Molecular phylogenetic analyses identify the process of speciation of endemic willow species in the Japanese Archipelago

Satoshi Kikuchi (kikusato@affrc.go.jp)
Forestry and Forest Products Research Institute: Shinrin Sogo Kenkyujo https://orcid.org/0000-0003-2858-9681
Suzuki Setsuko
Forestry and Forest Products Research Institute: Shinrin Sogo Kenkyujo
Teruyoshi Nagamitsu
Forestry and Forest Products Research Institute: Shinrin Sogo Kenkyujo
Wajiro Suzuki
Tadami Beech Center

Research Article

Keywords: Salix, low-copy nuclear genes, chloroplast, phylogenetic networks, divergence time estimation, Japanese Archipelago

Posted Date: June 26th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3070336/v1

Abstract

Here we examine the evolutionary history of *Salix* in Japan, and in particular the subg. *Vetrix/Chamaetia* complex. To do so, we performed molecular phylogenetic analyses covering all available native species, using multilocus datasets of low-copy nuclear genes and chloroplast sequences. Using phylogenetic network analysis and divergence time estimation, we identified three major lineages within Japanese subg. *Vetrix*, confirmed the polyphyly of subg. *Chamaetia*, and further resolved the taxonomic status of various taxa at the section to species levels. Moreover, this study also highlighted the speciation processes for many endemic species. These include *S. hukaoana*, a novel monotypic section of *Hukaoana* that distinctly shows ancient divergence and not hybrid speciation, and *S. miyabeana* (sect. *Helix*), which shows evidence of genomic and morphological differentiation from subsp. *miyabeana*via intersectional hybridization and introgression with the sympatric species *S. schwerinii* (sect. *Viminella*). Finally, we also identified local endemics classified into sect. *Hastatae*(i.e., *S. rupifraga*, *S. shiraii*, and *S. japonica*) which show evidence of radiative speciation from a single lineage descended from *S. vulpina*(sect. *Cinerella*).

Introduction

The genus *Salix* (Salicaceae) comprises ca. 400 tree species mainly distributed in the temperate, boreal, and arctic regions of the Northern Hemisphere. They are present in diverse ecological niches, including wetlands, riparian vegetation, uplands, and alpine/arctic tundra (Newsholme 1992; Dickmann and Kuzovkina 2014), and have diverse economic and ecological uses, including biomass production, ecological restoration, and various wood products (Pučka and Lazdiņa 2013).

A comprehensive classification of *Salix* has been attempted by many taxonomists, however, this has proven to be difficult due to the dimorphic sexual system, simple flowers, large phenotypic variation, and frequent hybridization and polyploidization of this genus (Argus 1997; Cronk et al. 2015). Current classification of *Salix* is based on several authoritative taxonomic opinions proposed for each region (Argus 1997; Fang et al. 1999; Skvortsov 1999; Dickmann and Kuzovkina 2014; Ohashi 2019).

Many molecular phylogenetic studies have been conducted to untangle the taxonomic and systematic complexity of willows: early studies based on chloroplast and nuclear ribosomal genes revealed key evolutionary trends in the genus *Salix* and contributed to proposed taxonomic rearrangements at subgenus levels. In particular, evidence confirmed that subg. *Vetrix/ Chamaetia* was a joint clade, suggesting their recent diversification and the repeated evolution of dwarf arctic/alpine willows (subg. *Chamaetia*) (Azuma et al. 2000; Chen et al. 2010; Hardig et al. 2010; Lauron-Moreau et al. 2015; Acar et al. 2022). Recent advances in genomic sequencing have enabled high-throughput phylogenetic analysis of willows, thereby unraveling the complex evolutionary history of this clade at lower taxonomic levels, particularly within the *Vetrix/Chamaetia* complex in Europe and China (Wagner et al. 2018, 2021; He et al. 2021).

The most updated classification of *Salix* in the Japanese Archipelago by Ohashi (2001) described 27 native species (Table 1). The majority (17) belonged to subg. *Vetrix*, with three alpine dwarf willows included in subg. *Chamaetia*, while subg. *Salix* included four species and subg. *Protitea* has only one. *S. arbutifoli*a Pall. and *S. cardiophylla* Trautv. & Mey., which had previously been given the rank of genus (*Chosenia* and *Toisusu*, respectively), are now affiliated with the subgenera *Chosenia* and *Pleuradenia*, respectively.

Table 1
List of *Salix* species native to Japan and their taxonomic status (Ohashi 2001), along with the number of samples analyzed in this study.

Subgenus	Section	Species	Subspecies	N	Comments
Pleuradenia Kimura		*Salix cardiophylla* Trautv. & Mey.		3	
Chosenia (Nakai) H.Ohashi		*Salix arbutifolia* Pall.		1	
Protitea Kimura		*Salix chaenomeloides* Kimura		4	
Chamaetia (Dumortier) Nasarov	*Herbella* Seringe	*Salix nummularia* Andersson		2	
	Myrtilloides (Borrer) Andersson	*Salix fuscescens* Andersson		-	Not collected in this study.
	Glaucae (Fries) Andersson	*Salix nakamurana* Koidz.	subsp. *nakamurana*	-	Endemic to Japan; Not collected in this study.
			subsp. *kurilensis*	2	
Salix	*Triandrae* Dumortier	*Salix triandra* L.		3	
	Subalbae Koidz.	*Salix eriocarpa* Franch. & Sav.		1	
		Salix pierotii Miq.		2	
		Salix jessoensis Seemen		5	Endemic to Japan
Vetrix Dumortier	*Hastatae* (Fries) A.Kerner	*Salix japonica* Thunb.		2	Endemic to Japan
		Salix shiraii Seemen		1	Endemic to Japan
		Salix rupifraga Koidz.		2	
	Sieboldianae C.K.Schneid.	*Salix sieboldiana* Blume		4	Endemic to Japan
		Salix reinii Seemen		2	
	Helix Dumortier	*Salix miyabeana* Seemen	subsp. *miyabeana*	2	
			subsp. *gilgiana*	2	
		Salix integra Thunb.		2	

Subgenus	Section	Species	Subspecies	N	Comments
	Incubaceae A.Kerner	*Salix subopposita* Miq.		1	
	Subviminales C.K.Schneid.	*Salix gracilistyla* Miq.		7	Including one sample from Korea.
	Hukaoana Kimura	*Salix hukaoana* Kimura		15	Endemic to Japan
	Daphnella Seringe	*Salix rorida* Lacksch.		9	Including f. pendula (1) and f. roridaeformis (1)
	Viminella Seringe	*Salix schwerinii* E. Wolf		4	
		Salix udensis Trautv. & Mey.		13	
	Cinerella Seringe	*Salix taraikensis* Kimura		1	
		Salix caprea L.		11	Including samples from Korea (2) and European variety var. coaetanea (1)
		Salix futura Seemen		2	Endemic to Japan
		Salix vulpina Andersson		4	

These taxa apparently have floristic links with continental east Asia, including Northeast China, the Korean Peninsula, Far Eastern Russia, and—in rare cases—Europe. However, they also show a moderate level of endemism (30%), harboring eight endemic species and two subspecies, some of which are of uncertain taxonomic affinity. For example, *S. hukaoana* Kimura, although given a novel monotypic section based on its unique morphological traits (Kimura 1973, 1974), is suspected to have originated from hybrid speciation. Moreover, three species (i.e., *S. rupifraga* Koidz., *S. shiraii* Seemen, and *S. japonica* Thunb.) of sect. *Hastatae* are all local endemics found within a narrow geographic range (Ohashi and Yonekura 2006) and thought to be of derivative origin.

Here we aim to reveal the taxonomic status and the evolutionary history of *Salix* in Japan, particularly with respect to species within the subg. *Vetrix/Chamaetia* complex. In the present study, we perform phylogenetic inference covering available native *Salix* species based on multilocus datasets of low-copy nuclear genes. These were then embedded into a phylogenetic network to resolve their evolutionary history at a lower taxonomic level. Finally, we conducted divergence time estimation to reveal the timescale of their diversification.

Materials and Methods

Data collection

This study targeted 26 willow species native to Japan (Table 1, Online Resource 1), which included all species except *S. fuscescens* Andersson, an endemic alpine species narrowly restricted to Mt. Daisetsu in Hokkaido Island. Moreover, we included 18 foreign willows and five (i.e., two native and three foreign) poplars as outgroups. Leaf samples were collected either from botanical gardens, herbarium specimens, or from individuals in the field (Table 1). This required intensive field sampling at several locations for *S. hukaoana*, *S. gracilistyla* Miq. and *S. rorida* Lacksch. to test the hypothesis of past

reticulate evolution. DNA was extracted from leaves using a DNeasy Plant Mini Kit (Qiagen, Maryland, USA) and diluted to a concentration of ~ 1 ng μL^{-1}.

We obtained sequences from three chloroplast intergenic regions (i.e., *trnL-trnF*, *trnR-trnN*, and *atpB-rbcL*) and three low-copy nuclear (COS) genes, i.e., chloroplast-expressed glutamine synthetase (*ncpGS*), glucose-6-phosphate isomerase (*PGI*), and 6-phosphogluconate dehydrogenase (*6PG*). Universal primers (Taberlet et al. 1991; Terachi 1993; Suyama et al. 2000) were used to amplify chloroplast regions, while specific primers for amplification of nuclear genes were designed to target exonal regions (Table 2) using OLIGO version 6.65 (Molecular Biology Insights, Inc.). These primers were developed for targeted sequences of *Populus trichocarpa* Torr. & Gray retrieved from the JGI PhycoCosm database (Grigoriev et al. 2021; https://phycocosm.jgi.doe.gov/phycocosm/home).

Table 2
List of the amplifying and reading primers developed for this study.

	Gene	Type	Primer Name	Sequence (5'-3')	Location
PGI	glucose-6-phosphate isomerase	amplifying/reading primer (Forward)	Poptr_PGI + 2151	AAATGTAGATCCTATTGATGTTG	CDS(exon)
		amplifying/reading primer (Reverse)	Poptr_PGI – 2976	GCTGATCAATGCTTGATGCTCC	CDS(exon)
		internal primer (Reverse)	Poptr_PGI – 3442	TTGTTAGGATCAATGCCAAACT	CDS(exon)
ncpGS	glutamine synthetase leaf isozyme	amplifying/reading primer (Forward)	Poptr_ncpGS + 1490	GATGCACATTATAAGGCTTG	CDS(exon)
		amplifying/reading primer (Reverse)	Poptr_ncpGS – 2449	AATGTGTTCCTTATGGCGAAG	CDS(exon)
		internal reading primer (Reverse)	Poptr_ncpGS – 2252	GGTGTGGCATCCAGCACC	CDS(exon)
		internal reading primer (Forward) specific for subg. Vetrix/Chamaetia	Poptr_ncpGS + 1848	CAGTATCCTTGTCAAAGATTTG	intron
6PG	6-phosphogluconate dehydrogenase	amplifying/reading primer (Forward)	Poptr_6PG + 67	GCCCTTAATATCGCAGAG	CDS(exon)
		amplifying/reading primer (Reverse)	Poptr_6PG – 1195	TGGCAAGATCAGGATTCCTATCA	CDS(exon)

PCR was performed using a PerkinElmer 9700 Thermocycler in 10µL reaction mixtures. These consisted of ~ 0.5 ng template DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 0.15 µm of each primer and 0.25 U *Taq* polymerase. The PCR conditions were as follows: 94°C for 3 min, followed by 35–40 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 2 min, with a final extension of 72°C for 5 min. PCR products were then purified (ExoSAP-IT, Amersham Biosciences) before being subjected to cycle sequencing using an ABI Big Dye™ Terminator Cycle Sequencing Kit version 3.1 (Applied Biosystems, Foster City, USA), and finally analyzed on an ABI 3100 automated sequencer (Applied Biosystems). Sequences were read in both directions using forward and reverse amplifying primers; if needed, we also used internal sequencing primers designed for this study (Table 2).

Sequence editing and assembly were performed using CodonCode Aligner version 3.7.1 (CodonCode Corporation) to generate consensus sequences. Heterozygous substitutions in nuclear genes were coded using IUPAC ambiguity codes. Sequences with multiple heterozygous indels were not successfully assembled and were therefore excluded from further analysis. All assembled sequences were registered in the DNA Data Bank of Japan (Online Resource 1). Multiple sequence alignment was performed using the ClustalW2 algorithm as implemented in the SeaView alignment editor (Gouy et al. 2010). Nuclear sequences were then phased into haplotypes using the PHASE algorithm as implemented in DNAsp version 6 (Librado and Rozas 2009). To test for non-neutral evolution, haplotypic sequence data for each gene were used to compute haplotype diversity (*Hd*), nucleotide diversity (*Pi*), and Tajima's *D* (Tajima 1989) using DNAsp.

Phylogenetic reconstruction

Phylogenetic trees for each gene ("gene trees") were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) methods. The exonic sequences for nuclear ncpGS and PGI genes were first identified using the JGI PhycoCosm database, then removed to avoid violations of the substitution model. In contrast, the sequences obtained for the gene 6PG, which were all located within exons, were analyzed directly. ML trees were constructed using RAxML version 8.2.10 (Stamatakis 2014) with a raxmlGUI 2 (Edler et al. 2021) graphical interface and applying a GTR + I substitution model. The bootstrap confidence values of nodes were evaluated by generating one thousand bootstrap replicates. The BI method was executed using MrBayes version 3.2.7 (Ronquist et al. 2012). Two independent runs containing four (one hot and three cold) Markov chain Monte Carlo (MCMC) chains were run until the average standard deviation of split frequencies fell below 0.01. We saved trees every 500 generations and discarded the first 10% as burn-in. The optimal substitution models for chloroplast and nuclear ncpGS and PGI genes were then selected by comparing corrected Akaike Information Criterion (cAIC) as calculated by jModelTest version 2.1.10 (Posada 2008). In contrast, we ran a reversible-jump MCMC to explore suitable substitution modes for the exclusively exonic sequences of the 6*PG* gene; here data was partitioned by codon-position.

Finally, the evolutionary relationships of the genus *Salix* were visualized using phylogenetic network analysis. A combined data matrix was generated by concatenating the phased sequences of all genes, including both exonic and intronic sequences, from 68 samples that were associated with a complete data set. A phylogenetic network was constructed using the NeighborNet algorithm based on uncorrected P distance as implemented in SplitsTree version 6 (Huson and Bryant 2006).

Divergence time estimation

Divergence time was estimated using a Bayesian inference method implemented in BEAST version 2.6.7 (Drummond and Rambaut 2007). We targeted only highly variable nuclear genes and excluded exonic sequences to avoid violations of the substitution model violation. Subsequently, the intronic sequences of the ncpGS and PGI genes were only analyzed. We performed species tree analysis using *Beast (StarBEAST) and estimated posterior mean values and 95% highest posterior density (HPD) intervals of divergence time for all nodes via MCMC analysis for 10 million generations. Here we sampled every 5,000 generations and discarded the first 10% as burn-in. For this analysis we used a calibrated Yule model of speciation and an HKY model of nucleotide substitution. The optimal molecular clock model was selected through a Bayes factor analysis; we compared marginal likelihoods of four clock models, including a strict clock (STR), an uncorrelated log-normal relaxed clock (UCLN), an uncorrelated exponential relaxed clock (RCEP), and a random local clock model (RLC). The marginal likelihoods were evaluated using nested sampling with 10 particles, a subchain length of 10,000, and an epsilon value of 1×10^{-6}. The molecular dating was calibrated with the fossil records as follows (Wu et al. 2015). The first calibration point was set at 48 Ma (normal distribution, SD = 0.3) for the root node of Salicaceae *sensu strict.* This was based on an approximately 48-million-year-old fossil from the early Eocene in North America (*"Populus tidwellii"*), which most likely represents the stem lineage leading to *Populus* and *Salix* (Manchester et al. 2006). The second calibration was set at 23 Ma (normal distribution, SD = 0.3) for the root node of the *Vetrix/Chamaetia* clade, based on the earliest reliable *Salix* fossils from Late Oligocene deposits (23 Ma) in Alaska, which were found to be affiliated with subg. *Vetrix* (Wolfe

1987). This method of calibration differed from that used in previous studies (Wu et al. 2015; He et al. 2021), which used the age of the earliest *"Vetrix"* fossils to calibrate the nodes for the divergence between *Vetrix* and *Chamaetia*. Instead, we performed fossil calibration at the most recent common ancestor of the *Vetrix/Chamaetia* clade, since subg. *Chamaetia* has been found to be polyphyletic.

Results

After omitting sequences containing multiple heterozygous indels, which are often detected in *Populus* and subg. *Salix* species, especially in the *ncpGS* and *PGI* genes, we obtained 675 assembled sequences with a total of 5,809 aligned base pairs (bp) from chloroplast and three nucleotide genes. We then submitted these sequences to GenBank (i.e., accession numbers LC757833-758526, Online Resource 1). Subsequently, 713 phased/haplotypic sequences were subjected to further analysis.

Nucleotide diversity ranged from 0.00522 (chloroplasts) to 0.02828 (*ncpGS*), while gene diversity ranged from 0.894 (chloroplasts) to 0.959 (*ncpGS*). Tajima's *D* only significantly deviated for 6*PG* ($p < 0.05$). The diversity and neutrality indices for each gene are summarized in Table 3.

Table 3
Diversity and neutrality indices for each gene. Shown are the number of phased/haplotype sequences analyzed in this study, measures of the nucleotide and gene diversity, and Tajima's *D*.

Measure	*6PG*	*ncpGS*	*PGI*	Chloroplast (concatenated)
Number of phased/haplotypic sequences	252	182	152	127
Number of sites	957	719	1033	3100
Pi: nucleotide diversity (per site)	0.01034	0.02828	0.02383	0.00522
Hd: haplotype (gene) diversity	0.932	0.959	0.953	0.894
Tajima's *D*	−1.91698	−1.49705	−1.4177	−1.4532
Statistical significance	(*) $P < 0.05$	$P > 0.10$	$P > 0.10$	$P > 0.10$

We used Hasegawa–Kishino–Yano models with a gamma distribution (*HKY* + G) as the best-fit model of nucleotide substitution for nuclear intronic sequences in the *ncpGS* and PGI genes. In contrast, a Felsenstein's 1981 model (F81), a transversional model (*TVM*) with invariable sites (*TVM* + I), and a *HKY* + G model were chosen for the chloroplast *atpB-rbcL*, *trnR-trnN*, and *trnL-trnF* intergenic regions, respectively.

Gene trees

The phylogenetic trees for each gene generated by the BL and ML methods are shown in Fig. 1 and Online Resource 2, respectively. The chloroplast phylogeny (Fig. 1a and Online Resource 2a) indicated two diverging clades in *Salix*. One comprises subg. *Salix* (except *S. triandra* L. and *S. eriocarpa* Franch. & Sav.) and subg. *Protitea*, and the other included subg. *Vetrix/Chamaetia*, with *S. triandra* as its first diverging clade, followed by the divergence of *S. arbutifolia* (subg. *Chosenia*) and *S. cardiophylla* (subg. *Pleuradenia*). The phylogenetic relationships at lower taxonomic levels were not resolved, except that *S. hukaoana* was found to be located within a single lineage with some *S. rorida* specimens. *S. eriocarpa* (subg. *Salix*) was located close to *S. udensis* Trautv. & Mey. in this phylogeny as well as in trees for the other genes (therefore not specifically mentioned below).

The exonic sequences of 6*PG* also provided poorly resolved phylogenies (Fig. 1b and Online Resource 2b). Although our data confirmed the early divergence of subg. *Salix* and subg. *Protitea*, some operational taxonomic units (OTUs; i.e., *S.*

triandra and phased haplotypes of *S. jessoensis* Seemen) fell within the *Vetrix/Chamaetia*-dominated clade. The phylogenetic relationships within the *Vetrix/Chamaetia* clade were not resolved, except that *S. triandra* and *S. japonica* were shown to be monophyletic.

In contrast, intronic sequences in nuclear COS genes (i.e., *ncpGS* and *PGI*) provided highly resolved phylogenies, although the comparatively lower success of sequence assembly reduced the number of analyzed taxa, particularly for subg. *Salix*. Results for both genes indicated that a major clade was comprised of subg. *Vetrix/Chamaetia* along with *S. arbutifolia, S. cardiophylla*, and *S. triandra* as early diverging lineages.

The *PGI* gene tree (Fig. 1c and Online Resource 2c) showed the basal divergence of *S. arbutifolia/cardiophylla*, and split subg. *Vetrix/Chamaetia* into two major lineages. One comprised sect. *Viminella*, *Cinerella*, *Sieboldianae*, and *Hastate* (subg. *Vetrix*) as well as subg. *Chamaetia*. The other lineage included sect. *Hukaoana*, *Daphnella*, *Subviminales*, and *Helix*, which further fell into their respective clades. We note that our analysis suggested that *S. miyabeana* Seemen (sect. *Helix*) was polyphyletic, with subsp. *miyabeana* placed near sect. *Viminella*, positioned distantly from subsp. *gilgiana*. Moreover, another specimen of subsp. *miyabeana* that was not successfully assembled (and was therefore excluded from analyses) showed a mixed-base sequence that may be an intersectional hybrid between sect. *Helix* and sect. *Viminella* (Online Resource 3). Species monophyly was supported for *S. chaenomeloides* Kimura, *S. cardiophylla*, *S. japonica*, *S. gracilistyla*, and *S. hukaoana*.

The *ncpGS* gene tree (Fig. 1d and Online Resource 2d) was reconstructed using *S. chaenomeloides* as an outgroup. This tree confirmed the early divergence of *S. triandra*, and further divided all other trees into two lineages. One comprised sect. *Cinerella, Hastatae, Incubaceae*, and *S. nakamurana* Koidz. (subg. *Chamaetia*) along with *S. arbutifolia* and *S. cardiophylla*. Nonindigenous subg. *Chamaetia* (i.e., *S. myrtilloides* L.) was also nested within this clade. The other included sect. *Viminella* along with *S. miyabeana* subsp. *miyabeana* (sect. *Helix*), *S. nummularia* Andersson (subg. *Chamaetia*), and a nested clade including sect. *Hukaoana*, *Daphnella*, *Subviminales*, and *Helix* (except *S. miyabeana* subsp. *miyabeana*), which further fell into their respective clades. Species monophyly was supported for *S. chaenomeloides*, *S. cardiophylla*, *S. japonica, S. myrtilloides*, *S. gracilistyla*, *S. rorida*, and *S. hukaoana*.

Phylogenetic network

Next, we performed a NeighborNet analysis involving 67 samples from 24 species for which successfully assembled sequences of all genes were available. This dataset therefore lacked some native species (i.e., *S. subopposita* Miq., *S. sieboldiana* Blume, *S. reinii* Seemen, and *S. taraikensis* Kimura). Nevertheless, the phylogenetic network (Fig. 2) effectively resolved the taxonomic status of Japanese *Salix* species—especially subg. *Vetrix/Chamaetia*—on the sectional to species levels. Moreover, the network represented the three major lineage groups (Group I-III) that Japanese species of subg. *Vetrix/Chamaetia* evolved into. The first group (Group I) comprised sect. *Subviminales*, *Daphnella*, *Helix*, and *Hukaoana*, and their divergence was distinct without significant reticulation. The second group (Group II) consisted of sect. *Viminella* along with the alpine dwarf willow *S. nummularia* (subg. *Chamaetia* sect. *Herbella*) diverging at its basal position. *S. miyabeana* subsp. *miyabeana* (sect. *Helix*) joined in this group, as was positioned distantly from subsp. *gilgiana*. Moreover, a sample identified as *S. eriocarpa* (subg. *Salix*) was found to be almost identical with *S. udensis* (therefore denoted with "?"). The third group (Group III) included sect. *Cinerella* and *Hastatae* along with the alpine dwarf willow *S. nakamurana* (subg. *Chamaetia* sect. *Glaucae*) diverging from its root position. Sect. *Cinerella* and *Hastatae* showed progenitor-derivative relationships, insofar as species of sect. *Hastatae* (i.e., *S. japonica*, *S. rupifraga*, and *S. shiraii*: all endemic species) all descended from *S. vulpina*.

Divergent time estimation

The Bayes factor analysis favored the uncorrelated exponential relaxed clock (RCEP) model (Online Resource 4). The estimated divergence times at representative nodes are presented in Fig. 3 and Table 4. They suggest that the

Helix/Daphnella/Hukaoana/Subviminales lineage group was diversified at 18 Ma, followed by the divergence of *S. hukaoana* at 14 Ma. The endemic species group (classified as sect. *Hastatae*) was estimated to have diverged 7.8 Ma, with diversification starting around 4.5 Ma, and subg. *Chamaetia* was estimated to have appeared before 15 Ma.

Table 4
Mean divergence time estimates (Mya) of representative nodes (i.e., lettered nodes in Fig. 3) for Japanese native *Salix* species and lineage based on nuclear intronic sequences of the *ncpGS* and *PGI* genes. The 95% highest posterior density (HPD) interval is shown in parentheses.

Node	Event	Mean Divergence Time (95% HPD) (Ma)
A	Root node of Salicaceae *	48.02 (47.47–48.64)
B	Crown age of Salicaceae (Divergence of *S. chaenomeloides*)	44.22 (37.00–48.51)
C	Divergence of *S. triandra*	34.07 (25.37–42.48)
D	Divergence of *S. arbutifollia/cardiophylla* (Divergence of *Vetrix/Chamaetia*)	31.03 (24.46–39.22)
E	Divergence between *S. arbutifollia* and *S. cardiophylla*	12.93 (3.00–26.24)
F	Crown age of subg. *Vetrix/Chamaet*ia *	22.95 (22.41–23.57)
G	Crown age of *Daphnella/Subviminalis/Helix/Hukaoana* clade	17.94 (11.95–23.41)
H	Divergence of *S. hukaoana*	14.10 (7.81–19.92)
I	Divergence of *S. gracilistyla*	11.18 (5.88–16.67)
J	Crown age of sect. *Helix*	6.61 (2.72–12.67)
K	Divergence of Fossa Magna element (endemic species of subg. *Hastatae*)	7.75 (2.99–12.75)
L	Divergence between *S. japonica* and *S. rupifraga/shiraii*	4.49 (1.31–8.56)
M	Divergence of subg. *Chamaetia*	15.60 (8.89–23.48)
N	Crown age of sect. *Viminalis*	3.69 (0.88–7.25)
	* Calibrated nodes	

Discussion

The use of low-copy nuclear genes has recently become more widely used (Sang 2002; Zimmer and Wen 2013) as a tool for robust phylogenetic reconstruction and taxonomic resolution. Although individual loci may behave differently due to incomplete lineage sorting and interspecific hybridization, phylogenetic inference based on multilocus sequences has proven to be a robust tool for addressing evolutionary relationships (Huson and Scornavacca 2011).

None of the loci used in this study were able to provide high enough phylogenetic resolution to resolve the evolutionary history of *Salix* on the species or sectional levels. Low variability in chloroplast and exonic (6*PG*) gene sequences failed to disentangle the phylogenetic relationships within subg. *Vetrix/Chamaetia* (Fig. 1a,b; Online Resource 2a,b). Moreover, although the highly variable intronic sequences of the *ncpGS* and *PGI* genes generated well-resolved trees, they exhibited significant phylogenetic incongruence even at higher taxonomic levels—e.g., the placement of sections and the monophyly of subg. *Vetrix/Chamaetia* (Fig. 1c,d; Online Resource 2c,d). However, the NeighborNet analysis based on concatenated sequences (Fig. 2) resolved phylogenetic relationships of *Salix* well, although this analysis was limited to species with successfully assembled sequences.

Many previous studies have documented the existence of two major clades within the genus *Salix*. Usually, one includes subg. *Vetrix/Chamaetia* along with subg. *Chosenia/ Pleuradenia* (*S. arbutifolia/cardiophylla*) and sect. *Amygdalinae* of subg. *Salix* (e.g., *S. triandra*) (Azuma et al. 2000; Lauron-Moreau et al. 2015; Acar et al. 2022). However, the basal relationships among these subgenera had been incongruent on the genetic level: nuclear genes suggested a nested position within a non-*Vetrix/Chamaetia* taxon. The results of this study are suggestive of reticulate evolution due to hybridization and/or incomplete lineage sorting in the early divergence stage of these subgenera. Furthermore, the results of this study provide sufficient resolution to unravel the evolutionary history of Japanese willows of subg. *Vetrix/Chamaetia* on the sectional to species levels, especially the speciation processes of key endemic species (discussed below).

Ancient origin of S. hukaoana and related lineages

Previous studies have proposed that there is a high degree of relatedness among *Helix, Daphnella*, and *Subviminales* (Wagner et al. 2018, 2021; He et al. 2021). Here, we found that *S. hukaoana* (sect. *Hukaoana*) also belongs in this group (Group I). This group harbors morphologically diverse species, ranging from shrubs and tall trees, with lanceolate to elliptical and opposite to alternate leaves. Although there are no obvious synapomorphies, several traits are partially shared. This lineage group likely emerged during the early evolutionary stage of subg. *Vetrix/Chamaetia* and diverged into variable sections by the middle Miocene (Table 3). This finding was congruent with fossil records from the late Middle Miocene in Japan, which was thought to reflect extant species from sect. *Helix* and *Subviminales* (Narita et al. 2020).

Moreover, this group is characterized by a clear divergence among sections without apparent reticulate networks; this disconfirms the hybrid speciation hypothesis of *S. hukaoana*. *S. hukaoana* is an endemic species discovered in 1972 (Kimura 1973), that was later revealed to be a major component of mountainous riparian forests in northern Honshu Island (Kikuchi and Suzuki 2010). Although once considered to be related to sect. *Daphnella*, it was given a novel monotypic section *Hukaoana* based on its unique morphological traits (Kimura 1974). Our preliminary assumption for the hybrid origin of *S. hukaoana* (i.e., *S. gracilistyla* x *S. rorida*) was based on its morphological traits and its distribution. In particular, we note that: (1) connate stamens in male flowers are common to *S. hukaoana* and *S. gracilistyla* (and also to sect. *Helix*), (2) the yellow inner-bark is shared by *S. hukaoana* and sect. *Daphnella* (including *S. rorida*), (3) there is sympatric occurrence and hybrid formation between *S. hukaoana* and *S. gracilistyla* (*S.* × *sigemitui*), and (4) allopatrically distributed *S. hukaoana* and *S. rorida* occupy similar ecological niches, with both being tall trees in upper mountainous riparian forests in northern Japan.

Our results therefore indicate the phylogenetic distinctness of *S. hukaoana*, with an ancient origin and an approximate derivation of 14 Ma, suggests that the intermediate traits of *S. hukaoana* are not a byproduct of post-divergence hybridization, but are rather a combination of apomorphic and plesiomorphic characters. However, there remains the possibility of secondary hybridization and introgression between *S. hukaoana* and *S. rorida*, as indicated by the presence of shared haplotypes/genotypes in chloroplast and PGI genes (Fig. 1; Online Resource 2).

Evidence of genomic introgression in S. miyabeana subsp. miyabeana

Subg. *Vetrix* of Group II exclusively consists of narrow-leaved riparian species. Here we found that *S. miyabeana* subsp. *miyabeana* (sect. *Helix*) is distantly related to subsp. *gilgiana* (Group I) but is closely related to *S. schwerinii* E. Wolf. Consistent results from multiple samples rejected the possibility of misidentification.

S. miyabeana subsp. *miyabeana* is a type subspecies that can be identified by more highly elongated leaves, shorter styles, and (almost sessile) ovary stipes relative to subsp. *gilgiana.* Subsp. *miyabeana* is more northerly distributed (i.e., Hokkaido Island) than *gilgiana* (i.e., the south tip of Hokkaido to Honshu) and has a greater opportunity to grow sympatrically with *S. schwerinii* (northern Japan). We argue that intersectional hybridization and introgression with *S. schwerinii* has caused discordant phylogenetic signals for *S. miyabeana*, and may have caused morphological differentiation between these

subspecies. A short heterozygous sequence of the PGI gene obtained from *subsp. miyabeana* provides evidence supporting this argument (Online Resource 3). In contrast, the possibility of misidentification or specimen contamination cannot be denied for the disputable position of "*S. eriocarpa*," since its sequences were completely identical with *S. udensis*. For this reason it is noted with a question mark.

Radiative speciation of local endemic species of "sect. Hastatae"

Subg. *Vetrix* in Group III comprises round-leaved hillside willows of sect. *Cinerella* and *Hastatae*. Although not covered by the NeighborNet analysis, the hillside-willow section *Sieboldiana* may also be part of this group, as can be inferred from the placement of *S. reinii* in the PGI gene tree (Fig. 1c; Online Resource 2c), along with a recent phylogenomic study (Gulyaev et al. 2022) that showed a close relationship between *S. sieboldii/reinii* and sect. *Cinerella*.

In addition, the placement of *S. taraikensis* (sect. *Cinerella*) was also unclear: only inferences from PGI gene data showed its early divergence, although this was still close to other *Cinerella* species (Fig. 1c; Online Resource 2c). This may be noteworthy, since previous studies have identified the early divergence of *S. starkeana* Willd. (Wagner et al. 2020, 2021), which was classified together with *S. taraikensis* into subsect. *Substriatae* Goerz by Eurasian taxonomists (Skvortsov 1999).

Group III represents an ancestor-descendant relationship between subg. *Cinerella* and *Hastatae*, with the latter being a derived lineage that diverged from *S. vulpina* around 7.8 Ma and subsequently diversified at 4.5 Ma into three species (Fig. 3, Table 4). First, it should be noted that the classification of subg. *Hastatae* (i.e., *S. japonica*, *S. shiraii*, and *S. rupifraga*; Ohashi 2000, 2001) may be invalid, since Wagner et al. (2021) have proven the polyphyly of this section. Nevertheless, our results provide significant insight into the plant speciation process within the Japanese Archipelago: Three species of "subg. *Hastatae*" are all local endemics affiliated to a biogeographic group called the "Fossa Magna element." This is a group of endemic taxa distributed within a narrow range of the Fuji Volcanic Zone (extending from the Izu Islands to central Honshu), that is considered to have evolved via adaptation to volcanic environments (Takahashi 1971). These willows exhibit divergent ecological and morphological traits ranging from hillside shrubs to alpine dwarf species with free to connate stamens (Ohashi and Yonekura 2006).

That a single radiating lineage led to the establishment of these endemic willows suggested the occurrence of adaptive radiation, with the divergence time being highly consistent with the formation of the volcanic zone in Japan (Maruyama et al. 1997). The timing of divergence was also similar to the timing of the divergence of *Rubus trifidus* Thunb., another species of the Fossa Magna element(Kikuchi et al. 2022), which suggests that tectonic activities may be drivers of plant speciation.

Polyphyly and ancient divergence of subg. Chamaetia

Although the polyphyly of subg. *Chamaetia* had already been proven by previous studies (Lauron-Moreau et al. 2015; Wagner et al. 2018), this study further revealed the basal and ancient divergence of *Chamaetia* species (i.e., *S. nakamurana* and *S. nummularia*) by the middle Miocene (Table 4). This estimate agrees with the North American fossil record (Wolfe 1987), which shows the emergence of several sections of subg. *Chamaetia* during the Miocene.

Although the placement of *S. nummularia* (sect. *Herbella*) near sect. *Viminella* was supported by Wagner et al. (2021), the position of *S. nakamurana* (sect. *Glaucae*) near sect. *Cinerella* was incongruent with that report for species in sect. *Glaucae* (i.e., *S. pyrenaica* Gouan), suggesting the polyphyly of this section. Moreover, the phylogenetic position of *S. fuscescens* (sect. *Myrtosalix*), the only *Chamaetia* species not covered in this study, remains an issue for future research. Wagner et al. (2021) placed sect. *Myrtosalix* species near sect. *Daphnella*, and we therefore infer that *S. fuscescens* may be nested within Group I.

Conclusion

Despite the methodological limitation of using Sanger sequencing, here we performed phylogenetic analyses using nuclear COS genes and chloroplast sequences to resolve the taxonomic status of Japanese *Salix* species of subg. *Vetrix/Chamaetia* on the sectional to species levels. In doing so, we provide significant insight into their evolutionary relationships and the timescale of diversification. Furthermore, this study highlighted the origins of endemic taxa via various speciation process, including ancient divergence (i.e., *S. hukaoana*), genomic and morphological differentiation through intersectional hybridization and introgression (i.e., *S. miyabeana* subsp. *miyabeana*), and radiative speciation (i.e., *S. japonica*, *S. shiraii* and *S. rupifraga*).

Declarations

Competing Interests: The authors declare that they have no conflicts of interest.

Acknowledgements

We are grateful to Tohoku University Botanical Gardens (Sendai City, Miyagi Pref., Japan) and Mitsuo Yashima for support in obtaining samples from the *Salix* collection. We also thank Prof. Hiroyoshi Ohashi for providing valuable commentary on the taxonomy of *Salix*. We thank Dr. Hiroshi Yoshimaru, Kensuke Yoshimura, Yasuko Kawamata for providing DNA samples from the DNA-barcoding project. This study received financial support from the Japanese Society for the Promotion of Science (JSPS KAKENHI grant numbers 20248017, 25292098, and 24770081).

References

1. Acar P, Değirmenci FÖ, Duman H, Kaya Z (2022) Molecular phylogenetic analysis resolving the taxonomic discrepancies among Salix L. species naturally found in Turkey. Dendrobiology 87:13–26. https://doi.org/10.12657/denbio.087.002
2. Argus GW (1997) Infrageneric Classification of Salix (Salicaceae) in the New World. Syst Bot Monogr 52:1. https://doi.org/10.2307/25096638
3. Azuma T, Kajita T, Yokoyama J, Hiroyoshi O (2000) Phylogenetic relationships of Salix (Salicaceae) based on rbcL sequence data. Am J Bot 87:67–75
4. Chen J, Sun H, Wen J, Yang Y (2010) Molecular phylogeny of Salix L.(Salicaceae) inferred from three chloroplast datasets and its systematic implications. Taxon 59:29–37
5. Cronk Q, Ruzzier E, Belyaeva I, Percy D (2015) Salix transect of Europe: Latitudinal patterns in willow diversity from Greece to arctic Norway. Biodivers Data J 3:1–29. https://doi.org/10.3897/BDJ.3.e6258
6. Dickmann DI, Kuzovkina J (2014) Poplars and willows of the world, with emphasis on silviculturally important species. In: Poplars and willows: Trees for society and the environment. CAB International Rome, pp 8–91
7. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Bol 7:214. https://doi.org/10.1186/1471-2148-7-214
8. Edler D, Klein J, Antonelli A, Silvestro D (2021) raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol Evol 12:373–377. https://doi.org/10.1111/2041-210X.13512
9. Fang CF, Zhao SD, Skvortsov AK (1999) Salicaceae. Flora of China (4: 139-274). In: Wu Z-Y, Raven PH (eds) Flora of China, vol. 4. Science Press; Missouri Botanical Garden Press, Beijing; St Louis, pp 139–274
10. Gouy M, Guindon S, Gascuel O (2010) Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224. https://doi.org/10.1093/molbev/msp259

11. Grigoriev I V., Hayes RD, Calhoun S, et al (2021) PhycoCosm, a comparative algal genomics resource. Nucleic Acids Res 49:D1004–D1011. https://doi.org/10.1093/nar/gkaa898
12. Gulyaev S, Cai XJ, Guo FY, et al (2022) The phylogeny of Salix revealed by whole genome re-sequencing suggests different sex-determination systems in major groups of the genus. Ann Bot 129:485–498. https://doi.org/10.1093/aob/mcac012
13. Hardig TM, Anttila CK, Brunsfeld SJ (2010) A Phylogenetic Analysis of Salix (Salicaceae) Based on matK and Ribosomal DNA Sequence Data. J Bot 2010:1–12. https://doi.org/10.1155/2010/197696
14. He L, Wagner ND, Hörandl E (2021) Restriction-site associated DNA sequencing data reveal a radiation of willow species (Salix L., Salicaceae) in the Hengduan Mountains and adjacent areas. J Syst Evol 59:44–57. https://doi.org/10.1111/jse.12593
15. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–67. https://doi.org/10.1093/molbev/msj030
16. Huson DH, Scornavacca C (2011) A survey of combinatorial methods for phylogenetic networks. Genome Biol Evol 3:23–35. https://doi.org/10.1093/gbe/evq077
17. Kikuchi S, Mimura M, Naruhashi N, et al (2022) Phylogenetic inferences using nuclear ribosomal ITS and chloroplast seq uences provide insights into the biogeographic origins, diversification timescales and trait evolution of Rubus in the Japanese Archipelago. Plant Syst Evol. https://doi.org/10.1007/s00606-022-01810-6
18. Kikuchi S, Suzuki W (2010) Distribution of an endangered willow, Salix hukaoana, on the Japan Sea side of northeastern Honshu. Japanese J Conserv Ecol 15:89–99
19. Kimura A (1973) Salicis nava species ex regione Okutonensi in Japonia. J Japanese Bot 48:321–326
20. Kimura A (1974) De Salicis Hukaoanae Kimura systematico positu. J Japanese Bot 49:46
21. Lauron-Moreau A, Pitre FE, Argus GW, et al (2015) Phylogenetic relationships of American Willows (Salix L., Salicaceae). PLoS One 10:. https://doi.org/10.1371/journal.pone.0121965
22. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187
23. Manchester SR, Judd WS, Handley B (2006) Foliage and fruits of early poplars (Salicaceae: Populus) from the eocene of Utah, Colorado, and Wyoming. Int J Plant Sci 167:897–908. https://doi.org/10.1086/503918
24. Maruyama S, Isozaki Y, Kimura G, Terabayashi M (1997) Paleogeographic maps of the Japanese Islands: Plate tectonic synthesis from 750 Ma to the present. Isl Arc 6:121–142. https://doi.org/10.1111/j.1440-1738.1997.tb00043.x
25. Narita A, Yabe A, Uemura K, Matsumoto M (2020) Late middle Miocene Konan flora from northern Hokkaido, Japan. Acta Palaeobot 60:259–295. https://doi.org/10.35535/acpa-2020-0012
26. Newsholme C (1992) Willows: the genus Salix. Timber Press, Inc.
27. Ohashi H (2001) Salicaceae of Japan. Sci report Toboku Univ 4th Ser Biol 40:269–396
28. Ohashi H (2019) A systematic enumeration of Japanese Salix (Salicaceae)
29. Ohashi H (2000) A systematic enumeration of Japanese Salix (Salicaceae). J Japanese Bot 75:1–41
30. Ohashi H, Yonekura K (2006) Additions and corrections for Salicaceae of Japan 2. J Japanese Bot 81:75–90
31. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–6. https://doi.org/10.1093/molbev/msn083
32. Pučka I, Lazdiņa D (2013) Review about investigations of Salix spp. in Europe. Res Rural Dev 2:13–19
33. Sang T (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit Rev Biochem Mol … 37:121–147
34. Skvortsov AK (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 Univ

Joensuu
35. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
36. Suyama Y, Yoshimaru H, Tsumura Y (2000) Molecular phylogenetic position of Japanese Abies (Pinaceae) based on chloroplast DNA sequences. Mol Phylogenet Evol 16:271–277. https://doi.org/10.1006/mpev.2000.0795
37. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–9
38. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
39. Takahashi H (1971) Fossa Magna element plants. Res Rep Kanagawa Prefect Museum Nat Hist 2:1–63
40. Terachi T (1993) Structural Alterations of Chloroplast Genome and Their Significance to the Higher Plant Evolution. Bull Inst Natl L Util Dev Kyoto Sangyo Univ 14:138–148
41. Wagner ND, 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. Ecol Evol 8243–8255. https://doi.org/10.1002/ece3.4360
42. Wagner ND, He L, Hörandl E (2021) The evolutionary history, diversity, and ecology of willows (Salix l.) in the european alps. Diversity 13:1–16. https://doi.org/10.3390/d13040146
43. Wagner ND, He L, Hörandl E (2020) Phylogenomic Relationships and Evolution of Polyploid Salix Species Revealed by RAD Sequencing Data. Front Plant Sci 11:1–15. https://doi.org/10.3389/fpls.2020.01077
44. Wolfe JA (1987) An Overview of the Origins of the Modern Vegetation and Flora of the Northern Rocky Mountains. Ann Missouri Bot Gard 74:785. https://doi.org/10.2307/2399450
45. Wu J, Nyman T, Wang D-C, et al (2015) Phylogeny of Salix subgenus Salix s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. BMC Evol Biol 15:31. https://doi.org/10.1186/s12862-015-0311-7
46. Zimmer EA, Wen J (2013) Reprint of: Using nuclear gene data for plant phylogenetics: Progress and prospects. Mol Phylogenet Evol 66:539–550. https://doi.org/10.1016/j.ympev.2013.01.005

Figures

Figure 1

Phylogenetic trees (gene trees) based on (a) chloroplast, (b) nuclear 6*PG*, (c) *PGI*, and (d) *ncpGS* gene sequences, as reconstructed using the Bayesian inference (BI) method. The numbers at each node refer to the posterior probabilities.

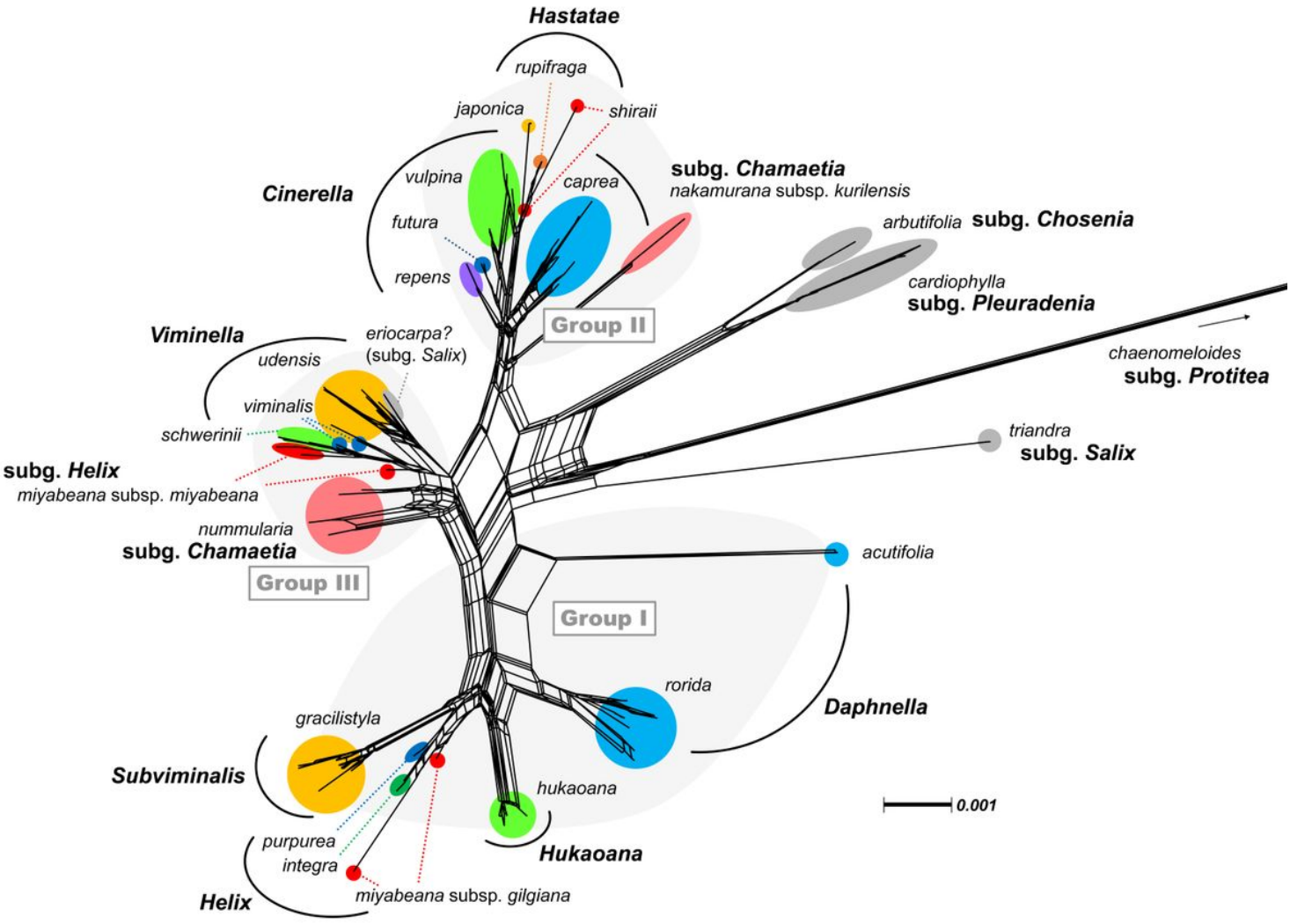

Figure 2

NeighborNet network of *Salix*, focusing on Japanese native species of the subg. *Vetrix/Chamaetia*complex. The network was constructed using uncorrected P distances based on concatenated sequences of chloroplast and three nuclear genes.

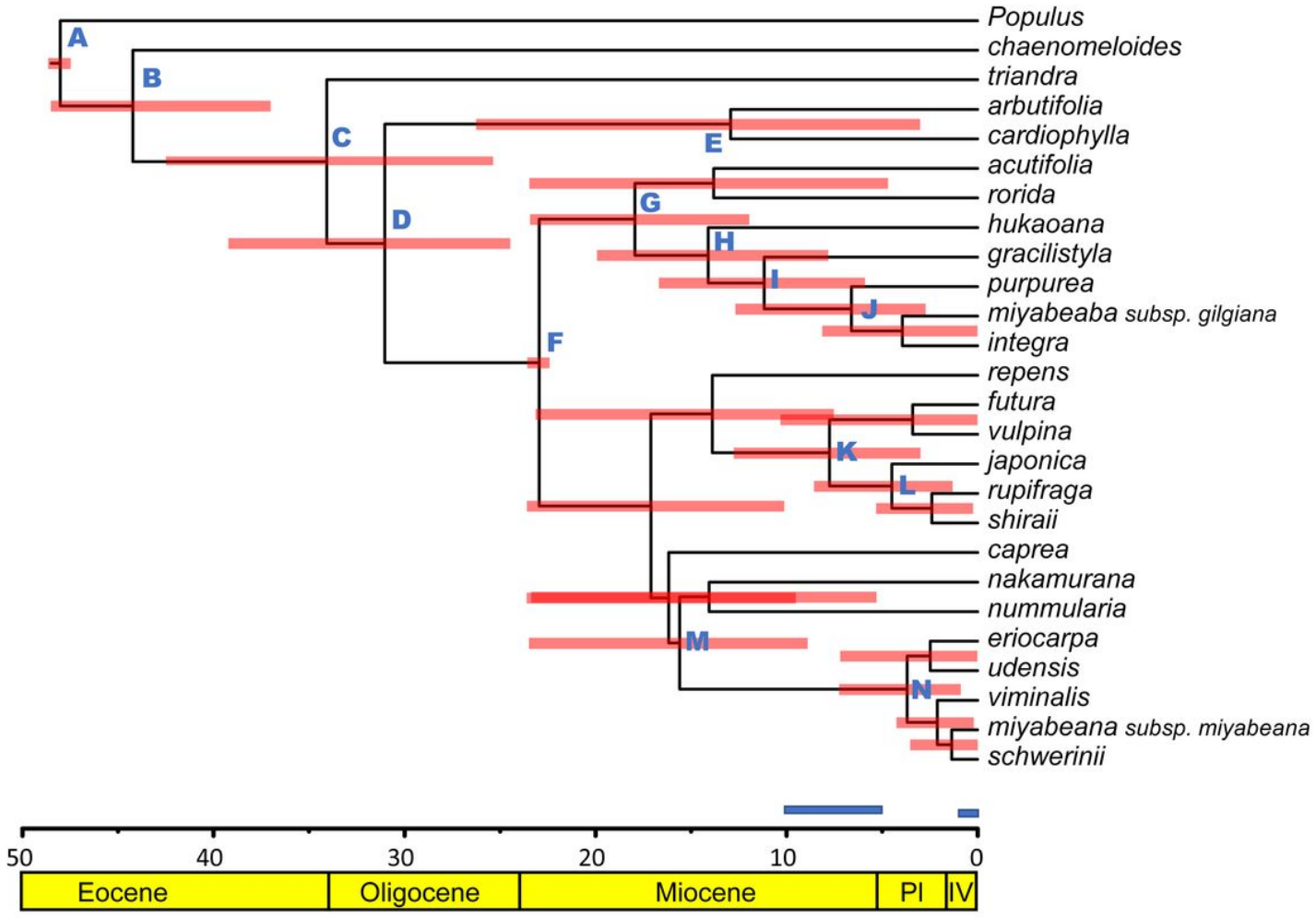

Figure 3

Bayesian divergence time estimates of Japanese native *Salix* species in millions of years ago (Mya), based on nuclear intronic sequences of the *ncpGS* and *PGI* genes. Pink bars at each node indicate the 95% highest posterior density (HPD) interval of divergence time. Mean divergence time and 95% HPD for lettered nodes (A–N) are listed in Table 4. Abbreviation of the periods: Pl—Pliocene, IV—Quaternary.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.xlsx
- ESM2.pdf
- ESM3.pdf
- ESM4.pdf