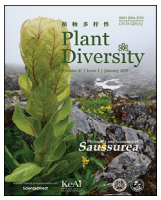
Plant Diversity 47 (2025) 21–33

Contents lists available at ScienceDirect

Plant Diversity

journal homepage: http://www.keaipublishing.com/en/journals/plant-diversity/
http://journal.kib.ac.cn



Research paper

New insights into the phylogeny and infrageneric taxonomy of *Saussurea* based on hybrid capture phylogenomics (Hyb-Seq)

Liansheng Xu [a, b], Zhuqiu Song [a, b], Tian Li [a, d], Zichao Jin [a, c], Buyun Zhang [a], Siyi Du [a, c], Shuyuan Liao [a, c], Xingjie Zhong [a], Yousheng Chen [a, b, *]

[a] *Plant Science Research Center, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*
[b] *Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, Guangzhou 510650, China*
[c] *University of the Chinese Academy of Sciences, Beijing 100049, China*
[d] *Central-Southern Safety and Environment Technology Institute Co., LTD, Wuhan 430064, China*

ARTICLE INFO

Article history:
Received 5 March 2024
Received in revised form
10 October 2024
Accepted 10 October 2024
Available online 18 October 2024

Keywords:
Asteraceae
Morphology
Next-generation sequencing
Phylogenomics
Subgenus

ABSTRACT

Saussurea is one of the largest and most rapidly evolving genera within the Asteraceae, comprising approximately 520 species from the Northern Hemisphere. A comprehensive infrageneric classification, supported by robust phylogenetic trees and corroborated by morphological and other data, has not yet been published. For the first time, we recovered a well-resolved nuclear phylogeny of *Saussurea* consisting of four main clades, which was also supported by morphological data. Our analyses show that ancient hybridization is the most likely source of deep cytoplasmic-nuclear conflict in *Saussurea*, and a phylogeny based on nuclear data is more suitable than one based on chloroplast data for exploring the infrageneric classification of *Saussurea*. Based on the nuclear phylogeny obtained and morphological characters, we proposed a revised infrageneric taxonomy of *Saussurea*, which includes four subgenera and 13 sections. Specifically, 1) *S.* sect. *Cincta*, *S.* sect. *Gymnocline*, *S.* sect. *Lagurostemon*, and *S.* sect. *Strictae* were moved from *S.* subg. *Saussurea* to *S.* subg. *Amphilaena*, 2) *S.* sect. *Pseudoeriocoryne* was moved from *S.* subg. *Eriocoryne* to *S.* subg. *Amphilaena*, and 3) *S.* sect. *Laguranthera* was moved from *S.* subg. *Saussurea* to *S.* subg. *Theodorea*.

1. Introduction

Saussurea DC. (Candolle, 1810) is one of the largest genera in the tribe Cardueae (Asteraceae), encompassing approximately 520 species in the Northern Hemisphere (Raab-Straube, 2017; Zhang et al., 2021d). These species exhibit extraordinary diversity in both morphologies and preferred habitats (Fig. 1; Lipschitz, 1979; Chen, 2015). The first comprehensive infrageneric classification of *Saussurea* was proposed by Lipschitz (1979), and included six subgenera and 19 sections (Fig. 2). However, recent molecular studies have raised concerns about the artificiality of this classification and recognized *Saussurea* as a monophyletic genus, resulting in a revised classification with four subgenera (*S.* subg. *Amphilaena* (Stschegl.) Lipsch., *S.* subg. *Eriocoryne* (DC.) Hook. f., *S.* subg. *Saussurea*, and *S.* subg. *Theodorea* (Cass.) Lipsch.) and 14 sections (Raab-Straube, 2003; Shi and Raab-Straube, 2011; Wang et al., 2013; Chen, 2015; Yuan et al., 2015). Herrando-Moraira et al. (2020) moved two of three species in *S.* sect. *Jurineiformes* (Lipsch.) Lipsch. to *Jurinea* Cass., including the type species, *S. chondrilloides* C. Winkl., thus, this section no longer belongs to *Saussurea* (Fig. 2). *Hemisteptia* Fisch. & C.A. Mey. and *Polytaxis* Bunge are sister groups to *Saussurea* and are not nested within *Saussurea* in the phylogenetic trees of previous molecular studies (Xu et al., 2019; Herrando-Moraira et al., 2020). Unlike Herrando-Moraira et al. (2020), we chose to focus on a narrower definition of *Saussurea*, excluding *Hemisteptia* and *Polytaxis*, to simplify the infrageneric classification in the present study.

Molecular evidence has challenged the existing *Saussurea* infrageneric classification (Raab-Straube, 2003; Kita et al., 2004; Wang et al., 2009, 2013; Xu et al., 2019; Zhang et al., 2021b). However, a comprehensive classification supported by robust phylogenetic trees, incorporating morphological and other data, is

* Corresponding author. Plant Science Research Center, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China.
E-mail address: yschen@scbg.ac.cn (Y. Chen).
Peer review under the responsibility of Editorial Office of Plant Diversity.

https://doi.org/10.1016/j.pld.2024.10.003

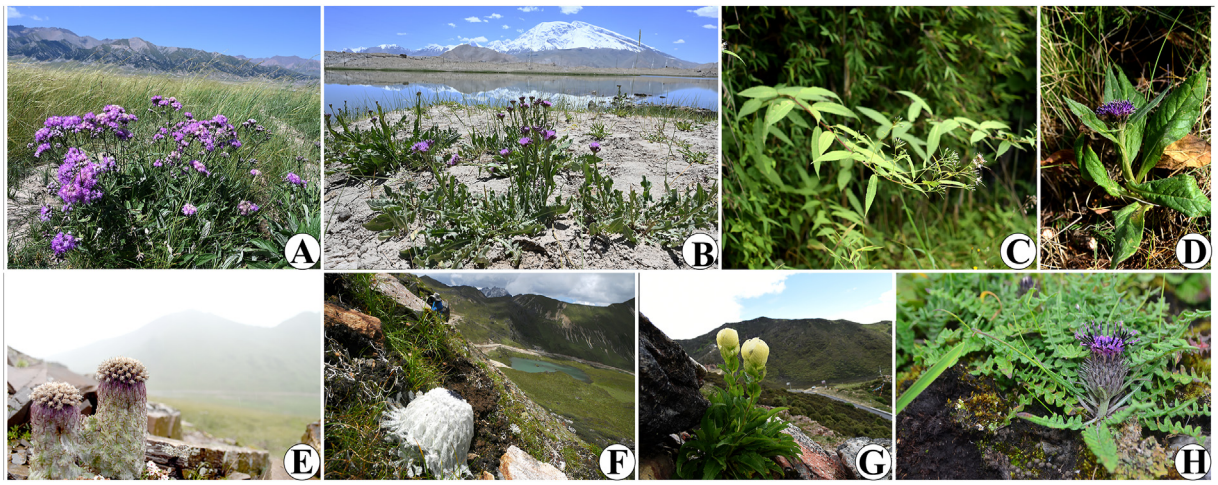


Fig. 1. *Saussurea* species diversity. A. *S. elegans* (CHINA. Xinjiang: Bole, 28 Jul. 2022); B. *S. salsa* var. *pamirica* (CHINA. Xinjiang: Akto, 18 Jul. 2022); C. *S. dolichopoda* (CHINA. Shaanxi: Xi'an, 29 Jul. 2021); D. *S. alpina* (CHINA. Xinjiang: Altay, 22 Aug. 2014); E. *S. medusa* (CHINA. Sichuan: Diebu, 17 Jul. 2023); F. *S. tridactyla* (CHINA. Xizang: Cona, 12 Aug. 2013); G. *S. glandulosissima* (CHINA. Xizang: Nyingchi, 18 Aug. 2023); H. *S. leontodontoides* (CHINA. Sichuan: Maoxian, 8 Aug. 2018). A, B, D, and F were photographed by Y.S. Chen, and the remaining by L.S. Xu.

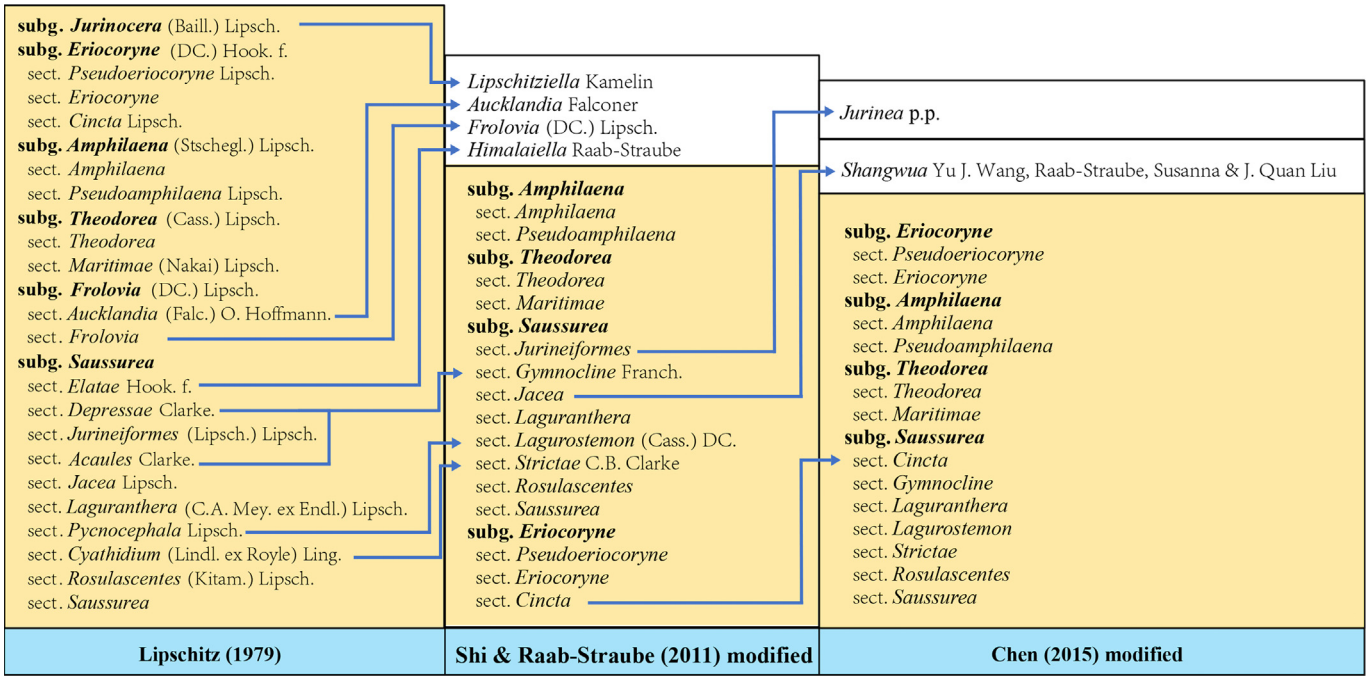


Fig. 2. Historical infrageneric classification schemes for *Saussurea*. Members of *Saussurea* are on yellow background. Main changes are marked with blue arrows. Groups removed from *Saussurea* are on white background. Study citations are on blue background.

lacking (Chen, 2015). Previous attempts, using a limited set of traditional Sanger sequence markers, failed to produce reliable trees with high support values (Fig. 3A–C; Raab-Straube, 2003; Kita et al., 2004; Wang et al., 2009; Wang et al., 2013) due to the extremely low levels of molecular divergence among closely related species and the insufficient number of informative sites, particularly in rapidly evolving groups. Phylogenomic studies of *Saussurea* have used three types of next-generation sequencing (NGS) data (Fig. 3D–F): chloroplast genome sequences, RNA-sequencing data (RNA-Seq), and hybrid capture phylogenomics (Hyb-Seq). Phylogenies based on chloroplast genomes have shown relatively high support values and revealed the artificiality of the current infrageneric classification (Fig. 3D). However, these phylogenies did not incorporate other forms of evidence, such as morphology (Xu et al., 2019; Zhang et al., 2021b), and were therefore unable to provide a cogent infrageneric classification. Hyb-Seq (Fig. 3E with 39 samples from 36 species; Herrando-Moraira et al., 2020) and RNA-Seq (Fig. 3F with 5 samples from 5 species; Zhang et al., 2021c) phylogenies to date have not included a broadly representative sample of the genus, and consequently, were unable to elucidate phylogenetic relationships of infrageneric groups.

Hyb-Seq, known as "target/hybrid enrichment" or "sequence capture", is a valuable technique in phylogenomic and evolutionary studies (Mandel et al., 2014). It allows the targeted capture of specific DNA regions (target regions) using probes or "baits" designed from the known genome or transcriptome of closely

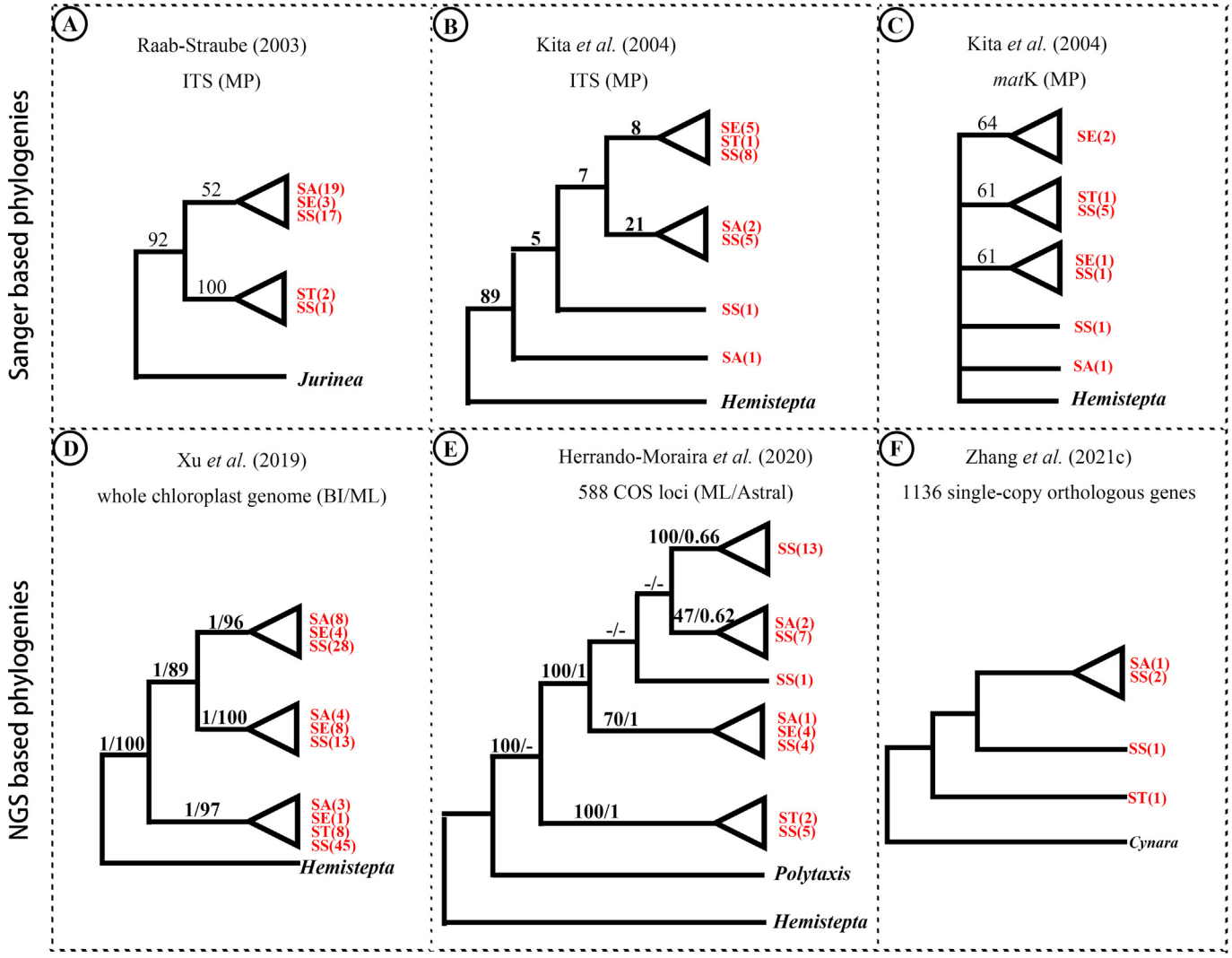


Fig. 3. Comparison of previous phylogenies performed on *Saussurea*. Below the study citation are specified the molecular markers and the phylogenetic inference method used. The branch support values correspond to bootstrap values in the case of maximum parsimony (MP) and maximum likelihood (ML) methods, posterior probabilities in Bayesian inference (BI), and local posterior probabilities in the coalescent Astral method (Astral). A dash (–) indicates inconsistency in topologies among different phylogenies. Red letters indicate the subgenus of *Saussurea* to which the species belongs (SA, *S.* subg. *Amphilaena*; SE, *S.* subg. *Eriocoryne*; SS, *S.* subg. *Saussurea*; ST, *S.* subg. *Theodorea*) and the number of samples.

related species, rather than sequencing the entire genome (Herrando-Moraira et al., 2018). For the Compositae, Mandel et al. (2014) developed a target enrichment method, which uses the Hyb-Seq approach, comprising a probe (myBaits COS Compositae1Kv1) set of 9678 baits targeting a total of 1061 conserved orthologous loci (COS), identified from thousands of expressed sequence tags (EST) across three available genomes of the family. Recently, Moore-Pollard et al. (2024) designed another probe set: Compositae-ParaLoss-1272. These probe sets have proven to be effective in generating well-resolved phylogenies at various taxonomic levels (Herrando-Moraira et al., 2018, 2019, 2020; Mandel et al., 2019; Siniscalchi et al., 2019; Xu and Chen, 2021; Hatami et al., 2022). Moreover, Hyb-Seq is cost-effective, requiring only silica gel-dried or herbarium tissues, making it more accessible compared to RNA-Seq, which necessitates fresh tissues.

This study used molecular (from Hyb-Seq) and morphological data to 1) determine which molecular data (nuclear or chloroplast) and morphological characters are suitable for exploring the infrageneric classification of *Saussurea*, 2) propose an infrageneric framework at the subgenus level, and 3) reclassify the existing sections to subgenus based on the sampled species.

2. Materials and methods

2.1. Taxon sampling

Our taxon sampling scheme of *Saussurea* (103 samples of 103 species) covered thirteen sections (Fig. 2) within *Saussurea* (Table S1 with 11 samples of *S.* sect. *Amphilaena,* three of *S.* sect. *Pseudoamphilaena* Lipsch., nine of *S.* sect. *Eriocoryne*, three of *S.* sect. *Pseudoeriocoryne* Lipsch., two of *S.* sect. *Cincta* Lipsch., six of *S.* sect. *Gymnocline* Franch., ten of *S.* sect. *Laguranthera* (C.A. Mey. ex Endl.) Lipsch, three of *S.* sect. *Lagurostemon* (Cass.) DC., four of *S.* sect. *Rosulascentes* (Kitam.) Lipsch., 33 of *S.* sect. *Saussurea*, 11 of *S.* sect. *Strictae* C.B. Clarke, one of *S.* sect. *Maritimae* (Nakai) Lipsch., and seven of *S.* sect. *Theodorea*) and encompassed all three *Saussurea* clades, as determined through studies of the chloroplast genome (Xu et al., 2019; Zhang et al., 2021b). We also incorporated four samples from three closely related genera of *Saussurea*, following previous studies (Mandel et al., 2019; Herrando-Moraira et al., 2020; Zhang et al., 2021a). In total, 107 samples were included in the present study. Of the 68 samples newly sequenced, materials were obtained from our field collections or herbaria (HUH, IBSC,

and PE). Sequence data for the remaining 39 samples were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/). For detailed sampling information, please refer to Table S1.

2.2. DNA extraction and sequencing

Total genomic DNA was extracted from either silica gel-dried tissue or herbarium tissue using a CTAB Plant Genomic DNA Extraction Kit (Biomed Beijing). Probe design, library construction, capture preparation, and sequencing were performed by iGeneTech Bioscience Co., Ltd. (Beijing, China). In summary, 13,623 custom RNA probes (Xu et al., 2024) were designed for 1061 COS described by Mandel et al. (2014). Then 200 ng of genomic DNA from each individual was sheared by a Biorupter (Diagenode, Belgium) to obtain fragments of 150–200 bp. The ends of these fragments were repaired and affixed with an Illumina Adaptor (Fast Library Prep Kit, iGeneTech, Beijing, China). Once the library was constructed, capture-enrichment experiments were carried out following the manufacturer's instructions (iGeneTech, Beijing, China). The captured libraries were mixed in equal molar amounts and then sequenced on partial lanes on the Illumina NovaSeq6000 platform (Illumina, San Diego, CA) with 150 base paired-end reads.

2.3. Sequence extraction and orthologous gene identification

We performed quality checks on raw data using FastQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and then cleaned the data with TrimGalore v.0.6.7 (https://doi.org/10.5281/zenodo.5127899), which was set to trim off ends with more than 5 bp of overlap with adapter sequences or low-quality (< 25) reads, in addition to adapter removal. Reads with a minimum length of 75 bp were retained with both corresponding forward and reverse pairs.

To extract sequence data for nuclear loci, we used targets from the most relevant reference genome, *Carthamus tinctorius* L., which contained a set of 475 COS created by Mandel et al. (2014). For chloroplast data, we used 245 chloroplast sequences (coding and noncoding sequences, extracted with Geneious 11.0.2 (Kearse et al., 2012)) from *Saussurea chabyoungsanica* Im (KX622799, Cheon et al., 2016) as targets. HybPiper v1.3.1 (Johnson et al., 2016) was utilized to extract the sequences: A targets file (Suppl. 1) with 720 sequences (including 475 nuclear and 245 chloroplast sequences) was used to extract nuclear and chloroplast sequences; Trimmed reads were first mapped to the targets using BWA 0.7.17 (Li and Durbin, 2009) and subsequently assembled into contigs using SPAdes 3.13.0 (Bankevich et al., 2012). "paralog_investigator.py" and "paralog_retriever.py" of HybPiper were employed to retrieve paralog sequences.

The sequences were aligned in MAFFT v.7.475 (Katoh and Standley, 2013) with default settings. FastTree v.2.1.11 (Price et al., 2010) was utilized to generate gene trees with SH-like local supports, with values above 0.95 considered statistically well-supported. PhyloPyPruner 1.2.4 (https://gitlab.com/fethalen/phylopypruner/) was used to identify orthologous genes using the alignments with paralog sequences and gene trees from FastTree as input (the nuclear and chloroplast alignments obtained were processed separately). The paralogy pruning method was set to 'LS' (the subtree that contains the highest number of sequences is the "largest subtree" and is retained as an ortholog) and nodes with a support value below 0.95 were collapsed. Alignments shorter than 100 bp, with fewer than 4 operational taxonomic units (OTUs), or with a branch length five times larger than the standard deviation of all branches within the tree were removed. To improve the accuracy of sequence alignment, the alignments of orthologues were checked and modified manually in BioEdit v.7.2.5 (Hall, 1999) and re-aligned in MAFFT with high accuracy (–maxiterate 1000 –localpair). Then we counted the parsimony-informative sites of each matrix with AMAS (Borowiec, 2016), the missing percentage of each sample with HybPhyloMaker v.1.6.4 (Fer and Schmickl, 2018), and the number of extracted sequences with HybPiper. Samples with less than 25 chloroplast sequences (10% of the total chloroplast target sequences) or 48 nuclear sequences (10% of the total nuclear target sequences) obtained were excluded from subsequent phylogenetic analysis to reduce the negative impact of missing data. HybPhyloMaker was employed to select the samples with a missing percentage of less than 90% in two supermatrices, which were concatenated with AMAS from chloroplast and nuclear alignments, respectively.

2.4. Phylogenetic analyses

We employed two complementary approaches for phylogenetic reconstruction: the concatenation approach, using the supermatrix dataset as input, and the coalescent approach, utilizing separate matrices for each locus. For the concatenation approach, we performed the ML analyses using IQ-TREE 2.0.3 (Minh et al., 2020) under Edge-linked partition model for 1000 ultrafast bootstraps (UFboot, Minh et al., 2013), as well as the Shimodaira-Hasegawa-like approximate likelihood-ratio test (SHaLRT, Guindon et al., 2010). Values of SHaLRT $\geq$ 80% and UFboot $\geq$ 95% indicate well-supported branches. For the coalescent approach, 1000 ultrafast bootstrap replicates were used to estimate individual gene trees with IQ-TREE. ASTRAL-III v.5.5.3 (Zhang et al., 2018) was then used to estimate the species tree based on local posterior probabilities (LPP) obtained from the previous set of gene trees (LPP > 0.7 indicates well-supported branches, Sayyari and Mirarab, 2016).

2.5. Cytoplasmic-nuclear discordance analyses

Although the most likely source of cytoplasmic-nuclear discordance is hybridization, incomplete lineage sorting (ILS) may produce a similar pattern (Folk et al., 2017; Zhou et al., 2022). We summarized the number of conflicting and concordant bipartitions with PHYPARTS (Smith et al., 2015), using the nuclear phylogeny estimated by ASTRAL-III and the individual nuclear gene trees, to test whether ILS occurs in the phylogeny. Then we reduced the dataset to a computationally tractable size (19 samples, 4 samples of each ntclade [nuclear tree's clade] of *Saussurea* and one of *Hemisteptia*, *Polytaxis*, and *Jurinea*, respectively). We extracted these 19 samples from the supermatrix using AMAS, then deleted the matrices with less than 4 samples and the samples with more than 80% missing data using the script modified from HybPhyloMaker5_missing_data_removal.sh in HybPhyloMaker. The individual gene trees for each locus were estimated with RAxML v.8.2.9 (Stamatakis, 2014) under the model GTRGAMMA. The method selected was a simultaneous rapid bootstrapping of 1000 replicates to assess branch support and best ML tree search with 10 randomized maximum parsimony starting trees. Individual gene trees (best trees) and bootstrap replicates were used to estimate a species tree in ASTRAL-III with 1000 coalescent bootstrap replicates. To test whether ILS alone explained the incongruence between plastome and nuclear phylogeny, we followed Folk et al. (2017) to simulate 10,000 plastome trees under the coalescent model using DendroPy v.4.1.015561 (Sukumaran and Holder, 2010). The species phylogeny from the ASTRAL-III tree was used as a guide tree for the simulation. To simulate plastome gene trees, branch lengths were scaled by a factor of four to account for the haploidy and maternal inheritance of the plastome. Clade frequencies of simulated trees were summarized using PHYPARTS. If ILS alone is responsible for cytoplasmic-nuclear discordance, the topology from our empirical

plastome gene tree should be present in simulated trees with high frequency; if gene flow is present, the topology recovered in our empirical tree should be absent or at very low frequency in the simulated trees.

To further explore reticulate evolutionary histories within *Saussurea*, we inferred species networks using SNaQ (Solis-Lemus and Ane, 2016), which was implemented in the package PhyloNetworks v.0.12.0 (Solis-Lemus et al., 2017). The species phylogeny from ASTRAL-III and individual gene trees were used as input, and nested analyses were performed allowing for zero (hmax = 0) to four (hmax = 4) hybridization events. Each nested analysis was optimized by 30 independent runs, and the best-fitting model was selected based on the phylogeny and the log pseudolikelihood score.

2.6. Morphological character analyses

To test which character state has a relatively simple evolutionary history and could serve as a diagnostic characteristic for each clade, we performed ancestral reconstruction for 18 morphological character states, which have been commonly used in the classification of *Saussurea* (Table 2). All morphological characters were initially checked against existing literature (Lipschitz, 1979; Shi and Raab-Straube, 2011; Chen, 2015; Raab-Straube, 2017) and subsequently confirmed in specimens (Table S2). Cytological data (Table S1) are primarily from Raab-Straube (2017) and Tropicos [https://www.tropicos.org/]). Bayesian Binary MCMC (BBM) analysis was carried out in RASP 4.3 (Yu et al., 2020), following the method described in Andrés-Sánchez et al. (2014). The input tree for the BBM was an ML tree from the concatenation approach containing only *Saussurea* species. The BBM was run with the Fixed Jukes-Cantor (Fixed(JC)) state frequencies model and Gamma + G among-Site rate variation model for 100 million generations, utilizing 10 chains and 2 parallel runs to obtain a combined result.

3. Results

3.1. Phylogeny

For all samples, the average number of reads per sample obtained was 15,305,930, ranging from 3,110,000 in *Saussurea taraxacifolia* (Lindl. ex Royle) Wall. ex DC. (SRR11926458, sequenced with Hyb-Seq method using myBaits COS Compositae set 1Kv1 (Hyb-Seq 1Kv1)) to 92,009,403 in *S. ochrochlaena* Hand.-Mazz. (SRR14280112, RNA-Seq). For 475 COS, HybPiper obtained 474 matrices from all samples (the average number of COS per sample was 469, ranging from 448 in *S. involucrata* (Kar. & Kir.) Sch. Bip. to 474 in four samples) and 451 matrices were retained after excluding paralogues (the average number of COS per sample was 339, ranging from 250 in *S. robusta* Ledeb. to 358 in *S. licentiana* Hand.-Mazz.; Table S1), resulting in a concatenated nuclear matrix of 306,289 bp with 28,137 (9.2%) parsimony-informative sites (Table S3). The missing percentage of each sample in the concatenated matrix ranged from 71.6% in *S. andryaloides* (DC.) Sch. Bip to 25.9% in *S. amurensis* Turcz. ex DC., with an average of 40.3% (Table S1). For 245 chloroplast sequences, HybPiper obtained 240 matrices from all samples (the average number of chloroplast sequences per sample was 212, ranging from 35 in *S. stubendorffii* Herder to 240 in 47 samples) and 191 matrices were retained after excluding paralogues (the average number of chloroplast sequences per sample in the final matrix was 160, ranging from 26 in *S. acromelaena* Hand.-Mazz. to 189 in five samples; Table S1), resulting in a concatenated matrix of 121,294 bp with 1,842 (1.5%) parsimony-informative sites (Table S3). The missing percentage of each sample in chloroplast supermatrix ranged from 89.5% in *S. acromelaena* to 5.4% in *S. globosa* F.H. Chen, with an average mean of 18.5% (Table S1).

Phylogenetic reconstructions of nuclear data, both through the concatenation and coalescent approaches, yielded highly similar topologies of four ntclades (Fig. S1); most discordances were within each ntclade. In the resulting trees (Figs. 4(left) and S1), *Saussurea* was found to be sister to *Hemisteptia* and contained four clades (ntclade 1–4), which does not fully support the current classification (Fig. 4(left) and Table 1). Ntclade 1 (SHaLRT = 100; UFboot = 100; LPP = 1.00) contained seven samples of *S.* sect. *Theodorea*, one of *S.* sect. *Maritimae*, and 10 of *S.* sect. *Laguranthera*; ntclade 2 (SHaLRT = 100; UFboot = 100; LPP = 1) contained four of *S.* sect. *Rosulascentes*, and 28 of *S.* sect. *Saussurea*; ntclade 3 (SHaLRT = 100; UFboot = 100; LPP = 0.99) contained nine of *S.* sect. *Eriocoryne* and one of *S.* sect. *Pseudoeriocoryne*; ntclade 4 (SHaLRT = 100; UFboot = 99; LPP = 0.91) contained two of *S.* sect. *Pseudoeriocoryne*, 11 of *S.* sect. *Amphilaena*, three of *S.* sect. *Pseudoamphilaena,* two of *S.* sect. *Cincta*, six of *S.* sect. *Gymnocline*, three of *S.* sect. *Lagurostemon*, 11 of *S.* sect. *Strictae*, and five of *S.* sect. *Saussurea*.

The phylogeny based on chloroplast data (Figs. 4(right) and S2) closely resembled that of previous studies (Xu et al., 2019; Zhang et al., 2021b), revealing three main clades (ctclade [chloroplast tree's clade] 1–3) within *Saussurea*, which also does not completely support the current classification (Table 1). Ctclade 1 (SHaLRT = 88; UFboot = 99) contained 12 samples of *S.* subg. *Amphilaena*, five of *S.* subg. *Eriocoryne*, and 18 of *S.* subg. *Saussurea*; ctclade 2 (SHaLRT = 99; UFboot = 100) contained one of *S.* subg. *Amphilaena*, seven of *S.* subg. *Eriocoryne* and five samples of *S.* subg. *Saussurea*; and ctclade 3 (SHaLRT = 74; UFboot = 99) contained one of *S.* subg. *Amphilaena*, 46 of *S.* subg. *Saussurea* and eight of *S.* subg. *Theodorea*.

For further details on phylogeny construction, please refer to Table 1.

3.2. Cytoplasmic-nuclear discordance

The discordance between the nuclear and chloroplast phylogenies was evident when we compared the two phylogenetic trees (Fig. 4). Ctclade 3 encompassed 48 out of 50 samples from ntclades 1 and 2, whereas ntclades 3 and 4 contained 46 out of 49 samples from ctclades 1 and 2. Notably, 1) ntclade 2 was closely related to ntclades 3 and 4 in the nuclear phylogeny; however, the samples within ntclade 2 did not form a clade and are mostly intermixed with ntclade 1 in the chloroplast phylogeny; 2) ntclades 3 and 4 were more evolved groups in the nuclear phylogeny, whereas they exhibited a different pattern in the chloroplast phylogeny.

Analysis of the phylogeny estimated by ASTRAL-III and the individual gene trees showed that most nodes of the nuclear phylogeny within *Saussurea* are not highly supported by nuclear gene trees (Fig. S3).

Analyses of cytoplasmic-nuclear discordance showed that the topology recovered in our empirical tree occurs at a very low frequency in the simulated trees of each clade of *Saussurea* (Fig. 5 (right)). Analysis of SNaQ (Figs. 5 (left) and S4) showed that when the number of hybridization events was two (hmax = 2), the network score was the best (-Ploglik = 1524.5). There were two gene flow events within *Saussurea:* between ntclades 1 and 2 (0.65/0.35) and within ntclade 1 (0.09/0.91).

3.3. Morphological character analyses

We examined 18 morphological character states (Table 2; Figs. S5–S10). Thirteen character states arose independently more than once, whereas the remaining five (highly branched from lower

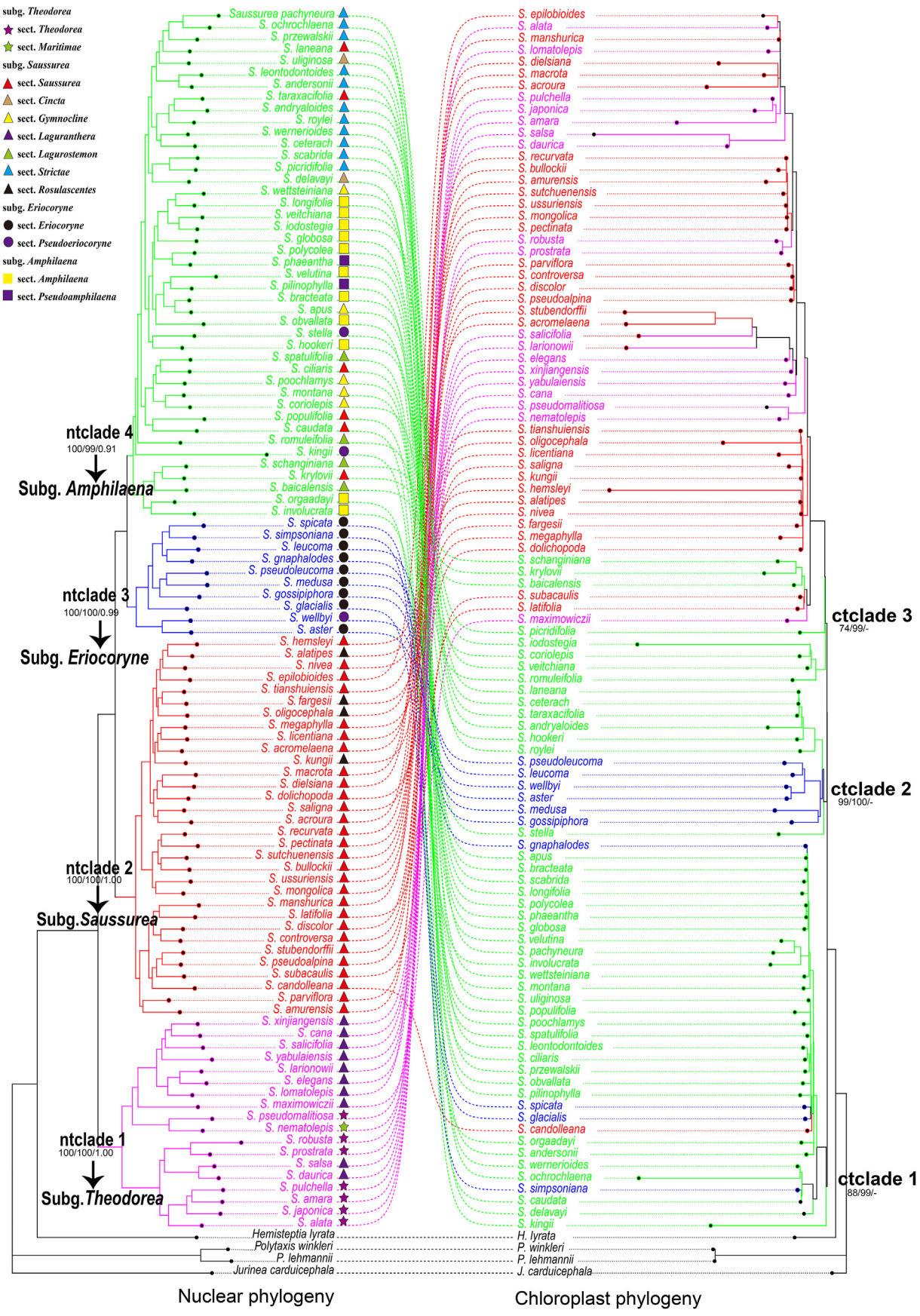


Fig. 4. Phylogenetic reconstruction obtained via the concatenation approach based on 451 nuclear loci [left] and 191 chloroplast sequences [right] from IQ-TREE. Ultrafast bootstrap, Shimodaira-Hasegawa-like approximate likelihood-ratio test values, and local posterior probabilities are marked above each branch [SHaLRT/UFboot/LPP]). Subgenus and sections are marked following the species name with a colored symbol. The color of branch lines, connecting lines, and species names in ntclade 1 is purple, ntclade 2 is red, ntclade 3 is blue, and ntclade 4 is green.

Table 1
Phylogenetic information obtained from nuclear and chloroplast data.

Clades	Support values (SHaLRT/UFboot/LPP)	Number of samples of each previous subgenus			
		Amphilaena	*Eriocoryne*	*Saussurea*	*Theodorea*
ntclade1	100/100/1.00	0	0	10	8
ntclade2	100/99/1.00	0	0	32	0
ntclade3	100/100/0.99	0	10	0	0
ntclade4	100/99/0.91	14	2	27	0
ctclade1	88/99/-	12	5	18	0
ctclade2	99/100/-	1	7	5	0
ctclade3	74/99/-	1	0	46	8

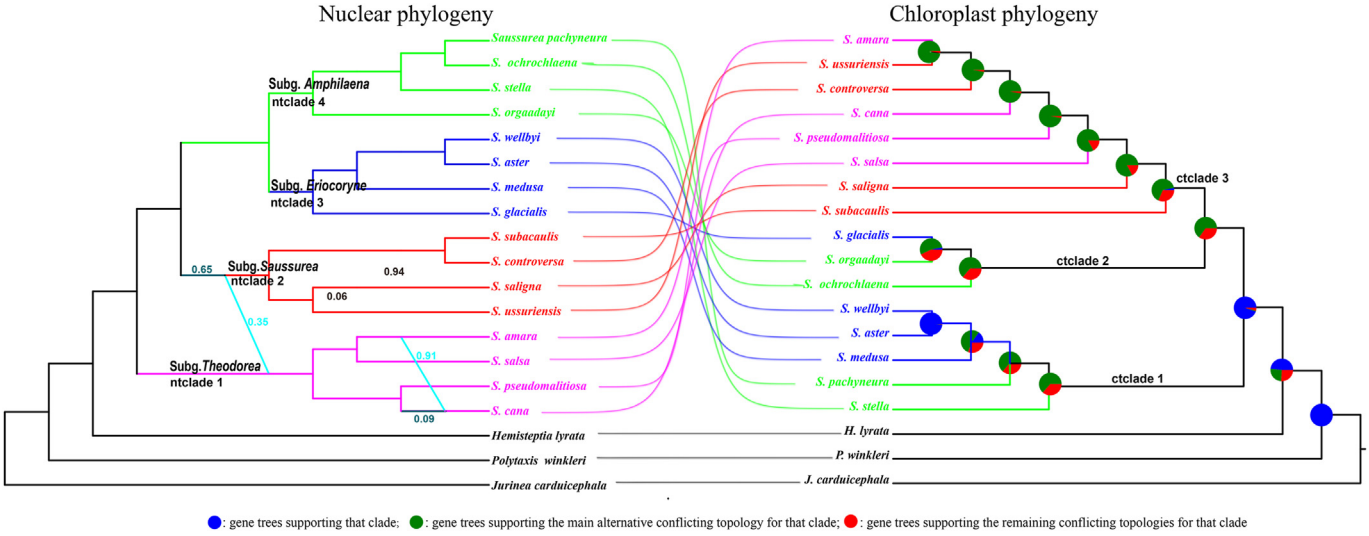


Fig. 5. Cytoplasmic-nuclear discordance of *Saussurea*. Left: gene flow among *Saussurea* inferred by SNaQ analyses. The light blue and dark blue lines indicate hybrid edges. The light blue and dark blue numbers indicate estimated inheritance probabilities from major and minor parental species, respectively. Right: conflict between plastome phylogeny and simulated plastid gene trees. Pie charts indicate the proportions of 10,000 simulated plastid gene trees supporting that clade (blue), the main alternative conflicting topology for that clade (green), the remaining conflicting topologies for that clade (red). The color of branch lines, connecting lines, and species names in ntclade 1 is purple, ntclade 2 is red, ntclade 3 is blue, and ntclade 4 is green.

part, highly branched on the upper part of the stems, uppermost leaves densely woolly on both surfaces, capitula arranged in clustered secondary heads, and phyllaries with colored apices) emerged independently only once (Table 2; Fig. 6A–D), and may serve as diagnostic characteristics. Moreover, the chromosome base numbers of almost all species recorded in ntclades 1 and 2 were 12–14, whereas those in 3 and 4 were 15–18 (Fig. S11).

4. Discussion

4.1. *Performance of Hyb-Seq of* Saussurea

To study the generic boundaries of subtribe Saussureinae, Herrando-Moraira et al. (2020) targeted 1061 loci, of which 1054 (99.3%) were recovered, and 588 (55.4%) were retained after

Table 2
Ancestral state reconstruction of 18 morphological character states.

Characters and coding		Number of originations
Stem	A: Short or nearly absent	≥ 2
	B: Solitary or rarely branched	≥ 2
	C: Highly branched from the lower part	1
	D: Highly branched on the upper part	1
Basal leaves (at anthesis)	A: Rosette	≥ 2
	B: Not rosette	≥ 2
Uppermost leaves	A: Membranous, surrounding synflorescence	≥ 2
	B: Similar to mid-stem leaves	≥ 2
	C: Densely woolly on both surfaces	1
Involucre diameter	A: < 10 mm	≥ 2
	B: ≥ 10 mm	≥ 2
Inflorescence	A: Capitula one or two	≥ 2
	B: Capitula arranged in compound corymbiform synflorescence	≥ 2
	C: Capitula arranged in simple corymbiform synflorescence	≥ 2
	D: Capitula arranged in clustered secondary heads	1
Phyllary	A: Apices colored	1
	B: Apices not colored	≥ 2

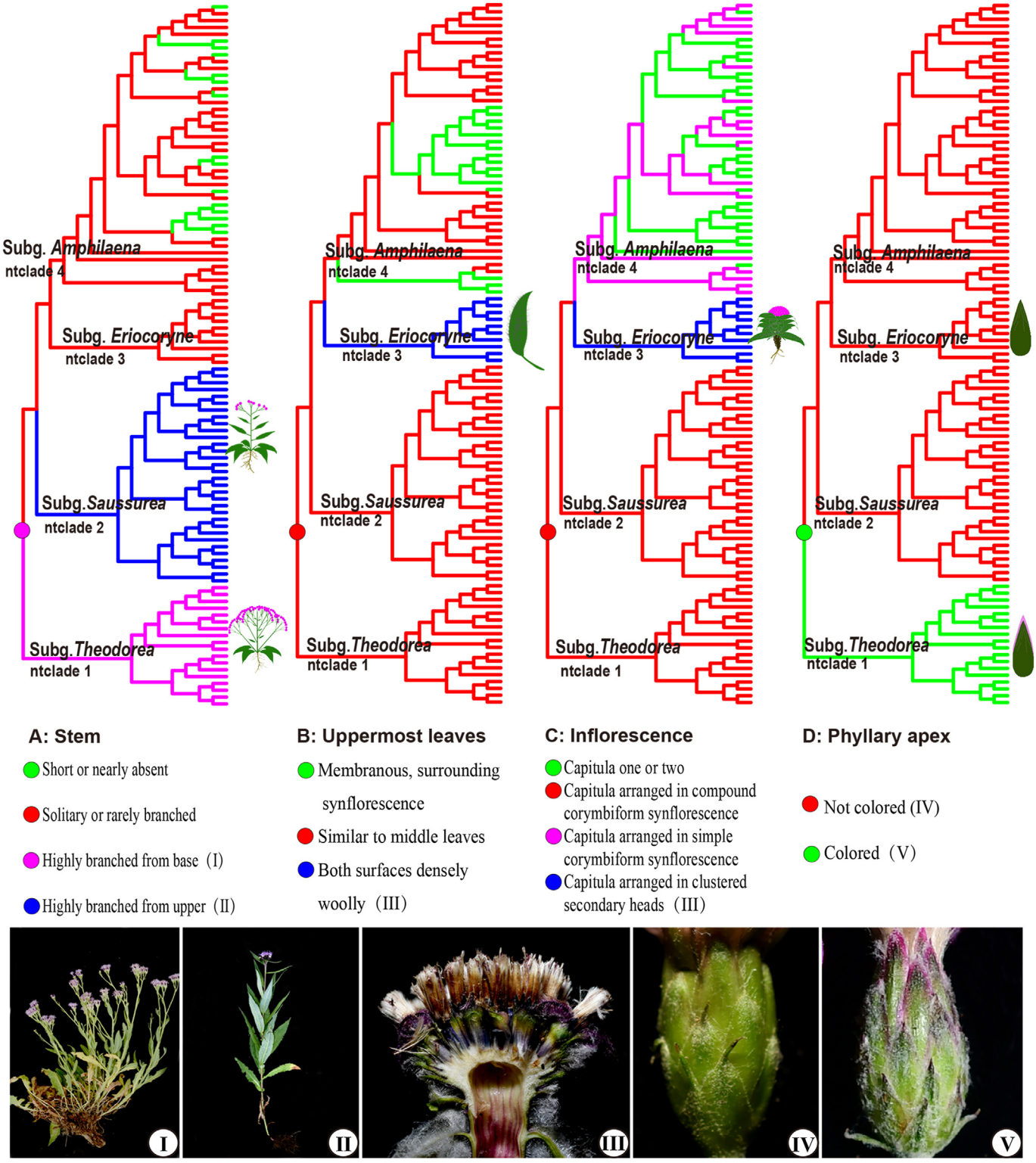


Fig. 6. Diagnostic characteristics of each ntclade. A. Stem. B. Uppermost leaves. C. Inflorescence. D. Phyllary apex. Detailed probabilities of characters for each node are shown in Figs. S5–S10. The colored line drawings indicate the diagnostic characteristics of each ntclade. The colors of dots on the main nodes indicate the characters states. I. *Saussurea salsa var. pamirica* (CHINA. Xinjiang: Akto, 18 Jul. 2022); II. *S. epilobioides* (CHINA. Gansu: Dingxi, 01Aug. 2023); III. *S. medusa* (CHINA. Sichuan: Diebu, 17 Jul. 2023); IV. *S. dolichopoda* (CHINA. Shaanxi: Xi'an, 29 Jul. 2021); V. *S. elegans* (CHINA. Xinjiang: Bole, 28 Jul. 2022). I, II, and V were photographed by Y.S. Chen, and the remaining by L.S. Xu.

filtering loci with paralog warnings. The present study used targets from the most relevant reference genome, *Carthamus tinctorius*, which includes 475 targets, of which 474 (99.8%) were identified (147 (31.0%) of them with paralog warnings). After excluding paralogue sequences with PhyloPyPruner, 451 (95.1%) loci were retained. We obtained a similar proportion of loci (99.8% vs 99.3%) and loci with paralog warnings (31.0% vs 44.6%) using HybPiper. However, we obtained a greater percentage of loci excluding paralog sequences using PhyloPyPruner, compared to directly filtering loci with paralog warnings (95.1% vs 55.4%), which is mainly due to the difference dealing with paralogs. However, we obtained a lower proportion of parsimony informative sites (9.2% vs 31.3%), which is mainly due to sampling (genus vs subtribe levels). Although we used fewer loci (451 vs 544), our nuclear phylogeny was more robust (all four main clades are well-supported (Fig. 4(left)) vs three main clades not well-supported (Fig. 3E)). The targets used and species sampled may be contributing to such differences. Our phylogeny was similar to the tree of the coalescent approach in Herrando-Moraira et al. (2020). Specifically, ntclade1 included all sampled species of the basal clade, ntclade 2 included 9 sampled species of the most evolved clades, ntclades 3 and 4 included 14/17 sampled species of the other two clades. However, our phylogeny was different from the tree of generated by the concatenation approach in Herrando-Moraira et al. (2020), especially in ntclades 2, 3, and 4. ILS may play important roles in the different topologies (see below).

4.2. Cytoplasmic-nuclear discordance

The different nuclear phylogenetic topologies between the concatenation and coalescent approach (Fig. S1), the low frequencies of nuclear gene trees in nuclear phylogeny (Fig. S3), and the occurrence of simulated organellar gene trees in chloroplast phylogeny (Fig. 5(right)) indicate that ILS impacted the nuclear phylogeny of *Saussurea*. However, we also found conflicting plastid bipartition frequencies at or near zero in the 10,000 simulated organellar gene trees, especially within *Saussurea* (Fig. 5(right)). ILS alone is therefore insufficient to explain the observed cytoplasmic-nuclear incongruence recovered in these datasets. Thus, historical gene flow must be invoked.

The ancient hybridization between ntclades 1 and 2 may have influenced cytoplasmic-nuclear discordance within *Saussurea* (Fig. 5(left)). The ancestor of *Saussurea* had the original chloroplast genotype. A new type resulting from the ancient hybridization between ntclades 1 and 2 was shared by ntclades1 and 2; thus, these two clades were most related to each other in the plastome phylogeny. Hybridization accelerates evolution, therefore, ctclade 3 (shared the hybrid genotype) was more evolved than ctclades 1 and 2 (shared the original genotype) in the plastome phylogeny. Therefore, ancient hybridization is the most likely source of deep cytoplasmic-nuclear conflict in *Saussurea*. Moreover, this indicates that nuclear data is more suitable than chloroplast data for exploring infrageneric classification in *Saussurea*.

To further address this discordance and determine which data is more suitable for exploring the infrageneric classification of *Saussurea*, we incorporated cytological and morphological characters. First, we examined 18 morphological characters commonly used in the classification of *Saussurea*. We found that the nuclear phylogeny for *Saussurea* was supported by at least five morphological characters (Fig. 6), whereas the chloroplast phylogeny lacked support. For example, characteristics such as highly branched from the lower part of stems and phyllaries with colored apices distinguished ntclade 1 from the other clades (Fig. 6A and D); highly branched on the upper part of the stems distinguished ntclade 2 from the other clades (Fig. 6A); uppermost leaves densely woolly on both surfaces and capitula arranged in clustered secondary heads distinguished ntclade 3 from ntclade 4 (Fig. 6B and C). Furthermore, almost all species with chromosome base numbers recorded in ntclades 1 and 2 shared 12—14 chromosomes, whereas ntclades 3 and 4 shared 15—18 (Fig. S11). We were unable to find cytological or morphological evidence to support the chloroplast phylogeny.

We believe that phylogenetic analysis based on nuclear data is better suited for exploring the infrageneric classification of *Saussurea*. Subsequent analyses of the phylogeny and infrageneric classification of *Saussurea* were primarily based on the nuclear phylogeny.

In addition to the factors we analyzed above, different sequences employed, unequal numbers of COS (250—358) and chloroplast sequences (26—189) used, insufficient sampling and other unpredictable reasons may also play important roles in the cytoplasmic-nuclear discordance that require further study.

4.3. Phylogeny

Numerous researchers have made substantial efforts to enhance the accuracy of *Saussurea*'s molecular phylogeny, progressing from Sanger-based to NGS-based data (Raab-Straube, 2003; Kita et al., 2004; Wang et al., 2009, 2013; Xu et al., 2019; Zhang et al., 2021b). In comparison to prior Sanger-based data, our study demonstrates enhanced stability in the topology and support values. Previous studies encountered issues with two or four clades lacking strong support (Fig. 3), whereas our research achieved a more precise nuclear phylogeny and identified four clades using two different methods (Fig. 4). When compared to chloroplast data (NGS-based sequencing, Xu et al., 2019), our data provided more reliable information and stronger support from various lines of evidence (see above). Moreover, compared to nuclear data from NGS-based studies with limited sample sizes of *Saussurea* (Herrando-Moraira et al., 2020, with 39 samples from 6 sections, and Zhang et al., 2021c, with 5 samples from 4 sections), our study improved sample representation by including 103 species across 13 sections (see taxon sampling).

4.4. Diagnostic characters of each clade

Morphological characters with complex evolutionary histories can be misleading and often lead to artificially assigned relationships, especially in rapidly evolving groups (Wang et al., 2009). A total of 13/18 morphological character states checked originated independently more than once (Table 2; Figs. S5—S10), and thus cannot serve as diagnostic characters for each clade. For example, the membranous uppermost leaves have been previously considered as a diagnostic character for *S.* subg. *Amphilaena*. However, our findings showed that this trait had independently evolved at least three times (Fig. 6B), and the members of this subgenus did not cluster together in the phylogeny. Similar patterns were observed for rosettes in *S.* sect. *Lagurostemon* (Fig. S6) and conspicuous phyllary appendages used to identify *S.* subg. *Theodorea*. In contrast, characters with simpler evolutionary histories (independently evolved once) can serve as diagnostic characters and help to distinguish clades from each other. Uppermost leaves densely woolly on both surfaces (Fig. 6B) and capitula arranged in clustered secondary heads (Fig. 6C) were found to distinguish ntclade 3 from other clades; highly branched from the lower part of stems and phyllaries with colored apices distinguished ntclade 1 (Fig. 6A and D); stems highly branched on upper parts distinguished ntclade 2 (Fig. 6A); capitula arranged in complex corymbiform synflorescence (compound corymbiform or corymbiform synflorescence) and highly branched stems were observed in ntclades 1 and 2,

whereas capitula arranged in simple synflorescence (solitary or simple corymbiform synflorescence) and not highly branched stems (stemless, short, solitary, or rarely branched) occurred in ntclades 3 and 4.

4.5. Subgenera

There are two sections in *Saussurea* subg. *Theodorea* as per Lipschitz (1979): *S.* sect. *Maritimae* and *S.* sect. *Theodorea*. All samples belonging to these two sections clustered into ntclade 1, but did not form a monophyletic group, instead mixing with samples belonging to *S.* subg. *Saussurea* sect. *Laguranthera* (Fig. 4(left)). Stems highly branched from the lower part and phyllaries with colored apices distinguished ntclade 1 from the other three clades (Fig.6A and D). The type species of *S.* subg. *Theodorea*, *S. amara* (L.) DC. was present in ntclade 1. Thus, we propose to treat species with the morphological characters described above as members of *S.* subg. *Theodorea*. Ntclade 1 included all samples of *S.* sect. *Maritimae*, *S.* sect. *Theodorea*, and *S.* sect. *Laguranthera*. We propose to treat these three sections as members of *S.* subg. *Theodorea*.

There are ten sections in *Saussurea* subg. *Saussurea* as per Lipschitz (1979). Four of these sections were later removed from *Saussurea* or treated as synonyms based on molecular and morphological data (Raab-Straube, 2003; Shi and Raab-Straube, 2011; Wang et al., 2013; Herrando-Moraira et al., 2020), leaving six remaining sections (see Fig. 2: *S.* sect. *Gymnocline*, *S.* sect. *Laguranthera*, *S.* sect. *Lagurostemon*, *S.* sect. *Rosulascentes*, *S.* sect. *Saussurea*, and *S.* sect. *Strictae*). Samples from the remaining six sections did not cluster together in our phylogenetic tree (Fig. 4(left)). Stems highly branched on upper parts distinguished ntclade 2 from the other clades (Fig. 6A). In addition, phyllaries lacking colored apices distinguish this clade from ntclade 1, whereas the presence of compound corymbiform synflorescence distinguished it from ntclades 3 and 4 (Fig. 6C and D). The type species of *S.* subg. *Saussurea* (*S. alpina* (L.) DC.) was found in ntclade 2. Thus, we propose to treat species with the morphological characters described above as members of *S.* subg. *Saussurea*. Ntclade 2 included all species of *S.* sect. *Rosulascentes* and the type species of *S.* sect. *Saussurea* (*S. alpina*). We, therefore, propose to treat these two sections as members of *S.* subg. *Saussurea* and to remove the other four sections from *S.* subg. *Saussurea*.

There are three sections in *Saussurea* subg. *Eriocoryne* as per Lipschitz (1979): *S.* sect. *Cincta*, *S.* sect. *Eriocoryne*, and *S.* sect. *Pseudoeriocoryne*. Samples from these three sections did not cluster together in our phylogenetic tree (Fig. 4(left)). All nine samples from *S.* sect. *Eriocoryne* and one sample from *S.* sect. *Pseudoeriocoryne* (*S. wellbyi* Hemsl.) were found in ntclade 3, all two samples from *S.* sect. *Cincta* (*S. alpina* Franch. and *S. uliginosa* Hand.-Mazz.) and two samples from *S.* sect. *Pseudoeriocoryne* (*S. kingii* J.R. Drumm. ex C.E.C. Fisch. and *S. stella* Maxim.) were found in ntclade 4. These phylogenetic results agreed partly with adjustments made by Chen (2015), who moved some species of *S.* sect. *Cincta* (including *S. uliginosa*, *S. fistulosa* J. Anthony, *S. delavayi*, and *S. bijiangensis* Y.L. Chen ex B.Q. Xu, N.H. Xia & G. Hao) and some species from *S.* sect. *Pseudoeriocoryne* out of *S.* subg. *Eriocoryne*. The type species of *S.* subg. *Eriocoryne* (*S.aussurea gossipiphora* D. Don) was included in ntclade 3, and all samples in ntclade 3 can be distinguished from the other three clades using the following traits: 1) uppermost leaves densely woolly on both surfaces and 2) capitula arranged in clustered secondary heads (Fig. 6B and C). Thus, we propose to treat species with the morphological characters described above as members of *S.* subg. *Eriocoryne* with only one section, *S.* sect. *Eriocoryne*.

There are two sections in *Saussurea* subg. *Amphilaena* as per Lipschitz (1979): *S.* sect. *Amphilaena* and *S.* sect. *Pseudoamphilaena*. All samples from these two sections belonged to ntclade 4 but did not cluster together (Fig. 4(left)). Morphological characters associated with ntclade 4 included not highly branched stems, phyllaries lacking colored apices, uppermost leaves not densely woolly on both surfaces, and capitula arranged in simple corymbiform synflorescence, or solitary (Fig. 6). The type species of *S.* subg. *Amphilaena* (*S. obvallata* (DC.) Sch. Bip.) was found in ntclade 4. Thus, we propose to treat species with the morphological characters described above as members of *S.* subg. *Amphilaena*. Ntclade 4 included type species of *S.* subg. *Amphilaena* sect. *Amphilaena* (*S. obvallata*) and sect. *Pseudoamphilaena* (*S. longifolia* Franch.), *S.* subg. *Eriocoryne* sect. *Pseudoeriocoryne* (*S. stella*), and *S.* subg. *Saussurea* sect. *Cincta* (*S. delavayi*), sect. *Gymnocline* (*S. ciliaris* Franch.) and sect. *Strictae* (*S. taraxacifolia*), as well as all samples of *S.* subg. *Saussurea* sect. *Lagurostemon*. We, therefore, propose to treat these seven sections as members of *S.* subg. *Amphilaena*. This subgenus was the most complex within all the four *Saussurea* clades, displaying great variations in morphology, habitat, and maximum elevational range (200–6200 m).

4.6. Revised infrageneric classification scheme of Saussurea

The commonly accepted classification of *Saussurea* proposed by Lipschitz (1979) is highly artificial both at the subgenus and section levels, as many of the morphological characters on which this scheme is based have been shown to have evolved multiple times (Raab-Straube, 2003; Kita et al., 2004; Wang et al., 2009, 2013; Xu et al., 2019; Zhang et al., 2021b). In this study, we reset the existing 13 sections into four subgenera (Fig. 7; *S.* subg. *Amphilaena*, *S.* subg. *Eriocoryne, S.* subg. *Saussurea,* and *S.* subg. *Theodorea*), mostly following how type species were places in our nuclear phylogeny based on the following key.

1a. Stems usually highly branched from the lower part; phyllaries or at least inner rows with colored apices (light green to purple) ……..Subg. *Theodorea*
1b. Stems branched on the upper part or not branched; phyllaries without colored apices.
2a. Stems usually highly branched on the upper parts; capitula usually arranged in compound corymbiform synflorescence, rarely solitary …… ……..Subg. *Saussurea*
2b. Stems usually not highly branched; capitula arranged in clustered secondary heads, simple corymbiform synflorescence or solitary.
3a. Capitula clustered in secondary heads, surrounded by or half-surrounded by densely woolly subtending leaves ……... Subg. *Eriocoryne*
3b. Capitula usually arranged in simple corymbiform synflorescence or solitary, not surrounded by or half-surrounded by densely woolly subtending leaves …… Subg. *Amphilaena*

Saussurea subg. ***Theodorea*** Cass. in Bull. Sci. Soc. Philom. Paris 1818: 168. Type: *S. amara* (L.) DC.

Herbs, biennial, perennial, or subshrubs. Stems usually tall, highly branched from the lower part. Capitula usually many to numerous, arranged in a compound corymbiform synflorescence. Involucres campanulate, cylindrical, or tubular. Phyllaries or at least inner rows with colored apices (light green to purple). Corolla light pink to purple, usually exserted from involucres. Pappus in 2 rows, inner bristles plumose, outer bristles scabrid.

This subgenus corresponds to ntclade 1, as revealed in this study (Figs. 4 and 7), and includes three sections: *Saussurea* sect. *Laguranthera*, *S.* sect. *Maritimae*, and *S.* sect. *Theodorea*. The species of this

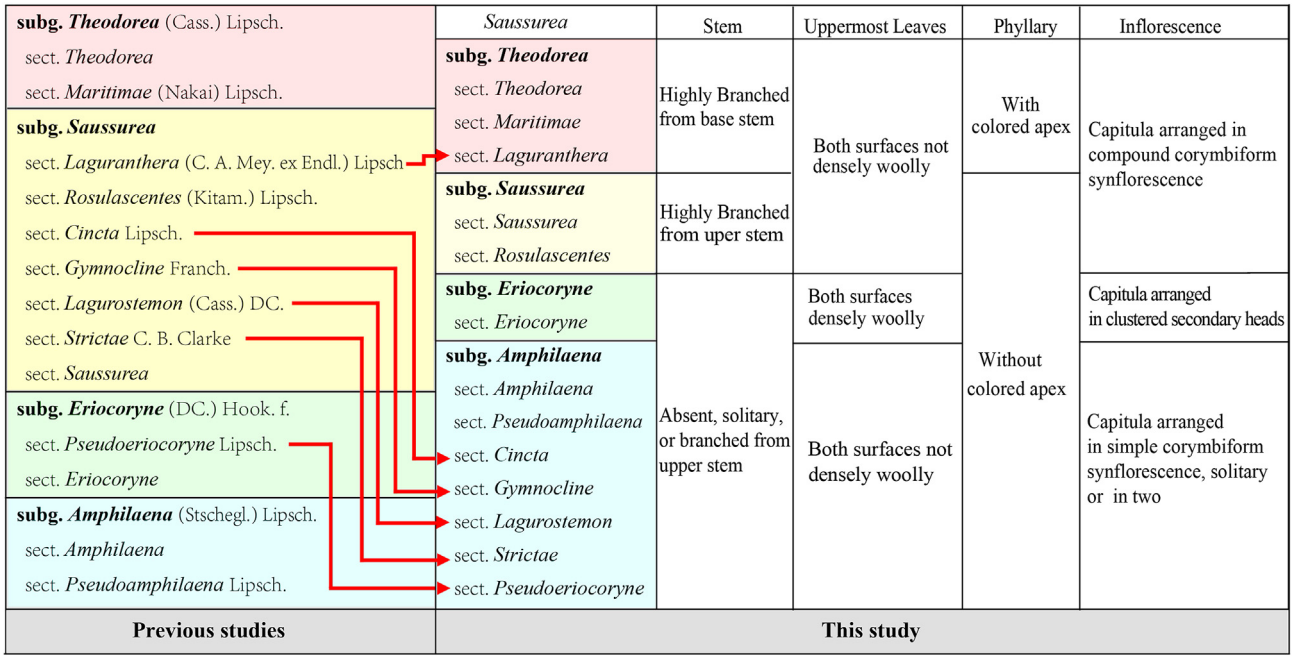


Fig. 7. A new infrageneric classification for *Saussurea*. The main changes in the present study are marked with red arrows.

subgenus are predominantly distributed across the steppes of Eurasia (Fig. 8).

Saussurea subg. ***Saussurea.*** Type: *S. alpina* (L.) DC.

Herbs or subshrubs, perennial. Stems usually highly branched on the upper part of the stems, rarely not branched. Capitula usually arranged in a compound corymbiform synflorescence, rarely one or two. Involucres campanulate, ovoid, cylindrical, tubular, or narrowly tubular. Phyllaries usually green, brown, or dark at apex. Corolla pink to purple, sometimes not obvious. Pappus dirty white, straw-colored, or yellowish brown, in 2 rows, inner bristles plumose, outer bristles scabrid.

This subgenus corresponds to ntclade 2, as revealed in this study (Figs. 4 and 7), and consists of two sections: *Saussurea* sect. *Rosulascentes* (Kitam.) Lipsch. and *S.* sect. *Saussurea*. Species of this subgenus are distributed in alpine grasslands in Europe, North America, north and midwestern Asia, and in forests in northeastern to southern China, Japan, the Korean peninsula, and the Russian Far East (Fig. 8).

Saussurea subg. ***Eriocoryne*** (Wall. ex DC.) Hook. f., in Fl. Brit. India 3: 376. 1881. ≡ *Aplotaxis* sect. *Eriocoryne* Wall. ex DC., Prodr. 6: 541. 1838. Type: *S. gossipiphora* D. Don.

Herbs, perennial, monocarpic or polycarpic. Stems usually not highly branched, erect, usually hollow, distally usually club-shaped and inflated. Capitula many, usually clustered in secondary heads at the center of leaf rosette or terminal on stem, enclosed or subtended by uppermost cauline leaves. Involucres usually campanulate, or cylindrical. Phyllaries usually green, brown, dark, or subhyaline at apex. Corolla white to dark purple, usually not much

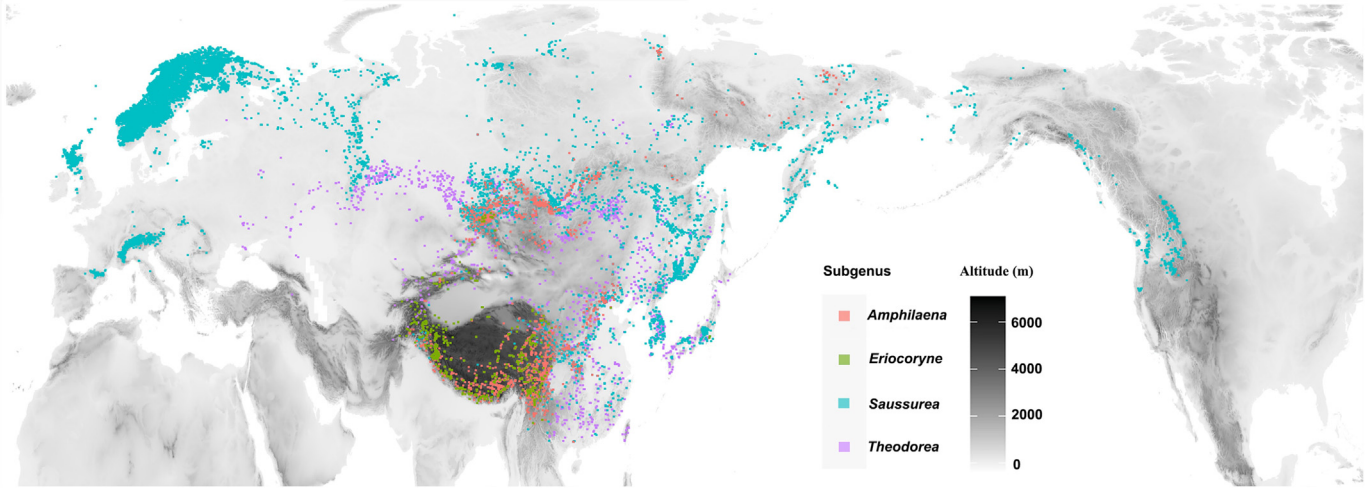


Fig. 8. Distributions of four subgenera of *Saussurea*. The locations of *Saussurea* species were mainly obtained from the Global Biodiversity Information Facility (GBIF, https://doi.org/10.15468/dl.avtup5) and Chinese Virtual Herbarium (CVH, https://www.cvh.ac.cn/), and modified (Table S4).

out of involucre. Pappus dirty white, straw-colored, light to dark brown, gray, or blackish, usually in 2 rows, outer bristles usually scabrid, sometimes absent.

This subgenus corresponds to ntclade 3, as revealed in this study (Figs. 4 and 7), and includes only one section: *Saussurea* sect. *Eriocoryne*. Species of this subgenus are primarily distributed within alpine alluvial fans or alpine scree vegetation on the Qinghai-Tibet Plateau (Fig. 8).

Saussurea subg. ***Amphilaena*** (Stschegl.) Lipsch. in Trudy Moskovsk. Obshch. Isp. Prir. 3: 182. 1960. ≡ *S.* sect. *Amphilaena* Stschegl. in Bull. Soc. Nat. Imp. Naturalistes Moscou 21(3): 244. 1848. Type: *S. obvallata* (DC.) Sch. Bip.

Herbs or subshrubs. Stems from developed to nearly absent, usually not highly branched or branched on the upper part. Capitula usually arranged in a simple corymbiform synflorescence or solitary. Involucres campanulate, globose, ovoid, cylindrical, or tubular. Phyllaries usually green, brown, or dark at apex. Corolla white to dark purple, usually slightly exserted from involucres. Pappus white to blackish, usually in 2 rows, outer bristles usually scabrid or rarely plumose, sometimes absent.

This subgenus corresponds to ntclade 4, as revealed in this study (Figs. 4 and 7), and includes seven sections: *Saussurea* sect. *Amphilaena*, *S.* sect. *Pseudoamphilaena*, *S.* sect. *Cincta*, *S.* sect. *Gymnocline*, *S.* sect. *Lagurostemon*, *S.* sect. *Strictae*, and *S.* sect. *Pseudoeriocoryne*. These species are broadly distributed across most habitats of the Qinghai-Tibet Plateau and adjacent areas (Fig. 8).

4.7. Relationships and evolution of the four subgenera

The relationships between the four subgenera are depicted in the nuclear phylogeny and were further supported by morphological findings. The ancestor of *Saussurea* subg. *Theodorea* exhibited traits similar to the ancestor of *Saussurea* (Fig. 6), which aligns with the basal position of *S.* subg. *Theodorea* in the nuclear phylogeny. *S.* subg. *Amphilaena*, *S.* subg. *Eriocoryne*, and *S.* subg. *Saussurea* are closely related and share similar traits, particularly the absence of colored apices on their phyllaries. Of these three subgenera, *S.* subg. *Saussurea* is the basal group, with its ancestor featuring highly branched stems and compound corymbs, reminiscent of the common ancestor of all *Saussurea* species. Relationships between these four subgenera were further substantiated by the evolution of various characteristics. Ancestral groups exhibited more primitive character states compared to more evolved groups. The evolution of *Saussurea* stems transitioned from complex to simple, with the ancestors of *S.* subg. *Theodorea* and *S.* subg. *Saussurea* (more basal groups), having highly branched stems, whereas *S.* subg. *Amphilaena* and *S.* subg. *Eriocoryne* (more evolved groups) have single stems or stemless. Similarly, the evolution of the number of capitula shifted from many to few, with the ancestors of the *S.* subg. *Theodorea*, *S.* subg. *Saussurea*, and *S.* subg. *Eriocoryne* (more basal groups) displaying various synflorescences composed of many capitula, whereas the ancestor of *S.* subg. *Amphilaena* (more evolved groups) featured a simple corymbiform synflorescence with a few capitula.

4.8. Sections

In our nuclear phylogeny, nearly no section of *Saussurea* was monophyletic (Fig. 4), and some species of a single section were distributed in different ntclades (subgenus), e.g., *Saussurea* sect. *Saussurea* and *S.* sect. *Pseudoeriocoryne*. This highlights the need for taxonomic revision within these sections. While our study aimed to revise the subgeneric classification of *Saussurea*, the limitations in our sampling scheme prevented us from addressing infrasectional classification issues. Consequently, we primarily relocated sections based on the position of each section's type species. The taxonomic revision of sections within *Saussurea* still necessitates comprehensive sampling across the genus, given the large number of *Saussurea* species and the complex evolutionary history of the genus, which makes achieving greater phylogenetic resolution a challenging endeavor.

5. Conclusions

We present the first well-resolved nuclear phylogeny of *Saussurea*, which consists of four main clades supported by morphological data. Our analyses show that ancient hybridization is the most likely source of deep cytoplasmic-nuclear conflict in *Saussurea*, and a phylogeny based on nuclear data is more suitable for exploring the infrageneric classification of *Saussurea* than one based on chloroplast data. We used our nuclear phylogeny and morphological characters to propose a revised infrageneric classification of *Saussurea*, which includes four subgenera and 13 sections.

CRediT authorship contribution statement

Liansheng Xu: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Zhuqiu Song:** Writing – review & editing, Resources. **Tian Li:** Resources, Investigation. **Zichao Jin:** Resources, Investigation. **Buyun Zhang:** Resources, Investigation. **Siyi Du:** Investigation. **Shuyuan Liao:** Investigation. **Xingjie Zhong:** Investigation. **Yousheng Chen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

The sequenced data have been deposited in the National Center for biotechnology information (NCBI), with public accession at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1107990.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Wei Wang, Ming Kang, Baosheng Wang, Chen Ren, Hongrui Zhang, and Long Wang for their helpful suggestions and comments on the manuscript. We thank Zaigang Wang, Zaiqiang Wang, and Liping Zou for participating in the fieldwork and material collections. We thank the editors and the reviewers for their valuable advice and comments that greatly improved our manuscript. We would like to thank Joseph Elliot, Emily Drummond, and Daniel Petticord for their assistance with the English language and grammatical editing of the manuscript. This work was supported by the National Natural Science Foundation of China (grant no. 32200172) and Guangzhou Science and technology project (grant no. 2023A04J0715).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pld.2024.10.003.

References

Andrés-Sánchez, S., Galbany-Casals, M., Bergmeier, E., et al., 2014. Systematic significance and evolutionary dynamics of the achene twin hairs in *Filago* (Asteraceae, Gnaphalieae) and related genera: further evidence of morphological homoplasy. Plant Syst. Evol. 301, 1653–1668. https://doi.org/10.1007/s00606-014-1185-7.

Bankevich, A., Nurk, S., Antipov, D., et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. https://doi.org/10.7287/peerj.preprints.1355v1.

Candolle, A. P. de, 1810. Observations sur les plantes Composées ou syngenèses. Premier mémoire. Sur les Composées et les Cinarocéphales en général. Ann. Mus. Hist. Nat. 16, 135–158.

Chen, Y.S., 2015. Asteraceae II: *Saussurea*. In: Hong, D.Y. (Ed.), Flora of Pan-Himalaya. Science Press, Beijing, pp. 1–135.

Cheon, K.S., Kim, H.J., Han, J.S., et al., 2016. The complete chloroplast genome sequence of *Saussurea chabyoungsanica* (Asteraceae), an endemic to Korea. Conserv. Genet. Resour. 9, 51–53. https://doi.org/10.1007/s12686-016-0617-9.

Fer, T., Schmickl, R.E., 2018. HybPhyloMaker: target enrichment data analysis from raw reads to species trees. Evol. Bioinf. Online 14, 1176934317742613. https://doi.org/10.1177/1176934317742613.

Folk, R.A., Mandel, J.R., Freudenstein, J.V., 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst. Biol. 66, 320–337. https://doi.org/10.1093/sysbio/syw083.

Guindon, S., Dufayard, J.F., Lefort, V., et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hatami, E., Jones, K.E., Kilian, N., 2022. New insights into the relationships within subtribe Scorzonerinae (Cichorieae, Asteraceae) using hybrid capture phylogenomics (Hyb-Seq). Front. Plant Sci. 13, 851716. https://doi.org/10.3389/fpls.2022.851716.

Herrando-Moraira, S., Calleja, J.A., Carnicero-Campmany, P., et al., 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). Mol. Phylogenet. Evol. 128, 69–87. https://doi.org/10.1016/j.ympev.2018.07.012.

Herrando-Moraira, S., Calleja, J.A., Galbany-Casals, M., et al., 2019. Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: a new subtribal classification and a temporal diversification framework. Mol. Phylogenet. Evol. 137, 313–332. https://doi.org/10.1016/j.ympev.2019.05.001.

Herrando-Moraira, S., Calleja, J.A., Chen, Y.S., et al., 2020. Generic boundaries in subtribe Saussureinae (Compositae: Cardueae): insights from Hyb-Seq data. Taxon 69, 694–714. https://doi.org/10.1002/tax.12314.

Johnson, M.G., Gardner, E.M., Liu, Y., et al., 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4, 1600016. https://doi.org/10.3732/apps.1600016.

Kearse, M., Moir, R., Wilson, A., et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kita, Y., Fujikawa, K., Ito, M., et al., 2004. Molecular phylogenetic analyses and systematics of the genus *Saussurea* and related genera (Asteraceae, Cardueae). Taxon 53, 679–690. https://doi.org/10.2307/4135443.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lipschitz, S., 1979. Genus *Saussurea* DC. (Asteraceae). Nauka, Leningrad.

Mandel, J.R., Dikow, R.B., Funk, V.A., et al., 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. Appl. Plant Sci. 2, 1300085. https://doi.org/10.3732/apps.1300085.

Mandel, J.R., Dikow, R.B., Siniscalchi, C.M., et al., 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. Proc. Natl. Acad. Sci. U.S.A. 116, 14083–14088. https://doi.org/10.1016/j.ympev.2011.12.007.

Minh, B.Q., Nguyen, M.A., Haeseler, A.V., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024.

Minh, B.Q., Schmidt, H.A., Chernomor, O., et al., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Moore-Pollard, E.R., Jones, D.S., Mandel, J.R., 2024. Compositae-ParaLoss-1272: a complementary sunflower-specific probe set reduces paralogs in phylogenomic analyses of complex systems. Appl. Plant Sci. 12, e11568. https://doi.org/10.1002/aps3.11568.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490. https://doi.org/10.1371/journal.pone.0009490.

Raab-Straube, E. von, 2003. 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., *Lipschitziella* and *Frolovia*. Willdenowia 33, 379–402. https://doi.org/10.1093/botlinnean/boy081.

Raab-Straube, E. von, 2017. Taxonomic Revision of *Saussurea* Subgenus *Amphilaena* (Compositae, Cardueae). Botanic Garden and Botanical Museum, Berlin.

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.

Shi, Z., Raab-Straube, E. von, 2011. *Saussurea* group. In: Wu, Z.Y., Raven, P.H., Hong, D.Y. (Eds.), Flora of China. Science Press, Beijing, pp. 54-149.

Siniscalchi, C.M., Loeuille, B., Funk, V.A., et al., 2019. Phylogenomics yields new insight into relationships within Vernonieae (Asteraceae). Front. Plant Sci. 10, 1224. https://doi.org/10.3389/fpls.2019.01224.

Smith, S.A., Moore, M.J., Brown, J.W., et al., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 150. https://doi.org/10.1186/s12862-015-0423-0.

Solis-Lemus, C., Ane, C., 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genetics 12, e1005896. https://doi.org/10.1371/journal.pgen.1005896.

Solis-Lemus, C., Bastide, P., Ane, C., 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34, 3292–3298. https://doi.org/10.1093/molbev/msx235.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312. https://doi.org/10.1093/bioinformatics/btu033. –1213.

Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.

Wang, Y.J., Milnes, R., Susanna, A., et al., 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai-Tibetan plateau. Bot. J. Linn. Soc. 97, 893–903. https://doi.org/10.1111/j.1095-8312.2009.01225.x.

Wang, Y.J., Raab-Straube, E. von, Susanna, A., et al., 2013. *Shangwua* (Compositae), a new genus from the Qinghai-Tibetan Plateau and Himalayas. Taxon 62, 984–996. https://doi.org/10.12705/625.19.

Xu, L.S., Chen, Y.S., 2021. Phylogeny, origin, and dispersal of *Dubyaea* (Asteraceae) based on Hyb-Seq data. Mol. Phylogenet. Evol. 164, 107289. https://doi.org/10.1016/j.ympev.2021.107289.

Xu, L.S., Herrando-Moraira, S., Susanna, A., et al., 2019. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. Mol. Phylogenet. Evol. 141, 106613. https://doi.org/10.1016/j.ympev.2019.106613.

Xu, L.S., Song, Z.Q., Liao, S.Y., et al., 2024. *Qineryangia*, a new genus from the Hengduan Mountains and new insights into the phylogeny of the subtribe Crepidinae (Cichorieae, Asteraceae). J. Syst. Evol. 13066. https://doi.org/10.1111/jse.13066.

Yu, Y., Blair, C., He, X.J., 2020. Rasp 4: ancestral state reconstruction tool for multiple genes and characters. Mol. Biol. Evol. 37, 604–606. https://doi.org/10.1093/molbev/msz257.

Yuan, Q., Bi, Y.C., Chen, Y.S., 2015. *Diplazoptilon* (Asteraceae) is merged with *Saussurea* based on evidence from morphology and molecular systematics. Phytotaxa 236, 53–61. https://doi.org/10.11646/phytotaxa.236.1.4.

Zhang, C., Rabiee, M., Sayyari, E., et al., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang, C.F., Huang, C.H., Liu, M., et al., 2021a. Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. J. Integr. Plant Biol. 63, 1273–1293. https://doi.org/10.1111/jipb.13078.

Zhang, X., Landis, J.B., Sun, Y.X., et al., 2021b. Macroevolutionary pattern of *Saussurea* (Asteraceae) provides insights into the drivers of radiating diversification. Proc. R. Soc. B-Biol. Sci. 288, 20211575. https://doi.org/10.1098/rspb.2021.1575.

Zhang, X., Sun, Y.X., Landis, J.B., et al., 2021c. Transcriptomes of *Saussurea* (Asteraceae) provide insights into high-altitude adaptation. Plants 10, 1715. https://doi.org/10.3390/plants10081715.

Zhang, Y.Z., Chen, J.G., Sun, H., 2021d. Alpine speciation and morphological innovations: revelations from a species-rich genus in the Northern Hemisphere. AoB Plants 13, plab018. https://doi.org/10.1093/aobpla/plab018.

Zhou, B.F., Yuan, S., Crowl, A.A., et al., 2022. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat. Commun. 13, 1320. https://doi.org/10.1038/s41467-022-28917-1.