Molecular Phylogenetics and Evolution 199 (2024) 108161

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic insights into the Salicaceae: The evolution of willows and beyond

Ezgi Ogutcen [a,b,*], Paola de Lima Ferreira [a,c], Natascha D. Wagner [d], Pia Marinček [d], Jing Vir Leong [a,e], Gibson Aubona [a,e], Jeannine Cavender-Bares [f], Jan Michálek [a,g], Lucy Schroeder [f], Brian E. Sedio [h,i], Radim J. Vašut [j,k], Martin Volf [a,e]

[a] Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic
[b] Department of Environment and Biodiversity, Paris Lodron University of Salzburg, Salzburg, Austria
[c] Department of Biology, Aarhus University, Aarhus, Denmark
[d] Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, Göttingen, Germany
[e] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic
[f] College of Biological Sciences, University of Minnesota, St. Paul, MN, United States
[g] Institute of Microbiology of the Czech Academy of Sciences, Centre Algatech, Trebon, Czech Republic
[h] Department of Integrative Biology, University of Texas at Austin, Austin, TX, United States
[i] Smithsonian Tropical Research Institute, Apartado, 0843–03092 Balboa, Ancón, Republic of Panama
[j] Department of Botany, Faculty of Science, Palacký University Olomouc, Olomouc, Czech Republic
[k] Department of Biology, Faculty of Education, Palacký University Olomouc, Olomouc, Czech Republic

ARTICLE INFO

Keywords:
Angiosperms353
Divergence time estimation
Phylogeny
RAD-Seq
Salix
Whole-genome sequencing

ABSTRACT

The Salicaceae includes approximately 54 genera and over 1,400 species with a cosmopolitan distribution. Members of the family are well-known for their diverse secondary plant metabolites, and they play crucial roles in tropical and temperate forest ecosystems. Phylogenetic reconstruction of the Salicaceae has been historically challenging due to the limitations of molecular markers and the extensive history of hybridization and polyploidy within the family. Our study employs whole-genome sequencing of 74 species to generate an extensive phylogeny of the Salicaceae. We generated two RAD-Seq enriched whole-genome sequence datasets and extracted two additional gene sets corresponding to the universal Angiosperms353 and Salicaceae-specific targeted-capture arrays. We reconstructed maximum likelihood-based molecular phylogenies using supermatrix and coalescent-based supertree approaches. Our fossil-calibrated phylogeny estimates that the Salicaceae originated around 128 million years ago and unravels the complex taxonomic relationships within the family. Our findings confirm the non-monophyly of the subgenus *Salix* s.l. and further support the merging of subgenera *Chamaetia* and *Vetrix*, both of which exhibit intricate patterns within and among different sections. Overall, our study not only enhances our understanding of the evolution of the Salicaceae, but also provides valuable insights into the complex relationships within the family.

1. Introduction

The members of the Salicaceae are trees or shrubs that are distributed in a wide range of tropical and temperate habitats (Stevens, 2001 onwards). Many Salicaceae species have important ecological, economical, and scientific roles. They provide vital habitat and food sources for diverse wildlife, making them essential components of natural and cultivated forest communities worldwide (Skvortsov, 1999; Narango et al., 2020). Many members of the family produce a variety of secondary plant metabolites that mediate complex plant-herbivore interactions (Boeckler et al., 2011). Notably, acetylsalicylic acid (aspirin), one of the most widely used medications, is originally derived from the bark of willow trees. Certain Salicaceae trees hold economic importance as sources of timber, pulp, and bioenergy production, contributing to their significance as forest trees (Tognetti et al., 2013). Due to their small genome size, rapid growth, feasibility of conducting controlled crosses and ability to clonally propagate from stems, some *Populus* and *Salix* species have become model organisms for genomic, biochemical,

* Corresponding author at: Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic.
E-mail address: ezgi.ogutcen@plus.ac.at (E. Ogutcen).

https://doi.org/10.1016/j.ympev.2024.108161
Received 1 March 2024; Received in revised form 5 July 2024; Accepted 24 July 2024
Available online 28 July 2024
1055-7903/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

quantitative genetics and ecophysiological research (Bradshaw et al., 2000, Cronk 2005; Jansson & Douglas, 2007).

Traditionally, Salicaceae s.s. included two broad genera, *Salix* and *Populus*, encompassing willows, poplars, aspens, and cottonwoods (Chase et al., 2002). Fossil records indicate that Salicaceae s.s. dates back to the Late Paleocene to Early Eocene in North America and the Late Eocene to Middle Oligocene in Europe (Collinson, 1992; Boucher et al., 2003). In the early 21st century, the circumscription of the family was expanded to include a significant portion of the former Flacourtiaceae, including, among others, the type genus *Flacourtia* and the large tropical genus *Casearia* (Chase et al., 2002; APG II, 2003; Alford, 2005). Currently the Salicaceae s.l. comprises approximately 54 genera with over 1,400 species (Stevens, 2001 onwards; POWO, 2023), and the evolutionary timeline of the family needs an update. The Salicaceae family has a cosmopolitan distribution, ranging from the tropics to polar regions. While *Salix* and *Populus* exhibit notable diversity in temperate regions of the Northern Hemisphere, the genera that were previously in the former family Flacourtiaceae (e.g., *Casearia, Homalium, Xylosma*) display high Pantropical diversity (Stevens, 2001 onwards; de Mestier et al., 2022).

Previous research on the phylogeny of Salicaceae relied on small sets of molecular and morphological markers, yielding conflicting topologies with limited resolution (Meeuse, 1975; Leskinen & Alstrom-Rapaport, 1999; Liu et al., 2016). Due to these limitations, earlier studies failed to offer a clear understanding of the familial relationships within the Salicaceae. High throughput sequencing provides a significantly greater resolution than traditional molecular markers and allows for a more comprehensive analysis of evolutionary relationships. Currently available genome-based Salicaceae phylogenies often focus on a small subset of taxa, primarily a single genus or its subclade (Zhang et al., 2018; Zong et al., 2019; Sanderson et al., 2023; Wagner et al., 2020; Volf et al., 2023; but also see Li et al., 2019). While the phylogenetic relationships within some large genera, such as *Salix*, *Populus* and *Casearia* have been repeatedly tackled, most of the remaining genera are understudied, and the overall family phylogeny and the evolutionary timeline of its diversification remains unresolved. Consequently, an updated family-level Salicaceae phylogeny is a necessary step towards a more accurate and up-to-date understanding of the relationships within the family.

As the largest genus of the family, *Salix* with its c. 450 species has a notoriously difficult taxonomy to resolve (Argus, 2010). *Salix* exhibits dioecious reproduction, frequent hybridization, high polyploidy rates, and significant intraspecific phenotypic variation (Hörandl et al., 2012; Gramlich et al., 2016; Wagner et al., 2021a). These characteristics are also present in other Salicaceae genera, making the study of the overall family phylogeny a complex task. However, due to their ecological importance, Salicaceae serve as essential models for studying plant evolution (Hardig et al., 2000; Jansson & Douglas, 2007; Volf et al., 2023). The temperate members of the family stand out as diverse woody plant lineages in colder regions of the Northern Hemisphere, making them key components of ecological studies focused on plant chemical and community ecology (Nyman & Julkunen-Tiitto, 2005; Savage-Cavender-Bares, 2012; Volf et al., 2022). Resolving the Salicaceae phylogeny may thus not only inform us on the evolution of this lineage but also provide an important tool for ecological studies.

This study aims to determine the phylogenetic relationships within the Salicaceae using RAD-Seq enriched whole-genome sequencing (WGS) and targeted-capture with sampling efforts focusing mainly on the largest genus *Salix,* but also the lesser-studied genera across the family. The resulting time-calibrated phylogeny will serve as a strong backbone for studying the evolution and diversity of the Salicaceae. Unlike many plant families, which are more diverse in the tropics, Salicaceae show comparable richness in temperate regions. Our findings will facilitate comparative studies investigating the mechanisms contributing to high rates of diversification, allowing us to explore potential differences between temperate and tropical ecosystems. As Salicaceae members play vital roles in plant-herbivore interactions globally, a comprehensive phylogenetic history of this family will also enhance our understanding of the evolution of these complex relationships.

2. Materials and methods

2.1. Sampling

Plant materials were collected from the field in various locations (Supplementary Table S1). The species identification in the field was performed using reference books and field guides (Smith, 2008; Hörandl et al., 2012; Vašut et al., 2013), and they were further confirmed by comparing the vouchers (photographs available upon request) to previously verified specimens deposited at the herbaria of University of Göttingen, University of South Bohemia, and University of Minnesota. Field sampling included 65 species, out of which five species from Achariaceae were selected as outgroups. The remaining 60 species included both temperate and tropical Salicaceae, and it mainly focused on the largest genus of the family, *Salix*. In order to enrich our taxonomic sampling, we also included WGS data for 9 species from the Tree of Life database (https://treeoflife.kew.org/tree-of-life). As a result, a total of 74 species were sampled for the study.

2.2. DNA extraction and sequencing

Shortly after being collected in the field, fresh leaves were either stored in silica gel or liquid nitrogen and freeze-dried later. Dried leaves were then transported in 96 % ethanol and stored at −20 °C in order to remove phenolic compounds. Prior to DNA extraction, ethanol-treated leaves were left to dry at room temperature for 30 min. Dried leaf materials were submerged in liquid nitrogen and homogenized using mortar and pestle. Genomic DNA was extracted using Qiagen DNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions with three modifications: i) we supplemented the Buffer AP1 with 1 % RNase A (Qiagen, Hilden, Germany), 5 % proteinase K (Qiagen, Hilden, Germany), and 2 % (w/v) polyvinylpolypyrrolidone (PVPP; Fluka Analytical Sigma Aldrich, St. Louis, MO, USA); ii) AP1 buffer incubation was increased to 16 h; and iii) DNA was eluted in 100 µl EB buffer instead of 200 µl. DNA quantification was performed using a Qubit 3.0 fluorometer with a dsDNA HS Assay kit (Invitrogen Thermo Fisher Scientific Life Technologies, Eugene, OR, USA). 20 µl of each sample with a minimum of 30 ng/µl DNA concentration was sent to Novogene (Kowloon, Hong Kong) for whole-genome library preparation (350 bp) and Illumina Sequencing PE150 (10G raw data per sample).

2.3. Rad-seq enriched whole-genome sequence data processing

Demultiplexing, adapter removal, and quality trimming steps of the raw sequences were completed by Novogene Sequencing Centre (Cambridge, UK). The bioinformatic analyses were performed using the Sequence Capture Processor pipeline (SECAPR v2.0.2; Andermann et al., 2018) following the process described in Volf et al. (2023) with some modifications. We checked the quality of the clean reads using FastQC v0.11.8 (Andrews, 2010). Clean reads were de novo assembled into contigs using SPAdes (Bankevich et al., 2012) with an automated k-mer selection and discarding contigs shorter than 50 bp. In order to identify putative regions with enough variation, we generated a consensus sequence for each of the 23,393 loci recovered in a RADseq study (Wagner et al. 2020) using Geneious version v2023.0.1 (Kearse et al., 2012). Then, assembled contigs were mapped against the 23,393 consensus sequences using default parameters in LASTZ (Harris, 2007). Mapped contigs were aligned using MAFFT v7.487 (Katoh & Standley, 2013) and trimmed using Gblocks v0.91b (Castresana, 2000) with default parameters. After removing the alignments with less than three samples and with 100 % identical sites, we retained 17,568 loci. Our mapping strategy was based on short reference loci, with an average length of 80 bp, which can pose challenges during the mapping step. To

address this, we used the reference-based assembly function in SECAPR to generate consensus sequences for each locus from the 17,568 multiple sequence alignments generated in the previous step. This new set of consensus sequences is now recovered with an average locus length of 488 bp, which provides larger numbers of phylogenetically informative sites. This updated reference library was then used as sequence targets and mapped against the clean reads using BWA v0.7.17 (Li & Durbin 2009), and aligned using MAFFT. This final reference assembly step generated 11,378 loci, out of which 10,867 produced alignments with at least three samples.

In order to compare the phylogenetic informativeness of trimmed versus non-trimmed alignments, we generated two RAD-Seq enriched WGS datasets (referred to as ReWGS hereafter): ReWGS I was maintained without any trimming, and ReWGS II was trimmed using Gblocks with default parameters. After this step, both datasets were imported to Geneious and filtered based on four parameters: i) loci with sequences for less than 50 % of the samples; ii) loci with more than 95 % sequence identity); iii) loci that are shorter than 100 bp and longer than 2000 bp; and iv) loci with more than 50 % of samples showing potential paralogs as detected by the "paralog retriever" function in HybPiper 2.0 (Johnson et al., 2016).

2.4. Targeted-capture sequencing data processing

We tested the phylogenetic performance of two different targeted-capture gene sets: 353 genes from the Universal Angiosperm bait kit (Johnson et al., 2018; called ANG353 from now on), and 1219 genes from the Salicaceae-specific bait kit (Sanderson et al., 2020; called SAL1219 from now on). After the quality-control step described above, we used HybPiper (Johnson et al., 2016) to extract the two targeted gene sets from the raw whole-genome sequences. The pipeline first maps the raw reads onto the reference target sequences using BWA, then assembles the mapped reads into contigs using SPAdes, and lastly it aligns the assembled contigs using Exonerate (Slater and Birney, 2005). The resulting alignments for both gene sets were filtered out based on the following parameters: Alignments with i) less than 50 % reference gene length; ii) less than 50 % sample coverage; and iii) more than 25 % samples with paralog warnings. In addition, samples with less than 25 % gene coverage were also removed from the sample set.

2.5. Phylogenetic analyses

Maximum-likelihood (ML)-based phylogenetic reconstructions were performed using two main approaches: the supermatrix and the coalescent-based supertree. These approaches were applied to four sequence datasets (reWGS I, reWGS II, ANG353, and SAL1219), resulting in a total of eight phylogenetic trees.

For the supermatrix approach, the aligned and filtered gene sets were concatenated using AMAS (Borowiec, 2016). In order to determine the best partitioning scheme with the best substitution model for each partition, the MF+MERGE function in IQ-TREE v2.2.0 (Minh et al., 2020) was used. This function uses an ultrafast and automated model selection tool ModelFinder (Kalyaanamoorthy et al., 2017), which finds the best-fit partitioning scheme based on the Bayesian information criterion (BIC) scores. The concatenated alignments were run using the best fit model followed by an ultrafast bootstrap with 1000 replicates (Chernomor et al., 2016, Hoang et al., 2018).

For the supertree approach, each gene tree was generated using the MFP function in IQ-TREE v2.2.0. This function performs the model selection described above and then reconstructs a phylogenetic tree using the best-fit model followed by ultrafast bootstrap with 1000 replicates. In order to estimate the species tree from the reconstructed gene trees in the previous step, ASTRAL v5.6.3 (Mirarab et al., 2014) was used as a coalescent approach. ASTRAL was run with default parameters to generate a species tree and to compute full scoring with quartet supports, which measure the level of congruence among the gene trees. The quartet scores were added onto the tree using a custom script (https://github.com/sidonieB/scripts/blob/master/plot_Astral_trees.R) in R v4.3.1 (R Core Team, 2023).

2.6. Divergence time estimation

To estimate divergence times, we inferred a time-calibrated tree based on the ANG353 supermatrix dataset (see Johnson et al., 2018) with BEAST v.2.5.1 (Bouckaert et al., 2019). The ANG353 dataset was selected due to its smaller size, as the computational intensity and the long runtime of the analysis made it impractical to use larger datasets.

We set four uniform priors using three fossils and secondary calibration information from a recently published angiosperm phylogeny (Zuntini et al., 2024) that utilized a well curated AngioCal v1.1 fossil calibration dataset (modified from Ramírez-Barahona et al., 2020) and a sampling of almost 8,000 plant genera. For our secondary calibrations, we opted to use the estimated ages of the "young tree" from Zuntini et al. (2024), which is more congruent with other *Salicaceae* studies (de Mestier et al., 2022; Liu et al., 2022). The four uniform priors were set as follows: i) the crown Salicoideae node was calibrated from a minimum of 48.5 mya (estimated age of †*Populus tidwelli* fossil from Manchester et al., 2006) to a maximum of 108.3 mya ("young tree" stem Salicaceae age from Zuntini et al., 2024); ii) the stem of the genus *Populus* was calibrated from a minimum of 48.5 mya (estimated age of †*Populus wilmattae* fossil from Manchester et al., 1986) to a maximum of 95.5 mya ("young tree" crown Salicaceae age from Zuntini et al., 2024); iii) the root was calibrated from a minimum of 48.5 mya (the estimated age of the crown Salicoidae fossil; see above) to a maximum of 136.3 mya ("young tree" crown Malphigiales age from Zuntini et al., 2024); and iv) the crown node of the genus *Casearia* was calibrated from a minimum of 37 mya (estimated age of a pollen fossil described in Graham (1985) and utilized in de Mestier et al., 2022) to a maximum of 95.5 mya ("young tree" crown Salicaceae age from Zuntini et al., 2024).

The BEAST analysis was run with the optimized relaxed molecular clock approach (Douglas et al., 2021) using a GTR+Γ substitution model with four rate categories and a Yule model prior on speciation. Posterior distributions of parameters were estimated using a MCMC analysis of 30,000,000 generations with a 25 % burn-in. Two replicates of the BEAST analysis were run with different seed values and the log files were examined with Tracer v. 1.7.2 (Rambaut et al., 2018). The trees of the two replicate runs were combined with LogCombiner and then used to generate a maximum clade credibility tree with median heights in TreeAnnotator v. 2.5.1 (Drummond & Rambaut, 2007).

We performed an additional time divergence estimation analysis on all eight of our phylogenetic trees using a penalized-likelihood (PL) approach (Sanderson, 2002) implemented in R package "ape" (Paradis & Schliep, 2019) with the "chronos" function. The PL analysis with 100 iterations was performed using a relaxed clock model with the same time calibration approach used in BEAST (see above). The smoothing parameter lambda was set to 0 after initial testing of values between 0 and 10. The ultrametric trees resulting from all the analyses described above were displayed and edited using FigTree v1.4.3 (Rambaut, 2014).

3. Results and discussion

3.1. Rad-seq enriched WGS versus targeted-capture datasets

We acquired de novo sequencing for 66 samples, and after quality trimming, we recovered an average of 44.26 ± 2.04 (95 % confidence interval) million reads per sample. All de novo sequences have been deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA1082499.

We attempted to further improve the RAD-Seq enriched WGS data by removing regions with poor alignments and divergent sequences that could potentially introduce phylogenetic noise (Talavera & Castresana, 2007). Based on two different filtering schemes (see Materials and

Methods for details), we generated two alignment datasets. ReWGS I and ReWGS II recovered 6,726 and 2,284 loci for an average of 52.47 and 51.50 samples, respectively (Table 1). While ReWGS I has an average locus length of 511.50 bp, ReWGS II has shorter loci with an average of 179.72 bp. Total alignment lengths for ReWGS I and II are 3,440,351 and 428,447 bp, respectively. Both datasets have comparable percentages of parsimony informative sites per locus (5.04 % for ReWGS I and 6.70 % for ReWGS II; Fig. 1), but the percentage of alignment gaps is much larger in ReWGS I (26.34 %) than ReWGS II (1.98 %). While quality trimming reduced the presence of gappy regions and slightly increased the percentage of parsimony informative sites, it also reduced the number of loci and locus length, thereby decreasing the phylogenetic utility of ReWGS II (Table 1 and Fig. 1). As a result, the phylogenies generated using ReWGS II have lower node support values and incongruent topologies.

We also wanted to demonstrate the utility of two different targeted-capture datasets (ANG353 and SAL1219) and to compare these gene-focused methods with RAD-Seq enriched WGS in generating phylogenies. While SAL1219 is designed specifically for Salicaceae-focused phylogenetic studies (Sanderson et al., 2020), ANG353 is a universal bait set designed to be used across all angiosperms (Johnson et al., 2018). After the filtering steps (see Materials and Methods for details), the final locus counts were 298 and 1,111 for ANG353 and SAL1219 datasets, respectively. Compared to the RAD-Seq enriched WGS datasets, both targeted-capture datasets have longer loci that are recovered in more samples: ANG353 loci have an average length of 845.42 bp in an average of 62.42 samples, and SAL1219 loci have an average length of 2,394.52 bp in an average of 58.28 samples (Table 1). The loci in both datasets also have more parsimony informative sites per locus (19.30 % for ANG353 and 23.31 % for SAL1219) than both ReWGS datasets (Fig. 1). The percentage of alignment gaps is 24.15 % for ANG353 and 12.60 % for SAL1219.

Overall, ANG353 and SAL1219 targeted-capture loci showed similar performances throughout the whole sample set (Fig. 2). On average, 84.15 % of the ANG353 loci and 78.45 % of the SAL1219 loci were recovered in our samples. Compared to the reference locus lengths, the average length of these recovered loci is 66.82 % for ANG353 and 73.22 % for SAL1219. For both ANG353 and SAL1219, the highest rates of locus recovery and locus length are in de novo ingroup samples, and the lowest rates are observed in the outgroups (Fig. 2). The performance of the Tree of Life samples varied depending on the targeted-capture datasets, which yielded more loci and longer locus lengths with SAL1219 than ANG353.

Compared to the ReWGS datasets, targeted-capture datasets had lower numbers of loci but with higher percentages of informative characters (Table 1). Considering the fact that SAL1219 has more loci with a higher percentage of parsimony informative sites and that it was designed for Salicaceae samples, it was not surprising that SAL1219 produced slightly higher node support values than ANG353. Despite this limitation, ANG353 resulted in fewer topological conflicts compared to SAL1219, possibly due to the latter having more variable loci. With its more conserved loci, the ANG353 bait set might be a better choice for clades with high rates of introgression and polyploidy, both of which are frequently observed in the Salicaceae. These results further prove that universal bait sets can be just as effective as taxon-specific bait sets (Larridon et al., 2020; Ogutcen et al., 2021).

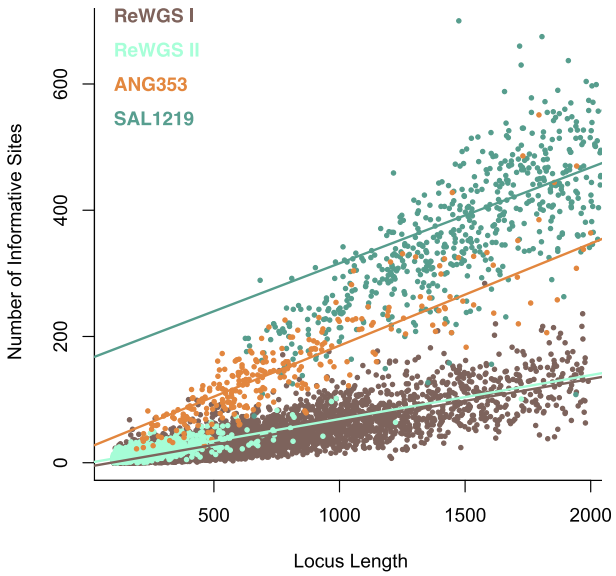

Fig. 1. Scatter plot showing the number of informative sites per locus length (in bp) for the four datasets used in this study. Each solid dot represents a locus, and the solid lines show the linear regression for each dataset.

Despite the accumulating examples of the use of universal bait sets on various angiosperm clades (Murphy et al., 2020; Shee et al., 2020; Giaretta et al., 2022; Haigh et al., 2023), previous studies showed that the performance of targeted-capture datasets varies depending on the similarity between the reference and the studied taxa (McLay et al., 2021), therefore occasionally favouring the use of clade-specific bait sets (Jantzen et al., 2020; Yardeni et al., 2022). We observed similar trends across our datasets: the highest locus recovery rates were observed with the de novo sequenced ingroup samples, and the outgroup samples had the lowest performance (Fig. 2). We also observed that the samples retrieved from the Tree of Life database had lower performance than our de novo sequenced samples in all our datasets. This could be due to the differences in sequencing technologies and the sequencing success,

Table 1
Summary statistics for the whole genome sequencing and targeted capture datasets.

	Whole Genome Sequencing		Targeted-Capture	
	Dataset I	Dataset II	ANG353	SAL1219
total number of loci	6,726	2,284	298	1,111
average locus length	511.50	179.72	845.42	2394.52
average number of samples per locus	52.47	51.50	62.42	58.28
average parsimony informative sites per locus	5.04%	6.70%	19.30%	23.31%
average gaps per locus	26.34%	1.98%	24.15%	12.60%

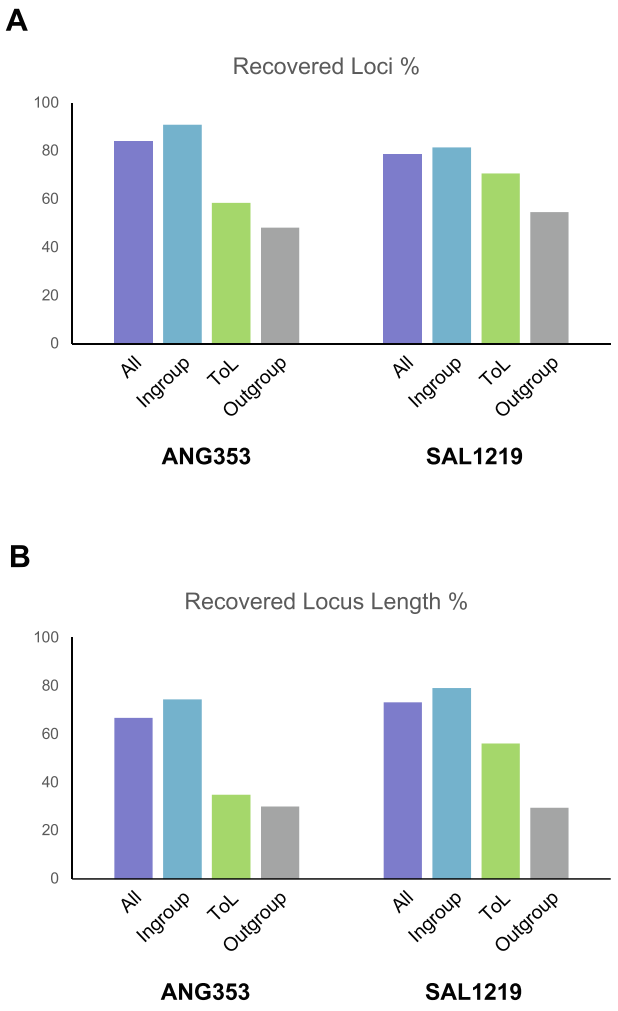

Fig. 2. Target recovery performances of ANG353 and SAL1219. A) Percentage of the recovered ANG353 and SAL1219 loci from each sample group. B) Percentage of the recovered ANG353 and SAL1219 locus length from each sample group. All: All 75 samples used in this study; Ingroup: 61 de novo sequenced samples that belong to the family Salicaceae; ToL: Nine samples for which the genomic DNA sequences were retrieved from the Tree of Life database; Outgroup: Five de novo sequenced Achariaceae samples that were collectively used as outgroup.

which depends on the sample quality and quantity (Johnson et al., 2018; Baker et al., 2022).

3.2. An updated Salicaceae phylogeny

Resolving the Salicaceae phylogeny has been a difficult task. Prior to the availability and cost efficiency of WGS, reconstructing the phylogeny long eluded researchers. The family is notorious for polyploidy and hybridization, and many species, especially in the genera *Populus* and *Salix,* have highly conserved plastomes with little variation (Gramlich et al., 2016; Wagner et al., 2021a; Zhang et al., 2021).

The use of RAD-Seq enriched WGS and targeted-capture datasets resulted in highly similar topologies for the family-level phylogenies, except for the genus *Salix* (see below for details). The overall tree topology was robust to different phylogenetic methodologies, as there were only minor discrepancies between the trees reconstructed with the supermatrix and supertree approaches (Supplementary Figures S1-S6).

Monophyly of the family was supported with 100 % bootstrap value in all phylogenies. Our largest dataset, reWGS I, analyzed with the supermatrix approach, produced a robust phylogeny with the highest support values. It demonstrated the highest consistency, sharing the most similarities with the other trees. Therefore, reWGS I supermatrix phylogeny is referred to as the main phylogeny hereafter (Figs. 3 and 4, Table 2). Overall, the main phylogenetic tree has well-defined clades with high support except for few nodes (Figs. 3 and 4).

3.2.1. Intrafamilial relationships within the Salicaceae

Based on morphological data, the Salicaceae is divided into three subfamilies: Samydoideae, Scyphostegioideae, and Salicoideae. Salicoideae is further split into six tribes: Abatieae, Bembicieae, Homalieae, Prockieae, Saliceae, and Scolopieae (Stevens, 2001 onwards). In the main phylogeny, the subfamily Samydoideae, which includes the large genus *Casearia*, and c. 11 other small, tropical genera, forms the outermost clade. The only member of the subfamily Scyphostegioideae, *Dianyuea turbinata* is recovered as sister to the subfamily Salicoideae. which is split into two main clades, but with low support (BP=0.38; Fig. 3). Clade I consists of two groups (BP=0.80): i) *Azara* and *Banara*; and ii) *Flacourtia*, *Xylosma*, and *Homalium*. Clade II consists of three groups: the largest is the *Populus-Salix* complex; sister to that is the *Idesia* and *Olmediella* group, and sister to these four genera is the *Carriera* and *Poliothyrsis* group. All bootstrap values are above 0.95, except for the ones reported above.

The main difference among the phylogenies generated with different datasets and methods (Supplementary Figures S1-S6) is the placement of *Homalium foetidum* and *Dianyuea turbinata*. While *Homalium foetidum* was often placed as sister to the *Flacourtia* and *Xylosma* clade with relatively high support, *Dianyuea turbinata* placement varied greatly among different trees and received much lower support. Besides the *Homalium foetidum* and *Dianyuea turbinata* placements, ReWGS I supermatrix and supertree, ANG353 supertree, SAL1219 supermatrix and supertree phylogenies have identical topologies. Among the rest of the phylogenies, there are four differences: i) ANG353 supermatrix tree places *Populus tremuloides* as sister to *P. tremula* and *P. alba*, as opposed to the other phylogenies placing *P. alba* as sister to the other two; ii) ReWGS II supertree and supermatrix trees place *Casearia clutiifolia* as sister to other *Casearia* species as opposed to other phylogenies placing *C. gossypiosperma* as sister to the others; iii) while the *Banara* and *Azara* clade is placed as sister to the *Flacourtia* and *Xylosma* clade in other phylogenies, ReWGS II supermatrix places this clade as sister to the rest of the Salicaceae except for the genus *Casearia*; and iv) in the outgroup, ReWGS II places *Pangium edule* as sister to *Trichadenia* and *Ryparosa* clade, whereas the other phylogenies place *Trichadenia philippinensis* as sister to *Pangium* and *Ryparosa* clade.

All of our Salicoideae samples were from the tribe Saliceae, except for two species (*Homalium foetidum* from Homalieae and *Banara tomentosa* from Prockieae), which were nested within the Saliceae in our phylogeny. Similar to our results, earlier research that founded the basis for the updated circumscription of the Salicaceae s.l. also placed *Homalium* and *Banara* within the paraphyletic Saliceae (Chase et al., 2002; Alford, 2005). However, our phylogeny and Alford (2005) both put *Homalium* as sister to the *Xylosma* – *Flacourtia* clade, whereas Chase et al. (2002) recovered *Homalium* as sister to the *Azara* – *Banara* clade, but with low support. More recent studies generated plastid-based Salicaceae phylogenies, and their placement of genera are identical to ours. While the sampling of Zhang et al. (2018) mostly included *Salix* and *Populus*, it also included four other genera. On the other hand, Li et al. (2019) had a broader taxonomic sampling with 18 genera, revealing the same intergeneric relationships as our phylogeny.

To date, this study has the largest genome-based sampling of the Salicaceae s.l., encompassing a wide range of genera throughout the family. Despite our sampling efforts, the phylogeny presented here is far from complete. We therefore avoid making broad inferences about the exact origins and the dispersal routes of the family. One thing we observed from the current phylogeny is that some genera with early divergence dates exhibit broad geographic ranges, spanning both the Old and the New World, as well as tropical and temperate regions. Notable examples include the pantropical *Casearia* and *Xylosma*, and

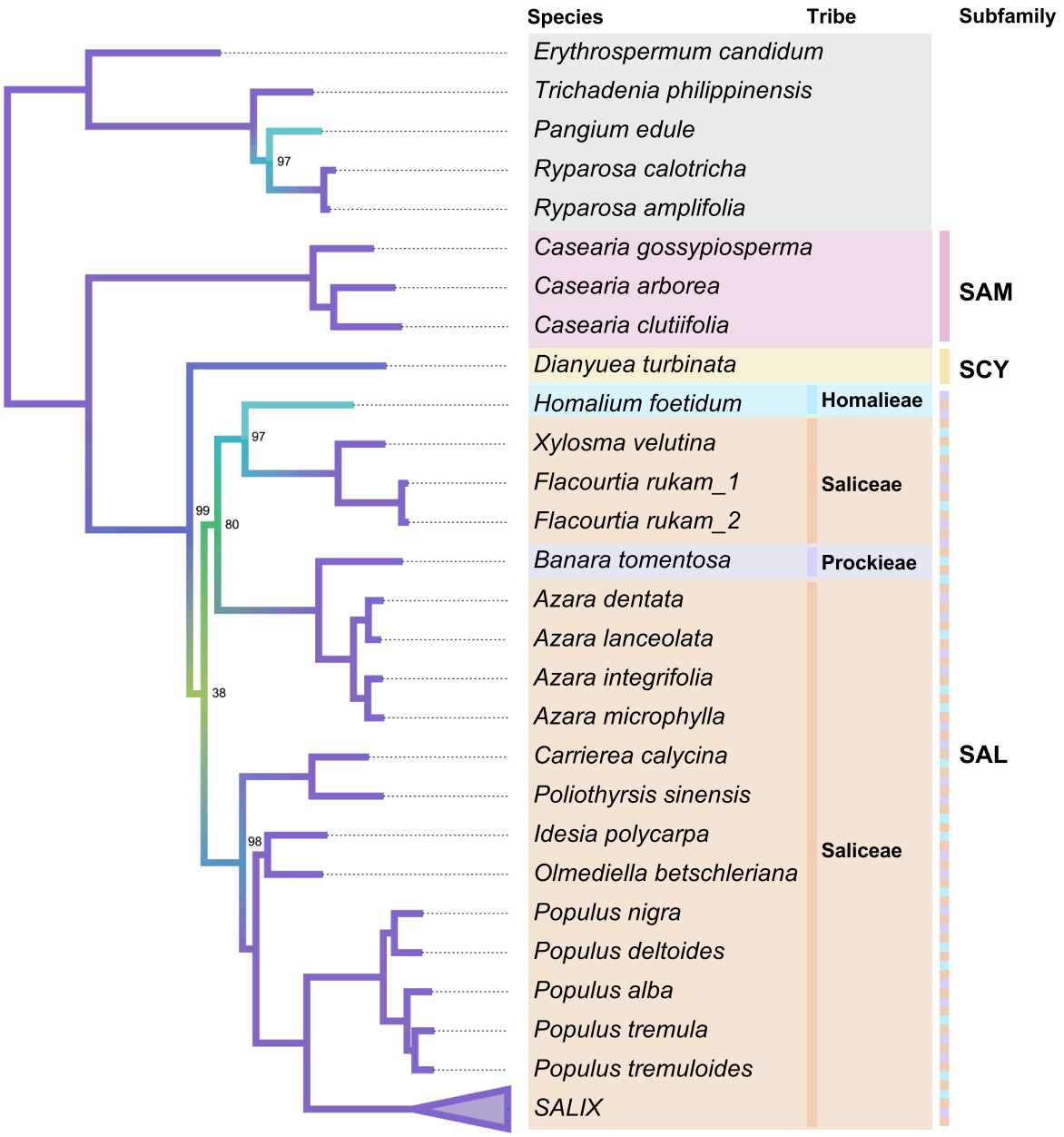

Fig. 3. Family-level Salicaceae phylogeny reconstructed using ReWGS I with the supermatrix approach. The columns on the right show subfamilies (SAM: Samydoideae, SCY: Scyphostegioideae, SAL: Salicoideae) and tribes. The genus *Salix* is displayed in Fig. 4. The outgroup (Achariaceae) is highlighted in gray. Branch colour gradient represents bootstrap support (BS) with purple (BS=100) to green (lower BS values). Only the nodes with BS<100 are labelled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cosmopolitan *Salix* and *Populus*, found in a wide range of habitats across the world. These findings suggest multiple geographic dispersals and historical presence of the Salicaceae across diverse geographic and climatic zones.

3.2.2. *Interspecies relationships within the genus Salix*

There have been repeated attempts to divide the large and complex genus *Salix* into subclades. Based on morphological characters, three subgenera are accepted: *Salix* s.l (sometimes divided into *Salix* s.s, and *Longifoliae*; Argus, 1997), *Chamaetia*, and *Vetrix* (Skvortsov, 1999). However, these traditional classifications have been disputed by molecular evidence, which suggests that most subgenera are not monophyletic, thus leaving the status of these clades unresolved (Chen et al., 2010; Lauron-Moreau et al., 2015; Sanderson et al., 2023).

Compared to the rest of the family, the relationships within the genus *Salix* had lower support values (Fig. 4), and the topologies recovered using different methods and datasets were moderately incongruent. In all phylogenies, the members of the subgenera *Chamaetia* and *Vetrix* were recovered as one monophyletic clade, but the monophyly of the subgenus *Salix* was not supported in any of the reconstructed phylogenies.

3.2.2.1. The subgenus Salix s.l. Overall, interspecific relationships within the subgenus are not well-supported. *Salix nigra* is recovered as sister to the rest of the genus *Salix* in all our phylogenies except for the ANG353 supermatrix tree, which clusters this species with most of the other members of the subgenus *Salix* s.l (Fig. 5). The position of *S. triandra* varies across the phylogenies generated with different datasets (Fig. 4, Supplementary Figures S1-S6), but they all agree that this species is not grouped in the *Salix* subgenus *Salix* s.l., unlike previous

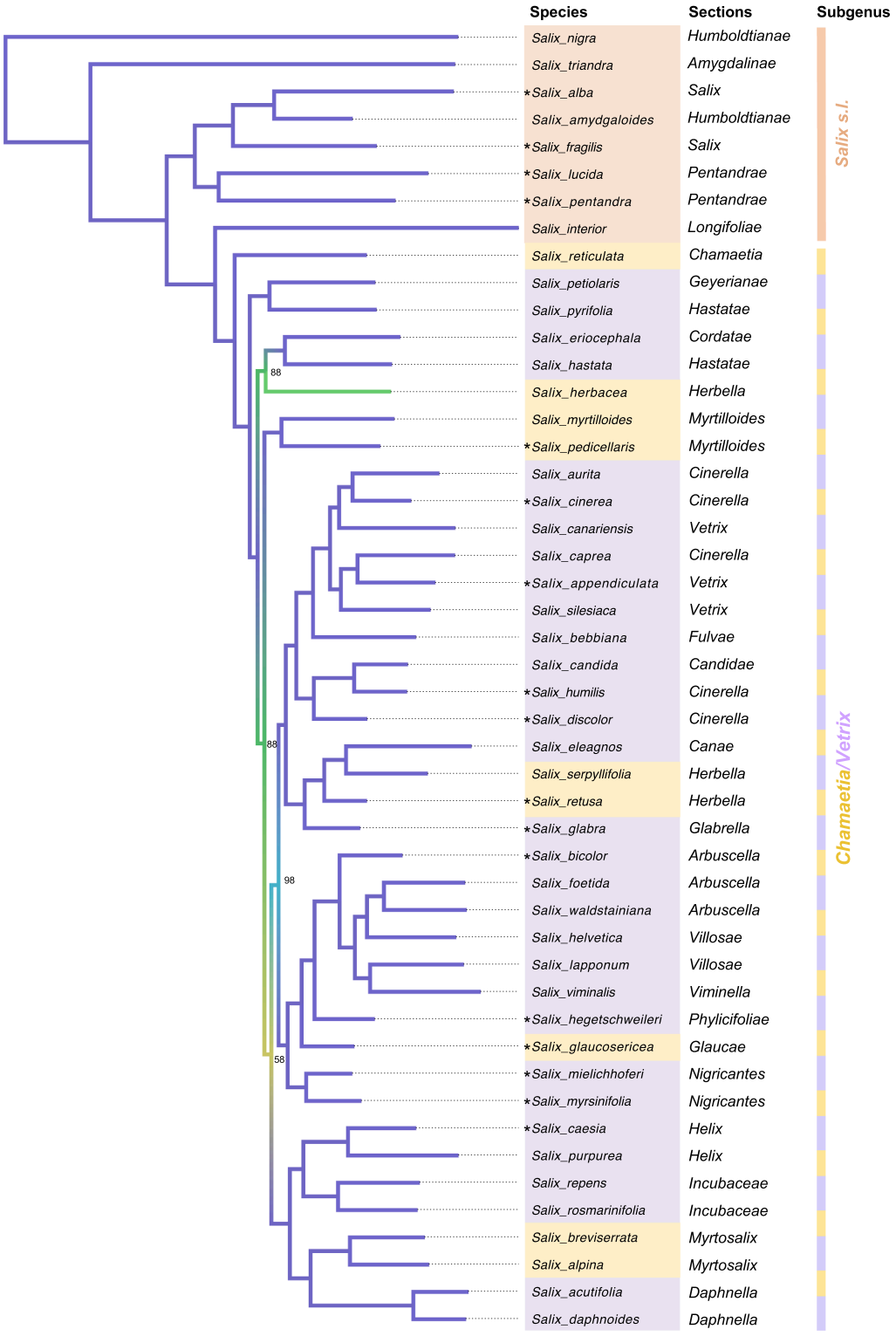

Fig. 4. *Salix* phylogeny reconstructed using ReWGS I with the supermatrix approach. The columns on the right show subgenera and sections. * indicates polyploid species. Branch colour gradient represents bootstrap support (BS) with purple (BS=100) to green (lower BS values). Only the nodes with BS<100 are labelled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

traditional morphology-based studies suggest (Argus, 1997; Skvortsov, 1999). While most of our phylogenies place *S. triandra* as the second outermost member of the whole genus, ANG353 and SAL1219 supermatrix trees and the ANG353 supertree recover this species as sister to the *Chamaetia/Vetrix* subgenus, which is in accordance with several studies (Trybush et al., 2008; Chen et al., 2010; Wu et al., 2015; Wagner et al., 2021a; Vašut et al. 2024).

Another species with inconsistent placement is *S. interior*, which is the only representative of the subgenus *Longifoliae* (classified as a section in Argus, 1997) in our sample set. In the ANG353 supermatrix phylogeny, *S. interior* is nested within the *Chamaetia/Vetrix* subgenus, whereas all other phylogenies recover it as sister to the *Chamaetia/Vetrix*

Table 2
The phylogenetic status of 22 clades at various taxonomic levels across eight phylogenetic trees. +++: monophyletic, —: polyphyletic. N/A indicates the clade is represented by less than two taxa. Gray cells indicate paraphyly (see notes below).

		reWGS I		reWGS II		ANG353		SAL1219	
		supermatrix	supertree	supermatrix	supertree	supermatrix	supertree	supermatrix	supertree
FAMILY	**Salicaceae**	+++	+++	+++	+++	+++	+++	+++	+++
	Achariaceae (outgroup)	+++	+++	+++	+++	+++	+++	+++	+++
SUBFAMILY	**Samydoideae**	+++	+++	+++	+++	+++	+++	+++	+++
	Salicoideae	+++	(1)	+++	(1)	+++	+++	+++	+++
TRIBE	**Saliceae**	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)
***SALIX* SUBGENUS**	***Salix s.l.***	---	---	---	---	---	---	---	---
	Chamaetia/Vetrix	+++	+++	+++	+++	(4)	+++	+++	+++
***SALIX* SECTION**	***Arbuscella***	---	---	---	---	---	---	---	---
	Cinerella	---	---	---	---	---	---	---	---
	Daphnella	+++	+++	+++	+++	+++	+++	+++	+++
	Hastatae	---	---	---	---	---	---	---	---
	Helix	+++	+++	+++	+++	---	+++	+++	+++
	Herbella	---	---	---	---	---	---	---	---
	Humboldtianae	---	---	---	---	---	---	---	---
	Incubaceae	+++	+++	+++	+++	+++	+++	+++	+++
	Myrtilloides	+++	+++	+++	---	---	---	+++	+++
	Myrtosalix	+++	+++	+++	+++	+++	+++	+++	+++
	Nigricantes	+++	+++	+++	+++	---	+++	+++	+++
	Pentandrae	+++	---	+++	---	(5)	---	---	---
	Salix	(6)	(6)	(6)	(6)	(6)	(6)	N/A	N/A
	Vetrix	---	---	---	---	---	---	---	---
	Villosae	---	---	---	---	---	---	---	---

(1) *Dianyuea* is included within the clade.
(2) *Banara* and *Homalium* are included within the clade.
(3) *Banara* is included within the clade.
(4) *Salix interior* is included within the clade.
(5) *Salix nigra* is included within the clade.
(6) *Salix amygdaloides is* included within the clade.

subgenus, which is concordant with a recent study based on genomic data (Sanderson et al., 2023). Based on plastid data, however, this species is quite isolated and in sister position to the subgenus *Salix* s.l. (Wagner et al., 2021b), hinting a complex evolutionary history including introgression and ancient hybridization.

None of our phylogenies recovered the members of the section *Humboldtianae* (*S. amygdaloides* and *S. nigra*) as monophyletic, and other studies confirm this finding (Lauron-Moreau et al., 2015; Wu et al., 2015, but also see Sanderson et al., 2023). On the contrary, morphologically similar but geographically isolated North American *S. lucida* and European *S. pentandra* were clustered together as members of the section *Salicaster* (named *Pentandrae* in Skvortsov, 1999) in the main phylogeny. These tetraploid species are morphologically distinct from other tetraploid species within the subgenus *Salix* (Argus, 1997).

Taken together, our study is concordant with previous findings that the subgenus *Salix* s.l. (or even *Salix* s.s. in some cases) does not form a monophyletic clade (Chen et al., 2010; Wu et al., 2015; Sanderson et al., 2023).

3.2.2.2. The subgenera Chamaetia and Vetrix. Consistent with our study, the non-monophyly of the subgenera *Chamaetia* and *Vetrix* has been repeatedly reported, and several studies suggested merging these two subgenera to form a monophyletic clade (Wu et al., 2015; Wagner et al., 2021a; Sanderson et al., 2023). Within the *Chamaetia/Vetrix* clade, most of the relationships within sections have high support values and they are mostly concordant among the ReWGS I, ReWGS II, ANG353, and SAL1219 phylogenies generated using supermatrix and supertree approaches. In all our phylogenies except for the ANG353 supermatrix tree, the section *Chamaetia* (represented by *S. reticulata*) is the outermost member of the *Chamaetia/Vetrix* clade, therefore supporting previous research (Wagner et al., 2018; Wagner et al., 2020; Sanderson et al., 2023; Marinček et al., 2024).

In all phylogenies except for the reWGS II supertree, sections *Arbuscella, Phylicifoliae* (part of *Arbuscella* in Skvortsov, 1999), *Villosae, and Viminella* (named *Vimen* in Skvortsov, 1999) form a clade, which is often recovered as sister to *Glaucae* (represented by *S. glaucosericea*). Section *Nigricantes* and *S. herbacea* are often recovered as part of this group as well. In the ReWGS I, ReWGS II, and SAL1219 phylogenies, sections *Daphnella, Helix, Incubaceae,* and *Myrtosalix* form a monophyletic clade with the *Daphnella-Myrtosalix* and *Incubaceae-Helix* groups forming sister clades, as previously reported (He et al., 2021; Volf et al., 2023). In the ANG353 supermatrix tree, however, *Daphnella-Myrtosalix* group is recovered as monophyletic, but the other sections are not clustered together.

Sections *Canae* (represented by *S. eleagnos*), *Herbella* (named *Retusae* in Skvortsov, 1999), and often *Glabrella* (represented by *S. glabra*) form a

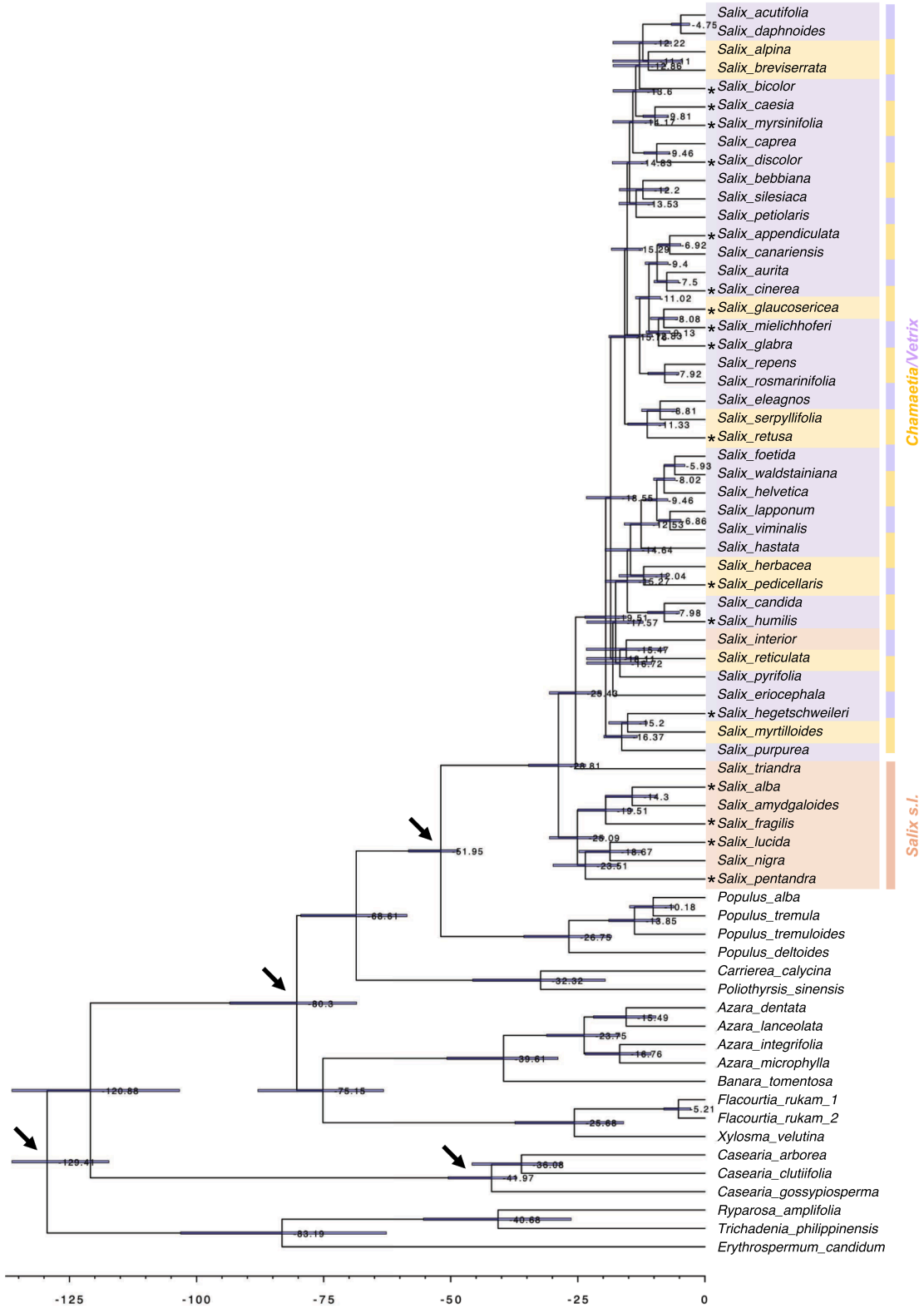

Fig. 5. Time-calibrated Salicaceae phylogeny reconstructed using ANG353 with the supermatrix approach and relaxed molecular clock. * indicates polyploid species. Arrows indicate the four calibration points (see Materials and Methods for details). Nodes are labelled with the mean age and the node bars represent the 95% highest posterior density (HPD). The timescale is in mya.

monophyletic clade across all phylogenies, except for the SAL1219 supertree. This *Canae/Glabrella/Herbella* clade is often placed as sister to the *Candidae*, *Cinerella, Fulvae, and Vetrix* clade (Wagner et al., 2020; Volf et al., 2023). It is worth noting that Skvortsov (1999) considered *Cinerella* and *Fulvae* as part of *Vetrix*, which, together with *S. candida,* form a monophyletic clade. Another notable difference in Skvortsov's (1999) classification is the placement of *Candidae* within the section *Villosae* which is not supported in our study.

The only representative of the section *Geyerianae*, *S. petiolaris* is sister to S. *pyrifolia* (section *Hastatae)* in all our phylogenies except for the ANG353 supermatrix tree. Sanderson et al. (2023) also clusters these two species together, further suggesting the re-evaluation of the taxonomic placements of these species. Out of all the sections sampled in this study, *Cordatae* (represented by *S. eriocephala*) and *Myrtilloides* have the most inconsistent placements among the generated phylogenies.

As we move from the older nodes to the tips, the phylogenetic relationships tend to stabilize. We had high support for the monophyly of the sections *Daphnella*, *Helix*, *Incubaceae*, and *Myrtosalix* in all our phylogenies. *Nigricantes* was recovered as a monophyletic section in all phylogenies except for the ANG353 supermatrix phylogeny. Similarly, the monophyly of the section *Myrtilloides* was recovered in all but the two ANG353 phylogenies and the ReWGS II supertree phylogeny. On the other hand, *Arbuscella*, *Cinerella*, *Hastatae*, *Herbella, Vetrix, and Villosae* are not recovered as monophyletic clades in any of the phylogenies. Our findings are similar to a previous research based on an extensive *Chamaetia/Vetrix* sampling that showed moderate overlap with our sample set (Wagner et al., 2020).

Overall, in terms of clade and section circumscriptions and their relationships among each other, the findings of Wagner et al. (2018) are identical to our ReWGS I supermatrix tree, but not to the ANG353 supermatrix tree. Furthermore, the ReWGS I phylogenies show more congruency between the trees generated using supermatrix and supertree approaches. This suggests that ReWGS I is better at resolving the phylogeny of the genus *Salix*. Since the species within this large genus show high genetic similarities, large numbers of loci are required to further investigate this challenging genus.

In sum, compared to the strong support for the groupings of sister species, the backbone of the clade has short branches with low support. Similar findings have been previously observed and it has been suggested that this phylogenetic pattern is linked with rapid diversification at the origin of the *Chamaetia/Vetrix* clade (Percy et al., 2014; Wagner et al., 2021a; Sanderson et al., 2023).

3.2.2.3. Ploidy variation within the genus Salix. 35 % of our *Salix* sampling included polyploid species with ploidy levels ranging from 3x to 8x (Hörandl et al., 2012; Supplementary Table S1). While polyploid taxa can pose challenges in phylogenetic reconstruction (Savage & Cavender-Bares, 2012; Rothfels, 2021), we implemented measures to mitigate their impact. We assumed that significant ploidy issues would manifest as more paralog warnings in polyploids than diploids. However, all samples showed similar sets of loci with paralog warnings, leading us to exclude these loci across all samples.

Across different datasets and phylogenetic reconstructions, the stability of the polyploid species were comparable to the diploid species, and incongruences in tree topology were observed equally in polyploid and diploid taxa, without higher frequency or unique patterns associated with polyploids. These results indicate that the sources of incongruence are likely related to factors affecting both polyploid and diploid species similarly and that polyploid species do not introduce additional phylogenetic instability. Additionally, the polyploid *Salix* species appear in multiple clades, suggesting several independent origins of polyploidy. Sanderson et al. (2023) reported high levels of ancient hybridization in *Salix,* which could explain the lack of increased conflicting signals in polyploids compared to diploids.

To conclude, our results demonstrate the robustness of our phylogenetic methodologies in handling polyploid species alongside diploids, providing valuable insights into the evolutionary dynamics of polyploid and diploid lineages.

3.3. The divergence timescale of the Salicaceae

After calibrating the ANG353 supermatrix phylogeny using four uniform priors (see Materials and Methods for details), the origin of the family Salicaceae was recovered as 120.88 mya with 95 % highest posterior density (HPD) intervals of 103.34 and 136.30 mya (Fig. 5). The Salicaceae origin estimates from Zuntini et al. (2024), which was used as a secondary calibration source, are slightly older (108.29 mya) but still fall within our estimated age range. Two previous studies, one on the order Malpighiales (Xi et al., 2012) and one on the family Salicaceae (Zhang et al., 2018) estimated the origin of the Salicaceae as 68.9 and 68.7 mya, respectively, but neither study included any samples from the genus *Casearia*. According to our study, the genus *Casearia* is the earliest divergent lineage of Salicaceae, with an estimated origin of 41.97 (95 % HPD: 37.00–50.50) mya, whereas a recent study on *Casearia* estimated a slightly younger origin of the genus (39.13 mya; de Mestier et al., 2022). In our phylogeny, the subfamily Salicoideae has an age of 80.30 (95 % HPD: 68.52–93.46) mya, which is a closer estimate to the findings of Xi et al. (2012) and Zhang et al. (2018). A more recent Malpighiales phylogeny estimated the age of the same clade as 52.1 mya, however, with only three Salicaceae species the study lacked a broad taxonomic sampling effort (Cai et al., 2019). Compared to another Salicaceae-focused study (Li et al., 2019), our phylogeny estimated much earlier origins for the Salicaceae and its major clades. While Li et al. (2019) included a wide range of genera in their phylogeny, the number of samples (26) was much lower than our dataset (75). Another major difference was their use of the Salicaceae − Passifloraceae split as their calibration point. As a result, their age estimates for the Salicaceae (87 mya) and the *Salix* – *Populus* split (34 mya) were much more recent than our estimates. On the other hand, Liu et al. (2022), with a dense *Populus* sampling and only four *Salix* species, estimated a much older age for the *Salix* – *Populus* split (68.67 mya).

A previously published phylogeny focused mainly on the *Salix* subgenus *Salix* estimated the origin of the genus as 43.87 mya (Wu et al., 2015). This is much older than our estimation of 28.81 mya (95 % HPD: 23.45–34.68). Our sample set densely represents the *Chamaetia/Vetrix* subgenus, which is estimated to have a more recent origin (19.51 mya with 95 % HPD: 15.44–23.61) than the subgenus *Salix* (25.09 mya with 95 % HPD: 19.85–30.58). An earlier origin of the *Chamaetia/Vetrix* subgenus is supported by previous studies (He et al., 2021; Wagner et al., 2021a), which repeatedly concluded that the interspecies relationships within this clade are difficult to resolve and that the clade has undergone recent radiation and geographical expansion, which gave rise to about 75 % of the species within the genus *Salix*. Skvortsov (1999) also observed that the *Salix* subgenus exhibits more plesiomorphic morphological characters, further pointing out the more recent origin of the *Chamaetia/Vetrix* subgenus.

Due to computational limitations of the Bayesian relaxed molecular clock approach with MCMC sampling, time calibration was performed on the ANG353 supermatrix phylogeny instead of the main phylogeny. ANG353 is our smallest dataset, which also resulted in the least concordant topology compared to our other trees. As a supplementary approach, we performed PL-based time-calibrations on all our trees using the same priors as our main approach. PL-based age estimates were mostly consistent across the eight phylogenies, except for the *Chamaetia/Vetrix* node that showed a wide range of estimated ages (0.68–17.05 mya). Overall, age estimates were similar between the PL and Bayesian-MCMC approaches, although some nodes had higher or lower age estimates with the PL approach compared to the Bayesian-MCMC approach (Table 3).

Table 3
Age estimations (in mya) for major clades calculated using penalized-likelihood (PL) and Bayesian-MCMC approaches. Besides the mean values, PL approach reports the range of these mean values across all eight trees, and the Bayesian-MCMC method reports 95% highest posterior density (HPD) intervals.

	Penalized-Likelihood		Bayesian-MCMC	
	mean	range	mean	95 % HPD
Root	110.83	110.02–112.23	129.41	117.27–136.30
Outgroup	57.76	55.01–74.82	83.19	62.66–103.12
Salicaceae	99.36	97.22–102.94	120.88	103.34–136.30
Casearia	66.66	62.35–72.74	41.97	37.00–50.50
Salicoideae	86.24	81.03–94.91	80.30	68.52–93.46
***Populus*-*Salix* complex**	67.08	63.92–73.59	51.95	48.50–58.3
Populus	43.97	33.82–49.06	26.75	18.39–35.62
Salix	35.94	32.23–51.14	28.81	23.45–34.68
Salix* subgenus *Chamaetia/Vetrix	4.33	0.68–17.05	19.51	15.44–23.61

3.4. Conclusions and future directions

Our study provides an overview of the Salicaceae phylogeny with the most taxonomically inclusive sampling up to date. While our findings are mostly consistent with former studies, they also add new insights on the position of early lineages within the Salicaceae, thus further improving our understanding of the evolution of the family.

Despite our extensive efforts, our sampling had some geographical and taxonomical limitations. Similar to many other studies on the Salicaceae (Leskinen & Alstrom-Rapaport, 1999; Zhang et al., 2018; Sanderson et al., 2023), our sample set mostly includes *Salix*. While we have representatives from almost all of the species-rich temperate and tropical genera, we had no access to the small tropical genera. For future research, our ongoing collaborations will help us to increase our sampling effort in the tropics, both in the New World and the Old World. Besides sampling limitations, the general lack of a substantial Salicaceae phylogeny is also due to the difficulties in obtaining high-quality molecular data. As genomic sequencing is becoming cheaper and faster with more tools available for non-model organisms (Johnson et al., 2016; Andermann et al., 2018; Johnson et al., 2018; Minh et al., 2020; Sanderson et al., 2020), this obstacle is soon to be overcome. However, the complex evolution of the Salicaceae s.s. remains a challenge for resolving the entangled relationships within this family.

Advancing our understanding of the evolutionary history of the Salicaceae is important, because this species-rich plant family is a good model system for studying many aspects of plant evolution (Wu et al., 2015; de Mestier et al., 2022; Sanderson et al., 2023). As the Salicaceae harbours an immense range of plant chemical defenses (Boeckler et al., 2011; Volf et al., 2015; Leong et al., 2022), a well-founded phylogeny would provide a strong backbone for a better understanding of the evolutionary factors contributing to this diversity and potential variation in drivers of diversification associated with climate and geography. Salicaceae is also an excellent study system for exploring how hybridization and polyploidy impact diversification and adaptation to various habitats (Gramlich et al., 2016; Wagner et al., 2020), and for studying the role of adaptation and trait evolution in community assemblages along environmental gradients (Savage and Cavender-Bares 2012, 2013). In summary, our study contributes to advancing our understanding of the evolutionary history of the Salicaceae, paving the path for further research in this ecologically significant, cosmopolitan family.

CRediT authorship contribution statement

Ezgi Ogutcen: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Paola de Lima Ferreira:** Writing – review & editing, Methodology. **Natascha D. Wagner:** Writing – review & editing, Formal analysis. **Pia Marinček:** Formal analysis, Methodology. **Jing Vir Leong:** Methodology, Data curation. **Gibson Aubona:** Data curation. **Jeannine Cavender-Bares:** Writing – review & editing, Data curation. **Jan Michálek:** Methodology, Data curation. **Lucy Schroeder:** Data curation. **Brian E. Sedio:** Writing – review & editing, Data curation. **Radim J. Vašut:** Writing – review & editing, Formal analysis. **Martin Volf:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests that could have influenced the work reported in this paper.

Acknowledgements

We thank Tereza Volfová, Petr Kozel, and Nela Urbanová for their help with laboratory procedures and sample collection. JVL and MV acknowledge funding by the Czech Science Foundation (grant number 23-06855L). EO and PLF acknowledge funding by the MEMOVA project (EU Operational Programme Research, Development and Education No. CZ.02.2.69/0.0/0.0/18_053/). NDW acknowledge funding by Deutsche Forschungsgemeinschaft (DFG) priority program Taxon-Omics SPP1991 project Wa3684/2-1. We also acknowledge NSF ASCEND BII (DBI 2021898), Cedar Creek LTER (DEB 1831944), and DEB 2240430 grants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108161.

References

Alford, M.H., 2005. Systematic studies in Flacourtiaceae. Cornell University. PhD dissertation.
Andermann, T., Cano, Á., Zizka, A., Bacon, C., Antonelli, A., 2018. SECAPR—a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. PeerJ 6, e5175.
Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.
Angiosperm Phylogeny Group (APG), 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society 141 (4), 399–436.
Argus, G.W., 1997. Infrageneric classification of *Salix* (Salicaceae) in the new world. Systematic Botany Monographs 1–121.
Argus, G. W., 2010. *Salix* L. in Flora of North America North of Mexico, Volume 7. Magnoliophyta: Salicaceae to Brassicaceae, eds. Flora of North America Editorial Committee. Oxford University Press, 23–162.
Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Forest, F., 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Systematic Biology 71 (2), 301–319.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19 (5), 455–477.
Boeckler, G.A., Gershenzon, J., Unsicker, S.B., 2011. Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. Phytochemistry 72 (13), 1497–1509.
Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660.
Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology 15 (4), e1006650.
Bradshaw, H.D., Ceulemans, R., Davis, J., Stettler, R., 2000. Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. Journal of Plant Growth Regulation 19 (3), 306–313.
Cai, L., Xi, Z., Amorim, A.M., Sugumaran, M., Rest, J.S., Liu, L., Davis, C.C., 2019. Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. New Phytologist 221 (1), 565–576.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17 (4), 540–552.
Chase, M.W., Zmarzty, S., Lledo, M.D., Wurdack, K.J., Swensen, S.M., Fay, M.F., 2002. When in doubt, put it in Flacourtiaceae: a molecular phylogenetic analysis based on plastid *rbcL* DNA sequences. KEW Bulletin 57 (1), 141–181.
Chen, J.H., Sun, H., Wen, J., Yang, Y.P., 2010. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. Taxon 59 (1), 29–37.
Chernomor, O., Von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Systematic Biology 65 (6), 997–1008.

Collinson, M.E., 1992. The early fossil history of Salicaceae: a brief review. Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences, 98, 155-167.
Cronk, Q.C.B., 2005. Plant eco-devo: the potential of poplar as a model organism. New Phytologist 166 (1), 39–48.
de Mestier, A., Brokamp, G., Celis, M., Falcón-Hidalgo, B., Gutiérrez, J., Borsch, T., 2022. Character evolution and biogeography of Casearia (Salicaceae): Evidence for the South American origin of a pantropical genus and for multiple migrations to the Caribbean islands. Taxon 71 (2), 321–347.
Douglas, J., Zhang, R., Bouckaert, R., 2021. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. PLoS Computational Biology 17 (2), e1008322.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7 (1), 1–8.
Giaretta, A., Murphy, B., Maurin, O., Mazine, F.F., Sano, P., Lucas, E., 2022. Phylogenetic relationships within the hyper-diverse genus *Eugenia* (Myrtaceae: Myrteae) based on target enrichment sequencing. Frontiers in Plant Science 12, 759460.
Graham, A., 1985. Studies in neotropical paleobotany. IV. The Eocene communities of Panama. Annals of the Missouri Botanical Garden, 504-534.
Gramlich, S., Sagmeister, P., Dullinger, S., Hadacek, F., Hörandl, E., 2016. Evolution in situ: hybrid origin and establishment of willows (*Salix* L.) on alpine glacier forefields. Heredity 116 (6), 531–541.
Haigh, A.L., Gibernau, M., Maurin, O., Bailey, P., Carlsen, M.M., Hay, A., Forest, F., 2023. Target sequence data shed new light on the infrafamilial classification of Araceae. American Journal of Botany 110 (2), e16117.
Hardig, T.M., Brunsfeld, S.J., Fritz, R.S., Morgan, M., Orians, C.M., 2000. Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. Molecular Ecology 9 (1), 9–24.
Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State University.
He, L., Wagner, N.D., Hörandl, E., 2021. RAD sequencing data reveal a radiation of willow species (*Salix* L., Salicaceae) in the Hengduan Mountains and adjacent areas. Journal of Systematics and Evolution 59, 44–57.
Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35 (2), 518–522.
Hörandl, E., Florineth, F., Hadacek, F., 2012. Weiden in Österreich und angrenzenden Gebieten (Willows in Austria and adjacent regions). University of Agriculture, Vienna.
Jansson, S., Douglas, C.J., 2007. *Populus*: a model system for plant biology. Annual Review of Plant Biology 58, 435–458.
Jantzen, J.R., Amarasinghe, P., Folk, R.A., Reginato, M., Michelangeli, F.A., Soltis, D.E., Soltis, P.S., 2020. A two-tier bioinformatic pipeline to develop probes for target capture of nuclear loci with applications in Melastomataceae. Applications in Plant Sciences 8 (5), e11345.
Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in Plant Sciences 4 (7), 1600016.
Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Wickett, N.J., 2018. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Systematic Biology 68 (4), 594–606.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14 (6), 587–589.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30 (4), 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649.
Larridon, I., Villaverde, T., Zuntini, A.R., Pokorny, L., Brewer, G.E., Epitawalage, N., Baker, W.J., 2020. Tackling rapid radiations with targeted sequencing. Frontiers in Plant Science 10, 1655.
Lauron-Moreau, A., Pitre, F.E., Argus, G.W., Labrecque, M., Brouillet, L., 2015. Phylogenetic relationships of American willows (*Salix* L., Salicaceae). PLoS One 10 (4), e0121965.
Leong, J.V., Jorge, L.R., Seifert, C.L., Volf, M., 2022. Quantity and specialisation matter: Effects of quantitative and qualitative variation in willow chemistry on resource preference in leaf-chewing insects. Insect Conservation and Diversity 15 (4), 453–460.
Leskinen, E., 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 Systematics and Evolution 215, 209–227.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14), 1754–1760.
Li, M.M., Wang, D.Y., Zhang, L., Kang, M.H., Lu, Z.Q., Zhu, R.B., Ma, T., 2019. Intergeneric relationships within the family Salicaceae s.l. based on plastid phylogenomics. International Journal of Molecular Sciences 20 (15), 3788.
Liu, X., Wang, Z., Wang, D., Zhang, J., 2016. Phylogeny of *Populus-Salix* (Salicaceae) and their relative genera using molecular datasets. Biochemical Systematics and Ecology 68, 210–215.
Liu, X., Wang, Z., Wang, W., Huang, Q., Zeng, Y., Jin, Y., Zhang, J., 2022. Origin and evolutionary history of *Populus* (Salicaceae): Further insights based on time divergence and biogeographic analysis. Frontiers in Plant Science 13, 1031087.
Manchester, S.R., Dilcher, D.L., Tidwell, W.D., 1986. Interconnected reproductive and vegetative remains of *Populus* (Salicaceae) from the middle Eocene Green River Formation, northeastern Utah. American Journal of Botany 73 (1), 156–160.
Manchester, S.R., Judd, W.S., Handley, B., 2006. Foliage and fruits of early poplars (Salicaceae: *Populus*) from the Eocene of Utah, Colorado, and Wyoming. International Journal of Plant Sciences 167 (4), 897–908.
Marinček, P., Léveillé-Bourret, É., Heiduk, F., Leong, J., Bailleul, S.M., Volf, M., Wagner, N.D., 2024. Challenge accepted: Evolutionary lineages versus taxonomic classification of North American shrub willows (*Salix*). American Journal of Botany e16361.
McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Jackson, C.J., 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. Applications in Plant Sciences 9 (7), e11420.
Meeuse, A.D.J., 1975. Taxonomic relationships of Salicaceae and Flacourtiaceae: their bearing on interpretative floral morphology and dilleniid phylogeny. Acta Botanica Neerlandica 24 (5/6), 437–457.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37 (5), 1530–1534.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30 (17), i541–i548.
Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., Cheek, M., 2020. A phylogenomic analysis of *Nepenthes* (Nepenthaceae). Molecular Phylogenetics and Evolution 144, 106668.
Narango, D.L., Tallamy, D.W., Shropshire, K.J., 2020. Few keystone plant genera support the majority of Lepidoptera species. Nature Communications 11 (1), 1–8.
Nyman, T., Julkunen-Tiitto, R., 2005. Chemical variation within and among six northern willow species. Phytochemistry 66, 2836–2843.
Ogutcen, E., Christe, C., Nishii, K., Salamin, N., Möller, M., Perret, M., 2021. Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. Molecular Phylogenetics and Evolution 157, 107068.
Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.
Percy, D.M., Argus, G.W., Cronk, Q.C., Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Graham, S.W., 2014. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? Molecular Ecology 23 (19), 4737–4756.
Powo, 2023. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew http://www.plantsoftheworldonline.org/.
R Core Team, 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67 (5), 901–904.
Rambaut, A., 2014. FigTree v1. 4.2, a graphical viewer of phylogenetic trees.
Ramírez-Barahona, S., Sauquet, H., Magallon, S., 2020. The delayed and geographically heterogeneous diversification of flowering plant families. Nature Ecology & Evolution 4 (9), 1232–1238.
Rothfels, C.J., 2021. Polyploid phylogenetics. New Phytologist 230 (1), 66–72.
Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19 (1), 101–109.
Sanderson, B.J., DiFazio, S.P., Cronk, Q.C., Ma, T., Olson, M.S., 2020. A targeted sequence capture array for phylogenetics and population genomics in the Salicaceae. Applications in Plant Sciences 8 (10), e11394.
Sanderson, B.J., Gambhir, D., Feng, G., Hu, N., Cronk, Q.C., Percy, D.M., Olson, M.S., 2023. Phylogenomics reveals patterns of ancient hybridization and differential diversification that contribute to phylogenetic conflict in willows, poplars, and close relatives. Systematic Biology syad042.
Savage, J.A., Cavender-Bares, J., 2012. Habitat specialization and the role of trait lability in structuring diverse willow (genus *Salix*) communities. Ecology 93, S138–S150.
Savage, J.A., Cavender-Bares, J., 2013. Phenological cues drive an apparent trade-off between freezing tolerance and growth in the family Salicaceae. Ecology 94, 1708–1717.
Shee, Z.Q., Frodin, D.G., Cámara-Leret, R., Pokorny, L., 2020. Reconstructing the complex evolutionary history of the Papuasian *Schefflera* radiation through herbariomics. Frontiers in Plant Science 11, 258.
Skvortsov, A. K., 1999. Willows of Russia and Adjacent Countries: Taxonomical and Geographical Revision (transl. from: Skvortsov AK (1968) Willows of the USSR: Taxonomic and Geographic Revision. Nauka, Moscow). Joensuu University, Joensuu.
Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31.
Smith, W.R., 2008. Trees and Shrubs of Minnesota. University of Minnesota Press, Minneapolis.
Stevens, P.F., 2001. onwards). Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/.
Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56 (4), 564–577.
Tognetti, R., Cocozza, C., Marchetti, M., 2013. Shaping the multifunctional tree: the use of Salicaceae in environmental restoration. iForest-Biogeosciences and Forestry 6 (1), 37.

Trybush, S., Jahodová, Š., Macalpine, W., Karp, A., 2008. A genetic study of a *Salix* germplasm resource reveals new insights into relationships among subgenera, sections and species. Bioenergy Research 1, 67–79.
Vašut, R.J., Sochor, M., Hroneš, M., 2013. Vrby české republiky (Willows of the Czech Republic). Univerzita Palackého v Olomouci, Olomouc.
Vašut, R.J., Pospíšková, M., Lukavský, J., Weger, J., 2024. Detection of hybrids in willows (*Salix*, Salicaceae) using genome-wide DArTseq markers. Plants 13 (5), 639.
Volf, M., Julkunen-Tiitto, R., Hrcek, J., Novotny, V., 2015. Insect herbivores drive the loss of unique chemical defense in willows. Entomologia Experimentalis et Applicata 156 (1), 88–98.
Volf, M., Volfová, T., Hörandl, E., Wagner, N.D., Luntamo, N., Salminen, J.P., Sedio, B.E., 2022. Abiotic stress rather than biotic interactions drives contrasting trends in chemical richness and variation in alpine willows. Functional Ecology 36, 2701–2712.
Volf, M., Leong, J.V., de Lima Ferreira, P., Volfová, T., Kozel, P., Matos-Maraví, P., Sedio, B.E., 2023. Contrasting levels of β-diversity and underlying phylogenetic trends indicate different paths to chemical diversity in highland and lowland willow species. Ecology Letters 26, 1559–1571.
Wagner, N.D., Gramlich, S., 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. Ecology and Evolution 8 (16), 8243–8255.
Wagner, N.D., He, L., Hörandl, E., 2020. Phylogenomic relationships and evolution of polyploid *Salix* species revealed by RAD sequencing data. Frontiers in Plant Science 11, 1077.
Wagner, N.D., He, L., Hörandl, E., 2021a. The evolutionary history, diversity, and ecology of willows (*Salix* L.) in the European Alps. Diversity 13 (4), 146.
Wagner, N.D., Volf, M., Hörandl, E., 2021b. Highly diverse shrub willows (Salix L.) share highly similar plastomes. Frontiers in Plant Science 12, 662715.
Wu, J., Nyman, T., Wang, D.C., Argus, G.W., Yang, Y.P., Chen, J.H., 2015. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. BMC Evolutionary Biology 15 (1), 1–13.
Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Davis, C. C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proceedings of the National Academy of Sciences 109 (43), 17519–17524.
Yardeni, G., Viruel, J., Paris, M., Hess, J., Groot Crego, C., de La Harpe, M., Leroy, T., 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. Molecular Ecology Resources 22 (3), 927–945.
Zhang, L., Xi, Z., Wang, M., Guo, X., Ma, T., 2018. Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. Ecology and Evolution 8 (16), 7817–7823.
Zong, D., Gan, P., Zhou, A., Zhang, Y., Zou, X., Duan, A., He, C., 2019. Plastome sequences help to resolve deep-level relationships of *Populus* in the family Salicaceae. Frontiers in Plant Science 10, 5.
Zuntini, A.R., Carruthers, T., Maurin, O., Bailey, P.C., Leempoel, K., Brewer, G.E., Knapp, S., 2024. Phylogenomics and the rise of the angiosperms. Nature 1–8.