

RESEARCH ARTICLE

Open Access



Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae)

Xu Zhang^{1,2,3†}, Tao Deng^{3†}, Michael J. Moore⁴, Yunheng Ji³, Nan Lin^{1,2,3}, Huajie Zhang^{1,2}, Aiping Meng¹, Hengchang Wang^{1*}, Yanxia Sun^{1*} and Hang Sun^{3*}

Abstract

Background: *Saussurea* DC. is one of the largest and most morphologically heterogeneous genera in Asteraceae. The relationships within *Saussurea* have been poorly resolved, probably due an early, rapid radiation. To examine plastome evolution and resolve backbone relationships within *Saussurea*, we sequenced the complete plastomes of 17 species representing all four subgenera.

Results: All *Saussurea* plastomes shared the gene content and structure of most Asteraceae plastomes. Molecular evolutionary analysis showed most of the plastid protein-coding genes have been under purifying selection. Phylogenomic analyses of 20 *Saussurea* plastomes that alternatively included nucleotide or amino acid sequences of all protein-coding genes, vs. the nucleotide sequence of the entire plastome, supported the monophyly of *Saussurea* and identified three clades within it. Three of the four traditional subgenera were recovered as paraphyletic. Seven plastome regions were identified as containing the highest nucleotide variability.

Conclusions: Our analyses reveal both the structural conservatism and power of the plastome for resolving relationships in congeneric taxa. It is very likely that differences in topology among data sets is due primarily to differences in numbers of parsimony-informative characters. Our study demonstrates that the current taxonomy of *Saussurea* is likely based at least partly on convergent morphological character states. Greater taxon sampling will be necessary to explore character evolution and biogeography in the genus. Our results here provide helpful insight into which loci will provide the most phylogenetic signal in *Saussurea* and Cardueae.

Keywords: *Saussurea*, Rapid radiations, Plastome, Genome, Phylogenomic analysis, Purifying selection

Background

Saussurea DC. is one of the largest genera in the family Asteraceae [1, 2]. It comprises approximately 300 species that are distributed in Asia, Europe and North America, with the highest diversity in the Himalayas and central Asia [3, 4]. *Saussurea* exhibits extreme morphological diversity and exists in habitats ranging from steppes to moist forests to cold and dry alpine meadows above 5000 m [3, 5].

Several phylogenetic studies have been conducted on *Saussurea* but the circumscription and infrageneric relationships of the genus remain controversial [5–12]. Lipschitz [13] recognized a total of 390 species belonging to six subgenera, namely subg. *Saussurea* DC., *Juriniocera* (Baill.) Lipsch., *Eriocoryne* (DC.) Hook. f., *Amphilaena* (Stschegl.) Lipsch., *Theodorea* (Cass.) Lipsch. and *Frolovia* (DC.) Lipsch. However, molecular evidence [7, 10, 14] has indicated that subg. *Juriniocera*, subg. *Frolovia* and sect. *Elatae* of subg. *Saussurea* should be excluded from *Saussurea* and treated as independent genera: *Lipschitzia* R.V. Kamelin, *Frolovia* (DC.) Lipsch. and *Himalaiella* Raab-Straube, respectively. Using molecular and morphological evidence, Shi and Raab-Straube [3] suggested *Saussurea* sect. *Aucklandia* should be treated as a new genus, *Aucklandia* Falc. Based on sequences of five loci (*rbcL*, *ndhF*, *matK*, *trnL-F* and ITS) and morphology, Wang et al. [11] established

* Correspondence: hcwang@wbpcas.cn; sunyanxia@wbpcas.cn; sunhang@mail.kib.ac.cn

†Xu Zhang and Tao Deng contributed equally to this work.

¹CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, China

³Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China

Full list of author information is available at the end of the article



the genus *Shangwua* Y. J. Wang, Raab-Straube, Susanna & J. Q. Liu from sect. *Jacea*, leaving four subgenera (*Saussurea*, *Eriocoryne*, *Amphilaena* and *Theodorea*) as constituting *Saussurea* s.s. Despite this progress, the relationships among and within these four subgenera have been poorly resolved due to a potentially rapid radiation, leaving insufficient phylogenetic signal at deeper nodes [5]. No phylogenomic studies have yet assessed these relationships, although a recent study using target enrichment of nuclear genes to resolve Cardueae relationships sampled 19 species of *Saussurea* representing two subgenera [15].

Plastomes have been proven to be powerful tools for exploring deep relationships in the plant Tree of Life [16–19]. They have helped resolve ambiguous relationships of particularly recalcitrant lineages, such as those that have undergone rapid evolutionary radiations (e.g. [20–23]). Complete plastome sequences also provide insight into the molecular evolutionary patterns associated with gene rearrangements, duplication and loss (e.g. [24–26]), and in some cases these structural changes are phylogenetically informative characters in and of themselves, as for example the two large inversions (~ 20 kb and ~ 3 kb inversions) that characterize the Large Single-Copy (LSC) region of most Asteraceae plastomes [27–29].

Different regions of the plastome have different selective constraints that may yield differing estimates of phylogeny, as for example noncoding versus coding regions [30, 31]. Selective forces may also play a role in driving plastome structure [32], including rearrangements [33] and gene loss [34, 35]. However, the effect of selective forces in plastome evolution within Asteraceae remains unclear.

To date, only three *Saussurea* plastomes have been reported: *S. involucrata* [36], *S. chabyoungsanica* [37] and *S. polylepis* [38]. Here, we sequenced 17 species representing all four subgenera of *Saussurea* in order to (1) elucidate plastome evolution, including structural variation and molecular signals of selection, (2) estimate the effectiveness of different plastome data sets in resolving relationships within this radiating lineage, and (3) investigate the backbone relationships within *Saussurea*.

Results

Characteristics of *Saussurea* plastomes

After de novo and reference-guided assembly, we obtained a single scaffold for each plastome. The sequencing and assembly information are provided in Tables 1 and Additional file 1: Table S2. The sizes of the 17 *Saussurea* plastomes were similar, ranging from 151,474 bp in *S. tridactyla* to 152,658 bp in *S. przewalskii*. All 17

Table 1 Characteristics of newly sequenced plastomes

Species	Plastome size (bp)	LSC length (bp)	IR length (bp)	SSC length (bp)	Gene Number	Protein coding genes	RNAs	GC content (%)
<i>Saussurea hookeri</i> C.B.Clarke	152,461	83,437	25,201	18,622	114	80	34	37.7
<i>Saussurea obvallata</i> (DC.) Sch.Bip.	152,544	83,460	25,196	18,692	114	80	34	37.7
<i>Saussurea pubifolia</i> S.W.Liu	152,622	83,541	25,227	18,627	114	80	34	37.7
<i>Saussurea</i> sp. nov	152,055	83,543	24,660	19,192	114	80	34	37.7
<i>Saussurea psudoleucoma</i> Y. S. Chen	152,412	83,450	25,183	18,596	114	80	34	37.7
<i>Saussurea lhozhangensis</i> Y.S.Chen	152,527	83,525	25,194	18,614	114	80	34	37.7
<i>Saussurea gossypiphora</i> D.Don	152,463	83,451	25,199	18,614	114	80	34	37.7
<i>Saussurea tridactyla</i> Sch.Bip. ex Hook.f.	151,474	82,564	25,193	18,524	114	80	34	37.7
<i>Saussurea gnaphalodes</i> (Royle) Sch. Bip.	152,475	83,379	25,201	18,694	114	80	34	37.7
<i>Saussurea salwinensis</i> J.Anthony	152,382	83,449	25,194	18,545	114	80	34	37.7
<i>Saussurea przewalskii</i> Maxim.	152,658	83,533	25,221	18,683	114	80	34	37.7
<i>Saussurea delavayi</i> Franch.	152,254	83,213	25,196	18,649	114	80	34	37.7
<i>Saussurea leontodontoides</i> (DC.) Sch.Bip.	152,387	83,433	25,201	18,552	114	80	34	37.7
<i>Saussurea durgae</i> C. Jeffrey & R.C.Srivast.	152,506	83,411	25,195	18,705	114	80	34	37.7
<i>Saussurea kingii</i> C.E.C.Fisch.	152,444	83,398	25,178	18,690	114	80	34	37.6
<i>Saussurea japonica</i> (Thunb.) DC.	152,612	83,502	25,140	18,830	114	80	34	37.7
<i>Saussurea tsoongii</i> Y.S.Chen	152,501	83,373	25,193	18,742	114	80	34	37.7

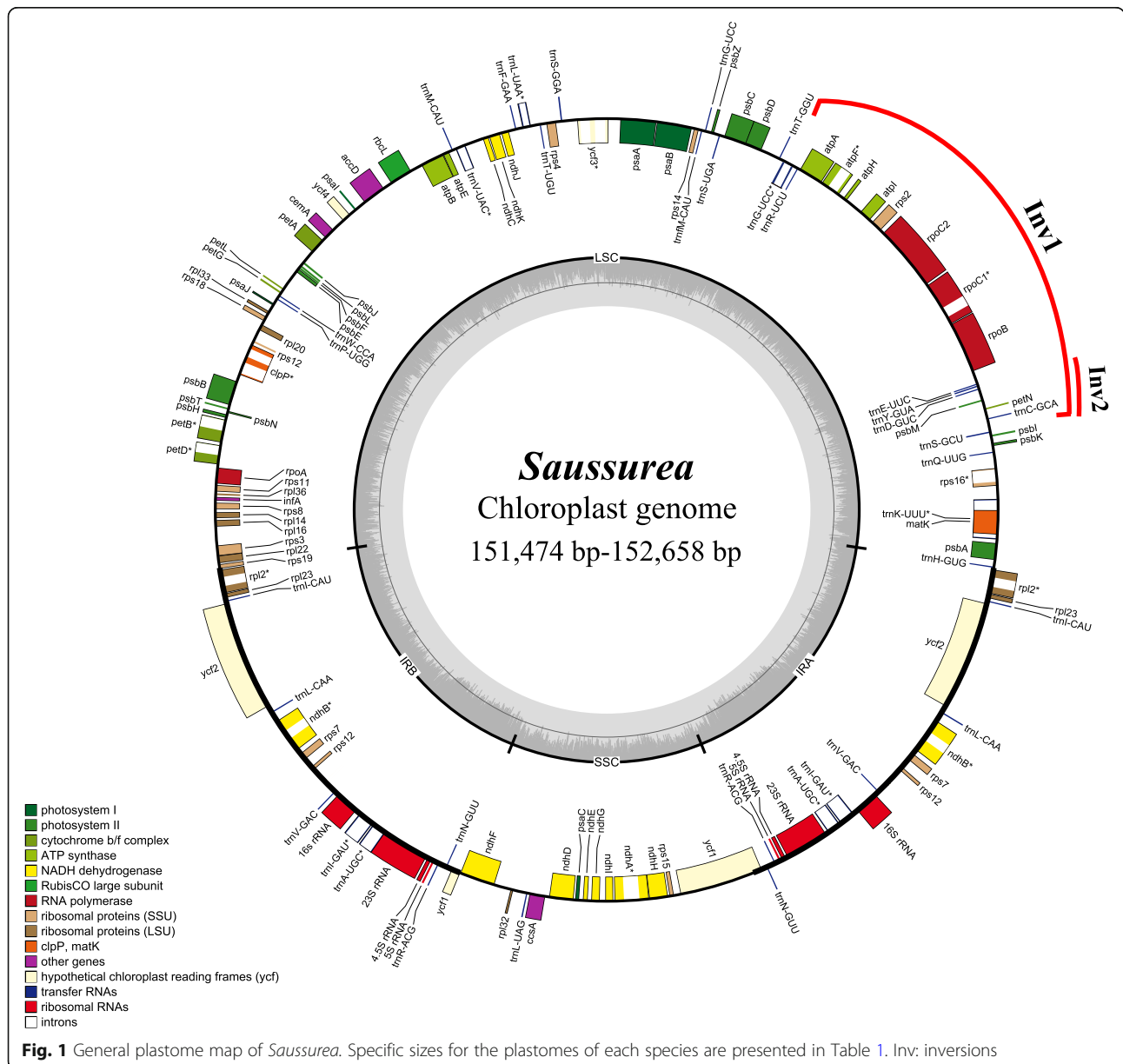
plastomes possessed the typical angiosperm quadripartite structure and contained 113 unique genes, including 79 protein-coding genes, 30 transfer RNA (tRNA) genes and four ribosomal RNA (rRNA) gene. A total of 18 genes (including 11 protein-coding genes and 7 tRNA genes) had introns, with 15 genes having one intron and three genes having two introns. The IR regions were also highly consistent, all of which included 17 genes (six protein-coding genes, seven tRNA genes, and four rRNA genes). In all plastomes, the *rps12* gene was found to be trans-spliced, with one of its exons located in the LSC region and the other duplicated in the IR (Fig. 1).

The ~ 20 kb and ~ 3 kb inversions (Inv1 and Inv2) of most Asteraceae were detected in all *Saussurea*

plastomes (Fig. 1). Inv1 was located between the *trnG-UCC* and *trnS-GCU* genes; Inv2 (located between the *trnS-GCU* and *trnE-UUC* genes) was nested within the large inversion and shared one end-point with Inv1 (Fig. 1). Sliding window analysis showed much higher proportions of variable sites in single-copy regions than in the IR regions. Seven relatively highly variable regions (*rps16-trnQ*, *trnS-trnC-petN*, *psbE-petL*, *ndhF-rpl32*, *rpl32-trnL*, *rps15* and *ycf1*) were identified from the plastome sequences (Fig. 3).

Selection analyses

Most protein coding genes showed a low dN/dS ratio (ω ; Additional file 2: Figure S1), indicating that they have



been under purifying selection. Only three genes (*psbL*, *psbZ* and *ycf2*) had $\omega > 1$, but the branch model results revealed no significant difference between foreground and background branches (Table 2).

Phylogenetic analyses

Our phylogenomic analyses substantially increased resolution and provided robust backbone relationships of *Saussurea* (Fig. 4, Additional file 3: Figure S2). Characteristics of the three concatenated data sets are presented in Table 3. Dataset-3 had the highest number of parsimony-informative (PI) characters, followed by dataset-1 and dataset-2. Centaureinae were resolved as sister to *Saussurea* in datasets-1 and -3 with strong support, but not in dataset-2. All three datasets also strongly supported the monophyly of *Saussurea* (BS = 100), while three (*Eriocoryne*, *Amphilaena*, and *Saussurea*) of the four traditional subgenera were resolved as paraphyletic. Three main clades of *Saussurea* were identified. Clade 1 included three species of subg. *Amphilaena* (*S. pubifolia*, *S. sp. nov.*, *S. involucrata*), one of subg. *Eriocoryne* (*S. lhozhagensis*) and five of subg. *Saussurea* (*S. durgae*, *S. przewalskii*, *S. salwinensis*, *S. delavayi*, *S. kingii*). Clade 2 included two species of subg. *Amphilaena* (*S. hookeri*, *S. obvalata*) and four of subg. *Eriocoryne* (*S. gnaphalodes*, *S. gossypiphora*, *S. psedoleucoma*, and *S. tridactyla*). Clade 3 included two species of subg. *Theodorea* (*S. japonica* and *S. tsoongii*) and two Korean species (*S. chabyoungsanica*, *S. polylepis*). Datasets-1 and -2 resolved subg. *Theodorea* as sister to remaining *Saussurea*, whereas dataset-3 resolved clade 2 as sister to remaining *Saussurea*, albeit with low support (Fig. 4, Additional file 3: Figure S2). The coalescent-based result yielded an almost identical topology with the concatenation-based phylogeny (dataset-1), except for the position of *S. kingii*, which was resolved as sister to clade 1 + clade 2 (Additional file 4: Figure S3).

Discussion

Plastome evolution

The 20 *Saussurea* plastomes in our analyses indicated that plastome evolution has been conservative within this genus. All *Saussurea* plastomes possessed the typical

plastome structure of most Asteraceae, including both LSC inversions that are present in nearly all Asteraceae, as for example in *Lactuca* [27], *Artemisia* [29], *Lasthenia* [28], *Taraxacum* [39] and *Mikania* [40]. The expansion and contraction of the IR region has been demonstrated to be a significant source of length variation in some plastomes, e.g. early-diverging eudicots [41, 42] and Apiales [43]. In the present study, however, no significant IR length variation was detected among *Saussurea* plastomes (Fig. 2).

In our molecular evolutionary analysis, most protein-coding genes were found to be under purifying selection (Additional file 2: Figure S1). This pattern has also been demonstrated in other Asteraceae plastomes, such as in *Mikania cordata* [40] and *Helianthus* [44], reflecting the typically conservative evolution of plastome genes in green plants. Indeed, the best evidence for relaxation of purifying selection is in plants that have lost photosynthesis, in which genes involved directly in photosynthesis evolve much faster due to loss of function, typically resulting in pseudogenization and eventual gene loss [32, 34, 35]. Nevertheless, complete genome- and transcriptome-based analyses are necessary to fully investigate the importance of selection at protein-coding loci in plastids, given that most plastid proteins are encoded in the nucleus.

Phylogenetically informative sites

To resolve relationships among closely related species, it is imperative to identify rapidly evolving loci. Previous phylogenetic studies of *Saussurea* mainly favored three plastid loci (*trnL-F*, *psbA-trnH*, and *matK*) but these have failed to resolve relationships across the genus (e.g. [5–8]). Our analyses revealed relatively low nucleotide diversity in these three regions (Fig. 3), explaining the low resolution in previous analyses and highlighting the importance of exploring more of the plastome to obtain additional informative sites and regions. We found seven relatively variable regions: *rps16-trnQ*, *trnS-trnC-petN*, *psbE-petL*, *ndhF-rpl32*, *rpl32-trnL*, *rps15* and *ycf1*. Of these, *rps16-trnQ*, *trnC-petN*, *psbE-petL*, *rpl32-trnL*, *rps15* and *ycf1* have been previously reported as hotspots of divergence and have been broadly used for

Table 2 Branch model results of three genes with $\omega > 1$. $\omega = dN/dS$, the ratio of nonsynonymous/synonymous substitution rates. np: number of parameters. ln L: log likelihood values. LRT: likelihood ratio test

Genes	Model	np	ln L	Estimates of parameters		LRT P value
<i>psbL</i>	Two ratio Model 2	67	-177.133731	$\omega_0 = 2.64875$	$\omega_1 = 1.71333$	0.9984
	Model 0	66	-177.133729	$\omega_0 = 2.64876$		
<i>ycf2</i>	Two ratio Model 2	67	-10,731.616197	$\omega_0 = 1.39172$	$\omega_1 = 1.98555$	0.7353
	Model 0	66	-10,731.67334	$\omega_0 = 1.40184$		
<i>psbZ</i>	Two ratio Model 2	67	-1424.307323	$\omega_0 = 2.73007$	$\omega_1 = 2.21091$	0.9972
	Model 0	66	-1424.307317	$\omega_0 = 2.73007$		

Table 3 Characteristics of the three different data sets

Dataset	Composition	Total number of characters (bp)	Variable sites (bp)	Parsimony-informative (PI) sites (bp)	Singleton sites (bp)	Proportion of PI sites (%)
(1)	Nucleotide sequences of all 79 protein-coding regions (CDS)	69,474	4071	2285	1782	3.29
(2)	Amino acid sequences of all 79 CDS (AA)	23,158	1835	1179	653	5.09
(3)	Complete plastome nucleotide sequences	134,955	10,622	5490	5043	4.07

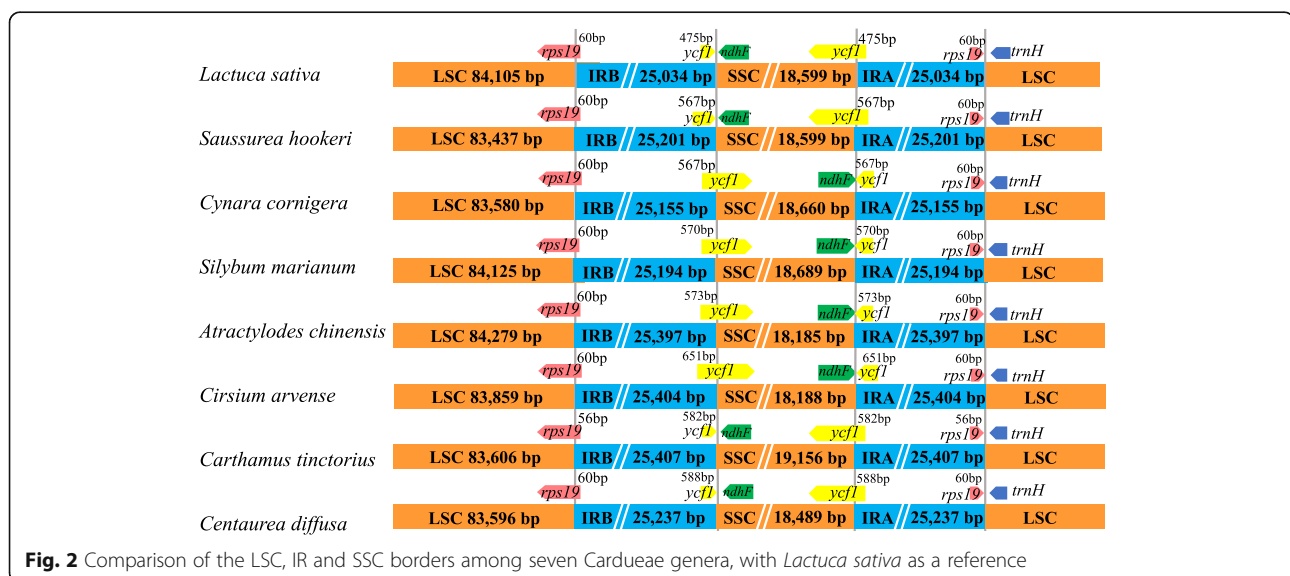
reconstructing phylogeny in plant taxa [40, 45–50]. The lineage-specific, rapidly evolving regions identified here will facilitate further phylogenetic resolution of the large and diverse *Saussurea*.

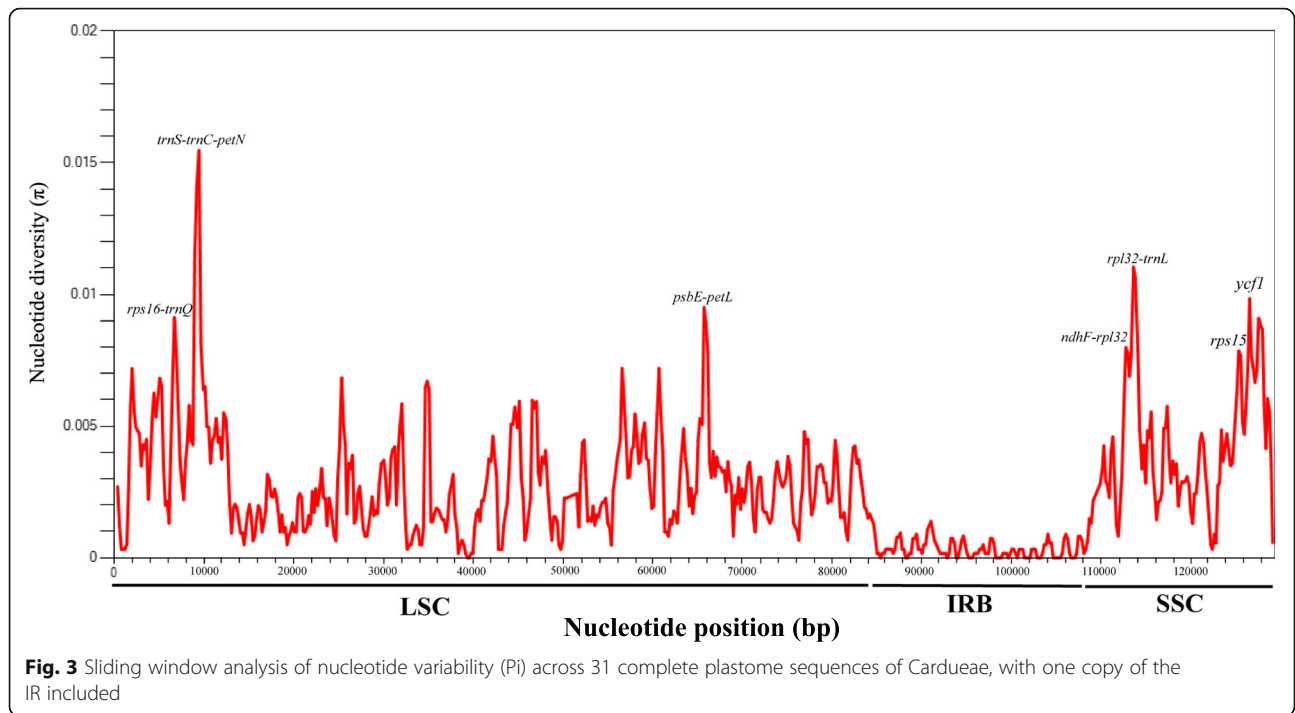
Phylogenetic relationships within *Saussurea*

The backbone relationships of *Saussurea* have been poorly resolved in previous molecular phylogenetic studies (e.g., [5–8, 10–12]). Our analyses greatly increased resolution with generally robust support (Fig. 4, Additional file 3: Figure S2). With the exception of subg. *Theodora* (the only monophyletic subgenus), there is relatively little concordance between the relationships recovered here and morphological characters used to define sections and subgenera [3, 4, 13]. In fact, these morphological characters have been shown to have adaptive value, as for example the dense woolly trichomes and colorful bracts that are used to circumscribe subg. *Eriocoryne* and subg. *Amphilaena* respectively. These two kinds of character states are prevalent among alpine species, and have been thought to protect plants from cold and UV-B radiation at high elevations [5, 51–53]. Hence, the discordance between phylogeny and morphology may reflect potential convergent evolution in *Saussurea*. It is also important to note that our estimate of phylogeny is based only on the plastome in a rapidly

radiating group. Given that incomplete lineage sorting (ILS) or hybridization are most likely to obscure the species phylogeny among close relatives, it is possible that the addition of nuclear phylogenomic data may result in a different estimate of relationships in *Saussurea*. Consequently, it is essential to expand taxon and locus sampling significantly within *Saussurea* to better understand patterns of character state evolution and biogeography.

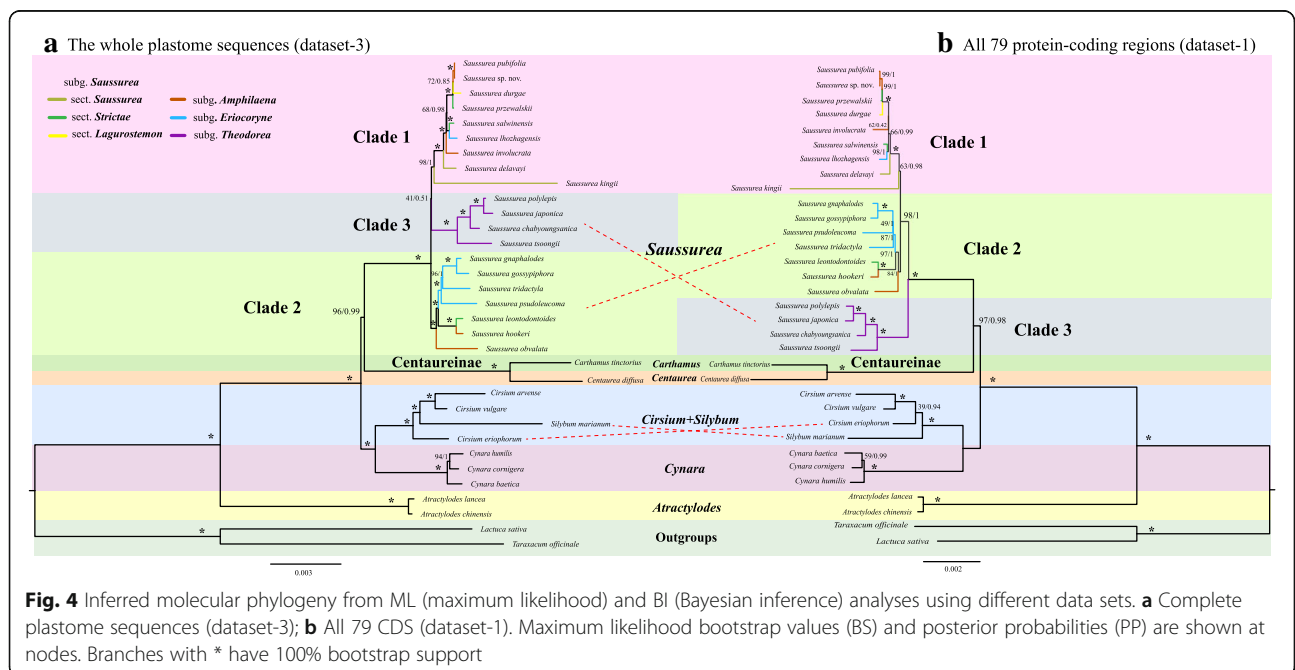
The clades formed by subg. *Theodorea* and sect. *Laguranthera* (*S. durgae*) were resolved as early-diverging groups in phylogenetic studies of *Saussurea* based on ITS and *trnL-trnF* [7]. In our concatenated datasets-1 and -2 and coalescent-based approach, the early-diverging position of subg. *Theodorea* was also supported, despite it being relatively distant phylogenetically from sect. *Laguranthera*. Across all concatenated datasets, *S. kingii* had the longest branch by far (Fig. 4, Additional file 3: Figure S2), which was also detected in the phylogenetic study of Wang and Liu [12]. As suggested there, this likely results from its biennial habit, as substitution rates are known to be higher in species with shorter generation times [54]. In addition, the systematic position of *S. kingii* was unstable between concatenated- and coalescent-based approaches, suggesting a further investigation may be required.





Incongruence at deeper levels among the trees resulting from our three concatenation-based analyses is likely related to differences in the number of parsimony-informative (PI) characters among data sets, with the highest number of PI characters in dataset-3 (Table 3). These differences likely explain the better overall support for the backbone of *Saussurea* in the tree based on dataset-3 (Fig. 4a) compared to the other trees.

Given the relatively low taxonomic level (within a genus) of our study, it makes sense that including nucleotide sequence, especially for noncoding regions, would maximize the power to resolve relationships. We therefore recommend complete plastome data sets in these situations. The incongruence at a few backbone nodes is not surprising given how short these branches are; it is likely that few PI characters ever existed at these



branches, and hence such nodes are sensitive to the conditions of phylogenetic analysis [23, 55].

Conclusions

Our analyses reveal both the structural conservatism and power of the plastome for resolving relationships in congeneric taxa. By examining signals of selection at protein-coding loci, we are able to eliminate systematic error due to selective biases as a source of topological incongruence. Hence, it is very likely that differences in topology among data sets are due primarily to differences in numbers of parsimony-informative characters. Our study further demonstrates that currently accepted subgeneric groups in *Saussurea* are likely based at least partly on convergent character states, and are therefore in need of revision. Moreover, greater taxon sampling is necessary to disentangle the patterns of character evolution and biogeography that are only hinted at here. Our results here provide helpful insight into which loci will provide the most PI sites in *Saussurea* and Cardueae, but they also suggest that complete plastome sequencing will be a valuable technique for resolving the relationships in this difficult genus.

Methods

Taxon sampling, chloroplast DNA isolation, high-throughput sequencing

We sequenced 17 new plastomes representing 16 currently described and one undescribed species of *Saussurea*; collection and voucher information are provided in Additional file 1: Table S1. These were added to the three previously reported plastomes available in GenBank (Additional file 1: Table S1). The circumscription and infrageneric treatment of *Saussurea* followed *Flora of China* and *Flora of Pan-Himalaya* [3, 4]. For all species, total DNA was extracted from fresh or silica gel-dried leaves with a modified CTAB (Cetyl trimethylammonium bromide) method [56]. Sequencing libraries were constructed and quantified following the methods introduced by Sun et al. [41]. For all plastomes, a 500-bp DNA TruSeq Illumina (Illumina Inc., San Diego, CA, USA) sequencing library was constructed using 2.5–5.0 ng sonicated DNA as input. Libraries were quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and by real-time quantitative PCR. Libraries were then multiplexed and sequenced using a 2 × 125 bp run on an Illumina HiSeq 2000 platform at Novogene in Kunming, Yunnan, China.

Plastome assembly, annotation, and comparative analyses

Raw sequence reads were subsequently filtered using Trimmomatic v.0.36 [57] with the following parameters: SLIDING WINDOW = 4:20, MINLEN = 50, LEADING =

3, TRAILING = 3, HEAD-CROP = 12, and AVGQUAL = 20. Remaining high-quality reads were assembled de novo into contigs with a minimum length of 1000 bp using CLC Genomics Workbench 11.0 (<https://www.qiagenbioinformatics.com/>) with default parameters. The resulting de novo contigs were then reference-assembled against the plastome of *S. chabyoungsanica*. Finished plastomes were annotated using DOGMA [58] and GeSeq [59]. Manual adjustments of start/stop codons and intron/exon boundaries were conducted in Geneious version 9.0.5 [60], using published plastomes of *Saussurea* as references. The tRNA genes were identified with tRNAscan-SE [61]. Physical maps of the circular plastomes were visualized with OGDRAW [62].

We performed plastome comparisons between *Saussurea polylepis* and six other Cardueae genera (*Cirsium arvense*, *Carthamus tinctorius*, *Cynara cornigera*, *Centaurea diffusa*, *Silybum marianum*, *Atractylodes chinensis*). All seven complete plastomes were aligned with ProgressiveMAUVE [63], assuming collinear genomes for the full alignment. To assess sequence divergence and determine highly phylogenetically informative sites, nucleotide variability (π) was calculated by sliding window analysis conducted in DnaSP version 6.11.01 [64] with all aligned plastome sequences of *Saussurea*. For the purposes of alignment, the SSC region was inverted manually in Geneious as necessary. The step size was set to 200 bp, with a 600 bp window length.

Phylogenetic analyses

Thirty-one taxa (Additional file 1: Table S1) of Cardueae (20 *Saussurea* + 11 outgroup genera from Cardueae) and two outgroup taxa of Cichorieae (*Lactuca sativa*, *Taraxacum officinale*) were included in phylogenetic analyses. Both concatenated and coalescent-based analyses were conducted. For concatenation-based approach, three datasets were analyzed: dataset-1 included the nucleotide sequences of all 79 protein-coding sequences (CDS); dataset-2 included the amino acid sequences of these 79 CDS; and dataset-3 included the complete plastome nucleotide sequences, including only one copy of the IR regions. Dataset-1 and -2 were created by concatenating alignments using PhyloSuite version 1.1.15 [65]. Characteristics of all three data sets were calculated using MEGA X [66]. For all concatenated data sets, Modeltest version 3.7 [67] was used to estimate the optimal model under the Akaike Information Criterion (AIC). Maximum likelihood (ML) analyses were conducted using RAxML version 8.2.10 [68] under the general time reversible model of nucleotide substitution, with the gamma model of rate heterogeneity (GTRGAMMA for dataset-1 and dataset-3; PROTGAMMAAUTO for dataset-2). Bootstrap (BS) support was estimated with 1000 bootstrap replicates using the “rapid bootstrap”

algorithm of RAxML. Bayesian inference (BI) was performed using MrBayes version 3.2.3 [69]. Two runs were conducted in parallel with four Markov chains (one cold and three heated), with each running for 5000,000 generations from a random starting tree and sampled every 5000 generations. Convergence was assessed by examining the average standard deviation of split frequencies (ASDF). After ASDF reached < 0.01, the first 25% of the trees were discarded as burn-in, and the remaining trees were used to construct majority-rule consensus trees.

For the coalescent-based analysis, ML unrooted trees for 79 CDS alignments were estimated separately using RAxML under the GTRGAMMA model with 500 bootstrap replicates. ASTRAL III version 5.6.2 algorithm [70] was used to estimate the species tree from 79 gene trees with node supports calculated as local posterior probabilities.

Analyses of signatures of selection

To test for evidence of selection in plastid protein coding genes, we estimated the ratio of nonsynonymous (dN) to synonymous (dS) substitutions (ω) for all 79 protein coding genes using CodeML in PAML version 4.9 [71] with the following settings: model = 0, seqtype = 1, NSsites = 0. Genes showing higher ω were identified with the branch model [72, 73] to determine lineage-specific selection in plastomes of *Saussurea*. Following the recommendations in CodeML, the best ML tree determined by RAxML with dataset-1 using concatenation-based approach was used as the input topology, and the clade formed by *Saussurea* was set as a foreground branch. The likelihood ratio and *P* value were used to test if a model (“model = 2”) of positive selection on the foreground branch was a significant improvement over a null model (“model = 0”) where no positive selection occurred on the foreground branch.

Additional files

Additional file 1: Table S1. Taxa included in present study. Collection locality and voucher information are provided for newly sequenced

Table S2. The sequencing and assembly information of newly sequenced plastomes. Q30: the percentage of bases with Phred quality score greater than 30 in the total base.plastomes. (DOCX 27 kb)

Additional file 2: Figure S1. The ratio of nonsynonymous and synonymous substitutions (ω , dN/dS) within each protein coding gene, as calculated by CodeML in PAML. Genes with $\omega > 1$ are colored in red. (PDF 1045 kb)

Additional file 3: Figure S2. Inferred molecular phylogeny from ML (maximum likelihood) and BI (Bayesian inference) analyses for the amino acid sequence of (79 CDS; data set 2). Maximum likelihood bootstrap values (BS) and posterior probabilities (PP) are shown at nodes. Branches with * have 100% bootstrap support and 1.0 posterior probability. (PDF 272 kb)

Additional file 4: Figure S3. Estimated species tree from 79 CDS alignment by coalescent-based approach. Local posterior probabilities are labeled at nodes. Branches with * have 1.0 posterior probability. The clade of *S. kingii* is colored in red. (PDF 548 kb)

Abbreviations

AIC: Akaike Information Criterion; ASDF: Average standard deviation of split frequencies; BI: Bayesian Inference; CDS: Protein-coding sequences; CTAB: Cetyl trimethylammonium bromide; DnaSP: DNA Sequences Polymorphism; DOGMA: Dual Organellar Genome Annotator; GTR: General time reversible; IR: Inverted repeat; ITS: Internal transcribed spacer of ribosomal DNA; LSC: Large single copy; ML: Maximum Likelihood; NCBI: National Center for Biotechnology Information; PI: Parsimony-informative; rRNA: Ribosomal RNA; SSC: Small single copy; tRNA: Transfer RNA

Acknowledgements

Not applicable.

Authors' contributions

HW, YS and HS conceived and designed the study. XZ performed de novo assembly, genome annotation, phylogenetic and other analyses. XZ, TD, MJM, YS, YJ, HW and HS drafted the manuscript. XZ, TD and NL collected the leaf materials. XZ, NL, HZ and AM performed the experiments. All authors discussed the results and helped shape the research, analyses and final manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the National Key R&D Program of China (2017YFC0505200), and grants-in-aid from the Major Program of the National Natural Science Foundation of China (31590823).

Availability of data and materials

All sequences used in this study are available from the National Center for Biotechnology Information (NCBI) (see Additional file 1: Table S1).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ³Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China. ⁴Department of Biology, Oberlin College, 119 Woodland St, Oberlin, OH, USA.

Received: 21 March 2019 Accepted: 19 June 2019

Published online: 02 July 2019

References

- Bremer K. Asteraceae: cladistics and classification. Portland: Timber Press; 1994.
- Susanna A, Garcia-Jacas N. Cardueae (Carduoideae). In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, editors. Systematics, Evolution, and Biogeography of Compositae. Vienna: IAPT; 2009. p. 293–313.
- Shi Z, Raab-Straube EV. Cardueae. In: Flora of China. Edited by Wu ZY, Raven, P. H. & Hong, D. Y., vol. 20–21. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press 2011: 42–194.
- Chen Y. Asteraceae II *Saussurea*, vol. 48(2). Beijing: Science Press; 2015.
- Wang Y-J, Susanna A, Von Raab-Straube E, Milne R, Liu J-Q. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai-Tibetan plateau. Biol J Linn Soc. 2009;97(4):893–903 <https://doi.org/10.1111/j.1095-8312.2009.01225.x>.
- Häffner E. On the phylogeny of the subtribe Carduinae (tribe Cardueae, Compositae). Englera. 2000;21:3–208 <https://doi.org/10.2307/3776757>.
- Raab-Straube EV. Phylogenetic relationships in *Saussurea* (Compositae, Cardueae) sensu lato, inferred from morphological, ITS and *trn L-trn F*

- sequence data, with a synopsis of *Himalaiella* gen. Nov., *Lipschitzella* and *Frolovia*. Willdenowia. 2003;33(2):379–402 <https://doi.org/10.3372/wi.33.33214>.
8. Susanna A, Garcia-Jacas N, Hidalgo O, Vilatersana R, Garnatje T. The Cardueae (Compositae) revisited: insights from ITS, *trnL-trnF*, and *matK* nuclear and chloroplast DNA analysis. Ann Mo Bot Gard. 2006;150–71 [https://doi.org/10.3417/00266493\(2006\)93\[150:TCCRIF\]2.0.CO;2](https://doi.org/10.3417/00266493(2006)93[150:TCCRIF]2.0.CO;2).
 9. Fu ZX, Jiao BH, Nie B, Zhang GJ, Gao TG. A comprehensive generic-level phylogeny of the sunflower family: implications for the systematics of Chinese Asteraceae. J Syst Evol. 2016;54(4):416–37 <https://doi.org/10.1111/jse.12216>.
 10. Kita Y, Kazumi F, Ito M, Ohba H, Kato M. Molecular phylogenetic analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). Taxon. 2004;53(3):679–90 <https://doi.org/10.2307/4135443>.
 11. Wang Y-J, von Raab-Straube E, Susanna A, Liu J-Q. *Shangwua* (Compositae), a new genus from the Qinghai-Tibetan plateau and Himalayas. Taxon. 2013; 62(5):984–96 <https://doi.org/10.12705/625.19>.
 12. Wang Y-J, Liu J-Q. Phylogenetic analyses of *Saussurea* sect. *Pseudoeriocoryne* (Asteraceae: Cardueae) based on chloroplast DNA *trnL-F* sequences. Biochem Syst Ecol. 2004;32(11):1009–23 <https://doi.org/10.1016/j.bse.2004.04.005>.
 13. Lipschitz S: Rod Saussurea DC. (Asteraceae). Leningrad: "Nauka", Leningradskoje Otdelenie; 1979.
 14. Wang Y-J, Liu J-Q, Miehle G. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. Ann Bot. 2007;99(2):311–22 <https://doi.org/10.1093/aob/mcl259>.
 15. Herrando-Moraira S, Calleja JA, Carnicero P, Fujikawa K, Galbany-Casals M, Garcia-Jacas N, Im H-T, Kim S-C, Liu J-Q, López-Alvarado J, López-Pujol J, Mandel JR, Massó S, Mehregan I, Montes-Moreno N, Pyak E, Roquet C, Sáez L, Sennikov A, Susanna A, Vilatersana R. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol Phylogenet Evol. 2018;128:69–87 <https://doi.org/10.1016/j.jmpev.2018.07.012>.
 16. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci. 2007; 104(49):19369–74 <https://doi.org/10.1073/pnas.0709121104>.
 17. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009;7(1):84 <https://doi.org/10.1186/1741-7007-7-84>.
 18. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci. 2007;104(49):19363–8 <https://doi.org/10.1073/pnas.0708072104>.
 19. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot. 2018;105(3):291–301 <https://doi.org/10.1002/ajb2.1048>.
 20. Wu ZQ, Ge S. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Mol Phylogenet Evol. 2012;62(1):573–8 <https://doi.org/10.1016/j.jmpev.2011.10.019>.
 21. Zhang N, Wen J, Zimmer EA. Another look at the phylogenetic position of the grape order Vitales: chloroplast phylogenomics with an expanded sampling of key lineages. Mol Phylogenet Evol. 2016;101:216–23 <https://doi.org/10.1016/j.jmpev.2016.04.034>.
 22. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol. 2016;209(2):855–70 <https://doi.org/10.1111/nph.13617>.
 23. Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. Chloroplast Phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). Syst Biol. 2014;63(6):933–50 <https://doi.org/10.1093/sysbio/syu054>.
 24. Sun Y, Moore MJ, Lin N, Adelalu KF, Meng A, Jian S, Yang L, Li J, Wang H. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family. BMC Genomics. 2017;18(1):592 <https://doi.org/10.1186/s12864-017-3956-3>.
 25. Yan M, Fritsch PW, Moore MJ, Feng T, Meng A, Yang J, Deng T, Zhao C, Yao X, Sun H. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. Mol Phylogenet Evol. 2018;121:198–211 <https://doi.org/10.1016/j.jmpev.2018.01.004>.
 26. Martin W, Deusch O, Stawski N, Grünheit N, Goremykin V. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. Trends Plant Sci. 2005;10(5):203–9 <https://doi.org/10.1016/j.tplants.2005.03.007>.
 27. Kim K-J, Choi K-S, Jansen RK. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol Biol Evol. 2005;22(9):1783–92 <https://doi.org/10.1093/molbev/msi174>.
 28. Walker JF, Zanis MJ, Emery NC. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). Am J Bot. 2014;101(4):722–9 <https://doi.org/10.3732/ajb.1400049>.
 29. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, Feng X, Gu YQ. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. PLoS One. 2013; 8(2):e57533 <https://doi.org/10.1371/journal.pone.0057533>.
 30. Lam VKY, Darby H, Merckx VSFT, Lim G, Yukawa T, Neubig KM, Abbott JR, Beatty GE, Provan J, Soto Gomez M, Graham SW. Phylogenomic inference in extremis: A case study with mycoheterotroph plastomes. Am J Bot. 2018; 105(3):480–94 <https://doi.org/10.1002/ajb2.1070>.
 31. Walker JF, Stull GW, Walker-Hale N, Vargas OM, Larson DA. Characterizing gene tree conflict in plastome-inferred phylogenies. bioRxiv. 2019:512079 <https://doi.org/10.1101/512079>.
 32. Bock DG, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants. Mol Ecol. 2014;23(20):4899–911 <https://doi.org/10.1111/mec.12920>.
 33. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2014;31(3):645–59 <https://doi.org/10.1093/molbev/mst257>.
 34. Su H-J, Barkman TJ, Hao W, Jones SS, Naumann J, Skippington E, Wafula EK, Hu J-M, Palmer JD, dePamphilis CW: novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. Proc Natl Acad Sci. 2019;116(3):934–43 <https://doi.org/10.1073/pnas.1816822116>.
 35. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Bellot S, Schneeweiss GM. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. Proc Natl Acad Sci. 2016;113(32):9045–50 <https://doi.org/10.1073/pnas.1607576113>.
 36. Xie Q, Shen KN, Hao X, Nam PN, Hieu BTN, Chen CH, Zhu C, Lin YC, Hsiao CD. The complete chloroplast genome of Tianshan snow Lotus (*Saussurea involucreata*), a famous traditional Chinese medicinal plant of the family Asteraceae. Mitochondrial DNA Part A. 2015;28(2):294–5 <https://doi.org/10.3109/19401736.2015.1118086>.
 37. Cheon KS, Kim HJ, Han JS, Kim KA, Yoo KO. The complete chloroplast genome sequence of *Saussurea chabyoungsanica* (Asteraceae), an endemic to Korea. Conserv Genet Resour. 2016;1–3 <https://doi.org/10.1007/s12686-016-0617-9>.
 38. Yun SA, Gil HY, Kim SC. The complete chloroplast genome sequence of *Saussurea polylepis* (Asteraceae), a vulnerable endemic species of Korea. Mitochondrial DNA Part B. 2017;2(2):650–1 <https://doi.org/10.1080/23802359.2017.1375881>.
 39. Salih RHM, Majesky L, Schwarzacher T, Gornall R, Heslop-Harrison P. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): identity and variation between three microspecies. PLoS One. 2017;12(2): e0168008 <https://doi.org/10.1371/journal.pone.0168008>.
 40. Su Y, Huang L, Wang Z, Wang T. Comparative chloroplast genomics between the invasive weed *Mikania micrantha* and its indigenous congener *Mikania cordata*: structure variation, identification of highly divergent regions, divergence time estimation, and phylogenetic analysis. Mol Phylogenet Evol. 2018;S1055790317307212 <https://doi.org/10.1016/j.jmpev.2018.04.015>.
 41. Sun Y, Moore MJ, Zhang S, Soltis PS, Soltis DE, Zhao T, Meng A, Li X, Li J, Wang H. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Mol Phylogenet Evol. 2016;96:93–101 <https://doi.org/10.1016/j.jmpev.2015.12.006>.
 42. Sun Y-X, Moore MJ, Meng A-P, Soltis PS, Soltis DE, Li J-Q, Wang H-C. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS One. 2013;8(4):e60429 <https://doi.org/10.1371/journal.pone.0060429>.

43. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Bot.* 2015;40(1):336–51 <https://doi.org/10.1600/036364415X686620>.
44. Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytol.* 2019;221(1):515–26. <https://doi.org/10.1111/nph.15386>.
45. Nolan K, Saemundur S, Hannes D, Yong YJ, Dapeng Z, Engels JMM, Quentin C. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Botany.* 2012;99(2): 320 <https://doi.org/10.3732/ajb.1100570>.
46. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep.* 2015;5:8348 <https://doi.org/10.1038/srep08348>.
47. Prince LM. Plastid primers for angiosperm Phylogenetics and Phylogeography. *Appl. Plant Sci.* 2015;3(6):1400085 <https://doi.org/10.3732/apps.1400085>.
48. Yu XQ, Drew BT, Yang JB, Gao LM, Li DZ. Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: insights into DNA barcoding and phylogeny. *PLoS One.* 2017;12(6):e0178026 <https://doi.org/10.1371/journal.pone.0178026>.
49. Gao X, Zhang X, Meng H, Li J, Zhang D, Liu C. Comparative chloroplast genomes of *Paris* Sect. *Marmorata*: insights into repeat regions and evolutionary implications. *BMC Genomics.* 2018;19(10):878 <https://doi.org/10.1186/s12864-018-5281-x>.
50. Neubig KM, Whitten WM, Carlswald BS, Blanco MA, Endara L, Williams NH, Moore M. Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Plant Syst Evol.* 2009;277(1):75–84 <https://doi.org/10.1007/s00606-008-0105-0>.
51. Omori Y, Takayama H, Ohba H. Selective light transmittance of translucent bracts in the Himalayan giant glasshouse plant *Rheum nobile* Hook.f. & Thomson (Polygonaceae). *Bot J Linn Soc.* 2000;132(1):19–27 <https://doi.org/10.1111/j.1095-8339.2000.tb01852.x>.
52. Tsukaya H, Tsuge T. Morphological adaptation of inflorescences in plants that develop at low temperatures in early spring: the convergent evolution of "downy plants". *Plant Biol.* 2010;3(5):536–43 <https://doi.org/10.1055/s-2001-17727>.
53. Yang Y, Körner C, Sun H. The ecological significance of pubescence in *Saussurea Medusa*, a high-elevation Himalayan "woolly plant". *Arct Antarct Alp Res.* 2008;40(1):250–5 [https://doi.org/10.1657/1523-0430\(07-009\)\[YANG\]2.0.CO;2](https://doi.org/10.1657/1523-0430(07-009)[YANG]2.0.CO;2).
54. Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK-S, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, Sage RF, Covshoff S, Hibberd JM, Nelson MN, Smith SA. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol.* 2015;32(8): 2001–14 <https://doi.org/10.1093/molbev/msv081>.
55. Valderrama E, Richardson JE, Kidner CA, Madiñán S, Stone GN. Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renalmia* L.f., Zingiberaceae). *Mol Phylogenet Evol.* 2018; 119:13–24 <https://doi.org/10.1016/j.ympev.2017.10.001>.
56. Yang J-B, Li D-Z, Li H-T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour.* 2014;14(5):1024–31 <https://doi.org/10.1111/1755-0998.12251>.
57. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20 <https://doi.org/10.1093/bioinformatics/btu170>.
58. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004;20(17):3252–5 <https://doi.org/10.1093/bioinformatics/bth352>.
59. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–W11 <https://doi.org/10.1093/nar/gkx391>.
60. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647–9 <https://doi.org/10.1093/bioinformatics/bts199>.
61. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25(5): 955–64 <https://www.ncbi.nlm.nih.gov/pubmed/9023104>.
62. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 2013;41(W1):W575–81 <https://doi.org/10.1093/nar/gkt289>.
63. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain. Loss and Rearrangement *Plos One.* 2010;5(6): e11147 <https://doi.org/10.1371/journal.pone.0011147>.
64. Rozas J, Ferrer-Mata A. SÁ nchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sà n-GA: DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol Biol Evol.* 2017;34(12):3299–302 <https://doi.org/10.1093/molbev/msx248>.
65. Zhang D, Gao F, Li WX, Jakovlić I, Zou H, Zhang J, Wang GT: PhyloSuite: an integrating and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *bioRxiv* 2018:489088. <https://doi.org/10.1101/489088>.
66. Kumar S, Stecher G, Li M, Nkay C, Tamara K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547–9 <https://doi.org/10.1093/molbev/msy096>.
67. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics (Oxford, England).* 1998;14(9):817–8 <https://doi.org/10.1093/bioinformatics/14.9.817>.
68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3 <https://doi.org/10.1093/bioinformatics/btu033>.
69. Ronquist F, Huelsenbeck JP. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17(8):754–5 <https://doi.org/10.1093/bioinformatics/17.8.754>.
70. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19(S6):153 <https://doi.org/10.1186/s12859-018-2129-y>.
71. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91 <https://doi.org/10.1093/molbev/msm088>.
72. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005;22(12):2472–9 <https://doi.org/10.1093/molbev/msi237>.
73. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 1998;15(5):568–73 <https://doi.org/10.1093/oxfordjournals.molbev.a025957>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

