



OPEN ACCESS

EDITED BY

Hans De Jong,
Wageningen University and Research,
Netherlands

REVIEWED BY

Takayoshi Ishii,
Tottori University, Japan
Paul Fransch,
Wageningen University and Research,
Netherlands

*CORRESPONDENCE

Terezie Mandáková
✉ terezie.mandakova@ceitec.muni.cz

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

RECEIVED 19 October 2023

ACCEPTED 12 March 2024

PUBLISHED 19 April 2024

CITATION

Hlavatá K, Závěská E, Leong-Škorničková J,
Pouch M, Poulsen AD, Šída O, Khadka B,
Mandáková T and Fér T (2024) Ancient
hybridization and repetitive element
proliferation in the evolutionary history of the
monocot genus *Amomum* (Zingiberaceae).
Front. Plant Sci. 15:1324358.
doi: 10.3389/fpls.2024.1324358

COPYRIGHT

© 2024 Hlavatá, Závěská, Leong-Škorničková,
Pouch, Poulsen, Šída, Khadka, Mandáková and
Fér. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Ancient hybridization and repetitive element proliferation in the evolutionary history of the monocot genus *Amomum* (Zingiberaceae)

Kristýna Hlavatá^{1†}, Eliška Závěská^{1,2†}, Jana Leong-Škorničková^{3,4}, Milan Pouch^{5,6}, Axel Dalberg Poulsen⁷, Otakar Šída⁸, Bijay Khadka¹, Terezie Mandáková^{5,9*} and Tomáš Fér¹

¹Department of Botany, Faculty of Science, Charles University, Prague, Czechia, ²Institute of Botany, Czech Academy of Science, Průhonice, Czechia, ³Herbarium, Singapore Botanic Gardens, National Parks Board, Singapore, Singapore, ⁴Department of Biological Sciences, National University of Singapore, Singapore, Singapore, ⁵Central European Institute of Technology, Masaryk University, Brno, Czechia, ⁶National Center for Biomolecular Research (NCBR), Masaryk University, Kamenice, Czechia, ⁷Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, ⁸Department of Botany, National Museum in Prague, Prague, Czechia, ⁹Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia

Genome size variation is a crucial aspect of plant evolution, influenced by a complex interplay of factors. Repetitive elements, which are fundamental components of genomic architecture, often play a role in genome expansion by selectively amplifying specific repeat motifs. This study focuses on *Amomum*, a genus in the ginger family (Zingiberaceae), known for its 4.4-fold variation in genome size. Using a robust methodology involving PhyloNet reconstruction, RepeatExplorer clustering, and repeat similarity-based phylogenetic network construction, we investigated the repeatome composition, analyzed repeat dynamics, and identified potential hybridization events within the genus. Our analysis confirmed the presence of four major infrageneric clades (A–D) within *Amomum*, with clades A–C exclusively comprising diploid species ($2n = 48$) and clade D encompassing both diploid and tetraploid species ($2n = 48$ and 96). We observed an increase in the repeat content within the genus, ranging from 84% to 89%, compared to outgroup species with 75% of the repeatome. The SIRE lineage of the *Ty1-Copia* repeat superfamily was prevalent in most analyzed ingroup genomes. We identified significant difference in repeatome structure between the basal *Amomum* clades (A, B, C) and the most diverged clade D. Our investigation revealed evidence of ancient hybridization events within *Amomum*, coinciding with a substantial proliferation of multiple repeat groups. This finding supports the hypothesis that ancient hybridization is a driving force in the genomic evolution of *Amomum*. Furthermore, we contextualize our findings within the broader context of genome size variations and repeatome dynamics

observed across major monocot lineages. This study enhances our understanding of evolutionary processes within monocots by highlighting the crucial roles of repetitive elements in shaping genome size and suggesting the mechanisms that drive these changes.

KEYWORDS

genome evolution, genome size, interspecific hybridization, repetitive DNA, repeatome, phylogeny, 5S rDNA, Zingiberaceae

1 Introduction

Genome size, also known as C-value or haploid nuclear DNA content (hereafter referred to as GS), is a fundamental parameter in the study of organismal evolution. In land plants, GS exhibits remarkable variation, spanning up to 2,400-fold (Pellicer et al., 2018). Both genome expansion and contraction have been recognized as major driving forces of diversification in land plants (Cheng et al., 2014; Meudt et al., 2015; Simonin and Roddy, 2018). Genome expansion, often linked to whole genome duplication events, has been a historical precursor to speciation and the emergence of novel morphological features in various plant lineages (Qiao et al., 2022). Another mechanism that is profoundly shaping GS is amplification of repetitive sequences, in which transposable elements play a pivotal role (Pulido and Casacuberta, 2023).

Repetitive elements, often referred to as “tuning knobs of evolution” (King et al., 1997; hereafter referred to as repeats), are integral components of plant genomes. They can constitute as little as 3% in *Utricularia gibba* or as much as 91% of the entire genome in *Allium sativum* (Sun et al., 2020). They play the key roles in gene expression regulation (Garrido-Ramos, 2012; Bennetzen and Wang, 2014) and can evolve into new genes due to their rapid evolutionary rates (Mehrotra and Goyal, 2014). From the evolutionary perspective, the proliferation of repeats has been associated with diversification of new phylogenetic groups (Gaiero et al., 2019; Hloušková et al., 2019) and facilitates adaptation to changing environments (Jansz, 2019; Kumar and Mohapatra, 2021). Given the linear relationship between repeat content and GS within specific ploidy level (Lee and Kim, 2014), it is plausible to hypothesize that factors influencing repeat amplification align with those governing changes in GS.

Interspecific hybridization, a widespread phenomenon in angiosperms (Mallet, 2005), may play an important role in GS increase triggered by so-called genomic shock after subgenome merger (O'Neill et al., 1998; Ungerer et al., 2006; Wei et al., 2021). However, hybridization events may also lead to repeat deactivation and genome downsizing (Renny-Byfield et al., 2013; Heyduk et al., 2021), through processes such as illegitimate recombination or unequal homologous recombination (Bennetzen et al., 2005; Staton et al., 2012). Recent advancements in phylogenetics and

phylogenomics, such as target enrichment techniques (Cao et al., 2019), now enable the robust identification of hybrid species and lineages and explore the role of hybridization in GS changes.

The tropical genus *Amomum*, comprising approx. 64 species (de Boer et al., 2018), represents a distinctive case study group for exploring processes associated with GS amplification. *Amomum* exhibits the most significant GS variation within the entire family Zingiberaceae, ranging from 1,731 to 7,656 Mb, representing a 4.4-fold difference (Záveská et al., 2024). In *Amomum*, two tetraploid species are known ($2n = 96$), with GS values of 6,254 and 7,656 Mb (Hlavatá et al., 2023). The diploid species ($2n = 48$) display substantial GS variation, ranging from 1,731 to 4,699 Mb, representing a 2.7-fold difference. A well-supported phylogeny, constructed using Hyb-Seq based on 449 nuclear genes, identified four main clades (A, B, C, and D; Hlavatá et al., 2023) and increasing trend of GS along the phylogeny. Evidence of cyto-nuclear discordance, possibly indicating hybridization, was detected and warrants further investigation (Hlavatá et al., 2023). In this context, we hypothesize that the enlargement of GS in diploid *Amomum* species is a result of an expansion of repeats triggered by interspecific hybridization. Particularly, we aim to answer the following questions: i) what is the repeatome composition in the genus *Amomum*?, ii) does interspecific hybridization play a role in the evolution of *Amomum* and its repeatome?; and iii) is the evolution of repeats correlated with phylogenetic relationships in *Amomum*? To answer these questions, we use a wide range of analyses starting with a revision of GS variation within the genus based on 52 *Amomum* accessions (33 species), continuing with a phylogenetic network reconstruction of 30 *Amomum* species, complemented with a qualitative and quantitative analysis of repeats in a subset of 11 *Amomum* species.

2 Methods

2.1 Plant material

A total of 52 accessions, corresponding to 30 distinct *Amomum* species and encompassing the documented morphological, phylogenetic, and cytological spectrum of the genus (Hlavatá et al., 2023), were employed in the present study to analyze

genome size (GS) data. Reticulate relationships were reconstructed for these 30 species, while a subset of 11 accessions (plus two outgroup species included for comparative purposes) was further designated for an in-depth examination of repeat content. The selection of these subsets was devised to ensure that they represented the following aspects i) the primary phylogenetic clades within the genus *Amomum*, ii) variability in GS within and among these clades and iii) variation in ploidy levels observed across the genus. A comprehensive listing of all samples and their characteristics is provided in [Supplementary Table 1](#).

2.2 Genome size estimation and chromosome counts

Nuclear GS (referred to as nuclear DNA 1C values in Mb) data were sourced from our previous study (Hlavatá et al., 2023) for a total of 52 *Amomum* accessions plus 2 outgroup species. Chromosome numbers, and ploidy levels were also sourced from Hlavatá et al. (2023) for 12 *Amomum* accessions representing 12 species ([Supplementary Table 1](#)).

2.3 Sequencing data from target enrichment (Hyb-Seq) for species networks reconstruction

Raw data derived from Hyb-Seq encompassing 30 *Amomum* accessions were obtained from Hlavatá et al. (2023) and were processed similarly as in the previous study using HybPhyloMaker 1.6.4 (Fér and Schmickl, 2018) up to the reconstruction of gene trees based on total of 448 loci employing RAxML 8.2.4 (Stamatakis, 2014) with 1,000 standard bootstrap replicates and per exon partitioning. In cases where gene trees contained uncertain nodes with bootstrap support below 50, branches were collapsed. These gene trees were then employed in the reconstruction of species networks using a maximum pseudo-likelihood (MPL) framework function ‘InferNetwork_MPL’ (Yu and Nakhleh, 2015) and implemented in PhyloNet 3.6.1 (Than et al., 2008). Since the comprehensive exploration of dataset comprising 30 accessions with a larger number of reticulations (>2) utilizing a MPL approach is limited by prohibitive runtime costs (Than et al., 2008; Skopalíková et al., 2023) we adopted a sequential, stepwise approach for the analysis of our dataset. Initially, we constructed a species network for the entire dataset of 30 accessions, hereafter referred to as ‘complete dataset’, allowing for a maximum of two reticulations. Subsequently, we conducted a separate analysis on a subset comprising 17 accessions, which represented 16 species belonging to clade D, hereafter termed the ‘clade D dataset’, again allowing for a maximum of two reticulations. Prior to the Phylonet analyses, the gene trees were rooted using Newick Utilities 1.6 (Junier and Zdobnov, 2010). For the complete dataset, *A. subulatum* and *A. petaloideum* served as rooting taxa, while *A. aff. curtisii*, *A. latiflorum* and *A. corrugatum* were employed for rooting the clade D dataset. Each analysis involved ten runs with default settings, resulting in the generation

of five optimal networks per analysis. The selection of the best-fitting network was accomplished by applying the Akaike information criterion (AIC, $AIC = 2 \cdot k - 2 \cdot L$). Here, ‘k’ represented the number of parameters, which included the number of branches and the number of reticulations, while ‘L’ denoted the likelihood value (Keuler et al., 2020). To present the findings effectively, we combined the best-fitting models from both datasets into a unified phylogenetic network.

2.4 Low-coverage sequencing

Genome skimming was conducted on a subset of 13 accessions, consisting of 11 *Amomum* species and two closely related outgroup species (*Aframomum melegueta* and *Renealmia polypus*, [Supplementary Table 1](#)). DNA sequencing libraries were prepared as described in [Supplementary Methods M1](#) and subjected to sequencing on an Illumina NextSeq, utilizing a 300-cycle sequencing kit to generate 150 bp paired-end reads. The sequencing process was conducted at the Central European Institute of Technology (CEITEC), Masaryk University in Brno, Czech Republic. Raw reads resulting from this process were subsequently uploaded to the Sequence Read Archive (SRA) under the BioProject designation (ID PRJNA1029323).

2.5 Repeatome analysis

Read clustering and subsequent automated quantification of repetitive elements (repeats) were performed on the Galaxy platform (Afgan et al., 2018; <https://repeatexplorer-elixir.cerit-sc.cz/>) following the established protocol as described by Novák et al. (2020). Classification of repeats was carried out in accordance with the automatic procedure of RepeatExplorer (REXdb; Neumann et al., 2019) and was subject to manual verification. For repeat identification, BLAST (Altschul et al., 1990) was employed to search against a comprehensive repeat library compiled from various publicly accessible sources, including msRep (Liao et al., 2022; <https://msrepdb.cbrc.kaust.edu.sa/>), PlantRep (Luo et al., 2022; <http://www.plantrep.cn/>), RepeatMasker (Smit et al., 2013; <https://www.repeatmasker.org/>), and Musaceae-specific repeat database (Novák et al., 2014; <https://olomouc.ueb.cas.cz/en/content/dna-repeats/>). The manually reviewed file for each accession was subsequently utilized for the quantification of repeats, considering the known GS of the respective accession. This process was carried out using the script available at <https://github.com/tomas-fer/scripts/blob/master/makeREsummary.sh>. Barplots representing the main repeat groups were constructed using Microsoft Excel (Microsoft Corporation, 2018).

An additional analysis of tandem repeats was conducted using Tandem Repeat Analyzer (TAREAN; Novák et al., 2017) for all accessions. This analysis aimed to identify potential satellite sequences that may not have been detected by the RepeatExplorer analysis and to provide insights into the presence and organization of 5S rDNA clusters. In TAREAN, the cluster size threshold was established at 0.01, and the processing queue was configured for “extra long” run times to accommodate the analysis of the

maximum feasible number of reads. In the context of 5S rDNA, diploid specimens typically exhibit as single-looped circular graphs, whereas accessions of hybrid and/or polyploid origin may present more complex multi-looped graphs, as detailed by Garcia et al. (2020). Hence, the examination of 5S rDNA can serve as an indicator of hybridization (Garcia et al., 2020).

2.6 Comparative analysis of repeats

A comparative analysis, involving the simultaneous clustering of reads from all accessions, was performed in RepeatExplorer. This analysis adhered to the established protocol of Novák et al. (2020) and employed default settings. A random subsample of 1,000,000 reads was selected from each accession for this analysis. From this analysis, the distribution of the 225 most prevalent comparative repeat clusters, hereafter referred to as “sub-lineages”, was graphically represented, excluding clusters originating from plastid-derived sequences. To integrate these repeat sub-lineage abundances onto the phylogeny, the “contMap” function from the “phytools” package within the R environment (Revell, 2012; R. 4.2.1, R Core Team, 2022) was employed.

2.7 Phylogenetic signal and correlation of repeat proportions

In the representation of phylogenetic relationships among the 13 accessions studied for repeats, the ASTRAL species tree, as constructed based on HybSeq data and presented in Hlavatá et al. (2023), served as the foundation. This tree was appropriately tailored by pruning using the “drop.tip” function in the “ape” package within the R environment to exclusively encompass the specific subset of sampled accessions. To assess the phylogenetic signal, represented as Pagel’s λ (Pagel, 1997, 1999), associated with the proportions of repeats (for lineages, superfamilies and groups), the “phylosig” functions within “phytools” package in R was employed. The degree of simple correlation (adjusted R-squared) between the quantity of repeats and 1C GS (both considering and not considering the phylogenetic context) was computed. This was achieved using the “geiger” (Pennell et al., 2014) and “caper” (Orme et al., 2018) packages in the R environment. The specific functions utilized for this purpose included “comparative.data”, “model.pgls”, and “anova”. These correlations were calculated for both the overarching repeat groups and individual repeat lineages.

2.8 Similarity based consensus network

The matrices, originally indicating the observed/expected number of edges between species as derived from RepeatExplorer clustering analysis, were transformed into distance matrices as described by Vitales et al. (2020). In this process, matrices that represented clusters without any edges connecting species were entirely excluded from the analysis. Additionally, both outgroup species were entirely omitted from consideration since they exhibited very few connections

with the ingroup. For the construction of neighbor-joining trees, we employed the “ape” package (Paradis and Schliep, 2019) within the R environment. Furthermore, a consensus network was established using the SplitsTree (Huson and Bryant, 2006), based on the method by Holland and Moulton (2003). Only splits that garnered support in a minimum of 10% of the trees were taken into account for subsequent analysis.

2.9 Fluorescent *in situ* hybridization

Genomic DNA was extracted from silica-dried or freshly collected leaves of selected accessions using the NucleoSpin Plant II kit (Macherey-Nagel). The highly variable GAG domains of retrotransposable elements (REs) SIRE (*Ty1-Copia*) and Tekay (*Ty3-Gypsy*) were sequentially chosen for FISH probe. The details of probe design is given in Supplementary Methods M2.

Mitotic chromosomes were isolated from root tips of *A. aff. repoeense* (clade B, 2n = 48), *A. aff. curtisii* (clade D1, 2n = 48), *A. trilobum* (clade D3, 2n = 48), and *A. cinnamomeum* (clade D3, 2n = 96) according to Mandáková and Lysák (2016a). The preparation and labeling of DNA probes and FISH followed the published protocols (Mandáková and Lysák, 2016b).

2.10 Comparative analysis of the repeatome structure across monocots

To contextualize the repeatome structure of *Amomum* in a broader context, we conducted an extensive data collection exercise encompassing various genomic and repeatomic characteristics across diverse monocot genera. This endeavor leveraged previously published studies employing diverse methodologies. Our primary data source included plant genome information available up to September 2023 from <https://www.plabipd.de>. We employed this resource to gather a comprehensive array of genomic features, conduct a repeatome analysis, and facilitate comparison across 17 monocot families. In the pursuit of comprehensive data, we thoroughly examined documented repeats from over 150 monocot plant species with a particular focus on multiple publications available for individual species when accessible. Notably, several species featured multiple publications, such as *Musa acuminata* (see Supplementary Table 4). To establish comparison data, we exclusively considered articles that presented information on the various repeat superfamilies and transposable elements, particularly LTR/*Ty1-Copia* and LTR/*Ty3-Gypsy*, which were annotated, quantified, and expressed as percentages relative to the entire genome. Publications utilizing the RepeatExplorer pipeline were excluded, as this method diverges from the approaches employed in the majority of other selected papers. For studies lacking essential details necessary for comparison, we meticulously reviewed the findings, supplementary materials, and other available data. In some instances, we recalculated several repeat families based on the published data. For the purposes of comparison, we used the percentage of the entire repeat content of the particular genome, as well as the percentages of LTR/*Ty1-Copia*

and LTR/*Ty3-Gypsy* elements, LTRs, LINEs and DNA transposons (Supplementary Table 4). Additionally, we extracted and log-transformed published GS for various monocot plant species and families from <https://cvalues.science.kew.org/> for further analysis (Supplementary Table 4). For Zingiberaceae and its subfamilies, we used our own GS measurements (Záveská et al., 2024).

3 Results

3.1 Revealing hybridization events in the evolutionary history of *Amomum*

The PhyloNet analyses of the ‘complete dataset’, comprising 30 *Amomum* species, consistently yielded the same topology with a single reticulation. The optimal model, determined by the highest log-probability and the lowest AIC score over multiple runs with varying priors on the number of reticulations, was obtained from a run specifying two predefined reticulations (Table 1; Supplementary Figure 3). The network’s topology closely aligns with the genus’s phylogeny as previously reported by Hlavatá et al. (2023), distinguishing four primary clades, denoted as A, B, C, and D, along with three subclades within clade D (D1–D3). Additionally, it introduces a reticulation indicating introgression from *Amomum* sp. 7 ($1-\gamma=0.1$, where γ represents the inheritance probability, Yu and Nakhleh, 2015) of clade C (or its ancestor) into ancestor of clade D. For the ‘clade D dataset’, the most suitable network also resulted from a run specifying two predefined reticulations (Table 1; Supplementary Figure 4). This network reveals i) introgression from the ancestor of D1 ($1-\gamma=0.2$) into a specific lineage within subclade D3, here referred to as ‘D3 hybrid’; and ii) introgression from the ancestors of *Amomum* sp. 6 and *A. unifolium* within the D3 subclade ($1-\gamma=0.3$) into the tetraploid *A. cinnamomeum*. The group of species that were not affected by hybridization and form a monophylum within the D3 subclade is further called the ‘D3 parental’ subclade. Figure 1A summarizes the outcomes of PhyloNet analyses on these two datasets, highlighting three significant hybridization events within the genus. As the hybridization events occurred prior to the diversification of specific groups (clade D and clade ‘D3 hybrid’), they align with the definition of ‘ancient hybridization’ proposed by Stull et al. (2023). Subsequently, in the following text, we also employ this term in the same context.

3.2 Genomic variation and repeat composition in *Amomum* and outgroup species

The GS (1C value) of *Amomum* species exhibited considerable variation, ranging from 1,731 Mb in *A. subulatum* to 7,656 Mb in *A. cinnamomeum* (Figure 1B), whereas the outgroup species displayed lower GS, with 1,006 Mb in *Aframomum melegueta* and 1,224 Mb in *Renalmia polypus*. Diploid *Amomum* species exhibited total repeat percentages ranging from 85% in *A. miriflorum* to 88% in *A. calcicola* and *A. subulatum* (Supplementary Table 2). The proportions of repeat content in tetraploids are well within the diploid range, with 87% in *A. aff. biphellum* and *A. cinnamomeum*.

Among the 13 species analyzed by RepeatExplorer (Figure 2), the repeat composition of *Amomum* exhibited significant divergence from that of the outgroup (Figures 2B, D). Several sub-lineages, prevalent in *Amomum*, either underwent substantial amplification or emerged anew within the genus. In the majority of *Amomum* genomes, a significant portion was found to be dominated by LTR retrotransposons of the *Ty1-Copia* superfamily, with *Ty3-Gypsy* lineages representing the second most prevalent element. Unclassified LTRs constituted a substantial portion of the genome in certain species, particularly in the tetraploid *A. cinnamomeum* and diploid *A. miriflorum*. Tandem repeats were more abundant in some species (*A. aff. curtisii*, *A. aff. biphellum*, *A. elan*) while their proportion remained notably low in others (*A. subulatum*, *A. aff. repense*, *A. unifolium*). *A. cinnamomeum* and *A. miriflorum* moreover exhibited a relatively high proportion of unclassified repeats. Single-copy genome content and “small clusters” (those comprising less than 0.01% of reads from the dataset) constituted substantial portions of the genome; nevertheless, their proportions displayed minimal variability among species. For detailed quantification data, see Supplementary Table 2.

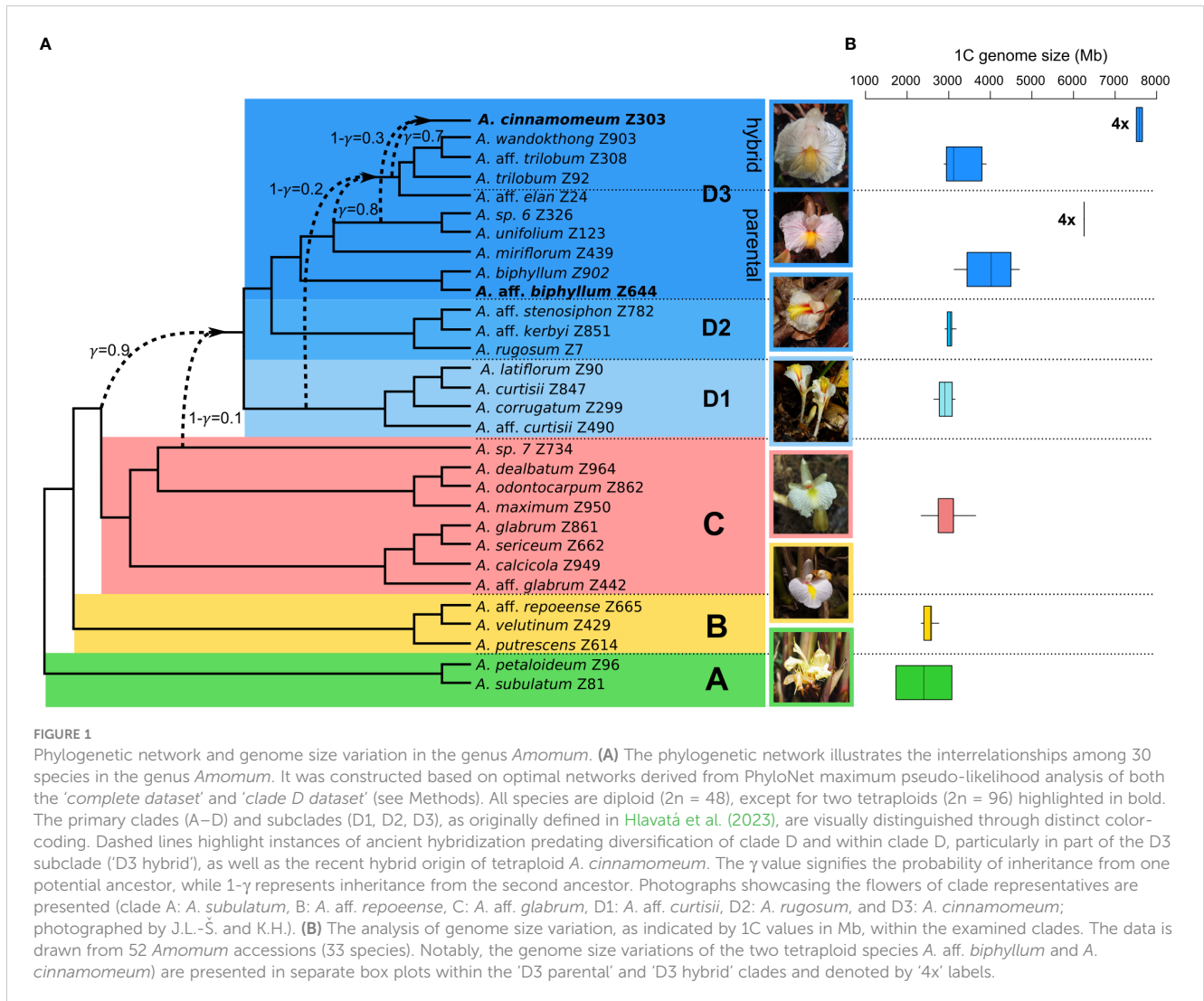
3.3 Comparative analysis reveals genomic distinctions and repeat composition in *Amomum*

The comparative analysis (Figure 2D) unveiled stark disparities between the outgroup and ingroup, as well as variations among

TABLE 1 PhyloNet outcomes and AIC assessments for the determination of the optimal network in the ‘complete dataset’ and ‘clade D dataset’.

	# reticulations	lnL	Δ lnL	# branch lengths	k	AIC	Δ AIC
<i>complete dataset</i>	0	-1374253.32	–	32.2	32.2	2748571.05	1069
	1	-1373843.13	410	35.6	36.6	2747757.46	255
	2	-1373715.25	128	35.8	37.8	2747502.11	0
<i>clade D</i>	0	-261422.00	–	18.8	18.8	522881.61	433
	1	-261291.72	130	21	22	522625.44	177
	2	-261202.79	89	21.4	23.4	522448.37	0

The optimal network is indicated in bold. In this context, ‘lnL’ denotes the likelihood value, while ‘k’ represents the cumulative count of reticulations and branch lengths, serving as the number of parameters involved in the AIC computation.

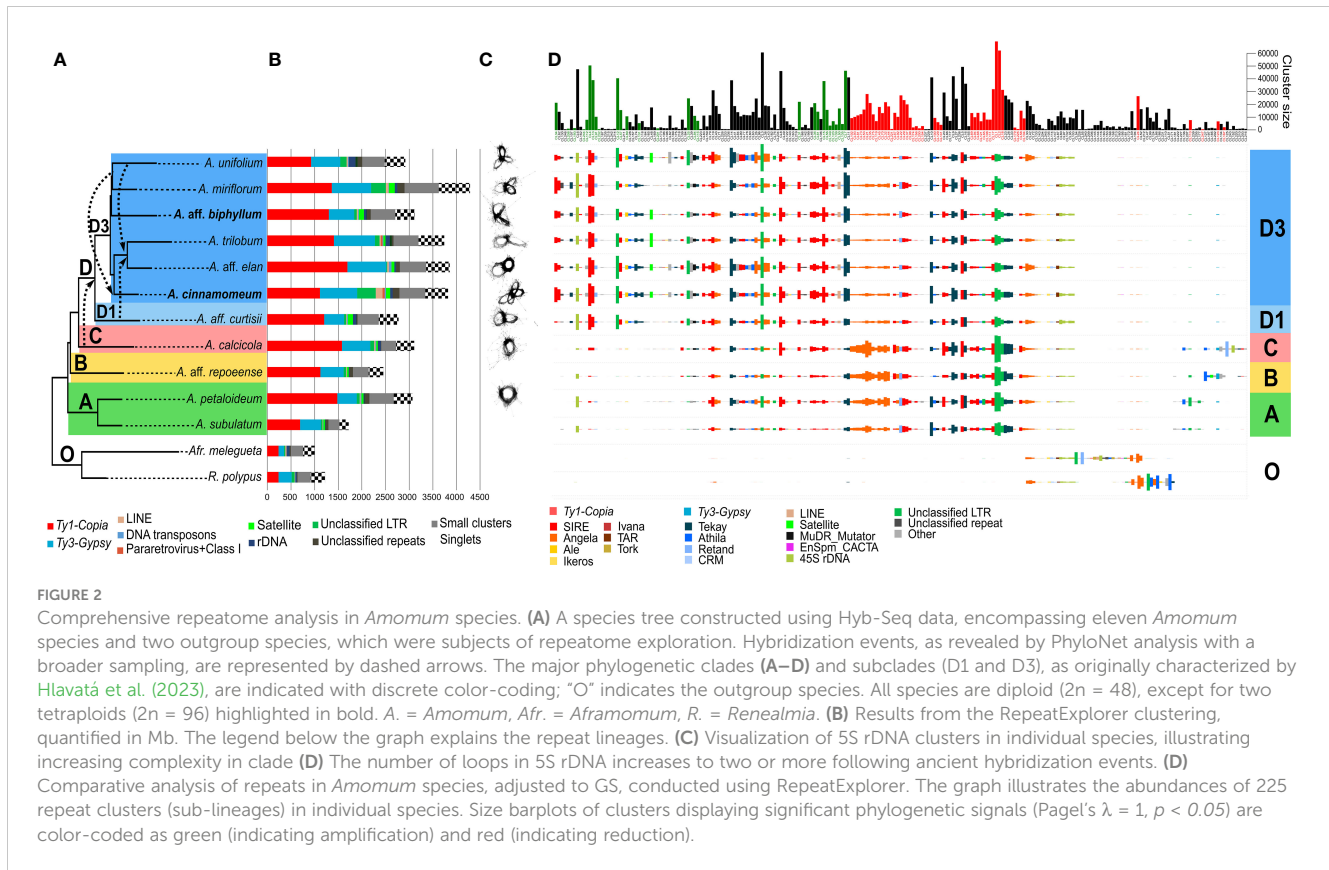


individual clades within the genus *Amomum*. Most repeat sub-lineages shared with the outgroup showed a reduction in *Amomum*, while few experienced amplification. Notably, certain lineages such as *Angela* (*Ty1-Copia*) or *Athila* (*Ty3-Gypsy*) exhibited different sub-lineage compositions in the outgroup compared to *Amomum*. *Amomum* featured several sub-lineages of unclassified LTR repeats not present in the outgroup. Clades A, B and C within *Amomum* exhibited highly similar repeat compositions, with minor distinctions in less abundant repeats, such as *Athila* and *Retand* (*Ty3-Gypsy*). In contrast, clade D showcased the emergence of a new, abundant, unclassified LTR sub-lineage, along with reductions in several other sub-lineages within this clade. Notably, clade D exhibited a pronounced amplification of numerous sub-lineages, including those of SIRE (*Ty1-Copia*), Tekay (*Ty3-Gypsy*), 45S rDNA, and unidentified LTRs, while experiencing reductions in other sub-lineages, particularly within *Angela*, and to a lesser extent, some SIRE and unidentified LTR sub-lineages. The distinctions in repeatome between subclades D1 (represented solely by *A. aff. curtisii*) and D3 were relatively minor, except for the notable amplification of specific SIRE and Tekay sub-lineages in subclade D3. Within subclade D3, unclassified LTRs and unclassified repeats

seemed to contribute to the observed increase in GS in select taxa, such as *A. miriflorum* and tetraploid *A. cinnamomeum*. Chromosome localization of SIRE and Tekay in *A. aff. repoeense* (clade B, $2n = 48$), *A. aff. curtisii* (clade D1, $2n = 48$), *A. trilobum* (clade D3, $2n = 48$), and *A. cinnamomeum* (clade D3, $2n = 96$) confirmed the abundant presence of both elements in the genomes. Their distribution appeared evenly dispersed across chromosomes of all species (Supplementary Figure 5).

The 45S rDNA displayed variable amplification within certain species in clade D, with larger genome sizes such as in *A. unifolium* and *A. trilobum*, but lesser amplification in others such as *A. miriflorum* and *A. aff. elan*. Surprisingly, *A. aff. curtisii*, despite having a smaller genome, exhibited notable 45S rDNA amplification (Figures 2B, D). Ribosomal DNA content varied across the genus, with the smallest amount observed in clade A.

From satellite regions a prominent cluster 48, was shared by most species in the D3 subclade, while a less abundant cluster 185 was exclusively found in *A. repoeense* of clade B (Supplementary Figure 6). While in the comparative analysis some species (e.g., *A. aff. curtisii*, *A. miriflorum*) showed no satellite presence, in a dedicated analysis employing TAREAN additional satellites were



identified, although none appeared to be shared among different *Amomum* species. Notably, the clusters of 5S rDNA, analyzed in relation to phylogeny and hybridization estimates (Figures 2A, C), exhibited increased complexity. In clades A and C, 5S rDNA clusters exhibited a one-looped configuration, whereas in subclade D1 the number of loops increased to two after a presumed ancient hybridization event. Within subclade D3, two and more loops were observed, and the maximum number of loops reached four in the tetraploid *A. cinnamomeum* (Figure 2B).

3.4 Assessing phylogenetic signal in repeat content and its correlations with GS

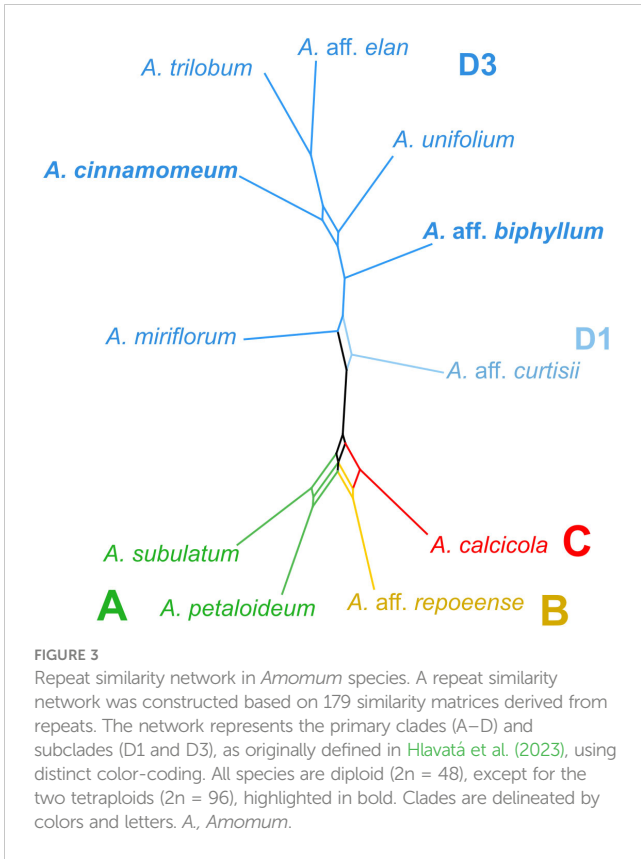
We conducted a comprehensive examination of the phylogenetic signal at various levels, encompassing overall repeat content, superfamilies, and specific lineages. The overall repeat content as well as the content of *Ty3-Gypsy* ($p < 0.05$) demonstrated significant phylogenetic signal (Supplementary Table 3, $p < 0.05$). On the level of repeat lineages we identified significant phylogenetic signals in the quantities of Ale ($p < 0.01$) and Ivana ($p < 0.05$) (both *Ty1-Copia*) and Tekay (*Ty3-Gypsy*, $p < 0.05$). Among the total 225 repeat clusters corresponding to sub-lineages in the comparative analysis, 75 (33.3%) displayed significant phylogenetic signals, as indicated by the presence of red and green bars in the barplot shown in Figure 2D. Of these, 12.4% clusters displayed an increasing trend, while 20.9% exhibited a decreasing trend from clade A towards clade D. The remaining 76.4% of clusters did not demonstrate any significant

phylogenetic signals (Supplementary Table 3). Diverse trends were observed within specific lineages, exemplified by the SIRE lineage, where 10 sub-lineages displayed an increase, while 12 showed a decrease (Figure 2D; Supplementary Figure 7). In certain lineages, all sub-lineages carrying phylogenetically significant signals displayed an increasing trend from clade A to subclade D3 (e.g., Tekay and LINES; Supplementary Table 3). Conversely, all 17 sub-lineages within Angela and both Athila sub-lineages, which conveyed phylogenetic signals, demonstrated a decreasing trend along the phylogeny.

Phylogenetically adjusted correlation tests were performed to assess the relationship between GS and the total amount of repeats, repeat superfamilies and lineages. Notably, a significant correlation was observed between GS and the overall quantity of repeats. Furthermore, significant positive correlations were found between GS and the repeat quantities at the superfamily level for, *Ty1-Copia*, *Ty3-Gypsy* and *LINES* (Supplementary Table 3). Similarly, positive correlations were shown for multiple lineages within above mentioned superfamilies, for satellites, for the group of pararetrovirus and unclassified Class I repeats, and for a group of unclassified repeats.

3.5 Consensus network analysis

In the creation of a consensus network, utilizing 179 matrices representing the observed/expected number of edges between species from the RepeatExplorer clustering analysis (Figure 3), distinct patterns emerged. Specifically, two accessions from

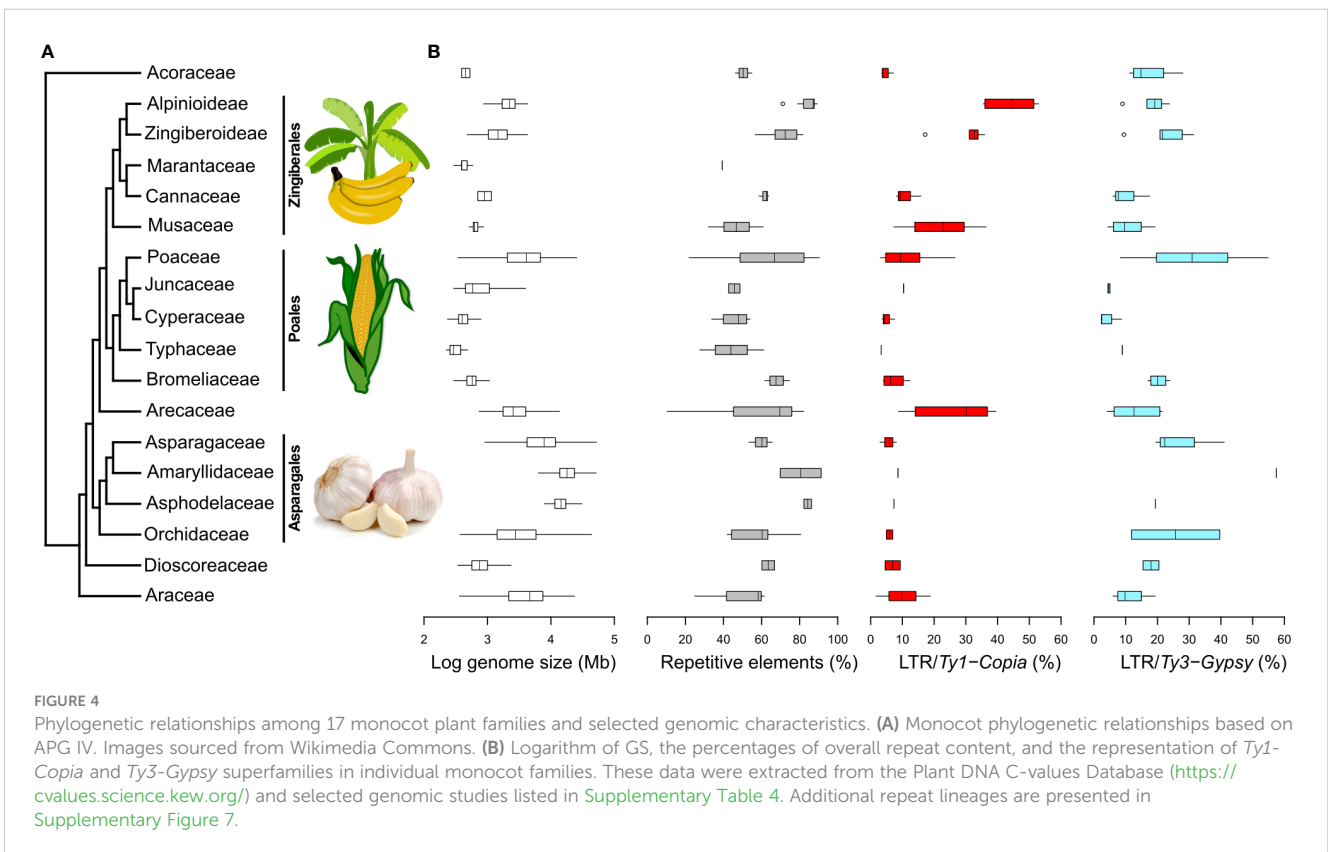


cladeA exhibited close clustering within the network, while accessions representing clades B and C similarly formed a consolidated cluster. Accessions originating from subclade D3 constituted a distinctive cluster, adjacent to the accession representing subclade D1. Notably, the consensus network, constructed based on cluster similarity, demonstrated a remarkable alignment with the nuclear-gene based phylogeny, exhibiting congruence across all major clades (Figure 3).

3.6 Genome size and repeatome structure across monocot families

Within the monocots compared in this study, the range of GS varies from 196 Mb (as observed in *Amorphophallus rivieri*; Zhang et al., 2013) to 80,343 Mb (as evidenced in *Galanthus lagodechianus*; Zonneveld et al., 2003), i.e., 2–5 on log GS scale (Figure 4B). Notably, among the subset of monocots examined in our comparison, GS ranges exhibit considerable diversity, with the most pronounced variations occurring within the Poaceae (Poales) and Asparagaceae (Asparagales) families. The highest absolute GS values are encountered in families belonging to Asparagales.

In our comparison of repeatomes in monocots (Figure 4B), we observe the lowest proportion of repeats among monocots (10.5%) in *Korthalsia laciniosa* (Arecaceae; Ghosh Dasgupta et al., 2021), while the highest proportion (91.3%) is observed in



Allium sativum (Amaryllidaceae; Sun et al., 2020). While garlic displays the most elevated proportion of repeats within the realm of monocots, the average repeat percentage within the Amaryllidaceae family ranks third, trailing behind the Asphodelaceae and Alpinioideae (Zingiberaceae).

The overall pattern of repeat proportion in monocot genomes is inherently reflected in the distribution of LTR proportions. While relatively less data is available for the proportion of LINES, it is generally observed that their proportions are lower, with the maximum proportion reaching less than 20% of the genome. Notably, the pattern of LINE proportions closely aligns with that of *Ty3-Gypsy* proportions, with a conspicuous divergence observed in Poaceae genomes, where both the proportion and range of LINES are notably lower than those of *Ty3-Gypsy* (Supplementary Figure 8).

Regarding DNA transposons, which have been more extensively documented, they are generally observed in relatively low proportions in most monocot families, typically within the range of up to 20% of the genome. Notably, the Cyperaceae genomes exhibit the highest proportion of DNA transposons, while the Bromeliaceae genomes showcase the most extensive range of proportions (Supplementary Figure 8).

The broadest ranges of repeatome proportions are observed in Poaceae and Areaceae genomes. The distribution of *Ty1-Copia* and *Ty3-Gypsy* superfamilies varies among different monocot groups, with Areaceae, Juncaceae, Musaceae, and Zingiberaceae displaying higher proportions of *Ty1-Copia*, while other families exhibit higher proportions of *Ty3-Gypsy*. The most significant quantities of *Ty1-Copia* are found in the family Zingiberaceae, while genomes with the highest *Ty3-Gypsy* proportions are identified in Poaceae.

Areaceae, Poaceae, and Musaceae appear to exhibit broader ranges of both *Ty1-Copia* and *Ty3-Gypsy* percentages in comparison to other monocot groups, although this observation may partly arise from the limited datasets available for some of these groups. Poaceae and Orchidaceae stand out with the widest ranges of *Ty3-Gypsy* proportions.

4 Discussion

4.1 Repeatome proportion in *Amomum* genome is among the largest within monocots

Our comprehensive analysis of repeat proportions across monocot families shows that *Amomum*'s repeat content (85–88%) stands out for its exceptional richness, rivaling representatives from Amaryllidaceae, such as *Allium sativum*, which exhibits one of the highest recorded repeat percentages at 91% (Sun et al., 2020). It's worth emphasizing that *Allium*'s genome is notably larger than other Amaryllidaceae species, as well as those of *Amomum*, with a 1C value of 15,844 Mb.

In contrast to Amaryllidaceae and broader groups such as Asparagales and Poales, where the *Ty3-Gypsy* superfamily predominates the repeatome, the *Ty1-Copia* superfamily takes precedence in shaping the repeatome in Alpinioideae, including *Amomum*. This predominance of *Ty1-Copia*, observed in our

comparative analysis, is also noted in closely related Musaceae (e.g., Novák et al., 2014) and distantly related Areaceae, suggesting a relatively rare pattern within monocots. The reasons for the abundance of one superfamily over the other in the genome are not fully understood due to the lack of literature on differences in functionality between *Ty1-Copia* and *Ty3-Gypsy*. It is speculated that this abundance may result from a combination of stochastic processes, selection, and phylogenetic relatedness (Venner et al., 2009).

Analysis of sequenced genomes of *Musa acuminata* (D'Hont et al., 2012) and *Areca catechu* (Yang et al., 2021) reveals that *Ty1-Copia* elements are concentrated along the centromeric regions. However, genomic studies on Zingiberaceae typically summarize only distribution densities of LTR elements along the genomes, showing either an even distribution of repeats along the chromosomes (e.g., in *Zingiber officinale*, Li et al., 2021) or a decreased concentration of repeats in regions with higher gene densities (e.g., *Curcuma longa*, Yin et al., 2022; *Wurfbainia villosa*, Yang et al., 2022). Comparisons of *Ty1-Copia* distribution between species indicate substantial differences, suggesting variation in plant genome structure (Heslop-Harrison et al., 1997). Our FISH results suggest that repeats in *Amomum* are also evenly distributed. Further detailed studies at the whole-genome level are necessary to better understand the role of *Ty1-Copia* in evolution.

4.2 The impact of repeat dynamics on genome size and evolution of *Amomum*

In our model plant genus *Amomum*, we observed a continuous increase in GS along the phylogeny, attributed to the amplification of *Ty1-Copia* elements, particularly the SIRE lineage. Similar patterns have been observed in the closely related Musaceae family (Novák et al., 2014) and the grass subtribe Loliinae (Moreno-Aguilar et al., 2022). Interestingly, the *Angela* sublineage (*Ty1-Copia*), often associated with significant GS amplification events in other species (e.g., in *Heloniopsis*, Pellicer et al., 2021, or *Passiflora*, Sader et al., 2021), showed a decreasing trend in *Amomum*, similar to observations in Musaceae (Novák et al., 2014). Our findings support previous suggestions that analyses at the sub-lineage level can elucidate the intricate mechanisms driving GS changes (e.g., Suguiyama et al., 2019).

To comprehend why GS increase in some groups while being constrained in others, Schley et al. (2022) investigated repeat-associated genome size expansion in light of ecological correlations. Their findings revealed that water stress inhibits repeat expansion through selection on upper genome size limits. This study builds upon previous suggestions that genome size may impact fitness, with larger genomes offering advantages in certain environments but disadvantages in others (Knight et al., 2005; Faizullah et al., 2021). Robust sampling is crucial to test such hypotheses, as demonstrated by Trávníček et al. (2019) and Carta and Peruzzi (2016). Within the Zingiberaceae family, Závěská et al. (2024) tested the hypothesis of large genome constraint and found that within the subfamily Alpinioideae, to which the genus *Amomum* belongs, plants with larger genomes thrive in shady habitats due to larger stomatal cells

and more efficient photosynthesis (Beaulieu et al., 2008; Hodgson et al., 2010). Light intensity acts as a stressor in these habitats, where selection favors plants with smaller genomes (Knight et al., 2005). In *Amomum*, we have demonstrated that genome size is primarily shaped by repeat dynamics. The next step is to conduct explicit tests on the correlation of particular sub-lineages with various ecological and climatic factors to determine which sub-lineages are directly responsive to environmental stresses. From our current results, we speculate that *Amomum* species from clade D, possessing the largest genomes in Zingiberaceae, are well adapted to shady habitats, while those from basal groups have adapted to cope with light stress by maintaining smaller genomes.

4.3 Hybridization as a potential trigger of repeat amplification

Our detailed examination of the repeatome composition within clade D revealed the most substantial increase in multiple repeat lineages (Figure 2D). Strikingly, clade D was also identified as having an ancient hybrid origin, as previously suggested by cyto-nuclear discordance observed in Hlavatá et al. (2023). While causality cannot be definitively proven, it is plausible that the significant increase in multiple repeat lineages within clade D is closely associated with its hybrid lineage origin. Another clue to this hypothesis comes from the pattern of 5S rDNA clustering. Cluster graphs of species within early derived clades, such as A and C, displayed a single loop of 5S rDNA, consistent with non-hybridogenous species (Garcia et al., 2020). In contrast, the analyzed species from clade D exhibited two or more loops, suggesting at least one hybridization event at the base of clade D, and potentially more within the diversification of subclade D3. Notably, tetraploid *A. cinnamomeum* displayed four loops of 5S rDNA, indicating recent (allopolyploid) as well as ancient hybridization events.

Genomic shock occurs when a significant portion of the repeatome is reactivated or activated anew following hybridization. This phenomenon is often triggered by the merging of sub-genomes, leading to DNA demethylation and the activation of previously silenced repeats (Michalak, 2009; de Tomás and Vicient, 2024). Reactivation of a broad spectrum of repeats post-hybridization typically arises from the breakdown or malfunction of regulatory mechanisms (e.g., Shan et al., 2005; Ungerer et al., 2006; Wei et al., 2021), conferring potential benefits in terms of adaptability (Schradler and Schmitz, 2019). However, this proliferation usually exhibits a finite duration, as transposable elements tend to amplify and diversify within the new genome until either losing the ability to transpose or reactivating the silencing mechanism (Liu et al., 2022).

In various studies, hybridization events have been associated with a burst of specific repeat lineages. Examples include the activation of the *Gorge3* element in *Gossypium* (Hawkins et al., 2009), chromovirus-like retro elements in *Nicotiana* (Renny-Byfield et al., 2013), two *Gypsy*-like retrotransposons in *Phalaenopsis* (Hsu et al., 2020), or one satellite in *Spartina* (Giraud et al., 2021). Our analysis of *Amomum* revealed a significant amplification of 12.4% of sub-lineages in clade D compared to other clades, namely SIRE sub-lineages from the *Ty1-Copia* superfamily.

4.4 Exploring the utility of repeats as molecular markers and phylogenetic tools in *Amomum*

Repeats have proven to be valuable resources in molecular biology and phylogenetics as species-specific or group-specific markers (Rebollo et al., 2010). In the Musaceae, repeats have served as effective molecular markers, as the proliferation of certain groups often accompanied speciation (Novák et al., 2014).

To address insufficient resolution when using other markers, Dodsworth et al. (2017) proposed the utilization of repeats as molecular markers. Vitales et al. (2020) recently reconstructed phylogenetic relationships by employing matrices of similarity between repeat clusters. In *Amomum*, despite the smaller sample size, the repeat-based method provided congruent results, affirming that cluster similarities within *Amomum* can be effectively used to estimate phylogeny or complement other phylogenetic markers. Furthermore, our phylogenetic network analysis provided compelling support for the hybrid origin of the D3 'hybrid' subclade, given its most distinct position within the network. This corroborates previous evidence indicating the hybrid origin of clade D3, strengthening our understanding of *Amomum*'s evolutionary history. On the other hand, we noted that this repeat-based phylogenetic method exhibited higher proportions of uncertainties in the relationships between clades A, B, and C (Figure 4), which mirrors the topological incongruences observed in previous studies involving chloroplast DNA, ribosomal DNA, and nuclear DNA (Hlavatá et al., 2023). These findings suggest that while the repeat-based phylogenetic approach is promising for resolving shallower evolutionary events, it may encounter limitations when addressing deeper phylogenetic relationships (Moreno-Aguilar et al., 2022).

5 Conclusions

In contrast to the predominant presence of *Ty3-Gypsy* elements in most monocot plant families, our study reveals that in *Amomum* (Zingiberaceae) and two other monocot families (Musaceae, Areaceae), *Ty1-Copia* elements are the prevailing component of the repeatome. While explaining why *Ty1-Copia* elements prevail over *Ty3-Gypsy* elements in *Amomum* poses a challenge, we propose that the genome's overall composition and the absolute amounts of repeats influence plant evolution by affecting cell size and photosynthetic efficiency, thereby impacting environmental stress tolerance. Specifically, we suggest that species with smaller genomes (from basal clades) may have faced selection pressure due to their distribution in more stressful (sunny) habitats, while species with larger genomes, facilitated by repeat amplification, are well adapted to shady habitats where selection against larger genomes is less stringent. In conclusion, the increase in genome size within terminal group D of the genus *Amomum* was likely triggered by ancient hybridization events, stimulating the amplification and diversification of various sub-lineages of both *Ty1-Copia* and *Ty3-Gypsy* subfamilies.

Data availability statement

The data generated in the study are deposited in the GenBank Short Reads Archive, accession number PRJNA1029323.

Author contributions

KH: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. EZ: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. JL-Š: Writing – review & editing, Resources. AP: Writing – review & editing, Resources. OŠ: Writing – review & editing, Resources. BK: Writing – review & editing, Investigation. TM: Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization, Investigation. TF: Writing – review & editing, Supervision, Investigation, Conceptualization. MP: Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Czech Science Foundation (projects 22-16826S to TM and GJ20-12579Y to EZ). Additional support was provided as part of a long-term research project of the Czech Academy of Sciences, Institute of Botany (project no. RVO 67985939), Ministry of Culture of the Czech Republic (project no. DKRVO 2019-2023/4.II.d and 2024-2028/4.I.a, National Museum, 00023272 to OŠ). The research of JL-Š is supported by the National Parks Board, Singapore. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure. Computational resources for RepeatExplorer analysis were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure. MetaCentrum provided under the program ‘Projects of Large Research, Development, and Innovations Infrastructures’ (CESNET LM2015042), is greatly appreciated.

References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., and Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytol.* 179, 975986. doi: 10.1111/j.1469-8137.2008.02528.x
- Bennetzen, J. L., MA, J., and Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127–132. doi: 10.1093/aob/mci008
- Bennetzen, J. L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65, 505–530. doi: 10.1146/annurev-arplant-050213-035811
- Cao, Z., Zhu, J., and Nakhleh, L. (2019). Empirical performance of tree-based inference of phylogenetic networks. *19th Int. Workshop Algorithms Bioinformatics (WABI 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi: 10.1101/693986
- Carta, A., and Peruzzi, L. (2016). Testing the large genome constraint hypothesis: plant traits, habitat and climate seasonality in liliaceae. *New Phytol.* 210, 709716. doi: 10.1111/nph.13769
- Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.* 1, 14024. doi: 10.1038/hortres.2014.24
- D'Hont, A., Denoeud, F., Aury, J. M., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213217. doi: 10.1038/nature11241
- de Boer, H., Newman, M. F., Poulsen, A. D., Fér, T., Hien, L., Hlavatá, K., et al. (2018). Convergent morphology in Alpinieae (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. *Taxon* 67, 6–36. doi: 10.12705/671.2

Acknowledgments

The authors would like to thank the Royal Botanic Garden Edinburgh (Sadie Barber and Simon Allen), Singapore Botanic Gardens (Sarah Lim), Prague Botanical Garden (K. Lorencová), V. Lamxay, W.J. Kress, M. Carlsen, D.J. Harris, J. Mood, M.F. Newman, S. Saensouk, H.Đ. Trần and E. Valderrama for providing sample tissues for DNA extractions. Lenka Flašková is acknowledged for wet-lab preparation, Blanka Hamplová and Štěpánka Hrdá for help with sequencing, and Vojtěch Zeisek for help with PhyloNet analyses. Jiří Macas is acknowledged for his advice on the repeatome analysis. Axel Dalberg Poulsen is supported by the players of the People's Postcode Lottery. The Royal Botanic Garden Edinburgh is supported by the Scottish Government's Rural and Environmental Science and Analytical Services Division.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1324358/full#supplementary-material>

- de Tomás, C., and Vicient, C. M. (2024). The genomic shock hypothesis: genetic and epigenetic alterations of transposable elements after interspecific hybridization in plants. *Epigenomes* 8, 2. doi: 10.3390/epigenomes8010002
- Dodsworth, S., Jang, T. S., Struebig, M., Chase, M. W., Weiss-Schneeweiss, H., and Leitch, A. R. (2017). Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). *Plant Syst. Evol.* 303, 1013–1020.
- Faizullah, L., Morton, J. A., Hersch-Green, E. I., Walczyk, A. M., Leitch, A. R., and Leitch, I. J. (2021). Exploring environmental selection on genome size in angiosperms. *Trends Plant Sci.* 26, 1039–1049. doi: 10.1016/j.tplants.2021.06.001
- Fér, T., and Schmickl, R. (2018). HybPhyloMaker: target enrichment data analysis from raw reads to species trees. *Evol. Bioinform.* 14, 1176934317742613. doi: 10.1177/1176934317742613
- Gaiero, P., Vaio, M., Peters, S. A., Schranz, M. E., de Jong, H., and Speranza, P. R. (2019). Comparative analysis of repetitive sequences among species from the potato and the tomato clades. *Ann. Bot.* 123, 521–532. doi: 10.1093/aob/mcy186
- García, S., Wendel, J. F., Borowska-Zuchowska, N., Aïnouche, M., Kuderova, A., and Kovarik, A. (2020). The utility of graph clustering of 5S ribosomal DNA homoecologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00041
- M. A. Garrido-Ramos (Ed.) (2012). *Repetitive DNA* (Basel; New York: Karger).
- Ghosh Dasgupta, M., Dev, S. A., Muneera Parveen, A. B., Sarath, P., and Sreekumar, V. B. (2021). Draft genome of *Korthalsia laciniosa* (Griff.) Mart., a climbing rattan elucidates its phylogenetic position. *Genomics* 113, 2010–2022. doi: 10.1016/j.jygeno.2021.04.023
- Giraud, D., Lima, O., Huteau, V., Coriton, O., Boutte, J., Kovarik, A., et al. (2021). Evolutionary dynamics of transposable elements and satellite DNAs in polyploid *Spartina* species. *Plant Sci. Int. J. Exp. Plant Biol.* 302, 110671. doi: 10.1016/j.plantsci.2020.110671
- Hawkins, J. S., Proulx, S. R., Rapp, R. A., and Wendel, J. F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17811–17816. doi: 10.1073/pnas.0904339106
- Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G., et al. (1997). The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* 100, 197–204. doi: 10.1023/A:1018337831039
- Heyduk, K., McAssey, E. V., Grimwood, J., Shu, S., Schmutz, J., McKain, M. R., et al. (2021). Hybridization history and repetitive element content in the genome of a homoploid hybrid, *Yucca gloriosa* (Asparagaceae). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.573767
- Hlavatá, K., Leong-Škorničková, J., Závěská, E., Šída, O., Newman, M., Mandáková, T., et al. (2023). Phylogenomics and genome size evolution in *Amomum* s.s. (Zingiberaceae): Comparison of traditional and modern sequencing methods. *Mol. Phylogenet. Evol.* 178, 107666. doi: 10.1016/j.ympev.2022.107666
- 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
- Hodgson, J. G., Sharafi, M., Jalili, A., Diaz, S., Montserrat-Martí, G., Palmer, C., et al. (2010). Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog? *Ann. Bot.* 105, 573–584. doi: 10.1093/aob/mcq011
- Holland, B., and Moulton, V. (2003). “Consensus networks: A method for visualising incompatibilities in collections of trees,” in *Algorithms in Bioinformatics*. Eds. G. Benson and R. D. M. Page (Springer Berlin Heidelberg, Berlin, Heidelberg), 165–176.
- Hsu, C.-C., Chen, S.-Y., Lai, P.-H., Hsiao, Y.-Y., Tsai, W.-C., Liu, Z.-J., et al. (2020). Identification of high-copy number long terminal repeat retrotransposons and their expansion in *Phalaenopsis* orchids. *BMC Genomics* 21, 807. doi: 10.1186/s12864-020-07221-6
- 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
- Jansz, N. (2019). DNA methylation dynamics at transposable elements in mammals. *Essays Biochem.* 63, 677–689. doi: 10.1042/EBC20190039
- Junier, T., and Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* 26, 1669–1670. doi: 10.1093/bioinformatics/btq243
- Keuler, R., Garretson, A., Saunders, T., Erickson, R. J., St. Andre, N., Grewe, F., et al. (2020). Genome-scale data reveal the role of hybridization in lichen-forming fungi. *Sci. Rep.* 10, 1497. doi: 10.1038/s41598-020-58279-x
- King, D. G., Soller, M., and Kashi, Y. (1997). Evolutionary tuning knobs. *Endeavour* 21, 36–40. doi: 10.1016/S0160-9327(97)01005-3
- Knight, C. A., Molinari, N. A., and Petrov, D. A. (2005). The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann. Bot.* 95, 177–190. doi: 10.1093/aob/mcl011
- Kumar, S., and Mohapatra, T. (2021). Dynamics of DNA methylation and its functions in plant growth and development. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.596236
- Lee, S.-I., and Kim, N.-S. (2014). Transposable elements and genome size variations in plants. *Genomics Inform.* 12, 87–97. doi: 10.5808/GI.2014.12.3.87
- Li, H.-L., Wu, L., Dong, Z., Jiang, Y., Jiang, S., Xing, H., et al. (2021). Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic. Res.* 8, 189. doi: 10.1038/s41438-021-00627-7
- Liao, X., Hu, K., Salhi, A., Zou, Y., Wang, J., and Gao, X. (2022). msRepDB: a comprehensive repetitive sequence database of over 80 000 species. *Nucleic Acids Res.* 50, D236–D245. doi: 10.1093/nar/gkab1089
- Liu, P., Cuerda-Gil, D., Shahid, S., and Slotkin, R. K. (2022). The epigenetic control of the transposable element life cycle in plant genomes and beyond. *Annu. Rev. Genet.* 56, 63–87. doi: 10.1146/annurev-genet-072920-015534
- Luo, X., Chen, S., and Zhang, Y. (2022). PlantRep: a database of plant repetitive elements. *Plant Cell Rep.* 41, 1163–1166. doi: 10.1007/s00299-021-02817-y
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010
- Mandáková, T., and Lysák, M. A. (2016a). “Chromosome preparation for cytogenetic analyses in *Arabidopsis*,” in *Current Protocols in Plant Biology*. Eds. G. Stacey, J. Birchler, J. Ecker, C. R. Martin, M. Stitt and J.-M. Zhou (John Wiley & Sons, Inc, Hoboken, NJ, USA), 43–51. doi: 10.1002/cppb.20009
- Mandáková, T., and Lysák, M. A. (2016b). “Painting of *Arabidopsis* chromosomes with chromosome-specific BAC clones,” in *Current Protocols in Plant Biology*. Eds. G. Stacey, J. Birchler, J. Ecker, C. R. Martin, M. Stitt and J.-M. Zhou (John Wiley & Sons, Inc, Hoboken, NJ, USA), 359–371. doi: 10.1002/cppb.20022
- Mehrotra, S., and Goyal, V. (2014). Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinf.* 12, 164–171. doi: 10.1016/j.gpb.2014.07.003
- Meudt, H. M., Rojas-Andrés, B. M., Prebble, J. M., Low, E., Garnock-Jones, P. J., and Albach, D. C. (2015). Is genome downsizing associated with diversification in polyploid lineages of *Veronica*? *Bot. J. Linn. Soc.* 178, 243–266. doi: 10.1111/boj.2015.178.issue-2
- Michalak, P. (2009). Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* 102, 45–50. doi: 10.1038/hdy.2008.48
- Microsoft Corporation (2018). Microsoft Excel 365. Available at: <https://office.microsoft.com/excel>.
- Moreno-Aguilar, M. F., Inda, L. A., Sánchez-Rodríguez, A., Arnelas, I., and Catalán, P. (2022). Evolutionary dynamics of the repeatome explains contrasting differences in genome sizes and hybrid and polyploid origins of grass liliinae lineages. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.901733
- Neumann, P., Novák, P., Hošťáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob. DNA* 10, 1. doi: 10.1186/s13100-018-0144-1
- Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 45, e111–e111. doi: 10.1093/nar/gkx257
- Novák, P., Hřibová, E., Neumann, P., Koblížková, A., Doležel, J., and Macas, J. (2014). Genome-wide analysis of repeat diversity across the family musaceae. *PLoS One* 9, e98918. doi: 10.1371/journal.pone.0098918
- Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi: 10.1038/s41596-020-0400-y
- O’Neill, R. J. W., O’Neill, M. J., and Graves, J. A. M. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72. doi: 10.1038/29985
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., et al. (2018). *caper: Comparative Analyses of Phylogenetics and Evolution in R*. Available at: <https://CRAN.R-project.org/package=caper>.
- Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zool. Scr.* 26, 331–348. doi: 10.1111/j.1463-6409.1997.tb00423.x
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 877. doi: 10.1038/44766
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Pellicer, J., Fernández, P., Fay, M. F., Micháková, E., and Leitch, I. J. (2021). Genome size doubling arises from the differential repetitive DNA dynamics in the genus *heloniopsis* (Melanthiaceae). *Front. Genet.* 12. doi: 10.3389/fgene.2021.726211
- Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes* 9, 88. doi: 10.3390/genes9020088
- Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., et al. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30, 22162218. doi: 10.1093/bioinformatics/btu181
- Pulido, M., and Casacuberta, J. M. (2023). Transposable element evolution in plant genome ecosystems. *Curr. Opin. Plant Biol.* 75, 102418. doi: 10.1016/j.pbi.2023.102418
- Qiao, X., Zhang, S., and Paterson, A. H. (2022). Pervasive genome duplications across the plant tree of life and their links to major evolutionary innovations and transitions. *Comput. Struct. Biotechnol. J.* 20, 3248–3256. doi: 10.1016/j.csbj.2022.06.026

- R Core Team. (2022). *R: A language and environment for statistical computing*. Available at: <https://www.R-project.org/>.
- Rebollo, R., Horard, B., Hubert, B., and Vieira, C. (2010). Jumping genes and epigenetics: Towards new species. *Gene* 454, 1–7. doi: 10.1016/j.gene.2010.01.003
- Renny-Byfield, S., Kovarik, A., Kelly, L. J., Macas, J., Novak, P., Chase, M. W., et al. (2013). Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J.* 74, 829–839. doi: 10.1111/tpj.12168
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–233. doi: 10.1111/j.2041-210X.2011.00169.x
- Sader, M., Vaio, M., Cauz dos Santos, L., Dornelas, M., Vieira, M.-L., Melo, N., et al. (2021). Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. *Planta* 253. doi: 10.1007/s00425-021-03598-0
- Schley, R. J., Pellicer, J., Ge, X.-J., Barrett, C., Bellot, S., Guignard, M. S., et al. (2022). The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. *New Phytol.* 236, 433–446. doi: 10.1111/nph.18323
- Schrader, L., and Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Mol. Ecol.* 28, 1537–1549. doi: 10.1111/mec.14794
- Shan, X., Liu, Z., Dong, Z., Wang, Y., Chen, Y., Lin, X., et al. (2005). Mobilization of the Active MITE Transposons mPing and Pong in Rice by Introgression from Wild Rice (*Zizania latifolia* Griseb.). *Mol. Biol. Evol.* 22, 976–990. doi: 10.1093/molbev/msi082
- Simonin, K. A., and Roddy, A. B. (2018). Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol.* 16, e2003706. doi: 10.1371/journal.pbio.2003706
- Skopalíková, J., Leong-Škorníčková, J., Šída, O., Newman, M., Chumová, Z., Zeisek, V., et al. (2023). Ancient hybridization in *Curcuma* (Zingiberaceae)—Accelerator or brake in lineage diversifications? *Plant J.* 116, 773785. doi: 10.1111/tpj.16408
- Smit, A., Hubley, R., and Green, P. (2013). *RepeatMasker open-4.0*. Available online at: <http://www.repeatmasker.org>.
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 13121313. doi: 10.1093/bioinformatics/btu033
- Staton, S. E., Bakken, B. H., Blackman, B. K., Chapman, M. A., Kane, N. C., Tang, S., et al. (2012). The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* 72, 142–153. doi: 10.1111/j.1365-3113X.2012.05072.x
- Stull, G. W., Pham, K. K., Soltis, P. S., and Soltis, D. E. (2023). Deep reticulation: the long legacy of hybridization in vascular plant evolution. *Plant J.* 114, 743–766. doi: 10.1111/tpj.16142
- Suguiyama, V., Vasconcelos, L., Rossi, M., Biondo, C., and Setta, N. (2019). The population genetic structure approach adds new insights into the evolution of plant LTR retrotransposon lineages. *PLoS One* 14, e0214542. doi: 10.1371/journal.pone.0214542
- Sun, X., Zhu, S., Li, N., Cheng, Y., Zhao, J., Qiao, X., et al. (2020). A Chromosome-Level Genome Assembly of Garlic (*Allium sativum*) Provides Insights into Genome Evolution and Allicin Biosynthesis. *Mol. Plant* 13, 1328–1339. doi: 10.1016/j.molp.2020.07.019
- Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinf.* 9, 322. doi: 10.1186/1471-2105-9-322
- Trávníček, P., Čertner, M., Ponert, J., Chumová, Z., Jersáková, J., and Suda, J. (2019). Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked to partial endoreplication, plant life-history traits and climatic conditions. *New Phytol.* 224, 1642–1656. doi: 10.1111/nph.15996
- Ungerer, M. C., Strakosh, S. C., and Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol. CB* 16, R872–R873. doi: 10.1016/j.cub.2006.09.020
- Venner, S., Feschotte, C., and Biémont, C. (2009). Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* 25, 317–323. doi: 10.1016/j.tig.2009.05.003
- Vitales, D., Garcia, S., and Dodsworth, S. (2020). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Mol. Phylogenet. Evol.* 147, 106766. doi: 10.1016/j.ympev.2020.106766
- Wei, G., Li, X., and Fang, Y. (2021). Sympatric genome size variation and hybridization of four oak species as determined by flow cytometry genome size variation and hybridization. *Ecol. Evol.* 11, 1729–1740. doi: 10.1002/ece3.7163
- Yang, Y., Huang, L., Xu, C., Qi, L., Wu, Z., Li, J., et al. (2021). Chromosome-scale genome assembly of areca palm (*Areca catechu*). *Mol. Ecol. Resour.* 21, 2504–2519. doi: 10.1111/1755-0998.13446
- Yang, P., Zhao, H.-Y., Wei, J.-S., Zhao, Y.-Y., Lin, X.-J., Su, J., et al. (2022). Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of *Wurfbainia villosa*. *Plant J.* 112, 630–645. doi: 10.1111/tpj.15968
- Yin, Y., Xie, X., Zhou, L., Yin, X., Guo, S., Zhou, X., et al. (2022). A chromosome-scale genome assembly of turmeric provides insights into curcumin biosynthesis and tuber formation mechanism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1003835
- Yu, Y., and Nakhleh, L. (2015). “A distance-based method for inferring phylogenetic networks in the presence of incomplete lineage sorting,” in R. Harrison, Y. Li and I. Mándou eds. *Bioinformatics Research and Applications. ISBRA 2015*. Lecture Notes in Computer Science. (Springer, Cham) 9096, 378–389. doi: 10.1007/978-3-319-19048-8_32
- Záveská, E., Šída, O., Leong-Škorníčková, J., Chumová, Z., Trávníček, P., Newman, M. F., et al. (2024). Testing the large genome constraint hypothesis in tropical rhizomatous herbs: life strategies, plant traits and habitat preferences in gingers. *Plant J.* 117, 12231238. doi: 10.1111/tpj.16559
- Zhang, L., Cao, B., and Bai, C. (2013). New reports of nuclear DNA content for 66 traditional Chinese medicinal plant taxa in China. *Caryologia* 66, 375–383. doi: 10.1080/00087114.2013.859443
- Zonneveld, B., Grimshaw, J., and Davis, A. (2003). The systematic value of nuclear DNA content in *galanthus*. *Plant Syst. Evol.* 241, 89102. doi: 10.1007/s00606-003-0016-z