molecular evolution, and the evolutionary dynamics of the rDNA units, including the number and position of rDNA sites in the genome, sequence homogenization, structure and organization of rRNA genes (Feliner and Rosselló, 2012). In addition, the terminology used by independent authors is somewhat vague, which often complicates comparisons.

This complexity does not appear to be fully appreciated by some systematic plant studies that use chromosomal, or cytological-based rDNA markers, alone or in combination with genomic data. Unfortunately, misconceptions and misuses may lead to far-reaching conclusions based on risky assumptions that do not have general value, or on the use of confusing terminology (**Tables 1–4**). To date, no efforts acknowledging the potential problems and limitations involved in generating, utilizing, and interpreting the raw ribosomal data have been reported.

This paper aims to discuss the limitations of nuclear 35S rDNA markers (i.e., site number, transcriptional activity, nucleolar dominance) based on cytological and karyological experimental work to draw sound biological and evolutionary conclusions in a systematic and phylogenetic context. Also, we provide clarification for some conceptual misconceptions usually found in published work that could help lead to an insightful utilization of the ribosomal world in plant evolution.

THE NUCLEAR 35S rDNA LOCUS

Terminology: A Primer of Confusion

Most of the following terms are often used interchangeably in scientific literature: secondary constriction, nucleolar constriction, satellite, sputnik, intercalary satellite, nucleolar organizing regions (NORs), satellite(d) chromosome, SAT-chromosome, nucleolar chromosome, NOR chromosome, and 35S rDNA locus (**Figure 1**). However, in some cases there are significant differences that might not be rightly appreciated. Thus, inappropriate or confusing denominations may cause incorrect judgment on technically accurate data (**Table 1**).

The association of non-centromeric (i.e., secondary) constrictions (Darlington, 1926) and nucleolus formation during the interphase was noted early on by Heitz (1931) and McClintock (1934). The latter is credited for coining the term nucleolar organizer (later changed to NOR) to describe the chromosome region in *Zea mays* that was involved in the formation of the nucleolus (Pikaard, 2000a). In addition, a heterochromatic knob of the chromosome, the satellite, is present at the telomeric-proximal site of the secondary constriction, but is not involved in the formation of the nucleolus (Chen et al., 2000).

The term satellite chromosome is commonly used in the same way as SAT-chromosome. However, this is a misinterpretation recognized early on by Berger (1940), whose efforts to distinguish both technical terms have been repeatedly ignored in cytological literature (Battaglia, 1955). SAT was coined by Heitz (1931) as an abbreviation of Sine Acido Thymonucleinico (Without Thymonucleic Acid, the early denomination of DNA) and it refers to secondary constrictions. The decondensed chromatin at the NOR observed by Heitz (1931) showed less stained intensity (achromatic) than other chromosomal regions and was wrongly interpreted as lacking DNA. Accordingly, and following the original meaning, SAT-chromosome is not a synonym for satellited-chromosome but implies either a satellited chromosome or a chromosome with a secondary constriction that is associated with the formation of the nucleolus but does not have a satellite (Berger, 1940). In addition, and in contrast

TABLE 1 | The nuclear 35S rDNA landmarks and significant terminology associated to its activity, detection and morphology.

Assumption	Comments	Selected references
Plant species show chromosome complements lacking secondary constrictions	The presence of rDNA genes in the nuclear genome is a requisite for viable cellular metabolism in eukaryotes. Therefore, a minimum number of active ribosomal 35S rDNA units should be present in the nuclear genome. However, recognizing their location by the observed of decondensed chromatin along the chromosome (secondary constriction) may be compromised due to the use of conventional techniques (standard stains) that lack sensitivity, their position at subterminal or terminal ends of the chromosomes and the number of rDNA units and their activity. Alternative, more powerful techniques (Ag-NOR, immunolocalization) are needed to locate the transcriptionally active ribosomal loci at the secondary constrictions.	Terasaka and Tanaka, 1974; Gao et al., 2012; Gürdal and Özhatay, 2018
All SAT-chromosomes are satellited chromosomes	SAT-chromosome is not a synonym for satellited-chromosome, but implies either a satellited chromosome or a chromosome with a secondary constriction associated with the formation of the nucleolus, which does not have a satellite.	Berger, 1940
Only satellited chromosomes show NOR loci	Active 35S rDNA loci may be present at the terminal ends of chromosomes.	Battaglia, 1955; Rosato et al., 2017; Fehrer et al., 2021
All ribosomal loci are NOR	Only transcriptionally active 35S rDNA loci (which are evaluated through silver staining, the immunolocalization of histone methylation or histone deacetylation, and DNA cytosine methylation) involved in the formation of the nucleolus are NORs.	Poczai and Hyvönen, 2010; Marques et al., 2011; Milioto et al., 2019
Ribosomal loci detected by FISH are NOR loci	Hybridization *in situ* techniques using radioactive or non-isotopic probes detect all 35S rDNA loci above a threshold of the number of repeats, irrespective of their transcriptional activity.	Cuñado et al., 2000; Abd El-Twab and Kondo, 2010; Milioto et al., 2019
The number of ribosomal loci can be inferred from the number of satellite-bearing chromosomes	Only active rDNA loci are located at secondary constrictions. Silenced and pseudogene loci are not transcribed and cannot be seen as decondensed chromatin near the satellite body.	Galián et al., 2012; Rosato et al., 2017; Báez et al., 2020
FISH signals using 35S rDNA probes always shows canonical ribosomal loci	The differential amplification of coding and spacer sequences of the rDNA cistron and their transposition to other chromosomes has been reported in several species. These sites can be detected as FISH signals if significant similarity exist between the DNA probes used and the target sequences, but they are not true canonical rDNA loci.	Maggini et al., 1991; Guimond and Moss, 1999; Macas et al., 2003

TABLE 2 | Patterns of intraindividual 35S rDNA and NOR variation in plants.

Assumption	Comments	Selected references
The number of 35S rDNA loci is constant between plant tissues from a single individual	Aberrant mitosis in the binucleate tapetal cells of some organisms results in rDNA instability regarding the number of locus within a single tissue.	Chiavarino et al., 2000; Mursalimov and Deineko, 2018
The number of NOR loci is invariant within individuals	The differential suppressing of duplicated NOR loci by epigenetic silencing (differential amphiplasty or nucleolar dominance) may differ between tissues of a single individual. Reports indicating combined uniparental and biparental tissue-specific expression are known.	Dobešová et al., 2015; Sochorová et al., 2017; Borowska-Zuchowska et al., 2021
The number of ribosomal loci is not affected by vegetative propagation	Tissue culture by *in vitro* techniques has been reported to induce drastic changes (somaclonal variation) regarding the number of repeats and loci in many species.	Bairu et al., 2011; Rosato et al., 2016b
Ribosomal loci are always located in chromosomes of the regular complement	B-chromosomes can show silenced or active rDNA sites in some species. Individual variation in such accessory chromosomes may result in contrasting numbers of ribosomal loci.	Jones et al., 2008; Houben et al., 2013; D'Ambrosio et al., 2017
Ribosomal loci are always present in autosome chromosomes	The presence of 35S rDNA loci have has also been reported in sexual chromosomes.	Nakayama et al., 2001; Fujisawa et al., 2003; Nakao et al., 2005; Deng et al., 2012
The number of NOR sites are not gender dependent in dioecious species	In addition to their presence in autosomes, 35S rDNA loci may be linked to sexual chromosomes. The differential silencing of transcriptional activity in individual sites is known for the male individuals of some species. In these cases, male and female plants differ in the overall number of secondary constrictions and Ag-NOR sites.	Nakao et al., 2005

to the original meaning coined by McClintock (1934), some authors broadly define NOR sites as the chromosomal segments that contain ribosomal genes, irrespective of their transcriptional activity (Trerè, 2000).

In maize, there is a single pair of chromosomes showing secondary constrictions (i.e., two nucleolar chromosomes) that are also satellited chromosomes, a single active 35S rDNA locus or NOR, and a single associated nucleolus (Chen et al., 2000). This chromosomal rDNA pattern shows a complete agreement between chromatin decondensation sites (secondary constrictions), transcriptional active regions (NOR), the physical location of 35S rDNA units and the number of nucleoli formed (McClintock, 1934; Khuong and Schubert, 1985; Sadder and Weber, 2001).

TABLE 3 | Evolutionary trends in the number and activity of rDNA and NOR loci: cautions and limitations.

Assumption	Comments	Selected references
Diploid species are characterized by the presence of a single 35 rDNA locus	It has been estimated that about 65% of the analyzed diploid species of spermatophytes show two or more 35S rDNA loci as assessed by FISH.	Roa and Guerra, 2012; García et al., 2017
The presence of a single NOR locus is the evolutionary derived state for seed plant lineages	Most data available for the number of NORs in plants have not been discussed against phylogenetic inferences. This precludes the building of solid hypotheses about patterns of rDNA site change and the identifying the ancestral and evolutionary derived states.	Díaz Lifante, 1996; Bisht et al., 1998
Within closely related lineages the number of satellited chromosomes is associated to the ploidy level of the species	The number of satellited chromosomes may vary between congeneric species showing the same ploidy level.	Díaz Lifante, 1996; Bisht et al., 1998
An increase in the number of rDNA loci is always linked to polyploidy	The amplification of ribosomal loci may take place within homoploid lineages in the absence of polyploidy by means of transposition, chromosomal translocations and disploidy.	Hidalgo et al., 2017; Rosato et al., 2017; Totta et al., 2017
Within lineages the ancestral number of 35S rDNA loci is usually one	The ancestral number of rDNA loci is variable between lineages. Dynamic and complex changes have been documented in their evolutionary history regarding the amplification and deletion of repeats and loci involving chromosome repatterning.	Ran et al., 2001; Mishima et al., 2002; Weiss-Schneeweiss et al., 2008; Rosato et al., 2015; Totta et al., 2017
The number, genomic location and activity of 35S rDNA loci are constant within species	Changes in the number of loci and sites, their chromosomal position and the number of repeats per locus have been detected in several species, even within populations.	Schubert, 1984; Schubert and Wobus, 1985; Weiss-Schneeweiss et al., 2013; Rosato et al., 2017

TABLE 4 | Intra- and interindividual variation in nucleoli number: assumptions about their use in plant evolution.

Assumption	Comments	Selected references
The number of nucleoli in interphase nuclei equates to the number of ribosomal loci	Only the transcriptionally active rDNA loci give rise to nucleoli. If active and inactive 35S rDNA loci are present in a species the number of nucleoli will be formed only by the active loci. Since nucleoli tend to fuse (mononucleolation) during the cell cycle, only the highest number of nucleoli detected should be taken as the number of NORs present in the chromosome complement.	Brasileiro-Vidal et al., 2003; Rosato et al., 2016b
The number of nucleoli is constant within all tissues of a single plant	Cytomixis (the migration of nuclei and their components, including nucleoli, between two cells) has been reported in several plant tissues. This could lead to the observation of different numbers of nucleoli, which may differ also in dimensions, because of intercellular migration. In hybrid and allopolyploid species, the number of nucleoli may vary between tissues (usually between somatic and reproductive tissues) due to nucleolar dominance, the uniparental expression of the 35S rDNA genes.	Mursalimov and Deineko, 2011, 2018; Kumar and Singhal, 2016 Hasterok and Maluszynska, 2000; Idziak and Hasterok, 2008; Borowska-Zuchowska and Hasterok, 2017
The number of nucleoli can be used as an alternative way to determine the ploidy level	Active rDNA loci in hybrids and allopolyploids can be transcriptionally silenced by epigenetic processes (nucleolar dominance). Thus, nucleolar suppression may lower the expected number of nucleoli in polyploids. In addition, the deletion of rDNA units in duplicate loci may eventually lead to the elimination of entire NOR loci and the associated nucleoli.	Pikaard, 2000a; González-Melendi et al., 2005; Rosato and Rosselló, 2009

Complexity may arise in lineages, however, when mechanisms involved in altering the number of rDNA loci have occurred along their evolutionary history. These bursts of rDNA amplification include the dispersion of loci by structural chromosome rearrangements at homoploid levels as well as the transposition and amplification of rDNA copy numbers (Datson and Murray, 2006; Rosato et al., 2017), and have involved both diploid lineages and complex scenarios of ancestral and more recent polyploid events (Rosato et al., 2015). These events may have resulted in karyotypes that eventually contain more 35S rDNA sites than expected when compared to the ancestral or parental lineages. When karyotypes contain more than one 35S rDNA locus, it clearly shows that care should be exercised to describe the chromosomal rDNA pattern, since not all the commonly used terms are equivalent and should not be broadly used as mere alternatives with similar or identical meanings (**Figure 2**).

As an example that illustrates this issue, let us consider that the chromosomal complement of a hypothetical diploid species contains two 35S rDNA loci. After taking into account all heretofore reported relevant genomic locations and transcriptional activity alternatives that occur in the karyograms of seed plants (i.e., the number of active and silenced sites, the number of chromosome pairs where they are located, the position of the loci along the chromosome's arms, the presence of sites in sexual chromosomes, the number of secondary constrictions and the presence of satellites), up to eight potential ribosomal phenotypes may be described (**Figure 2**) by common descriptors frequently used in the literature (the number of SAT-chromosomes and satellited chromosomes, secondary constrictions and NOR sites). Thus, for a given karyotype (1) the number of SAT-chromosomes, satellite(d) chromosomes and nucleolar chromosomes may differ, (2) the counting of secondary constrictions may not be predictive of the overall number of rDNA loci and satellite(d) chromosomes, (3) some 35S rDNA loci may not be transcriptionally active and are not, accordingly, NOR sites, and (4) an odd number of NOR sites does not

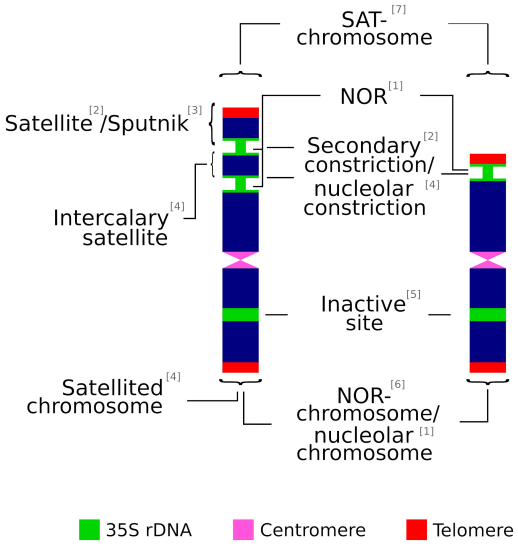


FIGURE 1 | Graphical summary of terms used to describe the chromosomic domains associated to the 35S rDNA locus. (1) McClintock, 1934; (2) Darlington, 1926; (3) Battaglia, 1999; (4) Battaglia, 1955; (5) Leitch and Heslop-Harrison, 1993; (6) Schwarzacher and Wachtler, 1983; (7) Heitz, 1931.

point to positional hemizygosity and may be attained instead by epigenetic silencing.

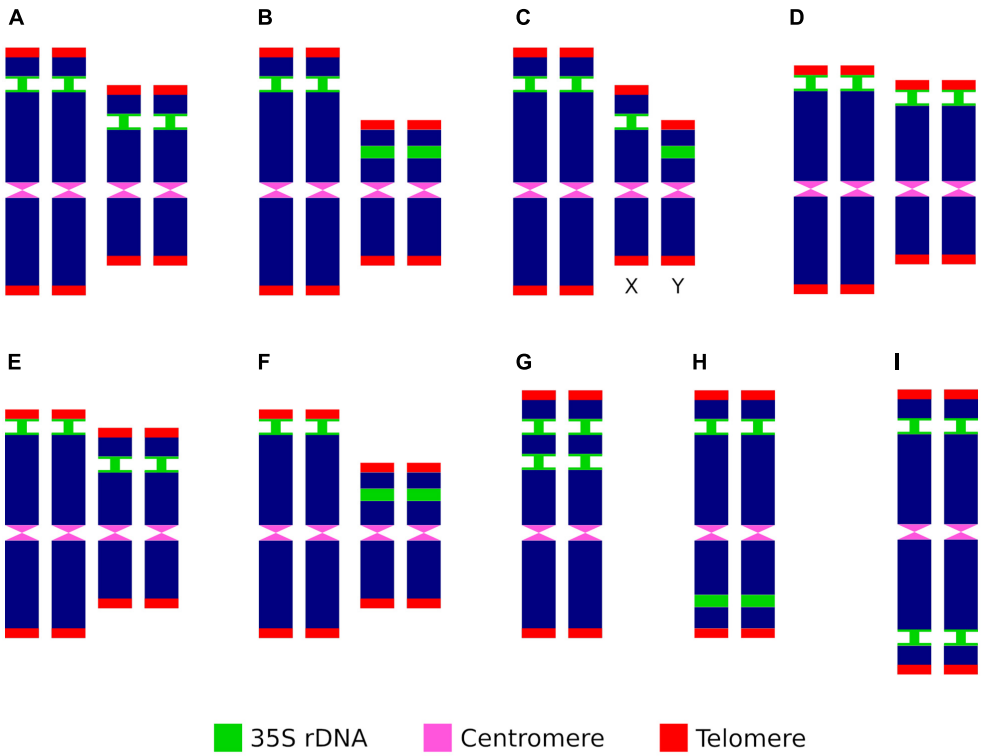
The Cytological Recognition of 35S rDNA Loci

Several methods are currently used to physically map the 35S rDNA units in plant cells. These include (1) the observation of chromosomes with conventional stains (e.g., acetic orcein, hematoxylin, Feulgen reagent, carmin acetic acid, Giemsa C-banding) or fluorescent dyes binding the DNA, e.g., DAPI (4',6-diamidino-2-phenylindole) and CMA (Chromomycin A3) (**Figures 3A,B**), looking for the presence of secondary constrictions that are associated with the nucleolus (Maluszynska et al., 1998), (2) the observation in secondary constrictions of acidic, non-histone proteins that bind silver ions (argyrophilic) and are differentially stained by silver impregnation (Goodpasture and Bloom, 1975; Tucker et al., 2010; **Figure 3C**), (3) the distribution of epigenetic marks at the rDNA loci such as methylation and deacetylation patterns of histone H3 and the pattern of DNA methylation (5-methylcytosine sites) (Marques et al., 2011; Borowska-Zuchowska and Hasterok, 2017), and (4) the detection of tandemly repeated rDNA copies using *in situ* hybridization techniques (Heslop-Harrison and Schwarzacher, 2011; Jiang, 2019; **Figure 3D**).

The above procedures differ in the level of specificity, sensitivity, and reproducibility in the required quality of the target cells, as well as in the time and complexity of the experimental work. Most importantly, the techniques differ in the type of information retrieved. The first three approaches (secondary constrictions, silver staining, and epigenetic patterns) basically identify the 35S rDNA sites which are transcriptionally active in the cell, and which may be cytologically visible from interphase to metaphase of mitosis and prophase I of meiosis. In contrast, *in situ* hybridization techniques on nuclei and chromosomes using DNA labeled probes identify both the active and inactive 35S rDNA sites. None of the available techniques are free of experimental drawbacks and limitations and the best results are usually obtained with a combination of methods.

The visualization of NORs by observing secondary constrictions using conventional and fluorochrome staining is a fast but rather crude technique which is still in use. It has been reported that the decondensation of the rDNA chromatin occurs in different ways depending on the cell type and the analyzed species (Leitch, 2000). It has even been suggested that the activity of rDNA at different loci may vary over the course of the cell cycle, with their transcription being determined in a time and region-specific manner (Li et al., 2006; Chandrasekhara et al., 2016). In addition, the length of the decondensed chromatin is connected to the number of rDNA units being transcribed. Thus, the observation of secondary constrictions on mitotic chromosomes may be difficult to obtain unless other confirmatory, more powerful methods are used. The highest level of chromatin decondensation displayed in meiosis, in contrast to mitotic chromosomes, makes prophase I (generally, from pachytene to diakinesis) an excellent stage for a highly accurate physical mapping of chromosomal landmarks (De Jong et al., 1999; Sepsi et al., 2018), including NORs. However, the association of the homologous chromosomes bearing NORs (nucleolar bivalent or NOR-bivalent) with the nucleolus may appear disconnected and is not always easily observed (McClintock, 1934) in a similar way as has been reported for somatic chromosomes. These explanations alone or in combination may be related to the unsuccessful cytological detection of secondary constrictions in plant chromosomes and should considerably mitigate their use when unorthodox observations are obtained. For instance, data on the presence of secondary constrictions in the karyotypes of the endemic flora of the Balearic Islands using conventional karyological techniques is known for only 13% of the species (Rosselló and Castro, 2008).

Nucleolar organizing regions contain argyrophilic proteins that are selectively stained by silver methods allowing their identification throughout the nucleolar area (Ag-NORs). Unfortunately, the use of such a sensitive stain is difficult to standardize and many technical improvements have been reported for plant and animal organisms from time to time (e.g., Biliński and Bilińska, 1996; Trerè, 2000). However, non-optimal silver impregnation resulting in a bright background of nuclear DNA (Goodpasture and Bloom, 1975) or unspecific results (for instance, the staining of centromeric regions; Báez et al., 2020) may compromise the reliability of the technique. Since the presence of NORs in a decondensed state is one of the prerequisites for positive silver staining (Jiménez et al., 1988), the detection of Ag-NORs is not always reliable. Research has revealed the presence of active rDNA sites which are barely visible or not reflected by secondary constrictions on metaphase chromosomes, even after Ag-NOR staining, in species from lineages that are not closely related (Berg and Greilhuber, 1993). These observations may have various origins, like a small



	Satellited chromosomes	NOR-chromosomes	Secondary constrictions	Active NORs	Inactive sites
A	4	4	4	4	0
B	2	2	2	2	2
C	3	3	3	3	1
D	0	4	4	4	0
E	2	4	4	4	0
F	0	2	2	2	2
G	2	2	4	4	0
H	2	2	2	2	2
I	2	2	4	4	0

FIGURE 2 | Hypothetical ribosomal phenotypes **(A–I)** and their associated ribosomal descriptors for individuals showing two rDNA loci and differing in transcriptional activity, the number of chromosome pairs where they are located, the position of the loci along the chromosome arms, the presence of sites in sexual chromosomes (X,Y), the number of secondary constrictions and the presence of satellites. Active sites are represented with narrow green stalks.

number of active ribosomal genes, a low transcriptional activity in some tissues or a different condensation level of chromatin in the NORs.

Fluorescence *in situ* hybridization (FISH) is by far the preferred karyological technique to assess 35S rDNA loci in plant chromosomes and very thorough compilations are available online (García et al., 2017). However, this approach alone does not differentiate the active and transcriptionally silent loci, and sequential methods (previous Ag-staining or the immunolocalization of epigenetic markers followed by FISH) are needed to obtain the most comprehensive results related to the number and functionality of 35S rDNA. FISH can reveal the major 35S rDNA loci of the genome, but it also has limitations or a lack of repeatability to detect minor loci characterized by a small number of rDNA units.

Intraindividual and Interspecific Variation in the Number and Activity of rDNA Sites: Sexual Chromosomes, Accessory Chromosomes, and Nucleolar Dominance

Assessments have reported that the number of secondary constrictions is gender dependent in several gymnosperm species, as *Ephedra foliata* (Mehra and Khitha, 1981), *Cycas* sp. pl. (Abraham and Mathew, 1962; Sangduen et al., 2007, 2009),

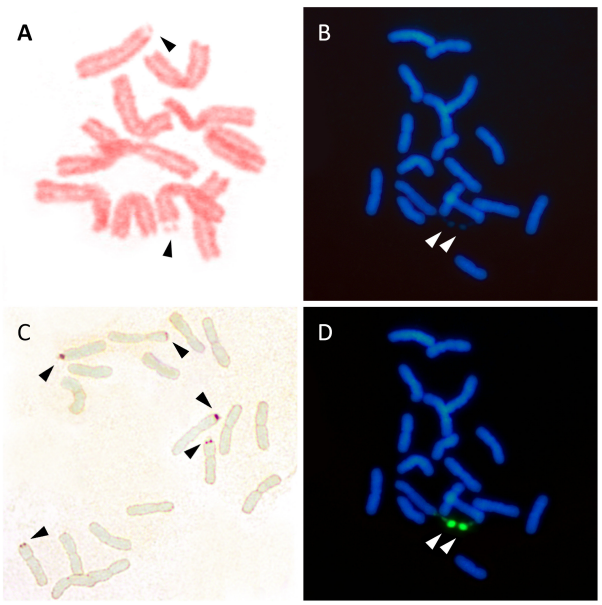


FIGURE 3 | Visualization of 35S rDNA locus using conventional staining **(A,B)**, silver impregnation **(C),** and FISH **(D)**. **(A)** *Lathyrus pisiformis*. A single chromosome pair with secondary constrictions after Feulgen staining. **(B)** *Achillea maritima*. The presence of two secondary constrictions (arrows) are observed using DAPI staining. **(C)** *Anacyclus homogamos*. Five NORs are present after Ag-NOR staining (arrows). **(D)** *Achillea maritima*. A single 35SrDNA locus is located on the secondary constrictions after FISH (green signals).

Ginkgo biloba (Nakao et al., 2005), and *Stangeria eriopus* (Kokubugata et al., 2002). Observations in *Ephedra* and *Cycas* were based on conventional cytological techniques suggesting that faint satellites located at the end of the postulated sexual chromosomes (XY system), differentiate between X (with a satellite at each arm end) and Y chromosomes (a satellite at one end only, or none). Sangduen et al. (2009) concluded that the karyotype of male and female *Cycas* species could be clearly distinguished by the number of homomorphic and heteromorphic chromosome pairs possessing secondary constrictions. Although uncertainties about the accuracy of these cytological observations have been expressed by some authors dealing with sex chromosomes in gymnosperms (Ming et al., 2011), no available experimental work has been produced to dispute these findings, some of which were corroborated in somatic and generative tissues (Abraham and Mathew, 1962). Two heteromorphic chromosome pairs connected to rDNA sites were detected in *Stangeria* using a homologous probe (Kokubugata et al., 2002). The failure to associate this finding with sex chromosomes was due to the fact that the sex of the individual investigated was unknown.

Similar observations were reported in *G. biloba* chromosomes, where male individuals showed three chromosomes with secondary constrictions whereas four were present in female plants (Nakao et al., 2005). Additional studies by Lan et al. (2008) reported that in male individuals, the satellites of chromosome 1 (the biggest of the complement) were homomorphic, while in females they were heteromorphic, and one appeared to be bigger than the other. Both male and female plants showed the same number of rDNA sites (four) when performing FISH using a homologous rDNA probe (Nakao et al., 2005). Therefore, the differential epigenetic silencing of a whole rDNA site associated to the presence of a different rDNA copy number in homologous chromosomes may explain these cytological singularities in *Ginkgo*.

Accessory or supernumerary B chromosomes (Bs) are one of the most captivating topics of the evolution of the nuclear eukaryote genome. Bs are additional dispensable genetic components that contribute as part of the genome of a great diversity of organisms including plants (D'Ambrosio et al., 2017; Houben et al., 2019; Pokorná and Reifová, 2021). Several studies have revealed the lack of essential genes in their composition, except for the eventual presence of 35S and 5S rDNA families (Sýkorová et al., 2003; Jones et al., 2008; Houben et al., 2013). The activity of 35S rDNA sites located on Bs have been analyzed, and diversity in the transcriptional activity has been revealed in a similar way as is known for the chromosomes of the regular chromosomal complement (A chromosomes).

In *Brachyscome dichromosomatica* the 35S rDNA sites located at the large and micro Bs are silenced and are not associated to the nucleolus (Marschner et al., 2007). In contrast, in *Plantago lagopus* and *Crepis capillaris* the Bs are transcriptionally active (NOR) and contribute to the genesis of the nucleolus (Jones, 1995; Dhar et al., 2002). In *Nierembergia aristata*, Bs possess not only strong nucleolar activity, but also show nucleolar competition with the A chromosomes (Acosta and Moscone, 2011). This phenomenon could be analogous to the nucleolar dominance that occurs in interspecific hybrids (see below). Moreover, *Secale cereale* show the presence of B chromosomes without 35S rDNA sites. Interestingly, it has been suggested that Bs changes the rDNA organization pattern in interphase nuclei as detected by a drastic increase of rDNA condensed blocks inside the nucleolus (Delgado et al., 2004). Available evidence suggests that the rDNA alteration is caused by the presence of the B chromosomes themselves rather than by an obvious dosage effect (Delgado et al., 2004). The singular nature of B chromosomes exhibiting specific genomic features illustrates the need for caution when analyzing the pattern of nucleolar activity (**Table 2**).

Differential amphiplasty, also known as nucleolar dominance or reversible NOR silencing, is a conspicuous cytological and complex molecular phenomenon which was known to early plant cytogeneticists (Navashin, 1934). When the nuclear genomes of related species are merged by hybridization processes their NORs may differ in their competitive ability to transcribe the ribosomal genes and form the nucleolus. The net result of this is that a set of NORs from one of the parental species is epigenetically suppressed and fails to form secondary constrictions, thus leading to a decrease in the number of nucleoli. Accordingly, the NORs of hybrid and allopolyploids may not be added together and thus fail to reveal all the active rDNA sites inherited from the two parents (Tucker et al., 2010).

Interestingly, in individuals of hybrid origin, the presence of nucleolar dominance may be tissue-specific and the rDNA

sites could be differentially expressed (Chen and Pikaard, 1997; Borowska-Zuchowska et al., 2021). These combined uniparental and biparental patterns of NOR silencing showing a contrasting number of NORs has detected in vegetative and reproductive tissues. In allotetraploid *Brassica* species (*B. carinata*, *B. juncea*, *B. napus*), rRNA genes silenced in leaves were found to be transcribed in all floral organs where biparental gene expression was maintained (Chen and Pikaard, 1997; Sochorová et al., 2017). However, Hasterok and Maluszynska (2000) reported that nucleolar dominance did not occur in root tip cells from these polyploid species. Biparental rDNA expression was found in roots, flowers and callus in the allotetraploid *Tragopogon mirus* (Dobešová et al., 2015). However, uniparental dominance was maintained in its leaves. These observations clearly indicate that silenced and derepressed rRNA genes may occur not only during developmental stages, but within vegetative and reproductive tissues. Since cytogenetic observations are preferentially limited to favorable cell types such as root tips and pollen mother cells, there may be contrasting reports on the number of active NOR sites (**Table 2**).

Ribosomal Loci as Evolutionary Markers: Some Cautions on Premises and Interpretations

Detailed knowledge of the number of rDNA loci, their genomic location and the linkage of the 35S and 5S rDNA units has been assessed for a substantial number of species (García et al., 2012, 2014; Roa and Guerra, 2012; Vitales et al., 2017). In contrast, compilations of NOR activity are very few and, unfortunately, this knowledge has not been updated since the work of Lima-de-Faria (1976). In addition, most information on the dynamics of rDNA loci variation has not been thoroughly analyzed under explicit phylogenetic frameworks, and the data has been subject to speculative interpretations. In sharp contrast with what is known in selected polyploid species, the evolutionary patterns of rDNA loci number in predominantly diploid lineages are insufficiently understood and have received less attention (e.g., Datson and Murray, 2006; Totta et al., 2017). Moreover, most previous studies on rDNA loci changes lack explicit temporal frames, and as a result, their dynamics could not be assessed with certainty. The fact that surprisingly few studies have addressed intrapopulation, interpopulation and interspecific levels of rDNA variability in non-model wild plants, may question the assumptions of generalized evolutionary trends that are usually based on very few case studies.

Thus, the pervading perceptions that (1) the number, genomic location, and activity of 35S rDNA loci are constant within species, (2) diploid species are usually characterized by a single rDNA locus, (3) the ancestral number of 35S rDNA loci is usually one, (4) the presence of a single NOR locus is the derived state for seed plant lineages and (5) the increase in the number of rDNA loci is mostly linked to polyploidy, should all be checked on a case-by-case basis (**Table 3**).

The cytogenetic research conducted on the Mediterranean *Anacyclus* (Asteraceae), a diploid genus comprising nine species of weedy annuals and a few perennials, has provided relevant results illustrating how the above statements are not generally applicable in plants. Available karyological rDNA data was first obtained by Schweizer and Ehrendorfer (1976) and Ehrendorfer et al. (1977), who determined the number of active rDNA loci for all species based on a small number of accessions. Later, Rosato et al. (2017) determined the number and chromosomal position of 35S rDNA sites in 196 individuals from 47 populations in all *Anacyclus* species using FISH. The following conclusions could be firmly established from the results obtained by both research teams. First, the level of rDNA site-number variation detected within most *Anacyclus* species was outstanding and included both intra-specific and intra-population polymorphisms that encompassed a large part of the range of variation found in all angiosperms. Second, no clear association could be established between the phylogenetic position of the species and the number of rDNA sites. Third, the cytogenetic changes underlying the inferred rDNA dynamism were not related to polyploidy and were likely triggered by genomic rearrangements resulting from contemporary hybridization. Finally, the number of NORs in the genus was not associated to the phylogenetic ancestry of the species; the perennial clade showed two loci whereas the most derived annual species presented three loci.

Inferring polyploidy based on the number of NOR and rDNA sites may be misleading. An increase of NORs, and thus of nucleoli, is accomplished not only by genome duplication (**Table 3**), as had been earlier postulated (Gates, 1942; Fankhauser and Humphrey, 1943). Additional processes including structural rearrangements, ectopic recombination and rDNA transposition have been proposed as alternative mechanisms to explain NOR amplification within genomes (Pikaard, 2000b; Cabrero and Camacho, 2008, and references therein). Specifically, the intragenomic mobility of rRNA genes because of transposon activity, which can produce a translocation of rDNA copies to new genomic sites has been substantiated in seed plants, and it has been hypothesized that it is one of the major forces driving rDNA locus evolution in connection with the origin of new 35S rDNA sites (Dubcovsky and Dvorák, 1995; Raskina et al., 2004; Datson and Murray, 2006).

The basal chromosome number for each major lineage of early land plants (liverworts, mosses, and hornworts) is not known with certainty, and the topic has been greatly debated. The fact that two NOR loci were reported in some liverwort species (*Riccardia pinguis*) led Berrie (1958a,b) to hypothesize that species with $n = 10$ were polyploids derived from $n = 5$ hornwort ancestors. He further suggested that liverworts with gametophytic numbers of chromosomes of $n = 8$, 9 or 10 are basically polyploids that evolved from $n = 10$ ancestors through disploidy. The hypothesis of Berrie (1958a,b) was based on the misconception that the presence of two nucleolar chromosomes in the haploid complement of a plant is a marker of polyploidy. Recent work has shown that in bryophytes the number of 35S rDNA loci and copies are not correlated to ploidy level (Rosato et al., 2016a). Also, a lack of association between the number of 35S rDNA sites and polyploidy is known in *Medicago* (Fabaceae) where some tetraploid species (*M. arborea*, *M. strasseri*) show a single rDNA site, the same number usually present in diploid lineages (Rosato et al., 2008; **Figure 4**).

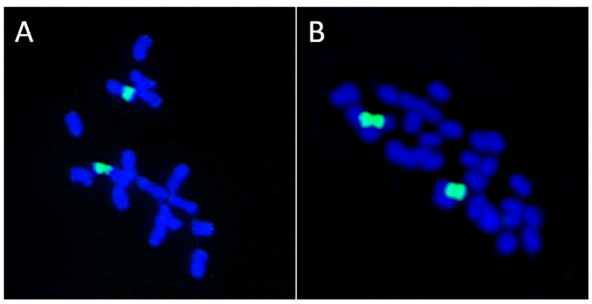


FIGURE 4 | Lack of association between the number of ribosomal loci and polyploidy. **(A)** *Medicago marina*, 2*n* = 2x = 16. **(B)** *M. arborea*, 2*n* = 4x = 32. Both species possess a single 35S rDNA locus as revealed by FISH (green signals).

It is known that several independent events of entire genome duplications are characteristic of almost all fundamental lineages of land plants, and they are considered a major driving force in species diversification (Clark and Donoghue, 2018; Ren et al., 2018). Hence, attempts made to associate species with low chromosome numbers to diploid entities are confirmed to be incorrect as more and more events of paleoploidization are discovered (Clark and Donoghue, 2018). This suggests that overall assumptions connecting chromosome number and the ancestral or derived state to the number of 35S rDNA loci and NOR loci should be viewed with utmost caution. Despite these caveats, it has been estimated that 64.9–66% of the analyzed *diploid* species of seed plants show more than a single 35S rDNA locus as assessed by FISH (Roa and Guerra, 2012; García et al., 2017). These figures clearly disagree with the assumptions that the presence of a single rDNA in the genome supports the diploid status of a species.

One of the flowering plant genera that has received many efforts to infer the patterns and evolutionary significance of satellite chromosomes (referred to as SAT-chromosomes in the literature) is *Taraxacum* (dandelions), a complex genus where sexual (diploid) and, mostly, apomitic (polyploid) lineages occur. Mogie and Richards (1983) noticed the presence of two markedly divergent satellited chromosomes in the genus differing in the location and visibility of the secondary constrictions as observed after conventional Feulgen staining. The *Taraxacum*-type patterns were characterized by the presence of a conspicuous and intercalary secondary constriction, whereas the identified so-called conventional type satellite had a subterminal location and an extremely small distal euchromatic region Mogie and Richards (1983). The karyological inspection of 123 dandelion species by their SAT-chromosomes showed that the species belonging to the 10 sections hypothesized to be most primitive in the genus lack satellited chromosomes. Surprisingly, most analyzed species lacking SAT-chromosomes were reported to be sexual diploids, whereas a variable and usually unstable number of *Taraxacum*-type or conventional-type satellites was found in polyploids. Additional observations revealed a most inconsistent pattern of satellites and secondary constrictions in a single species, involving their number, size and chromosomal location, even within individuals (Richards, 1989). The variation in these cytological rDNA markers did not appear to be associated with the geographic origin of the populations sampled or the taxonomic adscription of the species, and accordingly, a constant number of satellites was rarely recorded for any species (Den Nijs et al., 1978; Krahulcová, 1993). It was suggested that such uncertainties regarding secondary constrictions might partially be explained by the experimental vagueness associated to the lack of strict uniformity of the pretreatment procedures (Den Nijs et al., 1978). Despite these concerns, however, there has been an increase in the number of reports on the number, size and morphology of satellited chromosomes for taxonomic and evolutionary purposes in *Taraxacum* (e.g., Gürdal and Özhatay, 2018; Watanabe et al., 2021). Again, the apparent lack of secondary constrictions continues to be reported in some species (*T. scaturiginosum*) (Gürdal and Özhatay, 2018). A recent study investigated the position and number of 35S rDNA loci in 38 Taraxacum species covering different reproduction modes, geographical regions and putative phylogenetic groups using FISH (Macháčková et al., 2021). Interestingly, these authors do not support the presence of the previously differentiated *Taraxacum*-type and conventional-type satellites by Mogie and Richards (1983). Most importantly, all analyzed species showed at least one secondary constriction, refuting the view that the lack of SAT-chromosomes is a reliable karyological marker in dandelions, as previously indicated (Mogie and Richards, 1983; Gürdal and Özhatay, 2018). However, the work of Macháčková et al. (2021) raises additional questions. These authors reported and provided images of the presence of conspicuous secondary constrictions in chromosomes lacking FISH signals for 35S rDNA. These findings are surprising and require additional verification to use nucleolar chromosomes as a meaningful karyological and evolutionary marker in *Taraxacum*.

The location and number of secondary constrictions have also been used as relevant cytogenetic features to investigate the karyotype evolution in the Lilieae tribe, as assessed by conventional staining (Gao et al., 2012). These authors characterized the *Notholirion* genus by the absence of nucleolar constrictions in the three analyzed species, in contrast with the recognized presence in all species of the related *Cardiocrinum*, *Fritillaria*, *Lilium,* and *Nomocharis* genera (Gao et al., 2012). Since *Notholirion* showed a basal position in the inferred phylogenetic tree based on nuclear ribosomal Internal Transcribed Spacer sequences, Gao et al. (2012) hypothesized that the lack of secondary constrictions was the ancestral state in Lilieae and that they *emerged and evolved* as the apparition of genera took place over time. Since the secondary constrictions on chromosomes represent the expression of rRNA genes which were transcribed during the preceding interphase (Tucker et al., 2010), species in the Lilieae tribe for which no nucleolar constrictions were recorded by Gao et al. (2012) should obviously present at least one chromosome pair showing active rDNA units. Clearly, the evolutionary interpretations drawn from the data appear unreliable, and new research using FISH is imperative to assess the pattern of NOR evolution in the group.

THE NUCLEOLUS

The nucleolus is a conspicuous domain delimiting the nuclear territory of transcriptionally active and mostly de-condensed ribosomal 35S rRNA genes, where the ribosomal units are assembled (Bersaglieri and Santoro, 2019). Nucleoli can be easily observed and analyzed in interphase nuclei (also in meiotic prophase I) by the same conventional silver nitrate staining and protocols that are applied to NOR loci (Lacadena et al., 1984; Xu and Earle, 1996). In recent years, several *in situ* markers of plant nucleoli have been implemented which involve methods for tagging specific nucleolar proteins with fluorescent tags, by raising antibodies or through nascent DNA and RNA using the so-called click iT chemistry (Dvořáčková and Fajkus, 2018). Nevertheless, silver staining is a relatively easy and fast approach that is still in use and advantageous over many more demanding, expensive and time-consuming protocols. In addition, silver staining allows sequential FISH to be done in the same slides, which reveals the number, location and transcriptional activity of the 35S rDNA loci (e.g., Castro et al., 2018). Observing the nucleoli in interphase nuclei has its own value and may complement the information gained by analyzing the 35S rDNA loci in metaphase. However, care should be exercised when nucleolus descriptors (usually the number) are used as evolutionary markers in plants (**Table 4**) since nucleoli are dynamic substructures regarding their activity, size, position, and number (Leitch, 2000) and their morphology can be modified when several types of stress occur (Hayashi and Matsunaga, 2019).

The number of nucleoli has been correlated with the number of NORs present in the chromosome complement and in the ploidy level (Venkatesh et al., 2019). However, earlier claims suggesting a close connection (i.e., linearity) between the number of nucleoli and polyploidy (Gates, 1942; Fankhauser and Humphrey, 1943) are certainly not universally used. Thus, ploidy is not related to the nucleolar number but to the nucleolar size in the fruit cells of *Solanum lycopersicum* (Bourdon et al., 2012).

Counting nucleoli is one of the techniques used to estimate the ploidy level of individuals without having dividing cells, which may be useful when using tissues with a low mitotic index (Ochatt and Seguí-Simarro, 2021). Since nucleolar fusion frequently occurs during the cell cycle (mononucleation), only the highest number of nucleoli detected should be taken as evidence of the number of active rDNA loci (Ochatt and Seguí-Simarro, 2021; **Figures 5A–C**). A confident determination of the number of nucleoli should be estimated in a large sample size on this basis. Nucleoli have been detected in fossilized stems of a royal fern (Osmundaceae) dating back 180 million years (Bomfleur et al., 2014) and in herbarium specimens (Laane and Höiland, 1986), paving the way to assess the estimation of ploidy levels in extinct species and populations of well-preserved museum specimens.

The assumption that the number of nucleoli is constant within individuals and species is risky and should be validated on a case-by-case basis. The variability of NOR activity, expressed as the number and size of silver-stained active sites, has been detected between cells of single individuals, and also between individuals of the same population in cock's-foot (Besendorfer et al., 2002).

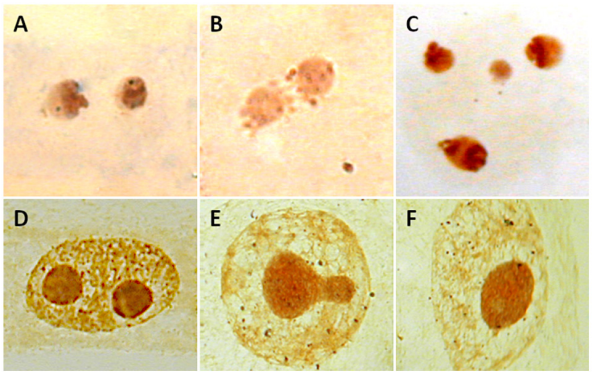


FIGURE 5 | Nucleoli in interphase nuclei of *Medicago* species **(A–C)** and *Cistus heterophyllus* **(D–F)** after Ag-NOR staining. **(A)** Two nucleoli are observed in diploid *M. marina* (2n = 2x = 16). **(B)** A single nucleoli is observed in *M. arborea* (2*n* = 4x = 32). **(C)** Four nucleoli are shown in hexaploid *M. citrina* (2*n* = 6x = 48). **(D–F)** Process of mononucleation (nucleoli fusion) in *C. heterophyllus*.

In addition, atypical nucleolar behaviors across the cell cycle, including the presence of numerous small silver-positive bodies in nuclei and cytoplasm, have been reported in some species and may also compromise the interpretation of the results (Dagne and Heneen, 1992). Moreover, the migration of nuclei, including nucleoli, between plant cells (cytomixis) is a cellular phenomenon frequently observable in the male meiosis of higher plants, but the causes and consequences of cytomixis are still not entirely understood (Mursalimov and Deineko, 2018). This intercellular migration could lead to the detection of a different number of nucleoli in cells associated with the production of pollen. Finally, in hybrid and allopolyploid species, the number of nucleoli may vary between tissues (usually between vegetative and reproductive tissues and at several developmental stages) due to nucleolar dominance, the uniparental expression of the NORs (Chen and Pikaard, 1997; Hasterok and Maluszynska, 2000; Dobešová et al., 2015) as indicated above.

Additional inconsistencies related to a lack of full correspondence between the number of NORs present in the chromosome complement and the number of nucleoli have been noted. For several species, there have been reports indicating that the number of Ag-NORs in metaphase may disagree with the maximum number of nucleoli recorded in interphase nuclei (e.g., Scaldaferro et al., 2016). In *Solanum lycopersicum* meiosis, pachytene chromosomes show five 35S rDNA loci as revealed by FISH (Xu and Earle, 1996). However, only the major rDNA locus is associated to the nucleolus by Ag-NOR staining, whereas the remaining four minor loci, which are also active after silver impregnation, are not. Thus, in this species there is a correspondence between the number of Ag-NOR pachytene chromosomes and the maximum number of nucleoli observed (five), but not all NOR-bivalents could be associated to a nucleolus (Xu and Earle, 1996).

Several hypotheses attempt to explain this fact, including nucleolar association, the merging of the nucleolus during

the interphase (Lacadena et al., 1984), or the occurrence of interchromosomal nucleolar dominance, where NORs from different pairs of chromosomes compete to make up the nucleoli (Tapia-Pastrana, 2020). The presence of cryptic NORs, chromosome regions apparently lacking rDNA loci, as revealed by FISH and silver staining, but which give rise to small nucleoli in interphase has been reported in plant and animal species (Sato et al., 1981; Cabrero and Camacho, 2008).

The biological reasons underlying the unsteadiness in the number of chromosomes with NOR sites and the number of nucleoli may be varied and are not fully understood. However, past and ongoing interspecific hybridization has been suggested as one of the most outstanding causes involved in the generation of cytological abnormalities that modify the regulatory system of the cell and contribute to decreasing the connection between the number of NORs and nucleoli (Levin, 1973; Tapia-Pastrana, 2020). For instance, in *Phlox hybrida* a high and significant correlation between the population hybridity and incidence of accessory nucleoli was detected (Levin, 1973). Supernumerary nucleoli were observed in *Crotalaria agatiflora* and their presence was attributed to the hybrid origin of this species (Verma and Raina, 1981). However, hybridization was rejected as the driving factor involved in the presence of accessory nucleoli in the diploid species *Trigonella foenum-graecum* although no alternative hypothesis was substantiated (Lakshmi and Raghavaiah, 1984).

Lack of association between number of nucleoli and polyploidy was detected in *Medicago* species (Rosato and Rosselló, 2009). A single nucleolus was present both in diploid (*M. marina*; 2n = 16) and tetraploid species (*M. arborea*, *M. strasseri;* 2n = 32). In addition, hexaploid *M. citrina* (2n = 48), for which the presence of three nucleoli should be theoretically expected, showed only two nucleoli (Rosato and Rosselló, 2009; **Figures 5D–F**). Inconsistencies also apply to non-flowering plants like bryophytes. Populations showing one or two nucleoli were registered in the gametophytic haploid (n = 9) *Pellia endiviifolia* noted (Newton, 1988). In contrast, a single nucleolus was present in diploid (n = 18) *P. borealis* (Newton, 1986).

AUTHOR CONTRIBUTIONS

JR: conceptualization, writing—original draft preparation, and funding acquisition. AM, MR, and JR: writing–review and editing. MR: supervision. All authors have equally contributed, read and agreed to the published version of the manuscript.

FUNDING

This study was supported by a grant from the Spanish Ministry of Economy and Competitivity through the project CGL2017-88500-P.

REFERENCES

Abd El-Twab, M. H., and Kondo, K. (2010). Characterization of chromosome complement in *Tridactylina kirilowii* (Turcz. ex DC.) Schultz-Bip. by aceto-orcein, CMA, DAPI and FISH. *Chromosom. Bot.* 5, 15–21. doi: 10.3199/iscb.5.15

Abraham, A., and Mathew, P. (1962). Cytological studies in the Cycads: sex chromosomes in *Cycas*. *Ann. Bot.* 26, 261–266. doi: 10.1093/oxfordjournals.aob.a083792

Acosta, M. C., and Moscone, E. A. (2011). B Chromosomes in *Nierembergia aristata* (Solanaceae): nucleolar activity and competition with the A chromosomes. *Cytogenet. Genome Res.* 132, 105–112. doi: 10.1159/000320705

Álvarez, I., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/S1055-7903(03)00208-2

Báez, M., Souza, G., and Guerra, M. (2020). Does the chromosomal position of 35S rDNA sites influence their transcription? A survey on *Nothoscordum* species (Amaryllidaceae). *Genet. Mol. Biol.* 43, e20180194. doi: 10.1590/1678-4685-GMB-2018-0194

Bairu, M. W., Aremu, A. O., and Van Staden, J. (2011). Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul.* 63, 147–173. doi: 10.1007/s10725-010-9554-x

Battaglia, E. (1955). Chromosome morphology and terminology. *Caryologia* 8, 179–187. doi: 10.1080/00087114.1955.10797556

Battaglia, E. (1999). The chromosome satellite [Navashin's sputnik or satelles]: a terminological comment. *Acta Biol. Crac. Ser. Bot.* 41, 15–18.

Berg, C., and Greilhuber, J. (1993). Cold-sensitive chromosome regions and heterochromatin in *Cestrum aurantiacum* (Solanaceae). *Plant Syst. Evol.* 185, 259–273. doi: 10.1007/BF00937662

Berger, C. A. (1940). SAT-chromosomes. *Science* 92, 380–381. doi: 10.1126/science.92.2391.380-b

Berrie, G. K. (1958a). The nucleolar chromosome in hepatics. I. *Trans. Br. Bryol. Soc.* 3, 422–426. doi: 10.1179/006813858804829451

Berrie, G. K. (1958b). The nucleolar chromosome in hepatics, II. A phylogenetic speculation. *Trans. Br. Bryol. Soc.* 3, 427–429. doi: 10.1179/006813858804829334

Bersaglieri, C., and Santoro, R. (2019). Genome organization in and around the nucleolus. *Cells* 8:579. doi: 10.3390/cells8060579

Besendorfer, V., Samardzija, M., Zoldos, V., Solic, M. E., and Papes, D. (2002). Chromosomal organization of ribosomal genes and NOR-associated heterochromatin, and NOR activity in some populations of *Allium commutatum* Guss. *Bot. J. Linn. Soc.* 139, 99–108. doi: 10.1046/j.1095-8339.2002.00047.x

Biliński, S. M., and Bilińska, B. (1996). A new version of the Ag-NOR technique. A combination with DAPI staining. *Histochem. J.* 28, 651–656. doi: 10.1007/BF02331386

Bisht, M. S., Kesavacharyulu, K., and Raina, S. N. (1998). Nucleolar chromosome variation and evolution in the genus *Vicia*. *Caryologia* 51, 133–147. doi: 10.1080/00087114.1998.10589128

Bomfleur, B., McLoughlin, S., and Vajda, V. (2014). Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in royal ferns. *Science* 343, 1376–1377. doi: 10.1126/science.1249884

Borowska-Zuchowska, N., and Hasterok, R. (2017). Epigenetics of the preferential silencing of *Brachypodium stacei*-originated 35S rDNA loci in the allotetraploid grass *Brachypodium hybridum*. *Sci. Rep.* 7:5260. doi: 10.1038/s41598-017-05413-x

Borowska-Zuchowska, N., Robaszkiewicz, E., Mykhailyk, S., Wartini, J., Piński, A., Kovarik, A., et al. (2021). To be or not to be Expressed: the first evidence of a nucleolar dominance tissue-specificity in *Brachypodium hybridum*. *Front. Plant Sci.* 12:768347. doi: 10.3389/fpls.2021.768347

Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Brown, S., et al. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. *Development* 139, 3817–3826. doi: 10.1242/dev.084053

Brasileiro-Vidal, A. C., Cuadrado, A., Brammer, S., Zanatta, A. C., Prestes, A., Moraes-Fernandes, M. I. B., et al. (2003). Chromosome characterization in

Thinopyrum ponticum (Triticeae, Poaceae) using *in situ* hybridization with different DNA sequences. *Genet. Mol. Biol.* 26, 505–510. doi: 10.1590/S1415-47572003000400014

Cabrero, J., and Camacho, J. P. M. (2008). Location and expression of ribosomal RNA genes in grasshoppers: abundance of silent and cryptic loci. *Chromosome Res.* 16, 595–607. doi: 10.1007/s10577-008-1214-x

Castro, C., Carvalho, A., Pavia, I., Leal, F., Moutinho-Pereira, J., and Lima-Brito, J. (2018). Nucleolar activity and physical location of ribosomal DNA loci in *Vitis vinifera* L. by silver staining and sequential FISH. *Sci. Hortic.* 232, 57–62. doi: 10.1016/j.scienta.2017.12.064

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12794–12797. doi: 10.1073/PNAS.0905845106

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., Van Den Berg, C., Madriñán, S., Petersen, G., et al. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon* 56, 295–299. doi: 10.1002/tax.562004

Chen, C. C., Chen, C. M., Hsu, F. C., Wang, C. J., Yang, J. T., and Kao, Y. Y. (2000). The pachytene chromosomes of maize as revealed by fluorescence *in situ* hybridization with repetitive DNA sequences. *Theor. Appl. Genet.* 101, 30–36. doi: 10.1007/s001220051445

Chen, Z. J., and Pikaard, C. S. (1997). Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3442–3447. doi: 10.1073/pnas.94.7.3442

Chiavarino, A. M., Rosato, M., Manzanero, S., Jiménez, G., González-Sánchez, M., and Puertas, M. J. (2000). Chromosome nondisjunction and instabilities in tapetal cells are affected by B chromosomes in maize. *Genetics* 155, 889–897. doi: 10.1093/genetics/155.2.889

Clark, J. W., and Donoghue, P. C. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23, 933–945. doi: 10.1016/j.tplants.2018.07.006

Cuñado, N., De la Herrán, R., Santos, J. L., Ruiz Rejón, C., Garrido-Ramos, M. A., and Ruiz Rejón, M. (2000). The evolution of the ribosomal loci in the subgenus *Leopoldia* of the genus *Muscari* (Hyacinthaceae). *Plant Syst. Evol.* 221, 245–252. doi: 10.1007/BF01089296

Dagne, K., and Heneen, W. K. (1992). The karyotype and nucleoli of *Guizotia abyssinica* (Compositae). *Hereditas* 117, 73–83. doi: 10.1111/j.1601-5223.1992.tb00010.x

D'Ambrosio, U., Alonso-Lifante, M. P., Barros, K., Kovaøík, A., de Xaxars, G. M., and García, S. (2017). B-chrom: a database on B-chromosomes of plants, animals and fungi. *New Phytol.* 216, 635–642. doi: 10.1111/nph.14723

Darlington, C. D. (1926). Chromosome studies in the Scilleae. *J. Genet.* 16, 237–251. doi: 10.1007/BF02983000

Datson, P. M., and Murray, B. G. (2006). Ribosomal DNA locus evolution in *Nemesia*: transposition rather than structural rearrangement as the key mechanism?. *Chromosome Res.* 14, 845–857. doi: 10.1007/s10577-006-1092-z

De Jong, J. H., Fransz, P., and Zabel, P. (1999). High resolution FISH in plants-techniques and applications. *Trends Plant Sci.* 4, 258–263. doi: 10.1016/S1360-1385(99)01436-3

Delgado, M., Caperta, A., Ribeiro, T., Viegas, W., Jones, R. N., and Morais-Cecilio, L. (2004). Different numbers of rye B chromosomes induce identical compaction changes in distinct A chromosome domains. *Cytogenet. Genome Res.* 106, 320–324. doi: 10.1159/000079306

Den Nijs, J. C. M., Sterk, A. A., and Van der Hammen, H. (1978). Cytological and ecological notes on the *Taraxacum* sections *Erythrosperma* and *Obliqua* of the coastal area of the Netherlands. *Acta Bot. Neerl.* 27, 287–305. doi: 10.1111/j.1438-8677.1978.tb00303.x

Deng, C. L., Qin, R. Y., Wang, N. N., Cao, Y., Gao, J., Gao, W. J., et al. (2012). Karyotype of asparagus by physical mapping of 45S and 5S rDNA by FISH. *J. Genet.* 91, 209–212. doi: 10.1007/s12041-012-0159-1

Dhar, M. K., Friebe, B., Koul, A. K., and Gill, B. S. (2002). Origin of an apparent B chromosome by mutation, chromosome fragmentation and specific DNA sequence amplification. *Chromosoma* 111, 332–340. doi: 10.1007/s00412-002-0214-4

Díaz Lifante, Z. (1996). A karyological study of *Asphodelus* L. (Asphodelaceae) from the Western Mediterranean. *Bot. J. Linn. Soc.* 121, 285–344. doi: 10.1111/j.1095-8339.1996.tb00760.x

Dobešová, E., Malinská, H., Matyášek, R., Leitch, A. R., Soltis, D. E., Soltis, P. S., et al. (2015). Silenced rRNA genes are activated and substitute for partially eliminated active homeologs in the recently formed allotetraploid, *Tragopogon mirus* (Asteraceae). *Heredity* 114, 356–365. doi: 10.1038/hdy.2014.111

Doveri, S., and Lee, D. (2007). Development of sensitive crop-specific polymerase chain reaction assays using 5S DNA: applications in food traceability. *J. Agric. Food Chem.* 55, 4640–4644. doi: 10.1021/jf063259v

Dubcovsky, J., and Dvorák, J. (1995). Ribosomal RNA multigene loci: nomads of the Triticeae genomes. *Genetics* 140, 1367–1377. doi: 10.1093/genetics/140.4.1367

Dvořáčková, M., and Fajkus, J. (2018). Visualization of the nucleolus using ethynyl uridine. *Front. Plant Sci.* 9:177. doi: 10.3389/fpls.2018.00177

Ehrendorfer, F., Schweizer, D., Greger, H., and Humphries, C. (1977). Chromosome banding and synthetic systematics in *Anacyclus* (Asteraceae Anthemideae). *Taxon* 26, 387–394. doi: 10.2307/1220037

Fankhauser, G., and Humphrey, R. R. (1943). The relation between number of nucleoli and number of chromosome sets in animal cells. *Proc. Natl. Acad. Sci. U. S. A.* 29, 344–350. doi: 10.2307/87559

Fehrer, J., Slavíková, R., Paštová, L., Josefiová, J., Mráz, P., Chrtek, J., et al. (2021). Molecular evolution and organization of ribosomal DNA in the hawkweed tribe Hieraciinae (Cichorieae, Asteraceae). *Front. Plant Sci.* 12:395. doi: 10.3389/fpls.2021.647375

Feliner, G. N., and Rosselló, J. A. (2012). "Concerted evolution of multigene families and homoeologous recombination," in *Plant Genome Diversity Volume 1*, ed. J. F. Wendel (Vienna: Springer), 171–193. doi: 10.1007/978-3-7091-1130-7_12

Fujisawa, M., Nakayama, S., Nishio, T., Fujishita, M., Hayashi, K., Ishizaki, K., et al. (2003). Evolution of ribosomal DNA unit on the X chromosome independent of autosomal units in the liverwort *Marchantia polymorpha*. *Chromosome Res.* 11, 695–703. doi: 10.1023/A:1025941206391

Galián, J. A., Rosato, M., and Rosselló, J. A. (2012). Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba*. *Heredity* 108, 640–646. doi: 10.1038/hdy.2012.2

Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., Ma, X., et al. (2010). Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *J. Ethnopharmacol.* 130, 116–121. doi: 10.1016/j.jep.2010.04.026

Gao, Y. D., Zhou, S. D., He, X. J., and Wan, J. (2012). Chromosome diversity and evolution in tribe Lilieae (Liliaceae) with emphasis on Chinese species. *J. Plant Res.* 125, 55–69. doi: 10.1007/s10265-011-0422-1

García, S., Gálvez, F., Gras, A., Kovaøík, A., and Garnatje, T. (2014). Plant rDNA database: update and new features. *Database* 2014:bau063. doi: 10.1093/database/bau063

García, S., Garnatje, T., and Kovaøík, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121, 389–394. doi: 10.1007/s00412-012-0368-7

García, S., Kovaøík, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Gates, R. R. (1942). Nucleoli and related nuclear structures. *Bot. Rev.* 8, 337–409. doi: 10.1007/BF02882158

Gemeinholzer, B., Oberprieler, C., and Bachmann, K. (2006). Using GenBank data for plant identification: possibilities and limitations using the ITS 1 of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. *Taxon* 55, 173–187. doi: 10.2307/25065539

González-Melendi, P., Ramírez, C., Testillano, P. S., Kumlehn, J., and Risueño, M. C. (2005). Three dimensional confocal and electron microscopy imaging define the dynamics and mechanisms of diploidisation at early stages of barley microspore-derived embryogenesis. *Planta* 222, 47–57. doi: 10.1007/s00425-005-1515-7

Goodpasture, C., and Bloom, S. E. (1975). Visualization of nucleolar organizer regions in mammalian chromosomes using silver staining. *Chromosoma* 53, 37–50. doi: 10.1007/bf00329389

Guerra, M. (2012). Cytotaxonomy: the end of childhood. *Plant Biosyst.* 146, 703–710. doi: 10.1080/11263504.2012.717973

Guimond, A., and Moss, T. (1999). A ribosomal orphon sequence from *Xenopus laevis* flanked by novel low copy number repetitive elements. *Biol. Chem.* 380, 167–174. doi: 10.1515/bc.1999.025

Gürdal, B., and Özhatay, N. (2018). Karyological study on 12 species of the genus *Taraxacum* (Asteraceae) grown in Turkey. *Flora* 28, 429–439. doi: 10.7320/FlMedit28.429

Hasterok, R., and Maluszynska, J. (2000). Nucleolar dominance does not occur in root tip cells of allotetraploid *Brassica* species. *Genome* 43, 574–579. doi: 10.1139/g00-005

Hayashi, K., and Matsunaga, S. (2019). Heat and chilling stress induce nucleolus morphological changes. *J. Plant Res.* 132, 395–403. doi: 10.1007/s10265-019-01096-9

Heitz, E. (1931). Die ursache der gesetzmässigen zahl, lage, form und grösse pflanzlicher nukleolen. *Planta* 12, 775–844. doi: 10.1007/BF01912443

Heslop-Harrison, J. S., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant J.* 66, 18–33. doi: 10.1111/j.1365-313X.2011.04544.x

Hidalgo, O., Vitales, D., Vallès, J., Garnatje, T., Siljak-Yakovlev, S., Leitch, I. J., et al. (2017). Cytogenetic insights into an oceanic island radiation: the dramatic evolution of pre-existing traits in *Cheirolophus* (Asteraceae: Cardueae: Centaureinae). *Taxon* 66, 146–157. doi: 10.12705/661.8

Houben, A., Banaei-Moghaddam, A. M., and Klemme, S. (2013). "Biology and evolution of B chromosomes," in *Plant Genome Diversity Volume 2*, ed. I. J. Leitch (Vienna: Springer), 149–165. doi: 10.1007/978-3-7091-1160-4

Houben, A., Jones, N., Martins, C., and Trifonov, V. (2019). Evolution, composition and regulation of supernumerary B chromosomes. *Genes* 10:161. doi: 10.3390/genes10020161

Idziak, D., and Hasterok, R. (2008). Cytogenetic evidence of nucleolar dominance in allotetraploid species of *Brachypodium*. *Genome* 51, 387–391. doi: 10.1139/G08-017

Jiang, J. (2019). Fluorescence in situ hybridization in plants: recent developments and future applications. *Chromosome Res.* 27, 153–165. doi: 10.1007/s00425-00018-03033-00424

Jiménez, R., Burgos, M., and de la Guardia, R. (1988). A study of the Ag-staining significance in mitotic NOR's. *Heredity* 60, 125–127. doi: 10.1038/hdy.1988.18

Jones, R. N. (1995). B chromosomes in plants. *New Phytol.* 131, 411–434. doi: 10.1111/j.1469-8137.1995.tb03079.x

Jones, R. N., Gonzalez-Sanchez, M., Gonzalez-García, M., Vega, J. M., and Puertas, M. J. (2008). Chromosomes with a life of their own. *Cytogenet. Genome Res.* 120, 265–280. doi: 10.1159/000121076

Khuong, N. T., and Schubert, I. (1985). Silver staining of nucleolus organizing regions in *Zea mays*. *Caryologia* 38, 331–334. doi: 10.1080/00087114.1985.10797757

Kokubugata, G., Hill, K. D., and Kondo, K. (2002). Ribosomal DNA distribution in somatic chromosomes of *Stangeria eriopus* (Stangeriaceae, Cycadales) and molecular-cytotaxonomic relationships to some other cycad genera. *Brittonia* 54, 1–5.

Krahulcová, A. (1993). New chromosome numbers in *Taraxacum* with reference to SAT-chromosomes. *Geobot. Phytotax.* 28, 289–294. doi: 10.1007/BF02853516

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374. doi: 10.1073/pnas.0503123102

Kumar, P., and Singhal, V. K. (2016). Nucleoli migration coupled with cytomixis. *Biologia* 71, 651–659. doi: 10.1515/biolog-2016-0076

Laane, M. M., and Höiland, K. (1986). Chromosome number and meiosis in herbarium specimens from the extinct Scandinavian population of *Crepis multicaulis*. *Hereditas* 105, 187–192. doi: 10.1111/j.1601-5223.1986.tb00660.x

Lacadena, J. R., Cermeno, M., Orellana, J., and Santos, J. L. (1984). Evidence for wheat-rye nucleolar competition (amphiplasty) in triticale by silver-staining procedure. *Theor. Appl. Genet.* 67, 207–213. doi: 10.1007/BF00317037

Lakshmi, N., and Raghavaiah, P. V. (1984). Accessory nucleoli in *Trigonella foenum-graecum* L. *Cytologia* 49, 401–405. doi: 10.1508/cytologia.49.401

Lan, T., Chen, R. Y., Li, X. L., Dong, F. P., Qi, Y. C., and Song, W. Q. (2008). Microdissection and painting of the W chromosome in *Ginkgo biloba* showed different labelling patterns. *Bot. Stud.* 49, 33–37.

Leitch, A. R. (2000). Higher levels of organization in the interphase nucleus of cycling and differentiated cells. *Microbiol. Mol. Biol. Rev.* 64, 138–152. doi: 10.1128/MMBR.64.1.138-152.2000

Leitch, A. R., and Heslop-Harrison, J. S. (1993). "Ribosomal RNA gene expression and localization in cereals," in *Chromosomes today Volume 11*, eds A. T. Sumner and A. C. Chandley (Dordrecht: Springer), 91–100. doi: 10.1007/978-94-011-1510-0

Levin, D. A. (1973). Accessory nucleoli in microsporocytes of hybrid *Phlox*. *Chromosoma* 41, 413–420. doi: 10.1007/BF00396499

Li, Z. Y., Fu, M. L., Hu, F. F., Huang, S. F., and Song, Y. C. (2006). Visualization of the ribosomal DNA (45S rDNA) of *indica* rice with FISH on some phases of cell cycle and extended DNA fibers. *Biocell* 30, 27–32. doi: 10.32604/biocell.2006.30.027

Lima-de-Faria, A. (1976). The chromosome field: I. Prediction of the location of ribosomal cistrons. *Hereditas* 83, 1–22. doi: 10.1111/j.1601-5223.1976.tb01565.x

Macas, J., Navrátilová, A., and Mészáros, T. (2003). Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma* 112, 152–158. doi: 10.1007/s00412-003-0255-3

Macháčková, P., Majeský, Ľ, Hroneš, M., Bílková, L., Hřibová, E., and Vašut, R. J. (2021). New insights into ribosomal DNA variation in apomictic and sexual *Taraxacum* (Asteraceae). *Bot. J. Linn.* 20, 1–26. doi: 10.1093/botlinnean/boab094

Maggini, F., Cremonini, R., Zolfino, C., Tucci, G. F., D'ovidio, R., Delre, V., et al. (1991). Structure and chromosomal localization of DNA sequences related to ribosomal subrepeats in *Vicia faba*. *Chromosoma* 100, 229–234. doi: 10.1007/BF00344156

Maluszynska, J., Hasterok, R., and Weiss, H. (1998). "rRNA genes-their distribution and activity in plants," in *Prace Naukowe Uniwersytetu Śląskiego w Katowicach*, ed. J. Małuszynska (Katowice: University of Silesia), 75–95.

Marques, A., Fuchs, J., Ma, L., Heckmann, S., Guerra, M., and Houben, A. (2011). Characterization of eu-and heterochromatin of *Citrus* with a focus on the condensation behavior of 45S rDNA chromatin. *Cytogenet. Genome Res.* 134, 72–82. doi: 10.1159/000323971

Marschner, S., Meister, A., Blattner, F. R., and Houben, A. (2007). Evolution and function of B chromosome 45S rDNA sequences in *Brachycome dichromosomatica*. *Genome* 50, 638–644. doi: 10.1139/G07-048

McClintock, B. (1934). The relationship of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Zeit. Zellforsch. Mik. Anat.* 21, 294–328. doi: 10.1007/BF00374060

Mehra, P. N., and Khitha, S. C. (1981). Karyotype and mechanism of sex determination in *Ephedra foliata* Boiss—a dioecious gymnosperm. *Cytologia* 46, 173–181. doi: 10.1508/cytologia.46.173

Milioto, V., Vlah, S., Mazzoleni, S., Rovatsos, M., and Dumas, F. (2019). Chromosomal localization of 18S-28S rDNA and $(TTAGGG)_n$ sequences in two South African dormice of the genus *Graphiurus* (Rodentia: Gliridae). *Cytogenet. Genome Res.* 158, 145–151. doi: 10.1159/000500985

Ming, R., Bendahmane, A., and Renner, S. S. (2011). Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* 62, 485–514. doi: 10.1146/annurev-arplant-042110-103914

Mishima, M., Ohmido, N., Fukui, K., and Yahara, T. (2002). Trends in site-number change of rDNA loci during polyploid evolution in *Sanguisorba* (Rosaceae). *Chromosoma* 110, 550–558. doi: 10.1007/s00412-001-0175-z

Mogie, M., and Richards, A. J. (1983). Satellited chromosomes, systematics and phylogeny in *Taraxacum* (Asteraceae). *Plant Syst. Evol.* 141, 219–229. doi: 10.1007/BF00989003

Müller, T., Philippi, N., Dandekar, T., Schultz, J., and Wolf, M. (2007). Distinguishing species. *RNA* 13, 1469–1472. doi: 10.1261/rna.617107

Mursalimov, S., and Deineko, E. (2018). Cytomixis in plants: facts and doubts. *Protoplasma* 255, 719–731. doi: 10.1007/s00709-017-1188-7

Mursalimov, S. R., and Deineko, E. V. (2011). An ultrastructural study of cytomixis in tobacco pollen mother cells. *Protoplasma* 248, 717–724. doi: 10.1007/s00709-010-0234-5

Nakao, Y., Taira, T., Horiuchi, S., Kawase, K., and Mukai, Y. (2005). Chromosomal difference between male and female trees of *Ginkgo biloba* examined by karyotype analysis and mapping of rDNA on the chromosomes by Fluorescence in situ Hybridization. *J. Jpn. Soc. Hortic. Sci.* 74, 275–280. doi: 10.2503/jjshs.74.275

Nakayama, S., Fujishita, M., Sone, T., and Ohyama, K. (2001). Additional locus of rDNA sequence specific to the X chromosome of the liverwort, *Marchantia polymorpha*. *Chromosome Res.* 9, 469–473. doi: 10.1023/A:1011676328165

Navashin, M. (1934). Chromosome alterations caused by hybridisation and their bearing upon certain general genetic problems. *Cytologia* 5, 169–203. doi: 10.1508/cytologia.5.169

Newton, M. E. (1986). *Pellia borealis* Lorbeer: its cytological status and discovery in Britain. *J. Bryol.* 14, 215–230. doi: 10.1179/jbr.1986.14.2.215

Newton, M. E. (1988). Cytological diversity in *Pellia endiviifolia* (Dicks.) Dum. *J. Bryol.* 15, 303–314. doi: 10.1179/jbr.1988.15.2.303

Nieto Feliner, G., and Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919. doi: 10.1016/j.ympev.2007.01.013

Ochatt, S. J., and Seguí-Simarro, J. M. (2021). "Analysis of ploidy in haploids and doubled haploids," in *Doubled Haploid Technology*, ed. J. M. Seguí-Simarro (New York: Humana Press), 105–125. doi: 10.1007/978-1-0716-1315-3_4

Ojha, S., Malla, S., and Lyons, S. M. (2020). snoRNPs: functions in ribosome biogenesis. *Biomolecules* 10:783. doi: 10.3390/biom10050783

Pikaard, C. S. (2000a). The epigenetics of nucleolar dominance. *Trends Genet.* 16, 495–500. doi: 10.1016/S0168-9525(00)02113-2

Pikaard, C. S. (2000b). Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids. *Plant Mol. Biol.* 43, 163–177. doi: 10.1023/A:1006471009225

Poczai, P., and Hyvönen, J. (2010). Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol. Biol. Rep.* 37, 1897–1912. doi: 10.1007/s11033-009-9630-3

Pokorná, M. J., and Reifová, R. (2021). Evolution of B Chromosomes: from dispensable parasitic chromosomes to essential genomic players. *Front. Genet.* 12:727570. doi: 10.3389/fgene.2021.727570

Ran, Y., Hammett, K. R., and Murray, B. G. (2001). Phylogenetic analysis and karyotype evolution in the genus *Clivia* (Amaryllidaceae). *Ann. Bot.* 87, 823–830. doi: 10.1006/anbo.2001.1422

Raskina, O., Belyayev, A., and Nevo, E. (2004). Activity of the En/Spm-like transposons in meiosis as a base for chromosome repatterning in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. *Chromosome Res.* 12, 153–161. doi: 10.1023/B:CHRO.0000013168.61359.43

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428. doi: 10.1016/j.molp.2018.01.002

Richards, A. J. (1989). A comparison of within-plant karyological heterogeneity between agamospermous and sexual *Taraxacum* (Compositae) as assessed by the nucleolar organiser chromosome. *Plant Syst. Evol.* 163, 177–185. doi: 10.1007/BF00936513

Roa, F., and Guerra, M. (2012). Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evol. Biol.* 12:225. doi: 10.1186/1471-2148-12-225

Rosato, M., Álvarez, I., Nieto Feliner, G., and Rosselló, J. A. (2017). High and uneven levels of 45S rDNA site-number variation across wild populations of a diploid plant genus (*Anacyclus*, *Asteraceae*). *PLoS One* 12:e0187131. doi: 10.1371/journal.pone.0187131

Rosato, M., Castro, M., and Rosselló, J. A. (2008). Relationships of the woody *Medicago* species (section *Dendrotelis*) assessed by molecular cytogenetic analyses. *Ann. Bot.* 102, 15–22. doi: 10.1093/aob/mcn055

Rosato, M., Kovaøík, A., Garilleti, R., and Rosselló, J. A. (2016a). Conserved organisation of 45S rDNA sites and rDNA gene copy number among major clades of early land plants. *PLoS One* 11:e0162544. doi: 10.1371/journal.pone.0162544

Rosato, M., Ferrer-Gallego, P., Totta, C., Laguna, E., and Rosselló, J. A. (2016b). Nuclear rDNA instability in in vitro-generated plants is amplified after sexual reproduction with conspecific wild individuals. *Bot. J. Linn. Soc.* 181, 127–137. doi: 10.1111/boj.12392

Rosato, M., Moreno-Saiz, J. C., Galián, J. A., and Rossello, J. A. (2015). Evolutionary site-number changes of ribosomal DNA loci during speciation: complex scenarios of ancestral and more recent polyploid events. *AoB Plants* 7:lv135. doi: 10.1093/aobpla/plv135

Rosato, M., and Rosselló, J. A. (2009). Karyological observations in *Medicago* section *Dendrotelis* (Fabaceae). *Folia Geobot.* 44, 423–433. doi: 10.1007/s12224-009-9048-7

Rosselló, J. A., and Castro, M. (2008). Karyological evolution of the angiosperm endemic flora of the Balearic Islands. *Taxon* 57, 259–273. doi: 10.2307/25065967

Sadder, M. T., and Weber, G. (2001). Karyotype of maize (*Zea mays* L.) mitotic metaphase chromosomes as revealed by fluorescence *in situ* hybridization (FISH) with cytogenetic DNA markers. *Plant Mol. Biol. Rep.* 19, 117–123. doi: 10.1007/BF02772153

Sangduen, N., Toahsakul, M., and Hongtrakul, V. (2007). Karyomorphological study of some selected Cycads. *Au J. T.* 11, 1–6.

Sangduen, N., Toahsakul, M., and Hongtrakul, V. (2009). Comparative karyomorphological study between male and female plants of some *Cycas* and *Zamia* species. *Nat. Resour. J.* 43, 476–485.

Sato, S., Matsumoto, E., and Kuroki, Y. (1981). Satellite association of the nucleolar chromosomes in a plant. *Protoplasma* 108, 139–147. doi: 10.1007/BF01276888

Scaldaferro, M. A., da Cruz, M. V. R., Cecchini, N. M., and Moscone, E. A. (2016). FISH and AgNor mapping of the 45S and 5S rRNA genes in wild and cultivated species of *Capsicum* (Solananceae). *Genome* 59, 95–113. doi: 10.1139/gen-2015-0099

Schubert, I. (1984). Mobile nucleolus organizing regions (NORs) in *Allium* (Liliaceae s. lat.)? —Inferences from the specifity of silver staining. *Plant Syst. Evol.* 144, 291–305. doi: 10.1007/BF00984139

Schubert, I., and Wobus, U. (1985). In situ hybridization confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* 92, 143–148. doi: 10.1007/BF00328466

Schwarzacher, H. G., and Wachtler, F. (1983). Nucleolus organizer regions and nucleoli. *Hum. Genet.* 63, 89–99. doi: 10.1007/BF00291525

Schweizer, D., and Ehrendorfer, F. (1976). Giemsa banded karyotypes, systematics, and evolution in *Anacyclus* (Asteraceae-Anthemideae). *Plant Syst. Evol.* 126, 107–148. doi: 10.1007/BF00981668

Sepsi, A., Fábián, A., Jäger, K., Heslop-Harrison, J. S., and Schwarzacher, T. (2018). ImmunoFISH: simultaneous visualisation of proteins and DNA sequences gives insight into meiotic processes in nuclei of grasses. *Front. Plant Sci.* 9:1193. doi: 10.3389/fpls.2018.01193

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Sochorová, J., Coriton, O., Kuderová, A., Lunerová, J., Chevre, A. M., and Kovaøík, A. (2017). Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of *Brassica napus*. *Ann. Bot.* 119, 13–26. doi: 10.1093/aob/mcw187

Sýkorová, E., Lim, K. Y., Fajkus, J., and Leitch, A. R. (2003). The signature of the *Cestrum* genome suggests an evolutionary response to the loss of $(TTTAGGG)_n$ telomeres. *Chromosoma* 112, 164–172. doi: 10.1007/s00412-003-0256-2

Tapia-Pastrana, F. (2020). Differential amphiplasty and nucleolar dominance in somatic metaphase cells as evidence of hybridization in *Prosopis juliflora* (Leguminosae, Mimosoideae). *Cytologia* 85, 295–299. doi: 10.1508/cytologia.85.295

Terasaka, O., and Tanaka, R. (1974). Absence of the nucleolar constriction in the division of generative nucleus in the pollen of some angiosperms. *Cytologia* 39, 97–106. doi: 10.1508/cytologia.39.97

Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. *J. Cell Sci.* 126, 4815–4821. doi: 10.1242/jcs.111948

Totta, C., Rosato, M., Ferrer-Gallego, P., Lucchese, F., and Rosselló, J. A. (2017). Temporal frames of 45S rDNA site-number variation in diploid plant lineages: lessons from the rock rose genus *Cistus* (Cistaceae). *Biol. J. Linn. Soc.* 120, 626–636. doi: 10.1111/bij.12909

Trerè, D. (2000). AgNOR staining and quantification. *Micron* 31, 127–131. doi: 10.1016/S0968-4328(99)00069-4

Tucker, S., Vitins, A., and Pikaard, C. S. (2010). Nucleolar dominance and ribosomal RNA gene silencing. *Curr. Opin. Cell Biol.* 22, 351–356. doi: 10.1016/j.ceb.2010.03.009

Venkatesh, K. H., Pragathi, G. S., and Shivashankar, M. (2019). Meiotic behavior and chromosomal association of two diploid mulberry varieties, (Moraceae). *Chromosome Bot.* 13, 68–70. doi: 10.3199/iscb.13.68

Verma, R. C., and Raina, S. N. (1981). Cytogenetics of *Crotalaria* V. Supernumerary nucleoli in *C. agatiflora* (Leguminosae). *Genetica* 56, 75–80. doi: 10.1007/BF00126933

Vitales, D., D'Ambrosio, U., Gálvez, F., Kovaøík, A., and García, S. (2017). Third release of the plant rDNA database with updated content and information on telomere composition and sequenced plant genomes. *Plant Syst. Evol.* 303, 1115–1121. doi: 10.1007/s00606-017-1440-9

Watanabe, K., Shibaike, H., Suzuki, T., Ito, M., and Hoya, A. (2021). DNA Contents and karyotypes of the natural hybrids in *Taraxacum* (Asteraceae) in Japan. *Acta Phytotaxon. Geobot.* 72, 135–144. doi: 10.18942/apg.202013

Weiss-Schneeweiss, H., Emadzade, K., Jang, T. S., and Schneeweiss, G. M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenet. Genome Res.* 140, 137–150. doi: 10.1159/000351727

Weiss-Schneeweiss, H., and Schneeweiss, G. M. (2013). "Karyotype diversity and evolutionary trends in angiosperms," in *Plant Genome Diversity Volume 2*, ed. I. J. Leitch (London: Springer), 209–230. doi: 10.1007/978-3-7091-1160-4_13

Weiss-Schneeweiss, H., Tremetsberger, K., Schneeweiss, G. M., Parker, J. S., and Stuessy, T. F. (2008). Karyotype diversification and evolution in diploid and polyploid South American *Hypochaeris* (Asteraceae) inferred from rDNA localization and genetic fingerprint data. *Ann. Bot.* 101, 909–918. doi: 10.1093/aob/mcn023

Xu, J., and Earle, E. D. (1996). High resolution physical mapping of 45S (5.8 S, 18S and 25S) rDNA gene loci in the tomato genome using a combination of karyotyping and FISH of pachytene chromosomes. *Chromosoma* 104, 545–550. doi: 10.1007/BF00352294

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpls.2022.819750

The Ribosomal DNA Loci of the Ancient Monocot *Pistia stratiotes* L. (Araceae) Contain Different Variants of the 35S and 5S Ribosomal RNA Gene Units

Anton Stepanenko[1†‡], Guimin Chen[1‡], Phuong T. N. Hoang[2,3†‡], Jörg Fuchs[2†], Ingo Schubert[2†] and Nikolai Borisjuk[1†]*

[1] Jiangsu Key Laboratory for Eco-Agricultural Biotechnology Around Hongze Lake and Jiangsu Collaborative Innovation Centre of Regional Modern Agriculture and Environmental Protection, School of Life Sciences, Huaiyin Normal University, Huai'an, China, [2] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, [3] Faculty of Biology, Dalat University, Đà Lạt, Vietnam

The freshwater plant water lettuce (*Pistia stratiotes* L.) grows in warm climatic zones and is used for phytoremediation and biomass production. *P. stratiotes* belongs to the Araceae, an ecologically and structurally diverse early monocot family, but the phylogenetic relationships among Araceae members are poorly understood. Ribosomal DNAs (rDNAs), including the 35S and 5S rDNA, encode the RNA components of ribosomes and are widely used in phylogenetic and evolutionary studies of various plant taxa. Here, we comprehensively characterized the chromosomal locations and molecular organization of 35S and 5S rDNA genes in water lettuce using karyological and molecular methods. Fluorescence *in situ* hybridization revealed a single location for the 35S and 5S rDNA loci, each on a different pair of the species' 28 chromosomes. Molecular cloning and nucleotide sequencing of 35S rDNA of *P. stratiotes*, the first representative Araceae *sensu stricto* in which such a study was performed, displayed typical structural characteristics. The full-length repeat showed high sequence conservation of the regions producing the 18S, 5.8S, and 25S rRNAs and divergence of the internal transcribed spacers ITS1 and ITS2 as well as the large intergenic spacer (IGS). Alignments of the deduced sequence of 18S rDNA with the sequences available for other Araceae and representatives of other clades were used for phylogenetic analysis. Examination of 11 IGS sequences revealed significant intra-genomic length variability due to variation in subrepeat number, with four types of units detected within the 35S rDNA locus of the *P. stratiotes* genome (estimated size 407 Mb/1C). Similarly, the 5S rDNA locus harbors gene units comprising a conserved 119-bp sequence encoding 5S rRNA and two types of non-transcribed spacer (NTS) sequences. Type I was classified into four subtypes, which apparently originated via progressive loss of subrepeats within the duplicated NTS region containing the 3' part of the 5S rRNA

OPEN ACCESS

Edited by:
Timothy E. Audas, Simon Fraser University, Canada

Reviewed by:
Leonor Morais-Cecilio, University of Lisbon, Portugal
Jana Lunerova, Academy of Sciences of the Czech Republic, Czechia

****Correspondence:***
Nikolai Borisjuk
nborisjuk@hytc.edu.cn

†ORCID:
Anton Stepanenko
orcid.org/0000-0003-2326-6613
Phuong T. N. Hoang
orcid.org/0000-0002-7418-9091
Jörg Fuchs
orcid.org/0000-0003-4171-5371
Ingo Schubert
orcid.org/0000-0002-6300-2068
Nikolai Borisjuk
orcid.org/0000-0001-5250-9771

‡These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *22 November 2021*
Accepted: *10 February 2022*
Published: *03 March 2022*

Citation:
Stepanenko A, Chen G, Hoang PTN, Fuchs J, Schubert I and Borisjuk N (2022) The Ribosomal DNA Loci of the Ancient Monocot Pistia stratiotes L. (Araceae) Contain Different Variants of the 35S and 5S Ribosomal RNA Gene Units. Front. Plant Sci. 13:819750. doi: 10.3389/fpls.2022.819750

gene. The minor Type II NTS is shorter than Type I and differs in nucleotide composition. Some DNA clones containing two or three consecutive 5S rDNA repeats harbored 5S rDNA genes with different types of NTSs, confirming the mosaic composition of the 5S rDNA locus.

Keywords: *Pistia stratiotes*, FISH, gene organization, molecular evolution, 35S rDNA, 5S rDNA

INTRODUCTION

Water lettuce (*Pistia stratiotes*) belongs to a monospecific genus in the subfamily Aroideae of the ecologically and structurally diverse ancient monocot family Araceae *sensu lato* (also known as aroids), a group of 118 genera comprising approximately 3,800 species (Henriquez et al., 2014). Most aroids are tropical and subtropical species, but some members inhabit temperate regions, displaying broad habitat diversity, including geophytes, epiphytes, helophytes, climbers, and floating aquatics (Croat, 1988; Mayo et al., 1998; Keating, 2004). Bayesian analysis of divergence times based on multiple fossil and geological calibration points revealed that the *Pistia* lineage is 76–90 million years old (Renner and Zhang, 2004).

Water lettuce, which floats in fresh water, displays rapid, mostly vegetative propagation, and high biomass accumulation. In many locations, these features qualify *P. stratiotes* as an invasive species that is difficult to eliminate (Paolacci et al., 2018). However, *P. stratiotes* plants have tremendous potential for water bioremediation due to their capacity for fast and efficient assimilation of nitrogen and phosphate, heavy metals, and other water contaminants. Therefore, *P. stratiotes*, like the aquatic duckweeds (Acosta et al., 2021), has been used to remediate different types of wastewater (Zimmels et al., 2006; Rezania et al., 2016; Zhou and Borisjuk, 2019). For example, Lu et al. (2010) determined that water lettuce was superior to most other plants for efficient wastewater bioremediation due to its capability to annually remove 190–329 kg/ha of nitrogen and 25–34 kg/ha of phosphorus. Additionally, its high amounts of proteins and carbohydrates make *P. stratiotes* a valuable biomass resource for use as a green fertilizer or soil amendment (Kodituwakku and Yatawara, 2020) or as feedstock for the production of nitrogen-doped biochar (Zhang et al., 2021).

Ribosomal DNA (rDNA) plays a pivotal role in organisms by producing the RNA components required to form ribosomes (Moss and Stefanovsky, 2002; Appels et al., 2021; Hemleben et al., 2021). In plants, as in most eukaryotes, the rDNA encodes four ribosomal RNAs (rRNAs), which serve as the major structural and functional components of the ribosome. Plant rRNA genes typically occur in two types of loci: 35S rDNA loci containing three tightly linked rRNA genes (18S-5.8S-25S), which are transcribed by RNA Polymerase I into a 35S rRNA precursor; and 5S rDNA loci encoding 5S rRNA transcribed by RNA Polymerase III. The 35S and 5S rDNA loci have clusters of tandemly repeated units composed of conserved coding sequences and diverse intergenic spacers (IGSs) (Volkov et al., 2003). Due to its high copy number, its conserved coding sequence, and its more rapidly evolving spacer sequences, rDNA has become a favorite subject of studies related to plant systematics, evolution, and biodiversity and is used as a genome-specific marker in allopolyploids and hybrids (Borisjuk et al., 1988; Stadler et al., 1995; Mahelka et al., 2017).

To date, the DNA sequence data for the Araceae family have mostly been obtained from chloroplasts and mitochondria (Renner et al., 2004; Rothwell et al., 2004; Cusimano et al., 2011; Henriquez et al., 2014; Choi et al., 2017; Gao et al., 2018; Tian et al., 2018). With the exception of the whole-genome sequences of five species of the remotely related duckweeds (Acosta et al., 2021), little information is available about the nuclear genes of Araceae species, including their rDNA. The nuclear genomes of several Araceae species have been studied by examining the 35S and 5S rDNA loci using fluorescence *in situ* hybridization (FISH) (Sousa et al., 2014; Lakshmanan et al., 2015; Sousa and Renner, 2015; Vasconcelos et al., 2018), and a few nuclear rDNA sequences, represented by a single 25S rDNA sequence for *Spathiphyllum wallisii* (Zanis et al., 2003) and several sequences for 18S rDNA and ITS1-5.8S-ITS2, have been deposited in GenBank.

In this study, to gain a deeper understanding of the molecular organization and functionality of plant rDNAs and of the phylogenetic relationships and evolutionary history of the Araceae, we sequenced and examined the chromosomal localizations of *Pistia stratiotes* rDNAs. The obtained data include the entire nucleotide sequence of 11 35S rDNA repeat units (18S-ITS1-5.8S-ITS2-25S rDNA and the IGS) and sequences of 63 clones representing the variability of the 5S rDNA units in the *P. stratiotes* genome.

MATERIALS AND METHODS

Plant Material

A *Pistia stratiotes* plant, designated as isolate TB-1, was purchased from an online seller (taobao.cn) and cultivated in fresh water under laboratory conditions for molecular and cytological analysis. The species' identity was confirmed by DNA barcoding using primers specific for chloroplast intergenic spacers atpF-atpH (ATP) and psbK-psbL (PSB) as previously described (Borisjuk et al., 2015).

Genome Size Measurement

For flow cytometric genome size measurements roughly 0.5 cm^2 of fresh leaf tissue of *P. stratiotes* and *Raphanus sativus* cv. Voran (2C = 1.11 pg; Genebank Gatersleben, accession number: RA 34) as internal reference standard were co-chopped with a sharp razorblade in a Petri dish using the "CyStain PI Absolute P" reagent kit (Sysmex-Partec) according to manufacturers' instructions. The samples were filtered through a 50 μm mesh

and measured on a CyFlow Space flow cytometer (Sysmex-Partec). The DNA content (pg/2C) was calculated based on the values of the G1 peak means and the corresponding genome size (Mbp/1C), according to Dolezel et al. (2003).

Mitotic Chromosome Preparation

The plants were grown in nutrient solution (Appenroth et al., 1996) until daughter plants with new roots had developed. The root tips were collected and treated in 2 mM 8-hydroxylquinoline at 37°C for 2 h and then fixed in fresh 3:1 (absolute ethanol: acetic acid) for 48 h. The fixed samples were washed twice in 10 mM Na-citrate buffer pH 4.6 for 10 min each before and after softening in 2 mL PC enzyme mixture (1% pectinase and 1% cellulase in sodium-citrate buffer) for 120 min at 37°C, prior to maceration and squashing in 60% acetic acid. After freezing in liquid nitrogen, the slides were treated with pepsin, (50 μg pepsin/mL in 0.01 N HCl, 5 min at 37°C), post-fixed in 4% formaldehyde in 2x SSC (300 mM Na-citrate, 30 mM NaCl, pH 7.0) for 10 min, rinsed twice in 2x SSC, 5 min each, dehydrated in an ethanol series (70, 90, and 96%, 2 min each) and air-dried.

Ribosomal DNA Probe Labeling

The *A. thaliana* BAC clone T15P10 (Arabidopsis Biological Resource Center, United States), labeled by nick-translation was used as 35S rDNA probe. Genomic DNA of the giant duckweed (*Spirodela polyrhiza*) was used for PCR-amplification of 5S rDNA with a primer pair listed in Hoang et al. (2019) and designed according to the 5S rDNA sequence of *Glycine max* (Gottlob-McHugh et al., 1990). The PCR product was used as template for PCR-labeling to generate the 5S rDNA FISH probe.

The 5S rDNA probe was labeled with Cy3-dUTP (GE Healthcare Life Science), and the 35S rDNA probe with Texas Red-12-dUTP (Life Technologies) and precipitated as described (Hoang and Schubert, 2017).

Fluorescence *in situ* Hybridization

Probes were denatured at 95°C for 5 min and chilled on ice for 10 min before adding 10 μL of each probe per slide. Then, the mitotic chromosome preparations were denatured together with the probes on a heating plate at 80°C for 3 min, followed by incubation in a moist chamber at 37°C for at least 16 h. Post-hybridization washing and signal detection were done as described (Lysak et al., 2006) with minor modifications. Widefield fluorescence microscopy for signal detection followed Cao et al. (2016). The images were pseudo-colored and merged using Adobe Photoshop software ver.12 (Adobe Systems).

Cloning and Sequence Analysis of 35S Ribosomal DNA

For analysis of rDNA genes, total DNA was isolated from the fresh biomass of *Pistia stratiotes* using the CTAB method (Murray and Thompson, 1980) modified according to Borisjuk et al. (2015). To clone the 35S rDNA, genomic DNA of *P. stratiotes* was digested with *Xba*I + *Mfe*1 restriction enzymes (Takara, China) and fractionated by agarose gel electrophoresis. DNA fragments of 2–6 kb were purified from the gel using AxyPrep™ DNA Gel Extraction Kit (Axygen, United States) and ligated into pUC18 plasmid, digested with *Xba*I + *EcoR*1. Following transformation of the plasmids into *E. coli*, about 230 of the obtained colonies were screened by PCR using three sets of primers (**Supplementary Figure 1**). One set of primers, specific for internal part of 18S rRNA gene was used to select clones containing the 18S-5.8S-25S rDNA; two additional pairs of primers, one specific for the 3′-end of 25S rRNA gene (Clo25Sfor and Clo25Srev), and one specific for the 5′-end of 18S rRNA gene (Clo18Sfor and Clo18Srev) were used for selecting clones containing the end of 25S rDNA, the intergenic spacer (IGS) and the 5′part of the 18S rDNA. The colony PCR screening resulted in selecting two clones containing the coding rDNA portion, Pi-rDNA-1 and Pi-rDNA-2, and one clone with the IGS, Pi-IGS-1. The isolated plasmids were custom sequenced by Sangon Biotech (Shanghai, China) using the combination of standard forward and reverse pUC18 primers and a number of the insert internal primers designed according the progress of the sequencing (**Supplementary Table 1**).

Based on the obtained sequence of Pi-IGS-1, primers specific for 18S and 25S rRNA genes were used to amplify the IGS region. The generated DNA fragments, were cloned into the vector pMD19 (Takara, Dalian, China) and the fragments of selected clones were checked by digestion using restriction enzymes *Eco*RI + *Hin*dIII and *Eco*RI + *Pst*I (Takara, Dalian, China). Ten clones with rDNA fragments of different length were custom sequenced (Sangon Biotech, Shanghai, China) using a combination of standard forward and reverse sequencing primers and a range of internal primers designed according the original sequence of the Pi-IGS-1 (**Supplementary Table 1**). The obtained nucleotide sequences were analyzed using the CLC Main Workbench (Version 6.9.2, Qiagen) software. The resulted sequences of the *P. stratiotes* rDNA fragments are deposited in the GenBank.

Cloning and Sequence Characterization of 5S Ribosomal RNA Genes

For analysis of *P. stratiotes* 5S rRNA genes, the specific DNA fragments were amplified from genomic DNA by PCR using two pairs of primers specific for 5S rRNA gene sequence DW-5S-F/DW-5S-R and cn-5S-for/cn-5S-rev (**Supplementary Table 1**) as previously described (Chen et al., 2021). In order to increase the chance of amplifying DNA fragments with multiple 5S rDNA units, the elongation time of PCR was prolonged to 1 min 30 s. The generated DNA fragments, cloned into the vector pMD19 (Takara, Dalian, China) were custom sequenced (Sangon Biotech, Shanghai, China), and the obtained nucleotide sequences were analyzed using the CLC Main Workbench (Version 6.9.2, Qiagen) software. Upon sequence analysis of the obtained clones, a second round of PCR amplification using primers specific for the revealed NTS sequences was performed. The primers used for amplification of 5S rDNA units are listed in **Supplementary Table 1**.

In silico Analysis of the Ribosomal DNA Sequences

The obtained *P. stratiotes* sequences for 18S rRNA were used to examine the phylogenetic relationships primarily within the Araceae family, but also with the representatives of

other plant clades. The corresponding sequences available for Araceae and representative species for monocots, magnoliids and eudicots were extracted from the GenBank[1] by blasting with the *P. stratiotes* sequences as a query, with cut off *E*-value equal to 0.05. The maximum-likelihood phylogenetic trees were constructed using NGPhylogeny webservice accessible through the https://ngphylogeny.fr using MAFFT Multiple Sequence Alignment and FastME algorithm (Lefort et al., 2015; Lemoine et al., 2019). iTOL[2] was used for displaying and annotating the generated phylogenetic trees (Letunic and Bork, 2021).

For detection of the DNA regions likely to fold into G-quadruplex structures, we have primarily used the *pqsfinder* prediction tool (Labudová et al., 2020) available at the website,[3] with further verification by the G4Hunter algorithm (Brázda et al., 2019), freely available at DNA Analyzer server.[4]

[1]https://www.ncbi.nlm.nih.gov
[2]https://itol.embl.de
[3]https://pqsfinder.fi.muni.cz/
[4]http://bioinformatics.ibp.cz

RESULTS

Characterization of the Chromosome Set and Visualization of 35S and 5S Ribosomal DNA Loci in *Pistia stratiotes*

The identity of the *Pistis stratiotes* TB-1 isolate used in this study (**Figure 1A**) was confirmed by examining the chloroplast DNA barcodes ATP (intergenic spacers atpF-atpH, Ac# OL435916) and PSB (intergenic spacers psbK-psbL, Ac# OL435917). Alignment of the obtained sequences showed that these sequences shared 100 and 99.6% similarity, respectively, with the corresponding ATP and PSB spacers of the *P. stratiotes* chloroplast genome deposited in GenBank (accession no. NC_048522).

The root tip meristems of *P. stratiotes* displayed a number of small chromosomes (2n = 28) after DAPI staining (**Figure 1B**). Due to their small size, structural details, such as primary (centromere) or secondary [nucleolus-organizing region (NOR)] constrictions, of these chromosomes were barely recognizable. Therefore, whether the chromosomes of this species are mono- or holocentric remains unknown. FISH with 5S and 35S rDNA

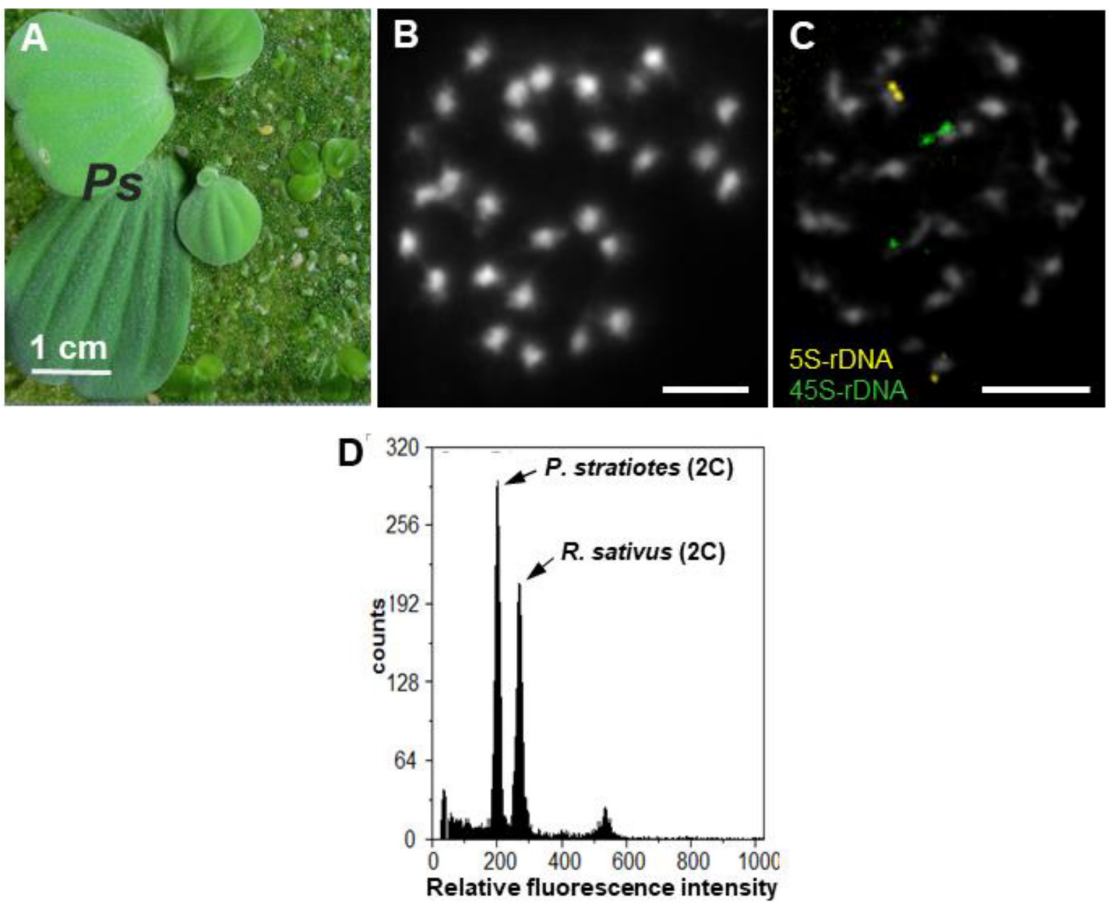


FIGURE 1 | *P. statiotes*, its chromosomes and nuclear DNA content. **(A)** Whole *P. stratiotes* plant (***Ps***) in the presence of different duckweed species. **(B)** Complete meristematic metaphase, with 28 small DAPI-stained chromosomes (scale bar: 5 μm). **(C)** FISH signals for 5S (yellow) and 35S rDNA (green) loci at one end of one pair of chromosomes each (scale bar: 5 μm). **(D)** Histogram of nuclear DNA of *P. stratiotes,* and *Raphanus sativus* as an internal reference standard.

probes uncovered one pair of chromosomes, each with terminal signals (**Figure 1C**), supporting the diploid nature of the species. Flow cytometric measurements of isolated nuclei (**Figure 1D**) revealed a nuclear genome size of 407 Mbp/1C (unreplicated haploid chromosome complement).

Nucleotide Sequence Analysis of 35S Ribosomal DNA

DNA Sequences Encoding 18S-5.8S-25S Ribosomal RNA

Analysis of the sequences of various duckweeds species, the only group of related aquatic plants with relatively well-characterized 18S and 25S rDNA genes (Tippery et al., 2015), and our restriction mapping of 35S rDNA in *S. polyrhiza* (Michael et al., 2017) revealed unique conservative restriction sites for *Xba*I in the 18S rDNA and for Mfe1 in the 25S rDNA. Our strategy to clone the entire 35S rDNA repeat unit was built on this finding, assuming a similar situation for *Pistia*. We characterized the entire 35S rDNA repeat sequence by sequencing three cloned genomic Mfe1 + *Xba*I restriction fragments (**Supplementary Figure 1**). Sequence comparison of the two clones, Pi-rDNA-1 and Pi-rDNA-2 (**Supplementary Figure 2**), containing part of the 35S rDNA repeat encoding 18S-5.8S-25S ribosomal RNA, revealed high nucleotide conservation with just eight single nucleotide polymorphisms (SNPs) over a length of 5,365 bp. Six of the detected SNPs were T↔C transitions, one was a C↔G transversion, and one was a nucleotide deletion located in ITS1, 5.8S, and 25S rRNA coding genes, with no variations in the sequences for 18S rDNA or ITS2. Due to this low sequence divergence, we used the 18S and 25S rDNA sequences of the clone Pi-rDNA-1, supplemented with the missing parts of the 5′-end of the 18S rDNA and the 3′-end of the 25S rDNA from clone Pi-IGS (**Supplementary Figure 1**), resulting in a 5,868-bp-long sequence covering the 18S-ITS1-5.8S-ITS2-25S rDNA for further analysis.

BLAST analysis showed that the *P. stratiotes* 18S rDNA sequence has the highest similarity (99.5%) to a previously sequenced but not published *Pistia* gene (accession no. AF168869), 96.98% similarity to 18S rDNA of the Araceae species *Orontium aquaticum* (Qiu et al., 2000), 97.5% to *Calla palustris* (accession no. AF168829), 96.99% to *Spathiphyllum wallisii* (accession no. AF207023), 96.14% to *Gymnostachys anceps* (accession no. AF069200), 97% to *Symplocarpus nipponicus* (accession no. MT247907) (Do et al., 2020), and 95.31–96.58% to the 18S rDNA sequences of the 36 duckweed species (Tippery et al., 2015; Hoang et al., 2020), whose 18S rDNA sequences are available in GenBank. The 18S rDNA-based phylogenetic tree is shown in **Supplementary Figure 3**. The *P. stratiotes* 5.8S rDNA sequence showed many more BLAST hits in the Araceae family than the 18S rDNA. The hit with the highest score (98.16%) was for the *Amorphophallus elliottii* gene (accession no. KR534451), followed by *Lasia spinosa* (Yeng et al., 2016) and numerous duckweed species (Tippery et al., 2015). When the query included ITS1 and/or ITS2 in addition to the 5.8S rDNA sequence, the list of meaningful BLAST hits primarily included species representing the *Schismatoglottidoideae* clade (genera *Aridarum*, *Bakoa*, *Hottarum*, *Ooia*, and *Piptospatha*) and the genus *Amorphophallus*. According to the advanced phylogenetics based on chloroplast and mitochondria DNA data (Henriquez et al., 2014), the *Schismatoglottidoideae* and *Amorphophallus* together with *Pistia* belong to the Aroidae subfamily of the Araceae. BLAST analysis of the 25S rDNA sequence of *P. stratiotes* primarily revealed the homologous sequences of duckweeds deposited by Tippery et al. (2015) and the 25S rDNA sequence of *Spathiphyllum wallisii* (Zanis et al., 2003), demonstrating the scarce representation of Araceae in the GenBank database.

Sequence Organization of the 35S Ribosomal DNA Intergenic Spacer Region of *Pistia stratiotes*

Sequencing of the Pi-IGS clone (**Supplementary Figures 1, 4**) revealed an IGS region of 2,644 bp with many features of molecular architecture previously described for other plants (Appels and Dvořák, 1982; Delcasso-Tremousaygue et al., 1988; Borisjuk and Hemleben, 1993; Borisjuk et al., 1997), as well as the classic plant rRNA transcription initiation site (TIS) signature TATAGGGGG located in the middle of the IGS, 1,270 bp upstream of the 18S rRNA gene. The *P. stratiotes* IGS has a relatively high average overall GC content of 61.7%, with an approximately equal percentage of AT and GC within the first half of the sequence and an irregular GC pattern in the second part of the IGS, reaching more than 75% close to the beginning of the 18S rRNA gene (**Figure 2A**). This GC-enriched part of the IGS is also predicted to form G-quadruplex structures, which might be involved in regulating the transcription and/or processing/stability of the transcribed 35S rRNA precursor (**Figure 2B**).

Sequence analysis showed that the *P. stratiotes* IGS can be subdivided into four distinct structural regions (I–IV, **Figure 2C**). Region I, with a length of 51 bp, represents a unique sequence, with two pyrimidine-rich motifs, CCCTGTCCCACCACCC and CCCCACTCACCCC, starting at nucleotide positions 1 and 39 relative to the end of the 25S rDNA gene (**Supplementary Figure 4**). Such motifs are believed to serve as transcription termination sites.

Region II (1,218 bp, GC content 56.3%) consists of seven units of repeated sequences. Each repeat is composed of subrepeats organized in a specific pattern of two major subrepeat types (A and B). This pattern resembles that found in rice, with the 253–264-bp subrepeats composed of three types of short related DNA elements (Cordesse et al., 1993). The Type A and B *Pistia* subrepeats are normally 19 bp long, with a certain level of divergence, which roughly divides them into subtypes A-1, A-2, B-1, B-2, and B-3. Each repeat contains two Type A subrepeats and three Type B subrepeats arranged in a A-1/B-1/B-2/A-2/B-3 pattern, with some sequence erosion in repeat 1, while repeat 7 is represented only by a combination of A-1 and B-1, as revealed by nucleotide alignment of the repeat sequences (**Figure 2C**). There is also a certain degree of variation between sequences of the same subtype, especially those located close to the repeat zone borders (**Supplementary Figure 5**). Repeat 1 (168 bp) is shorter than repeats 3–6 (191 bp) and demonstrates a higher degree of sequence variation. The Type A subtypes of repeat 1 are incomplete and are 13 and 16 nucleotides in size. The 5′ end

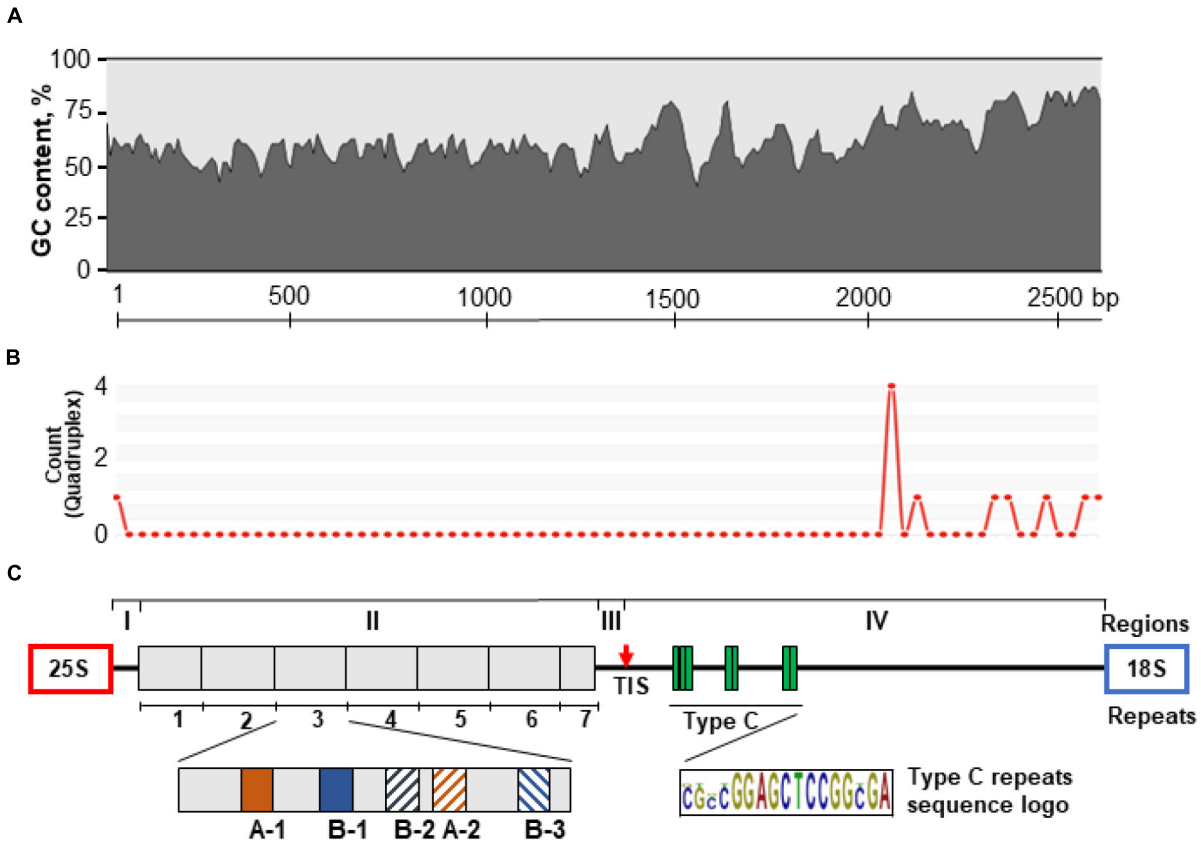


FIGURE 2 | Schematic representation of the molecular architecture of the 35S rDNA IGS of *P. stratiotes* obtained by sequencing the genomic clone Pi-IGS (accession no. OL409040). **(A)** Pattern of G + C nucleotide distribution along the IGS sequence calculated using window size of 45 nucleotides. **(B)** Patterns of G-quadruplex structures predicted for the IGS sequence using window size of 50 nucleotides; the heights of the peaks indicate the relative strength of each G-quadruplex structure. **(C)** The IGS is divided into four regions (I–IV); region II contains seven repeats (1–7) composed of subrepeats A-1, A-2, B-1, B-2, and B-3; region III has a transcription initiation site (TIS), marked by a red arrow; region IV, which corresponds to the 5′-ETS, contains Type C subrepeats with the consensus sequence CGCCGGAGCTCCGGCGA.

of repeat 2 (191 bp) is also quite polymorphic relative to other repeats. The last repeated unit, repeat 7, is incomplete, is 91 bp long, and contains only two subrepeats, which is characteristic of the first parts of repeats 1–6.

Region III, which is 202 bp long (GC content 61.4%), does not have any subrepeats and harbors a TIS with a signature typical for the majority of plant species examined (King et al., 1993; Volkov et al., 2003; Krawczyk et al., 2017). Region IV (1,155 bp) corresponds to the transcribed 5′-ETS and is characterized by the presence of three clusters of GC-rich C subrepeats, separated by unique sequences of lower GC percentage (**Figure 3**). The first cluster contains three subrepeats, and the other two contain only two subrepeats. The region of C-repeats is followed by a unique sequence characterized by higher GC content compared to the other regions.

To gain insight into the possible intragenomic heterogeneity of the individual repeats of 35S rDNA in *P. stratiotes*, we amplified the entire IGS region by PCR using primers specific for the 25S and 18S genes, followed by cloning and characterization of 10 individual clones by restriction enzyme analysis and nucleotide sequencing. Digestion of the clones with restriction enzymes *Eco*RI and *Hin*dIII, which cut out the entire insert, yielded fragments between 2 and 3 kb (**Figure 3A**). Additional digestion with *Eco*RI and *Pst*I, the latter having two recognition sites within the IGS according to the sequence of the original Ps-IGS clone (**Supplementary Figure 4**), revealed two invariable fragments covering the IGS region between the TIS and 18S rRNA gene and a fragment of variable size corresponding to the region of A-B subrepeats upstream of the TIS according to the scheme in **Figure 2**. Sequencing of the 10 clones revealed further details about their molecular architecture. Each clone starts with 53 nucleotides of the 25S rDNA and ends with 41 nucleotides of the 18S rDNA, with four major variants of IGS between the rRNA coding sequences. The three longest fragment variants in clones Pi-IGS_1, Pi-IGS_2, and Pi-IGS_3, which represent *P. stratiotes* IGS Type I, characterized by eight repeats upstream of the TIS, could be subdivided into subtypes I-A and I-B based on the copy numbers of C-repeats downstream of the TIS (**Figure 3**).

The two IGSs classified into Type I-A, Pi-IGS_1 and Pi-IGS_2, are 2,837 and 2,833 bp long, respectively, and contain seven

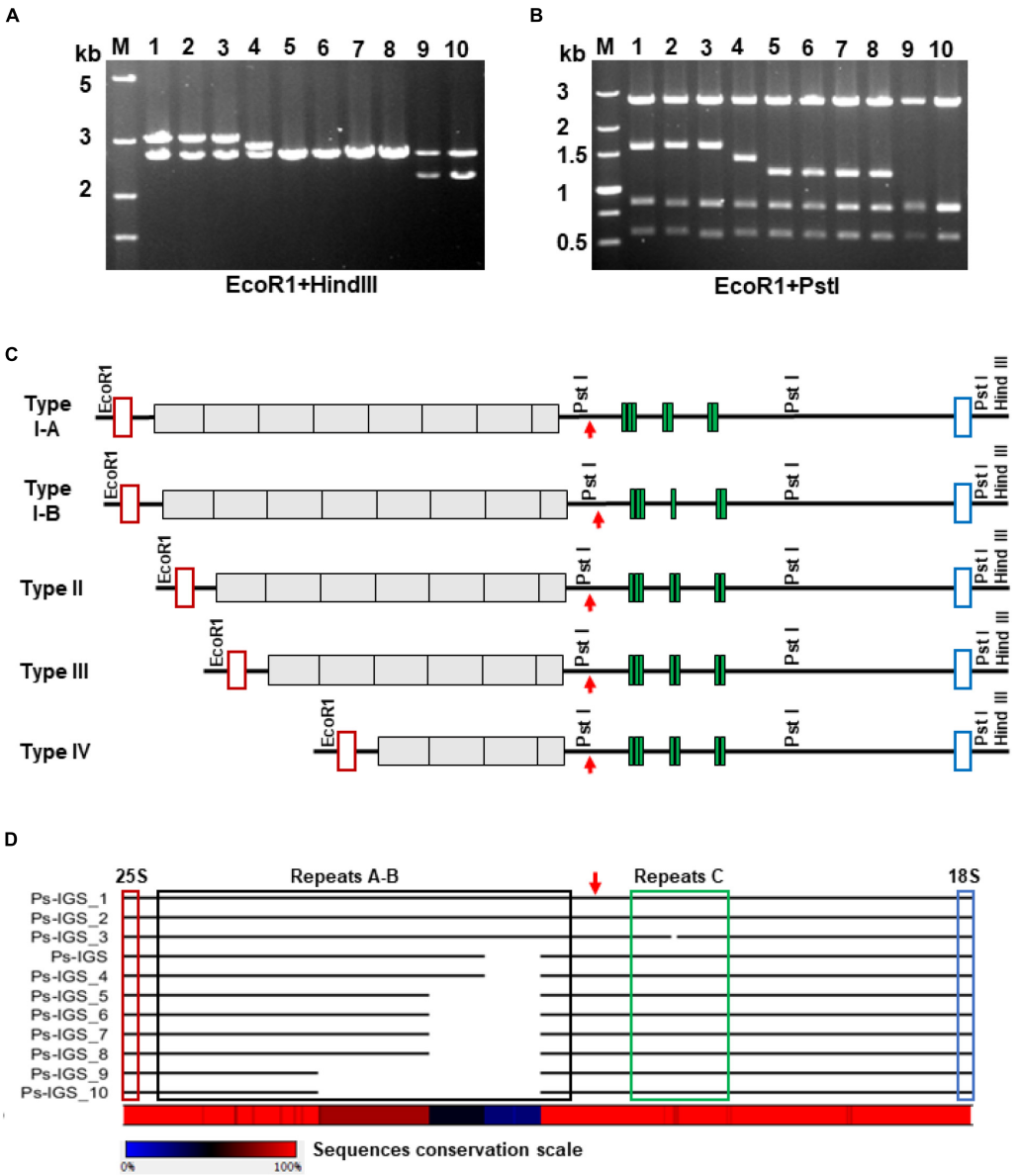


FIGURE 3 | Polymorphism of the *P. stratiotes* 35S rDNA repeat units, as revealed by analysis of IGS sequences in 10 clones containing PCR-amplified fragments. **(A,B)** Length variation of the cloned IGSs, as revealed by analyzing restriction fragment polymorphisms produced by digesting plasmid DNA with the restriction enzymes *Eco*RI plus *Hin*dIII **(A)** and *Eco*RI plus *Pst*I **(B)** and visualized by agarose gel electrophoresis. M—DNA fragments of molecular weight marker DL5000 (Takara) with their length in kilobases on the left site of the gel; numbers 1–10 above the gel images correspond to DNA clones Ps-IGS_1 to Ps-IGS_10. **(C)** Molecular structures of rDNA fragments depicted in A and B, based on full-length nucleotide sequences of the clones. The 10 clones were divided into four major groups based on the sequencing results: Type I–IV, depending on number of repeats in A-B and C. Clones Ps-IGS_1 and Ps-IGS_2 represent Type IA; Ps-IGS_3 represents Type IB; Ps-IGS_4 represents Type II; Ps-IGS_5, Ps-IGS_6, Ps-IGS_7, and Ps-IGS_8 represent Type III; Ps-IGS_9 and Ps-IGS_10 represent Type IV. Open red rectangles mark 25S rDNA sequences; gray blocks mark A-B repeats, green blocks mark C-repeats; open blue rectangles mark 18S rDNA sequences; the TIS is marked by a red arrow. **(D)** Simplified representation of the sequence alignment of the clones; the full nucleotide alignment and the sequences' accession numbers are shown in **Supplementary Figure 4**.

copies of C-repeats. By contrast, the IGS of clone Pi-IGS_3, which is 2,816 bp long, is classified as Type I-B, with six copies of C-repeats. The Type II IGS, represented by clone Pi-IGS_4, with seven repeats upstream of the TIS and a length of 2,643 bp, is very similar to the IGS of genomic clone Pi-IGS (2,644 bp long; **Figure 3**); the two sequences differ by just 11 SNPs, 7 of which are T↔C transitions. The Type III IGS is represented by four almost identical sequences featuring six repeats upstream of the TIS, with a length of 2,453 (clone Pi-IGS_5) and 2,452 bp (clones Pi-IGS_6, Pi-IGS_7, and Pi-IGS_8). Altogether, the four sequences share 41 SNPs dominated by 17 A↔G and 16 T↔C transitions.

The shortest IGS, Type IV, is represented by two clones 2,069 and 2,067 bp long (clones Pi-IGS_9 and Pi-IGS_10) with four repeats upstream of the TIS. Alignment of the obtained IGS sequences of *P. stratiotes* 35S rDNA revealed very high sequence conservation within each of the four IGS types as well as between types, with the major variations related to the number of internal repeats upstream of the TIS (**Figure 3D** and **Supplementary Figure 4**).

Characterization of the 5S Ribosomal DNA

To characterize the molecular structure of the 5S rDNA, we cloned PCR products amplified with two pairs of primers designed to cover neighboring 5S rDNA genes with the NTS between them and sequenced individual clones. The analysis of chimeric sequences reconstructed from parts of neighboring genes might lead to inaccurate conclusions (Galián et al., 2014a), so we used only the through-sequenced 5S rDNAs in subsequent analyses. Based on the sequencing results, we designed an additional pair of primers specific for the NTS sequences and used them to confirm the specific arrangements of the 5S rDNA units (**Supplementary Figure 7**). In total, we sequenced and analyzed 50 clones containing complete 5S rDNA repeats, composed of a sequence encoding 5S rRNA and an adjacent NTS, including 18 clones containing 2 5S rDNA units and 5 clones containing 3 sequential 5S rDNA units.

5S Ribosomal RNA Gene Sequences

Sequence analysis of the clones revealed 78 gene units containing the full-length sequence of the 5S rDNA gene. Only these through-sequenced 5S rDNA genes were used for further analysis (**Supplementary Figure 8**). Among these, 73 units contained a 5S rRNA gene sequence 119 bp long, which is characteristic of these sequences in the majority of plants analyzed (Vandenberghe et al., 1984; Zanke et al., 1995; Tynkevich and Volkov, 2014; Simon et al., 2018; Chen et al., 2021). The gene displayed several sequence variations, with 31 random nucleotide substitutions (mostly T↔C and G↔A transitions) and three more regular transitions: at positions 21 (T↔C, with a 85.9% frequency of T), 29 (G↔A, with a 73.1% frequency of A), and 117 (T↔C, with a 53.8% frequency of T). In the predicted 5S rRNA secondary structure, the substitutions at nucleotides 21 and 27 localize to the loops and are thus unlikely to affect rRNA folding, whereas the substitution at position 117 might make the intra-molecular pairing of the ends of the molecule more relaxed (**Figure 4**). Computer-aided folding of five shorter 5S rRNA gene sequences [114, 110, 106 (two), and 42 bp] revealed secondary structures highly divergent from that obtained for the full-length gene (119 bp), suggesting that these genes are not functional and represent pseudogenes (**Figure 4C**).

Variants of the 5S Ribosomal DNA Non-transcribed Spacers

Among the clones, we identified 64 full-length NTS sequences separating the 5S rDNA genes. These NTS sequences could be separated into two groups based on length and the nucleotide motifs at their 5′-ends (which usually define the termination of transcription of the 5S rRNA) and their 3′-ends (containing DNA elements that modulate gene transcription) (Hemleben and Werts, 1988; Cloix et al., 2003). Members of the major group of 61 NTS sequences (Type I) are longer and contain a TCGT motif at their 5′-terminus, which follows the end of 5S rRNA gene (**Figure 5A**). This represents a divergence from the classic TTTT transcription termination signal found in the majority of plants; however, a TCGT motif has been documented for some other monocot species: rice (*Oryza sativa*; GenBank: CP054686.1, positions 13743700–13751999) and *Scilla scilloides* (accession no. LC213012) and the eudicot *Gossypium* (Cronn et al., 1996). The minor group (Type II) is represented by two of the 64 NTSs (clones Ps19 and Ps20), which start with the classic 5S rDNA terminator motif TTTTT found in most of the plants analyzed, including the duckweeds *Spirodela polyrhiza* (Borisjuk et al., 2018) and *Landoltia punctata* (Chen et al., 2021). Types I and II also have a slightly different nucleotide arrangement at their 3′-termini in front of the 5S rRNA gene (**Figure 5A**).

The Type I NTSs can be further classified into four subtypes according to their length and molecular architecture (**Figure 5B**). The characteristic feature of the major subtype I-a, represented by 40 NTS sequences, is the tandem duplication of a 68-bp region that includes 15 bp of the 3′-terminal sequence of the 5S rRNA gene. The entire R-1 and R-2 sequences differ by a three-nucleotide deletion and five nucleotide substitutions. The 15-bp stretch of the 5S rRNA gene sequence in both duplicated regions (R-1 and R-2 in **Figure 5B**) is followed by a nearly perfect duplication of 13 nucleotides containing the transcription termination signal (**Figure 5**). The next part of the type I-a NTS is represented by either of two types of sequences: (i) a 122-nucleotide sequence with 17 nucleotide substitutions between 36 clones, or (ii) a 108-bp sequence with only a single T↔C transition found in four clones (**Figure 5** and **Supplementary Figure 8**).

Subtype I-b, represented by a single sequence of the clone Ps28, differs from subtype I-a sequences in that the 15-bp piece of the 5S rRNA gene sequence in R-2 was replaced by the shorter sequence GCAAGTCCT, with no obvious sequence homology to the 5S rRNA gene. Compared to subtype I-b, subtype I-c (20 sequences) is shorter by 13 bp, lacking one of the two duplicated elements following the 5S rRNA gene sequence. Subtype I-d (represented by a single sequence of Ps35, NTS-B) differs from subtype I-c due to the elimination of all sequence elements of the R-2 derivative.

The Type II NTS with the classic TTTT transcription termination site is represented in our survey by two clones of a size similar to that of Type I-d (**Figure 5C**). Both clones have almost identical sequences (**Supplementary Figure 8**). The main difference is that a TC-rich insertion of 19 nucleotides provides the longer variant with an additional G4 structure (**Supplementary Figure 9**).

The 5S Ribosomal DNA Unit Types Are Intermingled Within the Locus

The finding that *P. stratiotes* harbors a single 5S rDNA locus, as revealed by FISH, suggests that gene units with different NTS types are arranged in a certain order within this locus.

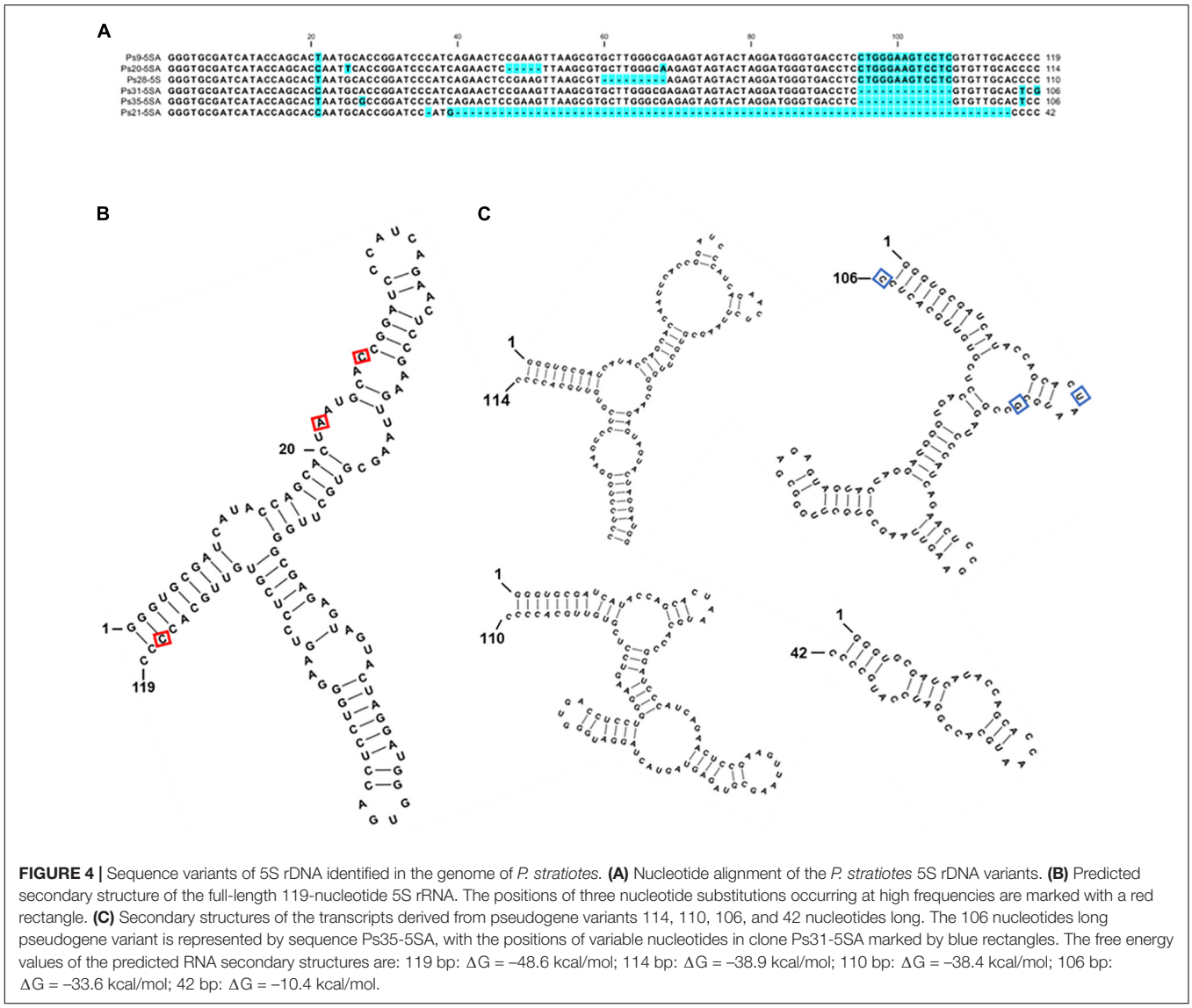


FIGURE 4 | Sequence variants of 5S rDNA identified in the genome of *P. stratiotes*. **(A)** Nucleotide alignment of the *P. stratiotes* 5S rDNA variants. **(B)** Predicted secondary structure of the full-length 119-nucleotide 5S rRNA. The positions of three nucleotide substitutions occurring at high frequencies are marked with a red rectangle. **(C)** Secondary structures of the transcripts derived from pseudogene variants 114, 110, 106, and 42 nucleotides long. The 106 nucleotides long pseudogene variant is represented by sequence Ps35-5SA, with the positions of variable nucleotides in clone Ps31-5SA marked by blue rectangles. The free energy values of the predicted RNA secondary structures are: 119 bp: ΔG = –48.6 kcal/mol; 114 bp: ΔG = –38.9 kcal/mol; 110 bp: ΔG = –38.4 kcal/mol; 106 bp: ΔG = –33.6 kcal/mol; 42 bp: ΔG = –10.4 kcal/mol.

Indeed, sequence analyses of the 10 clones containing two 5S rDNA repeats and 5 clones with three consecutive repeated units demonstrated different modes of unit arrangement. Of the 10 double unit clones, 7 contained rDNA units of the same type: 6 clones contained the dominant I-a type NTSs, and 1 contained type I-c units. The three remaining clones showed a mixed unit arrangement: I-a/I-c (clone Ps29), I-b/I-a (Ps28), and 1-c/1-a (Ps30) (**Figure 6**).

Of the five triple clones, one contained three 5S rDNA units with the same I-a type NTS (clone Ps32), and the others displayed different arrangements of NTS types I-a, I-c, and I-d: a/a/c (clone Ps33), a/c/a (clone Ps34), a/d/a (clone Ps35), and a/c/c (clone Ps36) (**Figure 6C**). To confirm the mixed patterns of type I-a and I-c NTSs, we developed a pair of specific primers for these NTS types and cloned and sequenced the amplified fragments. In total, we obtained six clones containing a 179-bp type I-c NTS, a 119-bp 5S rDNA, and an 81-bp type I-a NTS, confirming that 5S rDNA units with the two most common types of NTS are intermingled within the locus. Of special note is clone Ps21, containing a 5S rDNA pseudogene of 42 bp followed by an atypically short NTS (118 bp) and a regular 119-bp 5S rDNA sequence (**Supplementary Figure 10**). The sequence of this NTS is more similar to a Type II NTS than a Type I NTS but starts with TCGA motifs.

DISCUSSION

The identity of the Chinese *Pistia stratiotes* isolate examined in this study (**Figure 1A**) was confirmed by chloroplast DNA barcodes ATP and PSB. Alignment of the obtained sequences revealed 100 and 99.6% similarity, respectively, to the corresponding ATP and PSB spacers of the chloroplast genome deposited in GenBank (accession no. NC_048522) obtained from an ecotype that originated from North America. The high sequence similarity of the chloroplast DNA barcodes, combined

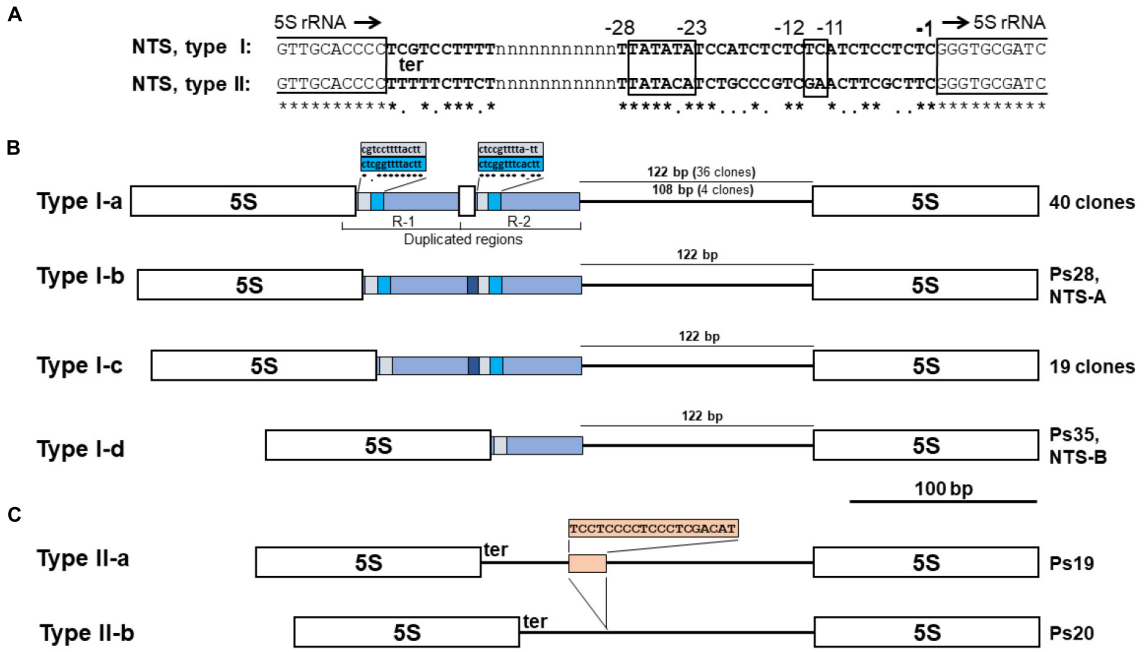


FIGURE 5 | NTS sequence variants of 5S rDNA in the genome of *Pistia stratiotes*. **(A)** Regulatory DNA elements downstream and upstream of the 5S rRNA genes. The end and beginning of 5S rRNA genes are marked by open boxes. Position -1 marks the first nucleotide upstream of the 5S rDNA transcription start; black arrows indicate the direction of gene transcription. **ter**, transcription terminator. The boxed nucleotide position -11 to -12 marks the usual location of the GC dinucleotide; positions -23 to -28 mark the TATA-like motif. **(B)** Schematic representation of 5S rDNA units with Type I NTSs. Open boxes define the sequences of 5S rRNA genes. **(C)** Schematic representation of 5S rDNA units with Type II NTSs. Open boxes define the sequences of 5S rRNA genes.

with the 99.5% similarity of 18S rDNA sequences from the Chinese (this study) and US (accession no. AF168869) isolates, suggests a relatively low genetic variability between ecotypes of this mostly vegetatively propagating aquatic plant.

Genome Size, Chromosome Number, and Localization of Ribosomal DNA Loci

In agreement with previous reports (Blackburn, 1933; Subramanyam, 1962; Krishnappa, 1971; Geber, 1989), our study revealed that *P. stratiotes* has a set of 2n = 28 small chromosomes. The estimated haploid genome size of 407 Mbp, is roughly 39% larger than previously reported by Geber (1989; 249 Mbp) The suggested reason for this discrepancy is that, in contrast to the presented data, previous measurements were performed by Feulgen microdensitometry. In any case, the obtained values are in the range of 100 Mbp—1 Gbp, typical for many small and fast-growing angiosperms (Leitch et al., 2019). FISH revealed a single terminal locus for both 5S and 35S rRNA genes, supporting the diploid status of *P. stratiotes*. Single rDNA loci are common in several diploid angiosperms, with 51.38% of these plants having a single 5S rDNA locus and 35.5% having a single 35S rDNA locus according to Garcia et al. (2017).

Karyologic data were available only for approximately one-quarter of the 3,300–3,600 species belonging to Araceae *sensu lato* (including Lemnaceae as Lemnoidea) (Cusimano et al., 2012), an ancient and very diverse family (Mayo et al., 1997; Boyce and Croat, 2013; Vasconcelos et al., 2018) whose history has been traced back to the early Cretaceous period (Friis et al., 2004), probably overlapping with of dinosaurs. Chromosome numbers ranging from 2n = 10 to 2n = 168 were reported for this group (Cusimano et al., 2012). However, increasing interest in the phylogeny and evolution of monocots in general (Barfod et al., 2010), and the Araceae family in particular (Cusimano et al., 2011) during the current decade, has stimulated research on Araceae cytogenomics, which led to the localization of 5S and 35S rDNA by FISH for several representative species. Two or four 35S rDNA loci have been mapped in six ornamental species of the genera *Anthurium*, *Monstera*, *Philodendron*, *Spathiphyllum*, *Syngonium*, and *Zantedeschia* (Lakshmanan et al., 2015). Mapping of 5S rDNA and 35S rDNA in 10 species of the genus *Typhonium* (Sousa et al., 2014) and in 14 species with chromosome numbers of 2n = 14 to 2n = 60, representing 11 different genera (Sousa and Renner, 2015), resulted in the visualization of 1 pair of 5S rDNA sites in subterminal or interstitial regions and (predominantly) 4 loci of 35S rDNA. By contrast, analysis of 29 species of *Philodendron* and 5 species of *Thaumatophyllum*, with chromosome numbers ranging from 2n = 28 to 2n = 36, revealed one or two loci for 5S rDNA and a wide range of 1–9 chromosomes with FISH signals for 35S rDNA (Vasconcelos et al., 2018). Localization of rDNA in 11 species representing all 5 genera of duckweed, with a 2n chromosome number ranging from 36 to 82, revealed 1–3 loci for 5S rDNA and 1 or 2 loci for 35S rDNA (Hoang et al., 2019). These increasing amounts of data on chromosomal rDNA representation should be extended by further investigating

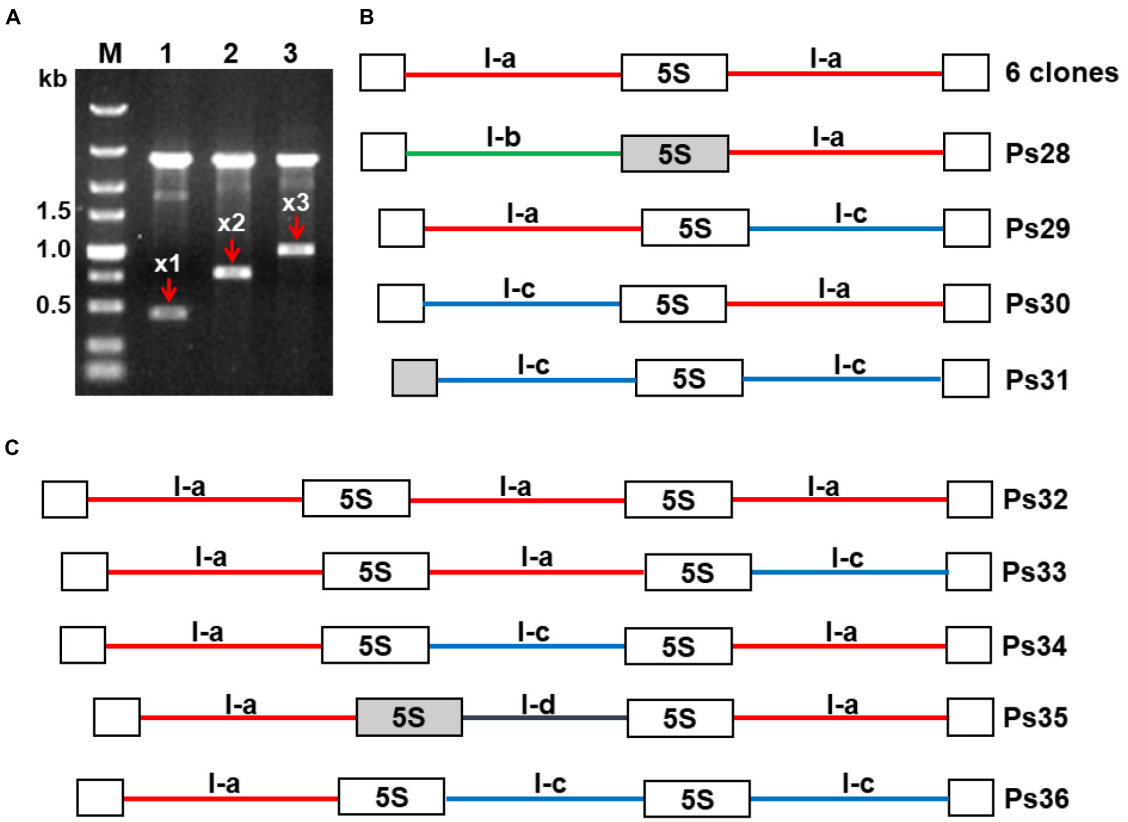


FIGURE 6 | Modes of repeat arrangement revealed in clones containing multiple 5S rDNA units. **(A)** Agarose gel showing digested plasmids containing DNA fragments composed of one (x1), two (x2), and three (x3) units of 5S rDNA in the representative clones Ps20, Ps25, and Ps35, respectively. **(B)** Schematic representation of the arrangement of rDNA units with NTS types I-a, I-b, and I-c in clones containing two 5S rDNA repeats. **(C)** Schematic representation of the arrangement of rDNA units with NTS types I-a, I-c, and I-d in clones containing three 5S rDNA repeats. All depicted DNA clones are in 5′–3′ orientation. Open boxes represent full-length (119 bp) 5S rDNA genes; gray boxes in clones Ps28, Ps31, and Ps35 represent 5S rDNA pseudogenes. The six identical clones with double NTSs of type I-a are represented in GenBank by the sequence Ps22 with accession number OL409056. The accession numbers of other sequences are as follows: Ps28—OL409060; Ps29—OL409057; Ps30—OL409058; Ps31—OL409059; Ps32—OL409061; Ps33—OL409062; Ps34—OL409063; Ps35—OL409064; Ps36—OL409065.

the functionality and molecular evolution of the Araceae 5S and 35S rDNA genes.

Molecular Organization and Evolution of 35S Ribosomal DNA Repeats in *Pistia stratiotes*

Deciphering the full-length nucleotide sequence of 35S rDNA in *P. stratiotes*, for the first time for the Araceae *sensu stricto*, revealed a molecular architecture typically found in other plant taxa (Hemleben and Zentgraf, 1994), with conserved sequences encoding 18S, 5.8S, and 25S rRNAs and diverse internal transcribed spacers, ITS1 and ITS2, and the IGS separating the coding sequences.

BLAST analysis of the rDNA gene sequences revealed the highest similarities (∼99.5%) to the single entry of 18S rDNA genes from *P. stratiotes* (sampled in the United States) in GenBank, demonstrating high sequence conservation among different *P. stratiotes* ecotypes. The phylogenetic tree based on 18S rDNA sequences (**Supplementary Figure 3**) groups *P. stratiotes* together with other Araceae species, with a certain distance from duckweeds, showing potential for further resolving phylogenetic relationships within the Araceae family (Tippery et al., 2021). No entry for the 5.8S rDNA of *P. stratiotes* was found in GenBank. The best hits for the 5.8S rDNA gene were the homologous sequences of *Amorphophallus elliottii*, a species that belongs to the same subfamily, *Aroideae*; however, the 5.8S rDNA sequence is too short and conserved for building a reliable phylogenetic tree. By contrast, the ITS1 and ITS2 sequences are more variable, making them a relatively popular tool in plant phylogenetic studies (Alvarez and Wendel, 2003). However, up to date there are not too many Araceae ITS sequences available in the GenBank. For example, when using "megablast" option for highly similar sequences, the *Pistia* ITS queries hit the homologous sequences representing a single genus of the whole Araceae family—*Amorphophallus*.

The IGS of *P. stratiotes* 35S rDNA has a structure typically observed in plants, comprising four structural/functional regions: (i) a 3′-end external transcribed sequence (3′-ETS) with the

presumed transcription termination site; (ii) a region of repeats, often functioning to enhance 35S rRNA precursor transcription; (iii) a region containing the transcription initiation site (TIS); and (iv) the 5′-external transcribed spacer (ETS), which stretches to the beginning of the sequence encoding 18S rRNA.

The arrangement of 3′-ETS in the 35S rDNA IGS of *P. stratiotes* (region I, **Figure 2C**) resembles that described for other monocotyledonous species such as rice, maize, sorghum, and *Brachypodium distachyon* (Krawczyk et al., 2017) as well as dicots (Borisjuk et al., 1997; Volkov et al., 2017). Between region II and the TIS (region III) with the sequence TATTATAGGGG, closely resembling that of other dicot and monocot plants (Volkov et al., 2003; Huang et al., 2017), is a unique 78-bp sequence that does not match any homologs deposited in GenBank. This one more time highlights the fact that no IGS sequences of related Araceae species are available for comparison in GenBank, considering that the IGS sequences upstream of TIS demonstrated significant similarity in other groups of species representing the same family, such as maize, wheat, and rice of Poaceae (Cordesse et al., 1993) or potato, tobacco, and tomato of Solanaceae (Borisjuk et al., 1997). The 5′-ETS (region IV), a 1,255-bp region in *Pistia*, contains clusters of 17-bp repeats downstream of the TIS and a gradually increasing GC content toward the beginning of the 18S rDNA, averaging 75.1% within the 600-bp region adjacent to the 18S rDNA gene. This GC-enriched region may form G4 structures (**Figure 2**), possibly contributing to the regulation of transcription and replication of rDNA (Havlová and Fajkus, 2020). BLAST analysis of the *Pistia* 5′-ETS sequence upstream of the 18S gene, an IGS region of high similarity between species of the same family (Cordesse et al., 1993; Borisjuk et al., 1997; Volkov et al., 2017), did not reveal any homologs in GenBank. This was surprising, as numerous sequences of this region from duckweed species (Tippery et al., 2015) were deposited in GenBank, and the classification based on morphological observations assumes a close relationship between duckweeds and core aroids including *Pistia* (Maheshwari, 1958). However, the finding agrees with the notion that duckweeds are not very closely related to the core Araceae and should be treated as a separate sister family (Lemnaceae) of Araceae (Tippery and Les, 2020; Acosta et al., 2021) rather than as the subfamily Lemnoidae within the Araceae (French et al., 1995; Cusimano et al., 2011).

Our data on the intragenomic heterogeneity of the 35S rDNA in *P. stratiotes*, which we generated by sequencing one genomic clone and 10 random IGS fragments obtained by PCR, shed additional light on the genome organization and evolution of plant rDNA. Nucleotide alignment of the obtained sequences revealed significant variation in IGS size, primarily due to variation in the number of repeats upstream of the TIS, which is well documented to occur at inter species level (Chang et al., 2010; Huang et al., 2017; Hemleben et al., 2021), between cultivars (Polanco and De La Vega, 1997) or between different rDNA loci of the same genome (Sáez-Vásquez and Delseny, 2019; Appels et al., 2021). Apart from the variation in repeat numbers, the IGSs demonstrate considerable sequence homogeneity, with 95 SNPs among the 11 IGS sequences (∼3 SNPs per kbp) (**Supplementary Figure 4**).

Remarkably, the variation in repeat copy number exclusively involves the most conserved repeated units in the middle of the track (**Figure 3**), while the more divergent flanking repeats remained intact in all 11 IGS variants. The area of the most homogenized repeats likely represents a recombination hot spot that is responsible for the IGS length variation in *P. stratiotes*. Another point to highlight is that the single 35S rDNA locus visualized by FISH contains at least five repeat variants (most likely more, considering that the 11 clones characterized here may not cover the entire IGS sequence variability). This observation supports the growing amount of data showing that the 35S rDNA repeats within a locus are not fully homogenized (Galián et al., 2014b; Sims et al., 2021).

Molecular Architecture of *Pistia stratiotes* 5S Ribosomal DNA

To the best of our knowledge, this work represents the first molecular study of 5S rDNA in Araceae considering the related duckweeds, for which data are available (Borisjuk et al., 2018; Hoang et al., 2020; Chen et al., 2021), as a separate family. The sequencing of multiple 5S rDNA gene units, containing 78 sequences of full length 5S rRNA gene and 64 full-length intergenic spacers of *P. stratiotes,* revealed a molecular arrangement common for the majority of analyzed plants (Zhu et al., 2008), with a conserved 119-bp sequence for 5S rRNA and a variable non-transcribed spacer (NTS). Certain nucleotide variations, mostly T↔C and G↔A transitions, which are often observed in other plants (Cronn et al., 1996), were detected in the 5S rDNA of *P. stratiotes*. Of special interest is the T↔C transition at position + 117 of the coding region. The 37 5S rRNA genes containing T at the + 117 position are followed by NTS Types I-c and 1-d, whereas the genes with C at this position are linked with NTS Types I-a and II.

In addition, analysis of the gene coding region showed some variants with deletions. Based on the deduced rRNA secondary structure (**Figure 4**), these variants appear to be pseudogenes. Similar degraded 5S rDNA sequences have been described for many eukaryotic organisms, especially fishes (Rebordinos et al., 2013; Barman et al., 2016), crustaceans (Perina et al., 2011), and plants (Sergeeva et al., 2017; Volkov et al., 2017). Other molecular pathways than those leading to "concerted evolution" likely contribute to shaping the landscape of plant 5S rDNA (Nei and Rooney, 2005).

Similar to the 5S rDNA of the duckweeds *S. polyrhiza*, *S. intermedia*, and *L. punctata* (Hoang et al., 2020; Chen et al., 2021), the *P. stratiotes* gene units can be divided into two major classes (Type I and II) and their subtypes (**Figure 5**) with slightly different nucleotide arrangements at their 5′- and 3′-termini. These differences affect G4 structures (**Supplementary Figure 9**), and can potentially influence a gene's functionality (Yadav et al., 2017; Havlová and Fajkus, 2020).

The intragenomic variability of 5S rDNA in *P. stratiotes* is quite intriguing, considering that a single 5S rDNA locus per genome was visualized by FISH (**Figure 1**). Even more interesting is our finding about the arrangement of different gene units within the locus uncovered by sequencing clones containing

double and triple 5S rDNA units (**Figure 6**) and of PCR clones generated with primers specific for different types of NTSs (**Supplementary Figure 11**). The data summarized in **Figure 6** suggest that different types of 5S rDNA units are randomly interlinked with each other along the rDNA stretch, forming various combinations between neighboring units, according to their frequency in the genome (**Supplementary Figure 8**). Fortunately, among these multiunit clones, we also identified less frequent types of rDNA units (NTS types 1b and 1d) and even one unit with a defective 5S rDNA sequences. These data further detail the mosaic arrangement of 5S rDNA revealed earlier in rice (Zhu et al., 2008, GenBank ID: CP054686.1, range: 13,743,700–13,751,999 bp), wheat (Sergeeva et al., 2017), and the duckweed *Landoltia punctata* (Chen et al., 2021), providing new insights into 5S rDNA organization and arrangement in plants.

Taken together, due to the generation of multiple sequences of both 35S and 5S rDNA units, our study sheds new light on the intra-genomic variability of rDNA and provides novel findings about the evolution of these genes in plants. In general, the newly obtained data support the recent trend suggesting separate modes of molecular evolution for 35S and 5S rDNA (Mahelka et al., 2013; Volkov et al., 2017).

CONCLUSION

Our combined molecular, cytogenetic, and phylogenetic data provide comprehensive characterization of rDNA loci of the ancient monocot plant *P. stratiotes*, representing the Araceae family. Our findings confirm the high conservation of sequences encoding 18S, 5.8S, 25S, and 5S rRNAs.

Nucleotide sequencing of multiple clones containing the most variable parts of rDNA, the IGS of 35S rDNA and the NTS of 5S rDNA, uncovered the scale of intra-genomic and intra-locus variation. Our data support the idea of a mosaic arrangement of multiple variants of 35S and 5S rDNA units in single loci as the rule rather than the exception. The *P. stratiotes* 35S rDNA locus displayed at least four length variants of the gene, which differ in the number of repeats within the IGS, but not in the promoter or transcribed sequences. The 5S rDNA locus displayed at least six types of functional gene units, intermingled with each other and with pseudogenes. Our findings confirm the notion that sequence homogenization through "concerted evolution," together with crossovers at recombination hot spots in IGS repeats, are major molecular forces shaping the 35S rDNA arrays in plants. At the 5S rDNA locus, an evolutionary shortening of the dominant type of gene units through gradual loss of repeated NTS elements apparently took place and likely generated non-functional pseudogene variants, best described by the "birth-and-death" model. Thus, our findings suggest that 35S rDNA and 5S rDNA in plants evolve via different mechanisms.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

IS and NB conceived the study. AS and GC conducted molecular analysis of rDNA and generated all rDNA sequencing data. PH and JF performed chromosome analysis. AS, IS, and NB organized, analyzed the data, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by an individual grant provided by the Huaiyin Normal University (Huai'an, China) to NB.

ACKNOWLEDGMENTS

We thank to Anton Peterson and Olena Kishchenko (both at the School of Life Sciences, Huaiyin Normal University, Huai'an, China) for purchasing, cultivation, and photographic images of *Pistia stratiotes*, as well as to a group of undergraduate students for their help in the laboratory.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.819750/full#supplementary-material

Supplementary Figure 1 | General organization of plant 35S rDNA repeats and representation of the cloned fragments of *Pistia stratiotes* rDNA units characterized in this study. Two genomic DNA regions encoding the major parts of 18S-5.8S-25S rRNA genes (Pi-rDNA-1 and Pi-rDNA-2) and a DNA region containing the 3′-part of the 25S rDNA, intergenic spacer (IGS), and 5′-part of the 18S rDNA (Pi-IGS-1) were cloned as *Xba*I + *Mfe*I restriction fragments. Ten additional fragments covering the IGS region (Pi-IGS_1 to Pi-IGS-10) were produced by PCR using primers specific for the coding sequences of 25S rRNA (F1) and 18S rRNA (R1) and through-sequenced with internal primers F2, F3, R2, and R3.

Supplementary Figure 2 | Sequence alignment of *P. stratiotes* rDNA units encoding 18S-5.8S-25S rRNAs represented by genomic clones Pi-rDNA-1 (accession no. OL375700) and Pi-rDNA-2 (accession no. OL375701). Variable residues are highlighted in light blue. Blue rectangles mark 18S rDNA sequences; yellow rectangles mark 5.8S sequences; red rectangles mark 25S rDNA sequences.

Supplementary Figure 3 | Phylogeny of monocots, eudicots, and magnoliids derived from 18S rDNA sequences. The phylograms shows a maximum likelihood tree obtained using different available 18S rDNA sequences from different plant clades. Branch lengths represent the expected numbers of substitutions per nucleotide site.

Supplementary Figure 4 | Sequence alignment of IGS variants revealed in *P. stratiotes* 35S rDNA. Variable residues and regions are highlighted in light blue. Red rectangle marks 25S rDNA sequences; black rectangles mark A-B repeats; green rectangles mark C repeats; blue rectangle marks sequences of 18S rDNA;

the TIS is marked by a red arrow; *Pst*I restriction sites are marked by black arrows. The sequences' accession numbers are as follows: Ps-IGS: OL409040; Ps-IGS_1: OL409041; Ps-IGS_: OL409042; Ps-IGS_3: OL409043; Ps-IGS_4: OL409044; Ps-IGS_5: OL409045; Ps-IGS_6: OL409046; Ps-IGS_7: OL409047; Ps-IGS_8: OL409048; Ps-IGS_9: OL409049; Ps-IGS_10: OL409050.

Supplementary Figure 5 | Sequence alignment of the seven repeats (1–7) comprising IGS zone II in **Figure 3C** with the deduced consensus sequence (**Cons**). Subrepeats A-I and A-2 are marked by red rectangles, and subrepeats B-I, B-2, and B-3 are marked by blue rectangles. Sequence alignment, the consensus sequence (**Cons**), and sequence logos (**Logo**) were produced using CLC Main Workbench (Version 6.9.2, Qiagen).

Supplementary Figure 6 | Schematic representation of 5S rDNA organization and the strategy used to generate DNA fragments specific for this genome region in *Pistia stratiotes*. The 5S rDNA repeats were cloned as single, double, or triple units following PCR amplification using primers specific for the 5S rRNA gene sequence (marked as F1, F2 and R1, R2) or to the newly discovered sequences of the NTS (F3 and R3).

Supplementary Figure 7 | Alignment of the complete 5S rDNA sequence of *P. stratiotes*. Matching residues are shown as dots. Gap fractions are highlighted in blue.

Supplementary Figure 8 | Nucleotide alignment of Type I-a, I-b, I-c, I-d, II-a, and II-b 5S NTS sequences from *P. stratiotes*. Variable residues and regions are highlighted in blue. The accession numbers of the representative sequences deposited in GenBank are as follows: Ps19: OL409055; Ps20: OL409052; Ps22: OL409056; Ps28: OL409060; Ps29: OL409057; Ps30: OL409058; Ps31: OL409059; Ps32: OL409061; Ps33: OL409062; Ps34: OL409063; Ps35: OL409064; Ps36: OL409065.

Supplementary Figure 9 | Patterns of G-quadruplex structures predicted for Type II 5S rDNA NTS sequences of *P. stratiotes*. Patterns of G-quadruplex structures predicted for Type II-a **(A)**, represented by clone Ps19 (accession no. OL409055) and Type II-b **(B)**, represented by clone Ps20 (accession no. OL409052). The heights of the peaks indicate the relative strength of each G-quadruplex structure; numbers indicate nucleotide positions relative to the transcription start of the 5S rDNA gene; the filled box corresponds to the position of the sequence TCCTCCCCTCCCTCGACAT in **Figure 5C**.

Supplementary Figure 10 | Sequence alignment of three PCR fragments containing the canonical 5S rDNA gene sequence (**Ps9**) and pseudogenes (**Ps20** and **Ps21)**. The 5S rDNA gene variants are marked by green rectangles. The identifications of the sequences are as follows: Ps9: OL450405; Ps20: OL409052; Ps21: OL409053.

Supplementary Figure 11 | Sequence alignment of the six PCR fragments amplified with primers specific for Type I-a and Type I-c NTSs. Type I-a NTS sequences are marked by red rectangles, 5S rDNA by green rectangles, and Type I-c NTSs by blue rectangles. The identifications of the representative sequences are as follows: Ps62—OL409066; Ps63—OL409067.

REFERENCES

Acosta, K., Appenroth, K. J., Borisjuk, L., Edelman, M., Heinig, U., Jansen, M. A. K., et al. (2021). Return of the Lemnaceae: duckweed as a model plant system in the genomics and postgenomics era. *Plant Cell* 33, 3207–3234. doi: 10.1093/plcell/koab189

Alvarez, I., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/s1055-7903(03)00208-2

Appels, R., and Dvořák, J. (1982). The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. *Theor. Appl. Genet.* 63, 337–348. doi: 10.1007/BF00303905

Appels, R., Wang, P., and Islam, S. (2021). Integrating wheat nucleolus structure and function: variation in the wheat ribosomal RNA and protein genes. *Front. Plant Sci.* 12:686586. doi: 10.3389/fpls.2021.686586

Appenroth, K.-J., Teller, S., and Horn, M. (1996). Photophysiology of turion formation and germination in *Spirodela polyrhiza*. *Biol. Plantarum* 38:95. doi: 10.1007/BF02879642

Barfod, A., Davis, J. I., Petersen, G., and Seberg, O. (2010). *Diversity, Phylogeny and Evolution in the Monocotyledons*. Aarhus: Aarhus University Press.

Barman, A. S., Singh, M., Singh, R. K., and Lal, K. K. (2016). Evidence of birth-and-death evolution of 5S rRNA gene in *Channa* species (*Teleostei*, *Perciformes*). *Genetica* 144, 723–732. doi: 10.1007/s10709-016-9938-6

Blackburn, K. B. (1933). Notes on the chromosomes of the duckweeds (Lemnaceae) introducing the question of chromosome size. *Proc. Univ. Durham Phil. Soc.* 9, 84–90.

Borisjuk, N., Chu, P., Gutierrez, R., Zhang, H., Acosta, K., Friesen, N., et al. (2015). Assessment, validation and deployment strategy of a two-barcode protocol for facile genotyping of duckweed species. *Plant Biol (Stuttg)* 17(Suppl. 1), 42–49. doi: 10.1111/plb.12229

Borisjuk, N., and Hemleben, V. (1993). Nucleotide sequence of the potato rDNA intergenic spacer. *Plant Mol. Biol.* 21, 381–384. doi: 10.1007/BF00019953

Borisjuk, N., Peterson, A. A., Lv, J., Qu, G., Luo, Q., Shi, L., et al. (2018). Structural and biochemical properties of duckweed surface cuticle. *Front. Chem.* 6:317. doi: 10.3389/fchem.2018.00317

Borisjuk, N. V., Davidjuk, Y. M., Kostishin, S. S., Miroshnichenco, G. P., Velasco, R., Hemleben, V., et al. (1997). Structural analysis of rDNA in the genus Nicotiana. *Plant Mol. Biol.* 35, 655–660. doi: 10.1023/a:1005856618898

Borisjuk, N. V., Momot, V. P., and Gleba, Y. (1988). Novel class of rDNA repeat units in somatic hybrids between *Nicotiana* and *Atropa*. *Theor. Appl. Genet.* 76, 108–112. doi: 10.1007/BF00288839

Boyce, P. C., and Croat, T. B. (2013). *The Überlist of Araceae, Totals for Published and 771 Estimated Number of Species in Aroid Genera*. Available online at: http://www.aroid.org/genera/ 772 130307uberlist.pdf (accessed November 19, 2021).

Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., et al. (2019). G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495. doi: 10.1093/bioinformatics/btz087

Cao, H. X., Vu, G. T. H., Wang, W., Appenroth, K. J., Messing, J., and Schubert, I. (2016). The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. *New Phytol.* 209, 354–363. doi: 10.1111/nph.13592

Chang, K. D., Fang, S. A., Chang, F. C., and Chung, M. C. (2010). Chromosomal conservation and sequence diversity of ribosomal RNA genes of two distant *Oryza* species. *Genomics* 96, 181–190. doi: 10.1016/j.ygeno.2010.05.005

Chen, G., Stepanenko, A., and Borisjuk, N. (2021). Mosaic arrangement of the 5S rDNA in the aquatic plant *Landoltia punctata* (Lemnaceae). *Front. Plant Sci.* 12:678689. doi: 10.3389/fpls.2021.678689

Choi, K. S., Park, K. T., and Park, S. (2017). The chloroplast genome of *Symplocarpus renifolius*: a comparison of chloroplast genome structure in Araceae. *Genes (Basel)* 8:324. doi: 10.3390/genes8110324

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., and Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the *Arabidopsis thaliana* 5S rRNA genes. *Plant J.* 35, 251–261. doi: 10.1046/j.1365-313x.2003.01793.x

Cordesse, F., Cooke, R., Tremousaygue, D., Grellet, F., and Delseny, M. (1993). Fine structure and evolution of the rDNA intergenic spacer in rice and other cereals. *J. Mol. Evol.* 36, 369–379. doi: 10.1007/BF00182184

Croat, T. B. (1988). Ecology and life-forms of Araceae. *Aroideana* 11, 4–56. doi: 10.3390/plants11030334

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Cusimano, N., Bogner, J., Mayo, S. J., Boyce, P. C., Wong, S. Y., Hesse, M., et al. (2011). Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies. *Am. J. Bot.* 98, 654–668. doi: 10.3732/ajb.1000158

Cusimano, N., Sousa, A., and Renner, S. S. (2012). Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in Araceae, with a critique of the bias introduced by 'x'. *Ann. Bot.* 109, 681–692. doi: 10.1093/aob/mcr302

Delcasso-Tremousaygue, D., Grellet, F., Panabieres, F., Ananiev, E. D., and Delseny, M. (1988). Structural and transcriptional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant. *Eur. J. Biochem.* 172, 767–776. doi: 10.1111/j.1432-1033.1988.tb13956.x

Do, H. D. K., Kim, C., Chase, M. W., and Kim, J.-H. (2020). Implications of plastome evolution in the true lilies (monocot order Liliales). *Mol. Phylogenet. Evol.* 148:106818. doi: 10.1016/j.ympev.2020.106818

Dolezel, J., Bartos, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry A* 51, 127–128. author reply 129, doi: 10.1002/cyto.a.10013

French, J. C., Chung, M. G., and Hur, Y. K. (1995). Chloroplast DNA phylogeny of the Ariflorae. *Monocotyledons: Syst. Evol.* 1, 255–275.

Friis, E. M., Pedersen, K. R., and Crane, P. R. (2004). Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16565–16570. doi: 10.1073/pnas.0407174101

Galián, J. A., Rosato, M., and Rosselló, J. A. (2014a). Partial sequence homogenization in the 5S multigene families may generate sequence chimeras and spurious results in phylogenetic reconstructions. *Syst. Biol.* 63, 219–230. doi: 10.1093/sysbio/syt101

Galián, J. A., Rosato, M., and Rosselló, J. A. (2014b). Incomplete sequence homogenization in 45S rDNA multigene families: intermixed IGS heterogeneity within the single NOR locus of the polyploid species *Medicago arborea* (Fabaceae). *Ann. Bot.* 114, 243–251. doi: 10.1093/aob/mcu115

Gao, Y., Yin, S., Yang, H., Wu, L., and Yan, Y. (2018). Genetic diversity and phylogenetic relationships of seven *Amorphophallus* species in southwestern China revealed by chloroplast DNA sequences. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 29, 679–686. doi: 10.1080/24701394.2017.1350855

Garcia, S., Kovařík, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Geber, G. (1989). *Zur Karyosystematik der Lemnaceae*. Ph.D. dissertation. Vienna: Universität Wien.

Gottlob-McHugh, S. G., Lévesque, M., MacKenzie, K., Olson, M., Yarosh, O., and Johnson, D. A. (1990). Organization of the 5S rRNA genes in the soybean *Glycine max* (L.) Merrill and conservation of the 5S rDNA repeat structure in higher plants. *Genome* 33, 486–494. doi: 10.1139/g90-072

Havlová, K., and Fajkus, J. (2020). G4 Structures in control of replication and transcription of rRNA Genes. *Front. Plant Sci.* 11:593692. doi: 10.3389/fpls.2020.593692

Hemleben, V., Grierson, D., Borisjuk, N., Volkov, R. A., and Kovarik, A. (2021). Personal perspectives on plant ribosomal RNA genes research – from precursor-rRNA to molecular evolution. *Front. Plant Sci.* 12:797348. doi: 10.3389/fpls.2021.797348

Hemleben, V., and Werts, D. (1988). Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). *Gene* 62, 165–169. doi: 10.1016/0378-1119(88)90591-4

Hemleben, V., and Zentgraf, U. (1994). "Structural organization and regulation of transcription by RNA Polymerase I of plant nuclear ribosomal RNA genes," in *Plant Promoters and Transcription Factors*, ed. L. Nover (Berlin: Springer), 3–24. doi: 10.1007/978-3-540-48037-2_1

Henriquez, C. L., Arias, T., Pires, J. C., Croat, T. B., and Schaal, B. A. (2014). Phylogenomics of the plant family Araceae. *Mol. Phylogenet. Evol.* 75, 91–102. doi: 10.1016/j.ympev.2014.02.017

Hoang, P. T. N., Fiebig, A., Novák, P., Macas, J., Cao, H. X., Stepanenko, A., et al. (2020). Chromosome-scale genome assembly for the duckweed *Spirodela intermedia*, integrating cytogenetic maps, PacBio and Oxford Nanopore libraries. *Sci. Rep.* 10:19230. doi: 10.1038/s41598-020-75728-9

Hoang, P. T. N., and Schubert, I. (2017). Reconstruction of chromosome rearrangements between the two most ancestral duckweed species *Spirodela polyrhiza* and *S. intermedia*. *Chromosoma* 126, 729–739. doi: 10.1007/s00412-017-0636-7

Hoang, P. T. N., Schubert, V., Meister, A., Fuchs, J., and Schubert, I. (2019). Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci among duckweeds. *Sci. Rep.* 9:3234. doi: 10.1038/s41598-019-39332-w

Huang, Y., Yu, F., Li, X., Luo, L., Wu, J., Yang, Y., et al. (2017). Comparative genetic analysis of the 45S rDNA intergenic spacers from three *Saccharum* species. *PLoS One* 12:e0183447. doi: 10.1371/journal.pone.0183447

Keating, R. (2004). Vegetative anatomical data and its relationship to a revised classification of the genera of Araceae. *Ann. Missouri Botanical Garden* 91, 485–494.

King, K., Torres, R. A., Zentgraf, U., and Hemleben, V. (1993). Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of cucurbitaceae. *J. Mol. Evol.* 36, 144–152. doi: 10.1007/BF00166250

Kodituwakku, K. A. R. K., and Yatawara, M. (2020). Phytoremediation of industrial sewage sludge with *Eichhornia crassipes*, *Salvinia molesta* and *Pistia stratiotes* in batch fed free water flow constructed wetlands. *Bull. Environ. Contam Toxicol.* 104, 627–633. doi: 10.1007/s00128-020-02805-0

Krawczyk, K., Nobis, M., Nowak, A., Szczecińska, M., and Sawicki, J. (2017). Phylogenetic implications of nuclear rRNA IGS variation in *Stipa* L. (Poaceae). *Sci. Rep.* 7:11506. doi: 10.1038/s41598-017-11804-x

Krishnappa, D. G. (1971). Cytological studies in some aquatic angiosperms. *Proc. Ind. Acad. Sci.-Sect. B* 73, 179–185. doi: 10.1007/bf03045290

Labudová, D., Hon, J., and Lexa, M. (2020). pqsfinder web: G-quadruplex prediction using optimized pqsfinder algorithm. *Bioinformatics* 36, 2584–2586. doi: 10.1093/bioinformatics/btz928

Lakshmanan, P. S., Van Laere, K., Eeckhaut, T., Van Huylenbroeck, J., Van Bockstaele, E., and Khrustaleva, L. (2015). Karyotype analysis and visualization of 45S rRNA genes using fluorescence in situ hybridization in aroids (Araceae). *Comp. Cytogenet.* 9, 145–160. doi: 10.3897/CompCytogen.v9i2.4366

Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150

Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O., and Bennett, M. D. (2019). *Plant DNA C-Values Database (Release 7.1)*. Available online at: https://cvalues.science.kew.org/ (accessed December 1, 2021).

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., et al. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 47, W260–W265. doi: 10.1093/nar/gkz303

Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Lu, Q., He, Z. L., Graetz, D. A., Stoffella, P. J., and Yang, X. (2010). Phytoremediation to remove nutrients and improve eutrophic stormwaters using water lettuce (*Pistia stratiotes* L.). *Environ. Sci. Pollut. Res. Int.* 17, 84–96. doi: 10.1007/s11356-008-0094-0

Lysak, M. A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5224–5229. doi: 10.1073/pnas.0510791103

Mahelka, V., Kopecky, D., and Baum, B. R. (2013). Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass *Thinopyrum intermedium* (Triticeae). *Mol. Biol. Evol.* 30, 2065–2086. doi: 10.1093/molbev/mst106

Mahelka, V., Krak, K., Kopeckı, D., Fehrer, J., Šafář, J., Bartoš, J., et al. (2017). Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1726–1731. doi: 10.1073/pnas.1613375114

Maheshwari, S. C. (1958). *Spirodela polyrhiza*: the Link between the Aroids and the Duckweeds. *Nature* 181, 1745–1746. doi: 10.1038/1811745a0

Mayo, S., Bogner, J., and Boyce, P. (1997). *The Genera of Araceae*. London: Royal Botanic Gardens, Kew, UK.

Mayo, S. J., Bogner, J., and Boyce, P. C. (1998). "Araceae," in *Flowering Plants · Monocotyledons: Alismatanae and Commelinanae (except Gramineae)*, ed. K. Kubitzki (Berlin: Springer), 26–74.

Michael, T. P., Bryant, D., Gutierrez, R., Borisjuk, N., Chu, P., Zhang, H., et al. (2017). Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* 89, 617–635. doi: 10.1111/tpj.13400

Moss, T., and Stefanovsky, V. Y. (2002). At the center of eukaryotic life. *Cell* 109, 545–548. doi: 10.1016/s0092-8674(02)00761-4

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321

Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. doi: 10.1146/annurev.genet.39.073003.112240

Paolacci, S., Jansen, M. A. K., and Harrison, S. (2018). Competition between *Lemna minuta*, *Lemna minor*, and *Azolla filiculoides*. Growing fast or being steadfast? *Front. Chem.* 6:207. doi: 10.3389/fchem.2018.00207

Perina, A., Seoane, D., González-Tizón, A. M., Rodríguez-Fariña, F., and Martínez-Lage, A. (2011). Molecular organization and phylogenetic analysis of 5S rDNA in crustaceans of the genus *Pollicipes* reveal birth-and-death evolution and strong purifying selection. *BMC Evol. Biol.* 11:304. doi: 10.1186/1471-2148-11-304

Polanco, C., and De La Vega, M. P. (1997). Intergenic ribosomal spacer variability in hexaploid oat cultivars and landraces. *Heredity* 78, 115–123. doi: 10.1038/hdy.1997.19

Qiu, Y., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., et al. (2000). Phylogeny of basal Angiosperms: analyses of five genes from three genomes. *Int. J. Plant Sci.* 161, S3–S27. doi: 10.1086/317584

Rebordinos, L., Cross, I., and Merlo, A. (2013). High evolutionary dynamism in 5S rDNA of fish: state of the art. *Cytogenet. Genome Res.* 141, 103–113. doi: 10.1159/000354871

Renner, S. S., and Zhang, L.-B. (2004). Biogeography of the *Pistia* clade (Araceae): based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. *Syst. Biol.* 53, 422–432. doi: 10.1080/10635150490445904

Renner, S. S., Zhang, L.-B., and Murata, J. (2004). A chloroplast phylogeny of *Arisaema* (Araceae) illustrates tertiary floristic links between Asia, North America, and East Africa. *Am. J. Bot.* 91, 881–888. doi: 10.3732/ajb.91.6.881

Rezania, S., Taib, S. M., Md Din, M. F., Dahalan, F. A., and Kamyab, H. (2016). Comprehensive review on phytotechnology: heavy metals removal by diverse aquatic plants species from wastewater. *J. Hazard Mater.* 318, 587–599. doi: 10.1016/j.jhazmat.2016.07.053

Rothwell, G. W., Van Atta, M. R., Ballard, H. E. J., and Stockey, R. A. (2004). Molecular phylogenetic relationships among Lemnaceae and Araceae using the chloroplast trnL-trnF intergenic spacer. *Mol. Phylogenet. Evol.* 30, 378–385. doi: 10.1016/s1055-7903(03)00205-7

Sáez-Vásquez, J., and Delseny, M. (2019). Ribosome biogenesis in plants: from functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell* 31, 1945–1967. doi: 10.1105/tpc.18.00874

Sergeeva, E. M., Shcherban, A. B., Adonina, I. G., Nesterov, M. A., Beletsky, A. V., Rakitin, A. L., et al. (2017). Fine organization of genomic regions tagged to the 5S rDNA locus of the bread wheat 5B chromosome. *BMC Plant Biol.* 17:183. doi: 10.1186/s12870-017-1120-5

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., et al. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Res.* 46, 3019–3033. doi: 10.1093/nar/gky163

Sims, J., Sestini, G., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). Sequencing of the Arabidopsis NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* 12:387. doi: 10.1038/s41467-020-20728-6

Sousa, A., Cusimano, N., and Renner, S. S. (2014). Combining FISH and model-based predictions to understand chromosome evolution in *Typhonium* (Araceae). *Ann. Bot.* 113, 669–680. doi: 10.1093/aob/mct302

Sousa, A., and Renner, S. S. (2015). Interstitial telomere-like repeats in the monocot family Araceae. *Botanical J. Linnean Soc.* 177, 15–26. doi: 10.1111/boj.12231

Stadler, M., Stelzer, T., Borisjuk, N., Zanke, C., Schilde-Rentschler, L., and Hemleben, V. (1995). Distribution of novel and known repeated elements of *Solanum* and application for the identification of somatic hybrids among *Solanum species*. *Theor. Appl. Genet.* 91, 1271–1278. doi: 10.1007/BF00220940

Subramanyam, K. (1962). *Aquatic Angiosperms*. New Delhi: CSIR.

Tian, N., Han, L., Chen, C., and Wang, Z. (2018). The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species. *PLoS One* 13:e0192956. doi: 10.1371/journal.pone.0192956

Tippery, N. P., and Les, D. H. (2020). "Tiny plants with enormous potential: phylogeny and evolution of duckweeds," in *The Duckweed Genomes*, eds X. H. Cao, P. Fourounjian, and W. Wang (Cham: Springer International Publishing), 19–38. doi: 10.1007/978-3-030-11045-1_2

Tippery, N. P., Les, D. H., Appenroth, K. J., Sree, K. S., Crawford, D. J., and Bog, M. (2021). Lemnaceae and Orontiaceae are phylogenetically and morphologically distinct from Araceae. *Plants (Basel)* 10:2639. doi: 10.3390/plants10122639

Tippery, N. P., Les, D. H., and Crawford, D. J. (2015). Evaluation of phylogenetic relationships in Lemnaceae using nuclear ribosomal data. *Plant Biol. (Stuttg)* 17(Suppl. 1), 50–58. doi: 10.1111/plb.12203

Tynkevich, I. O., and Volkov, R. A. (2014). Structural organization of 5S ribosomal DNA of Rosa rugosa. *Tsitol. Genet.* 48, 3–9.

Vandenberghe, A., Chen, M.-W., Dams, E., de Baere, R., de Roeck, E., Huysmans, E., et al. (1984). The corrected nucleotide sequences of 5 S RNAs from six angiosperms - with some notes on 5 S RNA secondary structure and molecular evolution. *FEBS Lett.* 171, 17–23. doi: 10.1016/0014-5793(84)80452-4

Vasconcelos, E. V., Vasconcelos, S., Ribeiro, T., Benko-Iseppon, A. M., and Brasileiro-Vidal, A. C. (2018). Karyotype heterogeneity in *Philodendron* s.l. (Araceae) revealed by chromosome mapping of rDNA loci. *PLoS One* 13:e0207318. doi: 10.1371/journal.pone.0207318

Volkov, R. A., Panchuk, I. I., Borisjuk, L. G., and Borisjuk, M. V. (2003). Plant rDNA: organization, evolution, and using. *Tsitol. Genet.* 37, 72–78.

Volkov, R. A., Panchuk, I. I., Borisjuk, N. V., Hosiawa-Baranska, M., Maluszynska, J., and Hemleben, V. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol.* 17:21. doi: 10.1186/s12870-017-0978-6

Yadav, V., Hemansi, Kim, N., Tuteja, N., and Yadav, P. (2017). G Quadruplex in plants: a ubiquitous regulatory element and its biological relevance. *Front. Plant Sci.* 8:1163. doi: 10.3389/fpls.2017.01163

Yeng, W. S., Meerow, A., and Croat, T. (2016). Resurrection and new species of the neotropical genus *Adelonema* (Araceae: philodendron Clade). *Syst. Bot.* 41, 32–48. doi: 10.1600/036364416X690732

Zanis, M. J., Soltis, P. S., Qiu, Y. L., Zimmer, E., and Soltis, D. E. (2003). Phylogenetic analyses and perianth evolution in basal Angiosperms. *Ann. Missouri Botanical Garden* 90, 129–150. doi: 10.2307/3298579

Zanke, C., Borisjuk, N., Ruoss, B., Schilde-Rentschler, L., Ninnemann, H., and Hemleben, V. (1995). A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum*. *Theor. Appl. Genet.* 90, 720–726. doi: 10.1007/BF00222139

Zhang, L., Cheng, H., Pan, D., Wu, Y., Ji, R., Li, W., et al. (2021). One-pot pyrolysis of a typical invasive plant into nitrogen-doped biochars for efficient sorption of phthalate esters from aqueous solution. *Chemosphere* 280:130712. doi: 10.1016/j.chemosphere.2021.130712

Zhou, Y., and Borisjuk, N. (2019). Small aquatic duckweed plants with big potential for production of valuable biomass and wastewater remediation. *Int. J. Environ. Sci. Nat. Resour.* 16, 1–4. doi: 10.19080/IJESNR.2019.16.555942

Zhu, X.-Y., Cai, D.-T., and Ding, Y. (2008). Molecular and cytological characterization of 5S rDNA in *Oryza species*: genomic organization and phylogenetic implications. *Genome* 51, 332–340. doi: 10.1139/G08-016

Zimmels, Y., Kirzhner, F., and Malkovskaja, A. (2006). Application of *Eichhornia crassipes* and *Pistia stratiotes* for treatment of urban sewage in Israel. *J. Environ. Manage.* 81, 420–428. doi: 10.1016/j.jenvman.2005.11.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fpls.2022.852406

5S Ribosomal DNA of Genus *Solanum*: Molecular Organization, Evolution, and Taxonomy

Yurij O. Tynkevich[1†], Antonina Y. Shelyfist[1†], Liudmyla V. Kozub[1†], Vera Hemleben[2†], Irina I. Panchuk[1,2†] and Roman A. Volkov[1†]*

[1] Department of Molecular Genetics and Biotechnology, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine,
[2] Center of Plant Molecular Biology (ZMBP), Eberhard Karls University of Tübingen, Tübingen, Germany

The *Solanum* genus, being one of the largest among high plants, is distributed worldwide and comprises about 1,200 species. The genus includes numerous agronomically important species such as *Solanum tuberosum* (potato), *Solanum lycopersicum* (tomato), and *Solanum melongena* (eggplant) as well as medical and ornamental plants. The huge *Solanum* genus is a convenient model for research in the field of molecular evolution and structural and functional genomics. Clear knowledge of evolutionary relationships in the *Solanum* genus is required to increase the effectiveness of breeding programs, but the phylogeny of the genus is still not fully understood. The rapidly evolving intergenic spacer region (IGS) of 5S rDNA has been successfully used for inferring interspecific relationships in several groups of angiosperms. Here, combining cloning and sequencing with bioinformatic analysis of genomic data available in the SRA database, we evaluate the molecular organization and diversity of IGS for 184 accessions, representing 137 species of the *Solanum* genus. It was found that the main mechanisms of IGS molecular evolution was step-wise accumulation of single base substitution or short indels, and that long indels and multiple base substitutions, which arose repeatedly during evolution, were mostly not conserved and eliminated. The reason for this negative selection seems to be association between indels/multiple base substitutions and pseudogenization of 5S rDNA. Comparison of IGS sequences allowed us to reconstruct the phylogeny of the *Solanum* genus. The obtained dendrograms are mainly congruent with published data: same major and minor clades were found. However, relationships between these clades and position of some species (*S. cochoae, S. clivorum, S. macrocarpon,* and *S. spirale*) were different from those of previous results and require further clarification. Our results show that 5S IGS represents a convenient molecular marker for phylogenetic studies on the *Solanum* genus. In particular, the simultaneous presence of several structural variants of rDNA in the genome enables the detection of reticular evolution, especially in the largest and economically most important sect. *Petota*. The origin of several polyploid species should be reconsidered.

Keywords: 5S rDNA, genomics, molecular evolution, hybridization, polyploidy, taxonomy, *Solanum*

OPEN ACCESS

Edited by:
Peter Poczai,
University of Helsinki, Finland

Reviewed by:
Natalia Borowska-Zuchowska,
University of Silesia in Katowice, Poland
Josep A. Rossello,
Jardí Botànic UV, Spain
Hannes Becher,
University of Edinburgh, United Kingdom

***Correspondence:**
Roman A. Volkov
r.volkov@chnu.edu.ua

†ORCID:
Yurij O. Tynkevich
orcid.org/0000-0002-0222-8098
Antonina Y. Shelyfist
orcid.org/0000-0001-5711-3362
Liudmyla V. Kozub
orcid.org/0000-0002-2675-5896
Vera Hemleben
orcid.org/0000-0002-5171-472X
Irina I. Panchuk
orcid.org/0000-0002-2837-4480
Roman A. Volkov
orcid.org/0000-0003-0673-2598

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *11 January 2022*
Accepted: *25 February 2022*
Published: *13 April 2022*

Citation:
Tynkevich YO, Shelyfist AY, Kozub LV, Hemleben V, Panchuk II and Volkov RA (2022) 5S Ribosomal DNA of Genus Solanum: Molecular Organization, Evolution, and Taxonomy. Front. Plant Sci. 13:852406. doi: 10.3389/fpls.2022.852406

INTRODUCTION

Regions coding for 5S rRNA (rDNA) are present in genomes of all cellular organisms. In eukaryotes, 5S rDNA belongs to the class of moderately repeated sequences and is represented by hundreds or thousands of copies of tandemly arranged repeated units (repeats). 5S rDNA clusters are mostly located in one or two chromosomes, although multiple loci are also found (Volkov et al., 2004; Hasterok et al., 2006; Garcia et al., 2012; Bustos et al., 2020; Vozárová et al., 2021). In contrast to majority of repeated sequences whose functions largely remain uncertain, the activity of 5S rDNA is vital for cells, providing rRNA indispensable for assembly of functional ribosomes. The copy number of rDNA repeats is higher than what is required for rRNA synthesis, and redundant copies of rDNA are transcriptionally silenced (Volkov et al., 2004; Tucker et al., 2010; Layat et al., 2012; Matyasek et al., 2016).

Each 5S rDNA repeat consists of an evolutionarily conserved region encoding 5S rRNA (coding sequence, CDS) and a rapidly evolving intergenic spacer (IGS) (Volkov et al., 2001; Ishchenko et al., 2018; Simon et al., 2018; Qin et al., 2019). The high evolutionary stability of the CDS is the result of purifying selection to maintain the function of 5S rRNA as a component of a large ribosome subunit (Vizoso et al., 2011; Mahelka et al., 2013).

Transcription of 5S rDNA is provided by RNA polymerase III (Pol III) and corresponding transcription factors (TFs). The Pol III promoter consists of internal and external elements. The internal elements of the promoter, A-Box, IE, and C-Box, are located in the CDS and represent targets for TFs, which are necessary for the recruitment of Pol III to the transcription initiation complex (Douet and Tourmente, 2007; Layat et al., 2012). Respectively, mutations in internal elements of the promoter should not only disturb the structure of 5S rRNA and ribosome but also affect binding of TFs to the promoter and, thus, the expression of 5S rDNA.

In contrast to the CDS, the main part of the IGS is not transcribed and probably does not have any function. Accordingly, it is believed that any mutation in the IGS is selectively neutral; therefore, this region evolves with a high rate. However, it was found that the IGS of *Arabidopsis thaliana* contains short-sequence motifs involved in initiation (external elements of the promoter) and termination (terminator) of 5S rDNA transcription (Douet and Tourmente, 2007; de Souza et al., 2020). Thus, one would expect that these regions are to be relatively conservative, as has been demonstrated in several taxonomic groups (Falistocco et al., 2007; Tynkevich and Volkov, 2014; Tynkevich et al., 2015; Mlinarec et al., 2016; Ishchenko et al., 2018; Alexandrov et al., 2021). However, the existing knowledge of the organization of external promoter elements and their molecular evolution is still incomplete.

In many diploid species, numerous copies of rDNA repeats in the same genome tend to be nearly identical because of sequence homogenization (Volkov et al., 1999b, 2001, 2003), i.e., individual copies of repeated elements do not evolve independently but in a concerted manner (Arnheim et al., 1980; Coen et al., 1982). To explain the high intragenomic similarity of 5S rDNA repeats, the "concerted evolution" and "birth and death" hypotheses were proposed. The mechanisms and intensity of homogenization may differ in different taxa and for different groups of repeated sequences, e.g., for 5S and 35S rDNA (Pinhal et al., 2011; Vizoso et al., 2011; Song et al., 2012; Mahelka et al., 2013; Galián et al., 2014; Barman et al., 2016; Volkov et al., 2007, 2017; Chen et al., 2021).

Additive inheritance of both parent variants of 5S and 35S rDNA is usually observed in the first generation of interspecific hybrids/allopolyploids. However, in ancient allopolyploids, 35S rDNA can be the subject of interlocus sequence conversion (Volkov et al., 1999a,b, 2007; Vizoso et al., 2011), while different variants of 5S rDNA can coexist in a plant genome for a long time without being homogenized (Fulnecek et al., 2002; Song et al., 2012; Mahelka et al., 2013; Mlinarec et al., 2016; Volkov et al., 2017; Ishchenko et al., 2018). Especially, a significant sequence divergence was found for spatially distant 5S rDNA variants that are located in different loci of the same chromosome set (Cronn et al., 1996; Vozárová et al., 2021) and for different parent loci in genomes of hybrid origin (Fulnecek et al., 2002; Matyasek et al., 2002), while repeats from the same locus appeared to be highly homogenized. Accordingly, 5S rDNA became an attractive focus for investigation of molecular evolution of repeated sequences, identification of hybrids, and phylogenetic studies on angiosperms (Blöch et al., 2009; Baum et al., 2012; Simeone et al., 2018; Tynkevich and Volkov, 2019; Ishchenko et al., 2020, 2021; Cardoni et al., 2021; Vozárová et al., 2021). However, 5S rDNA is still poorly characterized in many important plant groups such as the *Solanum* L. genus.

The *Solanum* (nightshade) genus is an attractive model for comparative genomics and investigation of molecular evolution of repeated sequences. With around 1,200 species, it belongs to the so-called "giant genera" and is the fifth largest genus of flowering plants (Frodin, 2004; Echeverría-Londoño et al., 2020). *Solanum* species are distributed worldwide from tropical to temperate areas and grow under diverse ecological conditions. Most *Solanum* species inhabit the New World, although secondary centers of diversity have been found in Africa, Asia, and Australia (D'Arcy, 1991). Overall, the *Solanum* genus is an example of unusual hyperdiversity in life forms, morphological features, and ecological preferences, representing a unique system for studying the diversification of plants (Knapp et al., 2004; Echeverría-Londoño et al., 2020). The genus includes important crops such as *S. tuberosum* L. (potato), *S. lycopersicum* L. (tomato), and *S. melongena* L. (brinjal eggplant, aubergine), about 20 cultivated species of local significance like *S. aethiopicum* L. (Ethiopian eggplant), *S. betaceum* Cav. (tamarillo), *S. muricatum* Aiton (pepino), and *S. quitoense* Lam. (lulo), as well as several medicinal and ornamental plants (*S. marginatum* L.f., *S. aviculare* G. Forst., *S. mammosum* L., and *S. pseudocapsicum* L.).

In the Solanaceae family, the *Solanum* genus belongs to the strongly supported large "x = 12" clade (Olmstead et al., 2008). The most common chromosome number in *Solanum* is x = 12, which occurs in 97% of species examined, such as diploids (77%), tetraploids (14%), hexaploids (4%), triploids (2%), and

octoploids (0.2%). Application of *in situ* hybridization showed that 55 out of 64 (85.9%) diploid species possess only one 5S locus per chromosome set (Chiarini et al., 2018). Up to now, the molecular organization and evolution of the 5S rDNA in the genus *Solanum* have only been analyzed in about 35 species and breeding lines (Volkov et al., 2001; Davidjuk et al., 2010; Sun et al., 2014). In this study, combining cloning and sequencing with analysis of available genomic data in the Sequence Read Archive (SRA) public database, we evaluate the molecular organization, diversity, and evolution of the IGS for 184 plant accessions, representing 137 species across the *Solanum* genus. Especially, our results shed a new light on the phylogeny of the genus and reticulate evolution of the largest and economically most important sect. *Petota*.

MATERIALS AND METHODS

Plant Material and DNA Extraction

A plant material of the *Solanum* species was obtained from several collections (see **Tables 1**, **2**). A plant material of out-group species, *Lycianthes lycioides* (L.) Hassl. and *Physalis peruviana* L. (acc. no. NK-03), was obtained from Orto Botanico di Padova (Italy) and National Botanical Garden of National Academy of Sciences of Ukraine (Kyiv, Ukraine), respectively.

Total genomic DNA was isolated from herbarium specimens according to the CTAB method of DNA extraction (Porebski et al., 1997). In addition, DNA was treated with Proteinase K (Sigma-Aldrich, United States).

Amplification, Cloning, and Sequencing of 5S rDNA Repeats

Repeated units of 5S rDNA were amplified using the primers Pr5S-L and Pr5S-R, complementary to the 5S rRNA CDS. These primers provide amplification of complete 5S IGS and flanking regions of the CDS (Volkov et al., 2001). PCR amplification was performed as described previously (Tynkevich and Volkov, 2019). PCR products were ligated into plasmid vector pJET 1.2/blunt using CloneJET PCR Cloning Kit (Thermo Fisher Scientific, United States). Screening of recombinant clones and size selection of inserts were performed by colony PCR with pJET 1.2 forward and reverse primers. Two to eight clones per plant accession were Sanger-sequenced by LGC Genomics (Germany). Primary processing of nucleotide sequences and calculation of sequence similarity levels were performed using the Chromas software and the DNASTAR software package. The obtained sequences were deposited in the GenBank database under the accession numbers listed in **Table 3**.

Assembly of 5S rDNA Repeats From Illumina Short Reads

De novo assembly of 5S rDNA repeats was performed using libraries of pre-filtered paired or single Illumina reads from raw data of *Solanum* species genomes available in SRA (**Tables 3**, **4**). Read filtering was carried out using the built-in tool on the sequence download page: https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=search_seq_name. To filter reads containing 5S rDNA fragments, 20-bp long fragments of CDS were used for matching. *De novo* assembly was conducted using SeqMan NGen 14 (DNASTAR Lasergene suite). Libraries of filtered reads were automatically trimmed for quality, and the following assembly parameters were used: mer size 31, minimum match percentage 100%, and coverage threshold 100 reads. In the obtained contigs with highest coverage from 2 to 12, 5S rDNA repeats that contain one full IGS flanked by two fragments of CDS were identified and collected.

Prediction of 5S rRNA Secondary Structure

Hypothetical secondary structures of potential 5S rRNA transcripts were predicted using the Fold online tool in the RNAstructure server (Reuter and Mathews, 2010).[1] Lowest free energy structures were calculated using the following default parameters: temperature (in K) 310.15; maximum loop size 30; minimum helix length 3.

Median-Joining Network and Phylogenetic Analysis

Relationships among IGS sequences of the *Solanum* species were analyzed applying the median-joining network approach implemented in SplitsTree 5 (Huson and Bryant, 2006). Alignments of the IGS sequences were performed in the MAFFT server using the G-INS-I method, which is most suitable for sequences with global homology (Katoh et al., 2019).

For alignment of the IGS sequences of *Solanum* species belonging to different taxonomic groups in the genus, we applied the E-INS-I method implemented in the MAFFT server (Katoh et al., 2019). The generated alignment was checked and adjusted manually with the UGENE software.

The best-fit nucleotide substitution model was estimated with the lowest value of Bayesian Information Criterion (BIC) using the Find Best-Fit Substitution Model tool in Mega X (Kumar et al., 2018). A maximum likelihood (ML) phylogenetic tree was generated with the PhyML plugin for Geneious Prime 2021.0.3.[2] The IGS sequences of *L. lycioides* and *Ph. peruviana* produced in this study (acc. nos. OM100793-4 and OM744711-3) as well as those of four *Capsicum* species available in GenBank (*C. baccatum* L.: AF217951, *C. frutescens* L.: AF217952, *C. chinense* Jacq.: AF217953, and *C. pubescens* Ruiz and Pav.: AF217954) were used as outgroups. Branch support was calculated by approximate likelihood ratio tests, aLRT-Chi2 (Anisimova and Gascuel, 2006), and bootstrap analysis with 1,000 resampling replicates. Phylogenetic analysis was also performed by Bayesian inference using the MrBayes 2.2.4 plugin for Geneious Prime 2021.0.3. Four independent Monte Carlo Markov Chains (MCMCs) of 1,000,000 iterations each were run to generate phylogenetic trees with Bayesian posterior probabilities. Trees were sampled every 500 generations. The resulting trees were exported in Newick format and annotated using "Interactive tree of life" (iTOL v6).

[1] https://rna.urmc.rochester.edu/RNAstructureWeb/

[2] www.geneious.com

TABLE 1 | List of *Solanum* species analyzed (excluding sect. *Petota*).

Species name	Taxonomy		Chromosome number, 2 n	Abbreviation	Plant material	
	Nee, 1999 (Subgenus-Section)	Särkinen et al., 2013 (Clades)			Accession No	Source
S. abutiloides (Griseb.) Bitter and Lillo	Solanum-Brevantherum	Leptostemonum-Brevantherum	24	abu	19682363	MBG
S. aculeatissimum Jacq.	Leptostemonum-Acanthophora	Leptostemonum-Acanthophora	24	acu	79p515	WABG
S. aethiopicum L.	Leptostemonum-Melongena	Leptostemonum Old World	24	aet	SAMN 10986202	PRJNA 523664
S. albostellatum R.W. Davis and P.J.H. Hurter	Not indicated	Not indicated	nd	als	SAMN 10969051	PRJNA 522689
S. americanum Mill.	Solanum-Solanum	Morelloid	24	ame1	SAMEA 3486921	PRJEB 9916
				ame2	SAMEA 3486922	PRJEB 9916
				ame3	SAMEA 7573861	PRJEB 38240
S. anguivi Lam.	Not indicated	Leptostemonum Old World	24	ang	SAMN 16746499	PRJNA 676007
S. anomalostemon S. Knapp and M. Nee	Not indicated	Not indicated	24	ano	SAMEA 7820352	PRJEB 42506
S. appendiculatum Dunal	Solanum-Anarrhichomenum	Potato-Anarrhichomenum	24	ape1	SAMN 12623209	PRJNA 561636
				ape2	SAMN 12623212	PRJNA 561636
S. aviculare G. Forst.	Solanum-Archaesolanum	Archaesolanum	46	avi	19771009	MBG
S. betaceum Cav.	Bassovia-Pachyphylla	Leptostemonum-Cyphomandra	24	bet	–	BGUT
S. chrysotrichum Schltdl.	Not indicated	Not indicated	24	chr	SAMN 08770449	PRJNA 438407
S. clarkiae Symon	Not indicated	Leptostemonum Old World	24	cla	SAMN 12161630	PRJNA 551615
S. cleistogamum Symon	Leptostemonum-Melongena	Leptostemonum Old World	24	cle	SAMN 10969163	PRJNA 522689
S. clivorum S. Knapp	Solanum-Holophylla	Not indicated	nd	cli	SAMEA 7820346	PRJEB 42506
S. cochoae G.J. Anderson and Bernardello	Solanum-Basarthrum	Not indicated	24	coc	SAMEA 7820347	PRJEB 42506
S. crinitum Lam.	Leptostemonum-Crinitum	Leptostemonum-Androceras/Crinitum	24	cri	74s1231	WABG
S. dimorphandrum S. Knapp	Not indicated	Not indicated	24	dim	SAMEA 7820348	PRJEB 42506
S. diversiflorum F. Muell.	Not indicated	Leptostemonum Old World	24	div	SAMN 10969025	PRJNA 522689
S. dulcamara L.	Solanum-Dulcamara	Dulcamaroid	24	dul	96065	BGUT
S. elatius A.R. Bean	Not indicated	Not indicated	24	ela	SAMN 10969339	PRJNA 522689
S. erianthum D. Don	Solanum-Brevantherum	Not indicated	24	eri	SAMN 08770591	PRJNA 438407
S. esuriale Lindl.	Not indicated	Leptostemonum Old World	24, 48	esu	SAMN 10969026	PRJNA 522689
S. ferocissimum Lindl.	Leptostemonum-Melongena	Leptostemonum Old World	24, 48	fer	SAMN 10969027	PRJNA 522689
S. guamense Merr.	Not indicated	Not indicated	nd	gua	81s39	WABG
S. hindsianum Benth.	Leptostemonum-Melongena	Leptostemonum-Elaeagnifolium	24	hin	–	LDZG
S. horridum Dunal ex Poir.	Not indicated	Leptostemonum Old World	24	hor	SAMN 10969028	PRJNA 522689
S. incanum L.	Not indicated	Leptostemonum Old World	24	inc	SAMN 07303451	PRJNA 392603

(Continued)

TABLE 1 | (Continued)

Species name	Taxonomy		Chromosome number, 2 n	Abbreviation	Plant material	
	Nee, 1999 (Subgenus-Section)	Särkinen et al., 2013 (Clades)			Accession No	Source
S. laciniatum Aiton	Solanum-Archaesolanum	Archaesolanum	92	lac	SAMEA 7820351	PRJEB 42506
S. lasiophyllum Humb. and Bonpl. ex Dunal	Not indicated	Not indicated	24, 48	las	SAMN 10969030	PRJNA 522689
S. linnaeanum Hepper and P.-M. L. Jaeger	Not indicated	Leptostemonum Old World	24	lin	SAMN 13023229	PRJNA 577305
S. macrocarpon L.	Leptostemonum-Melongena	Leptostemonum Old World	24	mac	SAMN 16746492	PRJNA 676007
S. mammosum L.	Leptostemonum-Acanthophora	Leptostemonum-Acanthophora	22	mam	–	BGUT
S. medicagineum A.R. Bean	Not indicated	Not indicated	nd	mdg	SAMN 12096241	PRJNA 533457
S. melongena L.	Leptostemonum-Melongena	Leptostemonum Old World	24	mel1	cultivar Black Beauty	VASSMA Ltd.
				mel2	SAMN 13023228	PRJNA 577305
				mel3	SAMN 07303456	PRJNA 392603
S. muricatum Aiton	Solanum-Basarthrum	Potato-Basarthrum	24	mur	–	BGUT
S. nigrum L.	Solanum-Solanum	Morelloid	24, 48, 72	nig	SAMN 17035829	PRJNA 683719
S. ossicruentum Martine and J. Cantley	Not indicated	Not indicated	48	oss	SAMN 12161629	PRJNA 533451
S. pachyandrum Bitter	Leptostemonum-Herposolanum	Not indicated	24	pac	SAMEA 7820344	PRJEB 42506
*S. paposanum*Phil.	Not indicated	Potato-Regmandra	24	pap	SAMEA 7820349	PRJEB 42506
S. phlomoides A. Cunn. ex Benth.	Not indicated	Leptostemonum Old World	24, 48	phl	SAMN 10969029	PRJNA 522689
S. pseudocapsicum L.	Solanum-Holophylla	Leptostemonum-Geminata	24	pse	–	BGChNU
S. pseudolulo Heiser	Not indicated	Leptostemonum-Lasiocarpa	24	psl	XX-GZU-88100737	BGUG
S. quitoense Lam.	Not indicated	Leptostemonum-Lasiocarpa	24	qui	XX-GZU-00120822	BGUG
S. scabrum Mill.	Solanum-Solanum	Morelloid	72	sca	SAMN 08456262	PRJNA 432637
*S. seaforthianum*Andrews	Solanum-Dulcamara	Not indicated	24	sea	74p1254	WABG
S. sejunctum Brennan, Martine and Symon		Leptostemonum Old World	24	sej	SAMN 12161632	PRJNA 551616
S. sisymbriifolium Lam.	Leptostemonum-Melongena	Leptostemonum-Sisymbriifolium	24	sis	SAMN 16746501	PRJNA 676007
S. spirale Roxb.	Not indicated	Not indicated	48	spi	SAMN 08770592	PRJNA 438407
S. torvum Sw.	Leptostemonum-Torva	Leptostemonum-Torva	24, 48	trv	SAMN 16746498	PRJNA 676007
S. valdiviense Dunal	Not indicated	Unclear	24	val	SAMEA 7820350	PRJEB 42506
S. vespertilio ssp. *vespertilio* Aiton	Leptostemonum-Melongena	Leptostemonum Old World	24	ves	–	BGUT
S. villosum Mill.	Solanum-Solanum	Morelloid	48	vil	–	BGChNU
S. wendlandii Hook.f.	Leptostemonum-Herposolanum	Leptostemonum-Allophyllum and Wendlandii	24	wen	77c37	WABG
S. wrightii Benth.	Leptostemonum-Crinitum	Leptostemonum-Androceras/Crinitum	24	wri	SAMN 16746495	PRJNA 676007

Taxonomy is shown according to Nee (1999) and Särkinen et al. (2013). Chromosome numbers are presented according to the Chromosome Counts Database (CCDB; http://ccdb.tau.ac.il/). Plant material sources: BGChNU, Botanical Garden of the Chernivtsi National University, Ukraine; BGUG, Botanical Garden of the University of Graz, Austria; BGUT, Botanical Garden of the University of Tübingen, Germany; LDZG, Living Desert Zoo and Gardens, California, United States; MBG, Meise Botanical Garden, Belgium; WABG, Waimea Arboretum and Botanical Garden, Hawaii, United States. PRJNA and PRJEB are the BioProject accession numbers in GenBank (https://www.ncbi.nlm.nih.gov/bioproject/).

TABLE 2 | List of analyzed *Solanum* species of sect. *Petota*.

Species name	Taxonomy		Chromosome number, 2 n	Plant material		
	Hawkes, 1990, Nee, 1999 (Subsection-Series)	Huang et al., 2019 (Clade)		Abbre-viation	Accession No	Source
S. abancayense Ochoa	Potatoe-Tuberosa (ii)	Clade 4 north	24	abn	SAMN 07540430	PRJNA 394943
S. acaule Bitter	Potatoe-Acaulia	Not indicated	48	acl	–	CIP
S. achacachense Cardenas	Potatoe-Tuberosa (iii)	Clade 4 north	24	ach	SAMN 07540512	PRJNA 394943
S. acroglossum Juz.	Potatoe-Piurana	Clade 3	24	acg	SAMN 07540377	PRJNA 394943
S. acroscopicum Ochoa	Potatoe-Tuberosa (ii)	Clade 3	24	acs	SAMN 07540369	PRJNA 394943
S. ahanhuiri Juz. and Bukasov	Potatoe-Tuberosa (cult.)	Not indicated	24	ajh	SAMN 12684889	PRJNA 556263
S. albornozii Correll	Potatoe-Piurana	Clade 3	24	abz	SAMN 07540378	PRJNA 394943
S. ambosinum Ochoa	Potatoe-Tuberosa (ii)	Clade 4 north	24	amb1	SAMN 07540480	PRJNA 394943
				amb2	SAMN 07540482	PRJNA 394943
S. andreanum Baker	Potatoe-Tuberosa (i)	Clade 3	24, 48	adr	SAMN 07540382	PRJNA 394943
S. arcanum Peralta	Estolonifera-Neolycopersicon	Not indicated	24	arc	SAMEA 2335233	PRJEB 5226
S. avilesii Hawkes and Hjert.	Potatoe-Tuberosa (iii)	Clade 4 south	24	avl	SAMN 07540476	PRJNA 394943
S. berthaultii Hawkes	Potatoe-Tuberosa (iii)	Clade 4 south	24	ber1	BGRC 18548	GDC
				ber2	SAMN 07540477	PRJNA 394943
S. blanco-galdosii Ochoa	Potatoe-Piurana	Clade 3	24	blg	SAMN 07540379	PRJNA 394943
S. boliviense Dunal	Potatoe-Megistacroloba	Not indicated	24	blv	SAMN 06564709	PRJNA 378971
S. brevicaule Bitter	Potatoe-Tuberosa (iii)	Clade 4 south	24, 48, 72	brc	SAMN 07540508	PRJNA 394943
S. bukasovii Juz. ex Rybin	Potatoe-Tuberosa (ii)	Clade 4 north	24	buk1	BGRC N 15424	GDC
				buk2	SAMN 07540400	PRJNA 394943
				buk3	SAMN 07540415	PRJNA 394943
				buk4	SAMN 07540419	PRJNA 394943
				buk5	SAMN 07540466	PRJNA 394943
				buk6	SAMN 07540519	PRJNA 394943
				buk7	SAMN 07540520	PRJNA 394943
S. bukasovii f. *multidissectum* (Hawkes) Ochoa	Potatoe-Tuberosa (ii)	Clade 4 north	24	bukm1	SAMN 07540390	PRJNA 394943
				bukm2	SAMN 07540456	PRJNA 394943
				bukm3	SAMN 07540457	PRJNA 394943
S. bulbocastanum Dunal	Potatoe-Bulbocastana	Clade 1+2	24	blb1	BGRC N 08006	GDC
				blb2	SAMN 07540359	PRJNA 394943
S. cajamarquense Ochoa	Potatoe-Tuberosa (ii)	Clade 4 south	24	cjm	SAMN 07540364	PRJNA 394943
S. canasense Hawkes	Potatoe-Tuberosa (ii)	Clade 4 north	24	can	SAMN 07540552	PRJNA 394943
S. candolleanum P. Berthault	Potatoe-Tuberosa (iii)	Not indicated	24	cnd	SAMN 06564692	PRJNA 378971
S. cardiophyllum Lindl.	Potatoe-Pinnatisecta	Clade 1+2	24, 36	cph	SAMN 07540547	PRJNA 394943
S. chacoense Bitter	Potatoe-Yungasensa	Clade 4 south	24	chc1	B2	MPI
			24, 36	chc2	SAMN 07540432	PRJNA 394943
S. chaucha Juz. and Bukasov	Potatoe-Tuberosa (cult.)	Not indicated	72	cha	SAMN 12684891	PRJNA 556263
S. cheesmaniae (L. Riley) Fosberg	Estolonifera-Neolycopersicon	Not indicated	24	che	SAMEA 2340812	PRJEB 5235
S. chilense (Dunal) Reiche	Estolonifera-Neolycopersicon	Not indicated	24	chi	SAMEA 2340822	PRJEB 5235
S. chmielewskii (C.M. Rick et al.) D.M. Spooner et al.	Estolonifera-Neolycopersicon	Not indicated	24	cml	SAMEA 2340810	PRJEB 5235
S. chomatophilum Bitter	Potatoe-Conicibaccata	Clade 3	24	chm	SAMN 07540374	PRJNA 394943
S. circaeifolium subsp. *quimense* Hawkes and Hjert.	Potatoe-Circaeifolia	Not indicated	24	crc	BGRC N 27036	GDC
S. commersonii Dunal	Potatoe-Commersoniana	Not indicated	24	cmm1	BGRC N 17654	GDC
			24, 36	cmm2	SAMN 06564712	PRJNA 378971

(Continued)

TABLE 2 | (Continued)

Species name	Taxonomy		Chromosome number, 2 n	Plant material		
	Hawkes, 1990, Nee, 1999 (Subsection-Series)	Huang et al., 2019 (Clade)		Abbre-viation	Accession No	Source
S. corneliomuelleri J.F. Macbr.	Estolonifera-Neolycopersicon	Not indicated	24	crm	SAMEA 2340786	PRJEB 5235
S. curtilobum Juz. and Bukasov	Potatoe-Tuberosa (cult.)	Not indicated	60	cur	SAMN 12684896	PRJNA 556263
S. demissum Lindl.	Potatoe-Demissa	Not indicated	72	dms	–	CIP
S. ehrenbergii (Bitter) Rydb.	Potatoe-Pinnatisecta	Not indicated	24	ehr	SAMN 06564745	PRJNA 378971
S. etuberosum Lindl.	Estolonifera-Etuberosa	Outgroup	24	etb	SAMN 07540542	PRJNA 394943
S. galapagense S.C. Darwin and Peralta	Estolonifera-Neolycopersicon	Not indicated	24	gal	SAMEA 2340846	PRJEB 5235
S. gourlayi Hawkes	Potatoe-Tuberosa (iii)	Clade 4 south	24	grl1	5.6	GFP
				grl2	SAMN 07540506	PRJNA 394943
S. habrochaites S. Knapp and D.M. Spooner	Estolonifera-Neolycopersicon	Not indicated	24	hab	SAMEA 2340830	PRJEB 5235
S. hondelmannii Hawkes and Hjert.	Potatoe-Tuberosa (iii)	Clade 4 south	24	hdm	SAMN 07540500	PRJNA 394943
S. huaylasense Peralta	Estolonifera-Neolycopersicon	Not indicated	24	hua	SAMEA 2340821	PRJEB 5235
S. hypacrarthrum Bitter	Potatoe-Piurana	Clade 3	24	hcr	SAMN 07540375	PRJNA 394943
S. incamayoense K.A. Okada and A.M. Clausen	Potatoe-Tuberosa (iii)	Clade 4 south	24	inm	SAMN 07540492	PRJNA 394943
S. infundibuliforme Phil.	Potatoe-Cuneoalata	Not indicated	24	ifd	SAMN 06564699	PRJNA 378971
S. iopetalum (Bitter) Hawkes	Potatoe-Demissa	Not indicated	72	iop	GLSK 161	IPK
S. jamesii Torr.	Potatoe-Pinnatisecta	Clade 1+2	24	jam1	BGRC N 10054	GDC
				jam2	SAMN 07540363	PRJNA 394943
S. juzepczukii Bukasov	Potatoe-Tuberosa (cult.)	Not indicated	36	juz	SAMN 12684892	PRJNA 556263
S. kurtzianum Bitter and Wittm.	Potatoe-Tuberosa (iii)	Clade 4 south	24	ktz	SAMN 07540435	PRJNA 394943
S. laxissimum Bitter	Potatoe-Conicibaccata	Clade 4 north	24	lxs1	GLKS 154.3	IPK
				lxs2	SAMN 07540550	PRJNA 394943
S. leptophyes Bitter	Potatoe-Tuberosa (ii)	Clade 4 north	24	lph	8.27	GFP
S. limbaniense Ochoa	Potatoe-Conicibaccata	Clade 4 north	24	lmb	SAMN 07540465	PRJNA 394943
S. lycopersicoides Dunal	Estolonifera-Juglandifolia	Not indicated	24	lpd	SAMN 10809628	PRJNA 516877
S. lycopersicum L.	Estolonifera-Neolycopersicon	Outgroup	24	lyc1	-	-
				lyc2	SAMN 15097861	PRJNA 637170
				lyc3	SAMN 11163599	PRJNA 527863
S. lycopersicum var. *cerasiforme* (Dunal) D.M. Spooner et al.	Estolonifera-Neolycopersicon	Not indicated	24	lycc1 lycc2	SAMN 09229594 SAMN 09229698	PRJNA 454805 PRJNA 454805
S. maglia Schltdl.	Potatoe-Maglia	Clade 4 south	36	mag	BGRC N032571	GDC
S. marinasense Vargas	Potatoe-Tuberosa (ii)	Clade 4 north	24	mrn	SAMN 07540408	PRJNA 394943
S. medians Bitter	Potatoe-Tuberosa (ii)	Clade 4 north	24, 36	med	SAMN 07540469	PRJNA 394943
S. megistacrolobum Bitter	Potatoe-Megistacroloba	Clade 4 south	24	mga	SAMN 07540385	PRJNA 394943
S. microdontum Bitter	Potatoe-Tuberosa (iii)	Clade 4 south	24	mcd1	BGRC 27351	GDC
			24, 36	mcd2	SAMN 07540501	PRJNA 394943
S. multiinterruptum Bitter	Potatoe-Tuberosa (ii)	Clade 3	24	mtp	SAMN 07540388	PRJNA 394943
S. neorickii D.M. Spooner, G.J. Anderson and R.K. Jansen	Estolonifera-Neolycopersicon	Not indicated	24	neo	SAMEA 2340816	PRJEB 5235
S. neorossii Hawkes and Hjert.	Potatoe-Tuberosa (iii)	Not indicated	24	nrs	11.42	GFP

(Continued)

TABLE 2 | (Continued)

Species name	Taxonomy		Chromosome number, 2 n	Plant material		
	Hawkes, 1990, Nee, 1999 (Subsection-Series)	Huang et al., 2019 (Clade)		Abbre-viation	Accession No	Source
S. okadae Hawkes and Hjert.	Potatoe-Tuberosa (iii)	Not indicated	24	oka1	BGRC 17550	GDC
				oka2	BGRC 24719	GDC
				oka3	SAMN 06564702	PRJNA 378971
S. palustre Poepp. ex Schltdl.	Estolonifera-Etuberosa	Outgroup	24	pal1	BGRC N 17441	GDC
				pal2	SAMN 07540543	PRJNA 394943
S. pampasense Hawkes	Potatoe-Tuberosa (ii)	Clade 4 north	24	pam	SAMN 07540427	PRJNA 394943
S. *paucissectum* Ochoa	Potatoe-Piurana	Clade 3	24	pcs	SAMN 07540376	PRJNA 394943
S. pennellii Correll	Estolonifera-Neolycopersicon	Not indicated	24	pen	SAMN 14984469	PRJNA 557253
S. peruvianum L.	Estolonifera-Neolycopersicon	Not indicated	24	per	SAMEA 2340809	PRJEB5235
S. phureja Juz. and Bukasov	Potatoe-Tuberosa (cult.)	Cultivated	24	phu1	IVP 101	CPBR
				phu2	SAMN 07540523	PRJNA 394943
S. pimpinellifolium L.	Estolonifera-Neolycopersicon	Not indicated	24	pim	SAMN 09229654	PRJNA 454805
S. pinnatisectum Dunal	Potatoe-Pinnatisecta	Clade 1+2	24	pnt1	BGRC N 08168	GDC
				pnt2	SAMN 07540354	PRJNA 394943
S. polyadenium Greenm.	Potatoe-Polyadenia	Clade 1+2	24	pld1	BGRC N 08176	GDC
				pld2	SAMN 07540357	PRJNA 394943
S. raphanifolium Cardenas and Hawkes	Potatoe-Megistacroloba	Not indicated	24	rap1	BGRC N 07207	GDC
				rap2	BGRC N 08189	GDC
				rap3	SAMN 06564696	PRJNA 378971
S. sitiens I.M. Johnst.	Estolonifera-Juglandifolia	Not indicated	24	sit	SAMN 14932980	PRJNA 633104
S. sogarandinum Ochoa	Potatoe-Megistacroloba	Clade 3	24	sgr1	SAMN 07540395	PRJNA 394943
				sgr2	SAMN 07540416	PRJNA 394943
S. sparsipilum (Bitter) Juz. and Bukasov	Potatoe-Tuberosa (ii)	Clade 4 south	24, 48	spl1	14.9	GFP
				spl2	SAMN 07540479	PRJNA 394943
S. spegazzinii Bitter	Potatoe-Tuberosa (iii)	Clade 4 south	24	spg1	17.45	GFP
				spg2	SAMN 07540411	PRJNA 394943
S. stenophyllidium Bitter	Potatoe-Pinnatisecta	Clade 1+2	24	ste	SAMN 07540355	PRJNA 394943
S. stenotomum Juz. and Bukasov	Potatoe-Tuberosa (cult.)	Cultivated	24	stn1	–	CIP
				stn2	SAMN 07540540	PRJNA 394943
S. stenotomum subsp. *goniocalyx* (Juz. and Bukasov) Hawkes	Potatoe-Tuberosa (cult.)	Cultivated	24	gon	SAMN 07540541	PRJNA 394943
S. stoloniferum Schltdl. and C.D.Bouché	Potatoe-Longipedicellata	Not indicated	48	sto	SAMEA 4949197	PRJEB 28862
S. tarijense Hawkes	Potatoe-Yungasensa	Clade 4 south	24	trj	SAMN 07540392	PRJNA 394943
S. tuberosum L.	Potatoe-Tuberosa (iii)	Breading lines	24	tbr1	B15	BLBP
			24	tbr2	R1	RAGIS
			24	tbr3	BP1076	Bio
			24	tbr4	B1	BLBP
S. tuberosum subsp. andigena (Juz. and Bukasov) Hawkes	Potatoe-Tuberosa (iii)	Not indicated	24, 36, 48	tbrA1	SAMN 06564721	PRJNA 378971
				tbrA2	SAMN 06564717	PRJNA 378971

(Continued)

TABLE 2 | (Continued)

Species name	Taxonomy		Chromosome number, 2 n	Plant material		
	Hawkes, 1990, Nee, 1999 (Subsection-Series)	Huang et al., 2019 (Clade)		Abbre-viation	Accession No	Source
S. venturii Hawkes and Hjert.	Potatoe-Tuberosa (iii)	Clade 4 south	24	vnt	SAMN 07540366	PRJNA 394943
				vrn1	–	GDC
				vrn2	SAMN 07540493	PRJNA 394943
				vrn3	SAMN 07540514	PRJNA 394943
S. verrucosum Schltdl.	Potatoe-Tuberosa (i)	Clade 4 south	24, 36, 48	ver	SAMN 07540496	PRJNA 394943
S. violaceimarmoratum Bitter	Potatoe-Conicibaccata	Clade 4 north	24	vio	SAMN 07540551	PRJNA 394943

Taxonomy and species name abbreviations are shown according to Hawkes (1990), Nee (1999), and Huang et al. (2019). Chromosome numbers are presented according to the Chromosome Counts Database (CCDB; http://ccdb.tau.ac.il/). Plant material sources: IPK, the Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany; GDC, German-Dutch Curatorium for Plant Genetic Resources, Braunschweig, Germany; MPI, Max-Planck-Institute für Züchtungforschung, Köln; GFP, Gesellschaft zur Förderung der Pflanzenzüchtung, Bonn, Germany; CPBR, Center for Plant Breeding and Reproduction Research CPRO, Wageningen, The Netherlands; CIP, Centro Internacional de la Papa, Lima Peru; BLBP, Bayrische Landesanstalt für Bodenkultur und Pflanzenbau, Freising, Germany. PRJNA and PRJEB are the BioProject accession numbers in GenBank (https://www.ncbi.nlm.nih.gov/bioproject/).

RESULTS

Cloning of 5S rDNA Repeats

5S rDNA repeats of 17 *Solanum* species representing different taxonomic groups were amplified by PCR using primers complementary to the coding region, cloned, and sequenced (**Table 3** and **Supplementary Material**). Analysis of the obtained sequences showed that majority of the clones contained IGS flanked on both sides by fragments of the coding region including the primers used for PCR. Besides, we obtained 5S rDNA clones of *S. vespertilio* and *S. pseudocapsicum* that contain rDNA dimers, i.e., two adjacent copies of IGS, and the whole sequence of the CDS between them. Also, two clones containing 5S rDNA dimers and one clone containing a trimer were sequenced for *S. wendlandii.*

Intragenomic Diversity of Intergenic Spacer: In-Depth Analysis of Sequence Read Archive Data

In order to assess the intragenomic variability of 5S rDNA, we evaluated how many different types/variants of repeated units (ribotypes) are present in genomes of the *Solanum* species. Genomes of three diploid species, *S. lycopersicum-3* (SRX5538725), *S. stenotomum-2* (SRX4645231) of sect. *Petota,* and *S. melongena-2* (SRX6995029) of sect. *Melongena,* were selected for detailed analysis. For these genomes, we assembled *de novo* 5S rDNA repeats composed of complete IGS and two flanking fragments of CDS. If the CDS contained indels or several SNP, the repeat was considered a pseudogene and excluded from further analysis. Variants of IGS that differed in at least one SNP were considered as distinct ribotypes. The total number of IGS ribotypes was 45, 177, and 31 in *S. lycopersicum-3, S. stenotomum-2*, and *S. melongena-2*, respectively. In order to visualize the intragenomic diversity of the ribotypes found in the three species, median-joining networks were constructed (**Figure 1**).

After that, we mapped the reads of complete genomic libraries to the reference sequences of all collected ribotypes in order to estimate their relative content in the genomes. The obtained results showed that the IGS ribotypes differ significantly in this parameter. Accordingly, we classified the ribotypes as major (≥10% of all IGS copies present in the genome), minor (<10 but ≥5%), or rare (<5%). The number of major, minor, and rare ribotypes is 5, 0, and 40 in *S. lycopersicum-3*: 3, 5 and 169 in *S. stenotomum-2*; 2, 5, and 24 in *S. melongena-2* (**Figure 1**). Altogether, the major and minor ribotypes represent 93, 68, and 78% of all rDNA repeats present in the genomes of these three species. Based on the results obtained, in the further analysis of 5S rDNA in other species, we considered only major and minor ribotypes. The variability of IGS sequences in each examined sample is given in **Supplementary Material**.

Length and GC Content of the 5S rDNA Repeated Units

Using sequences of clones and major + minor ribotypes, we determined GC content in the IGS of the *Solanum* species (**Tables 3**, **4**) and found that this value ranges from 40.5% in *S. seaforthianum* to 63.9% in *S. pseudocapsicum*. In 90% of the species, intragenomic difference in GC content between individual ribotypes and clones was less than 4%. A greater difference was observed in repeats that were subjected to deletions, particularly in the AT-rich region of the IGS. No significant changes in GC content were found for taxonomic groups in the *Solanum* genus, suggesting that this parameter remained relatively constant during evolution.

The typical length of IGS in members of the *Solanum* genus is about 190–220 bp (**Tables 3**, **4**). The shortest IGS were found in *S. cochoae*, 155–158 bp, and in *S. aethiopicum*, 162–175 bp. In *S. lasiophyllum*, however, one ribotype (las-C2R1) is even shorter, 115 bp, although five other ribotypes in this species are 180 bp in length. The longest IGSs were found in *S. melongena*, 344–360 bp, and in *S. lycopersicum*, 234–235 bp. The extremely long IGS length in *S. melongena* is associated with large duplication of the spacer sequence. There is no significant difference in IGS length among the taxonomic groups in the *Solanum* genus. In general,

TABLE 3 | Characteristics of the 5S intergenic spacer region (IGS) of *Solanum* species analyzed (excluding sect. *Petota*).

Species name	Abbreviation	Clade—Figure 7	Sequencing	TNS	SRA/clone No	GC content, %	IGS length, bp	SIM, %
S. abutiloides	abu	2.1	CS	2	OM100771-2	54.7	214	96.7
S. aculeatissimum	acu	2.3.1	CS	5	OM100773-7	48.24	189	96.8–100
S. aethiopicum	aet	2.3.3D	GA	4	SRX5438534	48.25	165	90.3–98.9
S. albostellatum	als	2.3.3B	GA	3	SRX5462807	55.37	193	92.4–94.4
S. americanum	ame1	1.1	GA	7	ERX1043111	49.41	225	93.4–98.2
	ame2	1.1	GA	6	ERX1043123	48.8	223	92.5–99.6
	ame3	1.1	GA	2	ERX4706760	49.8	226	99.6
S. anguivi	ang	2.3.3D	GA	8	SRX9473543	49.3	189	89.6–99
S. anomalostemon	ano	2.1	GA	5	ERX4907182	57.02	215	96.7–99.5
S. appendiculatum	ape1	1.4.1	GA	3	SRX6763530	47.57	219	85.5–96.4
	ape2	1.4.1	GA	6	SRX6763552	47.42	220	84.1–99.1
S. aviculare	avi	1.3	CS	3	OM100778-80	42.13	208	98.6–99.5
S. betaceum	bet	2.2	CS	2	OM100795-6	53.2	187	98.4
S. chrysotrichum	chr	2.3.2	GA	3	SRX4043085	49	185	94.1–97.8
S. clarkiae	cla	2.3.3A	GA	5	SRX6376308	56.14	199	97.5–99
S. cleistogamum	cle	2.3.3C	GA	4	SRX5462725	52.48	174	78.8–95
S. clivorum	cli	1.4.1	GA	4	ERX4907176	43.75	209	98.6–99.5
S. cochoae	coc	2.2	GA	5	ERX4907177	61.28	156	96.2–99.4
S. crinitum	cri	2.3	CS	2	OM100781-2	53	184	97.3
S. dimorphandrum	dim	1.2	GA	4	ERX4907178	50.25	167	90.8–99.4
S. diversiflorum	div	2.3.3A	GA	4	SRX5462955	56.1	198	97–98.5
S. dulcamara	dul	1.2	CS	5	AJ226026-30	57.68	221	96–99.6
S. elatius	ela	2.3.3B	GA	4	SRX5462442	53.6	199	93.5–99
S. erianthum	eri	2.1	GA	6	SRX4043227	53.18	213	94.8–98.1
S. esuriale	esu	2.3.3B	GA	4	SRX5462952	55.9	184	96.2–98.4
S. ferocissimum.	fer	2.3.3C	GA	3	SRX5462953	49.73	203	98.5–99
S. guamense	gua	2.3.2	CS	4	OM100797-800	48.35	185	93–96.8
S. hindsianum.	hin	2.3.3	CS	2	OM100783, OM744710	55.45	258	99
S. *horridum*	hor	2.3.3C	GA	6	SRX5462950	51.38	175	82.5–97.8
S. incanum	inc	2.3.3D	GA	5	SRX2977430	50.78	206	96.6–99
S. laciniatum	lac	1.3	GA	7	ERX4907181	43.33	206	81.8–99.5
S. lasiophyllum	las	2.3.3C	GA	6	SRX5462948	48	169	59.4–98.9
S. linnaeanum	lin	2.3.3D	GA	3	SRX6995030	49.03	207	96.2–99.5
S. macrocarpon	mac	2.3.3B	GA	6	SRX9473554	46.42	177	91.6–97.8
S. mammosum	mam	2.3.1	CS	2	OM100801-2	54.45	203	99.5
S. medicagineum	mdg	2.3.3C	GA	7	SRX6095227	50.84	171	97.1–99.4
S. melongena	mel1	2.3.3D	CS	3	HM042870-1, OM100803	49.37	198	56–99.6
	mel2	2.3.3D	GA	8	SRX6995029	49.58	338	51.4–99.7
	mel3	2.3.3D	GA	3	SRX2977427	50	349	95.2–99.4
S. muricatum	mur	1.4.1	CS	2	OM100804-5	45.5	209	99
S. nigrum	nig	1.1	GA	4	SRX9654460	49.68	226	97.8–99.6
S. ossicruentum	oss	2.3.3B	GA	6	SRX6376307	52.82	193	85.4–98
S. pachyandrum	pac	2.2	GA	6	ERX4907174	48.58	210	83.4–98.6
S. paposanum	pap	1	GA	4	ERX4907179	53.23	226	97.8–99.1
S. phlomoides	phl	2.3.3A	GA	4	SRX5462951	53.7	179	87.4–98.4
S. pseudocapsicum	pse	2.2	CS	3	OM100784-5	63.9	173	95.3–97.1
S. pseudolulo	psl	2.3.1	CS	3	OM100806-8	49.87	219	81.2–98.3
S. quitoense	qui	2.3.1	CS	4	OM100809-12	49.6	198	71.4–99.6
S. scabrum	sca	1.1	GA	4	SRX3641602	46.85	227	92.5–96
S. seaforthianum	sea	1	CS	3	OM100813-5	40.47	230	89.6–93
S. sejunctum	sej	2.3.3A	GA	4	SRX6376309	53.63	202	97–99
S. sisymbriifolium	sis	2.3	GA	2	SRX9473545	54.75	211	99.5

(Continued)

TABLE 3 | (Continued)

Species name	Abbreviation	Clade—Figure 7	Sequencing	TNS	SRA/clone No	GC content, %	IGS length, bp	SIM, %
S. *spirale*	spi	2.3.3D	GA	6	SRX4043228	49.03	191	88.7–99
S. torvum	trv	2.3.2	GA	5	SRX9473542	49.16	185	95.1–99.5
S. valdiviense	val	1.3	GA	3	ERX4907180	51.5	211	98.6–99.5
S. vespertilio	ves	2.3.3D	CS	3	OM100816-7	49.17	204	92.2–93.7
S. villosum	vil	1.1	CS	2	OM100818-9	48.65	226	96.9
S. *wendlandii*	wen	2.2	CS	8	OM100786-9	48.99	220	58.4–96.9
S. wrightii	wri	2.3	GA	1	SRX9473557	53	183	100

Methods used for generation of sequences: CS, cloning and sequencing; DS, direct sequencing of PCR product; GA, genomic assembly; TNS, total number of 5S IGS sequences (clones or ribotypes) analyzed in this study; SIM, intragenomic similarity between clones/ribotypes. For GC content and length of IGS, average values are shown.

our data show that the length remained largely unchanged during the evolution of the *Solanum* genus.

Long Duplication in the Intergenic Spacer of *S. melongena*

Two structural variants of IGS, long (∼350 bp) and short (∼200 bp) were identified in *S. melongena-2* (mel2). The long variant was found in three accessions, mel1 (analyzed by cloning and sequencing) and in mel2 and mel3 (extracted from SRA), while the short variant was only detected in mel1 and mel2. In mel2, all major and minor as well as majority of rare ribotypes belong to the long variant, while the short variant is only represented by two rare ribotypes, M17 and M29 (**Figure 1B**), whose relative content in the genome is below 1%.

Numerous single nucleotide polymorphisms (SNPs) and two oligonucleotide indels are present in the M29 sequence, so this ribotype appears to be a pseudogene. The ribotype M17 (mel2-C17R1, see **Supplementary Material**) also contains several SNPs compared to other ribotypes of *S. melongena*. Interestingly, this ribotype is identical to the most common ribotype of a closely related species, *S. linnaeanum*.

A detailed sequence analysis showed that the long variant contains a 146-bp-long tandem duplication in the central part of the IGS (**Figure 2**). The duplicated region consists of a 32-bp-long 3′-fragment of the coding region and an adjacent 114-bp fragment of the IGS. Two copies of the duplicated segment differ by 6 SNPs and one 8-bp-long indel. All mutations are localized in the fragment of the IGS, not in the coding region.

High Diversity of 5S rDNA in *S. wendlandii*

For the 5S rDNA of *S. wendlandii*, we sequenced four clones, pSowen-3,-13,-14, and-18, which bore inserts of different lengths, 732, 912, 319, and 657 bp, respectively. Sequence analysis showed that the shortest insert contains one copy of IGS flanked by CDS fragments. The longer inserts represent two dimers and a trimer composed of adjacent copies of 5S rDNA repeats (**Figure 3A**).

The sequence alignment revealed an obvious difference among IGS sequences of the adjacent 5S rDNA copies (**Figure 3B**), which is due to numerous nucleotide substitutions and insertions of different lengths of 1–82 bp. The 82-bp-long insertion harbors three tandem copies of the adjacent sequence, which is normally present once in the IGS. The level of sequence similarity among the compared IGS copies ranges from 58.4 to 96.9%, which indicates high intragenomic heterogeneity of the IGS in *S. wendlandii*.

Comparison of the 5S rRNA CDS of *S. wendlandii* and several *Solanum* species representing different intrageneric clades revealed that the CDS is, as expected, highly conserved in the genus. Analysis of the 5S rDNA clones/ribotypes of several *Solanum* species showed that a single CDS usually contains no more than two mutations compared to the respective consensus sequence (data not shown), which agrees well with the observation on other plant taxa (Park et al., 2000; Mahelka et al., 2013). In contrast, the complete CDS sequences of *S. wendlandii* each contain 5–16 base substitutions (**Figure 4A**).

The presence of numerous mutations in the CDS suggests its transformation into a pseudogene. To test this possibility, we calculated the secondary structure for transcripts of the complete CDS from the clones pSowen-3,-13, and-18. For comparison, the secondary structure was also calculated for (i) the total consensus CDS of the *Solanum* genus, (ii) consensus CDS of *S. melongena*, which differs from the total consensus by one base substitution, and (iii) CDS of *S. pseudocapsicum* (dimer clone pSpse-5S7), which contains two base substitutions (**Figure 4B**). The sequences examined formed a secondary structure typical for 5S rRNA (Sun and Caetano-Anollés, 2009), with the exception of the CDS of *S. wendlandii*, which appeared to be significantly changed, suggesting that the transcripts cannot fulfill their function in the ribosome.

Hence, the 5S rDNA of *S. wendlandii* appears to be very heterogeneous in both the IGS and CDS regions and likely contains numerous pseudogenes. Unfortunately, the complete genome sequence of *S. wendlandii* is currently not presented in the GeneBank and cannot, therefore, be used to further elucidate the unusual organization of 5S rDNA in this species.

Intergenic Spacer Organization in Distantly Related *Solanum* Species

To reveal the molecular organization and evolution of IGS in *Solanum*, we compared the IGS sequences of 37 species representing distantly related groups (D'Arcy, 1991; Nee, 1999; Bohs, 2005; Särkinen et al., 2013) of the genus (**Figure 5**). The total length of the alignment obtained is 287 bp. Only 9

TABLE 4 | Characteristics of the 5S IGS of *Solanum* species of sect. *Petota*.

Species name	Abbreviation	Cluster—Figure 8	Sequencing	TNS	SRA/clone No	GC content, %	Length, bp	SIM, %
S. abancayense	abn	A3	GA	7	SRX4645060	49.09	222	96–99.6
S. acaule	acl	D10	CS	4	AJ226031-34	49.43	219	97.3–99.5
S. achacachense	ach	D1, D7	GA	5	SRX4645061	51.06	208	87.1–99.1
S. acroglossum	acg	A5	GA	7	SRX4645064	51.04	223	96.4–99.6
S. acroscopicum	acs	A4	GA	11	SRX4645063	50.76	224	95.6–99.6
S. ahanhuiri	ajh	D10	GA	4	SRX6963077	49.2	219	95.9–99.5
S. albornozii	abz	A1	GA	5	SRX4645065	46.92	195	94.1–99
S. ambosinum	amb1	D1	GA	4	SRX4645068	49.18	207	89.2–98.6
	amb2	D1, D10	GA	12	SRX4645070	49.25	218	94.1–99.5
S. andreanum	adr	A4	GA	3	SRX4645073	48.83	214	97.7–98.6
S. arcanum	arc	A6	GA	3	ERX376595	45.37	231	97.4–98.3
S. avilesii	avl	D1	GA	7	SRX4645077	52.07	203	72.5–99.1
S. berthaultii	ber1	D5	CS	5	AJ226037-41	50.9	213	98.1–100
	ber2	D5	GA	7	SRX4645079	49.24	206	73.3–98.6
S. blanco-galdosii	blg	A5	GA	5	SRX4645082	50.28	220	96.4–99.5
S. boliviense	blv	D6	GA	4	SRX2646030	50.4	214	96.3–98.1
S. brevicaule	brc	D6	GA	4	SRX4645091	50.13	211	93.4–99.1
S. bukasovii	buk1	A3	DS	1	AF332130	48.2	222	nd
	buk2	A3	GA	4	SRX4645092	48.3	222	98.6–99.5
	buk3	D8, D9	GA	5	SRX4645093	49.02	214	96.7–99.5
	buk4	D8, D9	GA	8	SRX4645094	48.75	208	90.2–99.5
	buk5	D4	GA	5	SRX4645095	50.14	213	98.6–99.5
	buk6	A3, D10	GA	5	SRX4645098	48.58	220	84.7–99.6
	buk7	D6, D8, D10	GA	12	SRX4645099	50.16	214	93.2–99.1
S. bukasovii f. *multidissectum*	bukm1	A3, D6	GA	5	SRX4645184	49.94	212	86.9–99.5
	bukm2	D1, D10	GA	5	SRX4645190	49.86	216	94.5–99.5
	bukm3	D7	GA	4	SRX4645191	50.13	214	97.7–99.5
S. bulbocastanum	blb1	nd	CS	3	AJ226012-14	50.73	189	98.4–99.5
	blb2	nd	GA	3	SRX4645100	51.93	188	97.9–98.4
S. cajamarquense	cjm	D2	GA	4	SRX4645102	50.9	223	96–99.6
S. canasense	can	D1, D7	GA	4	SRX4645113	51.83	205	90.6–99.1
S. candolleanum	cnd	D3, D6	GA	4	SRX2646047	50.1	213	94.4–98.6
S. cardiophyllum	cph	A1	GA	5	SRX4645116	51.86	224	96.4–99.1
S. chacoense	chc1	D3	DS	1	AF331055	50.7	213	nd
	chc2	D3	GA	9	SRX4645120	53.16	213	90.1–99.1
S. chaucha	cha	A3, D8, D10	GA	10	SRX6966567	49.02	217	83.4–99.6
S. cheesmaniae	che	A6	GA	4	ERX384387	46.08	232	97.4–99.1
S. chilense	chi	A6	GA	4	ERX384397	46.7	230	97–99.6
S. chmielewskii	cml	A6	GA	5	ERX384385	44.95	232	95.3–97
S. chomatophilum	chm	A5	GA	5	SRX4645123	51.32	223	96.4–99.6
S. circaeifolium subsp. *quimense*	crc	A4	CS	8	AJ226015-22	49.73	227	94.3–100
S. commersonii	cmm1	A5	DS	1	AF331056	51.3	224	nd
	cmm2	D3	GA	4	SRX2646027	50.05	220	97.3–99.1
S. corneliomuelleri	crm	A6	GA	3	ERX384361	46.87	222	89.5–96.1
S. curtilobum	cur	A2, D1, D2, D9, D10	GA	8	SRX6966568	50.09	218	83.3–99.6
S. demissum	dms	D10	CS	3	AJ226023-25	49.03	219	98.6–99.5
S. ehrenbergii	ehr	A2	GA	5	SRX2645991	50.08	208	83.1–99.1
S. etuberosum	etb	A1	GA	7	SRX4645124	48.21	223	94.6–99.1
S. galapagense	gal	A6	GA	1	ERX384421	46.4	233	100
S. gourlayi	grl1	D6	DS	1	AF331057	51.8	213	nd
	grl2	D6	GA	3	SRX4645138	50.7	214	97.2–99.5
S. habrochaites	hab	A6	GA	3	ERX384405	45.63	233	96.6–97.9
S. hondelmannii	hdm	D6	GA	5	SRX4645145	50.56	217	90.1–98.3

(Continued)

TABLE 4 | (Continued)

Species name	Abbreviation	Cluster—Figure 8	Sequencing	TNS	SRA/clone No	GC content, %	Length, bp	SIM, %
S. huaylasense	hua	A6	GA	1	ERX384396	46.7	229	100
S. hypacrarthrum	hcr	A1	GA	5	SRX4645148	45.96	202	89.3–99.5
S. incamayoense	inm	D3, D6	GA	4	SRX4645153	49.43	215	90.4–99.1
S. infundibuliforme	ifd	D6	GA	4	SRX2646040	49.9	213	97.2–99.5
S. iopetalum	iop	D6	CS	4	AJ226042-45	49.0	212	96.2–99.1
S. jamesii	jam1	B	DS	1	AF331058	50.2	213	nd
	jam2	B	GA	2	SRX4645155	50.0	212	99.1
S. juzepczukii	juz	A2, D2	GA	8	SRX6966566	51.68	221	82.9–99.6
S. kurtzianum	ktz	D3	GA	6	SRX4645157	51.28	214	78.2–99.1
S. laxissimum	lxs1	C	CS	5	AJ226046-50	49.1	202	97–100
	lxs2	C	GA	4	SRX4645163	49.33	202	94.6–99.5
S. leptophyes	lph	D6	DS	1	AF331059	49.8	213	nd
S. limbaniense	lmb	D1, D10	GA	8	SRX4645171	50.01	219	91.5–99.6
S. lycopersicoides	lpd	A6	GA	2	SRX5301957	45.25	229	92.4
S. lycopersicum	lyc1	A6	CS	1	X55697	46	235	nd
	lyc2	A6	GA	5	SRX8467710	45.3	233	93.6–98.7
	lyc3	A6	GA	5	SRX5538725	45.66	233	98.3–99.6
S. lycopersicum var. *cerasiforme*	lycc1	A6	GA	4	SRX4183310	46.05	234	97.4–99.6
	lycc2	A6	GA	1	SRX4183171	45.7	234	100
S. maglia	mag	D3	CS	8	AF331047-54	47.35	189	72.1–100
S. marinasense	mrn	D9, D10	GA	4	SRX4645173	46.35	201	66.7–98.2
S. medians	med	A2, D1	GA	11	SRX4645178	49.6	216	73.8–99.6
S. megistacrolobum	mga	D2	GA	9	SRX4645179	51.3	213	88.2–99.5
S. microdontum	mcd1	D3	CS	9	AJ226051-59	50.27	210	91.2–100
	mcd2	D3, D6	GA	7	SRX4645250	48.43	199	68.8–98.6
S. multiinterruptum	mtp	A3	GA	4	SRX4645183	48.63	222	97.3–99.1
S. neorickii	neo	A6	GA	5	ERX384391	45.5	230	95.7–98.7
S. neorossii	nrs	D3	DS	1	AF331060	50.2	213	nd
S. okadae	oka1	D4	CS	3	AJ226060-62	50.33	220	96.4–99.5
	oka2	D1, D4	CS	4	AJ226063-66	51.28	206	72.3–98.6
	oka3	D1	GA	3	SRX2646037	52.57	204	97.1–98
S. palustre	pal1	A1	CS	6	AJ226035-63	53.57	174	66.4–100
	pal2	A1	GA	4	SRX4645193	47.95	224	97.3–99.1
S. pampasense	pam	D1, D8	GA	5	SRX4645197	50.36	210	85.5–96.7
S. paucissectum	pcs	A5	GA	5	SRX4645198	49.26	210	69.1–99.1
S. pennellii	pen	A6	GA	4	SRX8371122	47.05	229	96.9–99.1
S. peruvianum	per	A6	GA	4	ERX384384	46.88	230	98.3–99.6
S. phureja	phu1	D1	DS	1	AF331061	50.0	212	nd
	phu2	D1, D8	GA	6	SRX4645199	49.55	213	96.2–99.5
S. pimpinellifolium	pim	A6	GA	4	SRX4183091	46.45	229	98.3–99.1
S. pinnatisectum	pnt1	B	CS	5	X82779, AJ226008-11	49.24	210	92.4–96.2
	pnt2	B	GA	18	SRX3115796	50.18	211	83.5–99.5
S. polyadenium	pld1	A1	CS	3	AF331044-6	49.97	197	86.9–96.6
	pld2	A1	GA	6	SRX4645210	50.17	204	82.2–99.1
S. raphanifolium	rap1	C	DS	1	AF332131	50.0	172	nd
	rap2	A3	DS	2	AF332132-3	50.45	201	73.5
	rap3	C	GA	8	SRX2646043	50.76	176	88.8–98.9
S. sitiens	sit	A6	GA	5	SRX8537919	45.4	229	97.4–99.6
S. sogarandinum	sgr1	A1	GA	10	SRX4645211	51.15	212	54.9–99.1
	sgr2	A1	GA	6	SRX4645212	45.47	204	83.5–99.6
S. sparsipilum	spl1	D6	DS	1	AF331062	49.1	216	nd
	spl2	D6	GA	5	SRX4645216	50.8	206	87.3–98.6
S. spegazzinii	spg1	D3	DS	1	AF331063	52.2	205	nd

(Continued)

TABLE 4 | (Continued)

Species name	Abbreviation	Cluster—Figure 8	Sequencing	TNS	SRA/clone No	GC content, %	Length, bp	SIM, %
	spg2	D3	GA	6	SRX4645219	51.28	219	95.4–99.1
S. stenophyllidium	ste	A2	GA	3	SRX3115797	50.47	219	98.6–99.5
S. stenotomum	stn1	D6	DS	1	AF331064	47.5	200	nd
	stn2	D1, D9	GA	8	SRX4645231	50.64	215	94–99.5
S. stenotomum subsp. *goniocalyx*	gon	D8	GA	4	SRX4645128	49.53	213	96.7–98.6
S. stoloniferum	sto	D1, D3	GA	6	ERX2825240	50.48	203	86–99.5
S. tarijense	trj	D5	GA	7	SRX4645232	51.16	211	92.5–99.1
S. tuberosum	tbr1	D3	CS	3	X82780, Y16650-51	52.37	205	99.5–100
	tbr2	D10	CS	1	X82781	49.1	216	nd
	tbr3	D3	CS	4	Y16652-55	52.33	205	99.5–100
	tbr4	D1	CS	4	Y16656-59	49.45	213	92–98.1
	tbrA1	D6, D8, D9	GA	5	SRX2646018	50.12	214	93.5–99.1
	tbrA2	D3, D10	GA	4	SRX2646022	51.1	212	90.4–99.5
S. venturii	vnt	D3, D4	GA	8	SRX4645146	49.8	218	93.6–99.5
S. vernei	vrn1	D7	DS	1	AF332129	50.0	202	nd
	vrn2	D7	GA	4	SRX4645247	51.45	201	83.1–98.6
	vrn3	D7	GA	9	SRX4645251	51.63	204	70.1–99.1
S. verrucosum	ver	D6	GA	7	SRX4645248	49.83	213	93.9–99.1
S. violaceimarmoratum	vio	C	GA	4	SRX4645256	48.4	203	91.1–98.5

Methods used for generation of sequences: CS, cloning and sequencing; DS, direct sequencing of PCR product; GA, genomic assembly; TNS, total number of 5S IGS sequences (clones or ribotypes) analyzed in this study; SIM, intragenomic similarity between clones/ribotypes. For GC content and length of IGS, average values are shown. Sequences generated by cloning for members of sect. Petota were obtained from our previous publication (Volkov et al., 2001).

identical nucleotides were found in the compared sequences, and average pairwise identity value was 55.1%, indicating significant divergence of the IGS in the genus. Multiple base substitutions and indels of various lengths are scattered along the entire IGS in the species studied compared to the consensus sequence. The largest 31-bp-long indel is located in the central part of the IGS between the positions 144 and 174 bp. Despite numerous species-specific mutations, the sequence of the central indel shows an obvious sequence similarity in the species compared. Analysis of the phylogenetic dendrogram obtained by comparing IGS sequences (**Figure 6**, see also below) revealed that the central indel is present in the species belonging to major clade 1 (with the exception of *S. muricatum*) but is partially or completely absent in members of clade 2 (with the exception of *S. anomalostemon*). Hence, the central indel was present in the common ancestor of the *Solanum* genus and was later lost in some species during the course of evolution.

Intergenic Spacer Organization in Sect. *Petota*

Analysis of IGS molecular organization in the species-rich sect. *Petota* was performed separately. By sequence comparison, numerous base substitutions and indels were detected, which appear to be randomly distributed along the IGS (**Figure 7**), except for the presumptive external promoter region just upstream of the CDS (see section "Discussion").

The alignment of the sequences revealed that in the central part of the IGS there are two group-specific indels, I and II. Also, a lot of species contain a GC-duplication (GC-DUP) in the IGS (**Figure 7**). It is likely that these structural rearrangements occurred in different stages during the evolution of sect. *Petota.* With regard to the presence/absence of these molecular features, four structural variants (SVs) of the IGS can be distinguished. The evolutionary ancestral SV-A contains both specific indels, while independent deletions of indels II and I resulted in the formation of the derived SV-B and –C/D, respectively. SV-D additionally contains a GC-DUP.

Phylogenetic Analysis

The IGS sequences obtained by cloning as well as major and minor ribotypes extracted from SRA were used to reconstruct the phylogenetic relationships among *Solanum* species representing different taxonomic groups of the genus. For sect. *Petota*, seven species were selected whose IGS sequences belong to different structural classes (see section "Discussion").

Multiple sequence alignment for the whole genus *Solanum* phylogeny was generated with the Mafft E-INS-I method and then manually corrected. The final 609-bp length alignment presented only one identical site, with an average pairwise identity of 54.7%. The best-fit phylogenetic model was estimated using Mega X to be general time-reversible (GTR) + gamma (G) (Kumar et al., 2018). The obtained maximum likelihood (ML) phylogenetic tree has 302 leaves, which correspond to the IGS sequences of 65 *Solanum* species (**Figure 6**). Calculating the statistical support applying the aLRT-Chi2 method and bootstrap analysis showed that majority of the tree's nodes have a high or moderate support. The ML tree mostly matched the dendrogram generated by Bayesian inference.

On the dendrograms, all investigated species of the *Solanum* genus form a well-supported monophyletic group with *L. lycioides* as sister taxon. In the ML-tree, the *Solanum*

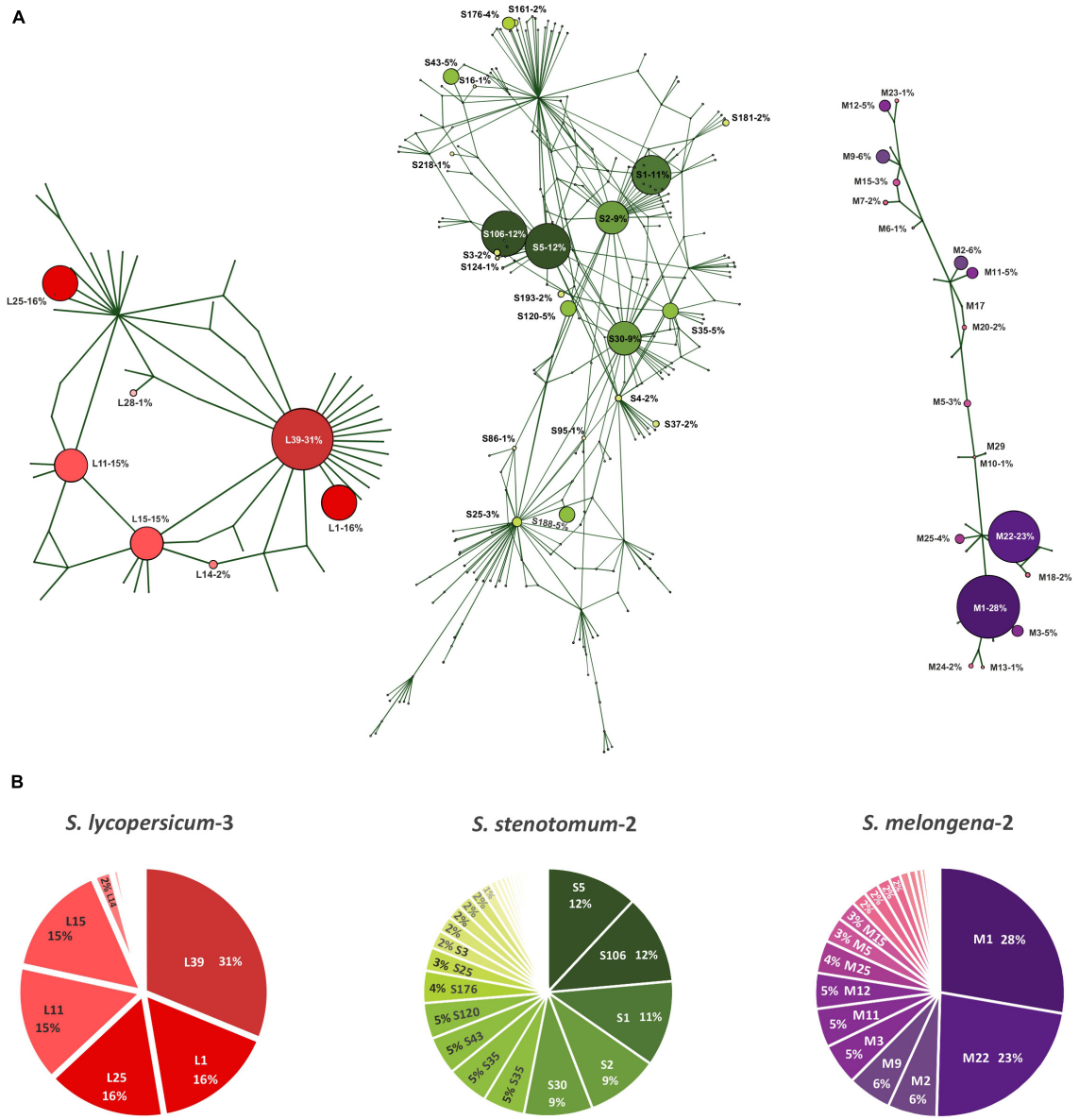


FIGURE 1 | Analysis of 5S rDNA intragenomic diversity in diploid *Solanum* species. **(A)** Median joining networks for IGS types/variants (ribotypes) of *S. lycopersicum-3*, *S. stenotomum-2*, and *S. melongena-2*. Ribotypes are designated by the first letter of corresponding species name with index numbers. The size of the circles is proportional to the relative content (in %) of each ribotype in the genome. **(B)** Relative content (in %) of ribotypes.

species are divided into two major clades, 1 and 2, with high statistical support. In the clades, several well-supported minor clades were found. The monophyly of the *Solanum* genus and clade 1 is also confirmed by Bayesian inference. In contrast, clade 2 is represented by polytomy in the Bayesian dendrogram.

DISCUSSION

Phylogeny of the *Solanum* Genus

Since the nineteenth century, the *Solanum* genus has been traditionally divided into two main groups, the so-called spiny and non-spiny solanums (Dunal, 1852; Seithe, 1962), which were further subdivided into sections, subsections, and series using morphological characters (D'Arcy, 1991; Nee, 1999). However, application of molecular methods shed a new light on the phylogeny of *Solanum*, demonstrating that these groups are mainly not monophyletic, and that the genus can be divided into 13 clades (Bohs, 2005; Weese and Bohs, 2007; Särkinen et al., 2013). Some of these clades have high statistical support, while the taxonomic placement and composition of the others are uncertain.

Analysis of several chloroplast genes and nuclear regions (e.g., ITS1/2 and *waxy*) is often performed in molecular phylogenetics.

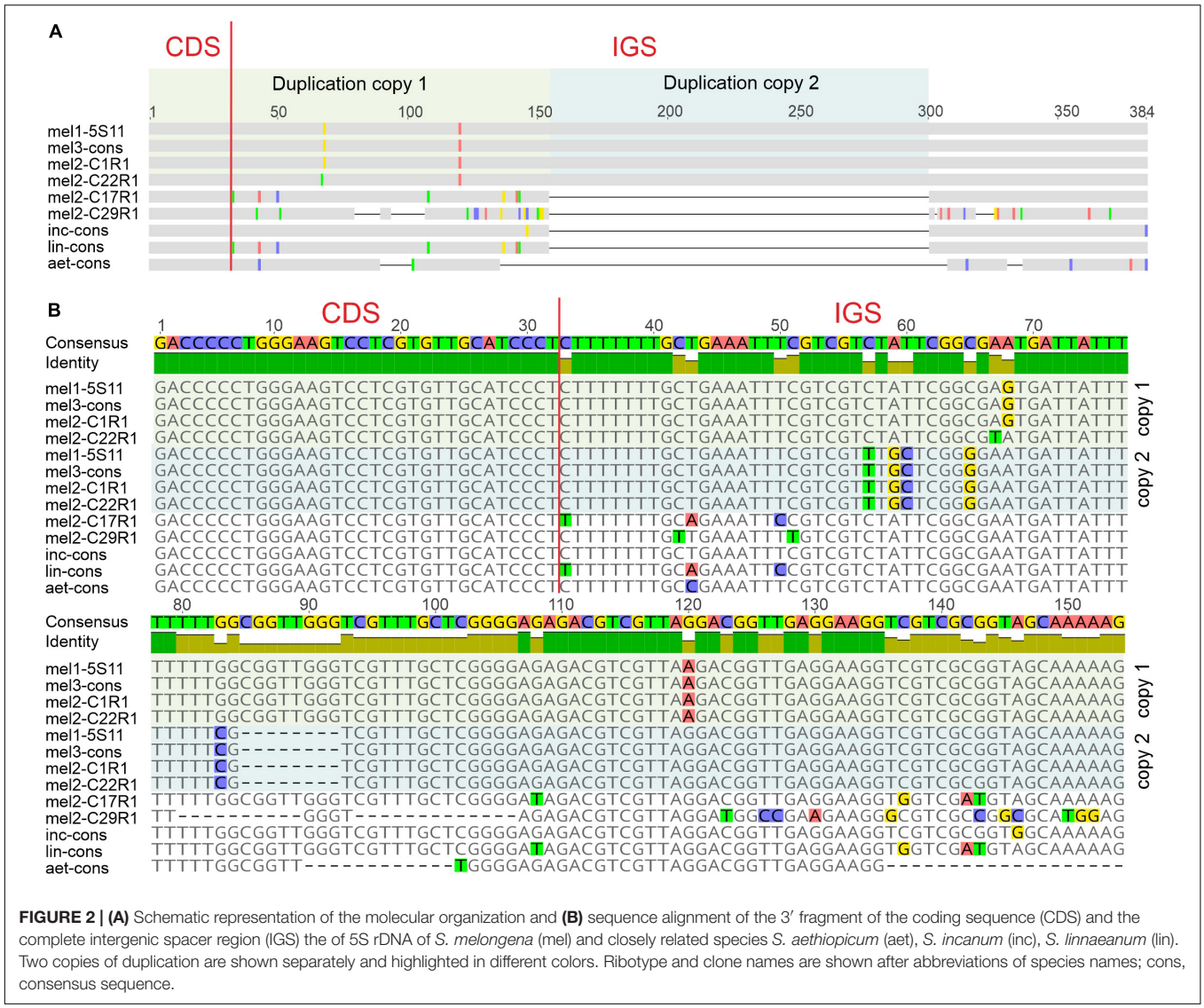


FIGURE 2 | (A) Schematic representation of the molecular organization and **(B)** sequence alignment of the 3′ fragment of the coding sequence (CDS) and the complete intergenic spacer region (IGS) the of 5S rDNA of *S. melongena* (mel) and closely related species *S. aethiopicum* (aet), *S. incanum* (inc), *S. linnaeanum* (lin). Two copies of duplication are shown separately and highlighted in different colors. Ribotype and clone names are shown after abbreviations of species names; cons, consensus sequence.

However, incongruence of results obtained by application of different markers is a well-known problem. Respectively, other genomic regions, particularly the 5S rDNA IGS, can additionally be used to clarify the phylogeny of lower-ranking taxa (Blöch et al., 2009; Tynkevich and Volkov, 2019; Cardoni et al., 2021; Ishchenko et al., 2021), including sect. *Petota* of the *Solanum* genus (Volkov et al., 2001). To evaluate the possibility of using this region to reconstruct phylogenetic relationships in the *Solanum* genus, we constructed an ML dendrogram that embraces 68 accessions from 63 species.

The ML dendrogram includes two major clades (**Figure 6**). Similar to our data, two clades in the *Solanum* genus were found by comparing sequences of plastid, nuclear ribosomal ITS and low-copy nuclear (*waxy*) genes (Särkinen et al., 2013, 2015). Particularly, four species, *S. abutiloides, S. erianthum, S. cochoa,* and *S. pseudocapsicum*, are included in Clade 2 of our dendrogram, which is in agreement with recent molecular data (Särkinen et al., 2013, 2015) but in contrast to the previous taxonomy of Nee (1999), who placed the species in the sections *Brevantherum, Basarthrum,* and *Holophylla* (see **Table 1**).

Clade 1 is composed of four smaller clades. Clade 1.1 contains four species of the *Morelloid* clade, *S. americanum, S. nigrum, S. scabrum,* and *S. villosum* (Särkinen et al., 2013, 2015). Two other species, *S. anomalostemon* and *S. valdiviense* previously associated with Morelloids are placed outside Clade 1.1, further supporting the phylogeny of the group proposed by Särkinen et al. (2015).

Clades 1.2–1.4 are combined in a polytomy. *S. dimorphandrum* of the *Thelopodium* clade (Bohs, 2005) and *S. dulcamara* of the *Dulcamaroid* clade (Bohs, 2005; Särkinen et al., 2013) belong to Clade 1.2, while another member of the *Dulcamaroid* clade, *S. seaforthianum,* occupies a basal position in Clade 1. *S. valdiviense* is included in Clade 1.3, which also comprises two species of sect. *Archaesolanum*, *S. aviculare* and *S. laciniatum.* The taxonomic position of *S. valdiviense* found in our analysis is fully consistent with previous data

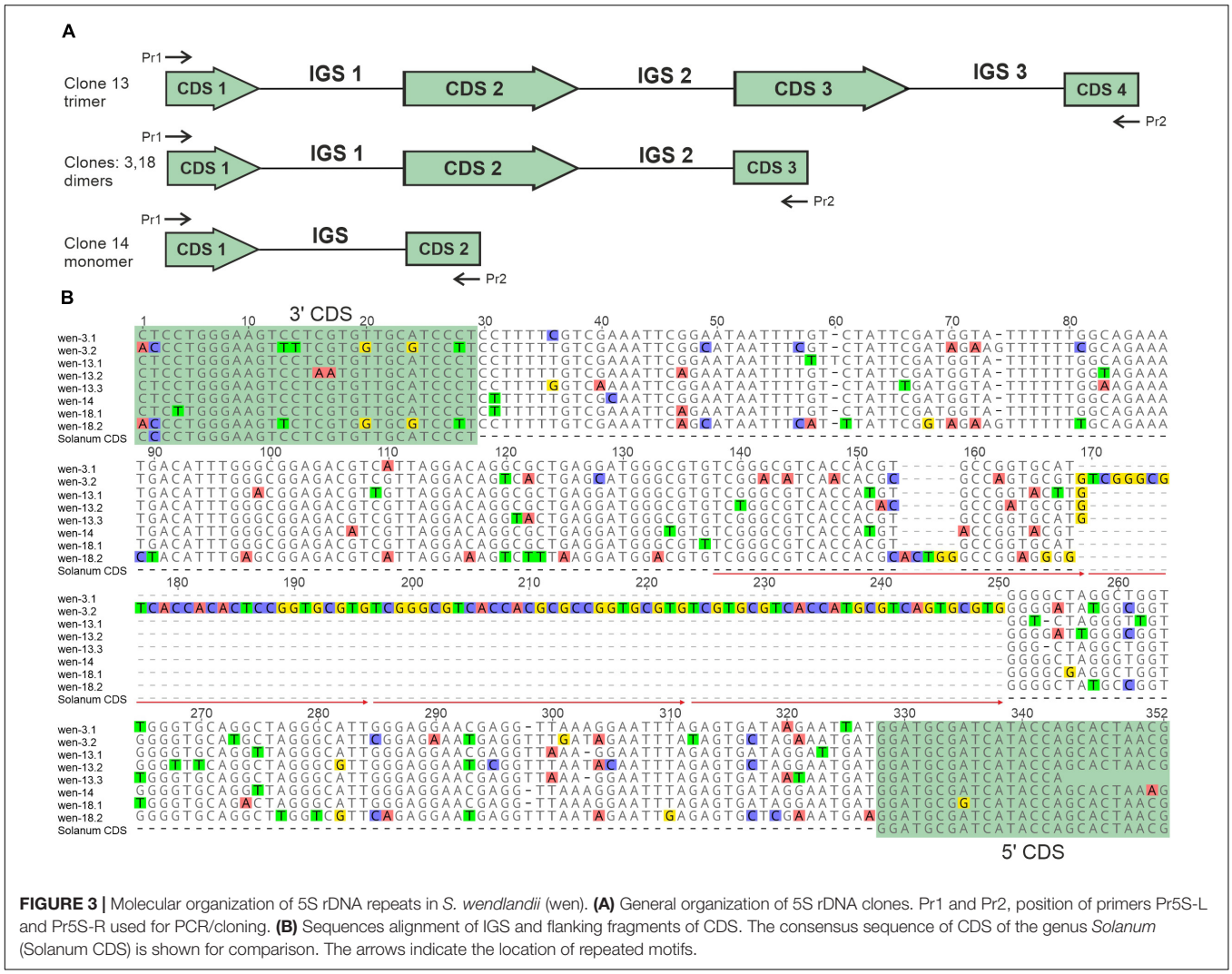


FIGURE 3 | Molecular organization of 5S rDNA repeats in *S. wendlandii* (wen). **(A)** General organization of 5S rDNA clones. Pr1 and Pr2, position of primers Pr5S-L and Pr5S-R used for PCR/cloning. **(B)** Sequences alignment of IGS and flanking fragments of CDS. The consensus sequence of CDS of the genus *Solanum* (Solanum CDS) is shown for comparison. The arrows indicate the location of repeated motifs.

(Särkinen et al., 2015). *S. aviculare* and *S. laciniatum* are closely related (**Figure 6**): There are several ribotypes in the genome of *S. aviculare* that are very similar and even identical to those of *S. laciniatum*. These data indicate incomplete lineage sorting during speciation or subsequent hybridization among these species. The close relationship between *S. aviculare* and *S. laciniatum* confirms the taxonomy derived from sequencing of three chloroplast and two nuclear regions in which these two species represent sister taxa (Poczai et al., 2011; Särkinen et al., 2015).

Clade 1.4.1 comprises Central American *S. appendiculatum* and South American *S. clivorum,* which were previously assigned, respectively, to sect. *Anarrhichomenum* and *Holophylla* (Nee, 1999) as well as *S. muricatum* of sect. *Basarthrum* (Nee, 1999; Särkinen et al., 2013), while Clade 1.4.2 embraces numerous species of sect. *Petota* (including tomato) (Hawkes, 1990; Nee, 1999; Komarova et al., 2008). According to a recent analysis (Särkinen et al., 2015; Gagnon et al., 2021), *S. appendiculatum* and *S. muricatum,* similar to our results, belong to the potato clade, in contrast to *S. clivorum*, which was placed outside clade I.

Clade 1 also includes the South American species *S. paposanum,* which represents the Regmandra clade (Bohs, 2005; Särkinen et al., 2013). It was found that this clade was resolved in different positions in three data sets used for comparison (Särkinen et al., 2015; Gagnon et al., 2021).

Clade 2 consists of three smaller clades, 2.1–2.3. Clade 2.1 comprises two closely related species, *S. abutiloides* and *S. erianthum*, which were assigned by Nee (1999) to sect. *Brevantherum* of the *Solanum* subgenus. Later, sect. *Brevantherum* was transferred to clade II consisting of predominantly spiny and shrubby species (Särkinen et al., 2013, 2015; Gagnon et al., 2021). Similarly, the third member of Clade 2.1, *S. anomalostemon*, was assigned to the Morelloid clade (Bohs, 2005) but later transferred to clade II (Särkinen et al., 2015; Gagnon et al., 2021). Accordingly, the inclusion of *S. abutiloides, S. erianthum,* and *S. anomalostemon* in clade II is further supported by our results.

Clade 2.2 contains five species, which were previously assigned to different taxonomic groups. According to Nee (1999), two Central/South American species, *S. wendlandii*

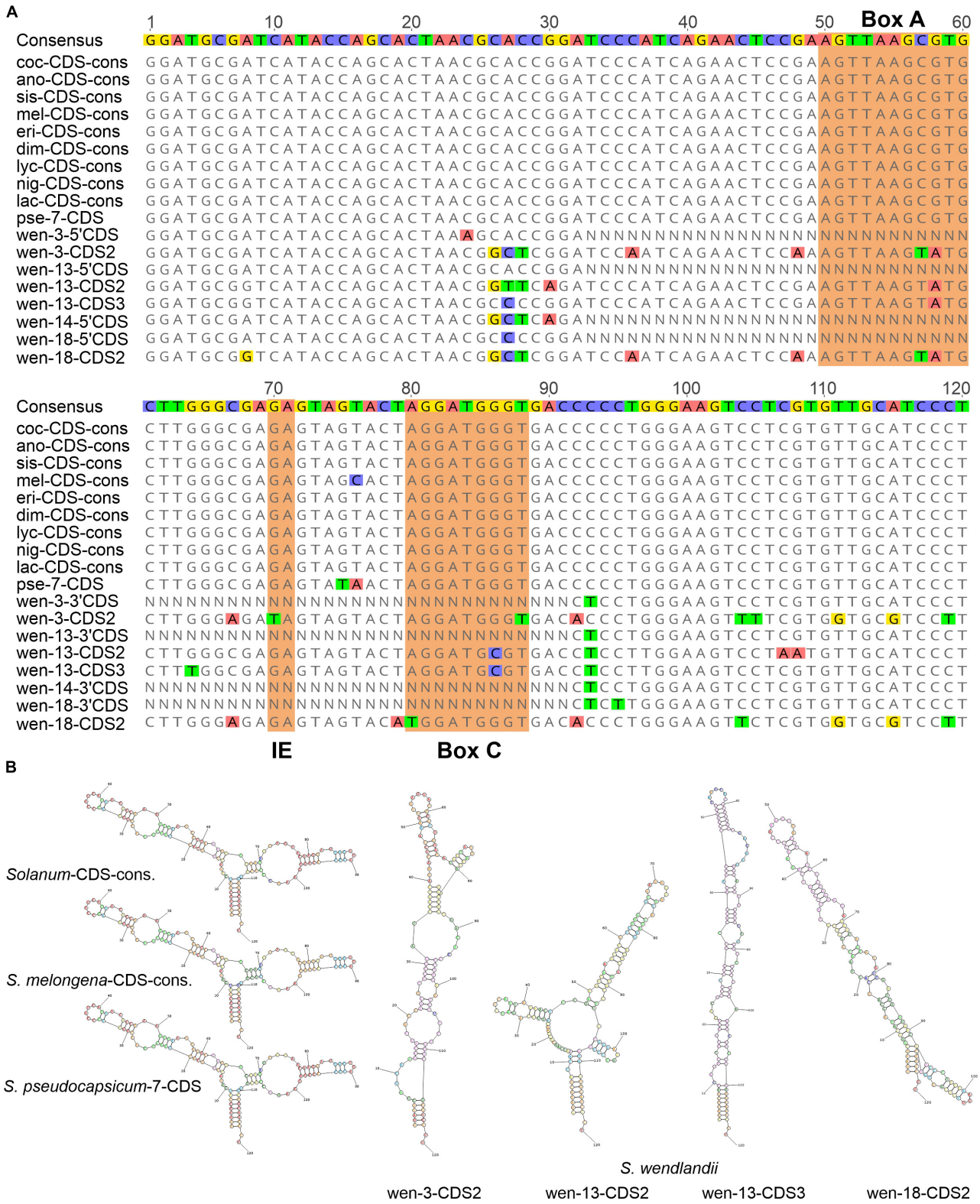


FIGURE 4 | Comparison of *Solanum* 5S rRNA CDS. **(A)** Alignment of the CDS of distantly related *Solanum* species and 5S rDNA clones of *S. wendlandii*. **(B)** Predicted secondary structures of 5S rRNA transcripts. Abbreviations of species names are given in **Table 1**.

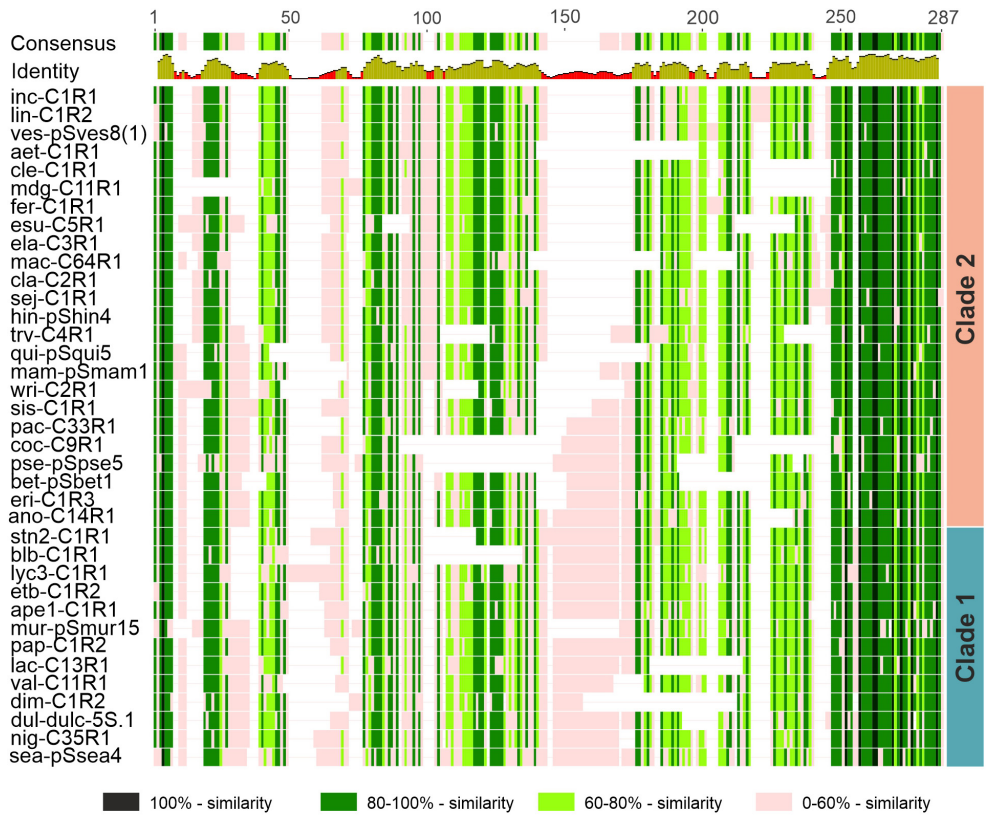


FIGURE 5 | Schematic representation of the sequence alignment of the 5S rDNA IGS of distantly related *Solanum* species. The level of sequence similarity is shown in different colors. Abbreviations of species names are given in **Table 1**. Taxonomic assignment of species to Clades 1 and 2 in the maximum likelihood (ML) dendrogram (see **Figure 7**) is shown.

and *S. pachyandrum,* are members of sect. *Herposolanum*. Later, it was shown that *S. wendlandii* belongs to clade *Wendlandii/Allophyllum*, while the position of *S. pachyandrum* appeared unclear (Bohs, 2005; Särkinen et al., 2013, 2015). Thereafter, both species were assigned to sect. *Aculeigerum* (Clark et al., 2015). Our data also confirm the phylogenetic affinity of *S. wendlandii* and *S. pachyandrum.*

The next two species, South American *S. cochoae* and *S. pseudocapsicum*, have been previously assigned to different sections, *Basarthrum* and *Holophylla* (Anderson and Bernardello, 1991; Nee, 1999). In contrast, *S. cochoae* and *S. pseudocapsicum* are combined in a well-supported clade in our ML dendrogram.

Originally, *S. cochoae* was included in sect. *Basarthrum* on the basis of morphological analyses and crossing experiments, although all crosses with related wild species were unsuccessful. Surprisingly, the only species crossed with *S. cochoae* was cultivated *S. muricatum*, despite large differences in karyotypes of these two species (Anderson and Bernardello, 1991). However, the possibility of obtaining hybrids cannot be seen as a decisive argument for the close relationship between these two species, as it is sometimes possible to successfully cross distant *Solanum* species (Daunay et al., 2019). The close relationship between *S. cochoae* and *S. muricatum* is also supported by recent molecular data (Gagnon et al., 2021). In our dendrogram, however, *S. cochoae* does not appear to be related to *S. muricatum* but to *S. pseudocapsicum*, a member of the Geminata clade (Gagnon et al., 2021).

It should be noted that the common feature of the 5S rDNA repeats of *S. cochoae* and *S. pseudocapsicum* is short length due to deletion in the central part of the IGS. In addition, each species possesses specific deletions in other IGS regions (**Figure 5**). Altogether, these structural features can affect the position of the species in the dendrogram. Accordingly, we believe that the taxonomic position of *S. cochoae* close to *S. pseudocapsicum* should be interpreted with appropriate reservation in this stage, and that further studies should be carried out in order to finally clarify the question.

The last member of Clade 2.2 is *S. betaceum*, which has been previously treated as a member of separate genus *Cyphomandra* (D'Arcy, 1991) and then later placed to *Solanum* (Bohs, 1995) and assigned to sect. *Pachyphylla* of the *Bassovia* subgenus (Nee, 1999) or clade *Cyphomandra* in clade II (Särkinen et al., 2013, 2015; Gagnon et al., 2021). In our dendrogram, *S. betaceum* is a sister taxon for the other members of Clade 2.2.

Clade 2.3 includes three clades of lower ranks, 2.3.1–2.3.3. Clade 2.3.1 comprises two pairs of species, the South American *S. aculeatissimum* and *S. mammosum* of the section/clade *Acanthophora* as well as Andean cultivated species *S. quitoense*

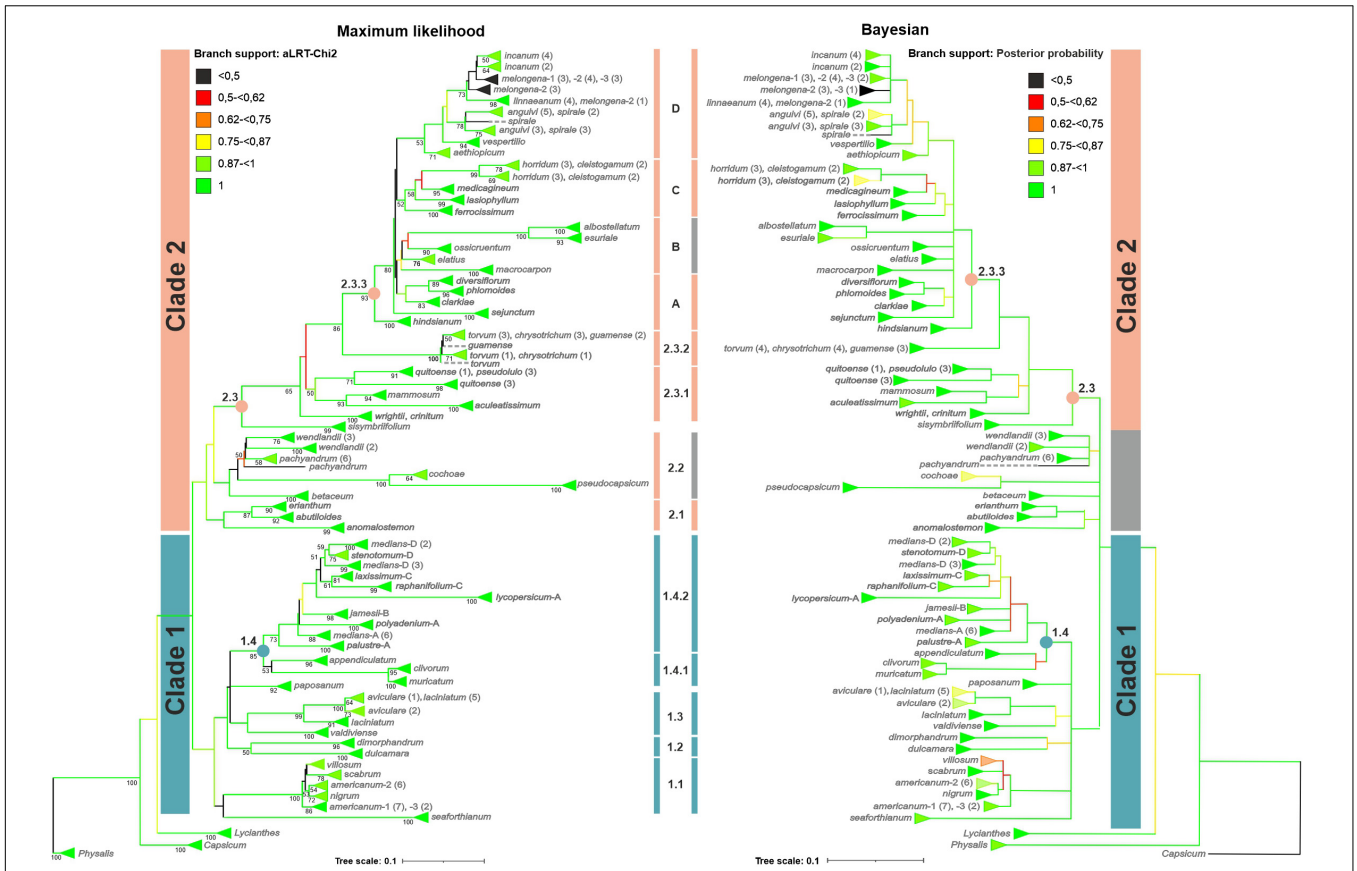


FIGURE 6 | Phylogenetic ML and Bayesian dendrograms constructed by comparison of 5S IGS sequences of *Solanum* species. The aLRT support and posterior probabilities are represented by different branch colors. Numbers in the ML dendrogram represent bootstrap support values. Terminal clades are collapsed to species level. Numbers in brackets near the collapsed nodes indicate the number of ribotypes used. IGS structural variants A-D are given for members of sect. *Petota* (Clade 1.4.2). ML clades not confirmed by Bayesian analysis are highlighted in gray color.

(naranjilla or lulo) and its wild relative *S. pseudolulo* of clade *Lasiocarpa* (Nee, 1999; Bohs, 2005; Levin et al., 2006; Särkinen et al., 2013). According to a molecular analysis, the clades *Acanthophora* and *Lasiocarpa* represent sister taxa (Särkinen et al., 2013; Gagnon et al., 2021). In the genome of *S. quitoense*, a ribotype similar to that of *S. pseudolulo* was detected, which could be due to hybridization between these species (Fory Sánchez et al., 2010).

Clade 2.3.2 comprises two Central/South American species, *S. chrysotrichum* and *S. torvum* of the section/clade *Torva*, as well as *S. guamense,* an endangered endemic species in Northern Mariana Islands (Stone, 1970) whose taxonomic status remains unclear (Nee, 1999; Bohs, 2005; Särkinen et al., 2013; Aubriot et al., 2016). Our analysis revealed that the three species share common ribotypes and are, therefore, unresolved in the dendrogram. The high genetic affinity of *S. chrysotrichum* and *S. torvum* agrees well with their morphological similarity. *S. guamense* also appeared to be closely related to these species.

Clade 2.3.3 includes *S. hindsianum* (clade Elaeagnifolium, Bohs, 2005; Särkinen et al., 2013), an endemic to the Sonoran Desert region of southern Arizona and northern Mexico (Knapp et al., 2017), and a well-supported monophyletic clade of 21 species, most of which have been assigned to sect. *Melongena* (Nee, 1999) or the Old World clade (Levin et al., 2006; Särkinen et al., 2013; Aubriot et al., 2016) of the *Leptostemonum* subgenus. However, the taxonomic position of five species (*S. albostellatum, S. elatius, S. lasiophyllum, S. medicagineum,* and *S. spirale*) has not yet been clarified, especially with molecular methods. In the Old World clade, there are four groups, A–D, of closely related species.

Clade 2.3.3A comprises members of "Dioicum Complex," a set of several dioecious species (Whalen, 1984; Bean, 2004) from tropical Australia. Our data show that *S. diversiflorum* and *S. phlomoides* are closely related, and that *S. clarkiae* is a more distant species. *S. sejunctum* is placed outside clade 2.3.3A. This result agrees with the phylogeny based on the analysis of *trnK–matK* and ITS data sets (Martine et al., 2006, 2009).

Four other Australian species, *S. albostellatum, S. esuriale, S. elatius, S. ossicruentum*, as well as *S. macrocarpon* (African eggplant), belong to the next clade, 2.3.3B, although with a moderate statistical support. The West African species *S. macrocarpon* was previously assigned to Anguivi Grade (Aubriot et al., 2016, 2018; Gagnon et al., 2021), a group of Old

FIGURE 7 | Schematic representation of sequence alignment of the 5S IGS of sect. *Petota* members. The level of sequence similarity is shown in different colors. Positions of group-specific indels I and II and GC-duplication (GC-DUP) are indicated. Abbreviations of species names are given in **Table 2**. Structural features of presented IGS sequences are indicated as structural variants (SVs) A-D.

World *Leptostemonum* species closely related to *S. melongena* (see our clade 2.3.3D below). Respectively, the phylogenetic affinity of *S. macrocarpon* to the Australian species seems somewhat unexpected and can be explained by the presumptive hybrid origin of *S. macrocarpon* (Daunay et al., 2019).

According to our data, *S. albostellatum* and *S. esuriale* from Western Australia show the closest relationship in Clade 2.3.3B, which is in good agreement with the high morphological similarity of these species (Davis and Hurter, 2012). *S. elatius* is also a member of the *S. esuriale* group (Bean, 2013).

S. ossicruentum represents a functionally dioecious bush tomato from northwestern Australia. Earlier, it was recognized as a variant of *S. dioicum,* a member of "Dioicum Complex." However, later molecular analysis shows that *S. ossicruentum* is either a sister taxon to the rest of this group or represents an independent dioecious lineage (Martine et al., 2016). Our data further supported the second opinion and indicate a phylogenetic affinity between *S. ossicruentum* and members of the *S. esuriale* group.

Clade 2.3.3C comprises five Australian species. Two species, *S. cleistogamum* and *S. horridum,* contain very similar sets of ribotypes in their genomes and appear unresolved in the dendrogram. A sister taxon to them is *S. medicagineum*, while *S. lasiophyllum* and *S. ferocissimum* are more distantly related species. A close relationship among *S. cleistogamum, S. horridum,* and *S. medicagineum* has been shown earlier (Bean, 2004, 2012; Levin et al., 2006).

Hence, the Australian *Solanum* species studied here belong to three clades, 2.3.3A, B, and C. Similarly, monophyly of the Australian species was not supported by the analysis of seven nuclear genes (Martine et al., 2019).

Clade 2.3.3D comprises seven species naturally distributed in Africa and Asia. In particular, this clade includes two very morphologically and genetically similar domesticated plants, *S. aethiopicum* (bitter tomato, Ethiopian eggplant) and *S. melongena* as well as their presumptive wild ancestors, *S. anguivi* and *S. incanum*. The second species is very similar and can even be confused with *S. linnaeanum* (Daunay et al., 2001; Doganlar et al., 2002; Prohens et al., 2012). *S. vespertilio*, a species endemic to the Canary Islands, appears to be closely related to the other members of clade 2.3.3D.

Previously, the phylogeny of Old World "spiny solanums" was clarified using plastid and nuclear markers (Aubriot et al., 2016, 2018; Vorontsova and Knapp, 2016; Knapp et al., 2019; Gagnon et al., 2021). It was demonstrated that *S. incanum, S. linnaeanum,* and *S. melongena* are closely related and belong to the Eggplant clade, and that *S. aethiopicum*, *S. anguivi, S. vespertilio* (and *S. macrocarpon*, which is placed to clade 2.3.3B in our dendrogram) are included in Anguivi Grade outside the Eggplant clade. Hence, our novel data mainly confirm these results.

Surprisingly, *S. anguivi* and morphologically different *S. spirale,* a tetraploid (Randell and Symon, 1976) species from East Asia, are not resolved in the dendrogram (see

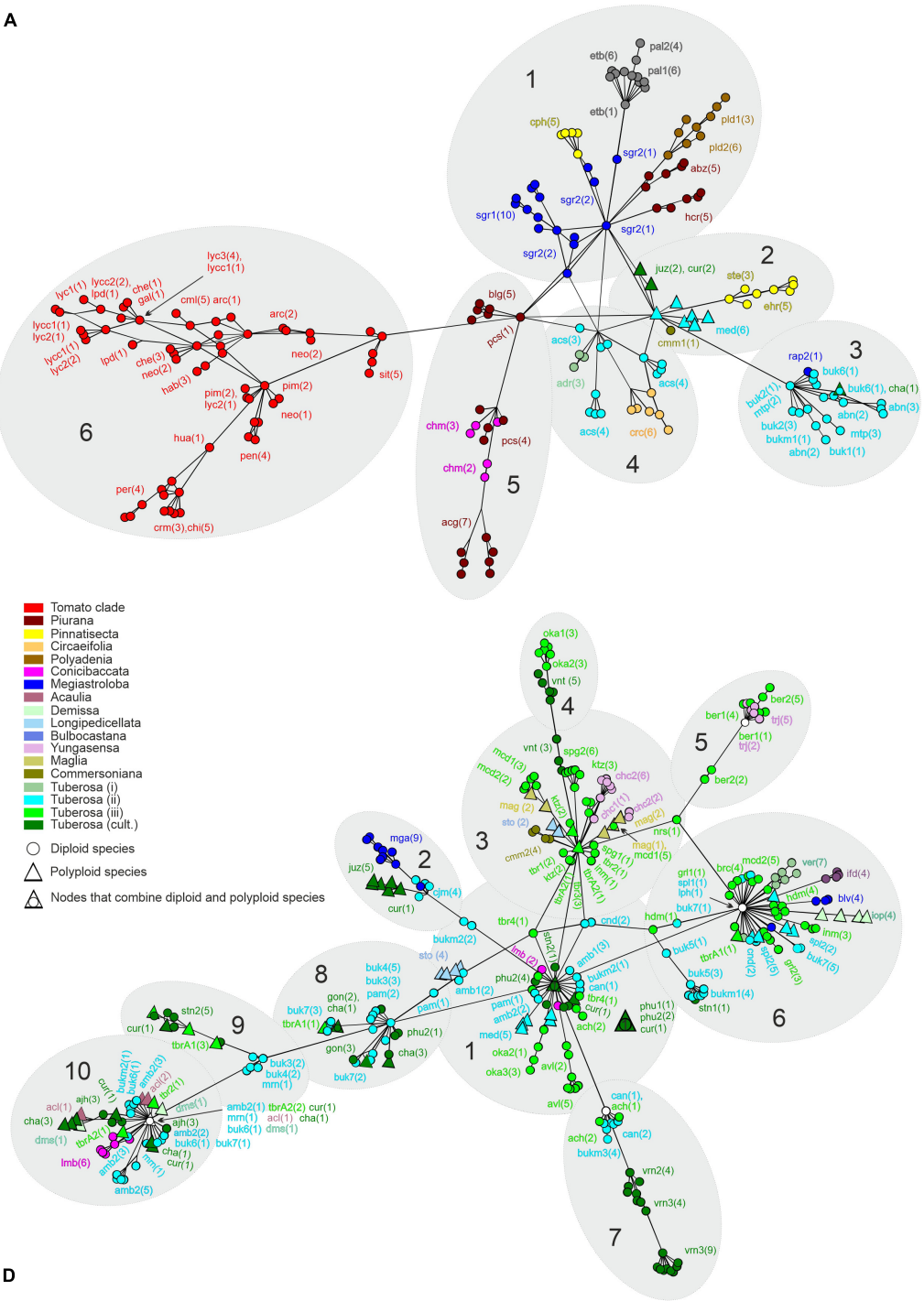


FIGURE 8 | Median-joining networks depicting relationships of 5S IGS structural variants A and D of the species of sect. *Petota*. Names of main clusters are given. The affiliation of the species to the series proposed by Hawkes (1990) is indicated by different colors. Numbers in brackets indicate the number of ribotypes used. Abbreviations of species names are given in **Table 2**.

Figure 6). A possible explanation for this result could be the allopolyploid origin of *S. spirale*. In this case, the 5S rDNA inherited from the parent related to *S. anguivi* could be retained in the genome, while the DNA of the other parent was lost. The uniparental inheritance of 5S rDNA in allopolyploids, both young and old, has been reported for several taxonomic groups including Solanaceae (Pontes et al., 2004; Volkov et al., 2017).

Clade 2.3 also comprises three species that do not belong to clades 2.3.1–2.3.3 presented above. Two species, *S. crinitum* and *S. wrightii*, represent the clade Androceras/Crinitum, while *S. sisymbriifolium* belongs to the clade Sisymbriifolium (Levin et al., 2006; Särkinen et al., 2013; Gagnon et al., 2021). The taxonomic position of these species in our dendrogram agrees well with previous results of molecular phylogenetics studies.

Majority of the clades identified in the ML tree was also recognized in the Bayesian dendrogram (**Figure 6**). However, the monophyly of Clade 2 was not confirmed by Bayesian inference: Clades 2.1, 2.2, and 2.3 are not combined with each other but belong to a basal polytomy in the *Solanum* genus.

Thus, in this study, we present the phylogeny of the *Solanum* genus derived from the analysis of 5S IGS sequences. Six species (*S. albostellatum, S. elatius, S. guamense, S. lasiophyllum, S. medicagineum,* and *S. spirale*) were characterized here for the first time using molecular taxonomy methods. The obtained dendrograms are mainly congruent with the published data for other regions of nuclear and plastid genomes: same major and minor clades were found for the species examined. However, taxonomic relationships between these clades and position of some species (e.g., *S. cochoae, S. clivorum, S. macrocarpon, S. spirale*) differ from previous results and require further clarification. Taken together, our results show that the 5S IGS represents a convenient molecular marker for phylogenetic studies on the *Solanum* genus. In particular, the simultaneous presence of several variants of rDNA in the genome enables the detection of cases of reticular evolution such as incomplete lineage sorting and interspecific hybridization.

Molecular Evolution and IGS Diversity in Sect. *Petota*

One of the species-rich groups in genus *Solanum* is sect. *Petota*, which has about 250 members (Hawkes, 1990; Nee, 1999). In our ML dendrogram (**Figure 6**), sect. *Petota* belongs to Clade 1.4.2.

To analyze the molecular evolution of 5S rDNA in this section and in more details, we assembled IGS ribotypes for 125 accessions representing 83 species (**Table 4**) and compared the results with our previous data, obtained by cloning and sequencing of 5S rDNA of 32 wild species and breeding lines of sect. *Petota* (Volkov et al., 2001).

Analysis of the IGS sequences revealed that they differ in base substitutions and indels (**Figure 7**). Same indels mostly occur in a single or some closely related species and, therefore, represent convenient molecular markers for their identification. For example, non-tuber-bearing species *S. etuberosum* and *S. palustre* (series *Etuberosa*; Hawkes, 1990) possess a common specific deletion at the beginning of the IGS, or *S. laxissimum* and *S. violaceimarmoratum* (series *Conicibaccata*) have a deletion in the central part of it (**Figure 7**). Several species-specific indels in the IGS of the *Solanum* species have already been described (Volkov et al., 2001), and our actual analysis additionally identifies new ones for the novel species. This finding further confirms our earlier assumption that indels are a characteristic feature of IGS evolution in sect. *Petota*. We have also argued that because of the high frequency of indels compared to base substitutions, IGS cannot be used for phylogenetic reconstruction of this section applying standard algorithms. However, the indels represent unique evolutionary events that should be considered in taxonomic studies.

Considering the location of group-specific indels I and II as well as GC-duplication (**Figure 7**), four major structural variants of the IGS were identified. Accordingly, members of sect. *Petota* can be divided into four groups, A–D.

Group A comprises species that belong to the subsection *Estolonifera* including the tomato group, and to the series *Pinnatisecta, Polyadenia, Commersoniana, Circaeifolia, Megistacroloba, Conicibaccata,* and *Piurana* of the subsection Potatoe (Hawkes, 1990; Nee, 1999), or to clades 1+2 and 3 (Spooner et al., 2014; Huang et al., 2019). Also, SV-A was found in four species (*S. acroscopicum*, *S. andreanum, S. abancayense, S. multiinterruptum*) that were assigned to ser. *Tuberosa* (Hawkes, 1990; Nee, 1999) or clade 4 (Spooner et al., 2014; Huang et al., 2019). Group B includes only two species, *S. jamesii* and *S. pinnatisectum* of ser. *Pinnatisecta* (Hawkes, 1990; Nee, 1999) or clade 1+2 (Huang et al., 2019). This means that the SR-B arose relatively recently during speciation in clade 1+2, just before the divergence of *S. jamesii* and *S. pinnatisectum* but after their separation from the sister taxon, which was similar to *S. stenophyllidium*.

Group C includes accessions of three species, *S. raphanifolium, S. laxissimum,* and *S. violaceimarmoratum*, which belong to the series *Megistacroloba* and *Conicibaccata* (Hawkes, 1990) or clade 4 (Spooner et al., 2014; Huang et al., 2019). Group D embraces numerous species that belong to the series *Yungasensa, Megistacroloba, Cuneoalata, Maglia,* and *Tuberosa* (Hawkes, 1990) or clade 4 (Spooner et al., 2014; Huang et al., 2019). SV-C and D were not found outside of clade 4, which, however, also includes four species possessing SV-A. Therefore, SV-C and D arose from SV-A after separation of clades 3 and 4.

Interestingly, in some cases, rDNA repeats representing different structural variants were found in the same plant accession (see below).

We have also found that the central part of the IGS is completely deleted in *S. bulbocastanum* of the series *Bulbocastana* (Hawkes, 1990) or clade 1+2 (Huang et al., 2019). Accordingly, the structural organization of ITS characteristic of this species cannot be assessed and used for phylogenetic reconstruction.

Analysis of our data showed that the most common IGS variants are SV-A and-D, and SV-B and-C were found only in four and three species, respectively. In order to assess the molecular diversity of SV-A and-D, we constructed median-joining networks for these two IGS variants using 204 and 353 sequences (**Figure 8**). In the median-joining networks, the sequences of SV-A and-D are distributed between the six and ten main clusters according to their similarity. In the vast majority of cases, each node corresponds to only one sequence, with the exception of one node in median-joining network A and seven nodes in median-joining network D. These nodes include two to nine ribotypes that mainly represent genomes of different species. Therefore, identical IGS sequence variants can be present in genomes of different species or plant accessions, suggesting their common origin.

SV-D sequences are distributed among ten clusters, D1-D10 (**Figure 8**). The largest clusters, D1, D3, D6, and D10, contain 50, 56, 68, and 51 sequences, respectively. The sequences included

in cluster D1 are nearly identical to SV-D consensus sequence and, therefore, represent evolutionary ancestral ribotypes, while the other clusters comprise derived sequences containing specific base substitution and indels. Starting from cluster D1, five evolutionary lineages can be distinguished.

Taken together, our data agree well with modern taxonomy, which is based on the application of molecular methods (Spooner et al., 2014; Huang et al., 2019) but are less consistent with the traditional classification of Hawkes (1990). In particular, the sections proposed by Hawkes (1990) are not confirmed, because species from different sections are mixed up and belong to different clusters in the median-joining network. In contrast, our results agree well with the molecular data, since clades 4 North and 4 South (Spooner et al., 2014; Huang et al., 2019) can also be recognized in our median-joining network: members of the North and South clades belong to clusters D1, D7-D10, and D2-D6.

Conserved Sequence Motifs in the Intergenic Spacer of *Solanum* Species

Comparative sequence analysis revealed that the most conservative regions of the IGS in *Solanum* species are the 7- and 40-bp-long fragments at the 5′ and 3′ ends (**Figure 5**). The evolutionary conservation of these regions has already been observed in other plants (Hemleben and Werts, 1988; Crisp et al., 1999; Tynkevich and Volkov, 2019; Ishchenko et al., 2020, 2021), and a possible reason for this seems to be their involvement in the transcription of 5S rDNA by RNA polymerase III (Pol III).

External elements of the Pol III promoter have been previously characterized in *Arabidopsis thaliana* (Douet and Tourmente, 2007; Vaillant et al., 2007; Layat et al., 2012; Simon et al., 2018). These signals include the TATATA motif (so-called TATA-box), GC-dinucleotide, and C nucleotide in positions-28,-13, and-1 bp, respectively. Similar sequences were also found in other plants (Tynkevich and Volkov, 2014; Tynkevich et al., 2015; Ishchenko et al., 2018, 2021). In representatives of the *Solanum* genus, as well as in *Quercus* (Tynkevich and Volkov, 2019) and *Rosa* (Tynkevich and Volkov, 2014), the TATA-box has a length of 7 bp and begins in position-30. Its sequence (TTTAATA) in *Solanum* is slightly different from that in other groups of plants.

Another external element of the Pol III promoter, the GC-dinucleotide (Douet and Tourmente, 2007) is duplicated in several *Solanum* species and is located, respectively, both in the typical position-12 and, additionally, in position-14. Similar to *Solanum*, duplication of this presumptive external promoter element was also found in the *Quercus* species (Tynkevich and Volkov, 2019).

The third conservative promoter element, cytosine, in position-1 (Douet and Tourmente, 2007; Simon et al., 2018), has been replaced by thymine in more than half of the *Solanum* species. In addition, we found that the dinucleotide GA in position-3 in the IGS is highly conserved, indicating its possible involvement in transcription initiation.

At the beginning of the IGS in *Solanum*, like in other genera, the oligo-T motif TTTTT was found, which probably represents a transcription termination site (Hemleben and Werts, 1988; Simon et al., 2018; de Souza et al., 2020).

The most variable central region of the IGS can be subdivided into (i) AT-rich and (ii) subrepeated regions. Previously, AT-rich regions were found in the IGS of Fabaceae (Hemleben and Werts, 1988) and Poaceae (Röser et al., 2001). AT-rich regions demonstrate a similarity to amplification-promoting sequences (Borisjuk et al., 2000), which may be involved in amplification of 5S rDNA repeats. Also, regions composed of subrepeats were described for the IGS of several plant taxa, e.g., Rosaceae (Tynkevich and Volkov, 2014) and Poaceae (Ishchenko et al., 2018, 2021). Previously, we have demonstrated that structural rearrangements of the variable central region of the IGS in *Solanum* species of sect. *Petota* as well as in distantly related *S. melongena* and *S. betaceum* are preferentially associated with four classes (A–D) of short direct subrepeats: the IGS evolved mainly by duplications of some sequence motifs, resulting in formation of several variants of subrepeats, which were independently amplified in different sections of the genus after radiation from a common ancestor (Volkov et al., 2001; Davidjuk et al., 2010, 2013).

Intragenomic Heterogeneity and Molecular Evolution of 5S rDNA

It is widely believed that 5S rDNA repeats present in the same genome (at least in diploid species) should be nearly identical because of concerted evolution (Coen et al., 1982; Tynkevich and Volkov, 2014; Barman et al., 2016). In our study, we performed a detailed analysis of 5S rDNA intragenomic sequence diversity and found several ribotypes in all the species studied. Comparative analysis of all available sequences showed that the IGS sequence similarity in *Solanum* species ranges from 51.4 to 100%. The highest levels were found in *S. wrightii* and four representatives of the tomato group, namely *lycopersicum var. cerasiforme-2*, *S. galapagense,* and *S. huaylasense*; each of which had only one ribotype detected. The high intragenomic homogeneity of IGS (over 95%) is also characteristic of other representatives of the tomato group with the exception of *S. lycopersicum-2* and *S. corneliomuelleri.* The relatively low IGS similarity in *S. corneliomuelleri* (89.5%) is due to the presence of 10-bp deletion in one ribotype, while no further indels are found in any of the other members of the tomato group. Hence, deletions are very rare during the evolution of IGS in the tomato group, which is in obvious contrast to other *Solanum* groups, especially to closely related tuber-bearing species of *Petota*.

Our calculations indicated that the lowest level of intragenomic IGS sequence similarity is demonstrated by *S. melongena-1* and *-2* (56 and 51.4%), *S. sogarandinum* (54.9%), *S. wendlandii* (58.4%), and *S. lasiophyllum* (59.4%). In *S. melongena-1* and *-2,* it is due to simultaneous existence of short and long (containing extra-long duplication, see **Figure 2** and **Supplementary Material**) repeats in the genome, while in *S. melongena-3,* which possesses only long repeats, the similarity amounts to 95.2–99.4%. Similarly, in *S. lasiophyllum,* three adjacent deletions (65 bp in total) present in one of six ribotypes is the main reason for the low level of intragenomic similarity. In contrast, two mechanisms contribute to the low similarity of

IGS in *S. sogarandinum:* (i) a long deletion in one ribotype and (ii) multiple base substitutions in another. In the second case, 33 of 210 bp in the same ribotype was changed compared to the consensus sequence. Notably, these mutations are present in the 3′ IGS region, which likely contains external promoter elements, suggesting putative pseudogenization of the ribotype. Putative 5S rDNA-related pseudogenes have already been described for members of *Solanum* and other genera of Solanaceae (Volkov et al., 2001, 2017).

In *S. wendlandii*, similar to *S. sogarandinum,* two mechanisms, an insertion and a large number of base substitutions, cause increased heterogeneity of IGS sequences (**Figure 3**). Accordingly, we excluded long (more than 5 bp) deletions and multiple base substitutions from our calculations and found that in this case the minimum level of intragenomic similarity of the IGS in *Solanum* species is around 85–90%.

In general, our results indicate that there are two mechanisms, long indels and multiple base substitutions, that significantly affect the heterogeneity of the IGS in *Solanum* species. Multiple base substitutions are rare events: out of about 900 analyzed sequences, only five ribotypes bearing multiple base substitutions were identified in four plant accessions (*S. kurtzianum, S. pinnatisectum-2*, *S. sogarandinum-1*, *and S. vernei-3*), while long indels are much more common.

The question, "what can be the source of the IGS intragenomic polymorphism?" arises. There are at least two possible options: (i) new variants emerge in the genome itself by accumulation of mutations and (ii) new variants appear in the genome as a result of introgression of genetic material due to interspecific hybridization. It is well known that in sect. *Petota,* especially in the *S. brevicaule* complex, interspecific hybridization is widespread at both the diploid and polyploid levels (Hawkes, 1990; Spooner et al., 2014). Among the 125 examined accessions representing sect. *Petota*, two or more structural variants of the IGS were found in 25 cases, and interspecific hybridization seems to be a plausible explanation for this polymorphism, especially when structurally different IGS variants (e.g., A and D) occur in the same genome. However, further research is needed to confirm this option.

Our data suggest that long indels and multiple base substitutions appeared repeatedly during the molecular evolution of IGS in the *Solanum* genus. However, it seems that they were mostly not conserved and eliminated. Accordingly, the length of IGS and contents of GC pairs did not change significantly during the course of speciation (see above). The likely reason for this negative selection could be the association between indels/multiple base substitutions and pseudogenization of 5S rDNA. Accordingly, it looks that the main road of the IGS molecular evolution seems to be step-wise accumulation of single base substitution or short indels.

Intraspecific 5S rDNA Heterogeneity

It could be anticipated that different accessions of same species possess identical/similar sets of ribotypes. To check this assumption, we examined two to three accessions for 25 diploid species (**Tables 1**, **2** and **Supplementary Material**), and in most cases, only one IGS variant was actually found. However, in seven species (*S. ambosinum, S. commersonii, S. microdontum, S. okadae, S. phureja, S. raphanifolium,* and *S. stenotomum*; **Figure 8** and **Supplementary Material**), one or two structural variants were detected in different accessions, indicating presumptive interspecific hybridization.

Among the species studied, we examined the highest number of accessions in *S. bukasovii,* which was considered one of the ancestors of cultivated potato (Ugent, 1970; Hawkes, 1990; Hosaka, 1995; Spooner et al., 2005; Hardigan et al., 2015). This close relationship was also confirmed in our previous study by analyzing the 5′-external transcribed spacer (ETS) region of nuclear 35S rDNA (Volkov et al., 2003). According to the comparison of whole plastid genomes, the species belongs to clade 4 North (Huang et al., 2019). Respectively, based on its taxonomic position, it might be expected that *S. bukasovii* should possess the IGS variant D. However, SV-A3 was previously found in the buk1 accession (Volkov et al., 2001), which indicates incongruence among different phylogenetic markers. To further clarify the issue, we analyzed ten accessions of *S. bukasovii* in this study and found an extreme variability of the 5S IGS set. Two accessions (buk1 and buk2) contain only SV-A3 in the genome, while six others (buk3, buk4, buk5, buk7, bukm2, and bukm3) possess different D variants, SV-D1,-D4,-D7,-D8,-D9, and-D10 (see **Table 4**). Two remaining accessions (buk6 and bukm1) have both SV-A3 and –D4 or-D10. The D variants detected in accessions of *S. bukasovii* belong to different clusters in the median-joining network (**Figure 8**) and are identical or very similar to the D variants of several species (e.g., *S. achacachense, S. ambosinum, S. canasense, S. marinasense, S. phureja, S. stenotomum*, etc.) belonging to clade 4 North + cultivated (Huang et al., 2019). Similarly, the SV-A3 found in *S. bukasovii* is identical/similar to that of *S. abancayense, S. multiinterruptum, S. raphanifolium,* and *S. chaucha.* This unusual diversity of IGS might indicate a complex hybridogenic origin: i.e., some of the examined accessions could represent natural hybrids between *S. bukosovii* and different *Petota* species. The genetic heterogeneity of *S. bukosovii* accessions and the putative hybrid origin of some of them by crossing with *S. sparsipilum* or *S. raphanifolium* have recently been demonstrated using plastid and mitochondrial markers (Achakkagari et al., 2020, 2021; Bozan, 2021).

Origin of Polyploid Species

There are several polyploids among the studied species. Two of them, *S. chaucha* and *S. juzepczukii,* are triploids. It was originally postulated that *S. chaucha* arose as a result of hybridization between diploid *S. phureja* and tetraploid *S. tuberosum* ssp. *andigena* (Bukasov, 1939), but later, *S. chaucha* was recognized as the autotriploid of *S. phureja* (Bukasov, 1978). In contrast, Hawkes (1962, 1990) suggested that *S. chaucha* originated from hybridization between *S. tuberosum* ssp. *andigena* and diploid species *S. stenotomum*. Our analysis indicates an allotriploid origin of *S. chaucha*, as three structural variants of IGS, SV-A3,-D8, and-D10, are found in its genome (**Table 4** and **Figure 8**). The molecular data confirm the close relationship among *S. chaucha, S. phureja* (accession phu2

but not phu1), *S. tuberosum* ssp. *andigena*, and a diploid species, *S. stenotomum* subsp. *goniocalyx;* all of which have SV-D8. However, the origin of *S. chaucha* appeared to be more complicated, because other structural variants of the IGS had to be inherited from species belonging to clusters A3 and D10.

It is widely believed that the triploid species *S. juzepczukii* originated from a natural cross between a cultivated diploid, *S. stenotomum,* and the wild tetraploid species *S. acaule* (maternal form), and pentaploid species *S. curtilobum* arose from a combination of non-reduced gamete of *S. jusepczukii* (3×, maternal form) with a reduced (2×) gamete of tetraploid *S. tuberosum* ssp. *andigena* (Bukasov, 1939; Hawkes, 1962, 1990; Schmiediche et al., 1980). The application of nuclear molecular markers confirmed this scenario (Rodríguez et al., 2010), but an alternative origin of *S. curtilobum* by hybridization of triploid species of the Andigenum group (non-reduced gamete, maternal form) and *S. acaule* was later proposed (Gavrilenko et al., 2013; Spooner et al., 2014). We found that the above-mentioned species contain the following IGS variants: *S. acaule*: SV-D10, *S. jusepczukii*: SV-A2 and-D2; *S. tuberosum* ssp. *andigena*: SV-D6,-D8, and-D9; *S. curtilobum*: SV-A2,-D1,-D2,-D9, and-D10. Three sets of IGS variants were identified for three accessions of *S. stenotomum* (stn1: SV-D1,-D9; stn2: SV-D4; gon: SV-D8). Remarkably, five IGS structural variants have been identified for pentaploid *S. curtilobum* that demonstrate the complex hybrid nature of this species. The analysis of the results showed that several common IGS structural variants are present in genomes of the examined species, which are correspondingly co-localized in the median-joining network (see **Figure 8**). However, no expected additivity of IGS structural variants from the presumptive parents in the derived alloploid progeny was found. It looks probable that the origin of *S. jusepczukii* and *S. curtilobum* may involve more parental diploids and requires further clarification using more plant accessions and additional molecular markers.

Two structural variants of IGS, SV-A2, and -D1, are present in the genome of *S. medians*, indicating its origin from a cross between potato species included in clusters A2 and D1. For *S. medians*, diploid and triploid populations were reported (Hijmans et al., 2007). Unfortunately, we could not find in SRA information about the ploidy level of the plant accession analyzed here.

According to a GISH analysis of meiotic preparations, *S. stoloniferum* appears to be an allotetrapolyploid species with a genomic constitution of AABB. It has been suggested that the species originated from *S. verrucosum* as the A genome donor and another North or Central American diploid species (e.g., *S. cardiophyllum, S. ehrenbergii,* or *S. jamesii*) as the B genome donor (Hijmans et al., 2007). In that case, regarding our data (**Table 4** and **Figure 8**), it is expected that *S. stoloniferum* inherited SV-A1, -A2, or -B from *S. cardiophyllum, S. ehrenbergii,* or *S. jamesii*, respectively, as well as SV-D4 from *S. verrucosum*. However, *S. stoloniferum* possesses only D-variants of IGS, namely, SV-D1 and SV-D3, that could be inherited from *S. ambosinum* and from one of the species that belong to cluster D3.

Previously, we have discussed the presumptive origin of a tetraploid, *S. acaule,* and *S. tuberosum* as well as hexaploids *S. demissum* and *S. iopetalum* (Volkov et al., 2001). The new data confirm the close relationship between *S. demissum* and *S. acaule* (cluster D10) and more distant position of *S. iopetalum* (cluster D6), and provide new information on putative diploid ancestors of the polyploid species. Our novel data also show that the IGS variants of two accessions of *S. tuberosum* ssp. *andigena* belong to different clusters (tbrA1: clusters D4, D8, and D9, and tbrA2: clusters D3 and D10), indicating that these accessions have an independent hybrid origin from different parental diploids.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

RV, YT, and IP conceived and designed the study. YT, AS, and LK performed the experiments and collected the material. YT, RV, AS, IP, and VH analyzed the data. RV and VH wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Ministry of Education and Science of Ukraine (Grant No. 0121U111109), Alexander v. Humboldt Foundation (Bonn, Germany) Fellowship for RV, and DAAD (German Academic Exchange Service) Research Fellowships for IP and YT.

ACKNOWLEDGMENTS

We are thankful to Frank van Caekenberghe (Meise Botanical Garden, Belgium), Tatiana Derevenko (Botanical Garden of the Chernivtsi National University, Ukraine), David Orr (Waimea Arboretum and Botanical Garden, Hawaii, United States), and Edith Stabentheiner (University of Graz, Austria) for providing the plant material. We are particularly grateful to the staff at Frontiers for exempting our article from paying the publication fee. We decided to donate this money to support the Ukrainian army and the needs of refugees from Russian aggression.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.852406/full#supplementary-material

REFERENCES

Achakkagari, S. R., Bozan, I., Anglin, N. L., Ellis, D., Tai, H. H., and Strömvik, M. V. (2021). Complete mitogenome assemblies from a panel of 13 diverse potato taxa. *Mitochondrial DNA B Resour.* 6, 894–897. doi: 10.1080/23802359.2021.1886016

Achakkagari, S. R., Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. V. (2020). Complete plastome assemblies from a panel of 13 diverse potato taxa. *PLoS One* 15, 894–897. doi: 10.1080/23802359.2021.1886016

Alexandrov, O. S., Razumova, O. V., and Karlov, G. I. (2021). A comparative study of 5S rDNA non-transcribed spacers in Elaeagnaceae species. *Plants* 10:4. doi: 10.3390/10010004

Anderson, G. J., and Bernardello, L. M. (1991). The relationships of *Solanum cochoae* (Solanaceae), a new species from Peru. *Novon* 1, 127–133. doi: 10.2307/3391369

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., and Zimmer, E. (1980). Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proc. Natl. Acad. Sci. U.S.A.* 77, 7323–7327. doi: 10.1073/pnas.77.12.7323

Aubriot, X., Knapp, S., Syfert, M. M., Poczai, P., and Buerki, S. (2018). Shedding new light on the origin and spread of the brinjal eggplant (*Solanum melongena* L.) and its wild relatives. *Am. J. Bot.* 105, 1175–1187. doi: 10.1002/ajb2.1133

Aubriot, X., Singh, P., and Knapp, S. (2016). Tropical Asian species show the Old World clade of "spiny *Solanums*" (the *Leptostemonum* Clade: Solanaceae) is not monophyletic. *Bot. J. Linn. Soc.* 180, 1–27. doi: 10.1111/boj.12412

Barman, A. S., Singh, M., Singh, R. K., and Lal, K. K. (2016). Evidence of birth-and-death evolution of 5S rRNA gene in *Channa* species (Teleostei, Perciformes). *Genetica* 144, 723–732. doi: 10.1007/s10709-016-9938-6

Baum, B. R., Edwards, T., Mamuti, M., and Johnson, D. A. (2012). Phylogenetic relationships among the polyploid and diploid *Aegilops* species inferred from the nuclear 5S rDNA sequences (Poaceae: Triticeae). *Genome* 55, 177–193. doi: 10.1139/g2012-006

Bean, A. R. (2004). The taxonomy and ecology of *Solanum* subg. *Leptostemonum* (Dunal) Bitter (Solanaceae) in Queensland and far north-eastern New South Wales. *Austrobaileya* 6, 639–816.

Bean, A. R. (2012). A taxonomic revision of the *Solanum echinatum* group (Solanaceae). *Phytotaxa* 57, 33–50. doi: 10.11646/phytotaxa.57.1.6

Bean, A. R. (2013). A taxonomic review of the *Solanum sturtianum* subgroup of subgenus *Leptostemonum* (Solanaceae). *Nuytsia* 23, 129–161.

Blöch, C., Weiss-Schneeweiss, H., Schneeweiss, G. M., Barfuss, M. H., Rebernig, C. A., Villaseñor, J. L., et al. (2009). Molecular phylogenetic analyses of nuclear and plastid DNA sequences support dysploid and polyploid chromosome number changes and reticulate evolution in the diversification of *Melampodium* (Millerieae, Asteraceae). *Mol. Phylogenet. Evol.* 53, 220–233. doi: 10.1016/j.ympev.2009.02.021

Bohs, L. (1995). Transfer of *Cyphomandra* (Solanaceae) and its species to *Solanum*. *Taxon* 44, 583–587. doi: 10.2307/1223500

Bohs, L. (2005). Major clades in *Solanum* based on *ndhF* sequences. A festschrift for William G. D'Arcy: the legacy of a taxonomist. *Monogr. Syst. Bot. Missouri Bot. Gard.* 104, 24–49.

Borisjuk, N., Borisjuk, L., Komarnytsky, S., Timeva, S., Hemleben, V., Gleba, Y., et al. (2000). The tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes. *Nat. Biotechnol.* 18, 1303–1306. doi: 10.1038/82430

Bozan, I. (2021). *Genome Analysis of the Diploid wild Potato Solanum bukasovii*. Master thesis. Montreal: McGill University, 13–61.

Bukasov, S. M. (1939). The origin of potato species. *Physis* 18, 41–46.

Bukasov, S. M. (1978). Systematics of the potato. *Trudy Prikl. Bot. Genet. Selekcii* 62, 3–35.

Bustos, A., Figueroa, R. I., Sixto, M., Bravo, I., and Cuadrado, A. (2020). The 5S rRNA genes in *Alexandrium*: their use as a FISH chromosomal marker in studies of the diversity, cell cycle and sexuality of dinoflagellates. *Harmful Algae* 98:101903. doi: 10.1016/j.hal.2020.101903

Cardoni, S., Piredda, R., Denk, T., Grimm, G. W., Papageorgiou, A. C., Schulze, E. D., et al. (2021). High-throughput sequencing of 5S-IGS rDNA in *Fagus* L. (Fagaceae) reveals complex evolutionary patterns and hybrid origin of modern species. *bioRxiv [Preprint]* doi: 10.1101/2021.02.26.433057

Chen, G., Stepanenko, A., and Borisjuk, N. (2021). Mosaic arrangement of the 5S rDNA in the aquatic plant *Landoltia punctate* (Lemnaceae). *Front. Plant Sci.* 12:678689. doi: 10.3389/fpls.2021.678689

Chiarini, F., Sazatornil, F., and Bernardello, G. (2018). Data reassessment in a phylogenetic context gives insight into chromosome evolution in the giant genus *Solanum* (Solanaceae). *Syst. Biodivers.* 16, 397–416. doi: 10.1080/14772000.2018.1431320

Clark, J. L., Nee, M., Bohs, L., and Knapp, S. (2015). A revision of *Solanum* section *Aculeigerum* (the *Solanum wendlandii* group, Solanaceae). *Syst. Bot.* 40, 1102–1136. doi: 10.1600/036364415X690148

Coen, E. S., Thoday, J. M., and Dover, G. (1982). Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature* 295, 564–568. doi: 10.1038/295564a0

Crisp, M. D., Appels, R., Smith, F. M., and Keys, W. M. S. (1999). Phylogenetic evaluation of 5S ribosomal RNA gene and spacer in the *Callistachys* group (Fabaceae: Mirbelieae). *Plant Syst. Evol.* 218, 33–42. doi: 10.1007/BF01087032

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendell, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

D'Arcy, W. G. (1991). "The solanaceae since 1976, with a review of its biogeography," in *Solanaceae III: Taxonomy, Chemistry, Evolution*, eds J. G. Hawkes, R. N. Lester, M. Nee, and N. Estrada (Kew: Royal Botanic Gardens), 75–137.

Daunay, M. C., Lester, R. N., and Ano, G. (2001). "Cultivated eggplants," in *Tropical Plant Breeding*, eds A. Charrier, M. Jacquot, S. Hamon, and D. Nicolas (Oxford: Oxford University Press), 200–225.

Daunay, M. C., Salinier, J., and Aubriot, X. (2019). "Crossability and diversity of eggplants and their wild relatives," in *The Eggplant Genome, Compendium of Plant Genomes*, ed. M. A. Chapman (Cham: Springer), doi: 10.1007/978-3-319-99208-2_11

Davidjuk, Y. M., Hemleben, V., and Volkov, R. A. (2010). Structural organization of 5S rDNA of eggplant, *Solanum melongena* L. *Biol. Syst.* 2, 3–6.

Davidjuk, Y. M., Moloda, O. O., and Volkov, R. A. (2013). Molecular organization of 5S rDNA of *Solanum betaceum* Cav. *Bul. Vavilov Soc. Genet. Breed. Ukr.* 11, 14–19.

Davis, R. W., and Hurter, P. J. H. (2012). *Solanum albostellatum* (Solanaceae), a new species from the Pilbara bioregion of Western Australia Nuytsia. *Nuytsia* 22, 329–334.

de Souza, T. B., Gaeta, M. L., Martins, C., and Vanzela, A. L. L. (2020). IGS sequences in *Cestrum* present AT-and GC-rich conserved domains, with strong regulatory potential for 5S rDNA. *Mol. Biol. Rep.* 47, 55–66. doi: 10.1007/s11033-019-05104-y

Doganlar, S., Frary, A., Daunay, M.-C., Lester, R. N., and Tanksley, S. D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. *Genetics* 161, 1697–1711. doi: 10.1093/genetics/161.4.1697

Douet, J., and Tourmente, S. (2007). Transcription of the 5S rRNA heterochromatic genes is epigenetically controlled in *Arabidopsis thaliana* and *Xenopus laevis*. *Heredity* 99, 5–13. doi: 10.1038/sj.hdy.6800964

Dunal, M. F. (1852). "Solanaceae," in *Prodromus Systematis Naturalis Regni Vegetabilis*, Vol. 13, ed. A. P. de Candolle (Paris: V. Masson), 1–690. doi: 10.2307/j.ctt16vj2hs.4

Echeverría-Londoño, S., Särkinen, T., Fenton, I. S., Purvis, A., and Knapp, S. (2020). Dynamism and context-dependency in diversification of the megadiverse plant genus *Solanum* (Solanaceae). *J. Syst. Evol.* 58, 767–782. doi: 10.1111/jse.12638

Falistocco, E., Passeri, V., and Marconi, G. (2007). Investigations of 5S rDNA of *Vitis vinifera* L.: sequence analysis and physical mapping. *Genome* 50, 927–938. doi: 10.1139/g07-070

Fory Sánchez, P. A., Sánchez Mosquera, I., Bohórquez Cháux, A., Ramírez, H., Medina Cano, C., and Lobo Arias, M. L. (2010). Genetic variability of the Colombian collection of Lulo (*Solanum quitoense Lam.*) and related species of section *Lasiocarpa*. *Rev. Fac. Nac. Agron. Medellín* 63, 5465–5476.

Frodin, D. G. (2004). History and concepts of big plant genera. *Taxon* 53, 753–776. doi: 10.2307/4135449

Fulnecek, J., Lim, K. Y., Leitch, A. R., Kovarik, A., and Matyasek, R. (2002). Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species. *Heredity* 88, 19–25. doi: 10.1038/sj.hdy.6800001

Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., et al. (2021). Phylogenomic data reveal hard polytomies across the backbone of the large genus *Solanum* (Solanaceae). *bioRxiv [Preprint]* doi: 10.1101/2021.03.25.436973

Galián, J. A., Rosato, M., and Rosselló, J. A. (2014). Partial sequence homogenization in the 5S multigene families may generate sequence chimeras and spurious results in phylogenetic reconstructions. *Syst. Biol.* 63, 219–230. doi: 10.1093/sysbio/syt101

Garcia, S., Garnatje, T., and Kovarik, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121, 389–394. doi: 10.1007/s00412-012-0368-7

Gavrilenko, T., Antonova, O., Shuvalova, A., Krylova, E., Alpatyeva, N., Spooner, D. M., et al. (2013). Genetic diversity and origin of cultivated potatoes based on plastid microsatellite polymorphism. *Genet. Resour. Crop Evol.* 60, 1997–2015. doi: 10.1007/s10722-013-9968-1

Hardigan, M. A., Bamberg, J., Buell, C. R., and Douches, D. S. (2015). Taxonomy and genetic differentiation among wild and cultivated germplasm of *Solanum* sect. *Petota*. *Plant Genome* 8:elantgenome2014.06.0025. doi: 10.3835/plantgenome2014.06.0025

Hasterok, R., Wolny, E., Hosiawa, M., Kowalczyk, M., Kulak-Ksiazczyk, S., Ksiazczyk, T., et al. (2006). Comparative analysis of rDNA distribution in chromosomes of various species of Brassicaceae. *Ann. Bot.* 97, 205–216. doi: 10.1093/aob/mcj031

Hawkes, J. G. (1962). The origin of *Solanum juzepczukii* Buk. and *S. curtilobum* Juz. et Buk. *Z. Pflanzenzüchtung* 47, 1–14.

Hawkes, J. G. (1990). The potato: evolution, biodiversity and genetic resources. *Am. J. Pot. Res.* 67, 733–735. doi: 10.1007/BF03044023

Hemleben, V., and Werts, D. (1988). Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). *Gene* 62, 165–169. doi: 10.1016/0378-1119(88)90591-4

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). *Glob. Ecol. Biogeogr.* 16, 485–495. doi: 10.1111/j.1466-8238.2007.00308

Hosaka, K. (1995). Successive domestication and evolution of the Andean potatoes as revealed by chloroplast DNA restriction endonuclease analysis. *Theor. Appl. Genet.* 90, 356–363. doi: 10.1007/BF00221977

Huang, B., Ruess, H., Liang, Q., Colleoni, C., and Spooner, D. M. (2019). Analyses of 202 plastid genomes elucidate the phylogeny of *Solanum* section *Petota*. *Sci. Rep.* 9:4454. doi: 10.1038/s41598-019-40790-5

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Ishchenko, O. O., Bednarska, I. O., and Panchuk, I. I. (2021). Application of 5S ribosomal DNA for molecular taxonomy of subtribe Loliinae (Poaceae). *Cytol. Genet.* 55, 10–18. doi: 10.3103/S0095452721010096

Ishchenko, O. O., Mel'nyk, V. M., Parnikoza, I. Y., Budzhak, V. V., Panchuk, I. I., Kunakh, V. A., et al. (2020). Molecular organization of 5S ribosomal DNA and taxonomic status of *Avenella flexuosa* (L.) Drejer (Poaceae). *Cytol. Genet.* 54, 505–513. doi: 10.3103/S0095452720060055

Ishchenko, O. O., Panchuk, I. I., Andreev, I. O., Kunakh, V. A., and Volkov, R. A. (2018). Molecular organization of 5S ribosomal DNA of *Deschampsia antarctica*. *Cytol. Genet.* 52, 416–421. doi: 10.3103/S0095452718060105

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* 20, 1160–1166. doi: 10.1093/bib/bbx108

Knapp, S., Aubriot, X., and Prohens, J. (2019). "Eggplant (*Solanum melongena* L.): taxonomy and relationships," in *The Eggplant Genome. Compendium of Plant Genomes*, ed. M. Chapman (Cham: Springer), 11–22. doi: 10.1007/978-3-319-99208-2_2

Knapp, S., Bohs, L., Nee, M., and Spooner, D. M. (2004). Solanaceae — a model for linking genomics with biodiversity. *Comp. Funct. Genomics* 5, 285–291. doi: 10.1002/cfg.393

Knapp, S., Sagona, E., Carbonell, A. K. Z., and Chiarini, F. (2017). A revision of the *Solanum elaeagnifolium* clade (Elaeagnifolium clade; subgenus *Leptostemonum*, Solanaceae). *PhytoKeys* 84, 1–104. doi: 10.3897/phytokeys.84.12695

Komarova, N. Y., Grimm, G. W., Hemleben, V., and Volkov, R. A. (2008). Molecular evolution of 35S rDNA and taxonomic status of *Lycopersicon* within *Solanum* sect. *Petota*. *Plant Syst. Evol.* 276, 59–71. doi: 10.1007/s00606-008-0091-2

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Layat, E., Sáez-Vásquez, J., and Tourmente, S. (2012). Regulation of Pol I-transcribed 45S rDNA and Pol III-transcribed 5S rDNA in *Arabidopsis*. *Plant Cell Physiol.* 53, 267–276. doi: 10.1093/pcp/pcr177

Levin, R. A., Myers, N. R., and Bohs, L. (2006). Phylogenetic relationships among the "spiny *Solanums*" (*Solanum* subgenus *Leptostemonum*, Solanaceae). *Am. J. Bot.* 93, 157–169. doi: 10.3732/ajb.93.1.157

Mahelka, V., Kopecký, D., and Baum, B. R. (2013). Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass *Thinopyrum intermedium* (Triticeae). *Mol. Biol. Evol.* 30, 2065–2086. doi: 10.1093/molbev/mst106

Martine, C. T., Anderson, G. J., and Les, D. H. (2009). Gender-bending aubergines: molecular phylogenetics of cryptically dioecious *Solanum* in Australia. *Aust. Syst. Bot.* 22, 107–120. doi: 10.1071/SB07039

Martine, C. T., Cantley, J. T., Frawley, E. S., Butler, A. R., and Jordon-Thaden, I. E. (2016). New functionally dioecious bush tomato from Northwestern Australia, *Solanum ossicruentum*, may utilize "trample burr" dispersal. *PhytoKeys* 63, 19–29. doi: 10.3897/phytokeys.63.7743

Martine, C. T., Jordon-Thaden, I. E., McDonnell, A. J., Cantley, J. T., Hayes, D. S., Roche, M. D., et al. (2019). Phylogeny of the Australian *Solanum dioicum* group using seven nuclear genes, with consideration of Symon's fruit and seed dispersal hypotheses. *PLoS One* 14:e0207564. doi: 10.1371/journal.pone.0207564

Martine, C. T., Vanderpool, D., Anderson, G. J., and Les, D. H. (2006). Phylogenetic relationships of andromonoecious and dioecious Australian species of *Solanum* subgenus *Leptostemonum* section *Melongena*: inferences from ITS sequence data. *Syst. Bot.* 31, 410–420. doi: 10.1600/036364406777585801

Matyasek, R., Dobesova, E., Huska, D., Ježková, I., Soltis, D. E., Soltis, P. S., et al. (2016). Interpopulation hybridization generates meiotically stable rDNA epigenetic variants in allotetraploid *Tragopogon mirus*. *Plant J.* 85, 362–377. doi: 10.1111/tpj.13110

Matyasek, R., Fulnecek, J., Lim, K. Y., Leitch, A. R., and Kovarik, A. (2002). Evolution of 5S rDNA unit arrays in the plant genus *Nicotiana* (Solanaceae). *Genome* 45, 556–562. doi: 10.1139/g02-017

Mlinarec, J., Franjević, D., Bočkor, L., and Besendorfer, V. (2016). Diverse evolutionary pathways shaped 5S rDNA of species of tribe Anemoneae (Ranunculaceae) and reveal phylogenetic signal. *Bot. J. Linn. Soc.* 182, 80–99. doi: 10.1111/boj.12452

Nee, M. (1999). "Synopsis of *Solanum* in the New World," in *Solanaceae IV: Advances in Biology and Utilization*, eds M. Nee, D. E. Symon, R. N. Lester, and J. P. Jessop. (Kew: Royal Botanic Gardens), 285–333.

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the solanaceae. *Taxon* 57, 1159–1181. doi: 10.1002/tax.574010

Park, Y. K., Park, K. C., Park, C. H., and Kim, N. S. (2000). Chromosomal localization and sequence variation of 5S rRNA gene in five *Capsicum* species. *Mol. Cells* 10, 18–24. doi: 10.1007/s10059-000-0018-4

Pinhal, D., Yoshimura, T. S., Araki, C. S., and Martins, C. (2011). The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evol. Biol.* 11:151. doi: 10.1186/1471-2148-11-151

Poczai, P., Hyvönen, J., and Symon, D. E. (2011). Phylogeny of kangaroo apples (*Solanum* subg. *Archaesolanum*, Solanaceae). *Mol. Biol. Rep.* 38, 5243–5259. doi: 10.1007/s11033-011-0675-8

Pontes, O., Neves, N., Silva, M., Lewis, M. S., Madlung, A., Comai, L., et al. (2004). Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 18240–18245. doi: 10.1073/pnas.0407258102

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/BF02772108

Prohens, J., Plazas, M., Raigón, M. D., Seguí-Simarro, J. M., Stommel, J. R., and Vilanova, S. (2012). Characterization of interspecific hybrids and first backcross generations from crosses between two cultivated eggplants (*Solanum melongena*

and *S. aethiopicum* Kumba group) and implications for eggplant breeding. *Euphytica* 186, 517–538. doi: 10.1007/s10681-012-0652-x

Qin, Q. B., Liu, Q. W., Wang, C. Q., Cao, L., Zhou, Y. W., Qin, H., et al. (2019). Molecular organization and chromosomal localization analysis of 5S rDNA clusters in autotetraploids derived from *Carassius auratus* Red Var. (♀) × *Megalobrama amblycephala* (♂). *Front. Genet.* 10:437. doi: 10.3389/fgene.2019.00437

Randell, B. R., and Symon, D. E. (1976). Chromosome numbers in Australian *Solanum* species. *Aust. J. Bot.* 24, 369–379. doi: 10.1071/BT9760369

Reuter, J. S., and Mathews, D. H. (2010). RNA structure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129. doi: 10.1186/1471-2105-11-129

Rodríguez, F., Ghislain, M., Clausen, A. M., Jansky, S. H., and Spooner, D. M. (2010). Hybrid origins of cultivated potatoes. *Theor. Appl. Genet.* 121, 1187–1198. doi: 10.1007/s00122-010-1422-6

Röser, M., Winterfeld, G., Grebenstein, B., and Hemleben, V. (2001). Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). *Mol. Phylogent. Evol.* 21, 198–217. doi: 10.1006/mpev.2001.1003

Särkinen, T., Barboza, G. E., and Knapp, S. (2015). True black nightshades: phylogeny and delimitation of the morelloid clade of *Solanum*. *Taxon* 64, 945–958. doi: 10.12705/645.5

Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13:214. doi: 10.1186/1471-2148-13-214

Schmiediche, P. E., Hawkes, J. G., and Ochoa, C. M. (1980). Breeding of the cultivated potato species *Solanum x juzepczukii* Buk. and *Solanum x curtilobum* Juz. et Buk. *Euphytica* 29, 685–704. doi: 10.1007/BF00023216

Seithe, A. (1962). Die Haararten der Gattung *Solanum* L. und ihre taxonomische Verwertung. *Bot. Jahrb. Syst. Pflanzengesch. Pflanzengeogr.* 81, 261–336.

Simeone, M. C., Cardoni, S., Piredda, R., Imperatori, F., Avishai, M., Grimm, G. W., et al. (2018). Comparative systematics and phylogeography of *Quercus* section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ* 6:e5793. doi: 10.7717/peerj.5793

Simon, L., Rabanal, F. A., Dubos, T., Oliver, C., Lauber, D., Poulet, A., et al. (2018). Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Res.* 46, 3019–3033. doi: 10.1093/nar/gky163

Song, J., Shi, L., Li, D., Sun, Y., Niu, Y., Chen, Z., et al. (2012). Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. *PLoS One* 7:e43971. doi: 10.1371/journal.pone.0043971

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot. Rev.* 80, 283–383. doi: 10.1007/s12229-014-9146-y

Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., and Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14694–14699. doi: 10.1073/pnas.0507400102

Stone, B. C. (1970). The flora of Guam: a manual for the identification of the vascular plants of the island. *Micronesica* 6, 1–659.

Sun, F. J., and Caetano-Anollés, G. (2009). The evolutionary history of the structure of 5S ribosomal RNA. *J. Mol. Evol.* 69, 430–443. doi: 10.1007/s00239-009-9264-z

Sun, Y. L., Kang, H. M., Kim, Y. S., Baek, J. P., Zheng, S. L., Xiang, J. J., et al. (2014). Tomato (*Solanum lycopersicum*) variety discrimination and hybridization analysis based on the 5S rRNA region. *Biotechnol. Biotechnol. Equip.* 28, 431–437. doi: 10.1080/13102818.2014.928499

Tucker, S., Vitins, A., and Pikaard, C. S. (2010). Nucleolar dominance and ribosomal RNA gene silencing. *Curr. Opin. Cell Biol.* 22, 351–356. doi: 10.1016/j.ceb.2010.03.009

Tynkevich, Y. O., Nevelska, A. O., Chorney, I. I., and Volkov, R. A. (2015). Organization and variability of the 5S rDNA intergenic spacer of *Lathyrus venetus*. *Bull. Vavilov Soc. Genet. Breed. Ukr.* 13, 81–87.

Tynkevich, Y. O., and Volkov, R. A. (2014). Structural organization of 5S ribosomal DNA in *Rosa rugosa*. *Cytol. Genet.* 48, 1–6. doi: 10.3103/S0095452714010095

Tynkevich, Y. O., and Volkov, R. A. (2019). 5S ribosomal DNA of distantly related *Quercus* species: molecular organization and taxonomic application. *Cytol. Genet.* 53, 459–466. doi: 10.3103/S0095452719060100

Ugent, D. (1970). The potato. *Science* 170, 1161–1166.

Vaillant, I., Tutois, S., Cuvillier, C., Schubert, I., and Tourmente, S. (2007). Regulation of *Arabidopsis thaliana* 5S rRNA genes. *Plant Cell Physiol.* 48, 745–752. doi: 10.1093/pcp/pcm043

Vizoso, M., Vierna, J., González-Tizón, A. M., and Martínez-Lage, A. (2011). The 5S rDNA gene family in mollusks: characterization of transcriptional regulatory regions, prediction of secondary structures, and long-term evolution, with special attention to mytilidae mussels. *J. Heredity* 102, 433–447. doi: 10.1093/jhered/esr046

Volkov, R. A., Borisjuk, N. V., Panchuk, I. I., Schweizer, D., and Hemleben, V. (1999b). Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. *Mol. Biol. Evol.* 16, 311–320. doi: 10.1093/oxfordjournals.molbev.a026112

Volkov, R. A., Bachmair, A., Panchuk, I. I., Kostyshyn, S. S., and Schweizer, D. (1999a). 25S-18S rDNA intergenic spacer of *Nicotiana sylvestris* (Solanaceae): primary and secondary structure analysis. *Plant Syst. Evol.* 218, 89–97. doi: 10.1007/bf01087037

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Syst. Biodivers.* 5, 261–276. doi: 10.1017/S1477200007002447

Volkov, R. A., Komarova, N. Y., Panchuk, I. I., and Hemleben, V. (2003). Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. *Petota* (genus *Solanum*). *Mol. Phylogenet. Evol.* 29, 187–202. doi: 10.1016/s1055-7903(03)00092-7

Volkov, R. A., Medina, F. J., Zentgraf, U., and Hemleben, V. (2004). Molecular cell biology: organization and molecular evolution of rDNA, nucleolar dominance, and nucleolus structure. *Prog. Bot.* 65, 106–146. doi: 10.1007/978-3-642-18819-0_5

Volkov, R. A., Panchuk, I. I., Borisjuk, N. V., Hosiawa-Baranska, M., Maluszynska, J., and Hemleben, V. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol.* 17:21. doi: 10.1186/s12870-017-0978-6

Volkov, R. A., Zanke, C., Panchuk, I. I., and Hemleben, V. (2001). Molecular evolution of 5S rDNA of *Solanum species* (sect. *Petota*): application for molecular phylogeny and breeding. *Theor. Appl. Genet.* 103, 1273–1282. doi: 10.1007/s001220100670

Vorontsova, M. S., and Knapp, S. (2016). A revision of the "spiny *Solanums*", *Solanum* subgenus *Leptostemonum* (Solanaceae) in Africa and Madagascar. *Syst. Bot. Monogr.* 99, 1–436. doi: 10.3897/phytokeys.145.48531

Vozárová, R., Herklotz, V., Kovaøík, A., Tynkevich, Y. O., Volkov, R. A., Ritz, C. M., et al. (2021). Ancient origin of two 5S rDNA families dominating in the genus *Rosa* and their behavior in the *Canina*-type meiosis. *Front. Plant Sci.* 12:643548. doi: 10.3389/fpls.2021.643548

Weese, T. L., and Bohs, L. (2007). A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Syst. Bot.* 32, 445–463. doi: 10.1600/036364407781179671

Whalen, M. D. (1984). Conspectus of species groups in *Solanum* subgenus *Leptostemonum*. *Gent. Herbarum* 12, 179–282. doi: 10.2307/3391723

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fpls.2021.768347

To Be or Not to Be Expressed: The First Evidence of a Nucleolar Dominance Tissue-Specificity in *Brachypodium hybridum*

Natalia Borowska-Zuchowska[1], Ewa Robaszkiewicz[1], Serhii Mykhailyk[1], Joanna Wartini[1], Artur Pinski[1], Ales Kovarik[2] and Robert Hasterok[1]*

[1] Plant Cytogenetics and Molecular Biology Group, Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, Katowice, Poland, [2] Department of Molecular Epigenetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia

OPEN ACCESS

Edited by:
Roman A. Volkov, Chernivtsi University, Ukraine

Reviewed by:
Zixian Zeng, Sichuan Normal University, China
Magdalena Vaio, Universidad de la República, Uruguay
Roman Matyasek, Institute of Biophysics, Academy of Sciences of the Czech Republic, Czechia

***Correspondence:**
Natalia Borowska-Zuchowska
natalia.borowska@us.edu.pl

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *31 August 2021*
Accepted: *09 November 2021*
Published: *06 December 2021*

Citation:
Borowska-Zuchowska N, Robaszkiewicz E, Mykhailyk S, Wartini J, Pinski A, Kovarik A and Hasterok R (2021) To Be or Not to Be Expressed: The First Evidence of a Nucleolar Dominance Tissue-Specificity in Brachypodium hybridum. Front. Plant Sci. 12:768347. doi: 10.3389/fpls.2021.768347

Nucleolar dominance (ND) is an epigenetic, developmentally regulated phenomenon that describes the selective inactivation of 35S rDNA loci derived from one progenitor of a hybrid or allopolyploid. The presence of ND was documented in an allotetraploid grass, *Brachypodium hybridum* (genome composition DDSS), which is a polyphyletic species that arose from crosses between two putative ancestors that resembled the modern *B. distachyon* (DD) and *B. stacei* (SS). In this work, we investigated the developmental stability of ND in *B. hybridum* genotype 3-7-2 and compared it with the reference genotype ABR113. We addressed the question of whether the ND is established in generative tissues such as pollen mother cells (PMC). We examined condensation of rDNA chromatin by fluorescence *in situ* hybridization employing state-of-art confocal microscopy. The transcription of rDNA homeologs was determined by reverse-transcription cleaved amplified polymorphic sequence analysis. In ABR113, the ND was stable in all tissues analyzed (primary and adventitious root, leaf, and spikes). In contrast, the 3-7-2 individuals showed a strong upregulation of the S-genome units in adventitious roots but not in other tissues. Microscopic analysis of the 3-7-2 PMCs revealed extensive decondensation of the D-genome loci and their association with the nucleolus in meiosis. As opposed, the S-genome loci were always highly condensed and localized outside the nucleolus. These results indicate that genotype-specific loss of ND in *B. hybridum* occurs probably after fertilization during developmental processes. This finding supports our view that *B. hybridum* is an attractive model to study ND in grasses.

Keywords: nucleolar dominance, 35S rDNA, secondary constriction, *Brachypodium*, allopolyploidy, 3D-FISH, rRNA gene expression

INTRODUCTION

Allopolyploidy is an interspecific hybridization followed by chromosome doubling and it is believed to play an essential role in angiosperm evolution (Mason and Wendel, 2020). The genomic investigations of many plant allopolyploids of different evolutionary ages have revealed that the newly formed allopolyploid is subjected to immediate (short-term) and long-term

changes that operate at both genomic and transcriptomic levels (Wendel et al., 2018). The short-term consequences of allopolyploidy, including meiotic irregularities, chromosomal aberrations, transposon proliferation, widespread loss of non-coding sequences, and gene expression alterations, may constitute a considerable challenge for most newly formed allopolyploids (Ozkan et al., 2001; Shaked et al., 2001; Nicolas et al., 2009; Chester et al., 2012; Yoo et al., 2014). However, once they stabilize over a longer evolutionary timeframe, allopolyploids may be characterized by an increased phenotypic variability compared to their diploid progenitors and, therefore, may have a higher capacity to reach new environmental niches (Soltis et al., 2015).

The tandemly repeated 35S rRNA genes have been the subject of many studies on plant hybrids and allopolyploids, particularly for inferring the molecular background of nucleolar dominance (ND; previously known as "differential amphiplasty"), a widespread phenomenon that was initially documented by Navashin (1934) in *Crepis* hybrids. In the allopolyploids and hybrids that exhibit ND, the 35S rDNA loci that are inherited from one progenitor are dominant over the others (Pikaard, 2000; Volkov et al., 2004; Ge et al., 2013; Symonova, 2019). It is well known that ND is maintained in an epigenetic manner and that the under-dominant rDNA loci are hypermethylated, especially within the rRNA gene promoters (Matyasek et al., 2016) and are characterized by repressive epigenetic marks, e.g., the dimethylated lysine 9 of histone H3 (H3K9me2) (Neves et al., 1995; Chen and Pikaard, 1997a; Houchins et al., 1997; Lawrence et al., 2004; Borowska-Zuchowska and Hasterok, 2017). Moreover, it was shown that the inactivation of the rRNA gene loci is accompanied by an RNA-dependent DNA methylation pathway (RdDM) (Preuss et al., 2008; Costa-Nunes et al., 2010) and the deacetylation of the histones (Earley et al., 2006). However, the question of how specific rDNA loci are selected for inactivation still remains unanswered.

Nucleolar dominance is a common phenomenon among grass allopolyploids, including the economically important cereals, e.g., bread wheat (*Triticum aestivum* L.) (Cermeno et al., 1984; Guo and Han, 2014) and triticale (Lacadena et al., 1984; Vieira et al., 1990; Neves et al., 1995). Therefore, most molecular and cytogenetic research concerning the molecular basis of ND among monocots has been limited to wheat and its derivatives. Considering the massive genome size of wheat of ∼16.6 Gbp/1C (Dolezel et al., 2018), which is composed of three subgenomes ($2n$ = 42; genome composition AABBDD) in which many 35S rDNA loci are present, studies of ND in this species are tremendously complicated. Thus, it is necessary to find a monocot model system that will be amenable in ND studies.

In 2008, the presence of ND was observed in the root-tip cells of a small- and simple-genome allotetraploid grass, *Brachypodium hybridum* ($2n$ = 30; DDSS). Its putative ancestral species are two diploids that resembled the modern *B. distachyon* ($2n$ = 10; DD) and *B. stacei* ($2n$ = 20; SS). It was documented using silver staining followed by fluorescence *in situ* hybridization (FISH) with 25S rDNA as the probe that the S-genome-derived 35S rDNA loci are not transcribed in the root-tip meristematic cells (Idziak and Hasterok, 2008). The further analyses of the ND in *B. hybridum* were mainly focused on the reference genotype ABR113. Cytomolecular studies that employed immunostaining approaches revealed that the D- and S-genome-inherited 35S rDNA loci were differentiated by distinct epigenetic patterns. The D-genome-originated loci were characterized by significantly lower DNA methylation levels compared to the S-genome ones and were enriched in the acetylated isoforms of histones, e.g., H4K5ac, H4K16ac, and H3K9ac. By contrast, the S-genome loci were usually located at the nuclear periphery and were enriched in the heterochromatic H3K9me2 mark (Borowska-Zuchowska and Hasterok, 2017). Studies on the developmental regulation in ABR113 did not reveal any stage in *B. hybridum* ontogenesis in which the ND is abolished (Borowska-Zuchowska et al., 2016, 2019). We recently observed a reduction in the S-genome-like 35S rDNA copy number in 16 *B. hybridum* genotypes. *B. stacei*-derived rDNA contributions to total rDNA varied from 7 to 39%, which suggests that the inactive loci may have gradually been eliminated during evolution (Borowska-Zuchowska et al., 2020). Such a variation in ancestral rDNA ratios may be related to the polyphyletic origin of *B. hybridum*. Recent studies of plastomes, nuclear single nucleotide variants, and k-mers associated with retrotransposons showed two independent origins for *B. hybridum* (∼1.4 and ∼0.14 million years ago; Gordon et al., 2020). Considering that *B. hybridum* has only one 35S rDNA locus per ancestral genome and that the whole genomic sequence of both putative ancestors and several *B. hybridum* genotypes have been published (IBI, 2010; Gordon et al., 2020), this allotetraploid grass may become a suitable model in the ND studies in monocots. However, such a model should also provide a wide range of genotypes in which there is a diversity in the rDNA structure and expression. In the current study, we applied a combination of cytogenetic and molecular methods in order to verify whether the expression status of the 35S rDNA homeologs is additive or biased at the various developmental stages in *B. hybridum* and revealed a genotype in which there is a differential expression of S-genome 35S rRNA genes in some organs.

MATERIALS AND METHODS

Plant Material

Two genotypes of *Brachypodium hybridum* were used in this study. The reference genotype ABR113 (Portugal) was obtained from the Institute of Biological, Environmental and Rural Sciences (Aberystwyth University, Aberystwyth, United Kingdom). The 3-7-2 genotype was derived from the T_1 generation of the germplasm that had been collected in Turkey (geographical location: N38°17′40.2″ E27°24′13.9″; altitude: 173 m). All of the plants were grown in pots with soil mixed with vermiculite (3:1, w/w) at 20–22°C and a 16 h photoperiod in a greenhouse.

To analyze mitotic metaphase chromosomes from the primary roots, dehusked seeds were grown on a filter paper that had been moistened with tap water for 3 days at room temperature in the dark. Whole seedlings with 1.5-3-cm-long roots were immersed in ice-cold water for 24 h and then fixed in methanol:glacial acetic

acid (3:1, v/v) at 4°C overnight and stored at −20°C. Immature spikes for the analysis of the meiocytes were collected from 2 to 2.5-month-old plants, fixed in fresh ethanol:glacial acetic acid (3:1, v/v) at room temperature. The fixative was replaced by a fresh one after 24 h. The material was stored at −20°C until use.

DNA Isolation and Southern Blot Hybridization

Total genomic DNAs (gDNAs) from the leaves from 1.5-month-old plants of *B. hybridum* were isolated using a CTAB buffer as described previously (Borowska-Zuchowska et al., 2020). The purified gDNAs (1 μg/sample) were digested with the methylation insensitive *Bgl*II restriction enzyme cutting the AGATCT motif. The digested gDNAs were separated by gel electrophoresis on 1% (w/v) agarose gel and alkali-blotted onto nylon membranes (Hybond XL, GE Healthcare Life Sciences). The membranes were hybridized with ^{32}P-labeled DNA probes (DekaLabel kit, ThermoFisher Scientific) using the ∼220 bp-long PCR product derived from the 3′ end of the 25S rDNA of tobacco as previously described (Kovarik et al., 2005). The hybridization bands were visualized with a phosphorimager (Typhoon 9410, GE Healthcare), and the radioactivity of the bands was quantified with ImageQuant software (GE Healthcare). The ancestral rDNA proportions were expressed as a percentage of signal in 6.7 kb (*B. distachyon*-specific band) or 7.8 kb + 9.2 kb (*B. stacei*-specific bands) fractions out of the total (6.7 kb + 7.8 kb + 9.2 kb) rDNA signal. At least two individuals (two biological replicates) from each *B. hybridum* genotype were analyzed.

Fluorescence *in situ* Hybridization on the Squashed Mitotic Preparations

The root meristem cytogenetic preparations, including the multi-substrate preparations, were made as described (Hasterok et al., 2006; Jenkins and Hasterok, 2007). A 2.3-kb *Cla*I subclone of the 25S rDNA of *A. thaliana* (Unfried and Gruendler, 1990) was used as the probe to detect the 35S rDNA loci. The clone was labeled with tetramethylrhodamine-5-dUTP (Roche) *via* nick translation (Hasterok et al., 2002). Fluorescence *in situ* hybridization (FISH) was performed as previously described (Idziak et al., 2011). After precipitation, the 25S rDNA probe was dissolved in a hybridization mixture containing 50% deionized formamide and 20% dextran sulfate in 2 × saline sodium citrate buffer (SSC). The hybridization mixture was pre-denatured at 75°C for 10 min and applied to the chromosome preparations. After denaturation at 75°C for 4.5 min, the preparations were allowed to hybridize in a humid chamber at 37°C for 16–20 h. Post-hybridization washes were performed in 10% formamide in 0.1 × SSC at 42°C (79% stringency). The chromosomes were mounted and counterstained in Vectashield (Vector Laboratories) containing 2.5 μg ml^{-1} of 4′,6-diamidino-2-phenylindole (DAPI; Serva). Three different individuals (three biological replicates) have been analyzed for both primary and adventitious roots. Only the metaphases with visible secondary constrictions on the D-genome chromosomes have been considered (15–26 metaphases/meristem). In the case of interphase nuclei, approx. 100 nuclei have been analyzed for each individual.

Fluorescence *in situ* Hybridization on the Polyacrylamide Pads

The meiocytes of *B. hybridum* were embedded in acrylamide gel to preserve their three-dimensional architecture. The procedure of embedding was adopted from Bass et al. (1997). Briefly, fixed anthers were collected into a 1 × Buffer A (2 × buffer A salts, 1 mM DTT, 0.2 mM spermine, 0.5 mM spermidine, 0.35 M sorbitol) and macerated using a brass tapper. Next, 10 μl of the meiocyte suspension was transferred onto a 24 × 24 mm coverslip, mixed with 5 μl of activated acrylamide stock, and immediately covered with another coverslip. After 1 h of polymerization, the coverslips were separated using a razor blade. The one containing the polyacrylamide pad was transferred to a Petri dish and the FISH reaction was performed.

FISH on the polyacrylamide pads was performed as was described in Borowska-Zuchowska et al. (2019). A pre-denatured hybridization mixture with 25S rDNA as the probe was applied to the pads, and they were denatured together at 75°C for 8 min. After renaturation (∼20 h), post-hybridization washes were performed in 20% formamide in 0.1 × SSC at 37°C. The pads were mounted in a mounting medium [1 μg ml^{-1} DAPI, 200 mM Tris-HCl pH 8, 2.3% DABCO (1,4-diazobicyclo(2,2,2)octane) and 78% glycerol] and stored at 4°C until the images were acquired.

Image Acquisition and Analysis

Images of the mitotic metaphase chromosomes after FISH were acquired using a Zeiss Axio Imager.Z.2 wide-field fluorescence microscope equipped with an AxioCam HRm monochromatic camera. The meiocytes that had been embedded in the polyacrylamide gel were optically sectioned using an Olympus FV1000 confocal microscope system equipped with a 60×/1.35 PlanApo objective. All of the image stacks were acquired by scanning from the top to the bottom of a meiocyte in 0.25 μm steps and then processed using MBF ImageJ (Wayne Rasband, National Institutes of Health, Bethesda, MD, United States).

35S rDNA Expression Analysis

The procedures followed those described by Borowska-Zuchowska et al. (2020). Briefly, the total RNA was isolated from (i) 1.5-3-cm-long primary roots that had been collected from 3-day-old seedlings (three biological replicates); (ii) fresh, greenish leaves and adventitious roots that had been collected from 1.5-month-old plants (three biological replicates); and (iii) immature spikes that had been collected from 2 to 2.5-month-old plants (two biological replicates). The RNA was isolated using a NucleoSpin® RNA Plant and Fungi kit (Macherey-Nagel). Any contaminating DNA was removed using an RNase-Free DNase set (Qiagen). Each reverse transcription reaction contained 1 μg of total RNA and 1 μl of Maxima H Minus Enzyme Mix (Thermo Fisher Scientific) and was performed according to the manufacturer's instructions. The obtained cDNAs were used as templates in the ITS1 amplification using PCR with the primers

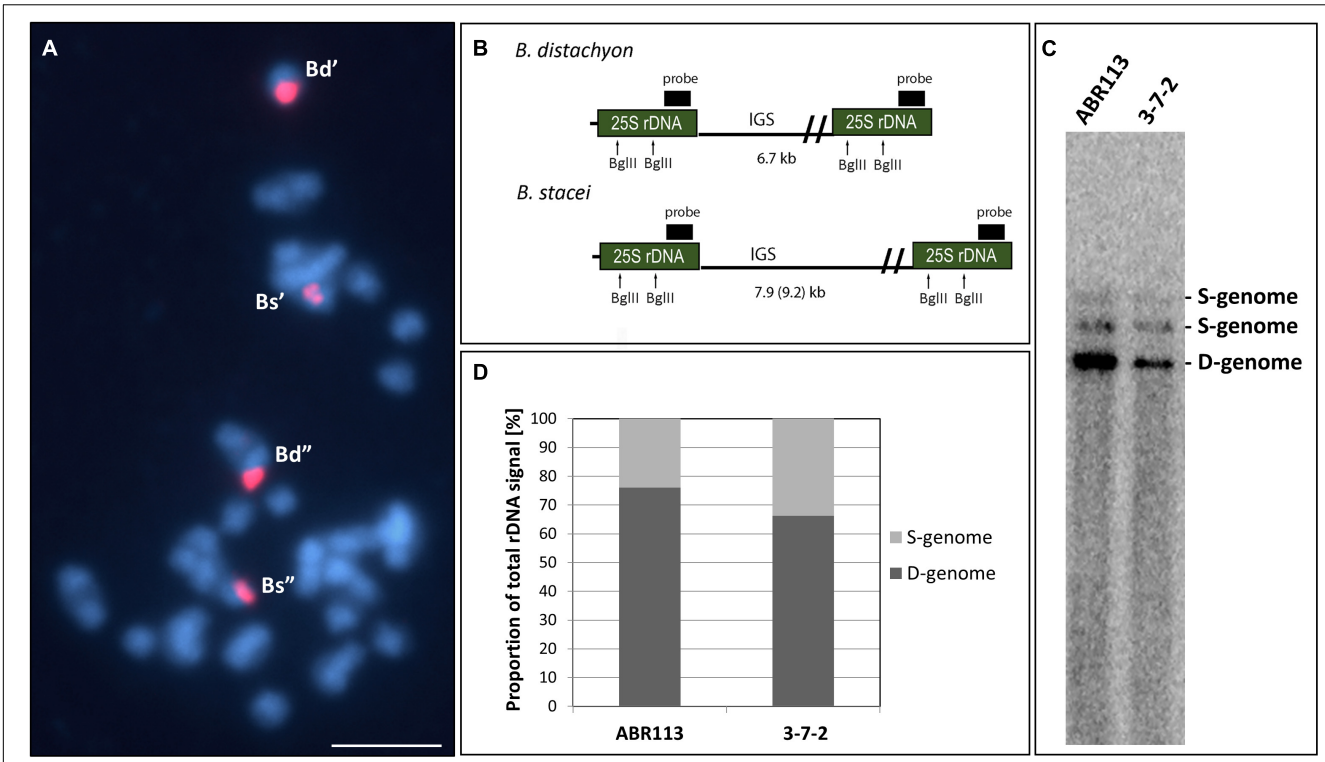


FIGURE 1 | The structure of the 35S rRNA gene loci in *B. hybridum* genotype 3-7-2. **(A)** FISH mapping of 25S rDNA (red fluorescence) in the highly condensed metaphase chromosomes of the genotype 3-7-2. Chromatin was stained with DAPI (blue fluorescence). Scale bar = 5 μm. **(B)** The *Bgl*II restriction maps in progenitor *B. distachyon* and *B. stacei* rDNA units. For simplicity, only the 25S rRNA genes are shown. **(C)** Southern blot hybridization of the genomic DNAs from the genotypes 3-7-2 and ABR113 (control) subjected to restriction with *Bgl*II. The blot was hybridized with a 220-bp-long fragment of 25S rDNA. **(D)** The quantification of 35S rDNA homeologs. The ancestral rDNA contributions are denoted as the proportion of the D- or S-genome-specific signals to the total rDNA signal.

18S and 5.8S rev (Kovarik et al., 2005). The ITS1 PCR products were subjected for restriction with *Mlu*I enzyme for 2 h at 37°C and separated using gel electrophoresis.

RESULTS AND DISCUSSION

The Structure of 35S rDNA Loci in *Brachypodium hybridum* 3-7-2

The FISH method with 25S rDNA as the probe was used to analyze the 35S rRNA gene loci number and chromosomal localization in the 3-7-2 genotype of *B. hybridum*. The analysis involved the meristematic cells from both, the primary and adventitious roots. Because the S- and D-genome-derived 35S rDNA loci occupy distinct and different positions on the chromosomes, no additional chromosome markers are required to differentiate the loci from the two ancestors. In our previous works, we showed that the *B. distachyon*-originated 35S rDNA locus is located on the terminal part of the short arm of chromosome Bd5, while the *B. stacei*-inherited locus occupies the proximal region of the significantly smaller Bs10 chromosome (Hasterok et al., 2004; Idziak and Hasterok, 2008; Borowska-Zuchowska et al., 2016; Lusinska et al., 2018). FISH revealed that the 3-7-2 genotype has two chromosomal pairs that bear 35S rDNA loci, one from each progenitor (**Figure 1A**). Their chromosomal positions were the same as in the reference genotype ABR113.

Our previous study showed that the intensity of the relative *B. stacei*-like 35S rDNA FISH signal corresponded to the S-genome rDNA contribution to the total rDNA, which was quantified based on the Southern blot hybridization results (Borowska-Zuchowska et al., 2020). To verify ancestral contributions of 35S rDNA, Southern blot hybridization with 25S rDNA as a probe was performed. The gDNAs from 3-7-2 and ABR113 (as a control) were digested with the *Bgl*II restriction enzyme. As was shown previously, there are two recognition sites for *Bgl*II in the 25S rDNA of both D- and S-genome rDNA units (**Figure 1B**; Borowska-Zuchowska et al., 2020). In **Figure 1C**, the 25S rDNA probe hybridized to the three *Bgl*II fragments in DNAs from both ABR113 and 3-7-2: (i) a fast-migrating, 6.7-kb-long fragment representing the D-genome rDNA units; (ii) 7.9-kb-long fragment representing the S-genome rDNA units in which both *Bgl*II sites were cut by the enzyme; and (iii) a slow-migrating fragment representing the full-size S-genome rDNA unit. The radioactivity of the hybridization bands was estimated by a phosphorimager revealing the quantitative relationships between the ancestral rDNAs (**Figure 1D**). The contributions of the *B. stacei*-derived rDNA units to the total rDNA were 33.8% and 23.9% in 3-7-2 and ABR113, respectively. In line with this observation, the FISH hybridization signals corresponding to the

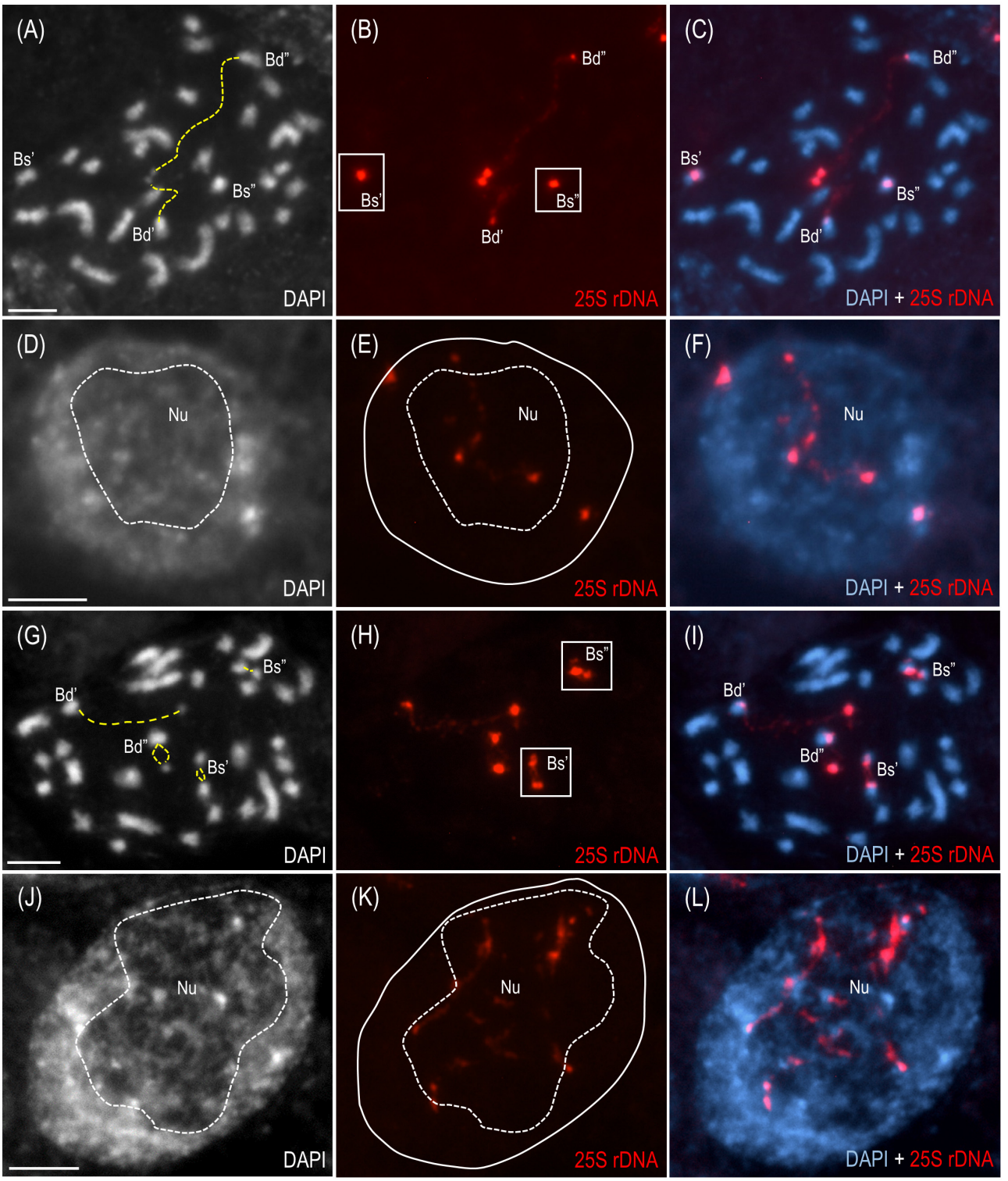


FIGURE 2 | The distribution of the 35S rDNA loci in the mitotic metaphase chromosomes and interphase nuclei of *B. hybridum* genotype 3-7-2. The cells originating from the primary **(A–F)** and adventitious **(G–L)** root apical meristems are presented.. FISH mapping of 25S rDNA (red fluorescence) in the metaphase chromosome complements **(A–C,G–I)** and interphase nuclei **(D–F,J–L)**. Bd, *B. distachyon*-inherited 35S rDNA loci; Bs, *B. stacei*-inherited 35S rDNA loci; Nu, nucleolus. The secondary constrictions on **(A,G)** are indicated by the yellow dashed lines. The position of the nucleolus on **(E,K)** is denoted by a white dashed line. Chromatin was stained with DAPI. Scale bar = 5 μm.

35S rDNA loci from the S-genome were more prominent in 3-7-2 than their counterparts in ABR113 (**Figure 1A**; data for ABR113 have already been published: Borowska-Zuchowska et al., 2016). Thus, the 3-7-2 genotype possesses a higher number of S-genome rDNA units than ABR113.

So far, the *B. stacei*-derived rDNA contributions to total rDNA vary from 7% to 39% among different *B. hybridum* genotypes (Borowska-Zuchowska et al., 2020). Since *B. hybridum* is a polyphyletic species that arose from multiple crosses between the two ancestors during Quaternary (Catalan et al., 2012; Diaz-Perez et al., 2018; Gordon et al., 2020), the fate of the rDNA homeologs in the different genotypes of this allotetraploid may be accompanied by various evolutionary scenarios. Therefore, the gradual elimination of the *B. stacei*-inherited rDNA units may be still in progress in some genotypes.

Nucleolar Dominance Is Stable in Leaves but Not in Roots of *Brachypodium hybridum*

The cytogenetic and molecular approaches were applied across four different tissues of *B. hybridum* 3-7-2 and, to the best of our knowledge, revealed that the ND in this species may be developmentally regulated for the first time. Because only the transcriptionally active 35S rDNA loci can form secondary constrictions on the metaphase chromosomes (Shaw, 2013), the presence of these structures constitutes indirect proof of the 35S rDNA activity. Thus, in the current studies, we verified whether the D- and S-genome 35S rDNA loci colocalized with the secondary constrictions on the chromosomes from the primary (**Figures 2A–C** and **Supplementary Figures 1A–C**) and adventitious roots of *B. hybridum* 3-7-2 (**Figures 2G–I** and **Supplementary Figure 2**) using FISH with 25S rDNA as a probe. Only the D-genome 35S rDNA loci formed secondary constrictions in the primary roots, while the *B. stacei*-inherited ones remained highly condensed (**Figures 2A–C** and **Supplementary Figures 1A–C**). Interestingly, all of the ancestral 35S rDNA loci formed secondary constrictions in the adventitious root-tip cells (**Figures 2G–I** and **Supplementary Figure 2**). Thus, ND was not stable in this type of roots. These observations were further corroborated at the level of the interphase nuclei from the root-tip cells. As was shown by FISH, only one pair of 35S rDNA loci corresponding to the D-genome was located adjacent to the nucleolus in the primary roots, whereas the *B. stacei*-like loci occupied the nuclear periphery and were located in the DAPI-positive, heterochromatic domains (**Figures 2D–F** and **Supplementary Figures 1D–F**). By contrast, all of the 25S rDNA FISH signals were located adjacent to or within the nucleolus in the interphase nuclei from the adventitious roots (**Figures 2J–L** and **Supplementary Figures 2A–C**), which strongly suggests that all of the 35S rDNA loci contributed to the formation of the nucleolus.

The reverse-transcription cleaved amplified polymorphic sequence (RT-CAPS) approach was used to examine the expression of the 35S rDNA homeologs at the different developmental stages of *B. hybridum*. The analysis involved RNA from the vegetative organs (primary and adventitious roots and leaves) and the generative organs (immature spikes; discussed further in the next paragraph). Because of the ITS1 sequence divergence between the two ancestral species, the rRNA precursors (pre-rRNA) that originated from the D- and S-genome 35S rDNA loci could be identified. The *Mlu*I enzyme only cuts the *B. distachyon*-like units producing two fragments (**Figures 3A,B**). Because there was no *Mlu*I restriction site in the *B. stacei*-like ITS1, its RT-PCR products were uncut (**Figures 3A,B**). In ABR113, only the D-genome *Mlu*I fragments were observed (the presence of two bands per each gel lane; **Figure 3C**), which indicated a strong uniparental silencing of the S-genome rRNA genes in all of the studied organs. However, in *B. hybridum* 3-7-2, both the D- and S-genome bands were present in the primary and adventitious roots, which implies a co-dominance of the ancestral rDNAs (**Figure 3C**). The lower intensity of the S-genome band in the sample from the primary root compared with that of the adventitious root was most probably correlated with the lower expression level of the *B. stacei*-inherited rDNA. The banding pattern in leaves and immature spikes of 3-7-2 indicated the expression dominance of the D-genome 35S rDNA (**Figure 3C**).

A tissue-specific expression pattern of the rDNA homeologs has been observed in many plant hybrids and allopolyploids (Volkov et al., 2007). For instance, in *Arabidopsis suecica,* the progressive silencing of the *A. thaliana*-originated rRNA genes occurred during the early postembryonic development in the tissues that had been derived from both the shoot and root apical meristems (Pontes et al., 2007). Although a fully established ND was observed in leaves, a trace expression of the *A. thaliana*-derived rRNA genes was detected in the root-tip cells of the mature *A. suecica* plants. Similarly, in several cultivars of the allotetraploid *Brassica napus*, the *B. oleracea*-inherited rDNA was stably repressed in the leaves (except for "Norin 9" cultivar, in which a co-dominance of ancestral rDNA was revealed in leaves and roots; Chen and Pikaard, 1997b; Sochorova et al., 2017). In the 2–3-day-old seedlings of *B. napus*, however, there was an expression of both ancestral rDNAs in the root-tip cells (Hasterok and Maluszynska, 2000b). Considering these observations and the current study, it can be concluded that the ND in natural allopolyploids is much more stable in the leaf tissue than in the roots. However, the establishment of ND may occur as early as 4–5 days after fertilization as was shown by silver staining in wheat-rye hybrids (Castilho et al., 1995). Thus, the mechanisms that determine the ancestral rDNA expression status at different ontogenetic stages may vary significantly between species.

To date, all of the studied *B. hybridum* genotypes have shown a strong, uniparental dominance toward the D-genome 35S rDNA in the roots and leaves as was revealed using the cytogenetic (Idziak and Hasterok, 2008; Borowska-Zuchowska et al., 2016) and molecular approaches (Borowska-Zuchowska et al., 2020). It is well known from many plant systems, however, that epigenetic mechanisms are behind the ND maintenance, making this phenomenon a reversible process (Vieira et al., 1990; Chen and Pikaard, 1997a; Lawrence et al., 2004; Earley et al., 2006; Preuss et al., 2008). Early

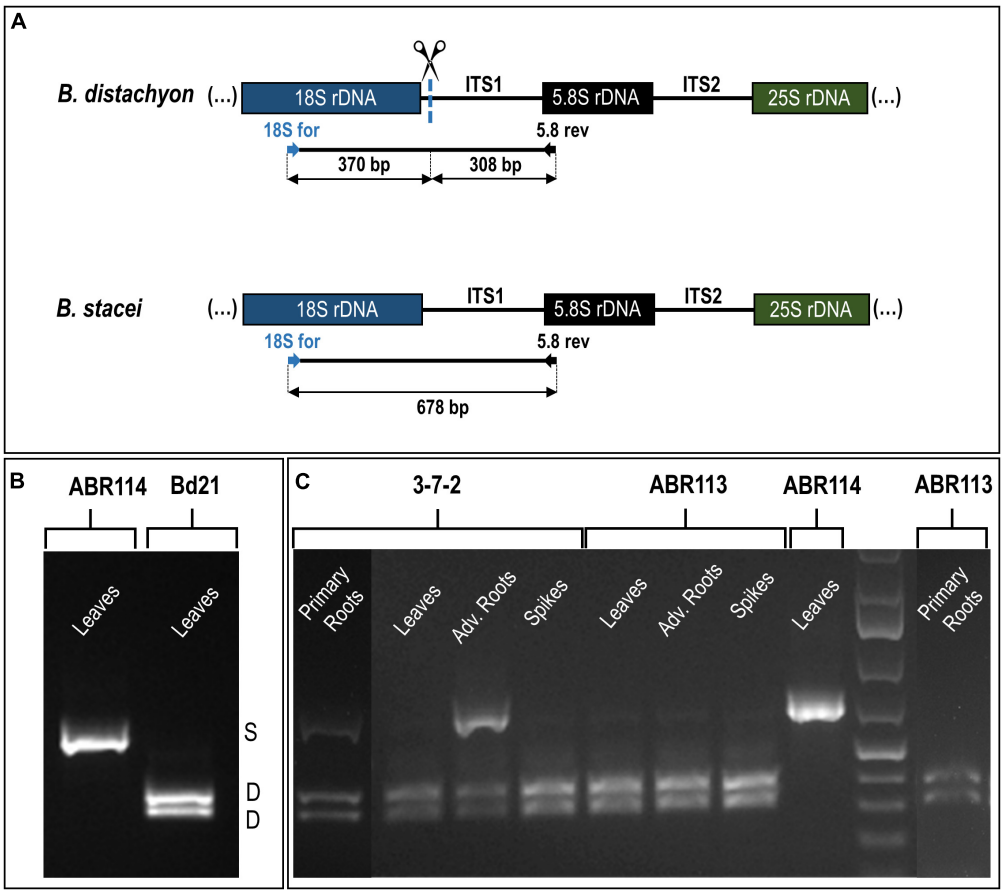


FIGURE 3 | The 35S rDNA expression analysis in the different tissues of *B. hybridum* (genotypes 3-7-2 and ABR113) and *B. stacei* (genotype ABR114; control) and *B. distachyon* (genotype Bd21; control) using the RT-PCR CAPS method. **(A)** The *Mlu*I restriction profile of the *B. distachyon*-like and *B. stacei*-like ITS1 PCR products. The expected sizes of the bands after *Mlu*I digestion are presented. **(B)** The *Mlu*I restriction profiles of ITS1 amplification products that were obtained from the leaves cDNAs of *B. distachyon* and *B. stacei*. **(C)** The *Mlu*I restriction profiles of ITS1 amplification products that were obtained from the primary roots, leaves, adventitious roots, and immature spikes cDNAs. Adv. Roots, adventitious roots.

studies using chemical agents that cause the hypomethylation of the genome (e.g., 5-azacytidine; 5-aza-2'-deoxycytidine) and/or histone deacetylation (trichostatin A) resulted in the transcriptional reactivation of the under-dominant rDNA loci (Vieira et al., 1990; Amado et al., 1997; Chen and Pikaard, 1997a; Lawrence et al., 2004). By contrast, the global hypomethylation of the *B. hybridum* ABR113 genome induced by 5-azacytidine was insufficient for the transcriptional reactivation of the S-genome loci (Borowska-Zuchowska and Hasterok, 2017). This observation suggested that the *B. stacei*-inherited rRNA genes in ABR113 may be irreversibly repressed, e.g., due to the accumulation of mutations. Thus, further investigations on the diverse *B. hybridum* genotypes are needed to find those that are characterized by a differential expression of the ancestral rDNA homeologs. In this study, a co-dominant expression of both ancestral rDNA loci in *B. hybridum* was revealed in the root tissue of the 3-7-2 genotype for the first time. Interestingly, such a pattern was not uniform across the whole root of this genotype. The cytological observation of highly condensed S-genome loci (**Figures 2A–C** and **Supplementary Figure 1**) and the absence of significant S-genome transcripts (**Figure 3**) indicate that the apical meristem of the primary root may not represent a tissue with impaired ND and that considerable rRNA silencing may occur in these cells. In contrast, loss of ND (codominant expression phenotype) seems to be highly pronounced in adventitious roots, which show secondary constrictions on both D- and S-type chromosomes (**Figures 2G–I** and **Supplementary Figure 2**) and strong expression of both homeologs (**Figure 3**). The differential expression of rDNA could be related to the different origins of primary and adventitious roots. In monocots, the primary root derives from an embryonic radicle and is often short-lived and replaced by adventitious roots, which are formed from any non-root tissue, usually stem, and are produced during normal development (Bellini et al., 2014). Previous analyses of Hasterok and Maluszynska (2000a) on *Allium cepa* showed that the 35S rRNA gene expression pattern can differ between the primary and adventitious roots. In their study, ND was manifested in adventitious and not primary roots. In contrast to the present study, in *A. cepa* more rDNA loci were transcriptionally active in the root-tip cells of the primary roots

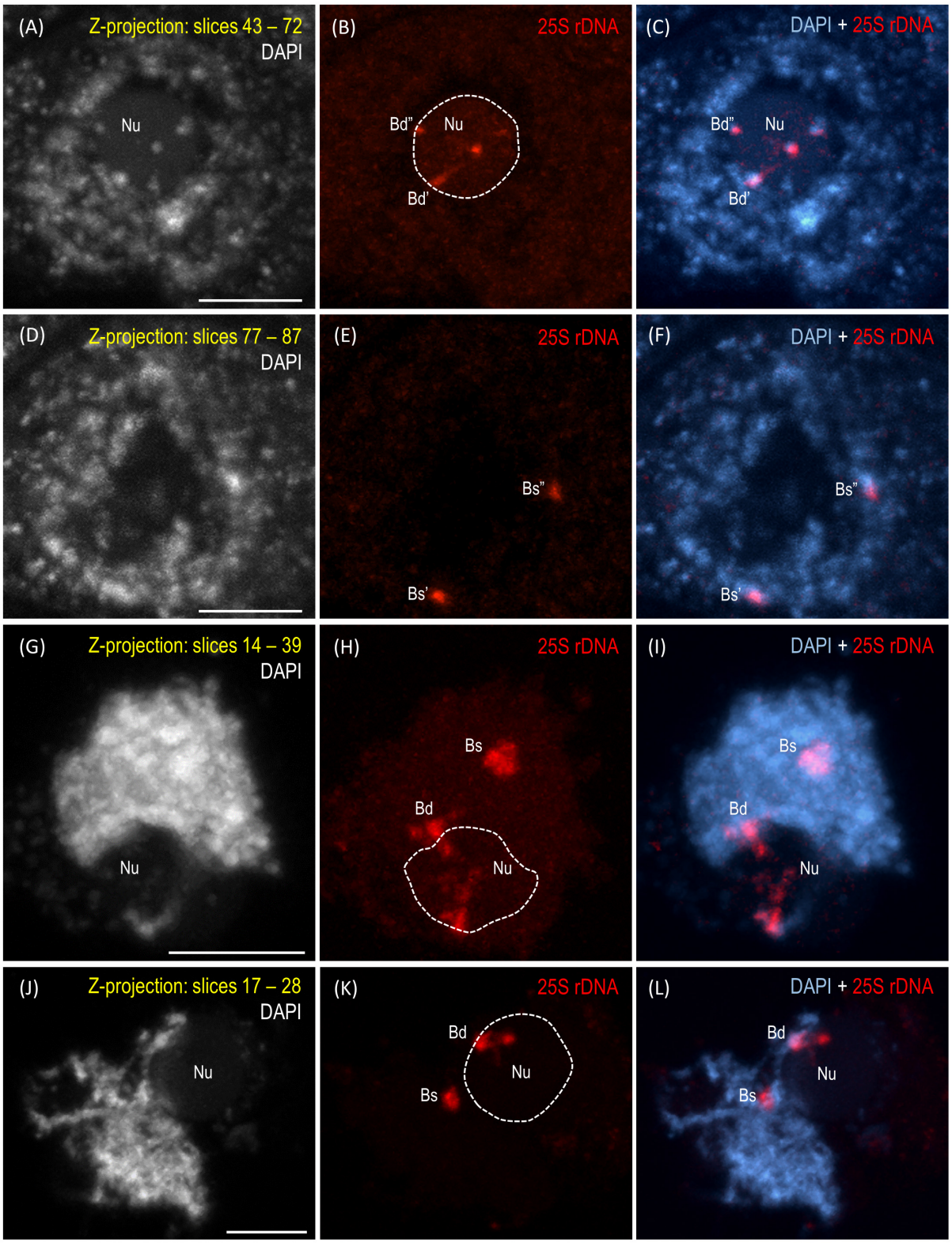


FIGURE 4 | Localization of the *B. distachyon*- and *B. stacei*-originated 35S rDNA loci in 3-D cytogenetic preparations of *B. hybridum* meiocytes (genotype 3-7-2) at leptotene, zygotene, and pachytene. Selected stacks that contain the 25S rDNA hybridization signals (red fluorescence) are presented. **(A–F)** Two different sections of one nucleus at leptotene. **(G–I)** Zygotene. **(J–L)** Zygotene/Pachytene. Bd, *B. distachyon*-inherited 35S rDNA loci; Bs, *B. stacei*-inherited 35S rDNA loci; Nu, nucleolus. The position of the nucleolus in the 25S rDNA channel **(B,E,H,K)** is denoted by a dashed line. Chromatin was counterstained with DAPI. Scale bar = 5 μm.

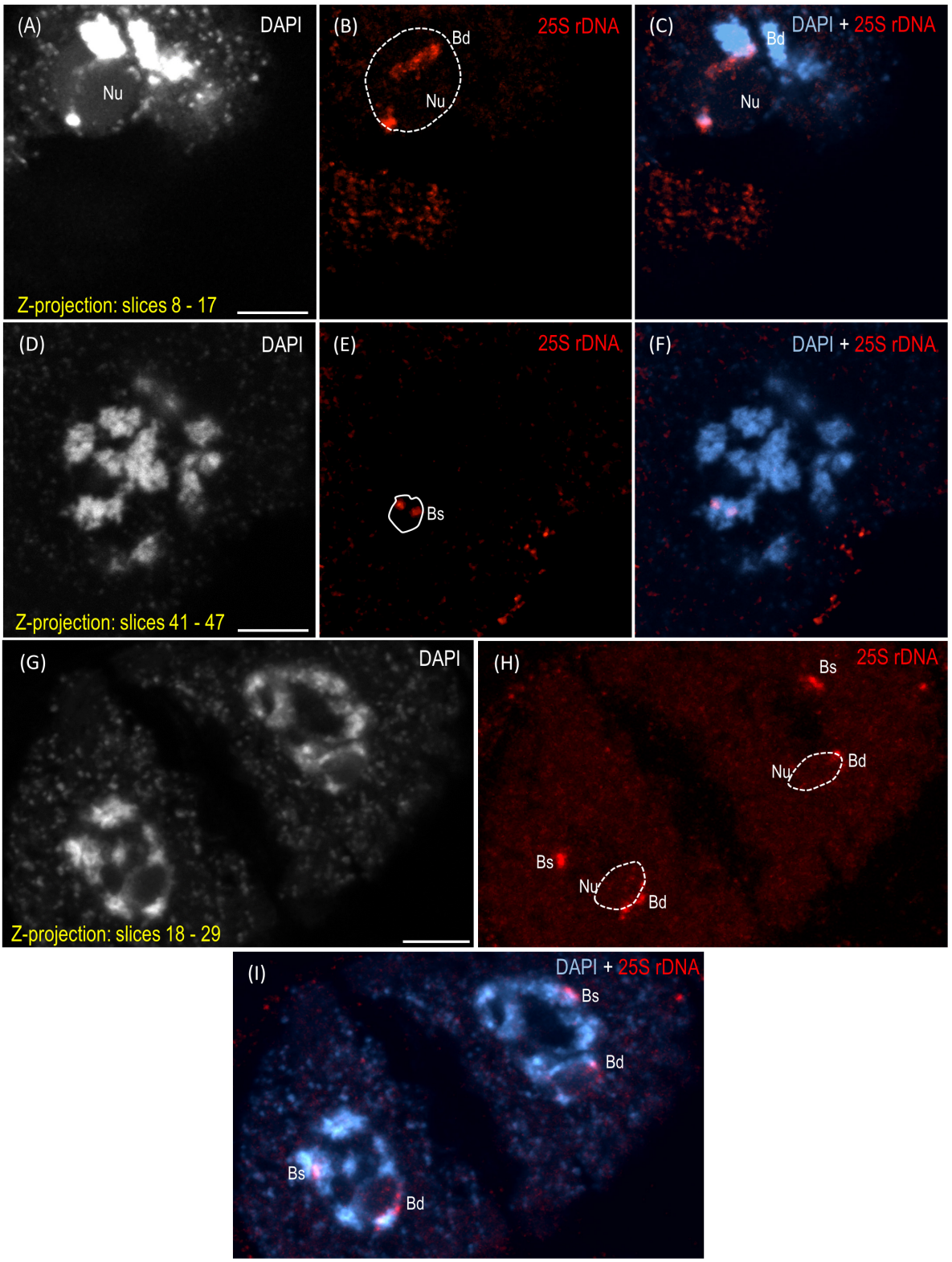


FIGURE 5 | Localization of the *B. distachyon*- and *B. stacei*-originated 35S rDNA loci in the 3-D cytogenetic preparations of *B. hybridum* meiocytes (genotype 3-7-2) at diakinesis and in dyads. Selected stacks that contain the 25S rDNA hybridization signals (red fluorescence) are presented. **(A–F)** Two different sections of one nucleus at diakinesis. **(G–I)** Dyad. Bd, *B. distachyon*-inherited 35S rDNA loci; Bs, *B. stacei*-inherited 35S rDNA loci; Nu, nucleolus. The position of the nucleolus in the 25S rDNA channel **(B,E,H)** is denoted by a dashed line. Chromatin was counterstained with DAPI. Scale bar = 5 μm.

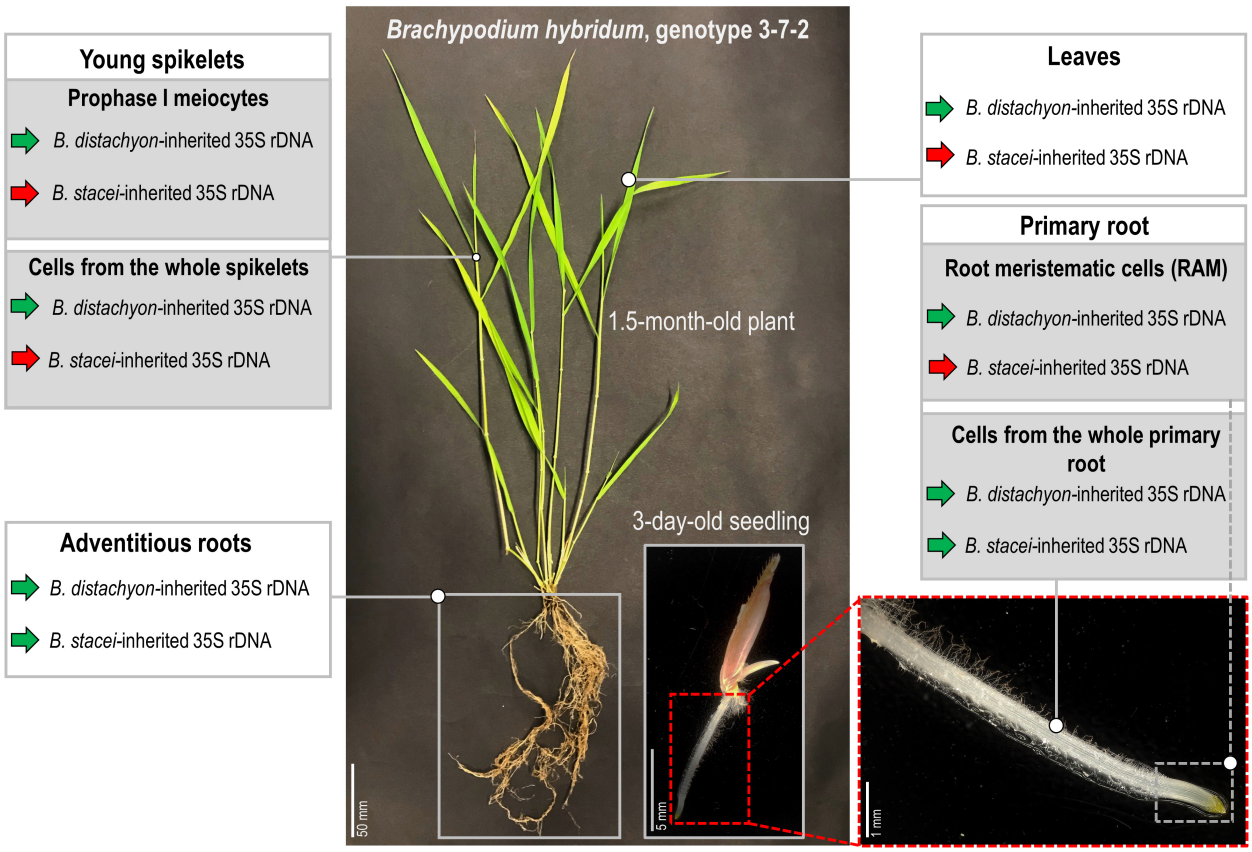


FIGURE 6 | Stability of the nucleolar dominance in different tissues of *B. hybridum* genotype 3-7-2. Green arrows denote the transcriptionally active 35S rDNA, while the red arrows indicate the silenced ones.

than in the adventitious ones, suggesting that the rRNA gene expression patterns might be species-specific.

Nucleolar Dominance Is Maintained in the Generative Organs of *Brachypodium hybridum*

The reprogramming of the rRNA gene expression may accompany the transition from the vegetative to the generative phase as was shown in *Brassica* (Chen and Pikaard, 1997b) and *Solanum* allopolyploids (Komarova et al., 2004). To verify whether such a transcriptional activation of the S-genome 35S rDNA loci occurs in the genotype 3-7-2 of *B. hybridum*, we determined (i) the localization of both ancestral homeologs in the meiocytes that had been isolated from the anthers and (ii) the origin of the pre-rRNA in the immature spikes. The FISH with 25S rDNA as the probe was performed on the *B. hybridum* 3-7-2 meiocytes at different substages of prophase I and in the dyads. Only one nucleolus per cell was observed in all of the studied stages of meiosis (**Figures 4**, **5** and **Supplementary Videos 1–5**). At the leptotene, two 35S rDNA loci of a D-genome origin were located within the nucleolus (**Figures 4A–C** and **Supplementary Video 1**), while the *B. stacei*-inherited loci were located at the nuclear periphery in the DAPI-positive domains (**Figures 4D–F** and **Supplementary Video 1**). Beginning with the zygotene to the latter substages of prophase I, only one 25S rDNA FISH signal per ancestral genome was observed after the formation of the bivalents. One bivalent with decondensed 35S rDNA loci that had been derived from the D-genome was associated with the nucleolus, while an S-genome bivalent with proximally located 35S rDNA loci was not attached to the nucleolus in the meiocytes at the stages from the zygotene to the diakinesis (**Figures 4G–L**, **Figures 5A–F** and **Supplementary Videos 2–4**). Moreover, only one 25S rDNA FISH signal was seen adjacent to/within the nucleolus in the dyads, which indicates that ND was maintained in this phase (**Figures 5G–I** and **Supplementary Video 5**). RT-CAPS analysis showed only the D-genome pre-rRNA precursors in the immature spikes of *B. hybridum* 3-7-2 (**Figure 3C**). Thus, ND was maintained not only in the meiocytes but in all of the floral organs. The exclusion of under-dominant S-genome 35S rDNA loci from the nucleolus contrasts with studies in *Arabidopsis thaliana*, whose all loci, irrespective of activity, seem to associate with the nucleolus in meiosis (Sims et al., 2021). Thus, it cannot be ruled out that the position of NORs is influenced by the nuclear topology, which may differ between diploid and allopolyploid species.

The studies of Silva et al. (1995) in the hexaploid triticale showed that the rRNA genes of a rye origin that were silent during the first meiotic division were transcriptionally activated in the microspores (Silva et al., 1995), thereby indicating that meiotic reprogramming may erase the preferential inactivation of rDNA via ND. Chen and Pikaard (1997b) used an S1 nuclease protection assay to determine the rRNA gene expression patterns in the different tissues of the allotetraploid *B. napus*. They observed the reactivation of the *B. oleracea*-inherited rDNA loci in all of the floral organs, including the sepals, petals, anthers, and siliques. Thus, the hypothesis that a derepression of the under-dominant rRNA genes occurs when both ancestral rDNA homeologs are segregated by meiosis as was shown previously in triticale (Silva et al., 1995) may not be universal for all plant allopolyploids (Reeder, 1985). The further studies of Sochorova et al. (2017) in *B. napus* showed only a trace expression of the *B. oleracea*-derived rDNA in the flower buds of two of the seven studied cultivars. These observations indicated that the ND regulation might even be genotype-specific. Such a specificity of the ND regulation among the floral organs in *B. hybridum* has not been observed yet. In addition to the current study on genotype 3-7-2, the ND establishment in the meiocytes was only analyzed in the reference genotype ABR113 (Borowska-Zuchowska et al., 2019). The S-genome rDNA homeologs could not form the nucleolus in either of the studied genotypes as they were transcriptionally silenced. The inactive state of the aforementioned loci in ABR113 was further confirmed using the silver staining method (Borowska-Zuchowska et al., 2019).

CONCLUSION AND PERSPECTIVES

To the best of our knowledge, this study revealed the first *B. hybridum* genotype with a co-dominance of the D- and S-genome rDNA homeologs in the primary and adventitious roots. **Figure 6** summarizes the rDNA expression patterns in the different tissues of *B. hybridum* 3-7-2 and shows the developmental regulation of ND in this species. Further comparative studies of the rDNA molecular structure in *B. hybridum* may shed more light on the specific mechanisms that shape ND in grasses. There is some evidence that both the chromosomal position and/or the presence of the control elements that are located within rDNA units may be responsible for the preferential expression of rDNA *via* ND (Chandrasekhara et al., 2016; Mohannath et al., 2016). Thus, using the new whole-genome sequencing strategies that permit a closer look at the complete rDNA units (McKinlay et al., 2021) may significantly improve our understanding of the rDNA evolution and behavior in allopolyploids.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

NB-Z conceived and designed the study. NB-Z, ER, SM, JW, AP, and AK performed the experiments. NB-Z, ER, AK, and RH analyzed the data. NB-Z and RH wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors gratefully acknowledge the financial support from the National Science Centre, Poland (grant 2018/31/B/NZ3/01761) and Czech Science Foundation (grants 19-03442S and 20-28029S). The publication was co-financed by funds granted under the Research Excellence Initiative of the University of Silesia in Katowice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.768347/full#supplementary-material

Supplementary Figure 1 | The rDNA-FISH analysis of primary root apical meristems of *B. hybridum* genotype 3-7-2. The distribution of the 35S rDNA loci in the mitotic metaphase chromosomes and interphase nucleus is shown in **A–F**, respectively. FISH mapping of 25S rDNA (red fluorescence) in the metaphase chromosome complement **(A–C)** and interphase nucleus **(D–F)**. Bd, *B. distachyon*-inherited 35S rDNA loci; Bs, *B. stacei*-inherited 35S rDNA loci. The secondary constrictions on **(A)** are indicated by the dashed lines. Chromatin was stained with DAPI. Scale bar = 5 μm. Note the absence of secondary constrictions on the Bs chromosomes.

Supplementary Figure 2 | The rDNA-FISH analysis of adventitious root apical meristems of *B. hybridum* genotype 3-7-2. The distribution of the 35S rDNA loci in the primary mitotic metaphase chromosomes and interphase nuclei is shown in **(A–F)**. FISH mapping of 25S rDNA (red fluorescence) in the metaphase chromosome complement **(A–F)** and interphase nucleus (**A–C**, left side of the photomicrograph). Bd, *B. distachyon*-inherited 35S rDNA loci; Bs, *B. stacei*-inherited 35S rDNA loci. The secondary constrictions on **(A,D)** are indicated by the dashed lines. Chromatin was stained with DAPI. Scale bar = 5 μm. Note the presence of secondary constrictions on both Bs and Bd loci.

Supplementary Video 1 | Three-dimensional distribution of the 25S rDNA hybridization signals (green) in *B. hybridum* 3-7-2 meiocyte at leptotene that is presented in **Figures 4A–F**. The cell shown in the figure is identified by a white rectangle. Chromatin is pseudocolored in red.

Supplementary Video 2 | Three-dimensional distribution of 25S rDNA hybridization signals (green) in *B. hybridum* 3-7-2 meiocyte at zygotene that is presented in **Figures 4G–I**. Chromatin is pseudocolored in red.

Supplementary Video 3 | Three-dimensional distribution of 25S rDNA hybridization signals (green) in *B. hybridum* 3-7-2 meiocyte at zygotene/pachytene that is presented in **Figures 4J–L**. Chromatin is pseudocolored in red.

Supplementary Video 4 | Three-dimensional distribution of 25S rDNA hybridization signals (green) in *B. hybridum* 3-7-2 meiocyte at diakinesis that is presented in **Figures 5A–F**. Chromatin is pseudocolored in red.

Supplementary Video 5 | Three-dimensional distribution of 25S rDNA hybridization signals (green) in *B. hybridum* 3-7-2 dyad that is presented in **Figures 5G–I**. Chromatin is pseudocolored in red.

REFERENCES

Amado, L., Abranches, R., Neves, N., and Viegas, W. (1997). Development-dependent inheritance of 5-azacytidine-induced epimutations in triticale: analysis of rDNA expression patterns. *Chromosome Res.* 5, 445–450. doi: 10.1023/a:1018460828720

Bass, H. W., Marshall, W. F., Sedat, J. W., Agard, D. A., and Cande, W. Z. (1997). Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. *J. Cell Biol.* 137, 5–18. doi: 10.1083/jcb.137.1.5

Bellini, C., Pacurar, D. I., and Perrone, I. (2014). Adventitious roots and lateral roots: similarities and differences. *Annu. Rev. Plant Biol.* 65, 639–666. doi: 10.1146/annurev-arplant-050213-035645

Borowska-Zuchowska, N., and Hasterok, R. (2017). Epigenetics of the preferential silencing of *Brachypodium stacei*-originated 35S rDNA loci in the allotetraploid grass *Brachypodium hybridum*. *Sci. Rep.* 7:5260. doi: 10.1038/s41598-017-05413-x

Borowska-Zuchowska, N., Kovarik, A., Robaszkiewicz, E., Tuna, M., Tuna, G. S., Gordon, S., et al. (2020). The fate of 35S rRNA genes in the allotetraploid grass *Brachypodium hybridum*. *Plant J.* 103, 1810–1825. doi: 10.1111/tpj.14869

Borowska-Zuchowska, N., Kwasniewski, M., and Hasterok, R. (2016). Cytomolecular analysis of ribosomal DNA evolution in a natural allotetraploid *Brachypodium hybridum* and its putative ancestors - dissecting complex repetitive structure of intergenic spacers. *Front. Plant Sci.* 7:1499. doi: 10.3389/fpls.2016.01499

Borowska-Zuchowska, N., Robaszkiewicz, E., Wolny, E., Betekhtin, A., and Hasterok, R. (2019). Ribosomal DNA loci derived from *Brachypodium stacei* are switched off for major parts of the life cycle of *Brachypodium hybridum*. *J. Exp. Bot.* 70, 805–815. doi: 10.1093/jxb/ery425

Castilho, A., Queiroz, A., Neves, N., Barao, A., Silva, M., and Viegas, W. (1995). The developmental stage of inactivation of rye origin rRNA genes in the embryo and endosperm of wheat x rye F1 hybrids. *Chromosome Res.* 3, 169–174. doi: 10.1007/BF00710710

Catalan, P., Muller, J., Hasterok, R., Jenkins, G., Mur, L. A., Langdon, T., et al. (2012). Evolution and taxonomic split of the model grass Brachypodium distachyon. *Ann. Bot.* 109, 385–405. doi: 10.1093/aob/mcr294

Cermeno, M. C., Orellana, J., Santos, J. L., and Lacadena, J. R. (1984). Nucleolar organizer activity in wheat, rye and derivatives analyzed by a silver-staining procedure. *Chromosoma* 89, 370–376. doi: 10.1007/Bf00331254

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Chen, Z. J., and Pikaard, C. S. (1997a). Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Gene Dev.* 11, 2124–2136.

Chen, Z. J., and Pikaard, C. S. (1997b). Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proc. Natl. Acad. Sci. U S A* 94, 3442–3447.

Chester, M., Gallagher, J. P., Symonds, V. V., da Silva, A. V. C., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. U S A* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Costa-Nunes, P., Pontes, O., Preuss, S. B., and Pikaard, C. S. (2010). Extra views on RNA-dependent DNA methylation and MBD6-dependent heterochromatin formation in nucleolar dominance. *Nucleus* 1, 254–259. doi: 10.4161/nucl.1.3.11741

Diaz-Perez, A., Lopez-Alvarez, D., Sancho, R., and Catalan, P. (2018). Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus Brachypodium using minimum evolution, coalescence and maximum likelihood approaches. *Mol. Phylogenet. Evol.* 127, 256–271. doi: 10.1016/j.ympev.2018.06.003

Dolezel, J., Cizkova, J., Simkova, H., and Bartos, J. (2018). One major challenge of sequencing large plant genomes is to know how big they really are. *Internat. Mole. Sci.* 19:11. doi: 10.3390/ijms19113554

Earley, K., Lawrence, R. J., Pontes, O., Reuther, R., Enciso, A. J., Silva, M., et al. (2006). Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Gene Dev.* 20, 1283–1293. doi: 10.1101/gad.1417706

Ge, X. H., Ding, L., and Li, Z. Y. (2013). Nucleolar dominance and different genome behaviors in hybrids and allopolyploids. *Plant Cell Rep.* 32, 1661–1673. doi: 10.1007/s00299-013-1475-5

Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., et al. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of Brachypodium hybridum and its diploid progenitors. *Nat. Comm.* 11:1. doi: 10.1038/s41467-020-17302-5

Guo, X., and Han, F. P. (2014). Asymmetric Epigenetic Modification and Elimination of rDNA Sequences by Polyploidization in Wheat. *Plant Cell* 26, 4311–4327. doi: 10.1105/tpc.114.129841

Hasterok, R., Draper, J., and Jenkins, G. (2004). Laying the cytotaxonomic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv. *Chromosome Res.* 12, 397–403. doi: 10.1023/B:CHRO.0000034130.35983.99

Hasterok, R., Dulawa, J., Jenkins, G., Leggett, M., and Langdon, T. (2006). Multi-substrate chromosome preparations for high throughput comparative FISH. *BMC Biotechnol.* 6:20. doi: 10.1186/1472-6750-6-20

Hasterok, R., Langdon, T., Taylor, S., and Jenkins, G. (2002). Combinatorial labelling of DNA probes enables multicolour fluorescence in situ hybridisation in plants. *Folia Histochem. Cytobiol.* 40, 319–323.

Hasterok, R., and Maluszynska, J. (2000a). Different rRNA gene expression in primary and adventitious roots of *Allium cepa* L. *Folia Histochem. Cytobiol.* 38, 181–184.

Hasterok, R., and Maluszynska, J. (2000b). Nucleolar dominance does not occur in root tip cells of allotetraploid *Brassica* species. *Genome* 43, 574–579.

Houchins, K., O'Dell, M., Flavell, R. B., and Gustafson, J. P. (1997). Cytosine methylation and nucleolar dominance in cereal hybrids. *Mol. Gen. Genet.* 255, 294–301.

IBI (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi: 10.1038/nature08747

Idziak, D., Betekhtin, A., Wolny, E., Lesniewska, K., Wright, J., Febrer, M., et al. (2011). Painting the chromosomes of *Brachypodium*-current status and future prospects. *Chromosoma* 120, 469–479. doi: 10.1007/s00412-011-0326-9

Idziak, D., and Hasterok, R. (2008). Cytogenetic evidence of nucleolar dominance in allotetraploid species of *Brachypodium*. *Genome* 51:5. doi: 10.1139/G08-017

Jenkins, G., and Hasterok, R. (2007). BAC 'landing' on chromosomes of *Brachypodium distachyon* for comparative genome alignment. *Nat. Protoc.* 2, 88–98. doi: 10.1038/nprot.2006.490

Komarova, N. Y., Grabe, T., Huigen, D. J., Hemleben, V., and Volkov, R. A. (2004). Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. *Plant Mole. Biol.* 56, 439–463. doi: 10.1007/s11103-004-4678-x

Kovarik, A., Pires, J. C., Leitch, A. R., Lim, K. Y., Sherwood, A. M., Matyasek, R., et al. (2005). Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* 169, 931–944. doi: 10.1534/genetics.104.032839

Lacadena, J. R., Cermeno, M. C., Orellana, J., and Santos, J. L. (1984). Evidence for wheat-rye nucleolar competition (amphiplasty) in triticale by silver-staining procedure. *Theor. Appl. Genet.* 67, 207–213. doi: 10.1007/BF00317037

Lawrence, R. J., Earley, K., Pontes, O., Silva, M., Chen, Z. J., Neves, N., et al. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell* 13, 599–609.

Lusinska, J., Majka, J., Betekhtin, A., Susek, K., Wolny, E., and Hasterok, R. (2018). Chromosome identification and reconstruction of evolutionary rearrangements in Brachypodium distachyon. B. stacei and B. hybridum. *Ann. Bot.* 122, 445–459. doi: 10.1093/aob/mcy086

Mason, A. S., and Wendel, J. F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* 11:1014. doi: 10.3389/fgene.2020.01014

Matyasek, R., Dobesova, E., Huska, D., Jezkova, I., Soltis, P. S., Soltis, D. E., et al. (2016). Interpopulation hybridization generates meiotically stable rDNA epigenetic variants in allotetraploid *Tragopogon mirus*. *Plant J.* 85, 362–377. doi: 10.1111/tpj.13110

McKinlay, A., Fultz, D., Wang, F., and Pikaard, C. S. (2021). Targeted Enrichment of rRNA Gene Tandem Arrays for Ultra-Long Sequencing by Selective

Restriction Endonuclease Digestion. *Front. Plant Sci.* 12:656049. doi: 10.3389/fpls.2021.656049

Mohannath, G., Pontvianne, F., and Pikaard, C. S. (2016). Selective nucleolus organizer inactivation in Arabidopsis is a chromosome position-effect phenomenon. *Proc. Natl. Acad. Sci. U S A* 113, 13426–13431. doi: 10.1073/pnas.1608140113

Navashin, M. (1934). Chromosomal alterations caused by hybridization and their bearing upon certain general genetic problems. *Cytologia* 5, 169–203.

Neves, N., Heslop-Harrison, J. S., and Viegas, W. (1995). rRNA gene activity and control of expression mediated by methylation and imprinting during embryo development in wheat x rye hybrids. *Theor. Appl. Genet.* 91, 529–533. doi: 10.1007/BF00222984

Nicolas, S. D., Leflon, M., Monod, H., Eber, F., Coriton, O., Huteau, V., et al. (2009). Genetic Regulation of Meiotic Cross-Overs between Related Genomes in Brassica napus Haploids and Hybrids. *Plant Cell* 21, 373–385. doi: 10.1105/tpc.108.062273

Ozkan, H., Levy, A. A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13, 1735–1747. doi: 10.1105/Tpc.010082

Pikaard, C. S. (2000). Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids. *Plant Mol. Biol.* 43, 163–177.

Pontes, O., Lawrence, R. J., Silva, M., Preuss, S., Costa-Nunes, P., Earley, K., et al. (2007). Postembryonic establishment of megabase-scale gene silencing in nucleolar dominance. *PLoS One* 2:e1157. doi: 10.1371/journal.pone.0001157

Preuss, S. B., Costa-Nunes, P., Tucker, S., Pontes, O., Lawrence, R. J., Mosher, R., et al. (2008). Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. *Mol. Cell* 32, 673–684. doi: 10.1016/j.molcel.2008.11.009

Reeder, R. H. (1985). Mechanisms of nucleolar dominance in animals and plants. *J. Cell Biol.* 101, 2013–2016. doi: 10.1083/jcb.101.5.2013

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13, 1749–1759. doi: 10.1105/tpc.010083

Shaw, P. (2013). The Plant Nucleolus. *Plant Gen. Div. Vol.* 2, 65–76.

Silva, M., Queiroz, A., Neves, N., Barao, A., Castilho, A., Morais-Cecilio, L., et al. (1995). Reprogramming of rye rDNA in triticale during microsporogenesis. *Chromosome Res.* 3, 492–496. doi: 10.1007/BF00713964

Sims, J., Rabanal, F. A., Elgert, C., von Haeseler, A., and Schlögelhofer, P. (2021). It is just a matter of time: balancing homologous recombination and non-homologous end joining at the rDNA locus during meiosis. *Front. Plant Sci.* 12:773052. doi: 10.3389/fpls.2021.773052

Sochorova, J., Coriton, O., Kuderova, A., Lunerova, J., Chevre, A. M., and Kovarik, A. (2017). Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of *Brassica napus*. *Ann. Bot.* 119, 13–26. doi: 10.1093/aob/mcw187

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003

Symonova, R. (2019). Integrative rDNAomics-Importance of the oldest repetitive fraction of the eukaryote genome. *Genes* 10:5. doi: 10.3390/genes10050345

Unfried, I., and Gruendler, P. (1990). Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Res.* 18:4011. doi: 10.1093/nar/18.13.4011

Vieira, R., Mellosampayo, T., and Viegas, W. S. (1990). 1R chromosome nucleolus organizer region activation by 5-azacytidine in wheat x rye hybrids. *Genome* 33, 707–712.

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: Inheritance, rearrangement, expression. *Systemat. Biodiv.* 5, 261–276. doi: 10.1017/S1477200007002447

Volkov, R. A., Medina, F. R., Zentgraf, U., and Hemleben, V. (2004). Molecular Cell Biology: Organization and molecular evolution of rDNA, nucleolar dominance, and nucleolus structure. *Prog. Bot.* 65, 106–146. doi: 10.1007/978-3-642-18819-0

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership