biologia plantarum
an international journal for experimental botany

BIOLOGIA PLANTARUM (2021) 65: 237-245 DOI: 10.32615/bp.2021.014

The complete chloroplast genome of *Tamarix ramosissima* and comparative analysis of *Tamaricaceae* species

L. WANG[1], L. WANG[2], and Z.-H. GUO[3,*]

[1] *State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xi'ning 810016, P.R. China*
[2] *Qinghai Academy of Agricultural Forestry Sciences, Qinghai University, Xi'ning 810016, P.R. China*
[3] *Gaolan Station of Agricultural and Ecological Experiment, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, P.R. China*

[*]Corresponding author: E-mail: guozhhong@lzb.ac.cn

Abstract

Tamarix ramosissima is a deciduous shrub that resides in arid and semi-arid regions. Although of ecological and medicinal values, some *Tamarix* species are considered invasive as they have dominated the riparian zones of dryland in some parts of the world. Here, the complete chloroplast (cp) genome of *T. ramosissima* was sequenced and analyzed, showing a size of 156 150 bp and a GC content of 36.5 %. The plastome displayed a typical quadripartite structure, consisting of a pair of inverted repeat (IR) regions of 26 554 bp, separated by a large single copy (LSC) region of 84 795 bp, and a small single copy (SSC) region of 18 247 bp. The cp genome encoded 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. A total of 32 repeat sequences and 64 simple sequence repeat (SSR) were identified in the plastome, and an obvious A/T bias was observed in the majority of the SSRs detected. By comparing the *T. ramosissima* cp genome with those of the other four *Tamaricaceae* species, a number of divergence hotspots were identified among these plastomes. Together with SSRs and long repeats identified, these divergence hotspots could be developed as potential molecular markers facilitating species discrimination and evolutionary studies. Using plastome sequences, we re-investigated the phylogenetic relationship among 19 species, and *T. ramosissima* was found to be a sister of *Tamarix chinensis*. Taken together, our study provides valuable genomic resources to deepen the understanding of plant photosynthetic mechanism and phylogenomics.

Keywords: chloroplast genome, phylogenomics, plant evolution, *Tamarix ramosissima*.

Introduction

Tamarix plants belong to the *Tamaricaceae* family and they are ancient species native to the Mediterranean region (Zhang *et al*. 2006). The *Tamaricaceae* family is composed of about 120 species distributed into 3 - 5 genera, among which *Tamarix* is the largest genus encompassing over 90 species (Crins 1989). Although several *Tamarix* species have been considered invasive in the United States, the supplementing of *Tamarix* plants in traditional medicine revealed their values in an application (Bahramsoltani *et al*. 2020). For instance, the leaves of *Tamarix* species have been reported to have pharmacological effects such as detoxification, rheumatism dispelling, and diuresis promotion in traditional Chinese medicine. In Middle East countries, the extract of *Tamarix* leaves has been used as an antiseptic agent (KalamUrfi *et al*. 2016). The bark of *Tamarix aphylla*, which differed in chemical constitution from the leaf, has been used as a herbal remedy for eczema capitis alleviating (Yusufoglu and Al-qasoumi 2011).

Tamarix species showed a wide range of variation in morphological traits, making species delimitation difficult (Sheidai *et al*. 2018). This was further complicated by the existence of interspecific hybridization (Mayonde *et al.*

Received 19 September 2020, *last revision* 25 February 2021, *accepted* 2 March 2021.
Abbreviations: BI - Bayesian inference; cp - chloroplast; IR - inverted repeat; LSC - large single copy; ML - maximum likelihood; NGS - next-generation sequencing; RSCU - relative synonymous codon usage; SSC - small single copy; SSR - simple sequence repeat.
Acknowledgements: This research was supported by the Science and Technology Project of Qinghai Province (2019-ZJ-962Q; 2016-ZJ-Y01), the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (2019-ZZ-05), and the Youth Foundation of Qinghai University (2019-QNY-2).
Conflict of interest: The authors declare that they have no conflict of interest.

2016). Using combinational barcodes of chloroplast and nuclear DNA, the morphological intermediates in the southwestern USA were identified as hybrids between *T. chinensis* or *T. ramosissima* and *T. aphylla* (Gaskin and Shafroth 2005).

Fragments of chloroplast (cp) DNA, and whole plastome sequences as well, have been extensively used as markers for phylogenetic inference in many plants (Dobrogojski *et al*. 2020). The chloroplast genome, which was also known as cpDNA, was inherited maternally in most plants. Double-stranded cpDNAs usually exhibited a circular structure, with genome sizes ranging from 107 to 218 kb (Turmel *et al*. 2015). Further alignment of published cp genomes revealed their conservation in gene arrangement. In angiosperms, most cps shared a quadripartite structure, consisting of two inverted repeat regions (IRa and IRb) separated by a small single copy (SSC) region and a large single copy (LSC) region (Palmer 1990). Sequences of LSC and SSC were conserved across most plant species. But in *Gymnospermae*, the inverted repeats could vary substantially between species and changes of the IR regions often leaded to massive adjustment in DNA arrangement (Yang *et al*. 2020). For example, when lacking a large inverted repeat, extensive rearrangements of chloroplast DNA were observed in two conifer plants (Strauss *et al*. 1988).

The moderate evolutionary rate of cp genomes made them potentially valuable resources for phylogenetic studies (Duan *et al*. 2020). Comparing with nuclear genomes, cpDNAs were smaller in size and contained more conserved sequences. With an increasing number of cp genomes being sequenced, plastome-based phylogenomics could provide novel solutions for resolving phylogenetic ambiguities in plants.

The present study aimed to report the complete cp genome of *Tamarix ramosissima* and compare it with those of other species of the family *Tamaricaceae*. Through comparison, we can unveil the differences among the cp genomes of five *Tamaricaceae* species. Data in this study could facilitate the development of cp-derived molecular markers and elucidate the phylogenetic relationship among *Tamaricaceae* species.

Materials and methods

Plant materials, DNA isolation, and next-generation sequencing: Fresh leaves of *Tamarix ramosissima* Kar. ex Boiss were collected from Gaolan Ecological and Agricultural Research Station, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences (36° 13′ N, 103° 47′ E). After washing with distilled water, sampled leaves were frozen immediately in liquid nitrogen and kept at -80 °C until DNA extraction. Subsequent genomic DNA extraction was performed using the *Tiangen* Plant Genomic DNA kit (*Tiangen Biotech Co*., Beijing, China) according to the manufacturer's instructions. We then submitted the extracted DNA for next-generation sequencing (NGS) library construction and sequencing using an *Illumina Hiseq 2500* platform (*Illumina Inc*., San Diego, CA, USA).

Genome assembly and annotation: The obtained raw data were trimmed with *Trimmomatic* software (*v. 0.36*) to remove adaptor sequence and low-quality reads. The resulting clean reads were mapped against the reference chloroplast genome (*Tamarix chinensis*) to extract cp-like reads. Reference-based assembly was initially performed with *MITObim v. 1.9* (Hahn *et al.* 2013). Then *de novo* assembly was performed using *NOVOPlasty v. 2.7.2* (Dierckxsens *et al*. 2017), with contigs assembled by *MITObim* as the seed and reference. The order and orientation of *NOVOPlasty* assemblies were then manually adjusted, and the draft genome from *MITObim* assembly was used as the evidence for adjustment when necessary. Finally, the draft assembly was polished with *Pilon v. 1.22* (Walker *et al*. 2014). All cp-like reads were re-mapped to the final assembly to calculate the value of coverage (Table 1 Suppl., Fig. 1 Suppl.).

The preliminary gene annotation of the draft *T. ramosissima* cp genome was performed using the *GeSeq* tool (Tillich *et al*. 2017). Then the annotations were further curated manually using the *CLC Sequence Viewer* (*version 8*). The map of the *T. ramosissima* cp genome was drawn using *Organellar Genome DRAW* software (Greiner *et al*. 2019). Annotated *T. ramosissima* plastome sequence was then submitted to *GenBank*.

Genome structure and genome comparison: To visualize the structural variations among the cp genomes of five *Tamaricaceae* species, the plastome of *T. ramosissima* was compared with those of *Reaumuria trigyna* (NC_041265), *Hololachna songarica* (NC_041273), *Myricaria paniculata* (NC_041270), and *T. chinensis* (NC_040943) by using the *mVISTA* program under Shuffle-LAGAN model (Mayor *et al*. 2000). The annotation of *T. chinensis* (NC_040943) was used as the reference.

For nucleotide variation analysis, the five cp genomes of *Tamaricaceae* were first aligned with *MAFFT v. 7.450* (Katoh and Standley 2013). Then, the nucleotide diversity values (Pi) among the cp genomes were calculated on *DnaSP6* (Rozas *et al*. 2017), with window length set to 800 bp and step size set to 200 bp.

The relative synonymous codon usage (RSCU) is the ratio of the observed frequency of specific codons to their expected frequency. When RSCU > 1, it means that this codon is used more frequently than expected. However, an RSCU value of less than 1 represents that a codon is used less frequently than expected. The RSCU value of each codon of the five *Tamaricaceae* cp genomes was calculated using *DAMBE v. 7.0.68* (Xia 2018).

Repeat analysis: *REPuter* program was used to identify the repetitive sequences within cp genomes (Kurtz *et al*. 2001). The selection criterion of a minimum length of 15 bp with sequence similarity of 90 % was applied to filter repeats in different types (forward, reverse, complement, and palindromic).

For simple sequence repeats (SSRs) analysis, the

prediction was performed with *MISA-web* (https://webblast.ipk-gatersleben.de/misa/). SSR motifs were searched within the cp genomes according to the criteria as follows: for mono-nucleotide repeats, ≥ 10 units of repeats are required; for di-nucleotide repeats, ≥ 8 units of repeats are required; for tri-nucleotide and tetra-nucleotide repeats, ≥ 4 units of repeats are required; and for penta-nucleotide and hexa-nucleotide repeats, ≥ 3 units of repeats are required (Wellington 2009).

Phylogenetic analysis: To investigate the phylogenetic relationships among *Tamaricaceae* species, a total of 18 plastome sequences were retrieved from *GenBank* and used for phylogeny construction. Both the coding regions (protein-coding genes, tRNAs, and rRNAs) and the non-coding regions (intergenic spacers and introns) were extracted with *PhyloSuite* (*v.1.2.2*) (Zhang *et al.* 2020). Sequences of each data matrix were then aligned using *MAFFT* under automatic model selection mode. These alignments were subsequently concatenated with *PhyloSuite*. The concatenated sequences were subjected to phylogenetic analyses using maximum likelihood (ML) and Bayesian (BI) methods. For a model selection of each data set, *PartitionFinder2* (*v. 2.1.1*) was used to identify the best partition scheme and to select the best-fit model (Lanfear *et al.* 2016). Identification was implemented under the Bayesian information criterion (BIC) using a heuristic search (search=rcluster). For partitioned Bayesian analysis, the inference was performed with *MrBayes* (*v. 3.2.6*) which was integrated with *PhyloSuite* as a plugin (Ronquist *et al.* 2012). Each data set was given a best-fit model of its own. The analysis was run for 2 000 000 generations, with the first 25 % of generations discarded as burn-in. Phylogenetic inference by the ML method was carried out using *IQ-TREE* (integrated with *PhyloSuite* as a plugin) (Nguyen *et al.* 2014), with the best-fit model selected for each partition loaded automatically.

Results

The complete cp genome of *T. ramosissima* was 156 150 bp in length, displayed a typical quadripartite structure, in which a small single-copy region (SSC; 18 247 bp) and a large single-copy region (LSC; 84 795 bp) were separated by two identical inverted repeats (IR; 26 554 bp) (Fig. 2 Suppl.). After comparing the size and structure of cp genomes from *Tamaricaceae* species, we found that the lengths of the five plastomes varied from 154 533 to 156 167 bp; *T. chinensis* had the largest, while *R. trigyna* had the smallest (Table 1). The overall GC content of the *T. ramosissima* plastome was 36.5 %, which was similar to those of the other four *Tamaricaceae* species. As shown in Table 1, the *T. ramosissima* cp genome encoded 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. The sequence of the newly assembled *T. ramosissima* plastome has been submitted to GenBank and deposited under the accession number MN726883.

All the genes annotated in the *T. ramosissima* cp genome were listed in Table 2. Of the 130 genes annotated, a total of 16 genes contained introns. Among these intron-containing genes, 14 genes contained one intron, including 6 tRNA genes (*trna-UUU*, *trna-CGA*, *trna-UUC*, *trna-UAA, trna-ACA, trna-UGC*) and 6 protein-coding genes (*ndhA, ndhB, atpF, rpoC1, rpl2, rps12*). Two genes contained two introns (*clpP, ycf3*). The *rps12* was the only trans-spliced gene in the *T. ramosissima* plastome.

Codon usage of protein-coding sequences in the *T. ramosissima* cp genome was analyzed with the *DAMBE* software. Overall, 64 codons, corresponding to the 20 amino acids, were found present in the *T. ramosissima* plastome. A total number of 24 724 codons were identified for all the protein-coding sequences (including the stop codons). Leucine (2 651 codons; 10.72 %) was the most abundant amino acid, whereas cysteine (283 codons; 1.14 %) was the least abundant. The relative synonymous codon usage (RSCU) value, which was positively

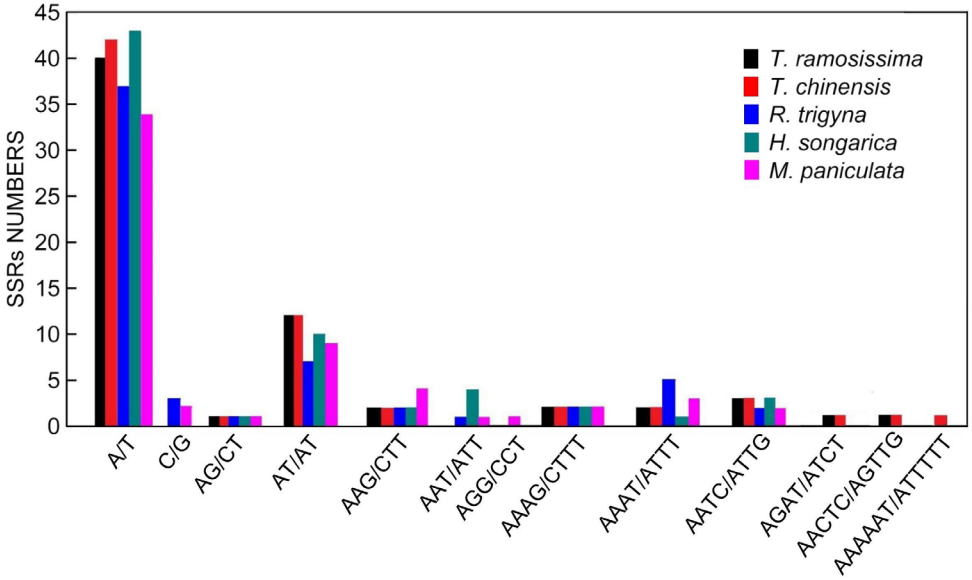

Fig. 1. Simple sequence repeats (SSRs) among the five *Tamaricaceae* cp genomes.

correlated with the number of codons, was calculated across the five *Tamaricaceae* species. As illustrated in Table 3, 30 codons exhibited high preferences (RSCU > 1) in all the *Tamaricaceae* plants, while 32 codons exhibited low preferences (RSCU < 1). The codon usage of methionine and tryptophan was unbiased (RSCU = 1).

The newly sequenced *T. ramosissima* cp genome was compared with those of the other four *Tamaricaceae* species using the *mVISTA* program (Fig. 3 Suppl.). The comparison revealed the high conservation existed between the cp genome of *T. ramosissima* and *T. chinensis*. Furthermore, coding regions were found to be more conserved than non-coding regions. Both the SSC and the LSC regions were more divergent than the two IR regions overall.

To reveal the divergence hotspots in the five *Tamaricaceae* chloroplast genomes, the nucleotide diversity values (Pi) were calculated using *DnaSP*. The Pi values for the five *Tamaricaceae* plastomes ranged from 0 to 0.195, and the average value was 0.02769. The

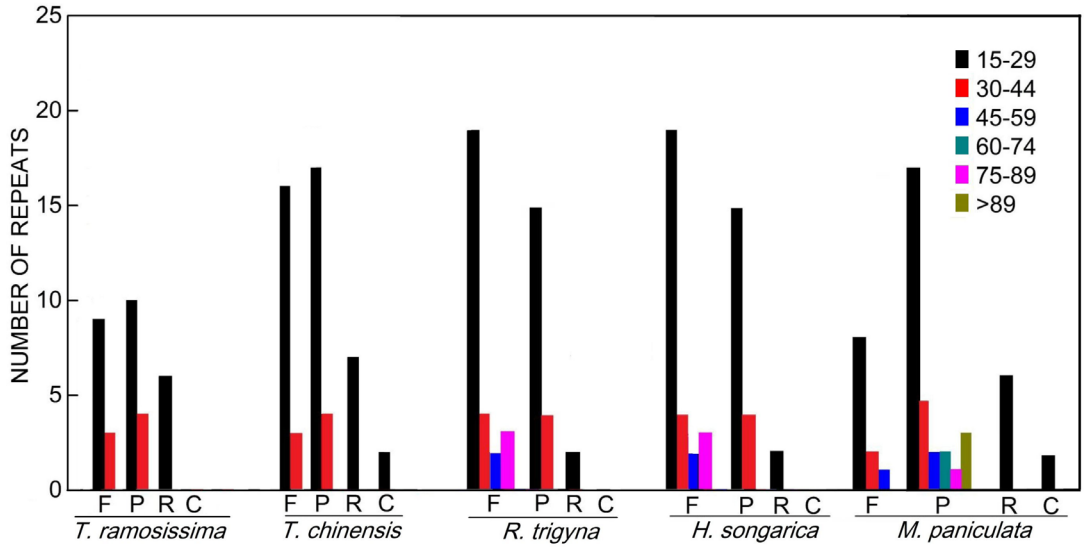

Fig. 2. Long repetitive sequences identified in the five *Tamaricaceae* cp genomes. F, P, R, and C represents the repeat types of forward (F), palindrome (P), reverse (R), and complement (C), respectively. Repeats with different lengths [bp] are coded with corresponding colours.

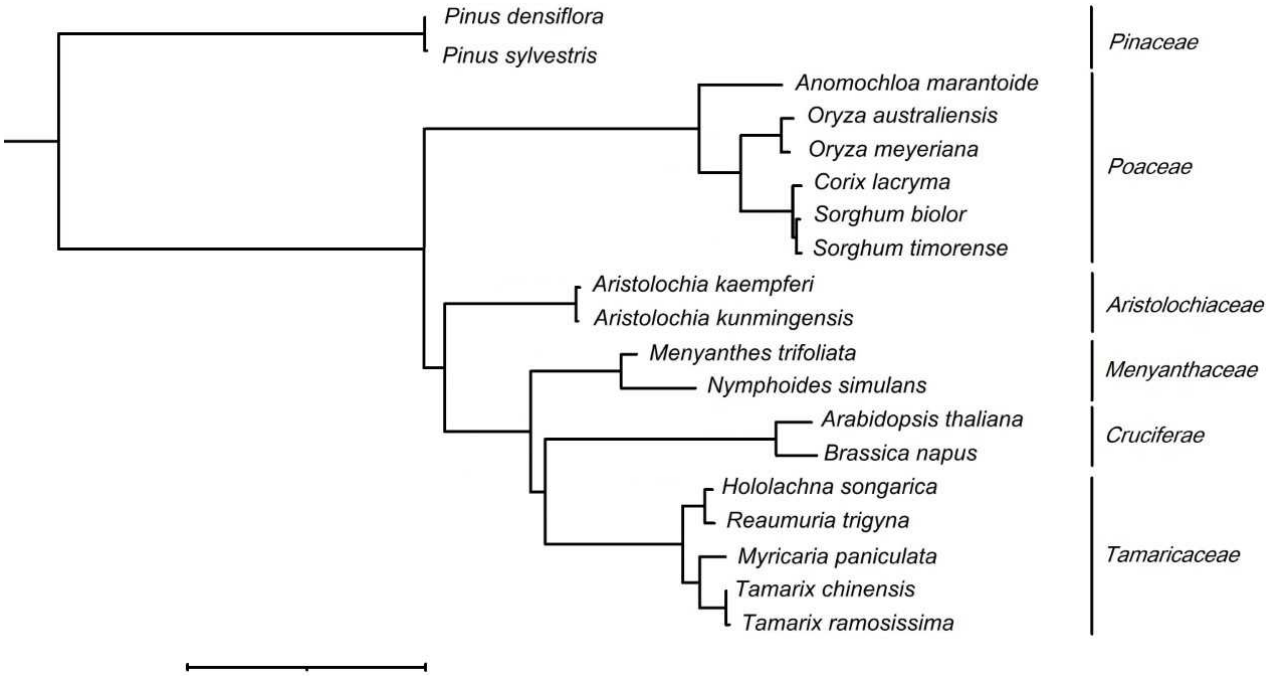

Fig. 3. Plastome-based phylogenetic relationship among the five *Tamaricaceae* species. *Pinus densiflora* and *Pinus sylvestris* were used as outgroups. Values beside branch nodes denote support values for bootstrap.

Table 1. Comparison of general features of five *Tamaricaceae* plastomes. LSC - large single copy, IR - inverted repeat, SSC - small single copy.

Species	Total [bp]	LSC [bp]	IR [bp]	SSC [bp]	Total genes	Protein coding genes	tRNA	rRNA	GC [%]
Tamarix ramosissima	156 150	84 795	53 108	18 247	130	85	37	8	36.5
Tamarix chinensis	156 167	84 768	53 152	18 247	130	85	37	8	36.5
Hololachna songarica	155 596	85 903	52 138	17 555	130	85	37	8	36.8
Reaumuria trigyna	154 533	84 811	52 116	17 607	130	85	37	8	37.0
Myricaria paniculata	154 651	84 379	49 588	20 684	130	85	37	8	36.3

Table 2. List of genes annotated in the chloroplast genome of *T. ramosissima.* * - indicates genes containing one intron; ** - indicates genes containing two introns; ^{T} - indicates trans-spliced genes; ×2 - indicates genes having two copies; ×3 - indicates genes having three copies; ×4 indicates genes having four copies.

Function	Gene names	Number
Photosystem I	psaA; psaB; psaC; psaI; psaJ	5
Photosystem II	psbA; psbB; psbC; psbD; psbE; psbF; psbH psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ	15
Cytochrome b/f complex	petA;petB;petD;petG;petL;petN	6
ATP synthase	*atpA;atpB;atpE;atpF*;atpH;atp*I	6
NADH dehydrogenase	*ndhA*; ndhB*(×2); ndhC; ndhD;ndhE; ndhF ndhG; ndhH; ndhI; ndhJ; ndh*K	12
Rubisco Large subunit	rbcL	1
Ribosomal RNAs	rrn4.5(×2); rrn5(×2); rrn16(×2); rrn23(×2)	8
Transfer RNAs	*trna-GUG; trna-UUU*;trna-UUG; trna-GCU; trna-CGA*;trna-UCU; trna-GCA; trna-GUC; trna-GUA; trna-UUC*(×3); trna-GGU; trna-UGA; trna-GCC; trna-CAU(×4); trna-GGA; trna-UGU; trna-UAA*;trna-GAA; trna-ACA*;trna-CCA; trna-UGG; trna-CAA(×2); trna-GAC(×2); trna-UGC*(×2); trna-ACG(×2); trna-GUU(×2); tr*na-UAG	37
DNA dependent RNA polymerase	*rpoA; rpoB; rpoC1*; rpoC*2	4
Small subunit of ribosome	*rps2; rps3; rps4; rps7(×2); rps8; rps11; rps14; rps12*^{T} (×2); rps16; rps15; rps18; rps19*	14
Large subunit of ribosome	*rpl2(×2)*; rpl14; rpl16; rpl20; rpl22; rpl23 (×2); rpl32; rpl33; rpl3*6	11
Proteins of unknown function	ycf1, ycf2 (×2), ycf3**, ycf4	5
Other genes	accD; ccsA; cemA; clpP**; matK; infA	6

LSC region and the SSC region showed higher nucleotide diversity than the two IR regions. Seven regions with high Pi values were identified as divergence hotspots (Fig. 4 Suppl.). The *rpl32-tRNA-UAG* region, with a Pi value of 0.195, was the most divergent part detected. Five intergenic regions (*tRNA-GCC-tRNA-CAU, psbK-psbI, tRNA-GAA-ndhJ, rps15-ycf1, rpl33-rps18*) and one gene region (*rpl16*) that had high Pi values were also identified as divergence hotspots. The divergence hotspots identified could be developed as potential markers for species delimitation of *Tamaricaceae* species.

The total number of SSRs identified in the five *Tamaricaceae* cp genomes ranged from 59 to 67 (Fig. 1). Among these SSRs, mono-nucleotide repeats were the dominant type, and A/T repeats accounted for nearly 60 % of all the SSRs identified. Di-nucleotide repeats were the second most abundant motif types identified, constituting 13.3 - 20.3 % of the total SSRs. Most of the di-nucleotide repeats were also AT-rich. As to tri-, tetra-, and penta-nucleotide repeats, they comprised a relatively small part of the SSRs detected (Fig. 1).

Long repeats in the five cp genomes were also analyzed with the *REPuter* software. As shown in Fig. 2, *T. ramosissima* had the smallest number of repeats in its plastome, consisting of 12 forward, 14 palindromic, and 6 reverse repeats (32 in total). More repetitive elements were identified in the chloroplast genomes of the other four *Tamaricaceae* plants (49 in each), but the types and sizes of the repetitive sequences varied in different species. The majority of the repeats identified were less than 29 bp. Repeats with lengths > 45 bp were only detected in the plastomes of *H. songarica, R. trigyna*, and *M. paniculate*.

The plastome-based phylogenomic inference was performed using both the ML and BI methods (Fig. 3). Among the five *Tamaricaceae* species, *T. ramosissima* and *T. chinensis* were clustered together. The *R. trigyna* and *H. songarica* were also monophyletic. *M. paniculata* was inferred to have a closer relationship with *Tamarix* species. The topological structure of the ML tree was consistent with the BI tree built.

Table 3. Codon content of 20 amino acids and stop codons in all protein-coding genes of the five *Tamaricaceae* cp genomes. RSCU - relative synonymous codon usage.

Amino acid	Codon	*T. ramosissima* RSCU	*T. chinensis*	*R. trigyna*	*H. songarica*	*M. paniculata*
Stop	UGA	0.622	0.679	0.714	0.786	0.532
Stop	UAG	0.732	0.679	0.714	0.679	0.646
Stop	UAA	1.646	1.643	1.571	1.536	1.823
A	GCU	1.792	1.799	1.765	1.766	1.810
A	GCG	0.348	0.335	0.344	0.350	0.348
A	GCC	0.643	0.637	0.635	0.621	0.609
A	GCA	1.217	1.230	1.256	1.263	1.234
C	UGU	1.534	1.547	1.572	1.553	1.598
C	UGC	0.466	0.453	0.428	0.447	0.402
D	GAU	1.577	1.578	1.577	1.571	1.566
D	GAC	0.423	0.422	0.423	0.429	0.434
E	GAG	0.460	0.447	0.473	0.471	0.457
E	GAA	1.540	1.553	1.527	1.529	1.543
F	UUU	1.329	1.346	1.305	1.305	1.377
F	UUC	0.671	0.654	0.695	0.695	0.623
G	GGU	1.346	1.342	1.285	1.302	1.371
G	GGG	0.619	0.608	0.633	0.633	0.574
G	GGC	0.372	0.374	0.399	0.382	0.374
G	GGA	1.663	1.676	1.684	1.682	1.681
H	CAC	0.470	0.479	0.430	0.432	0.458
H	CAU	1.530	1.521	1.570	1.568	1.542
I	AUU	1.507	1.501	1.487	1.491	1.555
I	AUA	0.915	0.926	0.939	0.942	0.919
I	AUC	0.578	0.574	0.573	0.568	0.526
K	AAA	1.494	1.518	1.489	1.482	1.543
K	AAG	0.506	0.482	0.511	0.518	0.457
L	CUA	1.109	1.090	1.121	1.134	1.179
L	CUC	0.600	0.597	0.600	0.606	0.529
L	CUG	0.515	0.518	0.521	0.498	0.479
L	CUU	1.775	1.796	1.758	1.761	1.814
L	UUA	1.229	1.244	1.194	1.201	1.260
L	UUG	0.771	0.756	0.806	0.799	0.740
M	AUG	1.000	1.000	1.000	1.000	1.000
N	AAC	0.459	0.440	0.450	0.450	0.433
N	AAU	1.541	1.560	1.550	1.550	1.567
P	CCA	1.143	1.168	1.157	1.165	1.142
P	CCC	0.666	0.656	0.702	0.703	0.697
P	CCU	1.604	1.608	1.566	1.564	1.661
P	CCG	0.587	0.569	0.575	0.568	0.500
Q	CAA	1.553	1.567	1.569	1.563	1.560
Q	CAG	0.447	0.433	0.431	0.437	0.440
R	AGA	1.471	1.472	1.446	1.445	1.439
R	AGG	0.529	0.528	0.554	0.555	0.561
R	CGA	1.555	1.593	1.554	1.573	1.531
R	CGC	0.365	0.354	0.407	0.379	0.370
R	CGG	0.490	0.486	0.545	0.543	0.469
R	CGU	1.590	1.568	1.494	1.504	1.630
S	AGC	0.418	0.418	0.455	0.447	0.433
S	AGU	1.582	1.582	1.545	1.553	1.567
S	UCA	1.151	1.148	1.121	1.125	1.108

S	UCC	0.768	0.772	0.818	0.814	0.793
S	UCG	0.452	0.457	0.466	0.464	0.450
S	UCU	1.628	1.623	1.595	1.597	1.648
T	ACC	0.667	0.671	0.677	0.673	0.670
T	ACA	1.240	1.242	1.250	1.253	1.210
T	ACG	0.427	0.416	1.410	0.409	0.385
T	ACU	1.666	1.670	1.663	1.665	1.735
V	GUU	1.510	1.508	1.476	1.489	1.544
V	GUG	0.522	0.507	0.532	0.517	0.488
V	GUC	0.439	0.441	0.460	0.444	0.434
V	GUA	1.529	1.544	1.533	1.550	1.534
W	UGG	1.000	1.000	1.000	1.000	1.000
Y	UAC	0.376	0.376	0.388	0.382	0.361
Y	UAU	1.624	1.624	1.612	1.618	1.639

Discussion

In the present study, we obtained the complete chloroplast genome of *T. ramosissima* by *Illumina* sequencing and compared it with those of the other four *Tamaricaceae* species. As shown in Table 1, the plastome size ranged from 154 533 bp to 156 167 bp with GC content slightly varied from 36.3 to 37.0 %. Each of the five *Tamaricaceae* cp genomes encoded 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. It was explicit that the five cp genomes were highly conserved in genome size and structure, especially for *T. ramosissima* and *T. chinensis.* However, boundary regions between SSC/IRs and LSC/IRs exhibited slight variations, which might be exerted by the expansion or contraction of IRs. Among the five *Tamaricaceae* species, plastome size was positively correlated with the length of IR, which was consistent with previous observations that changes in the IRs and their adjacent border regions were the main driving force for genome size variation and evolution (Fu *et al*. 2017, Xue *et al*. 2019).

SSRs or microsatellites are short tandem repeats that can be developed into molecular markers (Li *et al.* 2002). Chloroplast SSRs have been extensively used to study genetic diversity and phylogenetics in plants (Huang *et al*. 2015, Bi *et al*. 2018). Among the 64 SSRs identified in the *T. ramosissima* plastome, 40 (62.5 %) were A/T mononucleotide repeats, and 12 (18.75 %) were AT/AT di-nucleotide repeats. The high abundance of A and T in chloroplast SSRs were also observed in plastomes of the other four *Tamaricaceae* species. The findings in our study are consistent with those described previously in other species, including *Xanthium sibiricum* (Somaratne *et al*. 2019), *Populus* species (Gao *et al*. 2019), and *Lilium* plants (Du *et al*. 2017). The SSRs identified in *T. ramosissima* plastome, as well as those in other four *Tamaricaceae* species, could be developed as potential molecular markers facilitating future phylogenetic research.

Long repeats in plastid contribute to genome rearrangement and variation through unconventional combination at the repeat regions (Zhang *et al*. 2016), which might promote genetic diversity of the plastome (Timme *et al*. 2007). In the present study, we identified 32 - 49 repeats in the five *Tamaricaceae* cp genomes, the majority of which was localized in the LSC region. These repeat sequences, varying in types and sizes, may promote the evolution of the plastids of *Tamaricaceae* species by generating new variations.

Due to the scarcity of cp genomic data, phylogenetic study at the plastomic level was previously difficult to accomplish (Reginato *et al*. 2016). With the rapid development of high-throughput sequencing, plastome-based phylogenomics is emerging as a new tool for phylogenetics and evolutionary study in plants (McKain *et al*. 2018). We re-investigated the phylogenetic relationship in *Tamaricaceae* using the complete cp genome sequences available in a public database. Our analysis revealed that the two *Tamarix* species, *T. ramosissima* and *T. chinensis,* were clustered together. *Tamarix* plants were more closely related to *M. paniculata,* a species that shows resemblance in appearance to plants in the genus *Tamarix.* According to Wang *et al.* (2009), species of *Myricaria* were previously included in genus *Tamarix* taxonomically. Plastone-based phylogenomics and phylogeny inferred from the *pbsA-trnH* intergenic spacer also cofirmed the close relationship between these two genera (Channa *et al.* 2018, Yao *et al*. 2019). The complete cp genome sequence of *T. ramosissima* reported in our study will provide useful data resources for marker development and phylogenetics of *Tamaricaceae* species.

References

Bahramsoltani, R., Kalkhorani, M., Abbas Zaidi, S.M., Farzaei, M.H., Rahimi, R.: The genus *Tamarix*: traditional uses, phytochemistry, and pharmacology. - J. Ethnopharmacol. **246**:112245, 2020.

Bi, Y., Zhang, M.F., Xue, J., Dong, R., Du, Y.P., Zhang, X.H.: Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. - Sci. Rep. **8**:1184, 2018.

Crins, W.J.: The *Tamaricaceae* in the southeastern United States. - J. Arnold Arboretum **70**: 403-425, 1989.

Dierckxsens, N., Mardulyn, P., Smits, G.: NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. - Nucl. Acids Res. **45**: e18, 2017.

Dobrogojski, J., Adamiec, M., Luciński, R.: The chloroplast genome: a review - Acta Physiol. Plant. **42**: 98, 2020.
Du, Y.P., Bi, Y., Yang, F.P., Zhang, M.F., Chen, X.Q., Xue, J., Zhang, X.H.: Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. - Sci. Rep. **7**: 5751, 2017.
Duan, H., Guo, J., Xuan, L., Wang, Z., Li, M., Yin, Y., Yang, Y.: Comparative chloroplast genomics of the genus *Taxodium*. - BMC Genomics **21**: 114, 2020.
Fu, C.N., Li, H.T., Milne, R., Zhang, T., Ma, P.F., Yang, J., Li, D.Z., Gao, L.M.: Comparative analyses of plastid genomes from fourteen *Cornales* species: inferences for phylogenetic relationships and genome evolution. - BMC Genomics **18**: 956, 2017.
Gao, K., Li, J., Khan, W.U., Zhao, T.Y., Yang, X., Yang, X.Y., Guo, B., An, X.M.: Comparative genomic and phylogenetic analyses of *Populus* section Leuce using complete chloroplast genome sequences. - Tree Genet. Genomes **15**: 32, 2019.
Gaskin, J.F., Shafroth, P.B.: Hybridization of *Tamarix ramosissima* and *T. chinensis* (salt cedars) with *T. aphylla* (athel) (*Tamaricaceae*) in the southwestern USA determined from DNA sequence data. - Madroño **52**: 1-10, 2005.
Greiner, S., Lehwark, P., Bock, R.: OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. - Nucl. Acids Res. **47**: W59-W64, 2019.
Hahn, C., Bachmann, L., Chevreux, B.: Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. - Nucl. Acids Res. **41**: e129, 2013.
Huang, J., Yang, X., Zhang, C., Yin, X., Liu, S., Li, X.: Development of chloroplast microsatellite markers and analysis of chloroplast diversity in Chinese jujube (*Ziziphus jujuba* Mill.) and wild jujube (*Ziziphus acidojujuba* Mill.). - PloS ONE **10**: e0134519, 2015.
Channa, F.N., Shinwari, Z.K., Ali, S.I.: Phylogeny of *Tamaricaceae* using psbA-trnH nucleotide sequences. - Pak. J. Bot. **50**: 983-987, 2018.
KalamUrfi, M., Mujahid, M., Badruddeen, JuberAkhtar, Khalid, M., Khan, M.I., Usmani, A.: *Tamarix gallica*: for traditional uses, phytochemical and pharmacological potentials. - J. chem. pharm. Res. **8**: 809-814, 2016.
Katoh, K., Standley, D.M.: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. - Mol. Biol. Evol. **30**: 772-780, 2013.
Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R.: REPuter: the manifold applications of repeat analysis on a genomic scale. - Nucl. Acids Res. **29**: 4633-4642, 2001.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B.: PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. - Mol. Biol. Evol. **34**: 772-773, 2016.
Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E.: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. - Mol. Ecol. **11**: 2453-2465, 2002.
Mayonde, S.G., Cron, G.V., Gaskin, J.F., Byrne, M.J.: *Tamarix* (*Tamaricaceae*) hybrids: the dominant invasive genotype in southern Africa. - Biol. Invasions **18**: 3575-3594, 2016.
Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I.: VISTA: visualizing global DNA sequence alignments of arbitrary length. - Bioinformatics **16**: 1046-1047, 2000.
McKain, M., Johnson, M., Uribe-Convers, S., Eaton, D., Yang, Y.: Practical considerations for plant phylogenomics. - Appl. Plant Sci. **6**: e1038, 2018.
Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q.: IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. - Mol. Biol. Evol. **32**: 268-274, 2014.
Palmer, J.: Plastid chromosomes: structure and evolution. - In: Bogorad, L., Vasil, I.K. (ed): The Molecular Biology of Plastids. Pp. 5-53. Academic Press, San Diego - New York - Boston London - Sydney - Tokyo - Toronto 1990.
Reginato, M., Neubig, K., Majure, L., Michelangeli, F.: The first complete plastid genomes of *Melastomataceae* are highly structurally conserved. - Peer J. **4**: e2715, 2016.
Ronquist, F., Teslenko, M., Van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P.: MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. - Syst. Biol. **61**: 539-542, 2012.
Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A.: DnaSP 6: DNA sequence polymorphism analysis of large data sets. - Mol. Biol. Evol. **34**: 3299-3302, 2017.
Sheidai, M., Teymoori, H., Noormohammadi, Z., Mehrabian, A., Koohdar, F., Ghasemzadeh-Baraki, S.: Species delimitation in the genus *Tamarix*: morphological and molecular data. - Phytotaxa **343**: 101-115, 2018.
Somaratne, Y., Guan, D.L., Wang, W.Q., Zhao, L., Xu, S.Q.: Complete chloroplast genome sequence of *Xanthium sibiricum* provides useful DNA barcodes for future species identification and phylogeny. - Plant Syst. Evol. **305**: 949-960, 2019.
Strauss, S.H., Palmer, J.D., Howe, G.T., Doerksen, A.H.: Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. - Proc. nat. Acad. Sci. USA **85**: 3898-3902, 1988.
Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E., Fischer, A., Bock, R., Greiner, S.: GeSeq - versatile and accurate annotation of organelle genomes. - Nucl. Acids Res. **45**: W6-W11, 2017.
Timme, R.E., Kuehl, J.V., Boore, J.L., Jansen, R.K.: A comparative analysis of the *Lactuca* and *Helianthus* (*Asteraceae*) plastid genomes: identification of divergent regions and categorization of shared repeats. - Amer. J. Bot. **94**: 302-312, 2007.
Turmel, M., Otis, C., Lemieux, C.: Dynamic evolution of the chloroplast genome in the green algal classes *Pedinophyceae* and *Trebouxiophyceae*. - Genome Biol. Evol. **7**: 2062-2082, 2015.
Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Yong, S.K., Earl, A.M.: Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. - PloS ONE **9**: e112963, 2014.
Wang, Y., Liu,Y., Liu, S., Huang, H.: Molecular phylogeny of *Myricaria* (*Tamaricaceae*): implications for taxonomy and conservation in China. - Bot. Stud. **50**: 343-352, 2009.
Wellington, S.M.: WebSat – a web software for microsatellite marker development. - Bioinformation **3**: 282-283, 2009.
Xia, X.: DAMBE7: New and improved tools for data analysis in molecular biology and evolution. - Mol. Biol. Evol. **35**:1550-1552, 2018.
Xue, S., Shi, T., Luo, W.J., Ni, X.P., Iqbal, S., Ni, Z.J., Huang, X., Yao, D., Shen, Z.J., Gao, Z.H.: Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. - Hort. Res. **6**: 89, 2019.
Yang, X., Zhou, T., Su, X., Wang, G., Zhang, X., Guo, Q., Cao, F.: Structural characterization and comparative analysis of the

chloroplast genome of *Ginkgo biloba* and other gymnosperms. - J. Forest. Res. **32**: 765-778, 2020.

Yao, G., Jin, J.J., Li, H.T., Yang, J.B., Mandala, V.S., Croley, M., Mostow, R., Douglas, N.A., Chase, M.W., Christenhusz, M.J.M., Soltis, D.E., Soltis, P.S., Smith, S.A., Brockington, S.F., Moore, M.J., Yi, T.-S., Li, D.-Z.: Plastid phylogenomic insights into the evolution of *Caryophyllales*. - Mol. Phylogenet. Evol. **134**: 74-86, 2019.

Yusufoglu, H.S., Al-qasoumi, S.I.: Anti-inflammatory and wound healing activities of herbal gel containing an antioxidant *Tamarix aphylla* leaf extract. - Int. J. Pharmacol. **7**: 829-835, 2011.

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. **20**: 348-355, 2020.

Zhang, D., Zhang, Y., Gaskin, J.F., Chen, Z.: Systematic position of *Myrtama* Ovcz. & Kinz. based on morphological and nrDNA ITS sequence evidence. - Chinese Sci. Bull. **51**: 117-123, 2006.

Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., Zhang, W., Kim, K., Lee, S.C., Yang, T.J., Wang, Y.: The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. - Front. Plant. Sci. **7**: 306, 2016.