



Phylogenomic Relationships and Evolution of Polyploid *Salix* Species Revealed by RAD Sequencing Data

Natascha D. Wagner^{1*}, Li He² and Elvira Hörandl¹

¹ Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, Göttingen, Germany, ² College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

OPEN ACCESS

Edited by:

Marcial Escudero,
Universidad de Sevilla, Spain

Reviewed by:

Ales Kovarik,
Academy of Sciences of the Czech
Republic (ASCR), Czechia
Michael R. McKain,
University of Alabama, United States
Yingnan Chen,
Nanjing Forestry University, China

*Correspondence:

Natascha D. Wagner
natascha.wagner@uni-goettingen.de

Specialty section:

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

Received: 14 April 2020

Accepted: 30 June 2020

Published: 17 July 2020

Citation:

Wagner ND, He L and Hörandl E
(2020) Phylogenomic Relationships
and Evolution of Polyploid
Salix Species Revealed by
RAD Sequencing Data.
Front. Plant Sci. 11:1077.
doi: 10.3389/fpls.2020.01077

Polyploidy is common in the genus *Salix*. However, little is known about the origin, parentage and genomic composition of polyploid species because of a lack of suitable molecular markers and analysis tools. We established a phylogenomic framework including species of all described sections of Eurasian shrub willows. We analyzed the genomic composition of seven polyploid willow species in comparison to putative diploid parental species to draw conclusions on their origin and the effects of backcrossing and post-origin evolution. We applied recently developed programs like SNAPP, HyDe, and SNIploid to establish a bioinformatic pipeline for unravelling the complexity of polyploid genomes. RAD sequencing revealed 23,393 loci and 320,010 high quality SNPs for the analysis of relationships of 35 species of Eurasian shrub willows (*Salix* subg. *Chamaetia/Vetrix*). Polyploid willow species appear to be predominantly of allopolyploid origin. More ancient allopolyploidization events were observed for two hexaploid and one octoploid species, while our data suggested a more recent allopolyploid origin for the included tetraploids and identified putative parental taxa. SNIploid analyses disentangled the different genomic signatures resulting from hybrid origin, backcrossing, and secondary post-origin evolution in the polyploid species. Our RAD sequencing data demonstrate that willow genomes are shaped by ancient and recent reticulate evolution, polyploidization, and post-origin divergence of species.

Keywords: phylogenomics, polyploidy, restriction associated DNA sequencing, shrub willows, SNIploid

INTRODUCTION

Although the impact of polyploidy in plant evolution is evident (Soltis and Soltis, 2009; Wood et al., 2009; Jiao et al., 2011; Mayrose et al., 2011; Barker et al., 2016), there is much to discover about the impact of polyploidy on evolution of plant genomes. Polyploids may originate from combining genomes via hybridization (allopolyploids) or from intraspecific genome duplication (autopolyploids). Post-origin evolution and introgression may further leave genomic signatures (Peralta et al., 2013). This information, however, is crucial to understand the evolution of species. The potential offered by the rise of high-throughput sequencing tools to analyze the origin of natural polyploid plant species has not yet been fully exploited. The main reasons might be the difficulties in variant calling, especially the need to distinguish different alleles derived from the same parent

(homologs) from alleles originating from different parents (homeologs; Clevenger et al., 2015). Other difficulties are the necessity to filter paralogs (Hirsch and Buell, 2013), the uncertainty of allelic dosage as well as the lack of appropriate bioinformatic tools and models (Meirmans et al., 2018). Given a sufficient coverage, high-throughput sequencing methods are generally suited to detect all different alleles at a specific locus. However, to use a massive amount of short read data for the analysis of polyploids is still challenging, especially, if no reference genome is available.

Restriction Associated DNA (RAD) sequencing (Baird et al., 2008) is a widely used method that can harvest ten-thousands of informative SNPs representing the whole genome. The huge amount of biallelic SNPs is well suited to study evolutionary processes, resolve reticulate relationships, reconstruct species trees, and to analyze polyploid species in a phylogenetic framework (Bryant et al., 2012; Dufresne et al., 2014; Clevenger et al., 2015; Clevenger et al., 2018). However, there are only few studies using RAD sequencing data when dealing with polyploid non-model plants (Chen et al., 2013; Mastretta-Yanes et al., 2014; Qi et al., 2015; Brandrud et al., 2020; He et al., 2020).

The genus *Salix* L. (Salicaceae) comprises about 400-450 species of trees and shrubs that are mainly distributed in the Northern Hemisphere. About two thirds of the species occur in Eurasia, while about 140 species are distributed in America. Classical taxonomy and systematics in *Salix* have proven to be extremely difficult because of dioecy, simple, reduced flowers, common natural formation of hybrids, high intraspecific phenotypic variation, and the presence of polyploid species (Skvortsov, 1999; Hörandl et al., 2012; Cronk et al., 2015). Phylogenetic analyses using traditional Sanger sequencing and standard markers (like ITS) were able to separate subgenus *Salix* from a big clade uniting the two subgenera *Chamaetia* and *Vetrix*, but lacked any interspecific resolution (Leskinen and Alström-Rapaport, 1999; Chen et al., 2010; Percy et al., 2014; Lauron-Moreau et al., 2015; Wu et al., 2015). Especially in the case of polyploidization and frequent hybridization, the usage of uniparentally inherited plastid markers is useful for phylogenetic reconstruction. However, plastid markers show only little variation within willows (see Leskinen and Alström-Rapaport, 1999; Chen et al., 2010; Savage and Cavender-Bares, 2012; Percy et al., 2014; Lauron-Moreau et al., 2015; Wu et al., 2015). More recently, the complete plastomes of twelve *Salix* species were reconstructed and showed only 0.9% variability (N.D. Wagner, unpublished data). The *Chamaetia/Vetrix* clade comprises three quarters of species diversity in *Salix* ranging from creeping arctic-alpine dwarf shrubs to medium-sized trees (Wagner et al., 2018). In contrast to many other woody plant lineages about 40% of the species are polyploid (Suda and Argus, 1968), ranging from diploid to octoploid species (rarely deca-/dodecaploid), based on the chromosome number $x = 19$ (Argus, 1997). Previous studies proposed that ancient genome duplications occurred several times in *Salix* (Dorn, 1976; Leskinen and Alström-Rapaport, 1999) and in the Salicoid lineage (Tuskan et al., 2006). Paleopolyploid species are more difficult to recognize, because their cytological behaviour

has undergone 'diploidization', i.e. they have returned to regular chromosome pairing at meiosis as typical for diploid sexual species (Comai, 2005). Cold climatic regimes during the Pleistocene may have triggered spontaneous autopolyploidization by unreduced gamete formation (Ramsey and Schemske, 1998). Recent allopolyploidization ('young polyploidy') may have happened during the Pleistocene, when climatic oscillations caused range fluctuations of plant species and enhanced secondary contact hybridization of previously isolated species (Hewitt, 2004; Abbott et al., 2013; Kadereit, 2015). While homoploid hybridization may result in introgressive hybridization with the parental species (Hardig et al., 2010; Gramlich and Hörandl, 2016; Gramlich et al., 2018), allopolyploidy can establish a rapid crossing barrier against the parents and hence may result in the evolution of an independent lineage. In willows, natural hybridization is a frequent phenomenon (Argus, 1997; Skvortsov, 1999; Hörandl et al., 2012) and may occur even between distantly related species (e.g. Hardig et al., 2000; Gramlich et al., 2018). While in the 19th century experimental crosses in *Salix* were popular and even triple or multiple hybrids have been produced (Wichura, 1865), the origin of most natural polyploid species is still unknown. Population genetic markers like SSRs and AFLPs have been used so far to study polyploids (Barcaccia et al., 2014; Guo et al., 2016), but in the *Chamaetia/Vetrix* clade the polyploid species have never been analyzed with molecular markers. We expect to trace signatures of ancient and/or recent hybrid origin, secondary lineage-specific evolution, backcrossing, and introgression in their genomes.

RAD sequencing (Baird et al., 2008) was recently used to overcome the lack of phylogenetic information within the genus *Salix* (Gramlich et al., 2018; Wagner et al., 2018; He et al., 2020). Wagner et al. (2018) published the first well-resolved phylogeny of diploid European members of the *Chamaetia/Vetrix* clade, while He et al. (2020) recognized a recent adaptive radiation of willows in the Hengduan mountains. Gramlich et al. (2018) successfully used RAD sequencing to analyze recent homoploid hybridization patterns and introgression between two *Salix* species on alpine glacier forefields. RAD loci in *Salix* represent almost exclusively the nuclear genome, and provide 10-thousands of SNPs from both conservative and rapidly evolving non-coding genomic regions (Gramlich et al., 2018). Therefore, this method has the potential to resolve reticulate relationships and to analyze the origin of the polyploid species in *Salix*.

In this study we want to test the utility of RAD sequencing data for the analysis of polyploid willow species. We will use a comprehensive approach of different traditional and recently developed analytical tools to elucidate the evolutionary scenarios of polyploid origin in Eurasian shrub willows. By using various approaches, we will 1) establish a phylogenomic framework of the *Vetrix/Chamaetia* clade, 2) test for recent allopolyploidization by hybridization of two (or more) related diploid parents versus more ancient allopolyploidization of distantly related progenitors of extant species, and 3) test for autopolyploidization. Based on this information, we 4) try to get insights into post-origin evolution of polyploid species by analyzing their genomic composition.

MATERIAL AND METHODS

Sampling

For this study, we sampled 28 diploid species (incl. two subspecies), one triploid, three tetraploid, two hexaploid and one octoploid species representing 19 sections sensu Skvortsov (1999) of the *Chamaetia/Vetrix* clade. *Salix triandra* (subg. *Salix*) was included to serve as outgroup (Wu et al., 2015). Hence, the complete sample set consists of 36 *Salix* species. The samples were collected in Central and Northern Europe as well as in China and determined after Fang et al. (1999); Skvortsov (1999) and Hörandl et al. (2012). The diploid species represent all main European lineages (Rechinger and Akeroyd, 1993). Leaves were dried in silica gel and herbarium voucher specimens were deposited in the herbarium of the University of Goettingen (GOET). Two to five accessions per species were included in the analyses given a total of 133 samples (**Supplement Table S1**).

Ploidy Determination Using Flow Cytometry

The ploidy levels of seven species (23 samples) with unknown chromosome counts were estimated using flow cytometric analyses based on DAPI (4', 6-diamidino-2-phenylindole) fluorochrome applying a modified protocol of Suda and Trávníček (2006) as described in He et al. (2020). *Salix caprea* with known ploidy level ($2x=2n=38$) was used as an external standard.

Molecular Treatment and Analyses

The DNA of all samples was extracted using the Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions (Valencia, CA). After quality check, the DNA was sent to Floragenex, Inc. (Portland, Ore., USA) where the sequencing library preparation was conducted after Baird et al. (2008) using the restriction enzyme *Pst*I (see Wagner et al., 2018 for details). Polyploids require an increased depth of coverage based on the bigger genome size and the higher number of alleles (Hirsch and Buell, 2013; Clevenger et al., 2015). Thus, we sequenced polyploid and diploid taxa on different plates to avoid loss of coverage.

There are several tools available to analyse RAD sequencing data. While e.g. STACKS (Catchen et al., 2013) is more suited to population genetic analyses, ipyrad is predominantly used for phylogenetic approaches (see Eaton et al., 2017). Ipyrad conducts quality filtering and de novo locus identification and genotyping, with the advantage that it can handle insertion-deletion variation among alleles, and it is therefore better suited for studies of a broader taxonomic scale (Andrews et al., 2016). The quality of the resulting single-end 100bp long sequence reads was checked using FastQC v.0.10.1 (Andrews, 2010). After de-multiplexing, the reads were used to run ipyrad v.0.7.28 (Eaton and Overcast, 2020) with a clustering threshold of 85% and a minimum depth of eight reads for base calling. The clustering was done with VSEARCH as implemented in ipyrad v.0.7.28. The maximum number of SNPs per locus was set to 20, the maximum number of indels to 8. We set a threshold of maximal four alleles per site in the final cluster filtering. Ipyrad summarizes the underlying allelic information into a consensus sequence in the form of ambiguous sites at heterozygous positions, which is a common

strategy to face the problem of more than two alleles (Blischak et al., 2018b). The resulting sequence information can be used for downstream analyses. Initial settings for the minimum number of samples sharing a locus (m) ranged from $m4$ (average number of accessions per species) to $m120$ (loci present in 90% of accessions) and were optimized as described in Wagner et al. (2018). Statistics for the tested settings are summarized in **Supplement Table S2**. The $m40$ dataset (29.06% missing data) was used for phylogenetic analyses and the $m100$ dataset (9.19% missing data) for the genetic structure analysis. Additionally, we performed tree reconstruction with ipyrad for each clade separately that contains one or more polyploid samples applying the same settings.

We inferred phylogenetic relationships on concatenated alignments of the complete dataset as well as for each clade separately by using the GTR+ Γ model of nucleotide substitution implemented in RAxML v.8.2.4 (Stamatakis, 2014). We conducted for each ML analysis a rapid bootstrapping (BS) analysis with 100 replicates. Next to BS we applied *quartetsampling* (QS; Pease et al., 2018) with default settings to test statistical support of a given topology. We ran 300 replicates using the $-L$ option (minimum likelihood differential). QS is able to distinguish between conflicting signals and poor phylogenetic information. The Quartet Concordance score [QC, (1,-1)] gives the support of the current topology, the Quartet Differential score [QD, (0,1)] is an indicator of introgression and the Quartet Informativeness score [QI, (0,1)] quantifies the informativeness of each branch. For each phylogeny shown here, the observed QS values (QC/QD/QI) were visualized along with the BS values above and below branches, respectively.

Species Tree Estimation Using SNAPP

The SNAPP method (Bryant et al., 2012) estimates species trees directly from biallelic markers (e.g., SNP data), and bypasses the necessity of sampling the gene trees at each locus. This is specifically advantageous for RAD sequencing data sets based on short loci. For species tree calculation with SNAPP v1.4.2 we used the unlinked SNPs output (one SNP per locus, uSNPs) of the clade-specific ipyrad pipeline. We created the input file using BEAUti v2.5.1 (Bouckaert et al., 2014). The multiple accessions of a species were assigned to one taxon. As priors we used default settings for speciation rate (λ) (1/X). Non-polymorphic sites were excluded. We ran the analysis in BEAST2 (Bouckaert et al., 2014). The number of mcmc generations was up to 10,000,000 and we sampled every thousandth tree. The results were checked with TRACER 1.7.1 (Rambaut et al., 2018). The resulting trees were summarized in the SNAPP tree analyzer tool. The trees of all RAxML and SNAPP analyses were obtained in FigTree v1.4.3 (Rambaut, 2014).

Genetic Structure and Network Analyses

To test for an influence of reticulate evolution on the genetic composition of the included species, we explored the genetic structure of each sample. We simplified the data to be able to combine diploids and polyploids in the same STRUCTURE analyses by summarizing the observed allelic information of a polyploid to a "diploid" consensus sequence (e.g. the four alleles "AATT" will be reduced to "AT"). We used the uSNPs of the

m100 data set (< 10% missing data) for the complete dataset to avoid bias by too much missing data. The clade specific analyses were performed using the uSNP data of the clade specific ipyrad runs. We chose a burn-in of 10,000 and a MCMC of 100,000 replicates, with three iterations of each value of K (K=number of genotypic groups). After an initial test, the range of K was set from 2 to 7 in the final analysis. For the subclades, K was set from two up to the number of taxa included. The optimal K value was estimated by inferring the Evanno test revealing the optimal delta K value in Structure Harvester (Earl and vonHoldt, 2012).

Reticulate relationships resulting from hybridization or allopolyploidy are not well represented by bifurcating tree topologies (Huson and Bryant, 2006; McBreen and Lockhart, 2006). To overcome this problem, we utilized SplitsTree4 (Huson and Bryant, 2006) in order to reconstruct possible network-like evolutionary relationships among the species for each clade. Based on informative SNP data (see Feng et al., 2018), we generated the split network by implementing NeighbourNet analysis with variance of ordinary least squares complemented by a bootstrapping with 1000 replicates to test for statistical support. Missing data were treated as unknown in all analyses.

Hybridization Detection

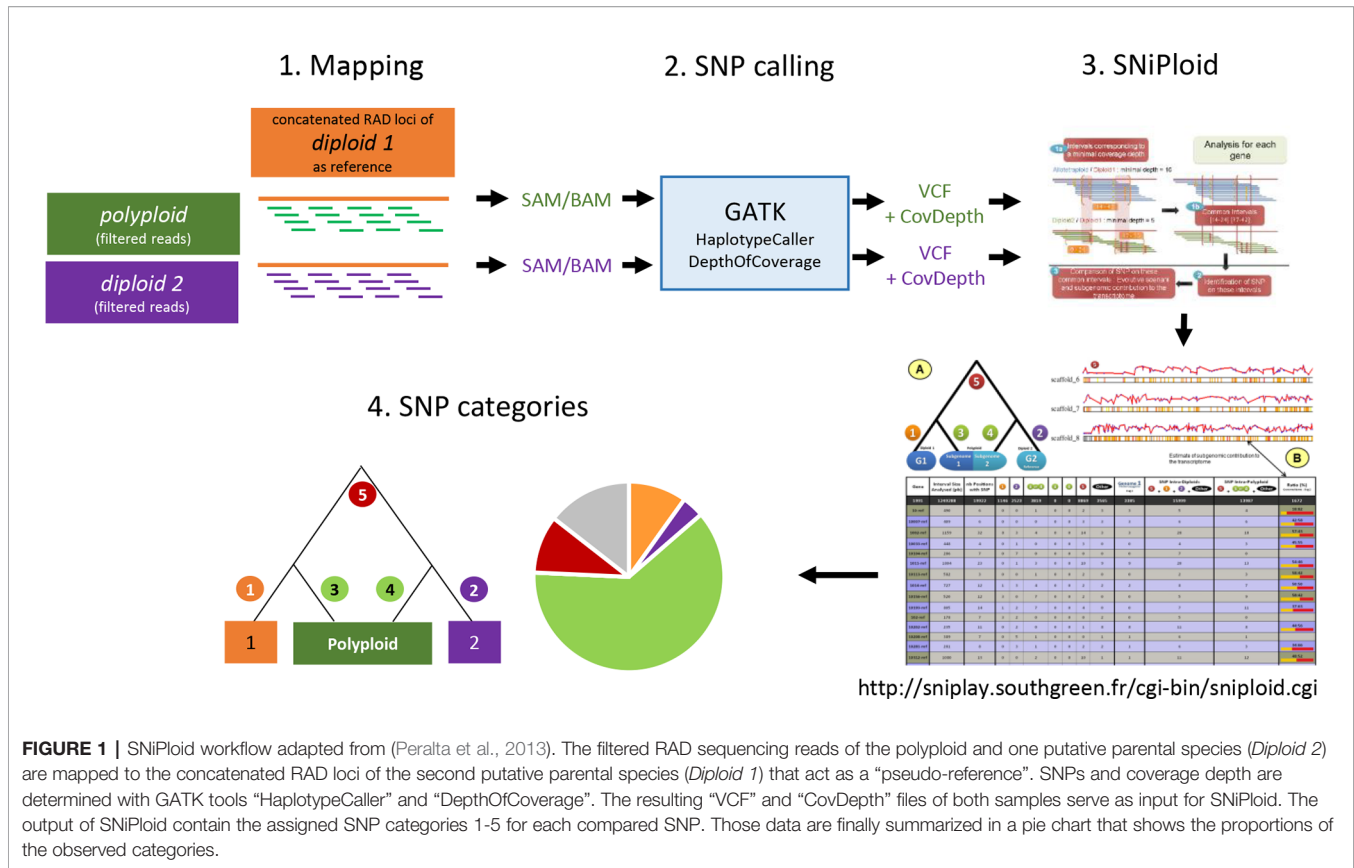
HyDe (Hybridization Detection; Blischak et al., 2018a) allows testing for hybridization and introgression at a population or species level based on D-Statistics by estimating the amount of admixture (γ). We applied the “run_hyde_mp.py” script to test for putative parent-hybrid combinations with respect to the polyploid individuals. HyDe uses p-values to test for the significance of results. While a 50:50 hybrid is characterized by a γ -value of about 0.5, very low levels of admixture (e.g. 0.01 = close to parent P1; 0.99 = close to parent P2) may be indicators for several processes such as ILS and more ancient hybridization. After initial tests, we used a range of $\gamma=0.4-0.6$ to identify recent hybridization events in the data set, and intermediate ranges of $\gamma= 0.1-0.4$ and $0.6-0.9$ for older events. We excluded significant values <0.1 and >0.9 . The complete sequences as well as the uSNP data of the clade-specific analyses were used as input data, as the authors stated this datatype as the “most appropriate input” (Blischak et al., 2018a). We tested two approaches: first, we assigned all individuals of a species to one taxon, and second, we tested all individuals as separate entities. The latter approach is better to detect individual admixture. To underline the usage of HyDe in RAD sequencing data we performed a preliminary test on six diploid naturally formed F1 hybrids between *S. helvetica* and *S. purpurea* (Gramlich et al., 2018). The results revealed γ -values of 0.4-0.5 (complete sequence data) and 0.3-0.4 (uSNPs). These values are in accordance with recent studies which showed similar γ -values for modelled and empirical hybrid data (Blischak et al., 2018a; Zhang et al., 2019).

Categorization of SNPs Using the SNIploid Pipeline

SNIploid (Peralta et al., 2013) is a tool developed to analyze the SNP composition of recently established allotetraploid taxa emerging from diploid parental lineages. Originally, this approach is based on RNA sequencing data. By mapping the

sequencing reads of an allotetraploid individual to a reference (diploid parent 1) and comparing the observed SNPs and read depth with the results of diploid parent 2, conclusions on the SNP composition of the tetraploid species can be drawn. SNIploid is able to distinguish five different SNP categories (Peralta et al., 2013, **Figure 5A**). Category 1/2 = “inter-specific” SNPs describes that the observed allele in the polyploid is identical with only one of the parental species. Category 3/4 = “derived” SNPs is attributed, when the variation observed in the tetraploid is not identified between parental genomes. This hints to a mutation that occurred after the polyploidization event. Finally, category 5 corresponds to putative homeo-SNPs, i.e., the tetraploid is heterozygous for homeologous alleles of both parental genomes (**Figure 5A**). SNPs that do not fall into the categories mentioned above are categorized as “other”.

To use the advantages of this pipeline and to endeavour the SNP composition of the tetraploids in our dataset we adapted the SNIploid pipeline to RAD sequencing data (**Figure 1**). For our study, we used the concatenated consensus RAD loci of one putative diploid parental species as pseudoreference (similar as in Chen et al., 2013) for each test. The obtained RAD loci for every single individual (SAMPLE.consens file) were transferred into FASTA format. The indexing of the reference was done in BWA/0.7.12 (Li and Durbin, 2009). The sequence dictionary was created with PICARD/2.10.5 using the CreateSequenceDictionary tool (<http://broadinstitute.github.io/picard/>). The trimmed and filtered reads of the tetraploid species as well as of the second putative diploid parental species were mapped to the indexed reference applying the BWA/0.7.12 mapping tool using the MEM algorithm. The option -M (mark shorter split hits) for compatibility with PICARD was used, and 12 threads were applied for calculations (-t 12). The PICARD suite was used to sort the SAM file and to add read groups to the mapped alignment (‘AddOrReplaceRead Groups’). The resulting BAM file was indexed with SAMTOOLS v. 1.8 (Li et al., 2009) and used as input for the Genome Analysis Toolkit [GATK 3.8; McKenna et al. (2010)] to analyze the read depth (“DepthOfCoverage”) and the observed SNPs (“HaplotypeCaller”). For the tetraploid samples the—ploidy argument of HaplotypeCaller was set to 4. The generated read depth and the vcf files were eventually used as input for the executable perl script SNIploid.pl (<http://sniplay.southgreen.fr/cgi-bin/sniploid.cgi>). The results of each tested combination were summarized in a table and pie diagram, given the percentage of each SNP category, and were then compared to each other. Recently established allotetraploid hybrids should show a high proportion of homeo-SNPs, because they should contain half of the alleles of diploid “parent1” and the other half of diploid “parent2” and should therefore be heterozygous for most loci. With ongoing speciation, we assume a higher proportion of cat 3/4 SNPs that are specific for the polyploid and originated after the polyploidization event. Backcrossing with one or both parents might lead to an increase of interspecific SNPs (cat 1/2), but these shared SNPs might also be the result of incomplete lineage sorting. The remaining “other” SNPs contain all other SNP combinations that cannot be assigned to any of the defined categories. This includes, e.g., SNPs that are heterozygous in the parental taxa and SNPs that show more than



two alleles. They partly represent shared ancient polymorphisms of the whole group.

We initially narrowed down the number of combinations by choosing the most likely putative parental species/lineages according to the results of RAxML, SNAPP, NeighbourNet, and STRUCTURE. These parental combinations were tested with SNIploid (see **Table 1**).

RESULTS

Ploidy Determination

For 29 species of our sample set we could derive the ploidy level from chromosome counts available in the literature (**Supplement Table S1**). To determine the ploidy level of the remaining samples, we analyzed seven species via flow cytometry. Six of them, *S. integra*, *S. nummularia*, *S. pyrenaica*, *S. pyrolifolia*, *S. rehderiana*, and *S. schwerinii*, were analyzed for the first time. All analyzed

species were diploid (**Table S1**). Selected histograms are available in **Supplement Figure S1**.

Phylogenetic Relationships

The average number of raw reads obtained from the RAD sequencing was 8.15 million reads per sample. The length after removal of adapters and barcode was 86 bp. After filtering, an average of 8.08 (+/-5.81) million reads per sample were used for clustering. An average of 161,233 clusters per sample was generated with an average depth of 52 reads per cluster. After optimization, we used a data set that consists of 23,393 RAD loci shared by at least 40 individuals and contained 320,010 variable sites of which 191,615 are parsimony informative. The concatenated alignment had a length of 1,931,205 bp with 29.06% missing data (**Supplement Table S2**). The analysis with a clustering threshold of loci shared by at least 100 individuals (m100) revealed 1,660 loci with 16,761 SNPs (**Supplement Table S2**). The amount of missing data was only 9.1%, however, the dataset was not suitable to resolve the phylogenetic relationships (see **Supplement Figure S5**). Therefore, we used this dataset only for a genetic structure analysis with STRUCTURE, which is sensitive to missing data. The HyDe results of the m40 dataset revealed 4,724 significant hybridization events of 21,420 tested combination in the taxon-assigned test based on the uSNP data. Of these, 38.6% belonged to combinations that treated a polyploid taxon as “hybrid”. The average γ -value of all observed events was 0.52.

TABLE 1 | Allotetraploid species and their putative diploid parental species according to results of the phylogeny and the combination of NeighborNet, HyDe, STRUCTURE, and SNIploid analyses.

Polyploid	Parent 1	Parent 2
<i>Salix caesia</i> (4x)	<i>S. purpurea</i>	<i>S. repens</i>
<i>Salix cinerea</i> (4x)	<i>S. aurita</i>	<i>S. appendiculata</i>
<i>Salix laggeri</i> (4x)	<i>S. appendiculata</i>	<i>S. caprea</i>

The RAxML phylogeny included 133 accessions representing 35 species of the *Chamaetia/Vetrix* clade from Europe and Asia as well as four accessions of *S. triandra* as outgroup (**Figure 2**, **Supplement Figure S2**). The phylogeny revealed that all species are clearly monophyletic. The topology showed *S. reticulata* in sister position to four well-supported clades. For consistency, we used the same Roman numerals for the main clades as in Wagner et al. (2018). Clade I comprised ten species, including two well-supported subclades. Clade II contained all members of section *Vetrix* as well as *S. eleagnos* and *S. serpyllifolia*, clade III comprised all members of sections *Villosae* and *Vimen* as well as triploid *S. bicolor*, and clade IV contained *S. hastata*, *S. herbacea*, *S. pyrenaica* and *S. lanata*. The tetraploid species fell

into clades I and II, the triploid *S. bicolor* into clade III. The octoploid *S. glaucosericea* was in sister position to clade III and IV while hexaploid *S. glabra* was sister to *S. eleagnos* and *S. serpyllifolia* in clade II. Hexaploid *S. myrsinifolia* was in sister position to the remaining accessions of clade II. The genetic structure analysis of the complete sampling based on 1,655 uSNPs (m100) reflected the four observed main clades (I-IV) and indicated genetic admixture of up to four clusters for the high polyploids *S. glaucosericea* and *S. myrsinifolia*. Hexaploid *S. glabra* showed admixture of “clade II” and “clade IV” (**Supplement Figure S4**).

Each clade that contained one or more polyploid samples was analyzed separately with the minimum number of accessions



FIGURE 2 | Simplified RAxML phylogeny of 133 accessions representing 35 species of *Salix* subgenus *Chamaetia/Vetrix* based on 23,393 RAD loci. Four clades (numbering follows Wagner et al., 2018) are highlighted in different colours, subclades for detailed analyses are also designated. Ploidy level is indicated behind species name, polyploid samples are highlighted in bold. *Salix triandra* (subgenus *Salix*) was used to root the tree. BS values below branches, QS values of selected clades and subclades above branches. A detailed phylogeny including QS support values for all branches is supplied in **Supplement Figure S2**.

sharing a locus (m) dependent on the different overall number of accessions in each clade. These were m15 for clade Ia, m23 for clade II and m10 for clade III, respectively.

Relationships of Polyploids in Clade I

In the RAxML phylogeny, clade I was divided into two subclades (Figure 2). The supported monophyletic subclade Ia (BS 100, QS 0.42/0.073/0.99) comprised the species *S. purpurea*, *S. caesia*, *S. integra* which belong to section *Helix* sensu Skvortsov (1999), as well as *S. gracilistyla*. They were in sister position to the closely related species *S. repens* (incl. *S. repens* ssp. *rosmarinifolia*; sect. *Incubaceae* sensu Skvortsov (1999)). Subclade Ib comprised the dwarf shrub species *S. breviserrata* and *S. alpina*, as well as the medium-sized species *S. pyrolifolia* and *S. daphnoides*. The tetraploid *S. caesia* appeared in a well-supported sister position to *S. purpurea* (BS 100, QS 0.18/0/0.99). The analysis of the 18 accessions of subclade Ia with two accessions of *S. reticulata* as outgroup based on loci shared by at least 15 accessions, contained 38,603 shared RAD loci with 238,614 SNPs. The SNAPP (Figure 3A) and RAxML topology (Figure 3B) show strong statistical support for the sister relationship of *S. purpurea* and *S. caesia*. Using 37,867 uSNPs as input and designating *S. reticulata* as outgroup, HyDe performed a test of 2,448 combinations in the individual approach of which 164 were detected as significant hybridization events. Only six of 30 tested events in the assigned taxa approach were significant. HyDe showed no significant hybridization event with *S. caesia* as hybrid. Thus, we repeated the analysis with the complete sequence data of the clade specific analysis comprising 238,614 SNPs. Here, HyDe detected 606 significant results of 2,448 tested combinations in the individual approach with an average γ -value of 0.59. In 85 hybridization events (14%) *S. caesia* was counted as “hybrid”. For these events the average observed γ -value was 0.65. The average γ -value was 0.8 in 20 significant combinations with *S. purpurea* and *S. repens* as parental species. In the m15 dataset, the 18 accessions of clade I (without outgroup) shared 35,078 RAD loci comprising 191,211 SNPs. The genetic structure analysis based on 34,071 uSNPs for the most likely K-value of four revealed genetic admixture in *S. caesia* with about 50% of each the genetic partition of *S. purpurea* and *S. repens* (Figure 3D). In the NeighbourNet analysis (Figure 3C), *S. caesia* was situated between *S. purpurea* and *S. repens*. To test a hypothesis of allopolyploid origin of *S. caesia* in more detail, we conducted a SNIploid analysis using *S. purpurea* and *S. repens* as putative diploid parents of putative allopolyploid *S. caesia*. The results revealed 5.78% cat 1 and 12.65% cat 2 interspecific SNPs, 44.52% homeo-SNPs (cat 5), and 19.17% cat 3/4 SNPs. About 17.89% of SNPs did not fall into the given categories and were treated as “others” (Figure 5B).

Relationships of Polyploids in Clade II

Clade II contained hexaploid *S. myrsinifolia* and *S. glabra*, both monophyletic. The latter was in sister position to *S. eleagnos* and *S. serpyllifolia* forming subclade IIb. *Salix myrsinifolia* was in sister position to the remainders of clade II, with good BS but a skew to an alternative topology (BS 97, QS 0.28/0.93/0.99; Figure 2). All included members of section *Vetrix* s.l. [sensu Skvortsov

(1999)] formed a well-supported monophyletic group (subclade IIa; BS 100, QS 0.92/0/1, Figure 2) including two tetraploid species, *S. laggeri* and *S. cinerea*. The clade-specific analysis of subclade IIa included 25 accessions and was based on 33,515 RAD loci comprising 254,819 SNPs shared by at least 23 samples (m23). The species were all monophyletic and well supported (Figure 4B). One accession each of *S. eleagnos* and *S. serpyllifolia* served as outgroup. *Salix starkeana* was in sister position to the remaining accessions of clade IIa (BS100, 1/NA/1). The backbone of the tree topology showed some less supported branches, especially with regard to the QS values that indicate incongruencies of different branching topologies.

In the taxon-assigned approach, HyDe tested 104 combinations and yielded only four significant events. When we treated the samples as individuals, 5,313 combinations were tested and revealed 106 significant hybridization events. In 24 events (23%), *S. cinerea* was treated as ‘hybrid’. The average γ -value for these events was 0.46. For the parental combination of *S. appendiculata* and *S. aurita*, the observed nine significant hybrid combinations showed an average γ -value of 0.49. No significant hybrid event was observed for tetraploid *S. laggeri*.

The tetraploid species *S. cinerea* was in well-supported sister-position to *S. aurita* and *S. salviifolia* in both the complete and the clade specific RAxML phylogeny (Figures 2 and 4B), while *S. appendiculata* was sister to this clade. *Salix cinerea* was in sister position to *S. aurita* in the SNAPP species tree (Figure 4A) and in close relationship to *S. aurita* in the NeighbourNet analysis (Figure 4C). The genetic structure analysis of 39,746 uSNPs (without outgroup) for the most likely K-value of three revealed genetic admixture for all samples (Figure 4D). *Salix cinerea* showed a similar structure as *S. aurita* and *S. salviifolia*. The SNIploid results for *S. cinerea* using *S. aurita* and *S. appendiculata* as potential diploid parental species, showed about 5.6% homeo-SNPs and 5.6% cat 1 and 5.7% cat 2 SNPs, respectively, whereas about 47.5% were heterozygous sites sharing one allele of only one parent (cat 3/4 SNPs); 35.3 % of SNPs did not fall into the given categories (Figure 5D).

Tetraploid *S. laggeri* appeared in moderately supported sister position to *S. caprea* in both the clade-specific and in the overall RAxML phylogeny (Figures 2 and 4B). The NeighbourNet showed *S. laggeri* in close relationship to *S. appendiculata* and *S. caprea* (Figure 4C). The genetic structure analysis of subclade IIa for K=3 revealed a similar genetic composition as *S. appendiculata*, with some admixture of the *S. caprea* specific partition (Figure 4D). The SNIploid analysis with *S. caprea* and *S. appendiculata* as putative diploid parents revealed 5.6% homeoSNPs and about 12.2% interspecific SNPs (5.3% cat 1, 6.9% cat 2). About 36.6% of SNPs were shared with one allele of one parent (cat 3/4), while 45.9% of SNPs were not categorized (Figure 5C).

Relationships of Polyploids in Clade III

Nine species grouped into clade III, which is monophyletic and well supported (BS 98, QS 0.43/0/1). The clade-specific analysis was based on 37 accessions with a minimum of 30 individuals sharing a locus and revealed 16,463 RAD loci containing 97,576 SNPs. *Salix reticulata* was used to root the RAxML tree. As

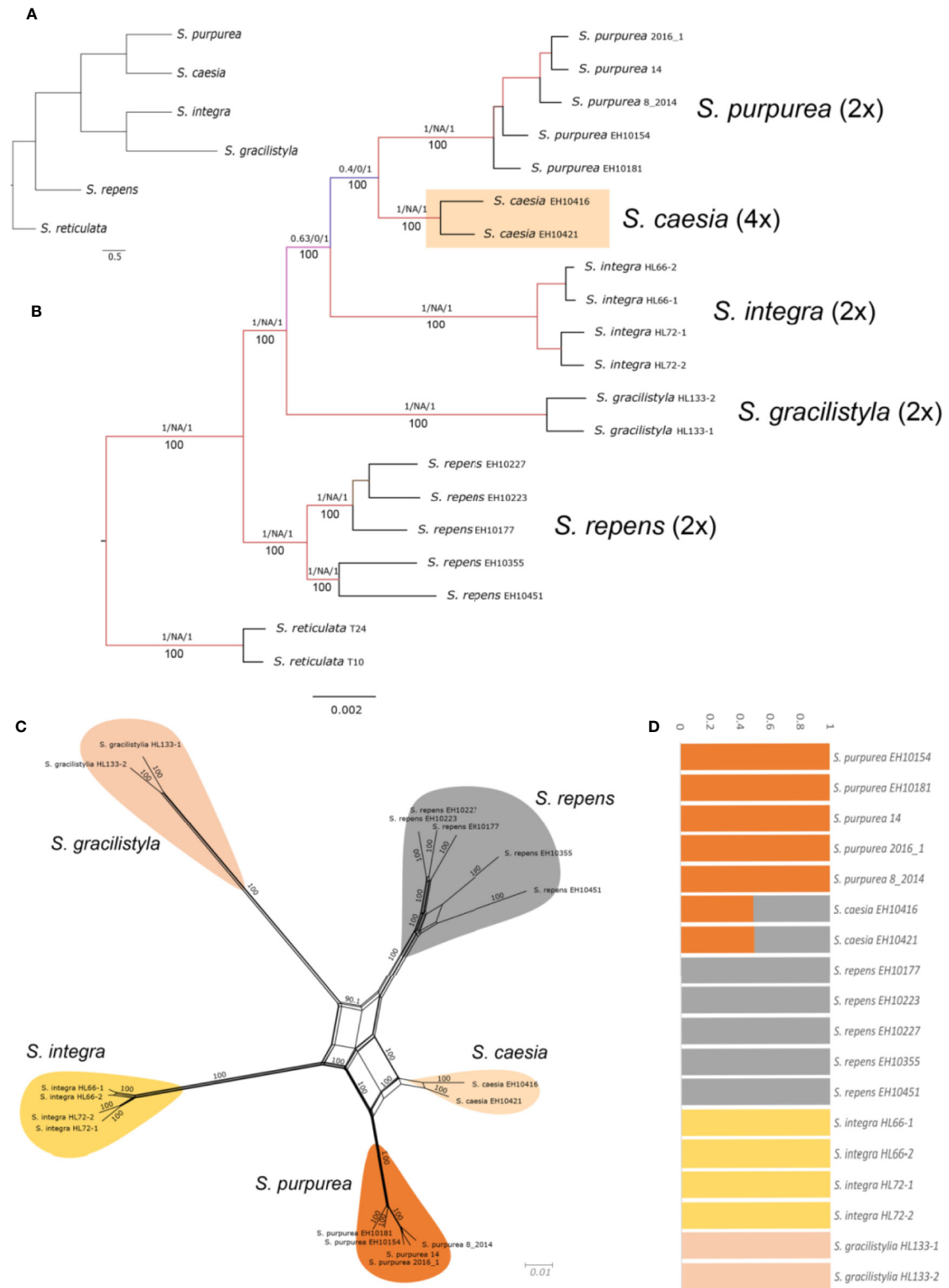
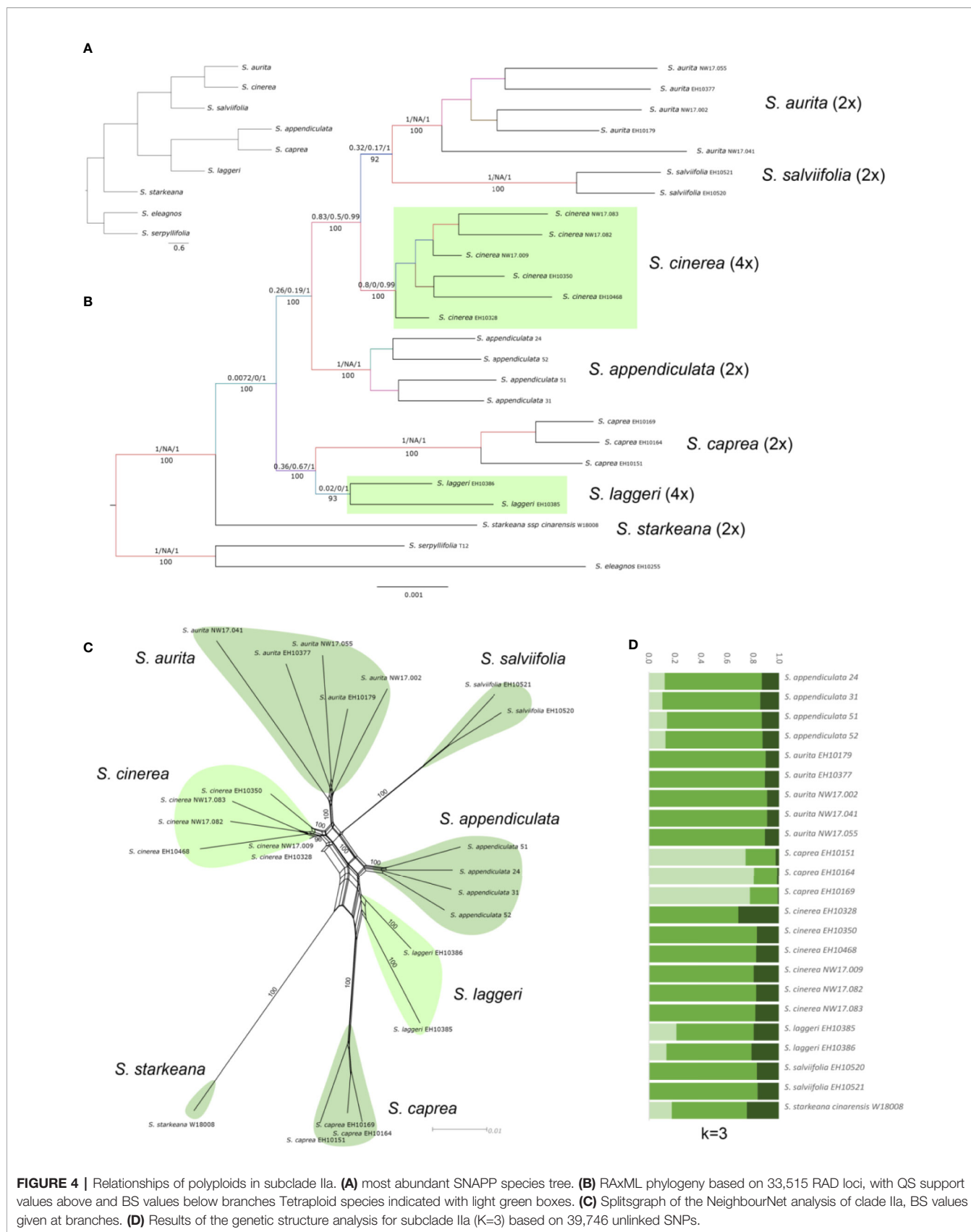
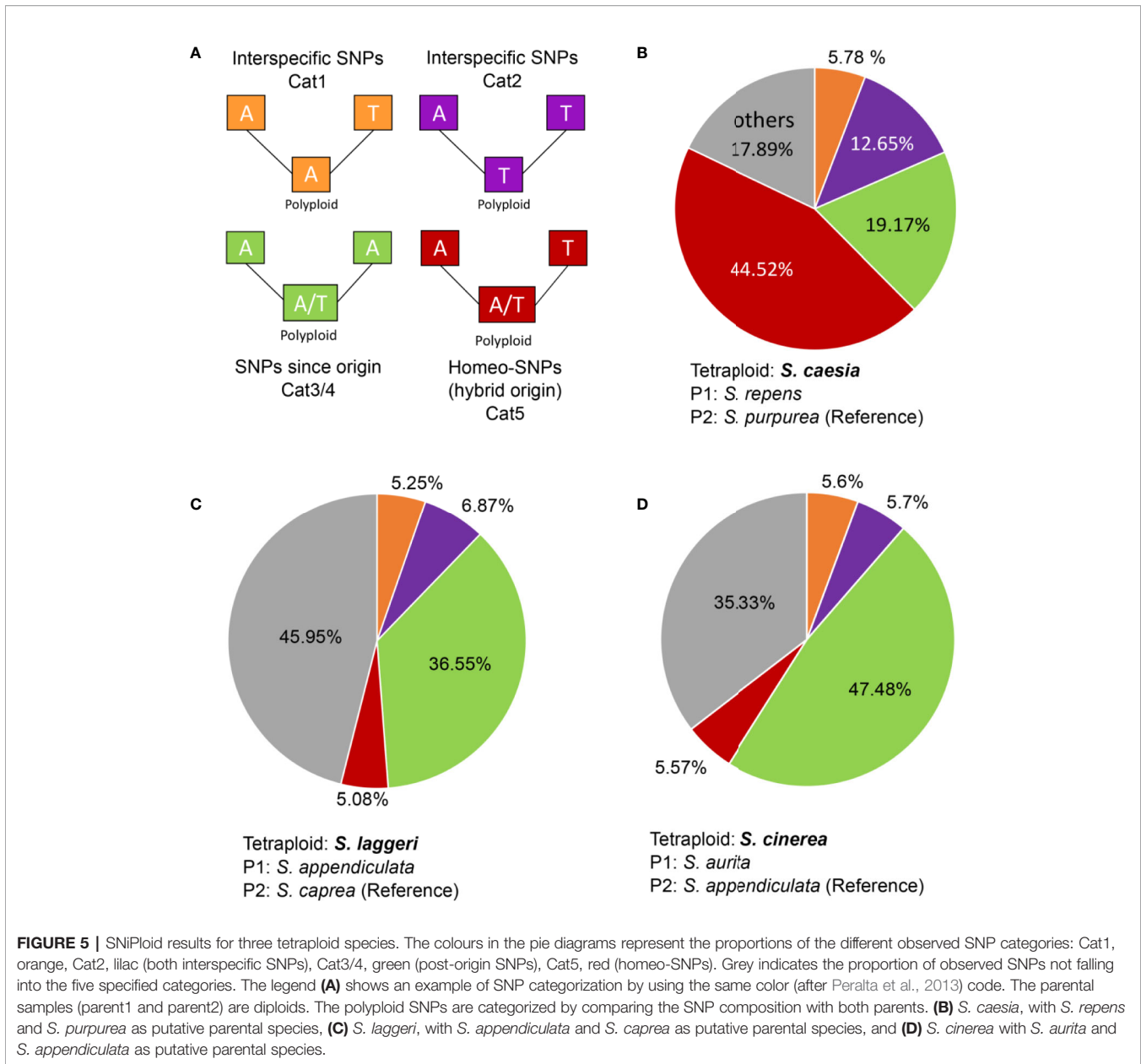


FIGURE 3 | Relationships of tetraploid *S. caesia* in clade Ia. **(A)** SNAPP species tree. **(B)** RAXML phylogeny based on 38,603 RAD loci and 238,614 SNPs, respectively. Bootstrap support values below, QS values above branches. **(C)** Resulting Splitsgraph of the NeighbourNet analysis of subclade Ia. Bootstrap values indicated at branches. **(D)** Genetic structure analysis for the subclade for the most likely value of K=4 based on 37,867 unlinked SNPs.





illustrated in **Figure S3**, *S. nummularia* was in sister position to all remaining members of this clade followed by *S. bicolor*, which is triploid and the only polyploid species in this clade. Two subclades diverged: subclade IIIa contained the species *S. helvetica*, *S. waldsteiniana*, *S. foetida*, and *S. lapponum*, while subclade IIIb consisted of *S. viminalis*, *S. schwerinii*, and *S. rehderiana*. Each species was clearly monophyletic, and morphological sister relationships were confirmed. In the clade-specific structure analysis for the most likely K-value of five (**Supplement Figure S3**), *S. bicolor* showed genetic admixture between *S. nummularia* and subclade IIIa. The amount of admixture differed between the two samples from the Eastern Alps in Austria (about two thirds shared with subclade IIIa) and the two samples from the Harz (Brocken) in

Germany (about half shared with subclade IIIa). The deep split between the two geographical regions is also reflected in the NeighbourNet network. SNIploid was not used to analyze the putative parenthood of *S. bicolor*, for the tool is optimized for tetraploid species (containing two subgenomes) and therefore not suitable for triploid species.

DISCUSSION

RAD Sequencing Data and Analysis Pipelines for Polyploids

Reduced representation libraries like RAD sequencing are frequently used to analyze intraspecific population structure,

closely related species groups and to infer phylogenetic relationships of diverged lineages based on high numbers of SNPs (e.g. Cariou et al., 2013; Andrews et al., 2016; Eaton et al., 2017). However, the use of RAD sequencing for polyploid species is still hampered by the lack of suitable tools and the statistical difficulties of dealing with more than two alleles [reviewed in Clevenger et al. (2015)]. Thus, only few studies using RAD sequencing on polyploid species are published so far (Mastretta-Yanes et al., 2014; Qi et al., 2015; Brandrud et al., 2017; Feng et al., 2018; Brandrud et al., 2020). Here we present a phylogenomic study on diploid and polyploid willow species by using a de-novo assembly of RAD sequencing data.

Our assembly and concatenation of short sequenced fragments representing the whole genome resulted in a robust, fully resolved phylogeny for the *Chamaetia/Vetrix* clade. The method of concatenation has been often criticized, especially in case of conflicting signal among genomic regions. However, Rivers et al. (2016) showed that the data output of the concatenated RAD loci including thousands of unlinked SNPs is able to recover a robust tree. Our findings support the suitability of RAD sequencing for interspecific phylogenetic inference and are in accordance with many studies on different levels of divergence (e.g. Hipp et al., 2014; Eaton et al., 2017; Wagner et al., 2018; He et al., 2020). The allelic information of the polyploid samples can be condensed to a single consensus sequence to circumvent the challenges of dealing with more than two alleles for phylogenetic approaches. However, the loss of information caused by this simplification might lead to wrong placement of the polyploid accessions in the phylogeny (Eriksson et al., 2018; Andermann et al., 2019). Our results showed that the RAD sequencing data contain enough information to resolve the phylogenetic relationships in *Salix* and that condensed allelic information was appropriate for the included allopolyploid species. While for phylogenetic approaches complete sequence information is preferable (Leaché et al., 2015), many unlinked loci are the preferable source of information for analyses, where the sites are treated as independently evolving genetic entities. Thus, we used the unlinked SNPs for network and genetic structure analyses. The uSNPs were also successfully used to reconstruct the species tree of each subclade using a coalescence-based method implemented in SNAPP (Bryant et al., 2012). Our results confirmed that the coalescent approach using RAD sequencing data is suitable for species delimitation, as recently shown by Brandrud et al. (2020).

HyDe (Blischak et al., 2018a) is a tool to detect hybridization and we showed that it is possible to use it with RAD sequencing data. We used HyDe to draw conclusions on the putative parental species of allopolyploids. First, we used the uSNP data, as suggested by Blischak et al. (2018a). The complete alignment using all SNPs, however, revealed a higher amount of significant results. Therewith we confirm the findings of Blischak et al. (2018a) that the number of significant hybridization events depends on the size of the input data and more input data reveal more accurate results. HyDe detected 4,724 significant hybridization events in the complete dataset.

About 40% of them were combinations including a polyploid as “hybrid”. That means that 60% were indicated as hybridization events between diploid samples. The high amount of natural hybridization in *Salix* is well known (Skvortsov, 1999; Argus, 2010), and homoploid hybridization even between distantly related species has been documented (Hardig et al., 2000; Gramlich et al., 2018). However, in this study we observe clearly distinct monophyletic groups for all included species and the sampled accessions represent morphologically definite individuals. We assume that HyDe uncovers a high amount of ancient gene flow that resulted in a shared polymorphism in the whole phylogeny and its subclades. We further suppose that non-visible introgression or, that the huge amount of input data cause significant number of “false” hybrid combinations.

SNiPloid (Peralta et al., 2013) was used to reconstruct the genomic constitution of the tetraploid species. SNiPloid was originally developed for transcriptomic data (Peralta et al., 2013). The application of this tool to RAD sequencing data provided us valuable information about the contributions of putative parental species to the tetraploid genome. In combination with genetic structure, HyDe and NeighborNet analyses, we could test for allopolyploid origin and revealed some insights into genome evolution after the polyploidization events. Unfortunately, the tool is only suitable for tetraploid species that consist of not more than two subgenomes (Peralta et al., 2013). However, for RAD sequencing data SNiPloid provides an alternative approach by using biallelic SNPs instead of sequence data to reveal potential parenthood of allotetraploids without previous knowledge. For the interpretation of the SNiPloid data it must be considered that in most cases, we did not test the “real” parental accessions. Our results are instead based on putative parental species/lineages, and we therefore expect a certain amount of SNPs that do not fall into the narrowly defined SNP categories. This will lead to an increase of proportions of non-categorized “other” SNPs. Furthermore, proportions of SNP categories must be interpreted with caution as also the intra-specific sampling can influence the actual amounts of detected SNPs for the respective categories. Nevertheless, we can present here a first overview of the origin and evolution of polyploid species within a phylogenetic framework. This increases the potential to analyze polyploid samples with reduced representation methods.

Phylogenetic Relationships and Origins of Polyploids in *Salix*

The RAD sequencing data revealed a well-resolved phylogeny of the *Chamaetia/Vetrix* clade including 35 Eurasian species plus *S. triandra* as outgroup and thus provide a sufficient framework for reconstructing the origin of the polyploid species. The presented phylogeny included members of all sections sensu Skvortsov (1999) and hence covered the morphological diversity and biogeographical range of Eurasian willows. All of the 36 included species are monophyletic. The isolated position of *S. reticulata* and the observed four clades are in accordance with a

former study on European diploid species by Wagner et al. (2018). The samples from Asia fall into the four clades.

About 40% of *Salix* species are polyploid (Suda and Argus, 1968) ranging from tetraploid to octoploid, rarely to decaploid. Our flow cytometry results showed a diploid level for the six included species from Asia that were previously not investigated (**Supplement Table 1, Supplement Figure S1**). Although some studies are published about tetraploid species in subgenus *Salix* s.l. (Triest et al., 2000; Triest, 2001; Barcaccia et al., 2014) no molecular studies existed so far on the origin of polyploid species of the *Chamaetia/Vetrix* clade. In this study we included seven polyploids from Europe with different ploidy levels to study different scenarios of their evolutionary origin. For the polyploid species included here, no diploid cytotypes have ever been reported, and they all exhibit very distinct morphologies and a genetic structure that is composed of several genetic clusters. These features make autopolyploid origins unlikely. The included polyploid species are all monophyletic. However, the species appear scattered over the phylogeny, indicating multiple independent origins of polyploids within the genus resulting from different parental combinations. Thereby *Salix* differs from plant genera in which a single allopolyploidization event resulted in post-origin adaptive radiation and speciation (e.g., the Hawaiian Silversword alliance, Seehausen, 2004). Willows rather resemble plant genera with repeated independent polyploidization events, like in *Nicotiana* (Chase et al., 2003), *Achillea* (Guo et al., 2013), *Ranunculus* (Baltisberger and Hörandl, 2016), and *Dactylorhiza* (Brandrud et al., 2020), among others.

The higher polyploid willows are in sister position to major clades or on basal branches. This could be explained by an old allopolyploid origin of these species. We would expect this pattern if two or more ancestral lineages of the extant major subclades (I-IV) had contributed to the polyploid genome. However, these ancestral lineages are probably not conspecific to any of the extant species. According to the crown group age of the *Chamaetia/Vetrix* clade (23.76 Ma, Wu et al., 2015), their origins may date back to the late Miocene/ early Pliocene. However, next to that, the hybrid nature of their origin might be responsible for the placement of the polyploids on basal branches of the subclades in the phylogeny. Our structure analyses indicated that the hexa- and octopolyploids, but also triploid *S. bicolor*, harbor two to four genetic partitions that characterize otherwise the main subclades I-IV (see **Figure S4**). Hence, we assume allopolyploid origins from divergent lineages. However, since the genetic structure analyses cannot disentangle ancient and recent hybridization or introgression events, further studies on the origin on the high polyploid willows are necessary.

Origin and Evolution of the Allotetraploid Species

The tetraploids fall into the observed clades (**Figure 2**). Since all included tetraploid species are monophyletic and do form well-supported entities in our analysis, we do not assume multiple origins with different parental combinations. Instead, we assume

a single or few hybrid (allopolyploid) origin(s) from a pool of related individuals (Abbott et al., 2013), followed by speciation. We are aware that extant species might not represent the true parental species, and that extinct taxa might have been also involved. Nevertheless, by testing different combinations of putative diploid parents that are the next extant relatives in our observed trees, we analyzed potential parental lineages that might have contributed to allotetraploid species formation. The comparison of the two respective subclades, however, suggests different evolutionary scenarios.

Clade Ia comprised species assigned to *Salix* sect. *Helix* sensu Skvortsov (1999), except for *S. gracilistyla*. However, all these species share the morphological character of connate filaments. The SNAPP species tree revealed a close relationship of *S. caesia* and *S. purpurea*, supporting the RAXML analyses. Based on our genetic structure analyses, tetraploid *S. caesia* combined the genetic partitions that occur in *S. purpurea* and *S. repens*, with about equal proportions, while the two Asian species *S. gracilistyla* and *S. integra* showed each a different genetic partition. The NeighbourNet analysis confirmed these findings and placed *S. caesia* between the *S. purpurea* and *S. repens* (**Figure 3C**). These two species are sympatric with *S. caesia* (Skvortsov, 1999). HyDe detected 14% significant hybridization events for *S. caesia* based on the complete sequence data. In case of a comparatively young hybridization event we would expect γ -values between 0.4 and 0.6. The high observed average γ -value of 0.8 for the combinations of *S. purpurea* and *S. repens* supports the hypothesis of an older event or other processes like incomplete lineage sorting. In so far, the HyDe analysis corroborates results of our SNiPloid analysis that revealed a high proportion of homeo-SNPs (44.52%) derived from the two parental species, whereas proportions of SNPs from post-origin interspecific events were much lower (18.53% for cat 1+2). However, the observed amounts of interspecific SNPs (cat1, cat2) may also be the result of incomplete lineage sorting. Thus, the observed results support the hypothesis of an allotetraploid origin from *S. purpurea* and *S. repens* s.l., with rather few post-origin backcrossing events. Indeed, no extant hybrids of *S. caesia* with the putative parental species have been reported so far, and hybridization with other species is extremely rare (Hörandl et al., 2012). The considerable proportion of post-origin SNPs (Cat 3/4, 19.17%) specific for the polyploid *S. caesia* indicates an independent evolution of the lineage. Occupation of niches at higher elevations may have contributed to reproductive isolation of *S. caesia* from the two parental species that occur in lowlands (Hörandl et al., 2012).

The tetraploid species *S. cinerea* and *S. laggeri* belong to subclade IIa that contain all included species of section *Vetrix* sensu Skvortsov (1999). The members of this section are medium sized shrubs with hairy leaves occurring from lowland to montane regions. They grow often in sympatry with other species in mixed populations and hybridize frequently, also with species of other sections (Hörandl et al., 2012). A big, shared gene pool is visible in high proportions of “other”, unassigned SNPs in SNiPloid. High ongoing gene flow may

explain that the clade-specific genetic structure analysis revealed no clear species-specific partitions within subclade IIa. However, both *Salix laggeri* and *S. cinerea* showed high proportions of species-specific SNPs (36.6% and 47.5%, respectively), indicating independent evolutionary lineages. The most likely parental combinations for their origin revealed in both cases around 5% of homeo-SNPs derived from hybrid origin, and also low proportions of post-origin hybridization with both parents, supporting allopolyploid origin. We cannot rule out that the true parental lineages were not included here – either due to extinction or to involvement of species that occur nowadays outside Europe (e.g., in adjacent Russia). A long evolutionary history with post-origin hybridization/ introgression events further affected the genomic composition of the hybrids. Interestingly, the two tetraploids differ strongly in the results of HyDe analyses: *S. cinerea* is responsible for 23% of significant hybridization events in the range of a more recent hybrid, while no significant hybrid event including *S. laggeri* was detected. Extant hybridization of *S. cinerea* with other co-occurring lowland species (*S. aurita*, *S. caprea*, *S. eleagnos*, *S. myrsinifolia*, *S. repens*, and a hybrid series with *S. viminalis*, Hörandl et al., 2012) may explain this result, while *S. laggeri* is a subalpine species that is reproductively better isolated from the other species and hybridizes only occasionally with the subalpine, sympatric *S. appendiculata* (Lautenschlager-Fleury and Lautenschlager-Fleury, 1986; Hörandl et al., 2012).

Our results on tetraploids confirmed an allopolyploid origin and a dynamic post-origin evolution of genomes, indicating speciation and evolution of independent lineages. Distribution ranges and ecological niches of the parental species, however, could have fluctuated from the origin of the clade onwards and may have caused various secondary contact hybridizations in different time periods of the Cenozoic (Hewitt, 2004; Hewitt, 2011). The relatively high age of the whole *Chamaetia/Vetrix* clade with a crown group age in the late Miocene, and the lack of a dated phylogeny makes it difficult to pinpoint hybridization events to certain geological time periods. According to extant hybridization patterns, isolation of the polyploid willows appears to be strongly influenced by the strength of habitat differentiation as discussed by Martini and Paiero (1988); Hörandl et al. (2012) and, Gramlich et al. (2016) for Central European species, and by Huang et al. (2015) on species from Taiwan. These findings support the notion that occupation of a separate niche is important for the establishment of a newly formed polyploid willow lineage.

CONCLUSIONS

Our data demonstrate that high-quality RAD sequencing data allow for the reconstruction of phylogenetic frameworks and give insights into origin and evolution of polyploid species. In willows, polyploidization appears to be predominantly connected to hybridization, i.e. to allopolyploid origin, as hypothesized by

Skvortsov (1999). Our data suggest that polyploids harbor considerable proportions of lineage-specific SNPs and managed to establish stable, self-standing evolutionary lineages after allopolyploid origin.

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/>, PRJNA433286.

AUTHOR CONTRIBUTIONS

NW and EH planned and designed research. NW, LH and EH conducted fieldwork. NW and LH performed experiments, NW analyzed data, and established new pipelines. NW and EH wrote the manuscript with contributions from LH.

FUNDING

This study was financially supported by the German research foundation DFG (Ho 5462/7-1 to EH), the National Natural Science Foundation of China (31800466 to LH), and the Natural Science Foundation of Fujian Province of China (2018J01613 to LH). LH was sponsored by China Scholarship Council for his research stay at the University of Goettingen (201707870015).

ACKNOWLEDGMENTS

We thank Jennifer Krüger and Susanne Gramlich for extracting the DNA and support in lab work. We also thank Susanne Gramlich as well as Andrea Danler, Katrin Scheufler, Claudia Pätzold, Jun Zhao, Wan-cheng Hou, and Gong-ru Zhou for collecting plant material and field assistance. We thank Ovidiu Paun for initial help with bioinformatic pipelines. We thank anonymous reviewers for their supportive comments.

This manuscript has been released as a pre-print at *BioRxiv*, (Wagner et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.01077/full#supplementary-material>

REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x
- Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., et al. (2019). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Syst. Biol.* 68, 32–46. doi: 10.1093/sysbio/syy039
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Argus, G. W. (1997). Infrageneric classification of *Salix* (Salicaceae) in the New World. *Syst. Bot. Monogr.* 52, 1–121. doi: 10.2307/25096638
- Argus, G. W. (2010). “*Salix*,” in *Flora of North America vol. 7 Magnoliophyta: Salicaceae to Brassicaceae*. Ed. E. Flora of North America Editorial Committee (New York: Oxford University Press), 23–51.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, 1–7. doi: 10.1371/journal.pone.0003376
- Baltisberger, M., and Hörandl, E. (2016). Karyotype evolution supports the molecular phylogeny in the genus *Ranunculus* (Ranunculaceae). *Perspect. Plant Ecol. Evol. Syst.* 18, 1–14. doi: 10.1016/j.ppees.2015.11.001
- Barcaccia, G., Meneghetti, S., Lucchin, M., and de Jong, H. (2014). Genetic segregation and genomic hybridization patterns support an allotetraploid structure and disomic inheritance for *Salix* species. *Diversity* 6, 633–651. doi: 10.3390/d6040633
- Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., and Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698
- Blischak, P. D., Chifman, J., Wolfe, A. D., and Kubatko, L. S. (2018a). HyDe: A python package for genome-scale hybridization detection. *Syst. Biol.* 67, 821–829. doi: 10.1093/sysbio/syy023
- Blischak, P. D., Kubatko, L. S., and Wolfe, A. D. (2018b). SNP genotyping and parameter estimation in polyploids using low-coverage sequencing data. *Bioinformatics* 34, 407–415. doi: 10.1093/bioinformatics/btx587
- Bouckaert, R., Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10 (4), e1003537. doi: 10.1371/journal.pcbi.1003537
- Brandrud, M. K., Paun, O., Lorenzo, M. T., Nordal, I., and Brysting, A. K. (2017). RADseq provides evidence for parallel ecotypic divergence in the autotetraploid *Cochlearia officinalis* in Northern Norway. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-05794-z
- Brandrud, M. K., Baar, J., Lorenzo, M. T., Athanasiadis, A., Bateman, R. M., Chase, M. W., et al. (2020). Phylogenomic relationships of diploids and the origins of allotetraploids in dactylohoriza (Orchidaceae). *Syst. Biol.* 69 (1), 91–109. doi: 10.1093/sysbio/sy035
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* 29, 1917–1932. doi: 10.1093/molbev/mss086
- Cariou, M., Duret, L., and Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol. Evol.* 3, 846–852. doi: 10.1002/ece3.512
- Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354
- Chase, M. W., Knapp, S., Cox, A. V., Clarkson, J. J., Butsko, Y., Joseph, J., et al. (2003). Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann. Bot.* 92, 107–127. doi: 10.1093/aob/mcg087
- Chen, J. H., Sun, H., Wen, J., and Yang, Y. P. (2010). Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. *Taxon* 59, 29–37.
- Chen, X., Li, X., Zhang, B., Xu, J., Wu, Z., Wang, B., et al. (2013). Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in allotetraploid *Brassica napus*. *BMC Genomics* 14, 346. doi: 10.1186/1471-2164-14-346
- Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., and Jackson, S. A. (2015). Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. *Mol. Plant* 8, 831–846. doi: 10.1016/j.molp.2015.02.002
- Clevenger, J. P., Korani, W., Ozias-Akins, P., and Jackson, S. (2018). Haplotype-Based Genotyping in Polyploids. *Front. Plant Sci.* 9, 1–6. doi: 10.3389/fpls.2018.00564
- Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711
- Cronk, Q., Ruzzier, E., Belyaeva, I., and Percy, D. (2015). *Salix* transect of Europe: latitudinal patterns in willow diversity from Greece to arctic Norway. *Biodivers. Data J.* 3, e6258. doi: 10.3897/BDJ.3.e6258
- Dorn, R. D. (1976). A synopsis of American *Salix*. *Can. J. Bot.* 54, 2769–2789. doi: 10.1139/b76-297
- Dufresne, F., Stift, M., Vergilino, R., and Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Mol. Ecol.* 23, 40–69. doi: 10.1111/mec.12581
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Eaton, D. A. R., and Overcast, I. (2020). iPYRAD: interactive assembly and analysis of RADseq data sets. *Bioinformatics* 36 (8), 2592–2594. doi: 10.1093/bioinformatics/btz966
- Eaton, D. A. R., Spriggs, E. L., Park, B., and Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Syst. Biol.* 66, 399–412. doi: 10.1093/sysbio/syw092
- 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. strasserii* (Fabaceae). *BMC Evol. Biol.* 18, 1–12. doi: 10.1186/s12862-018-1127-z
- Fang, C., Zhao, S., and Skvortsov, A. (1999). “Salicaceae,” in *Flora of China*, vol. 4. Eds. Z. Y. Wu and P. H. Raven (St. Louis: Science Press, Beijing & Missouri Botanical Garden Press), 139–274.
- Feng, J. Y., Li, M., Zhao, S., Zhang, C., Yang, S. T., Qiao, S., et al. (2018). Analysis of evolution and genetic diversity of sweetpotato and its related different ploidy wild species *I. trifida* using RAD-seq. *BMC Plant Biol.* 18, 1–12. doi: 10.1186/s12870-018-1399-x
- Gramlich, S., and Hörandl, E. (2016). Fitness of natural willow hybrids in a pioneer mosaic hybrid zone. *Ecol. Evol.* 6, 7645–7655. doi: 10.1002/ece3.2470
- Gramlich, S., Sagmeister, P., Dullinger, S., Hadacek, F., and Hörandl, E. (2016). Evolution in situ: Hybrid origin and establishment of willows (*Salix* L.) on alpine glacier forefields. *Heredity (Edinb.)* 116, 531–541. doi: 10.1038/hdy.2016.14
- Gramlich, S., Wagner, N. D., and Hörandl, E. (2018). RAD-seq reveals genetic structure of the F2-generation of natural willow hybrids (*Salix* L.) and a great potential for interspecific introgression. *BMC Plant Biol.* 18, 317. doi: 10.1186/s12870-018-1552-6
- Guo, Y. P., Tong, X. Y., Wang, L. W., and Vogl, C. (2013). A population genetic model to infer allotetraploid speciation and long-term evolution applied to two yarrow species. *New Phytol.* 199, 609–621. doi: 10.1111/nph.12262
- Guo, W., Hou, J., Yin, T., and Chen, Y. (2016). An analytical toolkit for polyploid willow discrimination. *Sci. Rep.* 6, 1–8. doi: 10.1038/srep37702
- Hörandl, E., Florineth, F., and Hadacek, F. (2012). *Weiden in Österreich und angrenzenden Gebieten (willows in Austria and adjacent regions)* (Vienna: University of Agriculture, Vienna).
- Hardig, T. M., Brunsfeld, S. J., Fritz, R. S., Morgan, M., and Orians, C. M. (2000). Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Mol. Ecol.* 9, 9–24. doi: 10.1046/j.1365-294X.2000.00757.x
- Hardig, T. M., Anttila, C. K., and Brunsfeld, S. J. (2010). A phylogenetic analysis of *Salix* (Salicaceae) Based on matK and ribosomal DNA sequence data. *J. Bot.* 2010, 1–12. doi: 10.1155/2010/197696
- He, L., Wagner, N. D., and Hörandl, E. (2020). RAD sequencing data reveal a radiation of willow species (*Salix* L., Salicaceae) in the Hengduan Mountains and adjacent areas. *J. Systematics Evol.* 00, 1–14. doi: 10.1111/jse.12593
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 183–195. doi: 10.1098/rstb.2003.1388
- Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139, 617–638. doi: 10.1007/s10709-011-9547-3

- Hipp, A. L., Eaton, D. A. R., Cavender-Bares, J., Fitzek, E., Nipper, R., and Manos, P. S. (2014). A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS One* 9 (7), e102272. doi: 10.1371/journal.pone.0093975
- Hirsch, C. N., and Buell, C. R. (2013). Tapping the promise of genomics in species with complex, nonmodel genomes. *Annu. Rev. Plant Biol.* 64, 89–110. doi: 10.1146/annurev-arplant-050312-120237
- Huang, C. L., Chang, C.-T., Huang, B. H., Chung, J.-D., Chen, J. H., Chiang, Y. C., et al. (2015). Genetic relationships and ecological divergence in *Salix* species and populations in Taiwan. *Tree Genet. Genomes* 11, 39. doi: 10.1007/s11295-015-0862-1
- 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
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chandrabali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Kadereit, J. W. (2015). The geography of hybrid speciation in plants. *Taxon* 64, 673–687. doi: 10.12705/644.1
- Lauron-Moreau, A., Pitre, F. E., Argus, G. W., Labrecque, M., and Brouillet, L. (2015). Phylogenetic relationships of American Willows (*Salix* L., Salicaceae). *PLoS One* 10, e0121965. doi: 10.1371/journal.pone.0121965
- Lautenschlager-Fleury, D., and Lautenschlager-Fleury, E. (1986). *Salix laggeri* Wimmer - Monographie einer wenig bekannten Weide. *Bauhimia* 8 (3), 149–156.
- Leaché, A. D., Banbury, B. L., Felsenstein, J., De Oca, A. N. M., and Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst. Biol.* 64, 1032–1047. doi: 10.1093/sysbio/syv053
- Leskinen, E., and Alström-Rapaport, C. (1999). Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: evidence from 5.8 S, ITS 1 and ITS 2 of the rDNA. *Plant Syst. Evol.* 215, 209–227. doi: 10.1007/BF00984656
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows – wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Martini, F., and Paiero, P. (1988). *I Salici d'Italia*. Ed. Lint: Trient.
- Mastretta-Yanes, A., Zamudio, S., Jorgensen, T. H., Arrigo, N., Alvarez, N., Pinero, D., et al. (2014). Gene duplication, population genomics, and species-level differentiation within a tropical mountain shrub. *Genome Biol. Evol.* 6, 2611–2624. doi: 10.1093/gbe/evu205
- Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., et al. (2011). Recently formed polyploid plants diversify at lower rates. *Science* (333), 1257. doi: 10.1126/science.1207205
- McBreen, K., and Lockhart, P. J. (2006). Reconstructing reticulate evolutionary histories of plants. *Trends Plant Sci.* 11, 398–404. doi: 10.1016/j.tplants.2006.06.004
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Meirmans, P. G., Liu, S., and Van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *J. Hered.* 109, 283–296. doi: 10.1093/jhered/esy006
- Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* 105, 385–403. doi: 10.1002/ajb.12016
- Peralta, M., Combes, M. C., Cenci, A., Lashermes, P., and Dereeper, A. (2013). SNIploid: A utility to exploit high-throughput SNP data derived from RNA-Seq in allopolyploid species. *Int. J. Plant Genomics* 2013, 890123. doi: 10.1155/2013/890123
- Percy, D. M., Argus, G. W., Cronk, Q. C., Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., et al. (2014). Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? *Mol. Ecol.* 23, 4737–4756. doi: 10.1111/mec.12837
- Qi, Z. C., Yu, Y., Liu, X., Pais, A., Ranney, T., Whetten, R., et al. (2015). Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RAD-tag based GBS - insights into species origin and effects of software pipelines. *J. Syst. Evol.* 53, 432–447. doi: 10.1111/jse.12176
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032
- Rambaut, A. (2014). FigTree v1.4.2, A graphical viewer of phylogenetic trees. Available from: <https://github.com/rambaut/figtree/releases>.
- Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467
- Rechinger, K. H., and Akeroyd, J. R. (1993). “Salicaceae,” in *Flora Europaea*, vol. 1. Eds. T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmondson, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb (Cambridge: Cambridge University Press), 53–66.
- Rivers, D. M., Darwell, C. T., and Althoff, D. M. (2016). Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. *Cladistics* 32, 672–681. doi: 10.1111/cla.12149
- Savage, J. A., and Cavender-Bares, J. (2012). Habitat specialization and the role of trait lability in structuring diverse willow (genus *Salix*) communities. *Ecology* 93, 138–150. doi: 10.1890/11-0406.1
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi: 10.1016/j.tree.2004.01.003
- Skvortsov, A. K. (1999). *Willows of Russia and adjacent countries. Taxonomical and geographical revision* (Joensuu, Finland: University of Joensuu).
- Soltis, P. S., and Soltis, D. E. (2009). The Role of Hybridization in Plant Speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Suda, Y., and Argus, G. W. (1968). Chromosome numbers of some North American *Salix*. *Brittonia* 20, 191–197. doi: 10.2307/2805440
- Suda, J., and Trávníček, P. (2006). Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. *Curr. Protoc. Cytom.* 38, 7.30.1–7.30.14. doi: 10.1002/0471142956.cy073038
- Triest, L., De Greef, B., De Bondt, R., and Van Slycken, J. (2000). RAPD of controlled crosses and clones from the field suggests that hybrids are rare in the *Salix alba*-*Salix fragilis* complex. *Heredity (Edinb.)* 84, 555–563. doi: 10.1046/j.1365-2540.2000.00712.x
- Triest, L. (2001). Hybridization in staminate and pistillate *Salix alba* and *S. fragilis* (Salicaceae): Morphology versus RAPDs. *Plant Syst. Evol.* 226, 143–154. doi: 10.1007/s006060170062
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* (313), 1596–1604. doi: 10.1126/science.1128691
- Wagner, N. D., Gramlich, S., and Hörandl, E. (2018). RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. *Ecol. Evol.* 8, 8243–8255. doi: 10.1002/ece3.4360
- Wagner, N. D., He, L., and Hörandl, E. (2019). Relationships and genome evolution of polyploid *Salix* species revealed by RAD sequencing data. Preprint at BioRxiv. doi: 10.1101/864504
- Wichura, M. (1865). *Die Bastardbefruchtung im Pflanzenreich erläutert an den Bastarden der Weiden* (Breslau: Morgenstern).
- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879. doi: 10.1073/pnas.0811575106
- Wu, J., Nyman, T., Wang, D.-C., Argus, G. W., Yang, Y.-P., and Chen, J.-H. (2015). Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evol. Biol.* 15, 31. doi: 10.1186/s12862-015-0311-7
- Zhang, B.-W., Xu, L.-L., Li, N., Yan, P.-C., Jiang, X.-H., Woeste, K. E., et al. (2019). Phylogenomics reveals an ancient hybrid origin of the Persian walnut. *Mol. Biol. Evol.* 36 (11), 2451–2461. doi: 10.1093/molbev/msz112

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.

Copyright © 2020 Wagner, He and Hörandl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.