Complete Chloroplast Genomes of Three *Salix* Species: Genome Structures and Phylogenetic Analysis

Xue-Jiao Zhang †, Kang-Jia Liu †, Ya-Chao Wang, Jian He, Yuan-Mi Wu and **Zhi-Xiang Zhang** *

Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Ecology and Nature, Conservation, Beijing Forestry University, Beijing 100083, China; zxj160501330@163.com (X.-J.Z.); liukangjia@bjfu.edu.cn (K.-J.L.); yachao_wang@bjfu.edu.cn (Y.-C.W.); j.he930724@gmail.com (J.H.); yuanmiwu@163.com (Y.-M.W.)
* Correspondence: zxzhang@bjfu.edu.cn
† These authors contributed equally to this work.

Abstract: High genetic diversity and low differentiation present challenges in taxonomy and systematics of *Salix*. Chloroplast (cp) genome sequencing is efficient for providing new genomic information and elucidating phylogenetic relationships. *Salix spathulifolia* Seemen, *S. cupularis* Rehder, and *S. annulifera* C.Marquand & Airy Shaw are three shrubby willows spread in high-altitude regions in western China. In this study, the integrated circular cp genomes were sequenced and analyzed, and a phylogeny of *Salix* was constructed on the basis of the cp genomes. The results of chloroplast assembly and annotation information were used to characterize genome feature and interspecific variation. The phylogenetic position of the three willows was evaluated using phylogenetic analysis. Full-length cp genomes were 155,566–155,680 bp with a typical double-stranded circular quadripartite structure, containing one large single-copy region (LSC, 84,431–4552 bp), one small single-copy region (SSC: 16,206–16,221 bp), and two inverted repeats (IR: 27,453–27,461 bp). The cp genomes encoded 130 genes, including 8 rRNA genes, 37 tRNA genes, and 85 protein-coding genes. The guanine-cytosine (GC) content of the overall genome was 36.7%. Comparison among the three willows' cp genomes revealed high similarity. Phylogenetic analysis indicated that *S. spathulifolia* was a basal taxon of clade I, while *S. annulifera* formed a monophyletic group with *S. rorida* Laksch.; *S. cupularis* was sister to *S. suchowensis* W.C. Cheng and *S. psammophila* Z. Wang & Chang Y. Yang. The complete chloroplast genomes of the three willows provides an additional sequence-based resource for studying the phylogeny and evolutionary history of Salicaceae.

Keywords: *Salix*; chloroplast genome; sequence variation; comparative analysis; phylogeny

1. Introduction

The chloroplast (cp) is an essential semi-autonomous organelle within the cells of green plants, responsible for carbon fixation, energy conversion, and other important biochemical pathways [1]. Moreover, it possesses its own independent circular genome [2]. The cp genomes of most higher plants are characterized by a quadripartite structure typically consisting of two inverted repeats (IRs) with a length of 20–28 kb, separated by a large single-copy (LSC) region with a length of 80–90 kb and a small single-copy (SSC) region with a length of 16–27 kb. Almost all cp genomes range in size from 120 to 160 kb [3] with 110–130 uniquely encoded genes [4], exhibiting highly conserved gene content and order, with most genes involved in major functions including photosynthesis, transcription, and translation [5,6]. Information provided by structural characteristics of chloroplast genomes can be helpful for inferring phylogenetic relationships among broad sets of plant taxa and enable studies on evolutionary forces that shape the plastome size and structure, such as gains and losses of genes and introns, expansion and contraction of the IRs, and inversion. Furthermore, analyzing the structure information of the chloroplast genome is of great significance for the study and utilization of chloroplast photosynthesis [7]. Furthermore,

chloroplast genomes are predominantly featured by maternal inheritance, but in most conifers are inherited paternally, and transmitted in a biparental manner in a minority of angiosperm species, revealing clues of hybrid speciation [8–10]. In addition, the cp genome can be used in species delimitation and phylogenetic analysis, owing to its small genome size, less recombination, and uniparentally inheritance in most species, as well as its moderate nucleic acid replacement rate compared with the nuclear genome for most angiosperms [6]. However, in contrast to acquiring limited coverage of cp genomes for small samples by conventional DNA sequencing technology, such as Sanger sequencing, in recent years, continuous development of sequencing technology has provided massive sequencing reads from multiple samples simultaneously; moreover, the number of sequenced chloroplast genomes has increased rapidly with low cost.

The genus *Salix* L. (Salicaceae) is the largest genus of Salicaceae, with the number of species ranging from 450 to 520, which can be divided into two categories according to their life forms: trees and shrubs [11–13]. *Salix* is widely distributed in the vast north temperate regions of Europe, Asia, and North America, and extends to tropical and subtropical alpine and Arctic regions [14]. Fast-growing shrub willows (*Salix* spp.) play an important role in the restoration of ecosystem habitats after disturbance and water and soil conservation, serving as important indicators of riparian habitats, and are also considered important energy crops [15,16], thus, portraying the cp genome of shrub willow will help to improve energy production and fully exploit the production and application value of this economic species.

Salix spathulifolia Seemen, *Salix cupularis* Rehder, and *Salix annulifera* C.Marquand & Airy Shaw (Salicaceae) are three shrubby willows endemic to western China, with the first two mainly distributed across Shaanxi, Gansu, Qinghai, northern Sichuan, and other provinces, and the latter mainly distributed across Yunnan and in the eastern Tibet, growing at high altitudes of more than 1800, 2500, and 3400 m, respectively [12]. Morphologically, the three willow species show evolutionary characteristics of shrub lifeform and flowers with two stamens, which may be the characterization of an alpine group that gradually adapted to the alpine cold and arid environment [12,17–19]. However, because the genus *Salix* shows high intraspecies genotypic polymorphism and large phenotypic variation, it is difficult to determine the phylogenetic status of species [11,13]. Currently, the nuclear internal transcribed spacer (ITS) sequences and some plastid DNA sequences have been applied to study the phylogeny of *Salix*, which contain limited phylogenetic information [20–23]. Although Li et al. (2021) employed the complete chloroplast genomes to characterize the phylogenetic relationships of *S. cupularis* in the genus *Salix*, the authors chose *Ginkgo bioloba* L. as the outgroup rather than a more related species [24]. Moreover, research on these three willows is mainly focused on morphology and ecology [25], but their phylogenetic position and evolutionary relationship remain unclear.

Here, we report three complete chloroplast genomes of *S. spathulifolia*, *S. cupularis*, and *S. annulifera*, by analyzing the structure and sequence characteristics of cp genomes and comparing them to each other. Furthermore, the construction of a phylogenetic tree of chloroplast genomes of 21 *Salix* species could shed light on their evolutionary relationship.

2. Materials and Methods

2.1. Plant Material

The experimental materials of *S. spathulifolia*, *S. cupularis*, and *S. annulifera* were obtained from fresh leaves dried with silica gel in the western Chinese provinces of Shaanxi and Xizang at an altitude of at least 2400 m. All the voucher specimens were deposited in the Herbarium of the Beijing Forestry University (Table 1).

2.2. DNA Extraction and Genome Sequencing

Whole-genomic DNA of all three species was extracted according to the CTAB method. After testing the quality of extracted DNA, a total amount of 1.5 µg DNA per sample was used to construct a 150 bp paired-end library, and the library was sequenced by the Illumina

HiSeq4000 platform with an insert size of around 350 bp at Shanghai Majorbio Bio-pharm Technology Co., Ltd, Shanghai, China. The raw data after sequencing the chloroplast genomes of three willows were taken from Gulyaev et al. in preparing [26]. In order to ensure the quality of subsequent analysis, we used Fastp with default settings to remove low-quality sequences and adapters from the raw data [27]. We deposited the samples of the complete chloroplast genome sequences to the NCBI GenBank (accessions MZ365445, MZ365446, MZ365447).

Table 1. Basic information on the sampling sites of the studied willow species.

Species	Collecting Number	Longitude	Latitude	Altitude	Province	Herbaria
Salix spathulifolia	HLCS19_15	108.80° E	33.86° N	2456 m	Shaanxi	BJFC
Salix cupularis	HLCS19_45	107.81° E	34.00° N	3377 m	Shaanxi	BJFC
Salix annulifera	ZZX2019091207	94.97° E	29.47° N	3559 m	Tibet	BJFC

2.3. Chloroplast Genome Assembly and Annotation

We used Geneious 10.2 to conduct the entire chloroplast genome assembly process [28], mainly following He et al. [29]. Low-quality sequencing bases at both ends were clipped out with an error probability limit of 0.05. Then, the Map to Reference function was used to map filtered reads which excludes nuclear and mitochondrial reads to published plastid genomes of *S. babylonica* L. (MF189167) [30], *S. chaenomeloides* Kimura (MG262362), *S. rehderiana* C.K.Schneid. (MG262367), *S. rorida* (MG262368), and *S. taoensis* Goerz ex Rehder & Kobuski (MG262369), which were set as references (iterate up to 10 times). The subsequent chloroplast reads were used for de novo assembly of the contigs with a medium-low sensitivity setting. Usually, there is only one contig (approximately 130 kb) that results from de novo assembly. The filtered reads were repeatedly mapped to the contigs to increase length; if more than one smaller contig was obtained, overlapping regions were used to enable all the contigs to be concatenated to a 130 kb contig with four junctions between two single-copy and IRs regions being confirmed. The Find Repeats Plug-in function was used to define the IR regions, which were copied and inverted manually to construct complete chloroplast genome sequences.

Plastid Genome Annotator (PGA) was used to perform initial annotations of complete chloroplast genome [31] with *Salix suchowensis* [32] as a reference and manually corrected with Plann 1.1.2 [33] and Geneious [28]. Finally, the circular graphical map of the cp genomes was drawn by the OGDRAW online tool [34].

2.4. Repeat Sequence and Codon Usage Analysis

We applied MISA-web to predict repeat sequences of chloroplast microsatellites (cpSSRs) [35] with parameters set to: minimum number of eight repeated motifs for mononucleotide, five repeated motifs for dinucleotide, four repeated motifs for trinucleotide, and three repeated motifs for tetra- and pentanucleotide repeats. Online REPuter software was employed to discover interspersed repeats in whole cp genomes, which included forward, reverse, complement, and palindromic repeats [36] with a minimum repeat size of 30 bp and the Hamming distance set to three. We used Geneious 10.2 [28] to calculate the percent of guanine-cytosine (GC). The relative synonymous codon usage (RSCU) determining the preference for the use of a codon was generated using MEGA 7 [37] based on 85 protein-coding genes (PCG) sequences. An RSCU not greater than 1.0 means no preference, while an RSCU greater than 1 means preference.

2.5. Genome Comparison and Sequence Divergence

Three willow cp genomes were globally compared with the Shuffle-LAGAN mode in online mVISTA software with default settings [38] by employing the *S. spathulifolia* as an annotation reference. Moreover, we conducted a DNA polymorphism analysis to detect nucleotide diversity (π) with a 100 bp window size and a 25 bp step size by using DnaSP

version 6 [39]. IRscope online [40] was used to analyze and visualize the borders between LSC/IRs and SSC/IRs among the three willow cp genomes.

2.6. Phylogenetic Research

The maximum likelihood (ML) phylogeny was applied to investigate the phylogenetic status of *S. spathulifolia*, *S. annulifera*, and *S. cupularis*. The cp genome sequences of 18 *Salix* and two *Populus* species were obtained from the NCBI genome database (National Center for Biotechnology Information). After manually checking the region direction, a total of 20 cp genomes along with three target willows were ultimately aligned using MAFFT online version [41] and then by Phylosuite V1.2.2 [42]. The ML tree was estimated using the best nucleotide replacement model GTR + F + I estimated by ModelFinder [43] implemented in Phylosuite [42]. The two *Populus* species were set as outgroups. We used bootstrap replicates of 5000 to test the confidence of each branch.

3. Results and Discussion

3.1. Genome Features

Full-length complete cp genomes of the three *Salix* species were accomplished varying from 155,566 to 155,680 bp and the genome structure comprised two IR regions (range from 27,453 to 27,461 bp), separated by an LSC region (84,431–84,552 bp) and an SSC region (16,206–16,221 bp) (Figure 1, Table 2). The overall GC content for all three willow species was about 36.7%, while IR regions displayed a higher GC content (41.9%) than the GC content of LSC (34.4%) and SSC (31.0%). This GC content distribution pattern seemed to be common in other plants [44], which is caused by a relatively high GC content in rRNA and tRNA genes [45], occupying more area than that of protein-coding genes in IR regions.

Table 2. Summary of chloroplast genome features of *S. spathulifolia*, *S. cupularis*, and *S. annulifera*.

		S. spathulifolia	*S. cupularis*	*S. annulifera*
Accession Number		**HLCS19_15**	**HLCS19_45**	**ZZX2019091207**
Genome	Length (bp)	155,680	155,566	155,626
	GC (%)	36.7	36.7	36.7
LSC	Length (bp)	84,552	84,431	84,499
	GC (%)	34.4	34.4	34.4
	Length (%)	54.31	54.27	54.30
SSC	Length (bp)	16,206	16,217	16,221
	GC (%)	31.0	31.0	30.9
	Length (%)	10.41	10.42	10.42
IR	Length (bp)	27,461	27,459	27,453
	GC (%)	41.9	41.9	41.9
	Length (%)	17.64	17.65	17.64
No. of genes (duplicated in IR)	Genes	130 (19)	130 (20)	130 (21)
	PCGs	85 (8)	85 (9)	85 (10)
	tRNA	37 (7)	37 (8)	37 (9)
	rRNA	8 (4)	8 (5)	8 (6)
	With introns	17 (5)	17 (6)	17 (7)

Notes: GC: guanine-cytosine; LSC: large single-copy; IR: invert repeat; SSC: small single-copy; PCG: protein-coding gene.
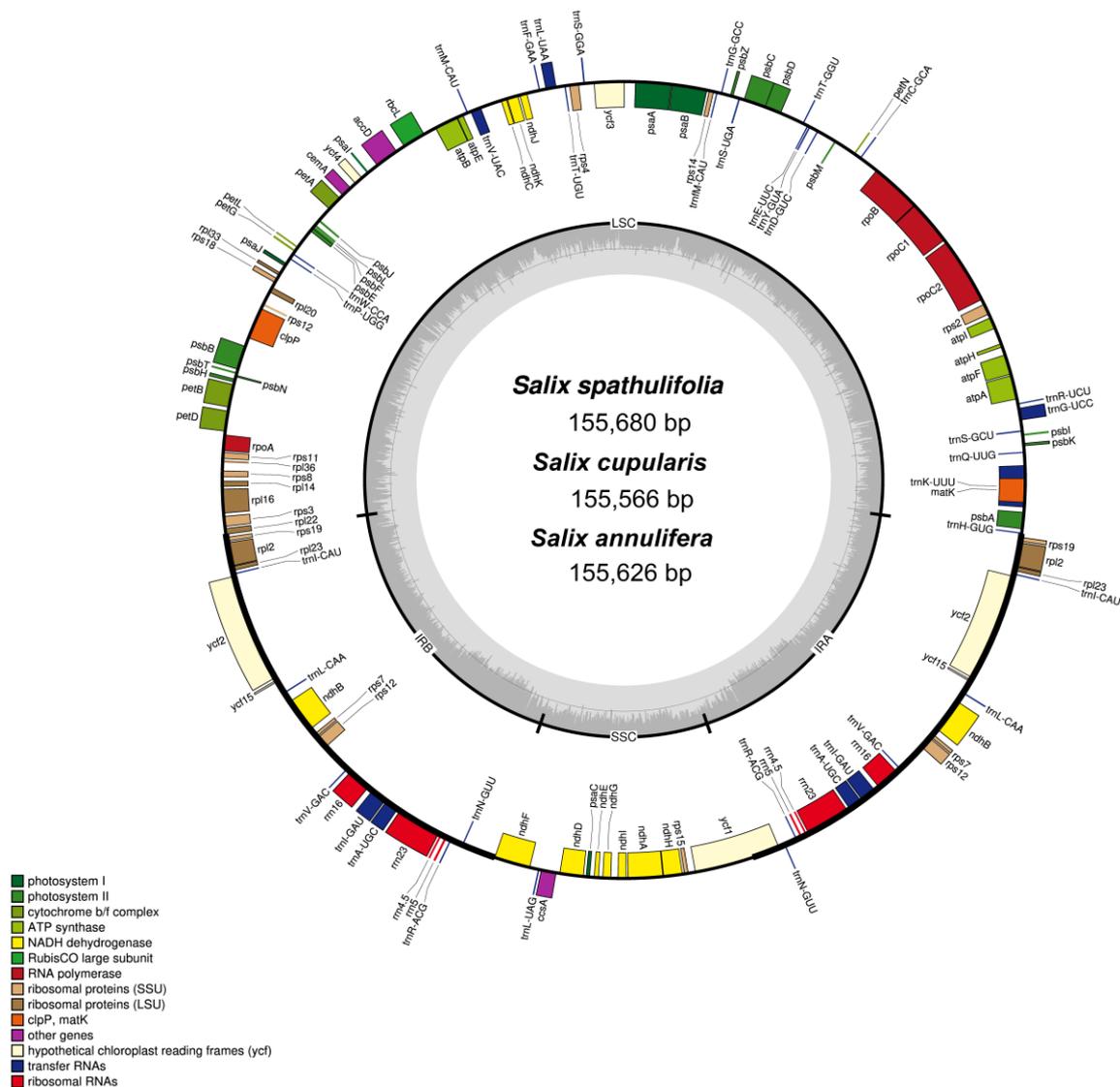
Figure 1. Circular map of cp genomes *S. spathulifolia*, *S. cupularis*, and *S. annulifera*. Genes transcribed clockwise are drawn outside the circle, while those transcribed counterclockwise are drawn inside. Genes are encoded in diversified colors according to their different functional groups. GC content of the chloroplast genome of *Salix spathulifolia* is represented by grey area in the inner circle.

In each of the three cp genomes, 130 genes were predicted (85 protein-coding genes, 37 tRNAs, and 8 rRNAs), of which 19, 20, and 21 genes were duplicated in IR regions, respectively (Table 2). Seventeen genes containing introns were identified, which was the same as in the previous study [30].

According to the genes' functions, they can be divided into 17 groups, of which 57 genes were found to be related to self-replication: 11, 8, 4, 4, 30 genes were responsible for encoding small ribosomal subunit protein, large ribosomal subunit protein, and RNA polymerase subunits, rRNAs, and tRNAs respectively. Another category contained 46 genes related to photosynthesis, involving 7 photosystem I genes, 15 photosystem II genes, 6 cytochrome b/f complex genes, 6 ATP synthase genes, 11 *ndh* genes, and 1 large subunit Rubisco gene (Table 3).

Table 3. Gene composition in the chloroplast genomes of *S. spathulifolia*, *S. cupularis*, and *S. annulifera*.

Groups of Genes	Name of Genes
Transfer RNAs	*trnA-UGC* *, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*, *trnH-GUG*, *trnI-CAU* *, *trnI-GAU* *, *trnK-UUU*, *trnL-CAA* *, *trnL-UAA*, *trnL-UAG*, *trnM-CAU*, *trnN-GUU* *, *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* *, *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* *, *trnV-UAC*, *trnW-CCA*, *trnY-GUA*
Ribosomal RNAs	*rrn4.5* *, *rrn5* *, *rrn16* *, *rrn23* *
Ribosomal protein small subunit	*rps2*, *rps3*, *rps4*, *rps7* *, *rps8*, *rps11*, *rps12* *, *rps14*, *rps15*, *rps18*, *rps19* *
Ribosomal protein large subunit	*rpl2* *, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23* *, *rpl33*, *rpl36*
Subunits of RNA polymerase	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *ycf3*, *ycf4*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Cytochrome b/f complex	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
NDH complex	*ndhA*, *ndhB* *, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Large subunit Rubisco	*rbcL*
Acetyl-CoA carboxylase	*accD*
Maturase	*matK*
Inner membrane protein	*cemA*
ATP-dependent protease	*clpP*
Cytochrome c biogenesis	*ccsA*
Conserved open reading frames	*ycf1*, *ycf2* *, *ycf15* *

Note: "*" means duplicated genes.

Seventeen genes contained introns, six of them were tRNA and the remaining encoded proteins. Among these intron-containing genes, 14 out of 17 included a single intron and 3 genes (*ycf3*, *clpP*, and *rps12*) had two introns (Table 4). In comparison with other introns, the longest intron was within the *trnK-UUU* gene in the LSC region, where the *matK* gene was located, reaching 2547 bp. By contrast, the intron of the *trnL-UAA* gene was the shortest, only 586 bp. The same as most species [46–48], the *rps12* gene was trans-spliced, which encoded the 40S ribosomal protein S12 because one exon in the 5′-end is located in the LSC region and the other two exons in the 3′-end are located in both IR regions. No pseudogenes were found because no stop codons were found in coding sequences.

Table 4. Characteristics of intron-containing genes in the chloroplast genome of *S. spathulifolia*.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Extron III (bp)
trnK-UUU	LSC	37	2547	35		
rpl16	LSC	9	1120	399		
trnG-UCC	LSC	23	693	48		
atpF	LSC	145	741	398		
rpoC1	LSC	453	777	1617		
ycf3	LSC	126	723	228	667	153
trnL-UAA	LSC	35	586	50		
trnV-UAC	LSC	39	609	35		
clpP	LSC	71	837	292	585	228
petB	LSC	6	811	642		
petD	LSC	8	788	490		
ndhA	SSC	552	1107	546		
rpl2	IR	397	668	434		
rps12	LSC&IR	114	~	232	536	26
ndhB	IR	723	682	756		
trnI-GAU	IR	37	949	35		
trnA-UGC	IR	38	802	35		

Features of the *S. cupularis*, *S. annulifera*, and *S. spathulifolia* chloroplast genomes were compared and the results showed that the cp genomes of *S. cupularis* and *S. annulifera* were 114 and 54 bp shorter than that of *S. spathulifolia*, respectively. Compared with *S. spathulifolia*, the LSC and IR region lengths of *S. cupularis* and *S. annulifera* were shorter, while the SSC region lengths were 11 and 15 bp longer. Furthermore, the whole genome GC content among the three willow species was almost the same, and the difference in GC content only exists in the SSC region. Moreover, there was no variation in gene content and order among the three willow species, including intron-containing genes.

3.2. Codon Usage and Repeat Sequence Analysis

A total of 26,115, 26,131, and 26,138 codons were identified in the cp genomes of *S. cupularis*, *S. annulifera*, and *S. spathulifolia*, respectively. The number of codons encoding leucine was the largest and the least abundant were those for cysteine, which was the same as for the other two species. Across the three genomes, there were 30 preferred synonymous codons (RSCU > 1); except for the UUG codon ending with G, the others ended with A or U. Moreover, the use of the two codons AUG and UGG had no preference for codon usage (RSCU = 1) (Table 5).

Table 5. RSCU in *S. spathulifolia*, *S. cupularis*, and *S. annulifera* chloroplast genomes.

Codon	Amino Acid	*S. spathulifolia*		*S. cupularis*		*S. annulifera*	
		Count	RSCU	Count	RSCU	Count	RSCU
UUU	Phe	987	1.31	988	1.31	992	1.31
UUC	Phe	516	0.69	517	0.69	517	0.69
UUA	Leu	892	1.91	894	1.91	894	1.91
UUG	Leu	569	1.22	570	1.22	570	1.22
CUU	Leu	586	1.25	586	1.25	586	1.25
CUC	Leu	178	0.38	180	0.38	179	0.38
CUA	Leu	398	0.85	397	0.85	397	0.85
CUG	Leu	179	0.38	179	0.38	180	0.38
AUU	Ile	1127	1.49	1129	1.49	1130	1.49
AUC	Ile	432	0.57	432	0.57	430	0.57
AUA	Ile	708	0.94	708	0.94	711	0.94
AUG	Met	620	1	621	1	621	1
GUU	Val	493	1.42	492	1.42	491	1.41
GUC	Val	167	0.48	167	0.48	167	0.48
GUA	Val	532	1.53	532	1.53	533	1.53
GUG	Val	199	0.57	199	0.57	199	0.57
UCU	Ser	572	1.69	573	1.69	573	1.69
UCC	Ser	332	0.98	329	0.97	329	0.97
UCA	Ser	409	1.21	408	1.2	409	1.21
UCG	Ser	184	0.54	186	0.55	185	0.55
CCU	Pro	418	1.56	418	1.57	420	1.57
CCC	Pro	198	0.74	197	0.74	199	0.74
CCA	Pro	308	1.15	306	1.15	308	1.15
CCG	Pro	145	0.54	145	0.54	143	0.53
ACU	Thr	522	1.6	524	1.6	524	1.6
ACC	Thr	238	0.73	236	0.72	238	0.73
ACA	Thr	423	1.29	424	1.29	423	1.29
ACG	Thr	126	0.39	126	0.38	124	0.38
GCU	Ala	620	1.83	620	1.83	620	1.83
GCC	Ala	202	0.6	205	0.61	204	0.6
GCA	Ala	385	1.14	385	1.14	385	1.14
GCG	Ala	145	0.43	143	0.42	145	0.43
UAU	Tyr	777	1.64	780	1.65	780	1.64
UAC	Tyr	171	0.36	168	0.35	170	0.36

Table 5. *Cont.*

| Codon	Amino Acid	*S. spathulifolia*		*S. cupularis*		*S. annulifera*	
		Count	RSCU	Count	RSCU	Count	RSCU
UAA	*	45	1.59	45	1.59	45	1.59
UAG	*	22	0.78	22	0.78	22	0.78
CAU	His	477	1.52	478	1.52	477	1.52
CAC	His	150	0.48	150	0.48	149	0.48
CAA	Gln	713	1.55	715	1.55	716	1.55
CAG	Gln	209	0.45	209	0.45	209	0.45
AAU	Asn	978	1.53	981	1.53	981	1.53
AAC	Asn	300	0.47	300	0.47	299	0.47
AAA	Lys	1051	1.47	1052	1.47	1050	1.47
AAG	Lys	376	0.53	376	0.53	375	0.53
GAU	Asp	833	1.58	832	1.58	834	1.58
GAC	Asp	223	0.42	223	0.42	222	0.42
GAA	Glu	1023	1.49	1022	1.49	1023	1.49
GAG	Glu	347	0.51	347	0.51	346	0.51
UGU	Cys	213	1.42	212	1.42	213	1.42
UGC	Cys	86	0.58	87	0.58	86	0.58
UGA	*	18	0.64	18	0.64	18	0.64
UGG	Trp	452	1	452	1	453	1
CGU	Arg	328	1.28	330	1.29	330	1.29
CGC	Arg	109	0.43	108	0.42	109	0.43
CGA	Arg	357	1.4	361	1.41	358	1.4
CGG	Arg	106	0.42	104	0.41	105	0.41
AGU	Ser	408	1.2	408	1.2	407	1.2
AGC	Ser	128	0.38	128	0.38	128	0.38
AGA	Arg	471	1.84	471	1.84	473	1.85
AGG	Arg	161	0.63	161	0.63	161	0.63
GGU	Gly	556	1.25	556	1.25	557	1.26
GGC	Gly	192	0.43	192	0.43	192	0.43
GGA	Gly	711	1.6	711	1.6	709	1.6
GGG	Gly	314	0.71	316	0.71	315	0.71

Note: "*" means terminal codons.

Using MISA-web, we detected 222 SSRs in the three cp genomes, of which the most were mono- and tetra-nucleotide types. In S. spathulifolia, the most abundant SSR type was mononucleotide repeats, with a total number of 197, followed by tetranucleotide (12), dinucleotide (10), and the least abundant were trinucleotide (1), pentanucleotide (1), and hexanucleotide (1) (Table S1 and Figure 2a). The numbers of mononucleotides in S. cupularis and S. annulifera were 199 and 196, respectively; however, the numbers of trinucleotide and pentanucleotide repeats were identical. Variations in the dinucleotide, tetranucleotide, and hexanucleotide repeats were no more than one. No hexanucleotide repeats were identified in the S. cupularis cp genome. In terms of base composition, among the three cp genomes, the most common type of SSRs was A/T type mononucleotide, accounting for 82.88–84.23% (Figure 2b). The longest SSR had a length of 18 bp while the shortest was 8 bp (Table S1). Comparison among the three willows showed high similarity in SSR type and quantity.
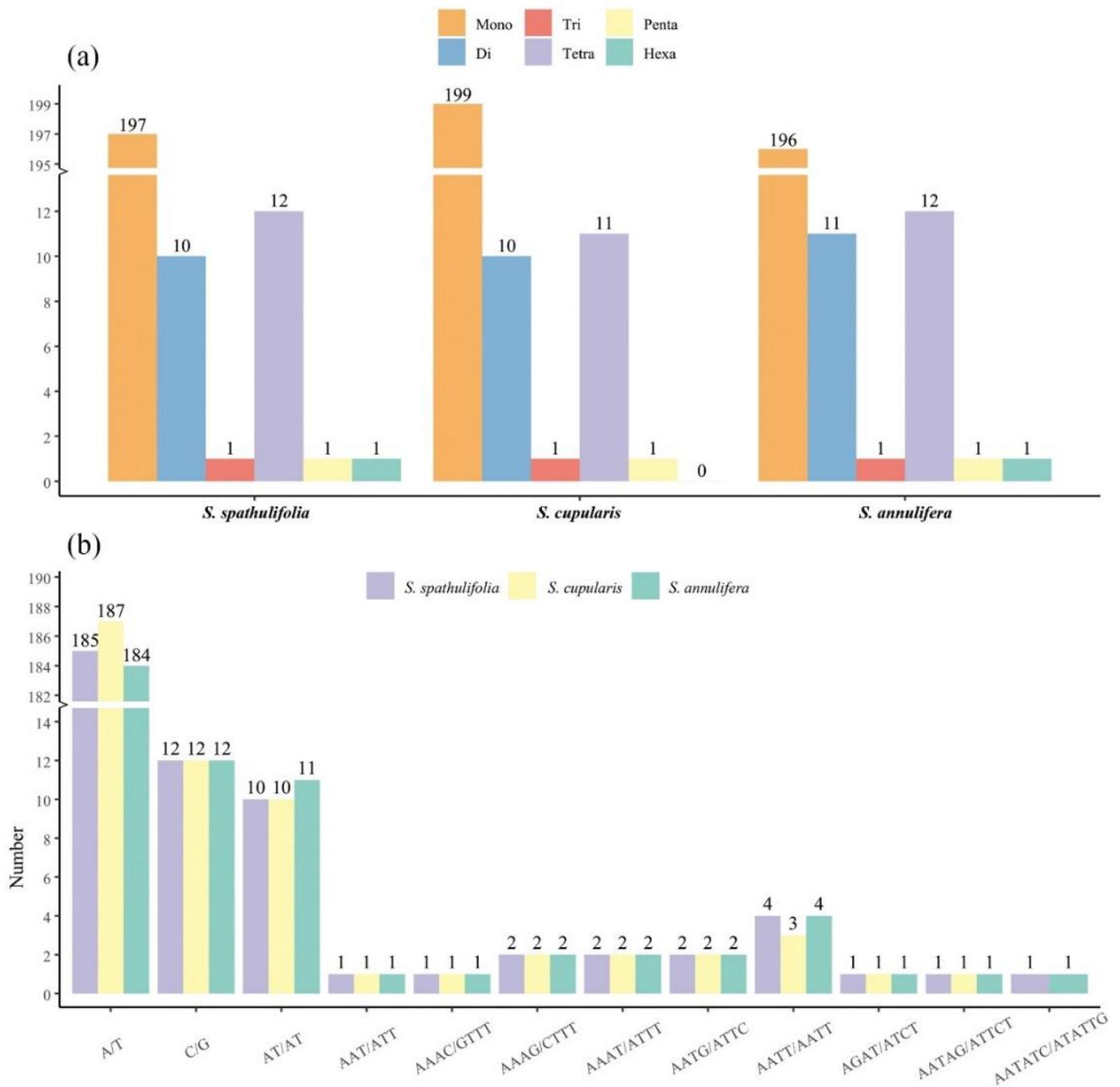
Figure 2. Simple sequence repeats (SSRs) in the chloroplast genomes of *S. spathulifolia*, *S. cupularis*, and *S. annulifera*. (**a**) Number of each repeat type; (**b**) type and number of SSR loci.

Based on the REPuter result, S. spathulifolia had the most repeats (46): 25 forward, 14 palindromic, 4 reverse, and 3 complement (Figure 3). The repetitive sequences ranged in size from 30 to 76 bp with a Hamming distance of 3. The most abundant repeats identified in the cp genome were 30 bp (12 sites) followed by 32 bp (10 sites) (Table S2). Comparatively, *S. annulifera* had the smallest number of repeats both in total (34) and in each group (21 forward, 11 palindromic, 1 reverse, 1 complement). The four repeat types were rank-order arranged according to the number of repeats of each type in three willows: Forward (F)—most; Palindromic (P)—second; Reverse (R)—third; and Complement (C)—least. In addition, no complement repeats were found in the cp genome of S. cupularis.
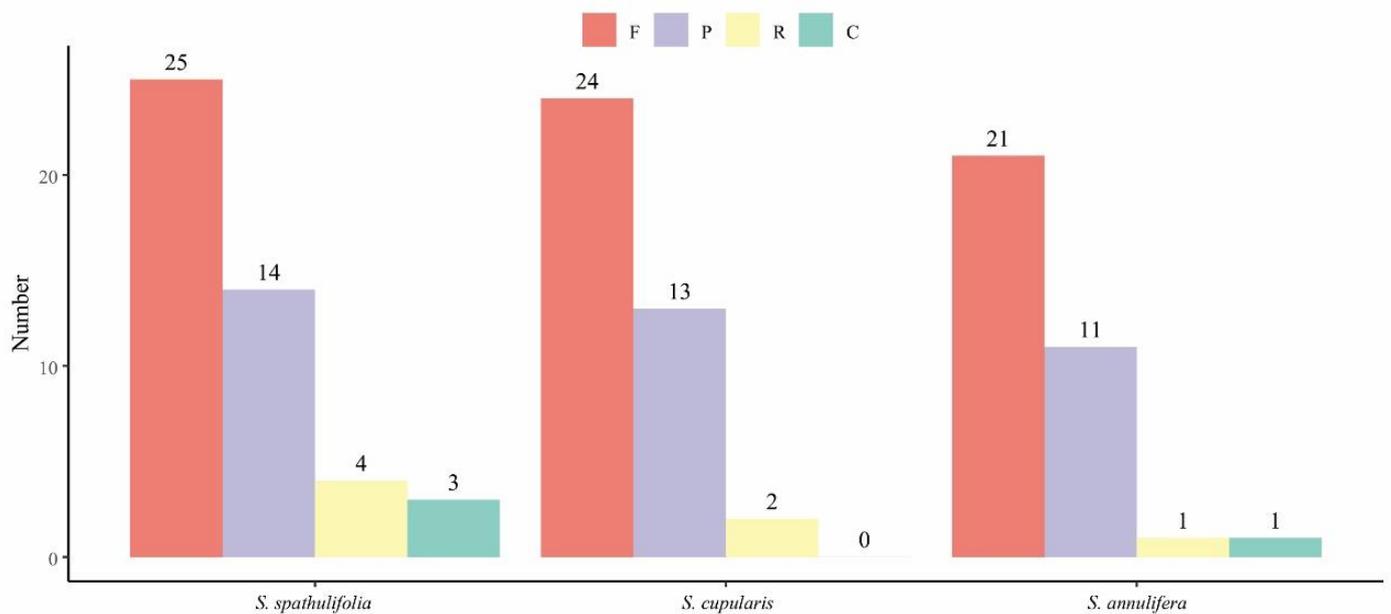
Figure 3. Number of repeats detected in the chloroplast genomes of *S. spathulifolia*, *S. cupularis*, and *S. annulifera*. (F: Forward; R: Reverse; C: Complement; and P: Palindromic).

3.3. Genome Comparison

The chloroplast genomes of the three *Salix* taxa were compared and analyzed using mVISTA online software with *S. spathulifolia* as a reference. Overall, the comparative genomic analysis showed that the three willows' cp genome sequences were relatively conserved. Generally, higher variation was found in non-coding regions. The most divergent regions in the coding sequences were *psbL*, *rpl16*, and *ycf1*. The regions with relatively higher variation were mainly concentrated in the LSC and SSC regions; however, the IR regions were more conserved (Figure 4). Similar results can be revealed by a sliding window analysis, which determines the nucleotide diversity in chloroplasts. The average Pi of the three chloroplast genomes was 0.00075, and high nucleotide variability (π) was exhibited at the SC regions in comparison to IR regions (Figure 5). Due to the difficulty in resolving the phylogeny of the genus *Salix*, these divergent regions—with the highest variation in *Salix* cp genomes—can be identified as a source of potential molecular markers.

The four junctions in the three willows' cp genomes were presented using IRscope (Figure 6). Overall, the distribution of genes on each border was almost identical, with only slight differences among the three willows. The LSC/IRb border was inside the *rpl22* gene in all three willows. The 347 bp fragment of the *rpl22* is located within the LSC region, whereas the remaining 52 bp section of this gene can be found within the IRb region. The IRb/SSC boundary is located inside the *ndhF* gene. The 10 bp fragment of the *ndhF* gene is located within the IRb region, whereas the remaining 2264 bp section of this gene can be found within the SSC region. The *ycf1* gene is located across the SSC/IRa junction. Compared to *S. annulifera*, the parts of the *ycf1* gene in the SSC of *S. cupularis* and *S. spathulifolia* were 6 bp shorter, whereas the parts of the *ycf1* gene in the IRa region were the same (1748 bp). The *rps19* gene was fully located in the IRa, while the *trnH* gene was fully located in the LSC, 1 bp away from the IRa/LSC border.
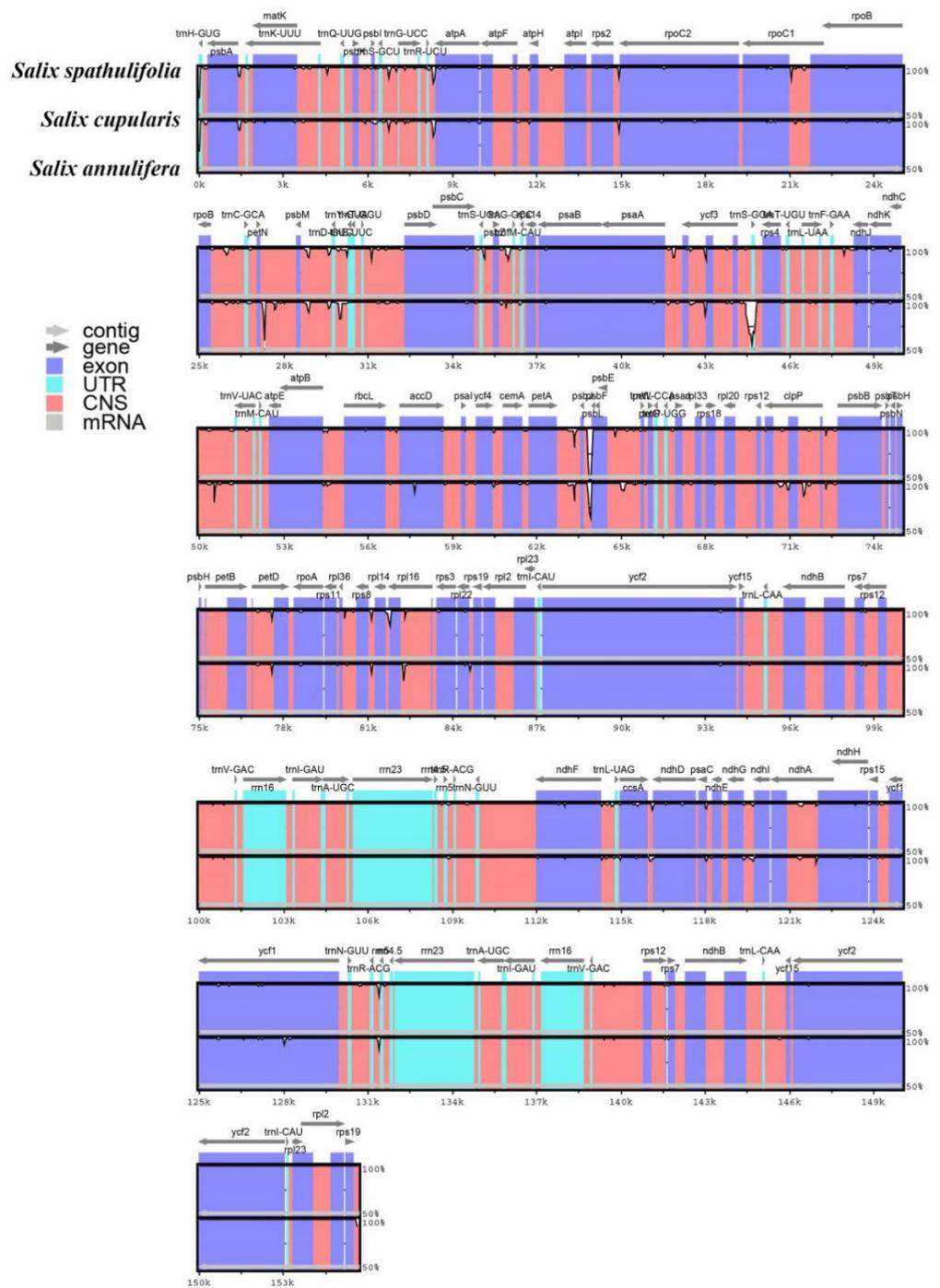
Figure 4. Sequence alignment of the three complete chloroplast genomes generated by mVISTA, with *S. spathulifolia* as reference. The *x*-axis represents the sequence length and the *y*-axis indicates the identity from 50 to 100%. Arrows indicate annotated genes and the direction of their transcription. UTR: untranslated regions; CNS: conserved non-coding sequences.
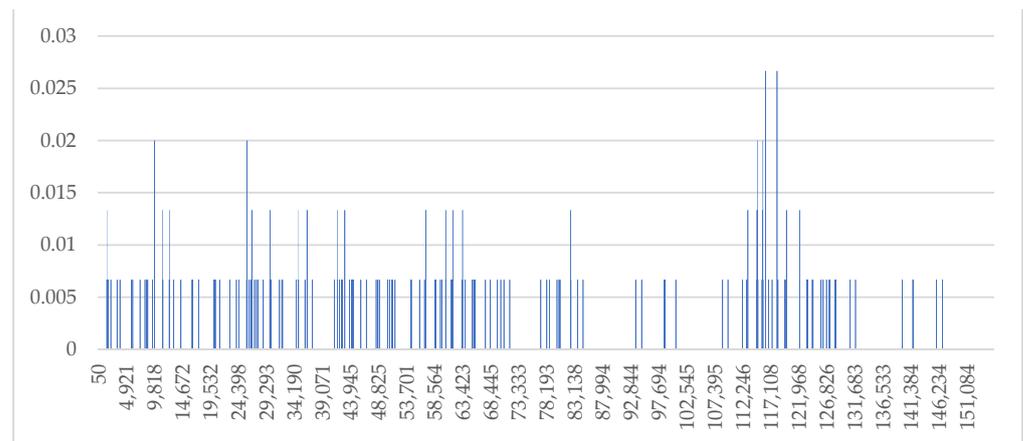
Figure 5. Sliding window analysis of the three willows whole plastid genomes (*X*-axis, position of the midpoint of a window; *Y*-axis, nucleotide diversity of each window).
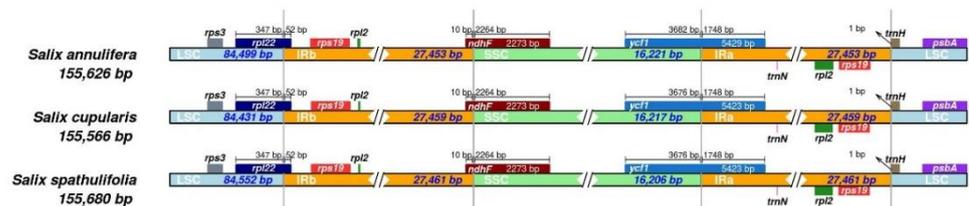
Figure 6. Comparison of the LSC, SSC, and IRs boundary regions among three *Salix* chloroplast genomes.

3.4. Phylogenetic Analysis

To explore the phylogenetic status of these three willows, an ML phylogeny was built using the complete cp genomes of 20 Salicaceae species published on NCBI and three in our study (Figure 7). As shown in Figure 7, all the available willows were evidently divided into two major groups, consistent with a previous study [49]. The three target willows were all in clade I. *S. spathulifolia* was at the base of clade I; *S. annulifera* was most closely to *S. rorida*, while *S. cupularis* was resolved as a separate branch and placed as a sister group to *S. suchowensis* and *S. psammophila*, which belonged to sect. Helix. The phylogenetic tree generated a total of 21 nodes, and bootstrap support was 100% in 11 out of 21 branch nodes.

However, in Li et al. (2021) the phylogenetic placement of *Salix* cupularis contradicted our result, which may be related to some reasons about constructing phylogenetic tree and origin of samples. Firstly, outgroup sampling was a crucial step in phylogenetic analyses, affecting homoplasy between ingroup and the outgroup and statistical error rate, thus, an outgroup with large genetic distance from the ingroup may share few relevant character states and took a long time to accumulate homoplasy, attach randomly to the ingroup, and bias ingroup relationships, and this was the reason why we selected genus *Populus*, the sister group of *Salix* to root the phylogenetic tree [50–52]. Meanwhile, methods of tree construction may have an effect on the phylogenetic results. The reconstruction methods of phylogeny can be divided into two categories: distance-based and sequence-based. The most popular distance method was the NJ method, which used the distance matrix as input to specify the distance between each pair of taxa, however, different from the NJ method, ML method was based on a clear sequence evolution model and likelihood function, one of the advantages of the ML method was that all its model assumptions are clear, so they can be evaluated and improved, generally speaking, if the model was appropriate, ML method outperformed other methods [53]. Furthermore, considering the considerable intraspecific genotypical polymorphisms, extensive interspecific hybridization, and chloroplast capture

result from reduced sterility barriers among taxa, establishing phylogeny of the genus *Salix* was difficult [54]. Our sample was taken from the vicinity of the type locality, which was some distance from the sampling sites mentioned in Li et al. [24]. We inferred that this may also be one of the reasons for the inconsistent phylogenetic results.
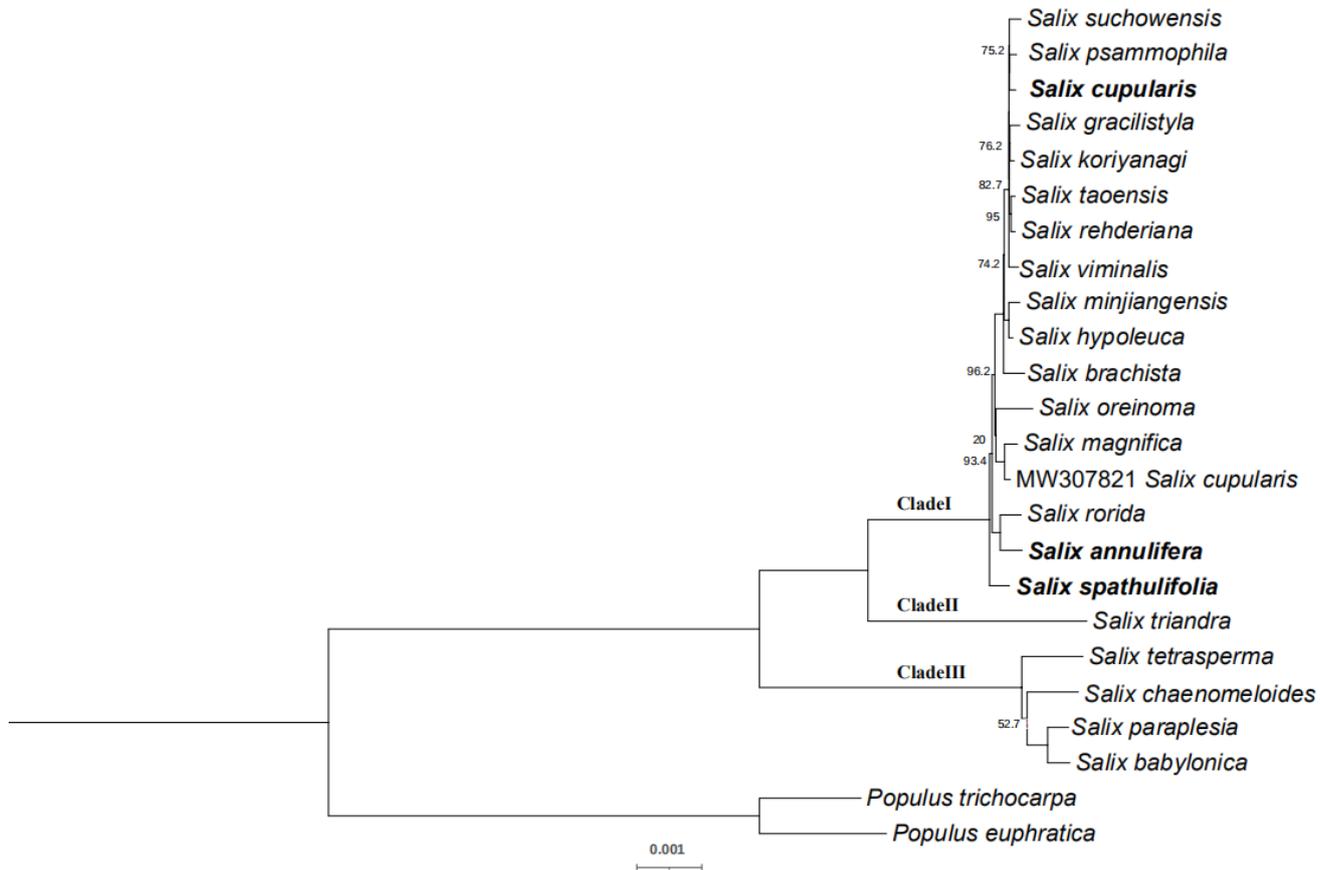
Figure 7. Maximum likelihood (ML) phylogeny of 23 species belonging to Salicaceae constructed based on their chloroplast genomes with 5000 bootstrap replicates. *Populus trichocarpa* Torr. & A.Gray ex Hook. and *Populus euphratica* Olivier were used as the outgroups. Bootstrap values are shown at branch nodes (only bootstrap values < 100% were shown). Accession numbers: *Salix magnifica* Hemsl. NC037424, *Salix hypoleuca* Seemen NC037423, *Salix rorida* NC037428, *Salix rehderiana* NC037427, *Salix viminalis* L. MN117720, *Salix koriyanagi* Kimura ex Goerz NC044419, *Salix psammophila* MN495627, *Salix suchowensis* NC026462, *Salix babylonica* L. NC028350, *Salix oreinoma* C.K.Schneid. NC035743, *Salix brachista* C.K.Schneid. CM018592, *Salix triandra* L. MK722343, *Salix tetrasperma* Roxb. NC035744, *Salix paraplesia* C.K.Schneid. NC037426, *Salix gracilistyla* Miq. NC043878, *Salix taoensis* Goerz ex Rehder & Kobuski NC037429, *Salix chaenomeloides* Kimura NC037422, *Salix minjiangensis* N.Chao NC037425, *Populus euphratica* NC024747, *Populus trichocarpa* NC009143. The three species used in this study are bold marked.

4. Conclusions

In this study, the whole chloroplast genomes of *S. annulifera*, *S. cupularis*, and *S. spathulifolia* were obtained by next-generation sequencing. Our results showed that the genomes ranged from 155,566 to 155,680 bp in length, encoding 130 genes. Comparison among these three willows' cp genomes showed similarity in their structure and organization. Phylogenetic analysis revealed that *Salix spathulifolia*, *S. cupularis*, and *S. annulifera* could be found within one clade among all studied *Salix* species. *S. spathulifolia* was located at the base of the clade, *S. annulifera* was sister to *S. rorida*, and *S. cupularis* was closest to *S. suchowensis* and *S. psammophila*. The complete cp genomes of these three willows may provide an opportunity for further molecular phylogenetic and evolutionary studies in Salicaceae.

References

1. Francisco-Ortega, J.; Wang, F.G.; Wang, Z.S.; Xing, F.W.; Liu, H.; Xu, H.; Xu, W.X.; Luo, Y.B.; Song, X.Q.; Gale, S.; et al. Endemic seed plant species from Hainan Island: A checklist. *Bot. Rev.* **2010**, *76*, 295–345. [CrossRef]
2. Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **2008**, *49*, 827–831. [CrossRef]
3. Palmer, J.D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **1985**, *19*, 325–354. [CrossRef]
4. Fan, W.B.; Wu, Y.; Yang, J.; Shahzad, K.; Li, Z.H. Comparative chloroplast genomics of dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front. Plant Sci.* **2018**, *9*, 689. [CrossRef]
5. Wicke, S.; Schneeweiss, G.M.; de Pamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [CrossRef] [PubMed]
6. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; de Pamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J.; et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **2005**, *395*, 348–384. [CrossRef] [PubMed]
7. Sloan, D.B.; Triant, D.A.; Forrester, N.J.; Bergner, L.M.; Wu, M.; Taylor, D.R. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenetics Evol.* **2014**, *72*, 82–89. [CrossRef]
8. Wang, Y.G. Natural hybridization and speciation. *Biodivers. Sci.* **2017**, *25*, 565–576. [CrossRef]
9. Ellis, R.J. *The Plastids: Their Chemistry, Structure, Growth and Inheritance*, 2nd ed.; JTO Kirk and RAE Tilney-Bassett, Elsevier/North-Holland Biomedical Press: Amsterdam, The Netherlands; New York, NY, USA; Oxford, UK, 1978.
10. Kuroiwa, T. The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei. *Int. Rev. Cytol.* **1991**, *128*, 1–62. [CrossRef]
11. Flora of North America Editorial Committee. *Flora of North America: Volume 7 Magnoliophyta: Salicaceae to Brassicaceae*; Oxford University Press: Oxford, UK, 2010.
12. Wu, Z.Y.; Raven, P.H. *Flora of China*; Missouri Botanical Garden Press: St. Louis, MO, USA, 1999; Volume 4.
13. Skvortsov, A.K. *Willows of Russia and Adjacent Countries: Taxonomical and Geographic Revision (Translated by n. Kadis, 1999)*; Joensuu University: Joensuu, Finland, 1999.
14. Fang, Z.F. On the distribution and origin of *Salix* in the world. *J. Syst. Evol.* **1987**, *25*, 307–313.
15. Myers-Smith, I.H.; Forbes, B.C.; Wilmking, M.; Hallinger, M.; Lantz, T.; Blok, D.; Tape, K.D.; Macias-Fauria, M.; Sass-Klaassen, U.; Lévesque, E.; et al. Shrub expansion in tundra ecosystems: Dynamics, impacts and research priorities. *Environ. Res. Lett.* **2011**, *6*, 045509. [CrossRef]
16. Smart, L.B.; Volk, T.A.; Lin, J.; Kopp, R.F.; Phillips, I.S.; Cameron, K.D.; White, E.H.; Abrahamson, L.P. Genetic improvement of shrub willow (*Salix* spp.) crops for bioenergy and environmental applications in the United States. *Unasylva* **2005**, *56*, 51.
17. Argus, G.W. The genus *Salix* (Salicaceae) in the southeastern United States. *Syst. Bot. Monogr.* **1986**, *9*, 1–170. [CrossRef]
18. Zhao, S.D. Distribution of Willows (*Salix*) in China. *J. Syst. Evol.* **1987**, *25*, 114–124.

19. Zhang, M.L. Studies on the distribution and differentiation of Willows (*Salix*) in Qinling Mountain. *Bull. Bot. Res.* **1993**, *13*, 136–145.
20. Chen, J.H.; Sun, H.; Wen, J.; Yang, Y.P. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. *Taxon* **2010**, *59*, 29–37. [CrossRef]
21. Lumbsch, H.T.; Lauron-Moreau, A.; Pitre, F.E.; Argus, G.W.; Labrecque, M.; Brouillet, L. Phylogenetic relationships of American Willows (*Salix* L., Salicaceae). *PLoS ONE* **2015**, *10*, e0121965. [CrossRef]
22. Azuma, T.; Kajita, T.; Yokoyama, J.; Ohashi, H. Phylogenetic relationships of *Salix* (Salicaceae) based on *rbcL* sequence data. *Am. J. Bot.* **2000**, *87*, 67–75. [CrossRef]
23. Barkalov, V.Y.; Kozyrenko, M.M. Phylogenetic relationships of *Salix* L. Subg. *Salix* species (*Salicaceae*) according to sequencing data of intergenic spacers of the chloroplast genome and its rDNA. *Russ. J. Genet.* **2014**, *50*, 828–837. [CrossRef]
24. Li, J.; Zhuo, Z.; Xu, D.; Yang, H.; Zhu, T. The complete chloroplast genome of *Salix cupularis* rehder, a sand binder in alpine hillslope, China. *Mitochondrial DNA Part B* **2021**, *6*, 2519–2520. [CrossRef]
25. Hu, Y.F.; Shu, X.Y.; He, J.; Zhang, Y.L.; Xiao, H.H.; Tang, X.Y.; Gu, Y.F.; Lan, T.; Xia, J.G.; Ling, J.; et al. Storage of C, N, and P affected by afforestation with *Salix cupularis* in an alpine semiarid desert ecosystem. *Land Degrad. Dev.* **2018**, *29*, 188–198. [CrossRef]
26. Gulyaev, S.; Cai, X.-J.; Guo, F.-Y.; Kikuchi, S.; Wendy, L.; Zhang, Z.-X.; Hörandl, E.; He, L. The phylogeny of *Salix* revealed by whole genome re-sequencing suggests different sex-determination systems in major groups of the genus. *Ann. Bot.* **2021**. submitted.
27. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An ultra-fast all-in-one fastq preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
28. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
29. He, J.; Yao, M.; Lyu, R.D.; Lin, L.L.; Liu, H.J.; Pei, L.Y.; Yan, S.X.; Xie, L.; Cheng, J. Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae). *Sci. Rep.* **2019**, *9*, 15285. [CrossRef]
30. Huang, Y.; Wang, J.; Yang, Y.; Fan, C.; Chen, J. Phylogenomic analysis and dynamic evolution of chloroplast genomes in Salicaceae. *Front. Plant Sci.* **2017**, *8*, 1050. [CrossRef]
31. Qu, X.J.; Moore, M.J.; Li, D.Z.; Yi, T.S. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **2019**, *15*, 50. [CrossRef]
32. Wu, Z.Q. The whole chloroplast genome of shrub Willows (*Salix suchowensis*). *Mitochondrial DNA Part A* **2016**, *27*, 2153–2154. [CrossRef] [PubMed]
33. Huang, D.I.; Cronk, Q.C. Plann: A command-line application for annotating plastome sequences. *Appl. Plant Sci.* **2015**, *3*, 1500026. [CrossRef]
34. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (ogdraw) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef]
35. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]
36. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef] [PubMed]
37. Kumar, S.; Stecher, G.; Tamura, K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
38. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]
39. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]
40. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [CrossRef] [PubMed]
41. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [CrossRef] [PubMed]
42. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [CrossRef] [PubMed]
43. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
44. Redwan, R.M.; Saidin, A.; Kumar, S.V. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biol.* **2015**, *15*, 196. [CrossRef]
45. Hu, Y.; Woeste, K.E.; Zhao, P. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. *Front. Plant Sci.* **2017**, *7*, 1955. [CrossRef] [PubMed]
46. Wang, W.; Yu, H.; Wang, J.; Lei, W.; Gao, J.; Qiu, X.; Wang, J. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [CrossRef] [PubMed]

47. Liu, X.F.; Zhu, G.F.; Li, D.M.; Wang, X.J. Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum* 'Parrish'. *PLoS ONE* **2019**, *14*, e0224038. [CrossRef] [PubMed]
48. Yang, J.B.; Tang, M.; Li, H.T.; Zhang, Z.R.; Li, D.Z. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Biol.* **2013**, *13*, 84. [CrossRef] [PubMed]
49. Chen, Y.; Hu, N.; Wu, H. Analyzing and characterizing the chloroplast genome of *Salix wilsonii*. *BioMed Res. Int.* **2019**, *2019*, 5190425. [CrossRef]
50. Lyons-Weiler, J.; Hoelzer, G.A.; Tausch, R.J. Optimal outgroup analysis. *Biol. J. Linn. Soc.* **2008**, *64*, 493–511. [CrossRef]
51. Wheeler, W.C. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **1990**, *6*, 363–367. [CrossRef]
52. Huelsenbeck, J.P.; Bollback, J.P.; Levine, A.M. Inferring the root of a phylogenetic tree. *Syst. Biol* **2002**, *51*, 332–343. [CrossRef]
53. Yang, Z.; Rannala, B. Molecular phylogenetics: Principles and practice. *Nat. Rev. Genet.* **2012**, *13*, 303–314. [CrossRef]
54. Rechinger, K.H. *Salix* taxonomy in Europe—Problems, interpretations, observations. *Proc. R. Soc. Edinb.* **1992**, *98*, 1–12. [CrossRef]