



The complete chloroplast genome of *Saxifraga sinomontana* (Saxifragaceae) and comparative analysis with other Saxifragaceae species

Yan Li^{1,2} · Liukun Jia^{1,2} · Zhihua Wang^{1,2} · Rui Xing¹ · Xiaofeng Chi¹ · Shilong Chen¹ · Qingbo Gao^{1,3}

Received: 10 July 2019 / Revised: 24 September 2019 / Accepted: 26 September 2019 / Published online: 25 October 2019
© The Author(s) 2019

Abstract

Saxifraga sinomontana J.-T. Pan & Gornall belongs to *Saxifraga* sect. *Ciliatae* subsect. *Hirculoideae*, a lineage containing ca. 110 species whose phylogenetic relationships are largely unresolved due to recent rapid radiations. Analyses of complete chloroplast genomes have the potential to significantly improve the resolution of phylogenetic relationships in this young plant lineage. The complete chloroplast genome of *S. sinomontana* was de novo sequenced, assembled and then compared with that of other six Saxifragaceae species. The *S. sinomontana* chloroplast genome is 147,240 bp in length with a typical quadripartite structure, including a large single-copy region of 79,310 bp and a small single-copy region of 16,874 bp separated by a pair of inverted repeats (IRs) of 25,528 bp each. The chloroplast genome contains 113 unique genes, including 79 protein-coding genes, four rRNAs and 30 tRNAs, with 18 duplicates in the IRs. The gene content and organization are similar to other Saxifragaceae chloroplast genomes. Sixty-one simple sequence repeats were identified in the *S. sinomontana* chloroplast genome, mostly represented by mononucleotide repeats of polyadenine or polythymine. Comparative analysis revealed 12 highly divergent regions in the intergenic spacers, as well as coding genes of *matK*, *ndhK*, *accD*, *cemA*, *rpoA*, *rps19*, *ndhF*, *ccsA*, *ndhD* and *ycf1*. Phylogenetic reconstruction of seven Saxifragaceae species based on 66 protein-coding genes received high bootstrap support values for nearly all identified nodes, suggesting a promising opportunity to resolve infrasectional relationships of the most species-rich section *Ciliatae* of *Saxifraga*.

Keywords Comparative genomics · Genome assembly · Phylogenetic analysis · *Saxifraga* sect. *Ciliatae*

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40415-019-00561-y>) contains supplementary material, which is available to authorized users.

✉ Qingbo Gao
qbgao@nwipb.cas.cn

Yan Li
liyan715@mails.ucas.ac.cn

Liukun Jia
jjaliukun@qq.com

Zhihua Wang
304339275@qq.com

Rui Xing
xingrui@nwipb.cas.cn

Xiaofeng Chi
xfchi@nwipb.cas.cn

Shilong Chen
slchen@nwipb.cas.cn

1 Introduction

Saxifraga L. is the largest genus of Saxifragaceae, consisting of 450–500 species in at least 13 sections (Tkach et al. 2015; Ebersbach et al. 2017a; Gao et al. 2017; Li et al. 2018). The most species-rich section of *Saxifraga* is

¹ Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China

² University of Chinese Academy of Sciences, Beijing 100039, China

³ Qinghai Provincial Key Laboratory of Crop Molecular Breeding, Xining 810001, China

Saxifraga sect. *Ciliatae* Haw. (ca. 175 species) (Pan et al. 2001), whose diversity center is the Qinghai-Tibetan Plateau (QTP) and Himalayas (Gao et al. 2015; Ebersbach et al. 2018). In addition, *Saxifraga* sect. *Ciliatae* is supposed to have experienced recent rapid radiations in the QTP region, particularly in *Saxifraga* sect. *Ciliatae* subsect. *Hirculoideae* Engl. & Irmsch. (Gao et al. 2015; Ebersbach et al. 2017b, 2018). Rapid radiations are usually associated with low genetic divergence among closely related species and clades (DeChaine et al. 2013), and thus, phylogenetic relationships could not be well resolved based on a limited number of DNA markers. This is also the case in *Saxifraga* sect. *Ciliatae*. Concerning to *Saxifraga* sect. *Ciliatae* subsect. *Hirculoideae*, relationships of its ca. 110 species are largely unresolved based on traditionally universal DNA markers (Gao et al. 2015). Furthermore, enigmatic species from *Saxifraga* sect. *Ciliatae* subsect. *Hirculoideae* could not be identified based on universal barcodes. New genomic resources are urgent to assist phylogenetic reconstruction and taxonomic classification of *Saxifraga* sect. *Ciliatae*. Here, we report the complete chloroplast genomes of seven species in *Saxifraga* sect. *Ciliatae*.

Analyses of complete chloroplast genomes have the advantage to significantly improve the resolution of phylogenetic relationships in large, complex plant lineages (Jansen et al. 2007; Doorduyn et al. 2011), even in enigmatic taxa (Dong et al. 2018). Chloroplast genomes are closed circular DNA molecules in most angiosperms, which have a conserved quadripartite structure consisting of a large single-copy (LSC) region and a small single-copy (SSC) region interspersed between a pair of inverted repeats (IR) (Palmer 1985; Wicke et al. 2011). The chloroplast genome size of angiosperms ranges from 120 to 160 kb, encoding 110–130 genes (Palmer 1985). Due to the characteristics of uniparental inheritance, haploid nature, conserved structure and gene content, small genome size, chloroplast genomes have been widely applied in phylogenetic reconstructions (Burke et al. 2016; Simmons 2017; Dong et al. 2018), molecular evolution (Huang et al. 2014; Walker et al. 2014) and superbarcoding studies (Hernández-León et al. 2013). With the recent advent of next-generation sequencing (NGS) techniques, the complete chloroplast genome sequences have grown rapidly. However, up to now, only one chloroplast genome of *Saxifraga* has been reported, *Saxifraga stolonifera* Curtis (GenBank accession no. NC_037882), a member of *S.* sect. *Irregulares* Haw. No chloroplast genome has been reported for the most species-rich sect. *Ciliatae* of *Saxifraga*.

Saxifraga sinomontana J.-T. Pan & Gornall, belonging to *S.* sect. *Ciliatae* subsect. *Hirculoideae* is a prominent element of the high-altitude habitats throughout the QTP-Himalayas region. This perennial herb is extraordinarily variable in morphology and shows high level of intraspecific genetic diversity, which was considered to have

experienced recent rapid intraspecific differentiation associated with Quaternary climatic oscillations (Li et al. 2018). Here, we de novo sequenced and assembled the complete chloroplast genome of *S. sinomontana* using an Illumina sequencing platform. Comparative analysis between this newly sequenced chloroplast genome and other six reported Saxifragaceae chloroplast genomes was implemented, and the putative performance for phylogenetic study on *Saxifraga* and Saxifragaceae was assessed.

2 Materials and methods

Sample collection, genome sequencing and assembly

Fresh leaves of *Saxifraga sinomontana* were sampled in Xuebudala pass, Longzi xian, Xizang Autonomous Region, China (28°37'58.6"N, 92°13'09.2"E). Leaves were collected from a single individual, then dried in silica gel. Voucher specimen was deposited in the herbarium of Northwest Institute of Plateau Biology (HNWP), Xining, Qinghai, China.

Total genomic DNA was isolated from silica-dried leaves using the modified CTAB method (Doyle and Doyle 1987). Genomic library was then prepared using the TruSeq Library Construction Kit (Illumina, San Diego, California, USA) following the manufacturer's instructions. Briefly, ca. 5 µg of genomic DNA was fragmented via ultrasound, followed by purification using CASpure PCR Purification Kit (Chaoshi-Bio, Shanghai, China) and end repair with poly-A on the 3' ends. The DNA fragments were then linked to adapters, extracted at specific size after agarose gel electrophoresis and amplified by PCR to generate a sequencing library. Genomic and chloroplast DNA were sequenced using the Illumina NovaSeq platform (Novogene, Tianjin, China), yielding ca. 5 Gb 150-bp paired-end reads from a library of approximately 350-bp DNA fragments.

The BWA-MEM algorithm was performed to map chloroplast DNA reads against the complete chloroplast genome sequence of *S. stolonifera* (NC_037882) as implemented in BWA 0.7.17 (Li and Durbin 2009). A total of 7,204,234 paired-end reads were recovered and then assembled using SPAdes 3.13.0 (Bankevich et al. 2012). Six contigs, ranging in size from 2993 to 60,980 bp, were obtained, followed by a further scaffolding using SSPACE-basic 2.0 (Boetzer et al. 2011). The gaps between the de novo contigs, as well as genomic regions located at the junction between the two contigs were filled and verified by Sanger sequencing. The primers used were designed using Primer-BLAST (Ye et al. 2012) and are listed in Supplementary Table S1.

Genome annotation and simple sequence repeats analysis

The genome was annotated using the program GeSeq (Tillich et al. 2017). The predicted annotations were verified

using BLAST search against cp genomes of other closely related species to validate positions of questionable start and stop codons. The circular cp genome map was drawn using OGDRAW (Greiner et al. 2019). The complete chloroplast genome of *S. sinomontana* was deposited into GenBank, with the accession of MN104589.

Simple sequence repeats (SSRs) were detected using MISA (Thiel et al. 2003) with minimal repeat numbers of 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexa-nucleotides, respectively.

Comparison with other Saxifragaceae chloroplast genomes – Sequence length of complete chloroplast genomes, as well as each individual regions were compared among *S. sinomontana* and six other Saxifragaceae species, *S. stolonifera* (NC_037882), *Bergenia scopulosa* T. P. Wang (NC_036061), *Chrysosplenium aureobracteatum* Y. I. Kim & Y. D. Kim (NC_039740), *Heuchera parviflora* Bartling var. *saurensis* R. A. Folk (KR478645), *Mukdenia rossii* (Oliv.) Koidz. (NC_037495), *Oresitrophe rupifraga* Bunge (NC_037514). The percentage of GC-content was calculated using MEGA version 7.0.26 (Kumar et al. 2016). The percentage of sequence identity among complete chloroplast genomes of the seven Saxifragaceae species was comparatively analyzed and plotted using the program mVISTA (Frazer et al. 2004), with alignment algorithm of LAGAN (Brudno et al. 2003), a cut-off of 70% identity and annotation of *S. sinomontana* as reference. Comparison of junction sites of LSC, IR and SSC regions among these Saxifragaceae chloroplast genomes was implemented using the program IRscope (Amiryousefi et al. 2018).

Phylogenetic analysis – To estimate the putative performance of chloroplast genomes on phylogeny resolution of Saxifragaceae species, phylogenetic relationships of the above-mentioned species were inferred, using *Ribes fasciculatum* Siebold & Zucc. var. *chinense* Maxim. (MH191388) and *Itea chinensis* Hook. & Arn. (NC_037884) as outgroups. A set of 66 protein-coding genes (*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *psbA*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*, *petA*, *petB*, *petD*, *petG*, *petL*, *petN*, *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *rbcL*, *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps14*, *rps15*, *rps16*, *rps18*, *rps19*, *rpoA*, *rpoB*, *rpoC1* and *rpoC2*) shared by all of the species were extracted from all the selected chloroplast genomes. Individual PCGs were concatenated into a single matrix and then aligned using MEGA version 7.0.26 (Kumar et al. 2016). Phylogenetic analysis was performed by mean of Maximum Likelihood (ML) using RAxML version 8.1.21 (Stamatakis 2014) implemented in raxmlGUI version 1.5b2

(Silvestros and Michalak 2012). The best-fit substitution model of GTR + I+G was selected by the Akaike Information Criteria (AIC) using jModelTest version 2.1.4 (Darrriba et al. 2012). Bootstrap support assessment was performed with 1000 replications.

3 Results

Genome content and structure of *S. sinomontana* – The chloroplast genome of *S. sinomontana* is a closed circular molecule of 147,240 bp with a typical quadripartite structure, including the LSC of 79,310 bp and SSC of 16,874 bp separated by a pair of IRs of 25,528 bp each (Fig. 1; Table 3). The GC-contents of the LSC, SSC, IR regions and the overall chloroplast genome of *S. sinomontana* are 36.2%, 32.0%, 42.9% and 38.0%, respectively (Table 3). The comparative analysis revealed that the chloroplast genome size of *S. sinomontana* is the smallest among all the seven Saxifragaceae chloroplast genomes, with 3826 bp shorter than its congeneric species, *S. stolonifera* (Table 3). Although the sizes of the overall genome, LSC, SSC and IR regions are different to some extent, the GC-contents of the complete chloroplast genomes and of each individual regions are similar among the seven Saxifragaceae species (Table 3).

The chloroplast genome of *S. sinomontana* contains 79 unique protein-coding genes, seven of which are duplicated (*ndhB*, *rpl2*, *rpl23*, *rps7* and *ycf2*) or partially duplicated (*rps19*, *ycf1*) in the IR. Four rRNA genes were identified in this genome, all of which are completely duplicated in the IR regions. In addition, 30 unique tRNA genes were annotated, among which, seven are duplicated. In total, the *S. sinomontana* chloroplast genome contains 131 genes (Table 1). Among them, 17 genes contain introns, including 11 protein-coding genes and six tRNA genes, almost all of which are single-intron genes except for *ycf3* and *clpP*. Like most other angiosperms (Nie et al. 2012; Hu et al. 2015; Wang et al. 2017; Yan et al. 2019), the *rps12* gene is trans-spliced, with its 5'-end exon located in the LSC region and 3'-end exon duplicated in the IRs. In general, both the content and order of genes present in *S. sinomontana* chloroplast genome are similar to those present in other Saxifragaceae species.

Simple sequence repeats analysis – Sixty-one SSRs, ranging from 10 to 17 bp in length, were detected in the *S. sinomontana* chloroplast genome (Table 2). The total length of the 61 SSRs found in *S. sinomontana* is 713 bp with an AT content of 95.1%, much higher than the rest of the genome. The mononucleotide repeat is the most abundant SSR type (41 out of 61) in *S. sinomontana* chloroplast genome, representing 67.2% of the total SSR loci. Among the 41 mononucleotide repeats, 40 are A or T types, whereas the number

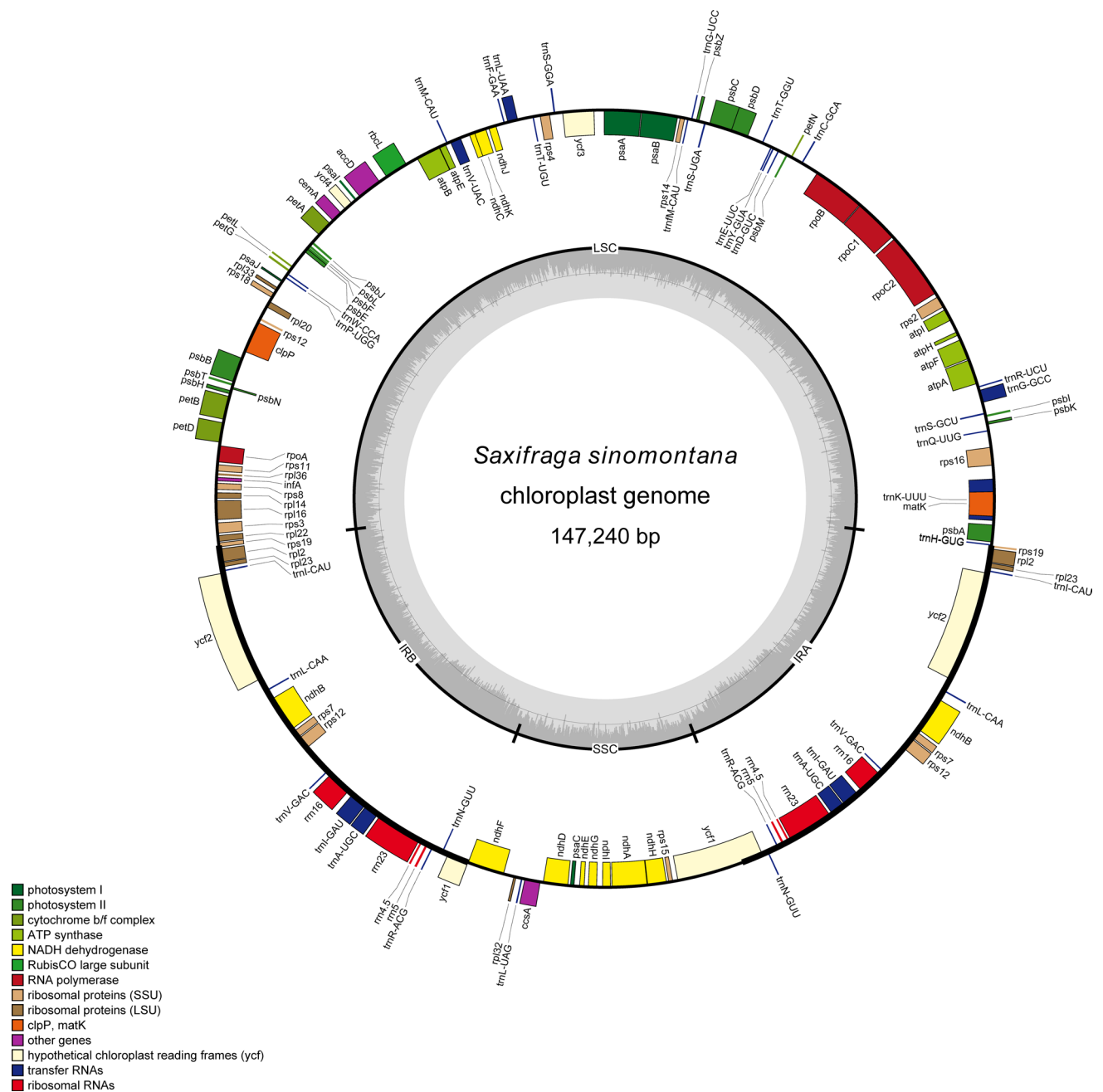


Fig. 1 Chloroplast genome map of *Saxifraga sinomontana*. Genes drawn inside of the outer circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are shown in different colors. The innermost darker gray corresponds to GC-content while the lighter gray corresponds to AT content

of di-, tri-, tetra-, and penta-nucleotide repeats detected in *S. sinomontana* chloroplast genome are 10, five, four and one, respectively. No hexanucleotide repeats were detected in this chloroplast genome. Among the 61 SSRs, 35 are located in the intergenic regions, 11 in introns, and 15 in protein-coding genes, including *ycf1*, *rpoB*, *atpB*, *ndhF*, *rpoC1*, *rps14*, *psbC*, *rpoC2*, *rpl22* genes (Table 2).

Chloroplast genome comparisons – The overall sequence identity of seven chloroplast genomes of Saxifragaceae species was plotted using mVISTA with the annotation of *S. sinomontana* chloroplast genome as reference (Fig. 2). The results showed that the Saxifragaceae chloroplast genomes exhibited a high level of sequence synteny, suggesting a conserved evolutionary pattern. In addition, the IR regions are less divergent compared to the LSC and SSC regions.

Table 1 Genes present in *Saxifraga sinomontana* chloroplast genome

No.	Gene types	Gene products
1	Photosystem I	psaA, B, C, I, J, ycf3 ^a
2	Photosystem II	psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
3	Cytochrome b6/f	petA, B ^b , D ^b , G, L, N
4	Atp synthase	atpA, B, E, F ^b , H, I
5	Rubisco	rbcL
6	NADH oxidoreductase	ndhA ^b , B ^{b,c} , C, D, E, F, G, H, I, J, K
7	Large subunit ribosomal proteins	rpl2 ^c , 14, 16 ^b , 20, 22, 23 ^c , 32, 33, 36
8	Small subunit ribosomal proteins	rps2, 3, 4, 7 ^c , 8, 11, 12 ^{b,c,d} , 14, 15, 16 ^b , 18, 19 ^c
9	RNAP	rpoA, B, C1 ^b , C2
10	Other proteins	accD, ccsA, cemaA, clpP ^a , infA, matK
11	Proteins of unknown function	ycf1 ^c , ycf2 ^c , ycf4
12	Ribosomal RNAs	rrn4.5 ^c , 5 ^c , 16 ^c , 23 ^c
13	Transfer RNAs	trnA(UGC) ^{b,c} , C(GCA), D(GUC), E(UUC), F(GAA), fM(CAU), G(GCC) ^b , G(UCC), H(GUG), I(CAU) ^c , I(GAU) ^{b,c} , K(UUU) ^b , L(CAA) ^c , L(UAA) ^b , L(UAG), M(CAU), N(GUU) ^c , P(UGG), Q(UUG), R(ACG) ^c , R(UCU), S(GCU), S(GGA), S(UGA), T(GGU), T(UGU), V(GAC) ^c , V(UAC) ^b , W(CCA), Y(GUA)

^aGene containing two introns^bGene containing a single intron^cTwo gene copies in IRs^dGene divided into two independent transcription units

The most divergent coding regions in the seven chloroplast genomes are *matK*, *ndhK*, *accD*, *cemaA*, *rpoA*, *rps19*, *ndhF*, *ccsA*, *ndhD*, *ycf1*. The most divergent regions among the seven chloroplast genomes are located in the intergenic spacers, including *trnK-UUU-rps16*, *rps16-trnQ-UUG*, *trnS-GCU-trnG-GCC*, *atpH-atpI*, *trnE-UUC-trnT-GGU*, *trnT-UGU-trnL-UAA*, *ndhC-trnV-UAC*, *petA-petJ*, *psbE-petL*, *ndhF-rpl32*, *rpl32-trnL-UAG*, *trnH-GUG-psbA* (Fig. 2).

In this study, the LSC/IRB/SSC/IRA boundaries and the adjacent genes were compared across the seven Saxifragaceae chloroplast genomes (Fig. 3). The *B. scopulosa* chloroplast genome showed dramatic variation in boundaries of IRB/SSC and SSC/IRA compared to the other chloroplast genomes, with its *ndhF* and duplicated *ycf1* genes completely located in the SSC region, resulting in the largest SSC of 21,920 bp but smallest IR of 23,811 bp among the seven Saxifragaceae chloroplast genomes (Table 3; Fig. 3). For the remaining six chloroplast genomes, expansion/contraction of IRs were also detected: the LSC/IRB border located within the coding region of *rps19* or *rpl22*, or within the intergenic spacer between *rps19* and *rpl2*; the IRB/SSC junction fell into the *ycf1* pseudogene and/or *ndhF* gene; the SSC/IRA border located within the *ycf1* gene, but with different extensions (Fig. 3). However, the IRA/LSC junction showed a complex pattern among the remaining six chloroplast genomes: (1) within the coding region of *trnH-GUG* (*S. sinomontana*); (2) within the intergenic spacer between

partially or completely duplicated *rps19* gene and *trnH-GUG* (*S. stolonifera* and *H. parviflora* var. *saurensis*); (3) within the intergenic spacer of *rpl2-trnH-GUG* without a duplication of *rps19* between them (*C. aureobracteatum*, *M. rossii* and *O. rupifraga*).

Phylogenetic analysis – A set of 66 protein-coding genes sharing among the seven Saxifragaceae chloroplast genomes were extracted, and phylogenetic relationships were reconstructed with *R. fasciculatum* var. *chinense* and *I. chinensis* as outgroups. The ML tree topology was consistent with previous studies, which confirmed *Saxifraga* sensu stricto as a monophyletic clade (Deng et al. 2015; Tkach et al. 2015). The ML bootstrap values based on the 66 shared genes were fairly high, and all but one node presented a bootstrap value of 100% (Fig. 4).

4 Discussion

The genome structure of *S. sinomontana* is consistent with that of most terrestrial plants (Palmer 1985), and the size (147,240 bp) falls well into the range of 120–160 kb of angiosperm chloroplast genomes (Palmer 1985). In general, gene content and genome organization of angiosperm chloroplast genomes are highly conserved compared to nuclear and mitochondrial genomes (Wicke et al. 2011). However, extensive gene losses and large inversions have

Table 2 Simple sequence repeat in *Saxifraga sinomontana* chloroplast genome

Unit	Length	SSR start	SSR end	location		
A	10	8313	8322	trnS-GCU-trnG-GCC		
		8779	8788	trnG-GCC intron		
		27,015	27,024	rpoB-trnC-GCA		
		51,951	51,960	atpB-rbcL		
		63,854	63,863	rps18-rpl20		
	11	10	71,934	71,943	petD intron	
			111,072	111,081	ndhD-psaC	
			4554	4564	matK-rps16	
			34,228	34,238	psbZ-trnG-UCC	
			76,514	76,524	rpl14-rpl16	
		12	11,421	11,432	atpA-atpF	
			15,696	15,707	rps2-rpoC2	
			44,897	44,908	trnT-UGU-trnL-UAA	
			66,120	66,131	clpP intron	
			118,540	118,553	ycf1	
	14	15	65,709	65,723	clpP intron	
			45,296	45,311	trnL-UAA intron	
			7765	7781	psbK-psbI	
			10	25,508	25,517	rpoB
				51,424	51,433	atpB
61,581		61,590		petL-petG		
73,215		73,224		petD-rpoA		
105,469		105,478		ndhF		
11		10	105,623	105,632	ndhF	
			4281	4291	matK-rps16	
	6231		6241	rps16-trnQ-UUG		
	12,792		12,802	atpF-atpH		
	20,271		20,281	rpoC1		
	12	27,141	27,151	rpoB-trnC-GCA		
		35,030	35,040	rps14		
		75,476	75,486	infA-rps8		
		77,866	77,876	rpl16-rps3		
		119,525	119,535	ycf1		
13	12	43,784	43,795	rps4-trnT-UGU		
		120,694	120,705	ycf1		
		15,667	15,679	rps2-rpoC2		
		40,151	40,163	psaA-ycf3		
		62,708	62,720	psaJ-rpl33		
14	15	22,255	22,268	rpoC1 intron		
		119,813	119,827	ycf1		
		11	32,461	32,471	psbC	
AT	19,242		19,251	rpoC2		
	30,503	30,514	trnT-GGU-psbD			
	107,171	107,184	ndhF-rpl32			

Table 2 (continued)

Unit	Length	SSR start	SSR end	location
TA	10	8246	8255	trnS-GCU-trnG-GCC
		44,365	44,374	trnT-UGU-trnL-UAA
		62,105	62,114	trnW-CCA-trnP-UGG
		63,293	63,302	rpl33-rps18
		12	105,611	105,622
TC	14	64,542	64,555	rpl20-rps12
		18,255	18,264	rpoC2
ATA	12	1488	1499	psbA-trnK-UUU
		41,487	41,501	ycf3 intron
TAT	12	45,252	45,263	trnL-UAA intron
		48,281	48,292	ndhC-trnV-UAC
ATG	12	76,082	76,093	rps8-rpl14
		41,127	41,138	ycf3 intron
TTTC	12	70,675	70,686	petB intron
		79,025	79,036	rpl22
ACAT	12	115,030	115,041	ndhA intron
		26,190	26,204	rpoB-trnC-GCA

been detected in several lineages, such as Gentianaceae (Fu et al. 2016; Sun et al. 2018), Asteraceae (Jansen and Palmer 1987; Liu et al. 2013; Walker et al. 2014), Leguminosae (Doyle et al. 1992, 1996). In the present study, no extensive gene losses and large inversions were detected among Saxifragaceae chloroplast genomes. In addition, comparing the complete chloroplast genome of *S. sinomontana* with that of other six Saxifragaceae species, we found that the size of *S. sinomontana* chloroplast genome is the smallest among the compared Saxifragaceae species, but the organization and gene content is highly similar.

Chloroplast SSRs usually exhibit high levels of variations and are widely used in polymorphism investigations, population genetics and phylogenetic analyses (Powell et al. 1995; Provan et al. 1999, 2001; Flannery et al. 2006; Xue et al. 2012; Li et al. 2019). For *S. sinomontana*, the number of cpSSRs (sixty-one) is moderate compared with other species of angiosperms (Liu et al. 2013; Fu et al. 2016; Yan et al. 2019). According to this and previous studies (Liu et al. 2013; Fu et al. 2016; Yan et al. 2019), extremely high AT contents of cpSSR sequences seem to be a common phenomenon in higher plant chloroplast genomes. In *S. sinomontana*, the mononucleotide repeat types are consistent with previous finding that short polyadenine or polythymine repeats are main contributors to SSRs in chloroplast genome (Kuang et al. 2011), and the distribution pattern of cpSSRs between coding and non-coding regions is consistent with most angiosperm species (Nie et al. 2012).

Although chloroplast genomes are considered to be rather conservative among angiosperm species, regions with high

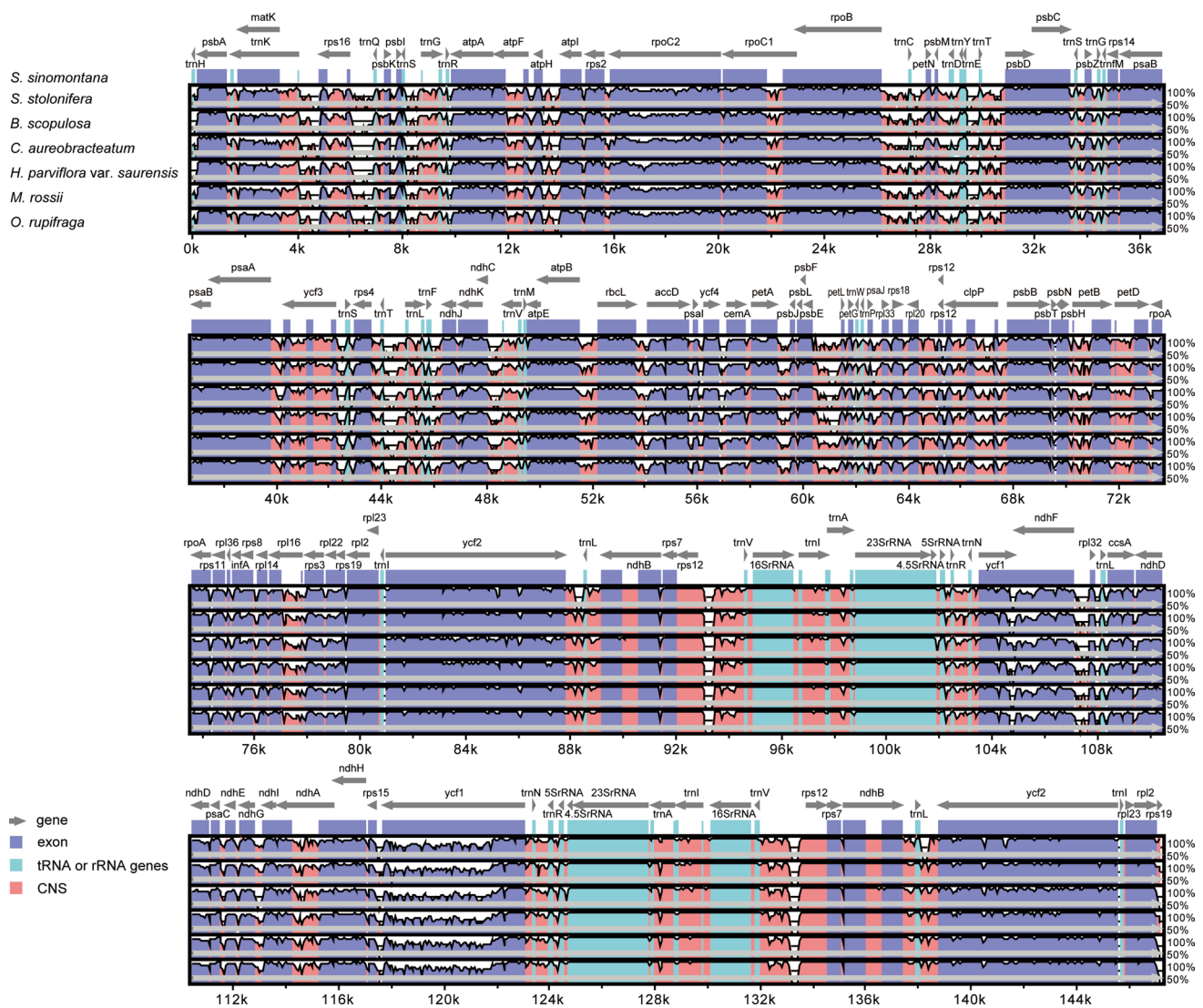


Fig. 2 Percentage of sequence identity between chloroplast genomes of *Saxifraga sinomontana* and other six Saxifragaceae species using mVISTA program. Gray arrows on the top line show transcriptional direction. The y-axis represents average percent identity between sequences of *S. sinomontana* and other six Saxifragaceae chloroplast genomes. The x-axis represents the coordinate in the chloroplast genome using *S. sinomontana* as reference. Genome regions are color coded as exon, tRNA, rRNA, and conserved non-coding sequences (CNS)

sequence polymorphisms are frequently observed even among closely related species (Kim and Jansen 1995). These highly divergent regions are widely used in plant phylogenetics, population genetics and DNA barcoding studies. Similar to other angiosperms (Nie et al. 2012; Liu et al. 2013; Chi et al. 2018; Yan et al. 2019), the coding regions of seven Saxifragaceae chloroplast genomes are more conservative than the non-coding regions. Among the ten most divergent coding regions in this study, *matK* has been considered as core universal DNA barcodes in many species (Li et al. 2019), and *ycf1* has recently been widely applied in plant phylogeny and DNA barcoding studies (Dong et al. 2015; Yang et al. 2017). Besides, *accD*, *rps19*, *ccsA* and *ndhF* have also been proved to exhibit large divergence among

various plant lineages and can be applied in phylogenetic studies (Ni et al. 2016; Ivanova et al. 2017). The boundaries of IRs and LSC/SSC differs among various plant species (Nie et al. 2012; Ni et al. 2016; Yan et al. 2019; Li et al. 2019), and expansion/extraction of the IR regions often lead to size variation of chloroplast genomes (Wang et al. 2008). Among the Saxifragaceae family, although various variations in the contraction/expansion of IRs have been detected, the IR sequences are not consistent with the total size of chloroplast genomes (Nie et al. 2012).

A number of phylogenetic studies have been conducted to clarify relationships within the family Saxifragaceae (Soltis et al. 2001, 2013; Xiang et al. 2012; DeChaine et al. 2013; Zhu et al. 2013; Deng et al. 2015), or within

Fig. 3 Comparison of the junction sites of LSC, IR and SSC regions among seven Saxifragaceae chloroplast genomes. Selected genes are indicated by boxes along the genome. Genes above the genome lines indicate their transcriptions in forward direction, while below in reverse direction. *JLB* junction between LSC and IRB, *JSB* junction between SSC and IRB, *JSA* junction between SSC and IRA, *JLA* junction between LSC and IRA

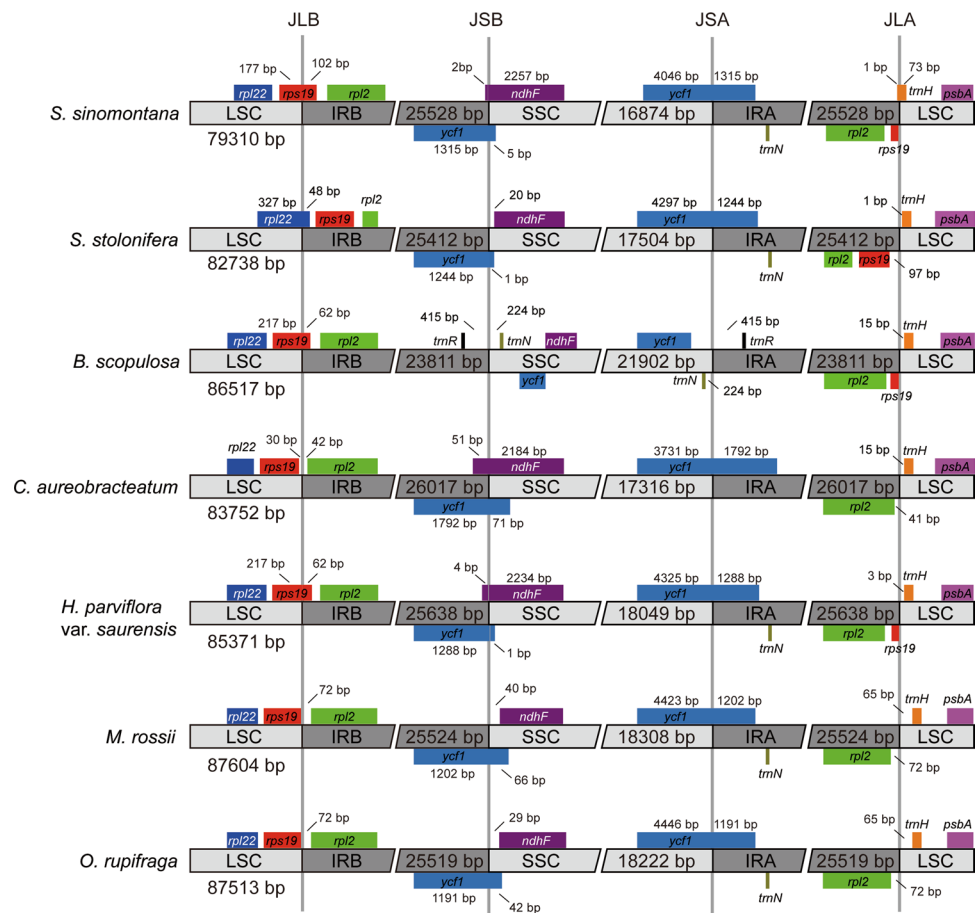


Table 3 Length and GC-content of the complete plastome, LSC, SSC and IR regions in seven Saxifragaceae species

Species	Complete genome		LSC		SSC		IR	
	Length (bp)	GC (%)	Length (bp)	GC (%)	Length (bp)	GC (%)	Length (bp)	GC (%)
<i>S. sinomontana</i>	147,240	38.0	79,310	36.2	16,874	32.0	25,528	42.9
<i>S. stolonifera</i>	151,066	37.8	82,738	35.9	17,504	32.2	25,412	43.0
<i>B. scopulosa</i>	156,041	37.8	86,517	35.8	21,902	33.0	23,811	43.5
<i>C. aureobracteatum</i>	153,102	37.3	83,752	35.2	17,316	31.1	26,017	42.8
<i>H. parviflora</i> var. <i>saurensis</i>	154,696	37.8	85,371	35.9	18,049	32.2	25,638	43.1
<i>M. rossii</i>	156,960	37.7	87,604	35.7	18,308	32.0	25,524	43.2
<i>O. rupifraga</i>	156,773	37.7	87,513	35.7	18,222	32.2	25,519	43.2

the genus *Saxifraga* (Zhang et al. 2008; Tkach et al. 2015; Gao et al. 2015). However, infrasectional relationships of the most species-rich section *Ciliatae* are still not well resolved (Zhang et al. 2008; Tkach et al. 2015; Gao et al. 2015), mostly due to the lack of resolutions within the recent divergent lineage of *Saxifraga* sect. *Ciliatae* subsect. *Hirculoideae*, in which informatively polymorphic sites are revealed to be limited (Gao et al. 2015). Chloroplast genomes comprise abundant phylogenetic information, which could have the potential to significantly advance our ability to resolve phylogenetic relationships

in this lineage. The result of phylogenetic analysis revealed that seven Saxifragaceae species based on 66 protein-coding genes received high bootstrap support values for nearly all identified nodes, suggesting a promising opportunity to resolve infrasectional relationships of the most species-rich section *Ciliatae*. This is the first chloroplast genome sequenced in *S.* sect. *Ciliatae* and also the second in the genus *Saxifraga*. More taxa, especially those of *S.* sect. *Ciliatae* subsect. *Hirculoideae* should be included in the chloroplast genome comparative analysis to realize the full

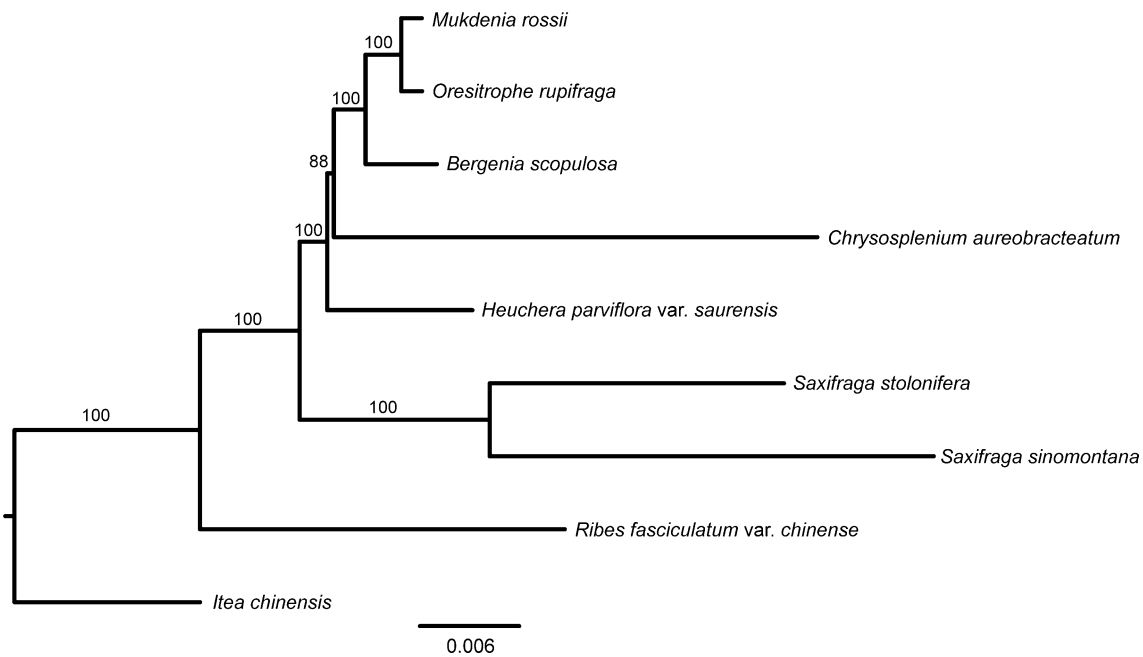


Fig. 4 The maximum-likelihood phylogenetic tree of seven Saxifragaceae species using 66 protein-coding genes of chloroplast genomes. Numbers above nodes are bootstrap support values

potential of chloroplast genomes in phylogenetic analysis of *S.* sect. *Ciliatae*.

Acknowledgements This study was supported by CAS Light of West China Program, Youth Innovation Promotion Association, CAS (Grant No. 2016378), Open Project of Qinghai Provincial Key Laboratory of Crop Molecular Breeding (Grant No. 2017-ZJY14), Application Fundamental Research Project of Qinghai Province (2019-ZJ-7019), High-end Innovative Talents Thousands of People Plan.

Author contributions Conceptualization: Qingbo Gao; Data Curation: Yan Li, Liukun Jia and Zhihua Wang; Formal Analysis: Rui Xing and Xiaofeng Chi; Resources: Shilong Chen; Writing—Original Draft Preparation: Yan Li; Writing—Review and Editing: Qingbo Gao; Funding Acquisition: Qingbo Gao and Shilong Chen.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Amiryousefi A, Hyvönen J, Poczar P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030–3031

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD et al (2012) SPAdes: a new genome assembly algorithm and its application to single-cell sequencing. *J Comput Biol* 19:455–477
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579
- Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, Green ED, Sidow A, Batzoglou S (2003) LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13:721–731
- Burke SV, Lin C-S, Wysocki WP, Clark LG, Duvall MR (2016) Phylogenomics and plastome evolution of tropical forest grasses (*Leptaspis*, *Streptochaeta*: Poaceae). *Front Plant Sci* 7:1993
- Chi XF, Wang JL, Gao QB, Zhang FQ, Chen SL (2018) The complete chloroplast genomes of two *Lancea* species with comparative analysis. *Molecules* 23:602
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models new heuristics and parallel computing. *Nat Methods* 9:772
- DeChaine EG, Anderson SA, McNew JM, Wendling BM (2013) On the evolutionary and biogeographic history of *Saxifraga* sect. *Trachyphyllum* (Gaud.) Koch (Saxifragaceae Juss.). *PLoS ONE* 8:e69814
- Deng JB, Drew BT, Mavrodiev EV, Gitzendanner MA, Soltis PS, Soltis DE (2015) Phylogeny divergence times and historical biogeography of the angiosperm family Saxifragaceae. *Mol Phylogenet Evol* 83:86–98
- Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL (2015) *ycf1* the most promising plastid DNA barcode of land plants. *Sci Rep* 5:8348
- Dong WP, Xu C, Wu P, Cheng T, Yu J, Zhou SL (2018) Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales. *Mol Phylogenet Evol* 126:321–330
- Doorduyn L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K (2011) The complete chloroplast genome of 17

- individuals of pest species *Jacobaea vulgaris*: SNPs microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res* 18:93–105
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull Bot Soc Am* 19:11–15
- Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ (1992) Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc Natl Acad Sci USA* 89:7722–7726
- Doyle JJ, Doyle JL, Ballenger JA, Palmer JD (1996) The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol Phylogenet Evol* 5:429–438
- Ebersbach J, Muellner-Riehl AN, Michalak I, Tkach N, Hoffmann MH, Röser M, Sun H, Favre A (2017a) In and out of the Qinghai-Tibet Plateau: divergence time estimation and historical biogeography of the large arctic-alpine genus *Saxifraga* L. *J Biogeogr* 44:900–910
- Ebersbach J, Schnitzler J, Favre A, Muellner-Riehl AN (2017b) Evolutionary radiations in the species-rich mountain genus *Saxifraga* L. *BMC Evol Biol* 17:119
- Ebersbach J, Muellner-Riehl AN, Favre A, Paule J, Winterfeld G, Schnitzler J (2018) Driving forces behind evolutionary radiations: *Saxifraga* section *Ciliatae* (Saxifragaceae) in the region of the Qinghai-Tibet Plateau. *Bot J Linn Soc* 186:304–320
- Flannery ML, Mitchell FJG, Coyne S, Kavanagh TA, Burke JI, Salamin N, Dowding P, Hodkinson TR (2006) Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. *Theor Appl Genet* 113:1221–1231
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279
- Fu PC, Zhang YZ, Geng HM, Chen SL (2016) The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ* 4:e2540
- Gao QB, Li YH, Gornall RJ, Zhang ZX, Zhang FQ, Xing R, Fu PC, Wang JL, Liu HR, Tian ZZ et al (2015) Phylogeny and speciation in *Saxifraga* sect. *Ciliatae* (Saxifragaceae): evidence from *psbA-trnH*, *trnL-F* and ITS sequences. *Taxon* 64:703–713
- Gao QB, Li Y, Gengji ZM, Gornall RJ, Wang JL, Liu HR, Jia LK, Chen SL (2017) Population genetic differentiation and taxonomy of three closely related species of *Saxifraga* (Saxifragaceae) from Southern Tibet and the Hengduan Mountains. *Front Plant Sci* 8:1325
- 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:W59–W64
- Hernández-León S, Gernandt DS, Pérez de la Rosa JA, Jardón-Barbolla L (2013) Phylogenetic relationships and species delimitation in *Pinus* section *Trifoliae* inferred from plastid DNA. *PLoS ONE* 8:e70501
- Hu S, Sablok G, Wang B, Dong Q, Barbaro E, Viola R, Li M, Varotto C (2015) Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genom* 16:306
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol* 14:151
- Ivanova Z, Sablok G, Daskalova E, Zahmanova G, Apostolova E, Yahubyan G, Baev V (2017) Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front Plant Sci* 8:204
- Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc Natl Acad Sci USA* 84:5818–5822
- Jansen RK, Cai Z, Raubeson LA, Daniel H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK et al (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104:19369–19374
- Kim KJ, Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. *Proc Natl Acad Sci USA* 92:10379–10383
- Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54:663–673
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Li Y, Gao QB, Gengji ZM, Jia LK, Wang ZH, Chen SL (2018) Rapid intraspecific diversification of the alpine species *Saxifraga sinomontana* (Saxifragaceae) in the Qinghai-Tibetan Plateau and Himalayas. *Front Genet* 9:381
- Li W, Zhang CP, Guo X, Liu QH, Wang KL (2019) Complete chloroplast genome of *Camellia japonica* genome structures comparative and phylogenetic analysis. *PLoS ONE* 14:e0216645
- Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, Feng X, Gu YQ (2013) Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS ONE* 8:e57533
- Ni LH, Zhao ZL, Xu HX, Chen SL, Dorje G (2016) The complete chloroplast genome of *Gentiana straminea* (Gentianaceae) an endemic species to the Sino-Himalayan subregion. *Gene* 577:281–288
- Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, Tan XF, Wan FH, Weining S (2012) Complete chloroplast genome sequence of a major invasive species crofton weed (*Ageratina adenophora*). *PLoS ONE* 7:e36869
- Palmer JD (1985) Comparative organization of chloroplast genomes. *Annu Rev Genet* 19:325–354
- Pan JT, Gornall RJ, Ohba H (2001) *Saxifraga*. In: Wu CY, Raven PH (eds) *Flora of China*. Science Press, Beijing; Missouri Botanical Garden Press, St Louis, pp 280–334
- Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci USA* 92:7759–7763
- Provan J, Russell JR, Booth A, Powell W (1999) Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Mol Ecol* 8:505–511
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147
- Silvestros D, Michalak I (2012) raxmlGUI: a graphical front-end for RAXML. *Org Divers Evol* 12:335–337
- Simmons MP (2017) Mutually exclusive phylogenomic inferences at the root of the angiosperms: *Amborella* is supported as sister and observed variability is biased. *Cladistics* 33:488–512
- Soltis DE, Kuzoff RK, Mort ME, Zanis M, Fishbein M, Hufford L, Koontz J, Arroyo MK (2001) Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. *Ann Mo Bot Gard* 88:669–693
- Soltis DE, Mort ME, Latvis M, Mavrodiev EV, O’Meara BC, Soltis PS, Burleigh JG, Rubio de Casas R (2013) Phylogenetic relationships and character evolution analysis of Saxifragales using a supermatrix approach. *Am J Bot* 100:916–929

- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Sun SS, Fu PC, Zhou XJ, Cheng YW, Zhang FQ, Chen SL, Gao QB (2018) The complete plastome sequences of seven species in *Gentiana* sect. *Kudoa* (Gentianaceae): insights into plastid gene loss and molecular evolution. *Front Plant Sci* 9:493
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). *Theor Appl Genet* 106:411–422
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11
- Tkach N, Röser M, Miehe G, Muellner-Riehl AN, Ebersbach J, Favre A, Hoffmann MH (2015) Molecular phylogenetics morphology and a revised classification of the complex genus *Saxifraga* (Saxifragaceae). *Taxon* 64:1159–1187
- Walker JF, Zanis MJ, Emery NC (2014) Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae Asteraceae). *Am J Bot* 101:722–729
- Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 8:36
- Wang WB, Yu H, Wang JH, Lei WJ, Gao JH, Qiu XP, Wang JS (2017) The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int J Mol Sci* 18:2288
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content gene order gene function. *Plant Mol Biol* 76:273–297
- Xiang CL, Gitzendanner MA, Soltis DE, Peng H, Lei LG (2012) Phylogenetic placement of the enigmatic and critically endangered genus *Saniculiphyllum* (Saxifragaceae) inferred from combined analysis of plastid and nuclear DNA sequences. *Mol Phylogenet Evol* 64:357–367
- Xue J, Wang S, Zhou SL (2012) Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am J Bot* 99:e240–e244
- Yan C, Du JC, Gao L, Li Y, Hou XL (2019) The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R Br): genome organization adaptive evolution and phylogenetic relationships in Cardamineae. *Gene* 699:24–36
- Yang J, Vázquez L, Chen XD, Li HM, Zhang H, Liu ZL, Zhao GF (2017) Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes. *Front Plant Sci* 8:816
- Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134
- Zhang DJ, Chen SY, Gao QB, Duan YZ, Chen SL (2008) Circumscription and phylogeny of *Saxifraga* sect. *Ciliatae*: evidence from nrDNA ITS sequences. *J Syst Evol* 46:667–675
- Zhu WD, Nie ZL, Wen J, Sun H (2013) Molecular phylogeny and biogeography of *Astilbe* (Saxifragaceae) in Asia and eastern North America. *Bot J Linn Soc* 171:377–394

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.