The Complete Chloroplast Genome of Tamarix Ramosissima and Comparative Analysis of Tamaricaceae Species

Le Wang
Qinghai University
Li Wang
Qinghai University
Zhihong Guo (guozhhong@lzb.ac.cn)
Gaolan Station of Agricultural and Ecological Experiment, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

Original Article

Keywords: Tamarix ramosissima, chloroplast genome, phylogenomics

Posted Date: August 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60520/v1

Abstract

Background: *Tamarix ramosissima* is a deciduous shrub resided 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. Chloroplast (cp) DNA is highly conserved in structure and gene arrangement, making cp genomic data valuable resources for species delimitation and phylogenetics. The cp genome of *T. ramosissima* was de novo assembled with the aim of providing reference and data resource for further cp-derived marker development and species delimitation of *Tamarix*.

Results: Here, the complete chloroplast (CP) genome of *T. ramosissima* was sequenced and analyzed, showing a size of 156150 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 26554 bp, separated by a large single copy (LSC) region of 84795 bp, and a small single copy (SSC) region of 18247 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 repeats (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 four other Tamaricaceae species, a number of divergence hotspots were identified among these plastomes. Together with the 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*.

Conclusions: Taken together, our study provides valuable genomic resources to deepen the understanding of the plant photosynthetic mechanism and phylogenomics.

Background

Tamarix plants belong to the Tamaricaceae family and are ancient species native to 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 (Gaskin et al. 2004). Although several *Tamarix* species are considered invasive in the United States, the supplementing of *Tamarix* plants in traditional medicine reveals their value in medical applications (Bahramsoltani et al. 2020). For instance, the leaves of *Tamarix* species are reported to have pharmacological effects such as detoxification, expelling rheumatism, and diuresis promotion in traditional Chinese medicine. While in Middle Eastern countries, an extract of *Tamarix* leaves has been used as an antiseptic agent (KalamUrfi et al. 2016). In addition, the bark of *Tamarix aphylla*, which differs in chemical constitution from the leaf, has been used as a herbal remedy for alleviating eczema capitis (Yu and Al 2011).

Chloroplasts (CPs) are photosynthetic organelles converting light energy from the sun into chemical energy stored in carbohydrates. Therefore, the cp proteome is abundant with enzymatic machineries

catalyzing photosynthesis, and proteins responsible for starch synthesis, amino acid synthesis, fatty acid synthesis, phytohormone synthesis, and DNA and RNA synthesis (Yu et al. 2019). The multifaceted roles cp played in plant metabolism make it a key organelle for plant productivity and survival. Metabolic processes in the cp are, however, regulated by both the nucleus and the cp genome itself (Barajas-López et al. 2013). The cp genome, which is also known as cpDNA, is inherited maternally in most plants. Double-stranded cpDNAs usually exhibit a circular structure, with genome sizes ranging from 107 to 218 kB (Turmel et al. 2015). Further alignment of published cp genomes reveals their conserveness in gene arrangement. In angiosperms, most plastomes share 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 (Bogorad and Vasil 1990). Sequences of LSC and SSC are consistently conserved across most plant species. In Gymnospermae, however, the inverted repeats could vary substantially between species, and the changes in inverted repeat regions often lead to massive adjustment in DNA arrangement (Yang et al. 2020). For example, when lacking a large inverted repeat, extensive rearrangements of chloroplast DNA are observed in two conifer plants (Strauss et al. 1988).

The moderate evolutionary rate of cp genomes makes them potentially valuable resources in phylogenetic studies (Duan et al. 2020). In comparison with nuclear genomes, cpDNAs are smaller in size, and contain 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.

In the present study, we report the first complete cp genome of *Tamarix ramosissima* and compare it with those of family Tamaricaceae. Through comparison, we elicidate the differences among the cp genomes of five Tamaricaceae species. Data in this study would facilitate the development of cp-derived molecular markers and the elucidation of phylogenetic relationship among Tamaricaceae species.

Materials And Methods

Plant materials, DNA isolation and next-generation sequencing

Fresh leaves of *T. ramosissima* 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 scale-like leaves were frozen immediately in liquid nitrogen and kept at -80 °C until DNA extraction. Subsequent genomic DNA extraction was performed per the manufacturer's instructions using the Tiangen Plant Genomic DNA Kit (Tiangen Biotech Co., Beijing, China). The extracted DNA was then submitted for NGS library construction and paired-end sequencing using an Illumina Hiseq 2500 platform (Illumina Inc., San Diego, CA, USA).

Genome assembly and annotation

The raw data was trimmed with Trimmomatic software 0.36 to remove adaptor sequence and low-quality reads. A total of 8,261,466,300 bases of clean data was generated after filtering, resulting in 27,538,221 clean reads. The resulting clean reads were mapped against the reference chloroplast genome (*T. chinensis*, GenBank accession number: NC_040943) 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 2.7.1 (Dierckxsens et al. 2016), 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 1.23 (Walker et al. 2014).

The preliminary gene annotation of the draft *T. ramosissima* cp genome was performed using the GeSeq tool (Tillich et al. 2017). The annotations were then 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). The 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 paniculate* (NC_041270), and *T. chinensis* (NC_040943) by using the mVISTA program in 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 V7.450 (Katoh and Standley 2013). The nucleotide diversity values (Pi) among the cp genomes were then calculated using DnaSP 6 (Rozas et al. 2017), with the window length set to 800 bp and step size set to 200 bp.

RSCU analysis

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, a RSCU value of less than 1 shows that a codon is used less frequently than expected. The RSCU value of each codon of the five Tamaricaceae cp genomes were 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, prediction was performed with MISA-web (MIcroSAtellite identification tool-web), an online identification tool. SSR motifs were searched within the cp genomes according to the criteria as follows: for mononucleotide repeats, ≥ 10 units of repeats are required; for dinucleotide repeats, ≥ 8 units of repeats are required; for trinucleotide and tetranucleotide repeats; ≥ 4 units of repeats are required; and for pentanucleotide and hexanucleotide repeats, ≥ 3 units of repeats are required (Wellington Santos 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. All sequences were first aligned with the MAFFT program. The nucleotide alignment was then subjected to phylogenetic analysis using the MEGA X program (Kumar et al. 2018). The phylogenetic relationship was inferred using both the neighbor-jointing (NJ) and the maximum likelihood (ML) methods.

Results

General features of the T. ramosissima chloroplast genome

The complete cp genome of *T. ramosissima* was 156,150 bp in length, displayed a typical quadripartite structure, in which a small-copy region (SSC, 18247 bp) and a large single-copy region (LSC, 84795 bp) were separated by two identical inverted repeats (IR, 26554 bp) (Fig. 1). After comparing the size and structure of cp genomes from Tamaricaceae species, we found that the lengths of the five plastomes varied from 154533 bp to 156167 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.

Table 1
Comparison of general features of five Tamaricaceae plastomes

Species	Total	LSC	IR	SSC	Total	Protein coding genes	tRNA	rRNA	GC%
Tamarix ramosissima	156150	84795	53108	18247	130	85	37	8	36.5%
Tamarix chinensis	156167	84768	53152	18247	130	85	37	8	36.5%
Hololachna songarica	155596	85903	52138	17555	130	85	37	8	36.8%
Reaumuria trigyna	154533	84811	52116	17607	130	85	37	8	37.0%
Myricaria paniculata	154651	84379	49588	20684	130	85	37	8	36.3%

All the genes annotated in the *T. ramosissima* cp genome are 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 8 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*). *rps12* was the only trans-spliced gene in the *T. ramosissima* plastome.

Table 2
List of genes annotated in the chloroplast genome of *T. ramosissima.*

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;atpI*	6
NADH dehydrogenase	*ndhA*; ndhB*(× 2); ndhC; ndhD;ndhE; ndhF ndhG; ndhH; ndhI; ndhJ; ndhK*	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); trna-UAG*	37
DNA dependent RNA polymerase	*rpoA; rpoB; rpoC1*; rpoC2*	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; rpl36*	11
Proteins of unknown function	*ycf1, ycf2 (× 2), ycf3**, ycf4*	5
Other genes	*accD; ccsA; cemA; clpP**; matK; infA*	6

* indicates genes containing one intron; ** indicates genes containing two introns;^{T} indicates trans-spliced Genes; ×2 indicates genes have two copies

Codon usage

Codon usage of protein coding sequences in the *T. ramosissima* cp genome was analyzed with DAMBE software. Overall, 64 codons, corresponding to the 20 amino acids, were found presence in the *T. ramosissima* plastome. A total of 24724 codons were identified for all the protein coding sequences

(including the stop codons). Leucine (2651; 10.72%) was the most abundant amino acid, whereas cysteine (283; 1.14%) was the least abundant. The relative synonymous codon usage (RSCU) value, which was positively correlated with the quantity 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).

Table 3
Codon content of 20 amino acid and stop codons in all protein-coding genes of the five Tamaricaceae cp genomes.

Amino acid	Codon	*T. ramosissima*	*T. chinensis*	*R. trigyna*	*H. songarica*	*M. paniculata*
		RSCU[a]				
Stop[b]	UGA	0.622	0.679	0.714	0.786	0.532
Stop[b]	UAG	0.732	0.679	0.714	0.679	0.646
Stop[b]	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

[a]Relative synonymous codon usage; [b]stop codon

Amino acid	Codon	*T. ramosissima*	*T. chinensis*	*R. trigyna*	*H. songarica*	*M. paniculata*
		RSCU[a]				
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

[a]Relative synonymous codon usage; [b]stop codon

Amino acid	Codon	*T. ramosissima*	*T. chinensis*	*R. trigyna*	*H. songarica*	*M. paniculata*
		RSCU[a]				
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

[a]Relative synonymous codon usage; [b]stop codon

Interspecific variation among Tamaricaceae cp genomes

The newly sequenced *T. ramosissima* cp genome was compared with those of the other four Tamaricaceae species using the mVISTA program (Fig. 2). The comparison revealed the high nucleotide conservation between the cp genome of *T. ramosissima* and *T. chinensis*. Furthermore, coding regions were found to be more conserved than non-coding regions, while the SSC and LSC regions were more divergent than the IRs regions.

To further 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. As illustrated in Fig. 3, the LSC region and the SSC region showed higher nucleotide diversity than the two IR regions. Six regions with high Pi values were identified as divergence hotspots (Fig. 3). The *rpl32-tRNA-UAG* region, with a Pi value of 0.195, was the most divergent part detected. Four intergenic regions (*tRNA-GCC-tRNA-CAU, psbK-psbI, tRNA-GAA-ndhJ, rps15-ycf1*) and one gene region (*rpl16*) had high Pi values and were also identified as divergence

hotspots. The divergence hotspots identified could be developed as potential markers for species delimitation of the Tamaricaceae species.

SSR and repeat structure analysis

The total number of SSRs identified in the five Tamaricaceae cp genomes ranged from 59 to 67 (Fig. 4). Among these SSRs, mononucleotide repeats were the dominant type, and A/T repeats accounted for nearly 60% of all SSRs identified. Di-nucleotide repeats were the second most abundant motif types identified, constituting 13.3–20.3 percent of the total SSRs. Most of the di-nucleotide repeats were also AT-rich. Tri-, tetra-, and penta-nucleotide repeats comprised a relatively small part of the SSRs detected (Fig. 4).

Long repeats in the five cp genomes were also analyzed with the REPuter software. As shown in Fig. 5, 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 the length > 45 bp were only detected in the plastomes of *H. songarica, R. trigyna*, and *M. paniculate*.

Phylogenetic relationship

A plastome-based phylogenomic tree was constructed with MEGA X to analyze the phylogenetic relationship (Fig. 6). Among the five Tamaricaceae species, two genus *Tamarix* plants, *T. ramosissima* and *T. chinensis*, were clustered together. The two genus *Reaumuria* species, *R. trigyna*, and *H. songarica*, were also monophyletic. *M. paniculate* was inferred to have a closer relationship with *Tamarix* species, according to the phylogeny. The topological structure of the ML tree was consistent with the constructed NJ tree.

Discussion

In the present study, we obtained the complete chloroplast genome of *T. ramosissima* by Illumina sequencing and compared it with those of other four Tamaricaceae species. As shown in Table 1, the plastome size ranged from 154533 bp to 156167 bp with GC content varying slightly from 36.3–37.0%. Each of the five Tamaricaceae cp genomes encoded 130 genes, including 85 protein coding genes, 37 tRNA genes, and 8 rRNA genes. 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 (Bi et al. 2018; Huang et al. 2015). 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 were localized in the LSC region. These repeat sequences, varying in types and sizes, may promote evolution of the plastids of Tamaricaceae species by generating new variation.

Due to the scarcity of cp genomic data, phylogenetic study at 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 the public database. Our analysis revealed that two *Tamarix* species, *T. ramosissima* and *T. chinensis*, were clustered together. *Tamarix* plants were more closely related to *M. paniculate*, a species that shows resemblance in appearance to plants in the genus *Tamarix.* As a genus erected from *Tamarix* (Liu 2009), the close relationship between *Tamarix* and *Myricaria* has also been confirmed in previous research (Naz 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.

Conclusions

In the present study, the complete cp genome of *T. ramosissima* was first obtained by Illumina sequencing, and comprehensively compared with those from other four *Tamaricaceae* species. The newly sequenced *T. ramosissima* plastome is 156150 bp in length, encoding 130 genes. The deciphered cp genome exhibits a typical quadripartite structure, consisting a large single copy, a small single copy, and two inverted repeats. Thirty-two repeat sequences and 64 SSRs are identified in the plastome, which could be used as potential molecular markers for species discrimination and evolutionary studies. Phylogenetic relationship among *Tamaricaceae* species were re-investigated using the complete cp genome sequences. *T. ramosissima* has a close relationship with *T. chinensis*, which is strongly supported by high bootstrap values. In summary, the complete *T. ramosissima* cp genome assembled

provides additional information for further studies, which will benefit future phylogenetic and evolutionary research.

Abbreviations

cp
Chloroplast; SSC:Small single copy; LSC:Large single copy; IR:inverted repeat; RSCU:relative synonymous codon usage; SSR:simple sequence repeat; NJ:neighbor-jointing; ML:maximum likelihood

Declarations

Author contributions

LW analyzed the data, and drafted the manuscript. LW provided advices on our analysis. ZHG provided the samples and supervised the work.

Funding

This research was supported by 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).

Availability of data and materials

The complete cp genome sequence of *T. ramosissima* has been submitted to GenBank and deposited under the accession number MN726883.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bahramsoltani R, Kalkhorani M, Abbas Zaidi SM, Farzaei MH, Rahimi R (2020) The genus Tamarix: Traditional uses, phytochemistry, and pharmacology. J Ethnopharmacol 246:112245

2. Barajas-López JdD, Blanco NE, Strand Å (2013) Plastid-to-nucleus communication, signals controlling the running of the plant cell. BBA-Mol Cell Res 1833(2):425–437
3. Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH (2018) Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on Fritillaria. Sci Rep-UK 8(1):1184
4. Bogorad L, Vasil IK (1990) The Molecular Biology of Plastids, 1st edn. Academic Press, San Diego
5. Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45(4):e18
6. Du YP, Bi Y, Yang FP, Zhang MF, Chen XQ, Xue J, Zhang XH (2017) Complete chloroplast genome sequences of Lilium: insights into evolutionary dynamics and phylogenetic analyses. Sci Rep-UK 7(1):5751
7. Duan H, Guo J, Xuan L, Wang Z, Li M, Yin Y, Yang Y (2020) Comparative chloroplast genomics of the genus Taxodium. BMC Genom 21(1):114
8. Fu CN, Li HT, Milne R, Zhang T, Ma PF, Yang J, Li DZ, Gao LM (2017) Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. BMC Genom 18(1):956
9. Gao K, Li J, Khan WU, Zhao TY, Yang X, Yang XY, Guo B, An XM Comparative genomic and phylogenetic analyses of Populus section Leuce using complete chloroplast genome sequences. Tree Genet Genomes 15(3):32
10. Gaskin JF, Ghahremani-nejad F, Zhang D-y, Londo JP (2004) A Systematic Overview of Frankeniaceae and Tamaricaceae from Nuclear rDNA and Plastid Sequence Data. Ann Mo Bot 91(3):401–409
11. Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47(W1):W59–W64
12. Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Res 41(13):e129
13. Huang J, Yang X, Zhang C, Yin X, Liu S, Li X (2015) 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(9):e0134519
14. KalamUrfi M, Mujahid M, Badruddeen, JuberAkhtar, Khalid M, Khan MI, Usmani A (2016) Tamarix gallica: For traditional uses, phytochemical and pharmacological potentials. J Chem Pharm Res 8(1):809–814
15. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780
16. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35(6):1547–1549

17. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29(22):4633–4642
18. Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11(12):2453–2465
19. Liu Y (2009) Molecular phylogeny of Myricaria (Tamaricaceae): implications for taxonomy and conservation in China. Bot Stud 50(3):343–352
20. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16(11):1046–1047
21. McKain M, Johnson M, Uribe-Convers S, Eaton D, Yang Y (2018) Practical considerations for plant phylogenomics. Appl Plant Sci 6(3):e1038
22. Naz F, Shinwari Z, Ali I (2018) Phylogeny of tamaricaceae using psbA-trnH nucleotide sequences. Pak J Bot 50(3):983–987
23. Reginato M, Neubig K, Majure L, Michelangeli F (2016) The first complete plastid genomes of Melastomataceae are highly structurally conserved. Peer J 4:e2715
24. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34(12):3299–3302
25. Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ (2019) Complete chloroplast genome sequence of Xanthium sibiricum provides useful DNA barcodes for future species identification and phylogeny. Plant Syst Evol 305:949–960
26. Strauss SH, Palmer JD, Howe GT, Doerksen AH (1988) Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. P Natl Acad Sci USA 85(11):3898–3902
27. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones E, Fischer A, Bock R, Greiner S (2017) GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45(W1):W6–W11
28. Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. Am J Bot 94(3):302–312
29. Turmel M, Otis C, Lemieux C (2015) Dynamic Evolution of the Chloroplast Genome in the Green Algal Classes Pedinophyceae and Trebouxiophyceae. Genome Biol Evol 7(7):2062–2082
30. Walker BJ et al (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PloS One 9(11):e112963
31. Wellington Santos M (2009) WebSat-a web software for microsatellite marker development. Bioinformation 3(6):282–283
32. Xia X (2018) DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution Mol Biol Evol 35(6):1550–1552

33. Xue S, Shi T, Luo WJ, Ni XP, Iqbal S, Ni ZJ, Huang X, Yao D, Shen ZJ, Gao ZH (2019) Comparative analysis of the complete chloroplast genome among Prunus mume, P. armeniaca, and P. salicina. Hortic Res 6:89
34. Yang X, Zhou T, Su X, Wang G, Zhang X, Guo Q, Cao F (2020) Structural characterization and comparative analysis of the chloroplast genome of Ginkgo biloba and other gymnosperms. J Forestry Res. doi:10.1007/s11676-019-01088-4
35. Yao G et al (2019) Plastid phylogenomic insights into the evolution of Caryophyllales. Mol Phylogenet Evol 134:74–86
36. Yu H, Al S (2011) Anti-inflammatory and Wound Healing Activities of Herbal Gel Containing an Antioxidant Tamarix aphylla Leaf Extract. Int J Pharmacol 7(8):829–835
37. Yu X, Zuo L, Lu D, Lu B, Yang M, Wang J (2019) Comparative analysis of chloroplast genomes of five Robinia species: Genome comparative and evolution analysis. Gene 689:141–151
38. Zhang D, Zhang Y, Gaskin JF, Chen Z (2006) Systematic position of Myrtama Ovcz. & Kinz. based on morphological and nrDNA ITS sequence evidence. 51(A01):117–123 Chinese Sci Bull
39. Zhang Y et al (2016) The Complete Chloroplast Genome Sequences of Five Epimedium Species: Lights into Phylogenetic and Taxonomic Analyses. Front Plant Sci 7:306

Figures

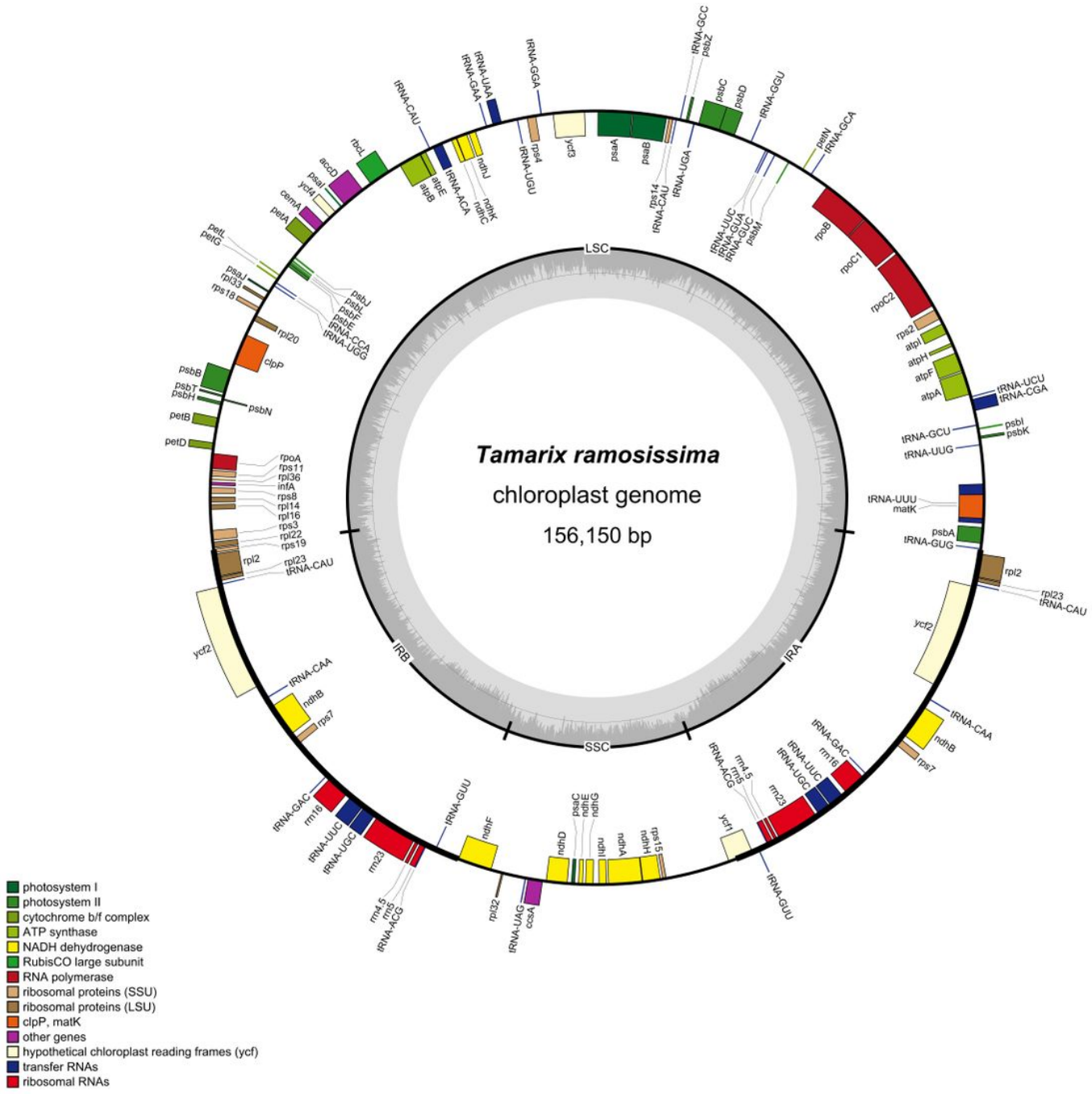

Figure 1

Structural map of the T. ramosissima chloroplast genome. Genes drawn inside of the circle are transcribed clockwise, while those outside are transcribed anticlockwise. The inverted repeat regions (IRa and IRb), which are separated by the large single copy (LSC) region and the small single copy (SSC) region, are denoted with thick lines. Genes belonging to different functional groups are color-coded accordingly.

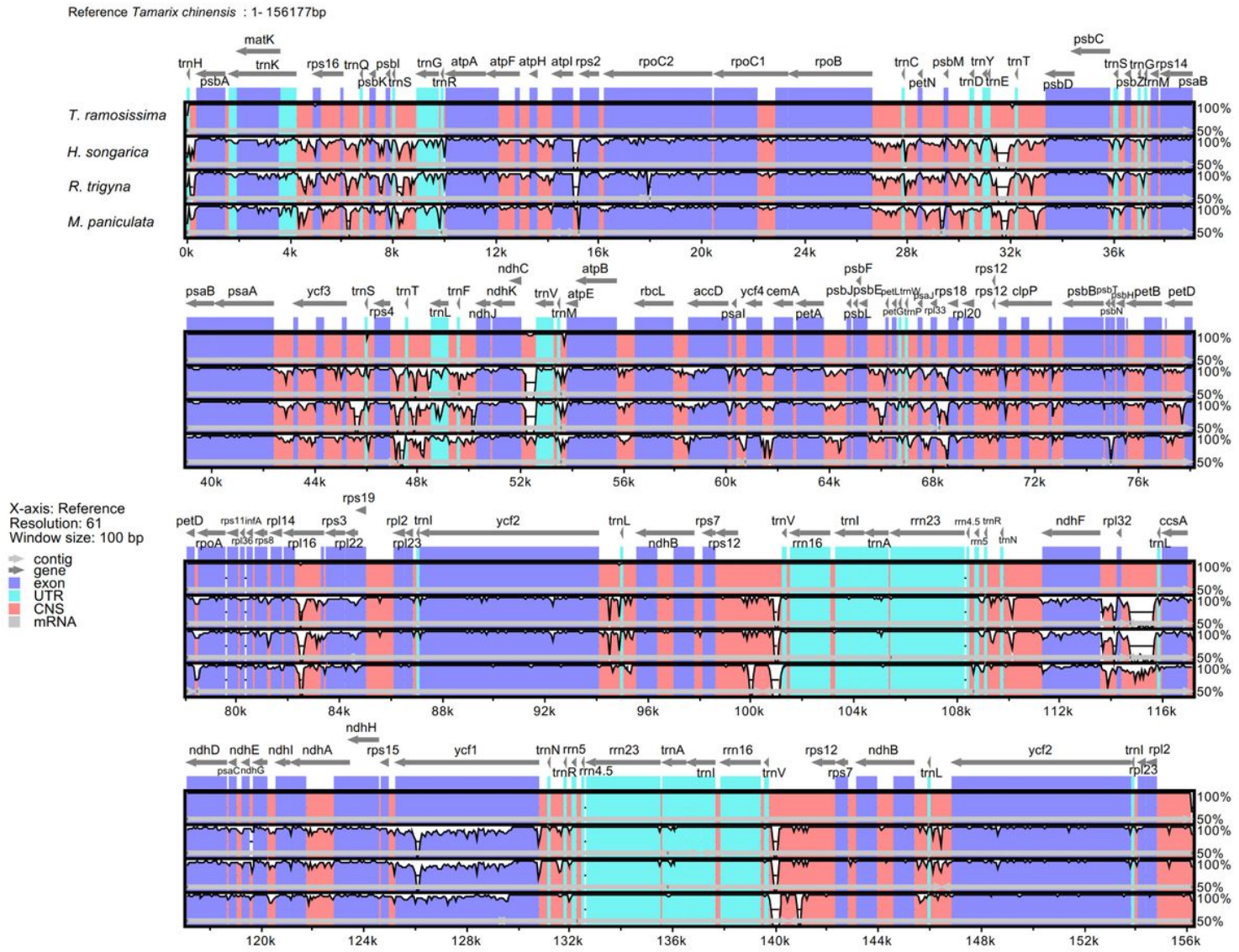

Figure 2

Comparsion of the five Tamaricaceae plastomes using the mVISTA program. The cp genome of T. chinensis was used as the reference. The x-axis represents the base sequence of alignment, and the y-axis represents the percentage of identity, ranging from 50 to100%.

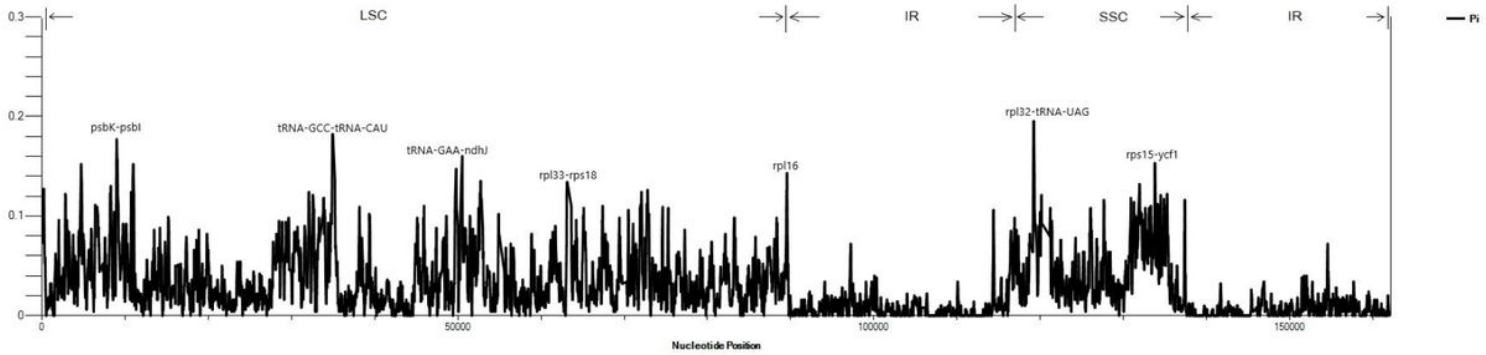

Figure 3

Nucleotide variability (Pi) values of the five Tamaricaceae chloroplast genomes. The x-axis represents the base sequence of alignment, and the y-axis represents Pi value.

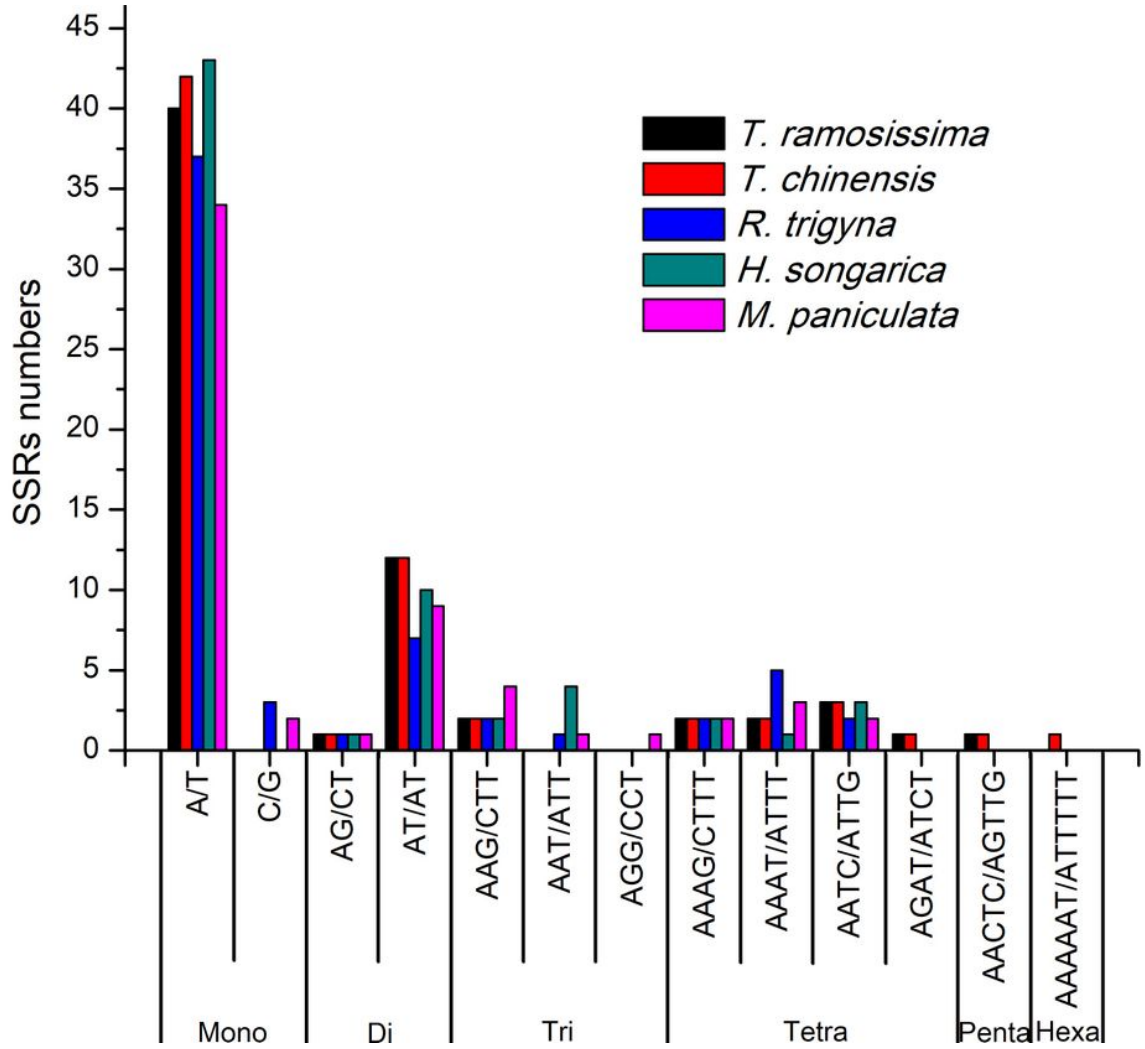

Figure 4

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

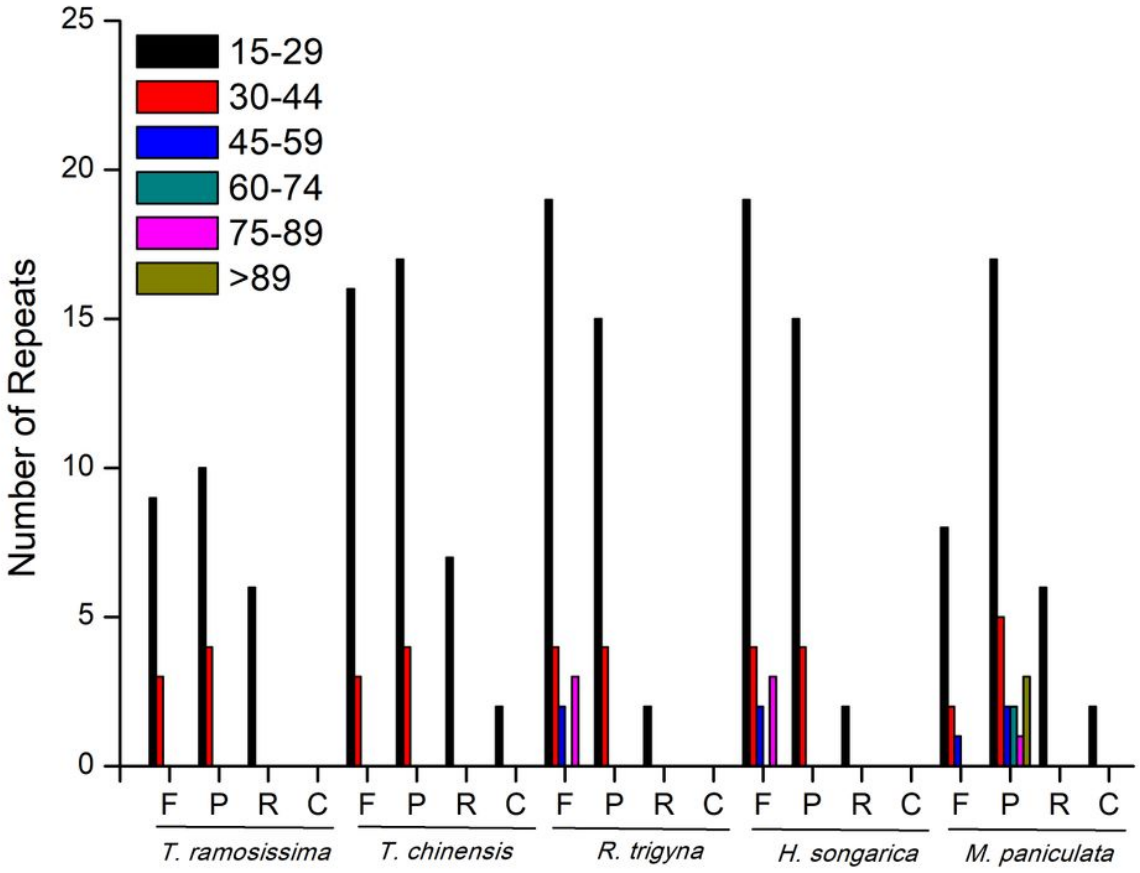

Figure 5

Long repetitive sequences identified in the five Tamaricaceae cp genomes. F, P, R, and C reprent the repeat types of forward (F), palindrome (P), reverse (R), and complement (C), respectively. Repeats with different lengths are coded with corresponding colors.

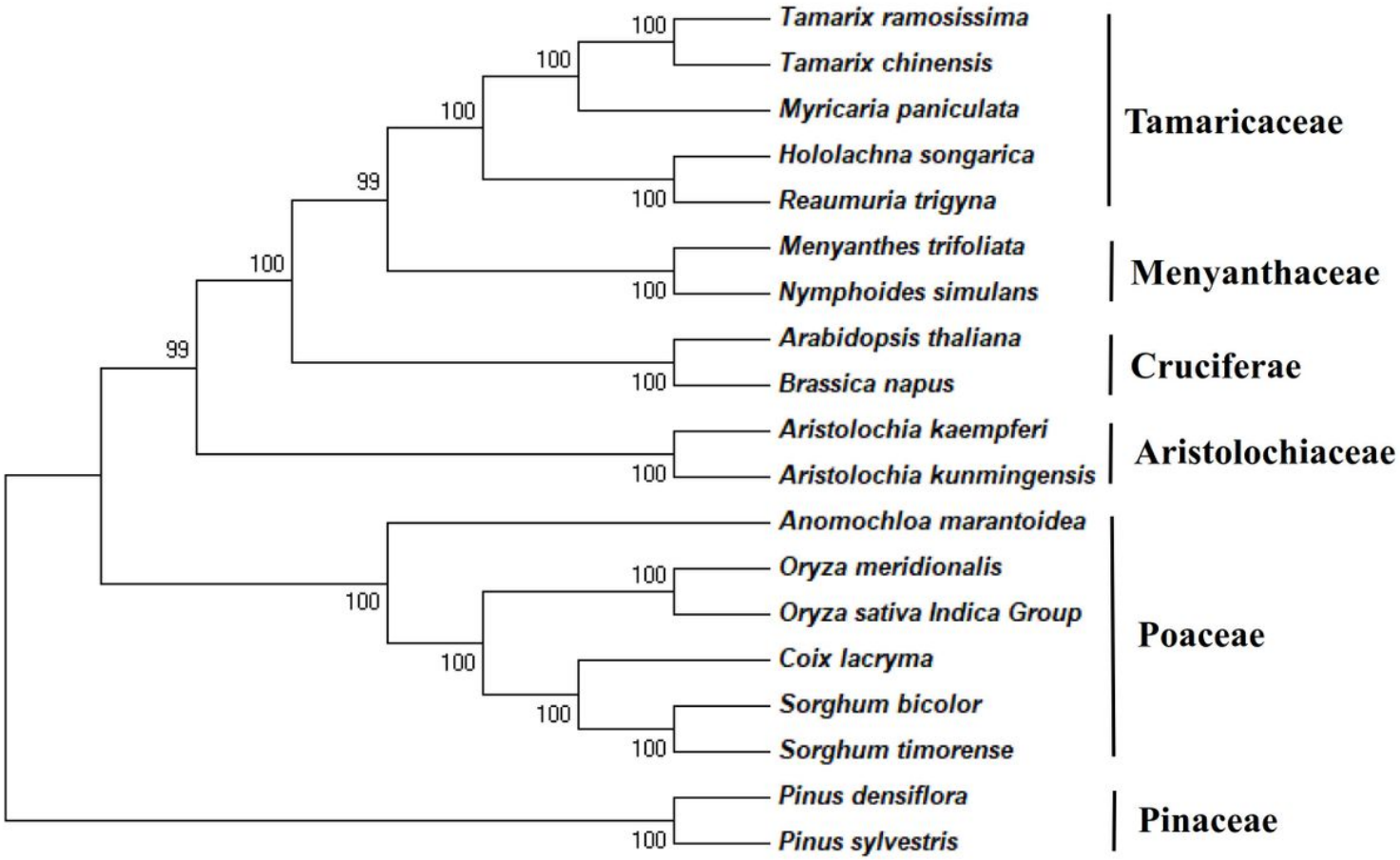

Figure 6

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