

RESEARCH

Open Access



Phylogenomics, plastome structure and species identification in *Mahonia* (Berberidaceae)

Ruchang Tong^{1,2}, Chaoxia Gui^{1,2}, Yu Zhang^{1,2}, Na Su^{1,2}, Xiaoqi Hou³, Meng Liu⁴, Zhaoping Yang⁵, Bing Kang^{1,2}, Zhaoyang Chang^{1,2}, Florian Jabbour⁶ and Liang Zhao^{1,2*}

Abstract

Background: Elucidating the phylogenetic relationships within species-rich genera is essential but challenging, especially when lineages are assumed to have been going through radiation events. *Mahonia* Nutt. (Berberidaceae) is a genus with cosmopolitan distribution, comprising approximately 100 species, two of which are known as *Caulis Mahoniae* (*M. bealei* and *M. fortunei*) with crucial pharmacological significance in Chinese herbal medicine. *Mahonia* is a taxonomically challenging genus, and intrageneric phylogenetic relationships still need to be explored using genome data. Universal DNA barcodes and floral morphological attributes have limited discriminatory power in *Mahonia*.

Results: We sequenced 17 representative plastomes and integrated three published plastome data together to conduct comparative and phylogenetic analyses. We found that *Mahonia* and *Berberis* share a large IR expansion (~12 kb), which is recognized as a typical character of Berberideae. Repeated sequences are revealed in the species of *Mahonia*, which are valuable for further population genetic studies. Using a comparative plastome analysis, we determined eight hypervariable regions whose discriminative power is comparable to that of the whole plastid genomes. The incongruence of the ITS and the plastome tree topologies may be ascribed to ancestral hybridization events and/or to incomplete lineage sorting. In addition, we suggest that leaf epidermal characters could help to distinguish closely related species in *Mahonia*.

Conclusions: We propose an integrative approach combining special barcodes and micromorphological traits to circumscribe *Mahonia* species. The results cast a new light on the development of an integrative method for accurate species circumscription and provide abundant genetic resources for further research on *Mahonia*.

Keywords: Berberidaceae, Leaf epidermal characters, *Mahonia*, Plastome, Special barcodes

Background

The Berberidaceae (Ranunculales) is an early-diverging eudicot plant family comprising 19 genera, including the newly proposed *Alloerberis* P.H. Raven ex C.C. Yu &

K.F. Chung and *Moranothamnus* P.H. Raven ex C.C. Yu & K.F. Chung [1, 2]. The 680+ Berberidaceae species are predominantly distributed in northern temperate zones extending to Andean South America and northern Africa [3–5]. The barberry family is traditionally known for its morphological diversity, intercontinental discontinuous distribution and medicinal utilization [5, 6].

Mahonia Nutt. is the second largest genus in Berberidaceae, comprising about 100 species [7]. However, the precise number of *Mahonia* species remains ambiguous,

*Correspondence: biology_zhaoliang@126.com

¹ College of Life Sciences, Northwest A&F University, Yangling 712100, Shaanxi, China

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



as 33 species were synonymized in the Flora of China [4]. Morphologically, the species of *Mahonia* are easily distinguished from other angiosperm species by their evergreen odd-pinnately compound leaves, their leaflets margins with spinose dentation, and their spineless stems [4]. The species of *Mahonia* are distributed in East Asia and Western North America [1, 8], making the genus an emblematic example of a biogeographic disjunction.

Besides, a few species of *Mahonia* are endemic to Europe, North Africa and South America [8]. Many species of *Mahonia* are broadly cultivated for horticulture (Fig. 1) and for their pharmacological properties [4, 7, 9]. For instance, the stems of *M. bealei* (Fortune) Carrière and *M. fortunei* (Lindl.) Fedde are known as *Caulis Mahoniae* with highly anti-inflammatory properties [10] and are included in the Chinese pharmacopoeia [9].



Fig. 1 Morphological diversity of *Mahonia* species. **A B** *M. bealei*. **A** fruits. **B** racemose inflorescence. **D M. hancockiana**. **C E H** *M. fortunei*. **C** a single flower. **E** compound leaves. **H** racemose inflorescence. **F M. fordii**. **G M. eurybracteata** subsp. *ganpinensis*. **I M. napaulensis**. **J M. bordinieri**. **K M. shenii**. **L M. breviracema**. **M M. oiwakensis**. **N M. duclouxiana**

The position of the genus *Mahonia* remains intractable and has been discussed for a long time [1, 11, 12]. Traditionally, morphological and molecular evidence indicated that *Mahonia* was paraphyletic, with *Mahonia* sect. *Horridae* being sister to *Berberis* L. sensu stricto [5, 12, 13]. Although several authors held a view that *Mahonia* should be subsumed under a broadly defined *Berberis* (*Berberis* sensu lato) [12–14], a great majority of researchers advocate for a paraphyletic *Mahonia* because of its compound leaves that are distinct from the simple leaves of *Berberis* [7, 11, 12, 15]. Yu and Chung [1] proposed a new classification that divided *Berberis* s.l. into four monophyletic clades and establishing four new genera (*Berberis* ≡ *Berberis* s.s., *Mahonia* ≡ core *Mahonia*, *Alloberberis* ≡ *Mahonia* sect. *Horridae*, and *Moranothamnus* ≡ *Berberis* *claireae*). This taxonomic treatment not only maintained the universally acceptable perception of *Berberis* but also resolved *Mahonia* as a monophyletic genus, which has been widely accepted by botanists in the fields of phylogenomics [16], taxonomy [17], and biogeography [8].

The genus *Mahonia* is taxonomically and phylogenetically challenging, owing to its considerable species richness, to the rapid diversification events that have punctuated its evolutionary history, and to the high similarity in the morphology of reproductive structures that hinders the easy and accurate identification of species [7, 8, 18]. Within *Mahonia*, floral organs are usually invariable in number and are arranged in whorls, and all *Mahonia* species bear similar yellow flowers and blue-black globose berries [4, 7].

Previously, a series of comparative plastome analyses found that a large IR expansion of over 12 kb occurred in *M. bealei*, which is unusual in plastome evolution [19, 20]. Using ITS (nuclear ribosomal DNA) and four DNA fragments (including the genes *accD*, *ndhF*, *rbcL* and the intergenic spacer *trnH-psbA*) of the plastid genome, Yu and Chung [1] proposed a new classification that recognized *Mahonia* as a distinctive monophyletic genus with strong support. The oldest reliable fossil records of *Mahonia* were collected in East Asia and used for biogeographic analyses [18]. On the basis of molecular dating estimates and comparison of leaf morphologies of extant *Mahonia* species, the researchers inferred that the genus *Mahonia* originated in Western North America and subsequently dispersed into East Asia. Notably, after the migration to East Asia, the genus *Mahonia* probably underwent a radiation event, leading to the current Eastern Asian biodiversity center [8, 18].

Phylogenetic analyses based on molecular datasets provide reasonable phylogenetic hypotheses [21–24]. However, it is disputable that using just a single line of evidence is sufficient to delimit species boundaries [25].

Therefore, multiple evidence (such as including plastome datasets, morphological traits, ecological traits) should be applied to modern systematics [26], in particular with respect to recently diverging lineages, such as *Mahonia*. Complete plastome data have proven to be effective in resolving phylogenetic relationships at a wide range of taxonomic levels [16]. The phylogenetic incongruence between the plastome tree and the nrDNA tree indicated that frequent hybridization has occurred between *Mahonia* and *Berberis* [2].

Characters of the apex of petals, the length of pedicels and bracts, the number of leaflets and spinose dentations, were used as critical morphological traits for discriminating the species of *Mahonia* [3, 4]. Given the stability and uniformity of micromorphological traits among taxa, researchers have undertaken a series of investigations to provide more evidence for the classification of *Mahonia* (e.g., floral anatomy [27]; seed micromorphology [28]; carpel micromorphology [29, 30]; sepal morphology [31]). Structural characters of leaf epidermis are usually constant and more accessible; they have been proven to possess great systematic significance in some complex taxa [32–34]. However, studies of leaf epidermal micromorphology with respect to the genus *Mahonia* is far from sufficient.

Here, we sequenced 17 representative complete plastomes of the genus *Mahonia*, and used 13 plastomes from GenBank to conduct comparative and phylogenetic analyses. Morphological and micromorphological traits of different species of *Mahonia* were recorded. We combined the evidence from the molecular and morphological data to resolve the phylogenetic relationships in *Mahonia*. Our goals are to 1) reconstruct phylogenetic relationships within *Mahonia* using nuclear internal transcribed spacer (ITS) and plastid genome sequences; 2) describe and interpret the plastome structure and evolution of *Mahonia*; 3) explore an integrative method for better distinguishing among *Mahonia* species.

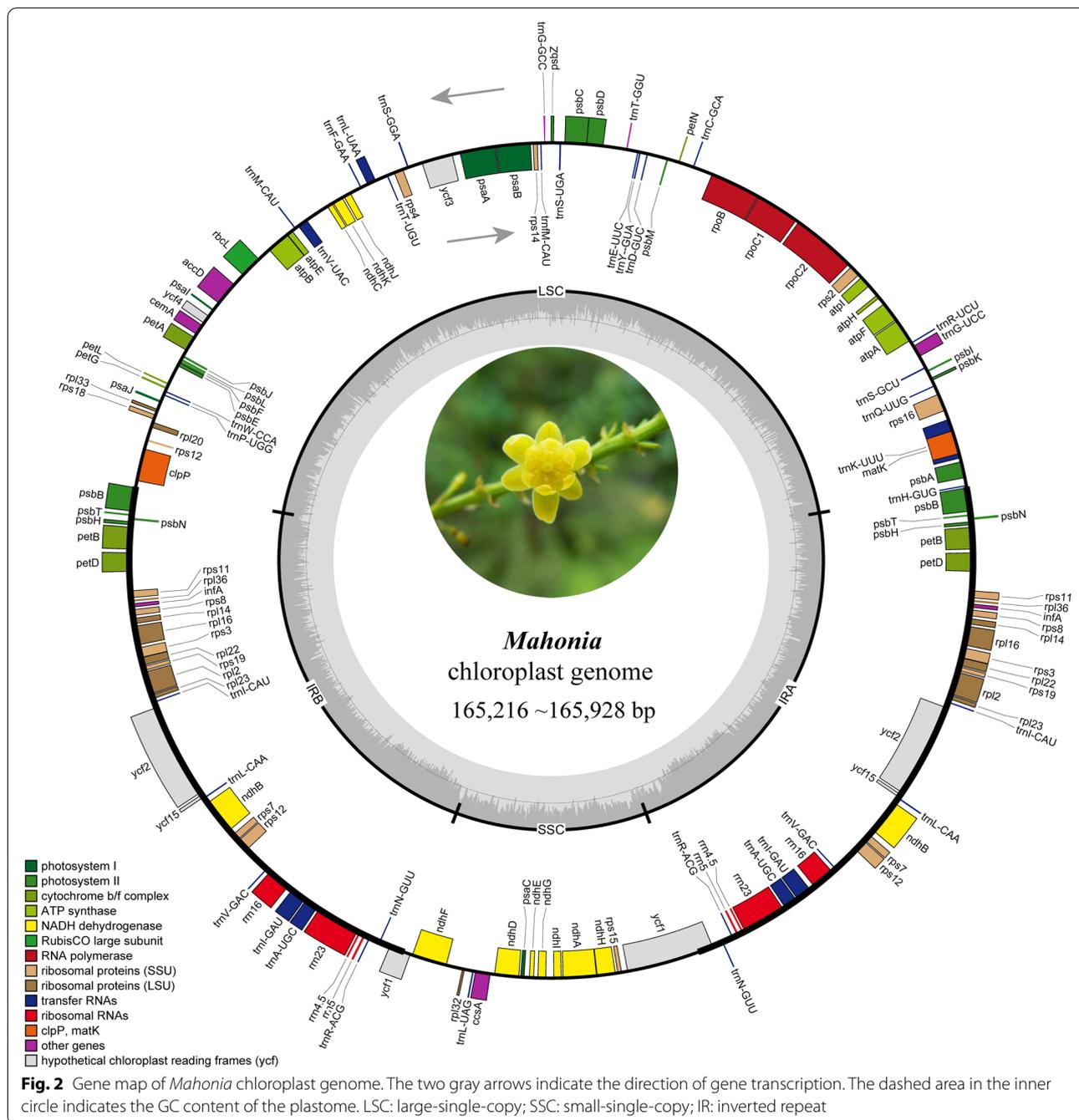
Results

Plastome features of *Mahonia*

The number of raw paired-end reads for each plastome ranges from 15,790,898 (*M. brevirostrata* Y.S. Wang & P.G. Xiao RC611 [MZ158268]) to 25,347,900 (*M. duclouxiana* Gagnep. RC602 [MZ086770]) (Table 1). The assembled plastid genomes range from 165,216 bp (*M. napaulensis* DC. RC603 [MZ158275]) to 165,928 bp (*M. shenii* Chun RC609 [MZ158280]) in length with 38% to 38.1% genomic GC contents overall. The GC contents in inverted regions (IR, 41.1%–41.2%) are much higher than in the large single copy (LSC) and in the small single copy (SSC) regions (Table 1). The typical quadripartite configuration of these plastid genomes consisted of

Table 1 Summary of 17 complete chloroplast genomes of *Mahonia*

Species	Voucher number	Number of reads	Average depth of coverage (×)	Total size (bp)	Total GC%	LSC		SSC		IR	
						size (bp)	GC%	size (bp)	GC%	size (bp)	GC%
<i>M. bealei</i>	RC601	18,802,778	147.3	165,713	38%	73,516	36.4%	18,773	32.2%	36,712	41.2%
<i>M. duclouxiana</i>	RC602	25,347,900	1099.9	165,384	38.1%	73,477	36.4%	18,563	32.4%	36,672	41.2%
<i>M. napaulensis</i>	RC603	16,952,962	275.3	165,216	38.1%	73,313	36.5%	18,563	32.4%	36,670	41.2%
<i>M. cardiophylla</i>	RC604	18,211,456	537.9	165,559	38.1%	73,353	36.4%	18,802	32.3%	36,702	41.2%
<i>M. nitens</i>	RC605	19,474,194	1456.3	165,692	38.1%	73,703	36.4%	18,655	32.4%	36,667	41.2%
<i>M. gracilipes</i>	RC606	17,893,824	575.3	165,909	38%	73,451	36.4%	18,730	32.4%	36,864	41.1%
<i>M. polyodonta</i>	RC607	17,570,436	779.1	165,526	38.1%	73,280	36.5%	18,804	32.4%	36,721	41.2%
<i>M. bodinieri</i>	RC608	16,171,718	395.3	165,697	38%	73,594	36.4%	18,675	32.4%	36,714	41.1%
<i>M. shenii</i>	RC609	17,804,842	778.8	165,928	38.1%	73,528	36.4%	18,764	32.3%	36,818	41.1%
<i>M. oiwakensis</i>	RC610	22,462,532	80.4	165,267	38.1%	73,198	36.5%	18,721	32.4%	36,674	41.2%
<i>M. breviracema</i>	RC611	15,790,898	408.3	165,555	38.1%	73,365	36.4%	18,706	32.3%	36,742	41.1%
<i>M. fordii</i>	RC612	23,844,540	165.2	165,689	38%	73,558	36.4%	18,705	32.4%	36,713	41.1%
<i>M. hancockiana</i>	RC613	18,513,816	150.6	165,495	38.1%	73,483	36.4%	18,670	32.4%	36,671	41.2%
<i>M. eurybracteata</i>	RC614	15,910,170	389.3	165,612	38.1%	73,474	36.4%	18,734	32.4%	36,702	41.2%
<i>M. japonica</i>	RC615	24,350,910	310	165,711	38%	73,685	36.4%	18,676	32.4%	36,675	41.1%
<i>M. aquifolium</i>	RC616	19,912,290	37.4	165,689	38.1%	73,212	36.5%	18,873	32.4%	36,802	41.1%
<i>M. pinnata</i>	RC618	18,121,752	35.8	165,308	38.1%	73,313	36.5%	18,713	32.4%	36,641	41.2%



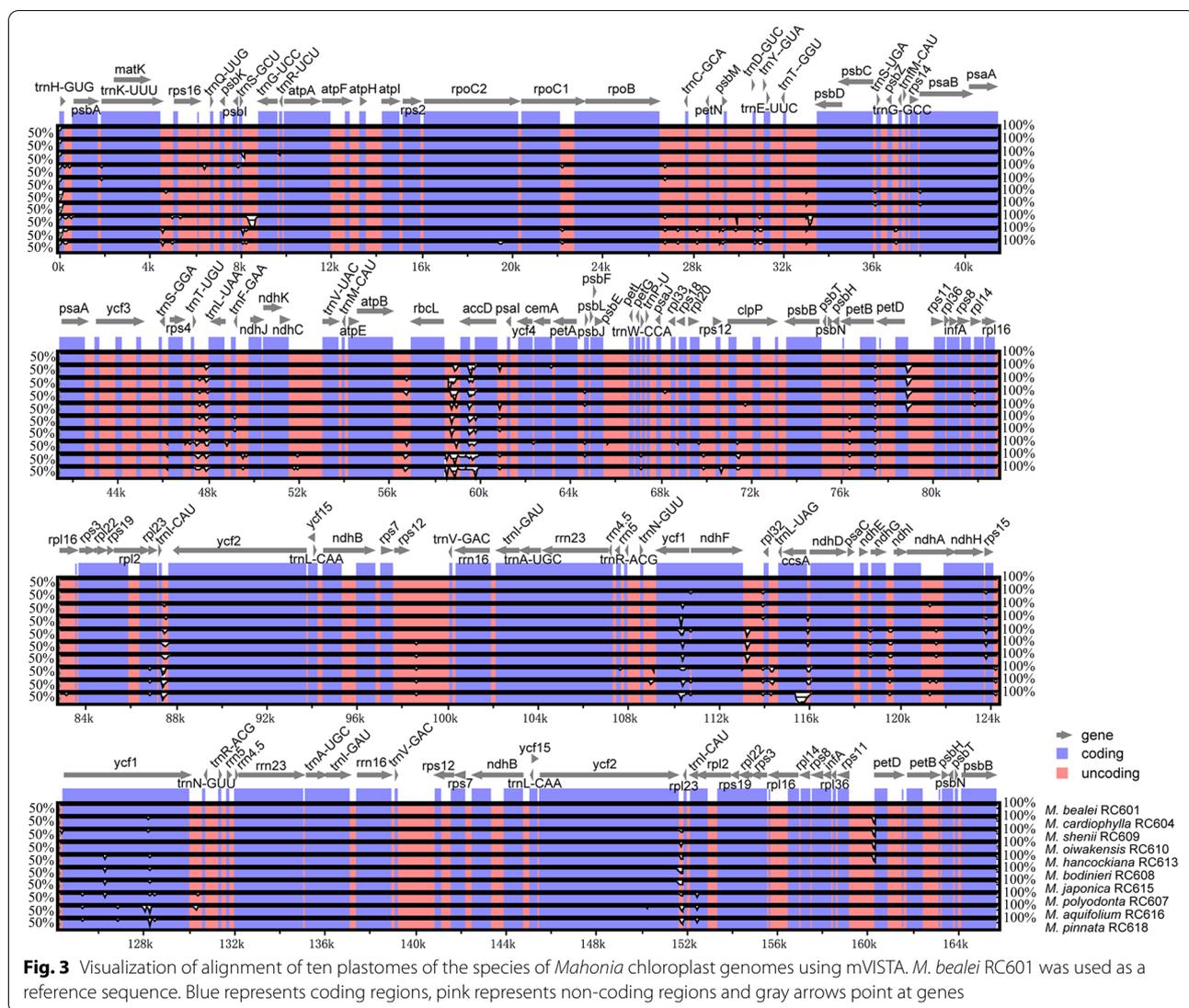
IR (36,641 bp–36,864 bp), which are separated by LSC (73,198 bp–73,703 bp) and SSC (18,563 bp–18,873 bp) regions (Fig. 2, Table 1).

The *Mahonia* species we sequenced encode 113 unique genes, 34 of which are duplicated in the IR. A total of 79 protein-coding, 30 transfer RNA (tRNA) and four ribosomal RNA (rRNA) genes are successfully predicted. Each IR copy contains 23 protein-coding, seven tRNA and four rRNA genes. In total, 147 genes are included in

the *Mahonia* plastid genomes we reconstructed (Table S1). There are 18 unique intron-containing genes in the plastid genomes. Sixteen genes (six tRNA and ten protein-coding genes) have a single intron, and the other two (*ycf3* and *clpP*) possess two introns.

Comparative plastid genome analyses

Using an annotated plastid genome (*Mahonia bealei* RC601 [MZ158266]) as reference, we plotted two graphs



for the overall sequence identity of ten *Mahonia* species and their outgroups using the program mVISTA (Figs. 3, S1). The results reveal that there are only slight variations within *Mahonia* plastid genomes. These variations are usually observed in the intergenic spacers (IGS) instead of coding-regions, which implies that coding regions are more conserved than non-coding regions (Fig. 3). The whole plastid genome of *M. bealei* RC601 [MZ158266] is also compared with those of *Berberis aristata* DC., *Ranzania japonica* (T. Itô ex Maxim.) T. Itô, *Gymnospermium kiangnanense* (P.L. Chiu) Loconte, *Leontice armeniaca* Boivin, *Caulophyllum robustum* Maxim. and *Nandina domestica* Thunb. However, the results reveal that there is a significant divergence in terms of sequence length, gene order and content among the genera related to *Mahonia* (Fig. S1). A large-scale IR expansion was found only in the plastomes of Berberideae, resulting in the additional

duplication of 15 genes compared with typical angiosperm plastomes (Figs. S1, S2). The plastome size of *M. bealei* is about 165 kb and harbored more genes than other genera. Due to incomplete duplication of the normal copy, the gene *ycf1* across the IRb-SSC boundary is truncated to ca. 1346 bp and recognized as a pseudogene ($\psi ycf1$). Out of the three exons of the trans-splicing gene *rps12*, two are duplicated in the IR. Gene rearrangement is not observed within *Mahonia* plastid genomes (Fig. S3). We compared the pairwise sequence distances and the number of nucleotide substitutions of 19 *Mahonia* species from 20 individuals. The highest level of pairwise sequence distance rate and the number of nucleotide substitutions (0.00546, 900 bp) is detected in the pair of *M. pinnata* and *M. japonica*. The lowest level (0, 0 bp) is observed between *M. bealei* RC601 [MZ158266] and *M. bealei* [MH795308] (Table S2).

Contraction and expansion of IR regions

We compared the IR-SC boundaries among seven plastid genomes from different genera of Berberidaceae, and showed that the contraction and expansion of IR varied among different genera of Berberidaceae. Visualizing these whole plastomes, we observed a large expansion of IR in *Mahonia bealei*, as well as in *Berberis aristata*. As a result, about 12 kb corresponding to 15 genes (including *rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14*, *rps8*, *infA*, *rpl36*, *rps11*, *petD*, *petB*, *psbH*, *psbN*, *psbT* and *psbB*) had suffered an additional duplication compared with the rest of the species we studied (Figs. S1, S2, Table S1). Thus, the IRb-LSC boundaries in these three species are located upstream of the *psbB* gene rather than located within the *rps19* gene, which is observed in the other four species we studied (*Caulophyllum robustum*, *Gymnospermium kiangnanense*, *Leontice armeniaca*, and *Nandina domestica*). The IRa-SSC and IRb-SSC boundaries are located within *ycf1* and *ycf1* pseudogenes ($\psi ycf1$), respectively. The *ndhF* genes, located downstream of the *ycf1* pseudogene, are 37 bp–540 bp away from the IRb-SSC boundaries. There are 7–77 bp from *trnH* genes to IRa-LSC boundaries (Fig. S2). In contrast, only slight shifts are observed in interspecies comparisons among ten *Mahonia* plastid genomes (Fig. 4).

Identification of hypervariable regions

Genome-wide sliding window analysis among 20 *Mahonia* individuals was performed in order to calculate the nucleotide diversity (Pi) values and identify the highly variable regions (mutational hotspots). The Pi values across the whole plastid genomes range from 0 to 0.06285 (mean = 0.00205), and the *accD* region exhibits the highest diversity level (Fig. 5). The eight most hypervariable regions (Pi > 0.008) were identified: five (*petN-psbM*, *ndhC-trnV*, *atpB-rbcL*, *accD*, *rpl20-clpP*) are located in the LSC, and the other three (*ycf1*, *ccsA-ndhD*, $\psi ycf1$) in the SSC. None was found in the IR (Fig. 5). The Pi values of the eight hypervariable regions we extracted range from 0.00311 to 0.0974 (Table 2).

Repeated sequence analyses

We used MISA to detect the simple sequence repeats (SSRs) among ten species of *Mahonia*. The number of SSRs in each *Mahonia* plastid genome varies from 81 in *M. aquifolium* to 94 in *M. cardiophylla* T.S. Ying & Boufford RC604 [MZ158269] and *M. shenii*. Within these SSRs, mononucleotides are the most abundant (86.1%), followed by hexanucleotides and then by dinucleotides and trinucleotides. In addition, tetranucleotides and pentanucleotides appear rarely in plastid genomes (Fig. 6A). The lengths of all the SSRs range from 10 to 28 bp, and a majority of the SSRs units possess 10 base pairs (Fig. 6B).

Most of the SSRs are distributed in LSC regions, and the SSRs located in SSC regions and IR are nearly equal in size (Fig. 6C).

Overall, a total of 208 tandem repeats were identified within the ten *Mahonia* plastid genomes. Each plastome contains 18 to 24 tandem repeats (Fig. 6D). We recorded 359 dispersed repeats in the ten plastid genomes of this genus. Each plastome includes 30 to 42 dispersed repeats. The forward repeats account for the largest proportion of dispersed repeats (59.6%), followed by palindromic repeats and then by reverse repeats. Moreover, complement repeats are often absent, except for *M. japonica* (Thunb.) DC. RC615 [MZ158274] and *M. hancockiana* Takeda RC613 [MZ158273] (Fig. 6D). The most common types of dispersed repeats range from 30 to 39 bp in length (Fig. 6E).

Phylogenetic analyses

In this study, nine alignment matrices were used to perform phylogenetic analyses using Bayesian inference (BI) and Maximum Likelihood (ML) method. The matrices consist of 20 ingroup accessions (*Mahonia*) and ten outgroup accessions. Notably, the ITS gene matrix includes only 28 samples, as the ITS sequences of *Leontice armeniaca* and *Ranzania japonica* are not available. The genus *Mahonia* is resolved as monophyletic and is sister to *Berberis* with strong support (bootstrap support (BS) \geq 99%, posterior probabilities (PP) \geq 0.99) across almost all trees (Figs. 7, S4, S5, S6). In the tree built using the complete plastid genome datasets, about 75 percent of the nodes are well supported (BS/PP = 99%/0.99). The phylogenetic trees exhibit that *Mahonia* species are grouped into four subclades. Subclade I comprises two species (*M. pinnata*, and *M. aquifolium*), which are both distributed in Western North America, while the species from the remaining subclades are native to East Asia. Subclade II contains a single species (*M. polyodonta* Fedde RC607 [MZ158279]). Subclade III consists of five species (*M. nitens* C.K. Schneid. RC605 [MZ158276], *M. fortunei* (Lindl.) Fedde [NC_042167], *M. japonica* (Thunb.) DC. RC615 [MZ158274], *M. fordii* C.K. Schneid. RC612 [MZ158271], and *M. bodinieri* Gagnep. RC608 [MZ158267]) with maximum PP support value (1.00). Regarding subclade III, the BI tree topology is not concordant with the ML tree topology. The remaining twelve individuals are gathered into subclade IV with high support values (BS/PP = 98%/1.00). Subclade I is the earliest-diverging lineage of *Mahonia*. Subclade IV is sister to subclade III, and together form a clade that is sister to subclade II (Fig. 7A).

To test the conflicting signals between plastomes and ITS sequences, both BI tree and ML tree based on ITS

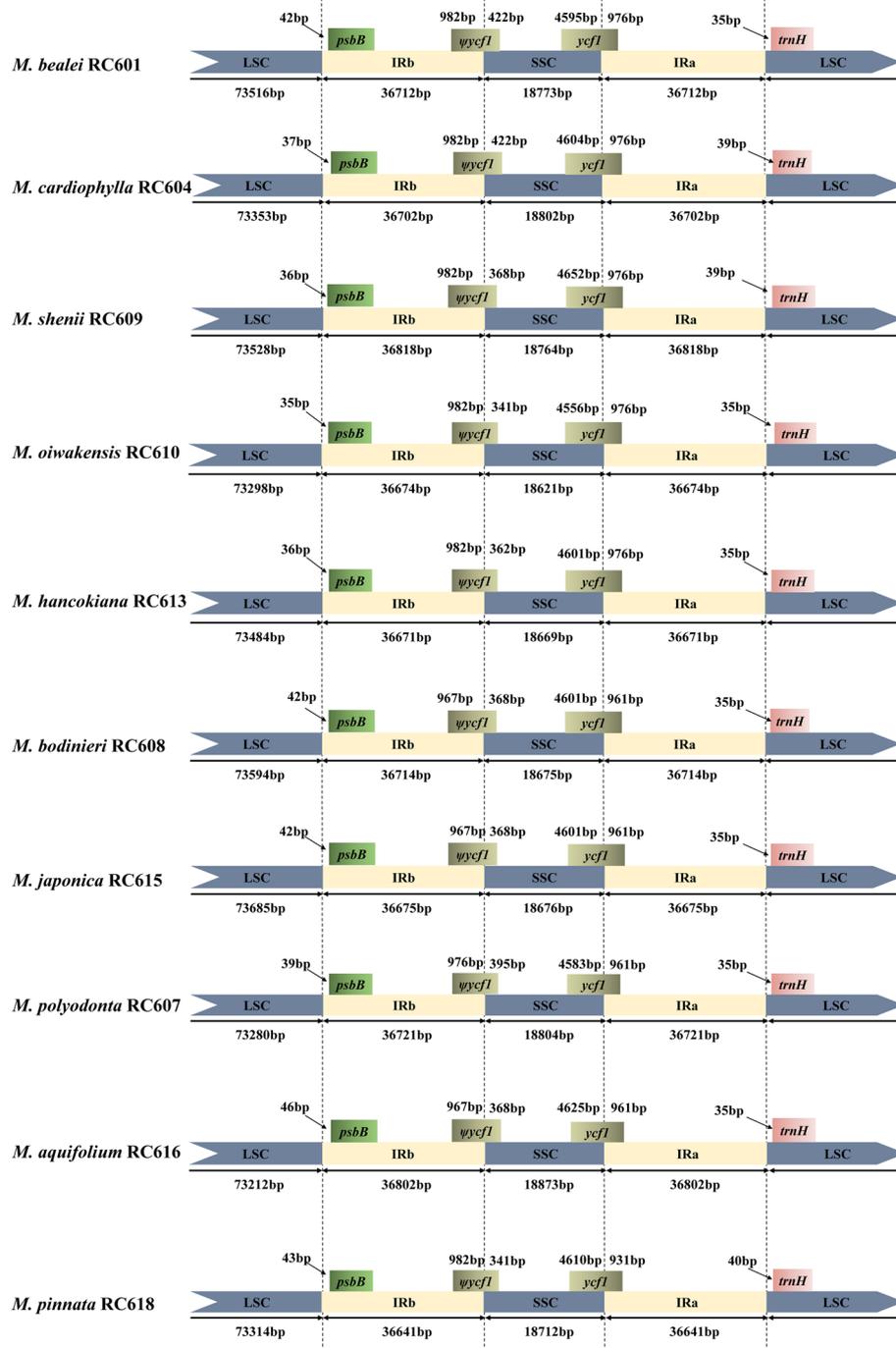


Fig. 4 Comparison of the LSC, IR and SSC boundary regions of ten plastomes of the species of *Mahonia*

datasets were generated and together compared with the trees based on plastomes (Fig. 7B). As shown in Fig. 7B, the genus *Mahonia* is also recovered as a clade with moderate support (BS/PP=0.96/85). Subclades I and II in the plastome tree are completely congruent with subclades A

and B in the ITS tree. The tree topologies outside subclades I and II are incongruent. Indeed, the tree based on ITS sequences possessed the highest number of polytomies and could not provide any valuable information to resolve the infrageneric relationships. However,

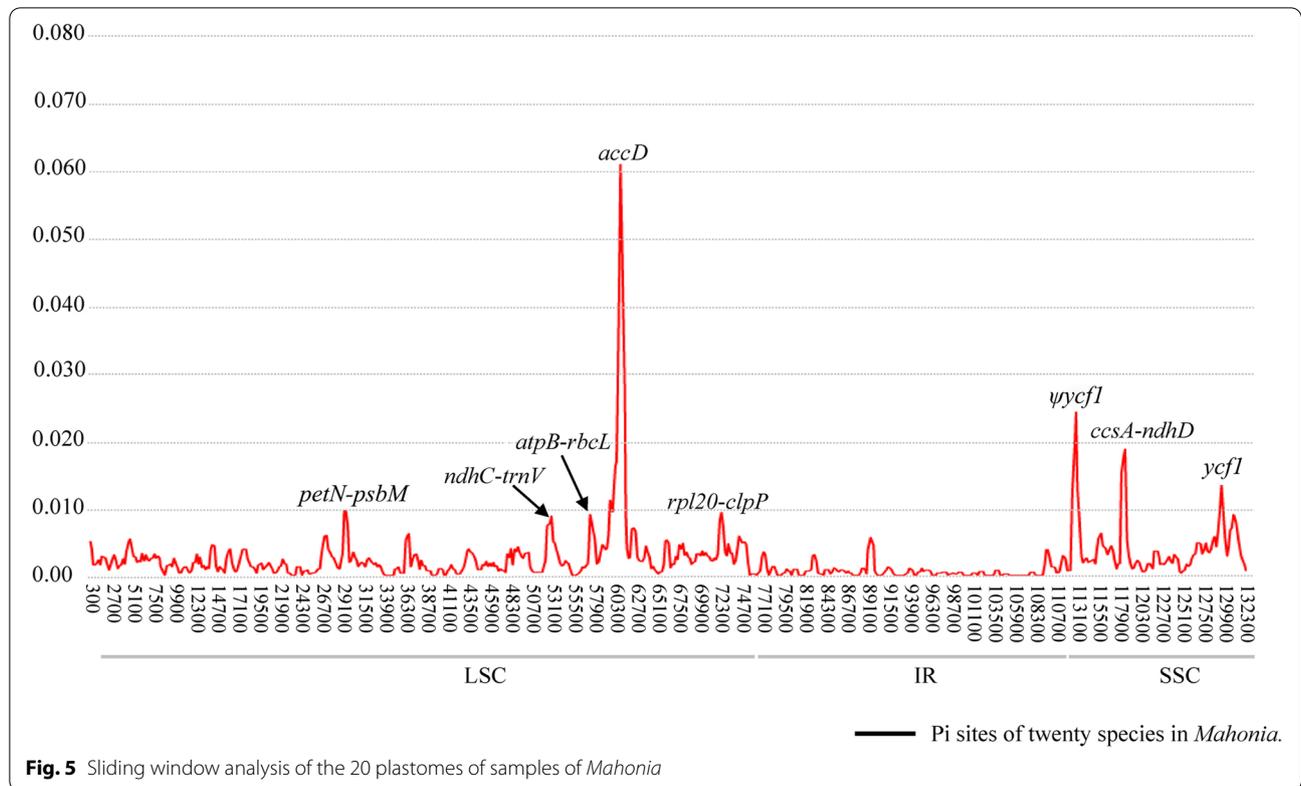


Table 2 Sequence characteristics of eight highly variable regions among 20 plastomes

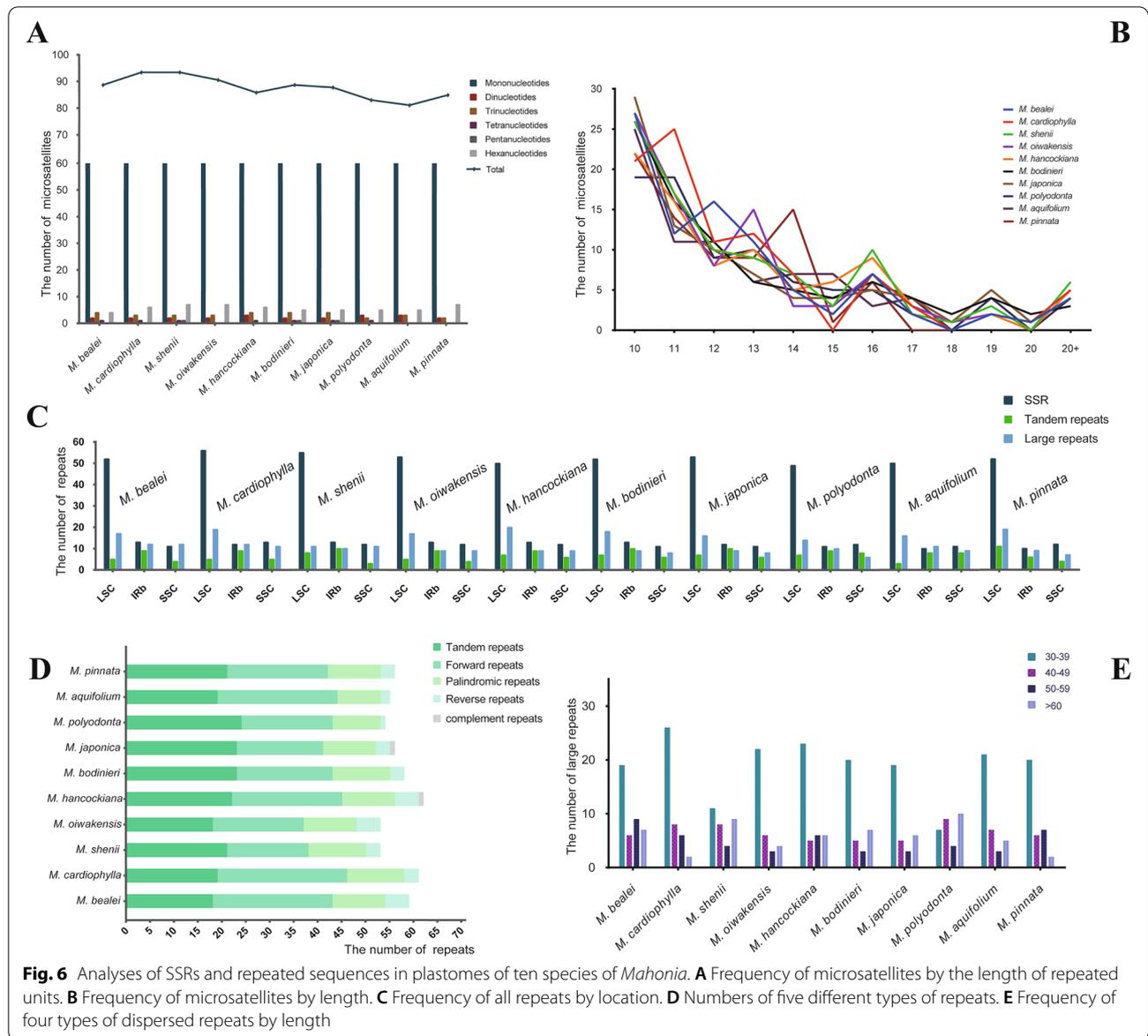
Region	Aligned length	Variable sites		Indels		Nucleotide diversity (Pi)
		No	%	No	Length range	
<i>petN-psbM</i>	661	27	4.08	24	1–9	0.0974
<i>ndhC-trnV</i>	1603	47	2.93	137	1–38	0.00580
<i>atpB-rbcL</i>	1134	11	0.970	405	1–171	0.00311
<i>accD</i>	1827	90	4.93	497	1–165	0.01391
<i>rpl20-clpP</i>	1429	41	2.87	55	1–13	0.00522
<i>ψycf1</i>	1425	43	3.02	118	1–54	0.00712
<i>ccsA-ndhD</i>	259	26	10.0	23	0–23	0.04237
<i>ycf1</i>	5820	114	1.96	352	1–78	0.00478

given that support values at internal nodes are much higher than the external, we can properly cluster the several subclades into subclades III and IV (in the plastome tree). Subclade III is largely identical to subclade G. Subclade IV gathers the remaining clades (C, D, F and H). Notably, the positions of *Mahonia fortunei* Chung 3342 [KX549421], *M. gracilipes* (Oliv.) Fedde RC606 [MZ158272] + *M. nitens* (subclade E) in the ITS tree are severely in conflict with the plastome tree.

It is noteworthy that the tree topology based on eight concatenated hypervariable regions is mainly identical to

the whole plastid genome tree. Whereas the phylogenetic relationships within subclade III and IV based on whole plastid genomes (Fig. 7A) are not fully consistent with the topology from the hotspots (Fig. 7C). The trees based on concatenated *rbcL*, *matK*, and *trnH-psbA*, have the lowest phylogenetic resolution (Fig. 7D). Furthermore, none of the phylogenetic reconstructions based on the concatenated *rbcL*, *matK*, and *trnH-psbA* datasets provides any evidence for the monophyly of *Mahonia* (Fig. 7D).

Based on the five datasets (coding, non-coding, LSC, SSC, and IR regions) extracted from the plastid genomes,



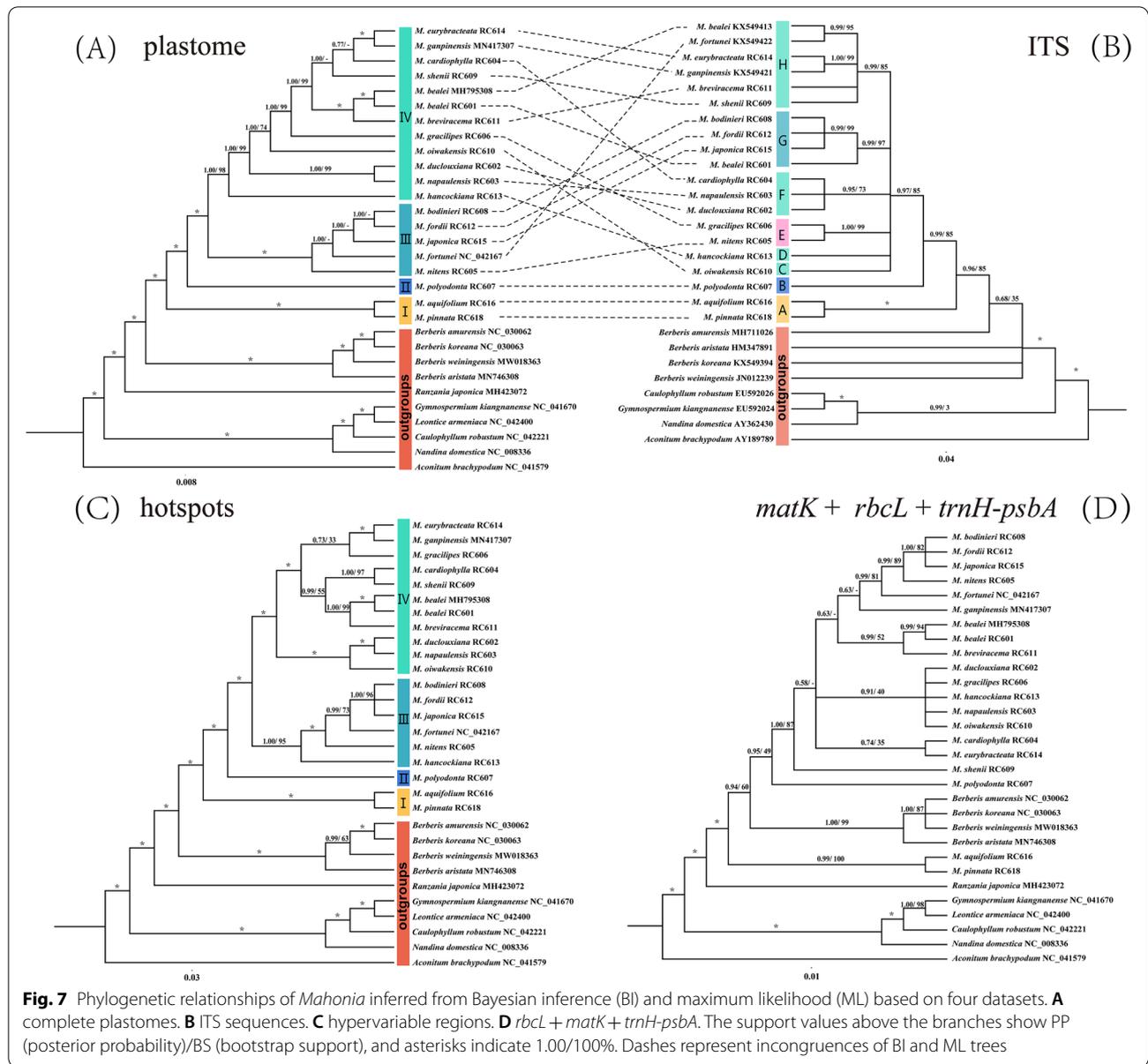
the overall topology is consistent with the topology retrieved from the complete plastid genome datasets. However, support values are high mostly at deep nodes (Fig. S4).

Leaf morphological and micromorphological characteristics

The leaves of *Mahonia* are odd-pinnately compound. The adaxial surfaces of mature leaves are glossy for most species (Fig. 8). The leaflets show substantial diversity with respect to the number and shape among different species. Margins of each leaflet are variously toothed with coarse or fine spined serrations (Figs. 9A1–F1, S7A1–F1, S8A1–F1).

The adaxial surfaces of epidermal cells are almost convex (Figs. 9A2–D2, S7A2, B2, D2), slightly convex (Figs. 9E2, F2, S7C2, E2, S8A2, B2). Fewer upper surfaces

are flat or nearly so (Figs. S7F2, S8C2). Seven species (*M. duclouxiana*, *M. cardiophylla*, *M. nitens*, *M. gracilipes*, *M. breviracema*, *M. eurybracteata* subsp. *ganpinensis*, and *M. pinnata*) show epicuticular waxes on the adaxial side of their leaves (Figs. S7D2–F2, S8C2–F2). On the abaxial surface of leaflets, cells with irregular shape and stomatal apparatus are found. The anticlinal walls of lower epidermal cells are either mostly inconspicuous or prominently sinuous, almost stellate in appearance (Figs. 9A3, C3, S7D3). Epidermal cells surrounding stomata are usually sunken, resulting in uneven lower epidermis of leaflets. Wax ornamentations in the form of strips is found on the abaxial surface of *M. hancockiana*, *M. breviracema* and *M. japonica* leaves (Figs. S7D3, F3, F4, S8B3, B4). All the leaves we observed are



hypostomatic (Figs. 9, S7, S8). The stomata are anomocytic (Figs. S7D₄, S8B₃–D₃, F₃), cyclocytic (Fig. 9D3) or actinocytic (Figs. 9A3–C3, S7A₄, B₄, E₄). In *M. bodinieri*, *M. polyodonta* and *M. nitens*, the abaxial surfaces of leaf epidermis are so flat that we could not detect the cell boundaries and determine the type of stomatal apparatus (Figs. 9E4, F4, S8E₄).

Discussion

Comparative plastome of *Mahonia*

In the vast majority of flowering plants, complete plastomes share a similar structure comprising a large inverted repeat (IR), a large single copy (LSC) and a small single

copy (SSC), respectively ~25 kb, ~87 kb and ~18 kb in length (e.g., *Diphylleia* Michaux, *Dysosma* Woodson, *Podophyllum* L., *Sinopodophyllum* Ying [(Berberidaceae)] [35]; *Maddenia* Hook. f. & Thoms, [36]; almost all genera of Styracaceae, [37]). The ebb and flow of IR are not unusual in evolutionary history [38]. IR recognition can display length divergence in angiosperm plastid genomes [39–41]. For instance, *Pelargonium transvaalense* R. Knuth possesses the largest known IR regions with ~88 kb in length [42]. More than 10 kb IR expansion was found in *Nicotiana acuminata* (Graham) Hook. (Solanaceae) [43]. In Trochodendrales, an IR expansion of about 4 kb was observed in both genera *Trochodendron*

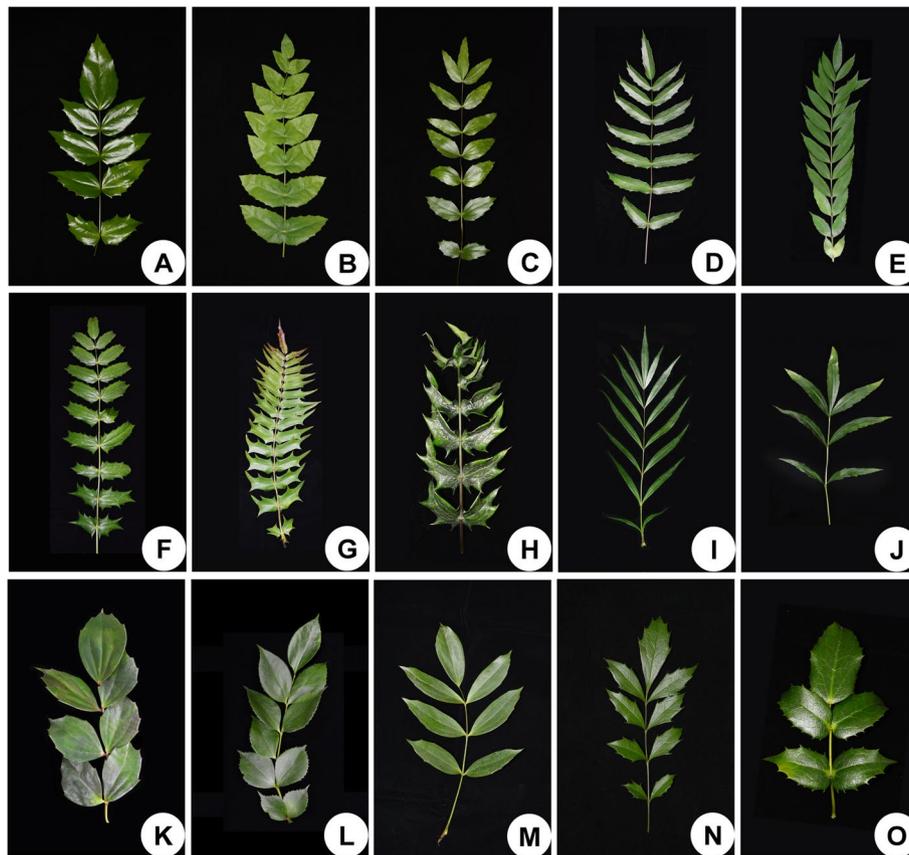


Fig. 8 Morphological variations of compound leaves in *Mahonia*. **A** *M. fordii*. **B** *M. napaulensis*. **C** *M. hancockiana*. **D** *M. duclouxiana*. **E** *M. eurybracteata*. **F** *M. bordineri*. **G** *M. oiwakensis*. **H** *M. cardiophylla*. **I** *M. eurybracteata* subsp. *ganpinensis*. **J** *M. fortunei*. **K** *M. nitenis*. **L** *M. gracilipes*. **M** *M. shenii*. **N** *M. brevircacema*. **O** *M. polyodonta*

Siebold & Zucc. and *Tetracentron* Oliv. [20, 44]. A possible mechanism for these large and small IR expansions is double-strand DNA break and combination, and gene conversion, respectively [38].

Based on a chloroplast restriction site mapping study, Kim and Jansen [19] proposed that there was a large-scale (ca. 10 kb) IR expansion in the plastomes of Berberideae (*Berberis* and *Mahonia*). Ma et al. [20] conducted a comparative plastid genome analysis among four species of Ranunculales, finding that the genome size of *Mahonia bealei* was about 4.9–9.7 kb larger than the other three species (*Nandina domestica* [Berberidaceae], *Megaleranthis saniculifolia* Ohwi and *Ranunculus macranthus* Scheele [Ranunculaceae]). They inferred that a large IR expansion is the main cause of the significant increase in genome size in *Mahonia bealei* [20]. A similar 10 kb IR expansion has been described in *Ranzania japonica* [45] (GenBank ID: MG234280), although this result was not supported in other studies [46]. The plastome structure of *Ranzania japonica* is controversial and deserves further investigation. We

present here the first comprehensive plastome analysis in *Mahonia* and show that there is a large IR expansion in the plastid genome of all species we investigated. Further intergeneric comparative plastome analyses have attested that a large-scale IR expansion was present in Berberideae [19, 46].

Although sequences in IR regions are commonly well conserved in comparison with single-copy (SC) regions, the IR-SC junctions are relatively variable. As shown in Fig. S2, neither large (>500 bp) expansions nor contractions are recorded in the plastome of *Mahonia* species, except within the junctions of IRb-LSC. The IRb regions deeply expand into LSC regions reaching 10 kb in length, resulting in the IRb-LSC boundaries being located upstream of the gene *psbB*. For many angiosperm plastomes, the IRb-LSC boundaries are located in the gene *rps19* [36, 37, 47]. Previous studies have concluded that these variations at IR-SC boundaries can provide more information for elucidating the evolutionary patterns of closely related species [48] and selecting potential phylogenetic molecular markers [49].

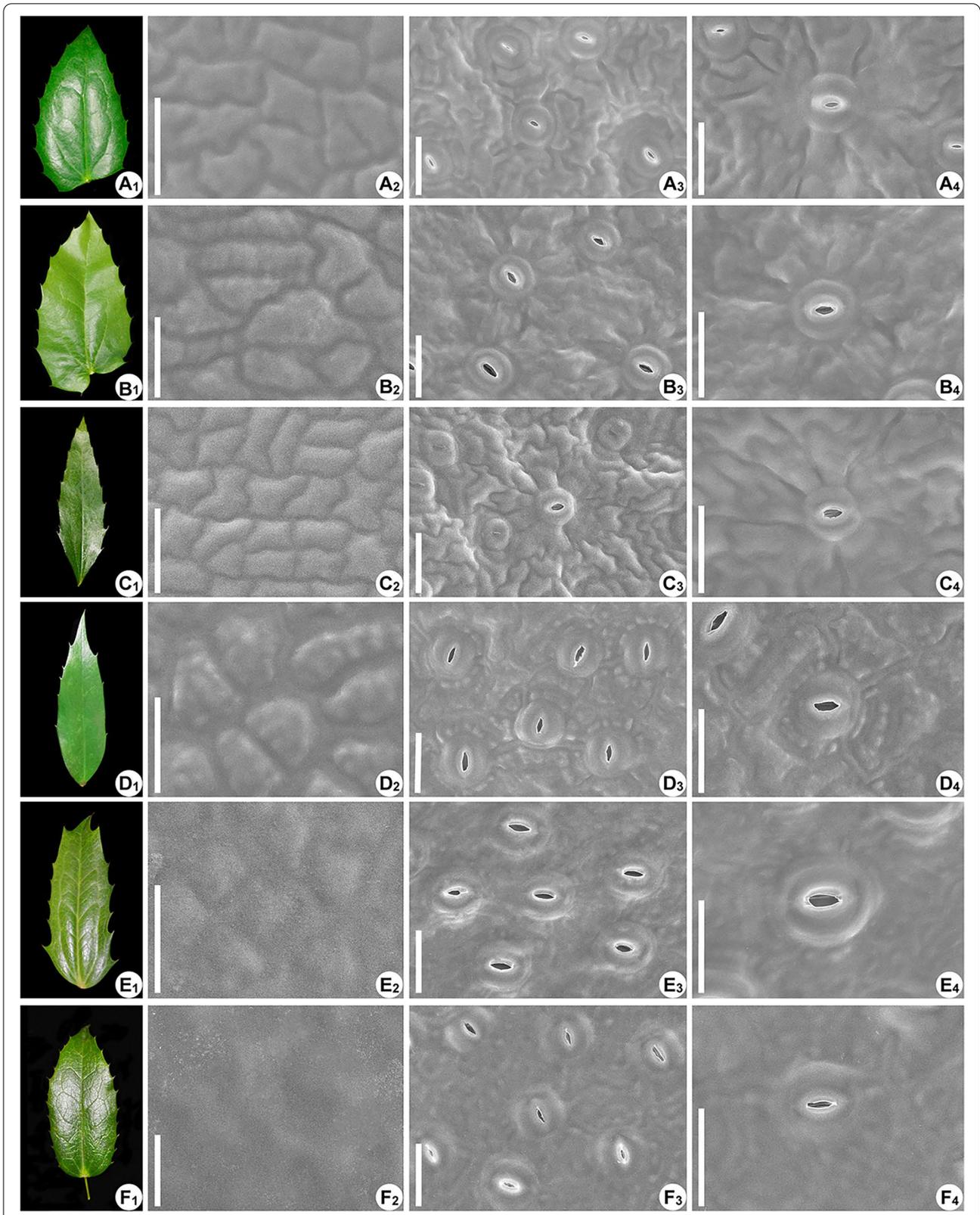


Fig. 9 Characteristics of leaflets and epidermal surface. **A₁-A₄** *M. bealei*. **B₁-B₄** *M. napaulensis*. **C₁-C₄** *M. fortunei*. **D₁-D₄** *M. eurybracteata*. **E₁-E₄** *M. bordinieri*. **F₁-F₄** *M. polyodonta*. The images show leaflets, adaxial leaves, abaxial leaves and magnifying stomatal apparatus on the abaxial surface in each row from the left to right, respectively

In *Mahonia*, as SSRs and dispersed repeats show abundant variations among different species, they could be developed into molecular markers in the future.

Phylogenetic analysis

In all phylogenetic trees we reconstructed (Figs. 7, S4, S5, S6), and except for the one based on the concatenation of the three common DNA barcodes (Fig. 7D), the monophyletic genus *Mahonia* was shown to be sister to *Berberis* with high support values, corroborating the results of previous studies [1, 2, 8]. Furthermore, in the plastome tree (Fig. 7A), the relationships among most clades are well resolved (PP=1.00) implying the great power of using complete plastomes to address intractable phylogenetic relationships. Plastid phylogenomics of the family Berberidaceae [16] and of its different subordinate taxa have been studied in depth, including Podophylloideae [35], *Epimedium* [6], *Berberis* and *Mahonia* [2, 50]. Given the phylogeny topologies with strong support values, these results demonstrated the power of plastid phylogenomics for improving plastome-based phylogeny, investigating early-divergent events, and conducting taxonomic and plastome evolution analyses. Hsieh et al. [2] used 93 plastomes representing all 19 genera of Berberidaceae to resolve the long-standing disputable taxonomic issues of Berberidaceae. They also paid attention to the phylogeny and plastome structure of the tribe Berberideae, corroborating the considerable topological discordance between nrDNA and plastomes. Our phylogenomic analysis of the genus *Mahonia* based on more representative taxon sampling than previous studies, provides valuable genetic resources and improves our understanding of the relationships among phylogenetically challenging groups.

Determining the discordance between the topologies generated based on plastome and nuclear DNA has profound significance for clarifying the evolutionary events and evaluating the current phylogenetic frameworks generated by plastome datasets [51, 52]. In our study, we find significant discordance throughout the topologies of plastome tree and ITS tree in particular at deep nodes (Figs. 7A, B). Some nodes unexpectedly clustered with strong support values (e.g., subclades G and E in the ITS tree). These conflicts may be ascribed to ancestral hybridization events and/or incomplete lineage sorting [18, 52]. Exploring the source of discordant relationships is challenging especially in the hyper-diverse taxa, since radiations create opportunities for the evolutionary processes abovementioned [53]. Besides, we observed that the distribution of morphological characters is more congruent with the nuclear-based topology than with the plastome-based topology. Focusing on the ITS tree, the leaflets in subclade H exhibit a continuous morphological

transition from linear to elliptic with several fine spined serrations at margins (Figs. 9C1, D1, S7D₁, E₁, S8A₁). The leaflets of the two species *M. gracilipes* and *M. nitens* in subclade E show highly similar shapes, which are distinctive from the leaf shapes of all species we investigated (Figs. S8E₁, F₁). These results indicate that nuclear datasets may have broader implications for morphological character evolution, hybridization and/or incomplete lineage sorting.

Standard DNA barcodes have been shown to often lack sufficient variable characters and thus often fail to discriminate species among many lineages [50, 54–56]. Chen et al. [8] added two hypervariable plastid genes (*accD* and *ndhF*) and combined them with ITS, *rbcL*, *matK*, *trnH-psbA*. The combined ITS and plastid DNA dataset was used to conduct phylogenetic analyses, revealing that this kind of dataset could significantly improve the intergeneric resolution but had rarely power to address the interspecific phylogenetic relationships. In this paper, the concatenated universal DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) expectedly failed to resolve the complex phylogenetic relationships among the species of *Mahonia*. On the contrary, based on eight concatenated hypervariable regions, the phylogenetic trees show a similar topological structure with the topology of the plastome trees and possess high support values. This phenomenon indicates that these hypervariable regions yield adequate information to address complicated phylogenetic relationships at the species level. Nowadays, an increasing number of studies consider the hypervariable regions to be valuable and introduce clade-specific barcodes (also named special barcodes) for phylogenetic purposes and even for the purpose of quick identification of medicinal plants [35, 36, 50, 57, 58]. Establishing clade-specific barcodes is far from easy as it depends on a series of factors, including the cost of whole plastome sequencing and sophisticated analytical tools [50]. Based on our results, we believe that the development of clade-specific barcodes has significant implications for species identification and biodiversity conservation in evolutionarily complex taxa [59].

An integrative method for distinguishing closely related species

Micromorphological characteristics are usually constant within species and could be used for detailed species identification [30, 33, 60–62]. Despite a high micromorphological similarity for vegetative and/or reproductive organs among closely related species, micromorphological characteristics, e.g., the structure of glandular trichomes (*Arnebia* and *Lithospermum* in Boraginaceae [63]), petal epidermal cell patterns (Berberidaceae [31]), palynological characters (*Sanguisorba*

(Rosaceae) [56]), have exhibited great diagnostic value. For instance, the patterns of lemma epidermis are taxonomically discriminant and frequently used to elucidate the phylogenetic relationships among different genera of Poaceae [64–66]. The different leaf epidermal characters are congruent with the different clades in *Cinnamomum* (Lauraceae) retrieved by Huang et al. (2016) [67], implying phylogenetic significance [68]. In addition, Wu et al. (2010) performed a set of investigations about seed morphology (i.e., seed size, color and shape, seed coat ornamentations) for 24 species of *Mahonia* [30]. They found that although seed morphological characters are conserved at the genus level, they are diversified enough to enable the division of the genus *Mahonia* into nine types for further systematic studies. In our micromorphological study, the type of stomata, the shape of epidermal cells, the pattern of anticlinal walls and cuticular ornamentation show high diversity among different species. These characters could be regarded as complementary evidence, in addition to molecular data, to distinguish among species of *Mahonia*.

Given the efficiency and objectivity of molecular data, standard DNA barcodes are used as essential elements for discriminating plants [56]. However, standard DNA barcodes (ITS, concatenated *rbcl*, *matK*, and *trnH-psbA*) could not be used to distinguish the species in the genus of *Mahonia*, due to the limited diagnostic information.

In the era of NGS technology, an increasing number of research groups can afford the cost of whole plastid genome sequencing and then employ the data sets to resolve challenging phylogenetic relationships [55, 62, 69, 70]. In this context, we generated a robust phylogenetic framework for the species of *Mahonia* using plastome datasets. Complete plastid genomes encompass adequate sequence variations for detailed identification, but their sequencing encounters some problems: high sequencing cost, huge-scale datasets and sophisticated computational process.

Based on eight concatenated hypervariable regions, the topology of the phylogenetic trees we reconstructed is mostly congruent with the topology of the tree based on the whole plastomes. This indicates that these hypervariable regions have adequate information which is almost equal to the information contained in the whole plastome. We extracted and developed these hypervariable regions into special barcodes, which combine the advantages of standard DNA barcodes and whole plastome [36, 71]. However, molecular data including standard and special barcodes could possibly remain unsuccessful at distinguishing among closely related species, especially in young lineages and lineages hosting an evolutionary radiation.

Micromorphological evidence can be used to address the different alternatives in resolving the polytomies in the tree built based on the special barcodes. For instance, the phylogenetic analysis could not differentiate the three species (*M. bordinieri*, *M. fordii*, *M. japonica*) from each other (Fig. 7C). However, we find that the adaxial surfaces of the epidermal cells of *M. bordinieri*, *M. fordii* and *M. japonica* are slightly convex, convex, and waxy, respectively (Figs. 9E2, S7A₂, F₂). In addition, the shape of epidermal cells of *M. bordinieri* and *M. fordii* is subquadrate and irregular, respectively. We could not determine the cell shape of *M. japonica* because of its invisible or obscure cell boundaries. On the abaxial surface of leaflets, we found actinocytic stomata in *M. fordii*. Several epidermal cells get organized in the form of a rosette around the stomata (Fig. S7A₄). Anomocytic stomata and annular stripe were observed on the abaxial leaflets of *M. japonica* (Fig. S7F₄). The epidermal cells on the abaxial surface of *M. bordinieri* appear to be considerably flat, which is obviously different from the other two species (Fig. 9E4). Therefore, these epidermal features could be used to distinguish the closely related species. Given the absence of wax ornamentations on the adaxial leaf surface of *M. bordinieri* and *M. fordii* compared to *M. japonica*, we suspect that the first two species above-mentioned might have a closer relationship.

Also, this study contributes micromorphological evidence to resolve polytomies in the ITS tree (Fig. 7B). The leaflets of *M. breviracema* are distinctly different from the leaflets of *M. shenii* regarding surface convexity. The former is convex with wax ornamentation (Fig. S7D₂), while the latter is slightly convex (Fig. S8A₂). The distinction indicates that *M. breviracema* may be closer to *M. bealei* and *M. fortunei*, and *M. shenii* may be closer to the *M. eurybracteata* RC614 [MZ158270] and *M. eurybracteata* subsp. *ganpinensis* (Fig. 7B). We can also apply micromorphological traits to differentiate *M. napaulensis*, *M. duclouxiana* and *M. cardiophylla* clearly (Figs. 9B2, S8C₂, D₂). The epidermal cells on the adaxial surface of leaves in *M. napaulensis* are convex without wax ornamentation, whereas the other two have substantial wax ornamentations in the form of stripes. More detailed comparison reveals that the leaf surfaces of *M. duclouxiana* are much flatter than the leaf surfaces of *M. cardiophylla*. However, we prefer to consider a closer relationship between *M. napaulensis* and *M. duclouxiana* based on the similar shape of leaflets (Figs. 9B1, S8C₁).

Conclusion

Based on the integration of molecular data from hypervariable regions and epidermal characters of leaflets, we can distinguish all the species we investigated within the genus *Mahonia*. Although our

sampling for next-generation sequencing is not extensive enough to delimit species boundaries clearly, our results shed a light on the taxonomic, phylogenetic, and evolutionary analysis of the genus *Mahonia*. For further investigations, on the basis of a more comprehensive sampling, we propose an integrative method based on special barcodes and broader macroscopical evidence (e.g., morphological, micromorphological, anatomical and even cytological characteristics) to distinguish closely related species. Furthermore, genetically variable hotspots could be developed as clade-specific barcodes for efficient and rapid species identification especially in medicinal plants. It deserves ongoing and concerted efforts of the "barcode" research community to build a comprehensive system for accurately identifying plant species.

Methods

Taxon sampling, DNA extraction and next-generation sequencing

Fresh leaves from adult plants were collected from Sichuan, Yunnan, and Guizhou provinces of China and immediately dried using silica gel (Table 3). Liang Zhao from Herbarium of Northwest A&F University (WUK) undertook the formal identification of the vouchers and the plant materials used in our study. Voucher specimens of the plant materials we collected have been deposited in WUK. Dried leaves of four species of *Mahonia* (*M. oiwakensis* Hayata RC610 [MZ158277], *M. japonica* (Thunb.) DC. RC615 [MZ158274], *M. aquifolium* (Pursh) Nutt. RC616 [MZ158265] and *M. pinnata* (Lag.) Fedde RC618 [MZ158278]) were taken from voucher specimens of the Herbarium of Northwest A&F University (WUK). The 17 species sampled represented all the subclades of *Mahonia* and these species are from East Asia, Western North America and Europe, where it was inferred to be the center of diversity for *Mahonia* [8]. For this study, complete plastomes were obtained from 17 species of *Mahonia*. In addition, we obtained from GenBank plastomes of three *Mahonia* species (*M. eurybracteata* subsp. *ganpinensis* (H. Lév.) Fedde, GenBank ID: MN417307, *M. fortunei*, GenBank ID: NC_042167, and *M. bealei*, GenBank ID: MH795308) and plastomes for ten outgroup species to conduct subsequent comparative and phylogenetic analyses (Table 3). Total genomic DNAs were extracted from dried leaves using Cetyltrimethylammonium Bromide (CTAB) method [72] and sequenced using the Illumina Miseq platform (Illumina, San Diego, California, USA) at the Beijing Genomics Institute (BGI). Paired-end sequence reads have been trimmed to remove low-quality reads and adapter sequences using Trimmomatic v0.40 [73].

Plastome assembly, annotation and visualization

We obtained approximately 2 GB high-quality data for each sample. The quality-filtered reads were then subjected to de novo assembling with GetOrganelle [74] or NOVOPlasty v4.3 [75], using *Mahonia oiwakensis* (GenBank ID: MN735221) as a reference for assembly. The 17 newly assembled plastome sequences were deposited in GenBank. We also submitted all the raw sequence data to GenBank and obtained SRA accessions (Table 3).

We annotated the 17 assembled plastomes with default parameters using Plastid Genome Annotator [76] (PGA) and inspected the accuracy of annotations with the annotation results from GeSeq [77]. On the basis of the results from PGA, we corrected the errors using Geneious v11.0.2 [78]. We checked the annotations of tRNA using tRNAscan-SE v2.0 [79]. The circular plastome maps of *Mahonia* were plotted using online OGDRAW [80].

Comparative genomic analysis

Plastomes of ten *Mahonia* species were selected for further comparative genomics analyses and also repeated sequence identification. The ten species were representative of the different clades of *Mahonia*. We aligned the ten representative plastomes of *Mahonia* using MAFFT v7.450 [81] and adjusted the boundaries in Geneious. Following the same procedure, we aligned seven plastomes from different genera of Berberidaceae for further comparison of intergeneric sequence identity (see the details from section Results). Then, the two aligned matrices were visualized using online mVISTA program [82] under Shuffle-LAGAN mode with default options for other parameters. In both cases, the reference sequence was *M. bealei* RC601 [MZ158266]. Gene rearrangement events in *Mahonia* were detected using Mauve v2.4.0 [83].

Using Geneious, we compared the construction of ten representative plastomes of *Mahonia* and seven plastomes from different genera in Berberidaceae mentioned above. The IR-SC boundaries of plastomes of the species of *Mahonia* and of the outgroup species were manually detected and plotted.

We employed DnaSP v5.10 [84] to detect the plastid genome divergence and parsimony informative sites among 20 individuals (19 species) of *Mahonia*. A sliding window analysis (window length = 600, step size = 200) allowed us to determine hypervariable regions and estimate the level of polymorphism for subsequent phylogenetic analyses.

Repeated sequence identification

Microsatellites (SSRs) were identified by MISA [85] with the thresholds of ten repeated units, and 6, 5, 5, 5, 4 repeated units for mono-, di-, tri-, tetra-, penta-,

Table 3 Voucher information and GenBank accession numbers for *Mahonia* and outgroups

Taxon	Location	Date	Herbarium	Voucher	Identifier	Accession	SRA Accession	Alt. (m)	Longitude (E)	Latitude (N)
<i>M. bealei</i> *	China:YunNan	2020	WUK	RC601	Liang Zhao	MZ158266	SRR14460604	1539 m	103°17'31.00"E	29°34'2.24"N
<i>M. dulcoulxiana</i> *	China:GuiZhou	2020	WUK	RC602	Liang Zhao	MZ086770	SRR14429474	1107 m	106°41'48.03"E	26°32'7.89"N
<i>M. napaulensis</i> *	China:YunNan	2020	WUK	RC603	Liang Zhao	MZ158275	SRR14460603	1539 m	103°17'31.00"E	29°34'2.24"N
<i>M. cardiophylla</i> *	China:SiChuan	2020	WUK	RC604	Liang Zhao	MZ158269	SRR14460596	-	-	-
<i>M. nitens</i> *	China:SiChuan	2020	WUK	RC605	Liang Zhao	MZ158276	SRR14460595	-	-	-
<i>M. gracilipes</i> *	China:SiChuan	2020	WUK	RC606	Liang Zhao	MZ158272	SRR14460594	-	-	-
<i>M. polydonata</i> *	China:SiChuan	2020	WUK	RC607	Liang Zhao	MZ158279	SRR14460593	-	-	-
<i>M. bodinieri</i> *	China:GuiZhou	2020	WUK	RC608	Liang Zhao	MZ158267	SRR14460592	1107 m	106°41'48.03"E	26°32'7.89"N
<i>M. shenii</i> *	China:GuiZhou	2020	WUK	RC609	Liang Zhao	MZ158280	SRR14460591	1107 m	106°41'48.03"E	26°32'7.89"N
<i>M. oiwakensis</i> *	China:YunNan	1952	WUK	RC610	Liang Zhao	MZ158277	SRR14460590	2400 m	-	-
<i>M. breviracema</i> *	China:YunNan	2020	WUK	RC611	Liang Zhao	MZ158268	SRR14460589	1539 m	103°17'31.00"E	29°34'2.24"N
<i>M. fordii</i> *	China:YunNan	2020	WUK	RC612	Liang Zhao	MZ158271	SRR14460602	1539 m	103°17'31.00"E	29°34'2.24"N
<i>M. hancockiana</i> *	China:YunNan	2020	WUK	RC613	Liang Zhao	MZ158273	SRR14460601	1539 m	103°17'31.00"E	29°34'2.24"N
<i>M. eurybracteata</i> *	China:SiChuan	2020	WUK	RC614	Liang Zhao	MZ158270	SRR14460600	-	-	-
<i>M. japonica</i> *	China:ShaanXi	1959	WUK	RC615	Liang Zhao	MZ158274	SRR14460599	-	-	-
<i>M. aquifolium</i> *	United Kingdom	1976	WUK	RC616	Liang Zhao	MZ158265	SRR14460598	500 m	-	-
<i>M. pinnata</i> *	USA	2000	WUK	RC618	Liang Zhao	MZ158278	SRR14460597	-	-	-
<i>M. bealei</i>	-	-	-	-	-	MH795308	-	-	-	-
<i>M. ganpinensis</i>	-	-	-	-	-	MN417307	-	-	-	-
<i>M. fortunei</i>	-	-	-	-	-	NC_042167	-	-	-	-
<i>Berberis koreana</i>	-	-	-	-	-	NC_030063	-	-	-	-
<i>Berberis amurensis</i>	-	-	-	-	-	NC_030062	-	-	-	-
<i>Berberis weiningensis</i>	-	-	-	-	-	MW018363	-	-	-	-
<i>Berberis aristata</i>	-	-	-	-	-	MN746308	-	-	-	-
<i>Ranzania japonica</i>	-	-	-	-	-	MH423072	-	-	-	-
<i>Leontice armeniaca</i>	-	-	-	-	-	NC_042400	-	-	-	-
<i>Gymnospermium kiangnanense</i>	-	-	-	-	-	NC_041670	-	-	-	-
<i>Caulophyllum robustum</i>	-	-	-	-	-	NC_042221	-	-	-	-
<i>Nandina domestica</i>	-	-	-	-	-	NC_008336	-	-	-	-
<i>Aconitum brachypodum</i>	-	-	-	-	-	NC_041579	-	-	-	-

Species with asterisks were collected by this study, whereas others were obtained from GenBank

and hexanucleotide SSRs, respectively. We used the online Tandem Repeats Finder [86] to find the tandem repeated sequences with the default settings. REPuter program [87] was used to identify the dispersed repeated sequences, including forward, reverse, complement, and palindromic repeats. The minimum repeated size and Hamming distance were set at 30 bp and three (i.e., 90% sequence identity), respectively.

Phylogenetic analyses

Phylogenetic analyses were made based on 17 newly sequenced complete plastomes of *Mahonia* and 13 already published plastid genomes (three species from *Mahonia* and ten species from Berberidaceae and Ranunculaceae for outgroup species). We aligned the plastid genomes and ITS sequences using MAFFT. Phylogenies were reconstructed based on the following datasets: (1) complete plastid genomes; (2) large-single-copy (LSC) region; (3) small-single-copy (SSC) region; (4) one inverted repeat (IR); (5) coding sequences; (6) non-coding sequences; (7) ITS; (8) concatenated sequences of *matK*, *rbcl*, and *trnH-psbA*; and (9) concatenated sequences of eight identified hypervariable regions. We applied the Maximum Likelihood (ML) method and Bayesian inference (BI) for each of the nine datasets to reconstruct phylogenetic trees, respectively. The ML analysis was carried out using RAxML-HPC Black Box [88] on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway [89], with 1000 bootstrap replicates and a GTR-GAMMA + I model to obtain support values. jModelTest [90] was utilized to compute the best-fit model using the Akaike information criterion (AICc) for each partition, which was also conducted at the CIPRES Science Gateway (Table S3). BI trees were generated with MrBayes v3.2 [91]. The Markov chain Monte Carlo (MCMC) analysis was run for 10,000,000 generations and sampled every 1,000 generations. The first 25% trees were discarded as burn-in. The remaining trees were used to estimate the consensus tree and the Bayesian posterior probabilities.

Recording of morphological and micromorphological character states

Images of mature leaves were taken with a Nikon 7100 camera (Nikon, Japan). Fresh leaves were first fixed in FAA (methanol: acetic acid: ethanol: water = 10:5:50:35). Next, small leaf pieces were dehydrated in an increasing alcohol series and isoamyl acetate series, and then, critical-point dried in CO₂ with a K850 critical-point dryer (EMITECH, Ashford, England). Leaf pieces were then mounted on stubs and sputter coated with gold-palladium using a JS-1600 sputter coater (HTCY, China). The materials were photographed with a Hitachi S-3400 scanning electron microscope (SEM, Hitachi, Japan) at 15 kV.

Abbreviations

s.l.: Sensu lato; *s.s.*: Sensu stricto; nrDNA: Nuclear ribosomal DNA; ITS: Internal transcribed spacer; LSC: Large single copy; SSC: Small single copy; IR: Inverted repeat; tRNA: Transfer RNA; rRNA: Ribosomal RNA; SSRs: Simple sequence repeats; BI: Bayesian inference; ML: Maximum likelihood; PP: Posterior probability; BS: Bootstrap support.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08964-0>.

Additional file 1: Fig. S1. Visualization of alignment of *M. bealei* RC601 and six outgroups. *Mahonia bealei* RC601 was used as a reference sequence. Blue represents coding regions, pink represents non-coding regions and gray arrows points at genes. **Fig. S2.** Comparison of the LSC, IR and SSC boundary regions of plastomes of *M. bealei* RC601 and six outgroups. **Fig. S3.** Structural variation between plastomes of ten species of *Mahonia* revealed by Mauve. **Fig. S4.** Phylogenetic relationships of *Mahonia* inferred from BI and ML based on six datasets. A complete plastomes. B coding regions. C large single copy region. D non-coding regions. E small single copy region. F inverted repeated region. The support values above the branches show PP (posterior probability)/BS (bootstrap support), and asterisks indicate 1.00/100%. Dashes represent incongruences of BI and ML trees. **Fig. S5.** Phylogenetic trees of *Mahonia* showed by branch lengths from BI based on four datasets. A complete plastomes. B ITS sequences. C hypervariable regions. D *rbcl+matK+trnH-psbA*. The support values above the branches show PP (posterior probability). Branches without values indicate 1.00. **Fig. S6.** Phylogenetic trees of *Mahonia* showed by branch lengths from ML based on four datasets. A complete plastomes. B ITS sequences. C hypervariable regions. D *rbcl+matK+trnH-psbA*. The support values above the branches show BS (bootstrap support). Branches without values indicate 100. **Fig. S7.** Characteristics of leaflets and epidermal surface. A₁–A₄ *M. fordii*. B₁–B₄ *M. oiwakensis*. C₁–C₄ *M. aquifolium*. D₁–D₄ *M. breviracema*. E₁–E₄ *M. eurybracteata* subsp. *ganpinensis*. F₁–F₄ *M. japonica*. The images show leaflets, adaxial leaves, abaxial leaves and magnifying stomatal apparatus on the abaxial surface in each row from the left to right, respectively. **Fig. S8.** Characteristics of leaflets and epidermal surface. A₁–A₄ *M. shenii*. B₁–B₄ *M. hancockiana*. C₁–C₄ *M. duclouxiana*. D₁–D₄ *M. cardiophylla*. E₁–E₄ *M. nitens*. F₁–F₄ *M. gracilipes*. The images show leaflets, adaxial leaves, abaxial leaves and magnifying stomatal apparatus on the abaxial surface in each row from the left to right, respectively. **Table S1.** Gene composition of the 20 complete *Mahonia* chloroplast genomes. **Table S2.** Numbers of nucleotide substitutions and pairwise sequence distance rate in *Mahonia* plastomes. **Table S3.** Akaike information criterion (AICc) selection results for nine datasets.

Acknowledgements

We thank Fu-zhen Guo, Min-rong Luo, Ying-cui Yu, Guo-yun Zhang, Ning-juan Fan and Xiao-hua He in Northwest A&F University for assistance with the SEM and LM. We thank Dr. Zhi-rong Zhang from Molecular Biology Experiment Center, Germplasm Bank of Wild Species in Southwest China for help with the analysis.

Statement of plant material collection and experimental methods

We have the formal permissions to collect all the materials used in this manuscript. All experimental methods were performed in accordance with the relevant guidelines and regulations.

Authors' contributions

Ru-chang Tong: Investigation, Methodology, Software, Formal analysis, Visualization, Writing-original draft, Writing-review & edition. Chao-xia Gui: Methodology, Software, Writing-original draft. Yu Zhang: Methodology, Software, Writing-original draft. Na Su: Formal analysis, Investigation, Methodology, Software, Visualization, Writing-original draft. Xiao-qi Hou: Investigation, Writing, Resources. Meng Liu: Investigation, Writing, Resources. Zhao-ping Yang: Investigation, Writing, Resources. Bing Kang: Investigation, Writing, Resources. Zhao-yang Chang: Investigation, Resources, Writing. Florian Jabour: Conceptualization, Methodology, Supervision, Validation, Writing-review

& edition. Liang Zhao: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing-review & edition. All the authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 32170381, 31770200 and 32060053], and the Chinese Universities Scientific Fund [grant number 2452020179].

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary materials. All Illumina data have been deposited in NCBI's Sequence Read Archive (SRA). Raw sequence reads are available on NCBI in the BioProject PRJNA727409 and PRJNA72753 (SAMN19020769–SAMN19020785). These sequence data have been submitted to the GenBank databases under accession number MZ086770, MZ158265–MZ158280.

Declarations

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

¹College of Life Sciences, Northwest A&F University, Yangling 712100, Shaanxi, China. ²Herbarium of Northwest A&F University, Yangling 712100, Shaanxi, China. ³Guiyang Botanical Garden, Guiyang 550009, Guizhou, China. ⁴Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China. ⁵College of Life Sciences, Tarim University, Alar 843300, Xinjiang, China. ⁶Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université Des Antilles, 75005 Paris, France.

Received: 1 August 2022 Accepted: 25 October 2022

Published online: 23 November 2022

References

- Yu CC, Chung KF. Why *Mahonia*? Molecular recircumscription of *Berberis s.l.*, with the description of two new genera, *Alloberberis* and *Moranothamnus*. *Taxon*. 2017;66(6):1371–92. <https://doi.org/10.12705/666.6>.
- Hsieh CL, Yu CC, Huang YL, Chung KF. *Mahonia* vs. *Berberis* unloaded: generic delimitation and infrafamilial classification of Berberidaceae based on plastid phylogenomics. *Front Plant Sci*. 2022;12:720171. <https://doi.org/10.3389/fpls.2021.720171>.
- Loconte H. Berberidaceae. In: Kubitzki K, Rohwer JG, Bittrich V, editors. The families and genera of vascular plants II. Berlin: Springer; 1993. p. 147–52.
- Ying TS. Berberidaceae. In: Flora of China Editorial Committee, editor. *Flora Reipublicae Popularis Sinicae*. vol. 29. Beijing: Science Press; 2001. p. 50–305.
- Wang W, Chen ZD, Liu Y, Li RQ, Li JH. Phylogenetic and biogeographic diversification of Berberidaceae in the northern hemisphere. *Syst Bot*. 2007;32(4):731–42. <https://doi.org/10.1600/036364407783390791>.
- Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, et al. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front Plant Sci*. 2016;7:306. <https://doi.org/10.3389/fpls.2016.00306>.
- Ahrendt LWA. *Berberis* and *Mahonia*: a taxonomic revision. *Bot J Linn Soc*. 1961;57(369):1–410. <https://doi.org/10.1111/j.1095-8339.1961.tb00889.x>.
- Chen XH, Xiang KL, Lian L, Peng HW, Erst AS, Xiang XG, et al. Biogeographic diversification of *Mahonia* (Berberidaceae): implications for the origin and evolution of East Asian subtropical evergreen broadleaved forests. *Mol Phylogenet Evol*. 2020;151:106910. <https://doi.org/10.1016/j.ympev.2020.106910>.
- China Pharmacopoeia Committee. *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Technology Press; 2015.
- Müller K, Ziereis K, Gawlik I. The antiproliferative activity against cell growth of human keratinocytes. *Planta Med*. 1995;61(1):74–5. <https://doi.org/10.1055/s-2006-958005>.
- Ying JS, Boufford DE, Brach AR. *Mahonia*. In: Wu ZY, Raven PH, editors. *Flora of China*. vol. 19. Beijing: Science Press; St. Louis, MO: Missouri Botanical Garden Press; 2011. p. 214–8.
- Adhikari B, Milne R, Pennington RT, Särkinen T, Pendry CA. Systematics and biogeography of *Berberis s.l.* inferred from nuclear ITS and chloroplast *ndhF* gene sequences. *Taxon*. 2015;64(1):39–48. <https://doi.org/10.12705/641.21>.
- Kim YD, Kim SH, Landrum LR. Taxonomic and phylogeographic implications from ITS phylogeny in *Berberis* (Berberidaceae). *J Plant Res*. 2004;117(3):175–82. <https://doi.org/10.1007/s10265-004-0145-7>.
- Kim YD, Kim SH, Kim CH, Jansen RK. Phylogeny of Berberidaceae based on sequences of the chloroplast gene *ndhF*. *Biochem Syst Ecol*. 2004;32(3):291–301. <https://doi.org/10.1016/j.bse.2003.08.002>.
- Terabayashi S. The comparative floral anatomy and systematics of the Berberidaceae II. Systematic considerations. *Acta Phytotax Geobot*. 1985;36(1–3):1–13. <https://doi.org/10.18942/bunruichiri.KJ00001078521>.
- Sun YX, Moore MJ, Landis JB, Lin N, Chen L, Deng T, et al. Plastome phylogenomics of the early-diverging eudicot family Berberidaceae. *Mol Phylogenet Evol*. 2018;128:203–11. <https://doi.org/10.1016/j.ympev.2018.07.021>.
- Colin O, Hingsinger DD, Strijk JS. *Mahonia lancasteri* (Berberidaceae), a new species originating from Sichuan (China) described from cultivation. *Phytotaxa*. 2021;482(1):45–54. <https://doi.org/10.11646/phytotaxa.482.1.5>.
- Huang J, Su T, Lebereton-Anberrée J, Zhang ST, Zhou ZK. The oldest *Mahonia* (Berberidaceae) fossil from East Asia and its biogeographic implications. *J Plant Res*. 2016;129(2):209–23. <https://doi.org/10.1007/s10265-015-0775-y>.
- Kim YD, Jansen RK. Chloroplast DNA restriction site variation and phylogeny of the Berberidaceae. *Am J Bot*. 1998;85(12):1766–78. <https://doi.org/10.2307/2446511>.
- Ma J, Yang BX, Zhu W, Sun LL, Tian JK, Wang XM. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene*. 2013;528(2):120–31. <https://doi.org/10.1016/j.gene.2013.07.037>.
- Hebert PDN, Ratnasingham S, de Waard JR. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B*. 2003;270(Suppl 1):96–9. <https://doi.org/10.1098/rsbl.2003.0025>.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci U S A*. 2005;102(23):8369–74. <https://doi.org/10.1073/pnas.0503123102>.
- Hollingsworth ML, Andra Clark A, Forrest LL, Richardson J, Pennington RT, Long DG, et al. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol Ecol Resour*. 2009;9(2):439–57. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>.
- CBOL Plant Working Group. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci U S A*. 2011;108(49):19641–6. <https://doi.org/10.1073/pnas.1104551108>.
- Carstens BC, Pelletier TA, Reid NM, Sattler JD. How to fail at species delimitation. *Mol Ecol*. 2013;22(17):4369–83. <https://doi.org/10.1111/mec.12413>.
- Wiens JJ. Species delimitation: new approaches for discovering diversity. *Syst Biol*. 2007;56(6):875–8. <https://doi.org/10.1080/10635150701748506>.
- Terabayashi S. Studies in the morphology and systematics of Berberidaceae: II. Floral anatomy of *Mahonia japonica* (Thunb.) DC. and *Berberis thunbergii* DC. *Acta Phytotax Geobot*. 1978;29(1–5):106–18. <https://doi.org/10.18942/bunruichiri.KJ00001078289>.
- Terabayashi S. Seedling morphology of the Berberidaceae. *Acta Phytotax Geobot*. 1985;38:63–74. <https://doi.org/10.18942/bunruichiri.KJ0002992235>.
- Brückner C. 2000. Clarification of the carpel number in Papaverales, Caparales, and Berberidaceae. *Bot Rev*. 2000;66(2):155–307. <https://doi.org/10.1007/BF02858151>.
- Wu JY, Qin HN, Xue DY, Zhou KX. Study on seed morphology of *Mahonia* (Berberidaceae). *Guihaia*. 2010;30(2):155–60.

31. Su S, Zhao L, Ren Y, Zhang XH. Diversity of petals in Berberidaceae: development, micromorphology, and structure of floral nectaries. *Protoplasma*. 2021;258:905–22. <https://doi.org/10.1007/s00709-021-01611-7>.
32. Baranova MA. Historical development of the present classification of morphological types of stomates. *Bot Rev*. 1987;53(1):53–79. <https://doi.org/10.1007/BF02858182>.
33. Wu D, Wang H, Lu JM, Li DZ. Comparative morphology of leaf epidermis in *Parnassia* (Parnassiaceae) from China. *Acta Phytotaxon Sin*. 2005;43(3):210–24. <https://doi.org/10.1360/aps040099>.
34. Shah SN, Celik A, Ahmad M, Ullah F, Zaman W, Zafar M, et al. Leaf epidermal micromorphology and its implications in systematics of certain taxa of the fern family Pteridaceae from Northern Pakistan. *Microsc Res Tech*. 2019;82(3):317–32. <https://doi.org/10.1002/jemt.23174>.
35. Ye WQ, Yap ZY, Li P, Comes HP, Qiu YX. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). *Mol Phylogenet Evol*. 2018;127:978–87. <https://doi.org/10.1016/j.ympev.2018.07.001>.
36. Su N, Liu BB, Wang JR, Tong RC, Ren C, Chang ZY, et al. On the species delimitation of the *Maddenia* group of *Prunus* (Rosaceae): evidence from plastome and nuclear sequences and morphology. *Front Plant Sci*. 2021;12:2135. <https://doi.org/10.3389/fpls.2021.743643>.
37. Cai CN, Ma H, Ci XQ, Conran JG, Li J. Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences. *J Syst Evol*. 2021;59(3):504–14. <https://doi.org/10.1111/jse.12556>.
38. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet*. 1996;252(1):195–206. <https://doi.org/10.1007/BF02173220>.
39. Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, et al. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol Gen Genet*. 1992;232(2):206–14. <https://doi.org/10.1007/BF00279998>.
40. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res*. 2004;11(4):247–61. <https://doi.org/10.1093/dnares/11.4.247>.
41. Abdullah Shahzadil, Mehmood F, Ali Z, Malik MS, Waseem S, et al. Comparative analyses of chloroplast genomes among three *Firmiana* species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae Plant. *Gene*. 2019;19:100199. <https://doi.org/10.1016/j.jplgene.2019.100199>.
42. Weng ML, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Delargonium* plastid genomes. *New Phytol*. 2017;214(2):842–51. <https://doi.org/10.1111/nph.14375>.
43. Shen GF, Chen K, Wu M, Kung SD. *Nicotiana* chloroplast genome. *Mol Gen Genet*. 1982;187(1):12–8. <https://doi.org/10.1007/BF00384377>.
44. Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, et al. 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>.
45. Wang M, Chen Y, Hina F, Ohi-Toma T, Li P. The complete chloroplast genome of *Ranzania japonica*, an endangered species native to Japan. *Conserv Genet Resour*. 2018;10(4):671–4. <https://doi.org/10.1007/s12686-017-0898-7>.
46. Kim YD, Jansen RK. Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in the Berberidaceae. *Plant Syst Evol*. 1994;193(1):107–14. <https://doi.org/10.1007/BF00983544>.
47. Yan MH, Fritsch PW, Moore MJ, Feng T, Meng AP, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. *Mol Phylogenet Evol*. 2018;2018(121):198–211. <https://doi.org/10.1016/j.ympev.2018.01.004>.
48. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci Rep*. 2018;8:2210. <https://doi.org/10.1038/s41598-018-20189-4>.
49. Do HDK, Kim C, Chase MW, Kim JH. Implications of plastome evolution in the true lilies (monocot order Liliales). *Mol Phylogenet Evol*. 2020;148:106818. <https://doi.org/10.1016/j.ympev.2020.106818>.
50. Kreuzer M, Howard C, Adhikari B, Bendry CA, Hawkins JA. Phylogenomic approaches to DNA barcoding of herbal medicines: developing clade-specific diagnostic characters for *Berberis*. *Front Plant Sci*. 2019;10:586. <https://doi.org/10.3389/fpls.2019.00586>.
51. Smith SA, Moore MJ, Brown JW, Yang Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol*. 2015;15(1):150. <https://doi.org/10.1186/s12862-015-0423-0>.
52. Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith S. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *Am J Bot*. 2020;107(5):790–805. <https://doi.org/10.1002/ajb2.1468>.
53. Fontaine MC, Pease JB, Steele A, Waterhouse M, Neafsey DE, Sharakhov IV, et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*. 2015;347(6217):1258524. <https://doi.org/10.1126/science.1258524>.
54. Simmonds SE, Smith JF, Davidson C, Buerki S. Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, *Piper* and *Peperomia* (Piperaceae). *Mol Phylogenet Evol*. 2021;163:107229. <https://doi.org/10.1016/j.ympev.2021.107229>.
55. Fu CN, Mo ZQ, Yang JB, Cai J, Ye LJ, Zou JY, et al. Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*. *Mol Ecol Resour*. 2022;22(1):404–14. <https://doi.org/10.1111/1755-0998.13479>.
56. Park I, Song JB, Yang SY, Choi G. A comprehensive study of the genus *Sanguisorba* (Rosaceae) based on the floral micromorphology, palynology, and plastome analysis. *Genes*. 2021;12(11):1764. <https://doi.org/10.3390/genes12111764>.
57. Vaughn JN, Chaluvadi SR, Tushar T, Rangan L, Bennetzen JL. Whole plastome sequences from five ginger species facilitate marker development and define limits to barcode methodology. *PLoS ONE*. 2014;9(10):e108581. <https://doi.org/10.1371/journal.pone.0108581>.
58. Manzanilla V, Kool A, Nguyen Nhat L, Van Nong H, Le Thi ThuH, de Boer HJ. Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species. *BMC Evol Biol*. 2018;18(1):44. <https://doi.org/10.1186/s12862-018-1160-y>.
59. Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Sci Rep*. 2018;8(1):1184. <https://doi.org/10.1038/s41598-018-91591-9>.
60. Kong HZ. Comparative morphology of leaf epidermis in the Chloranthaceae. *Bot J Linn Soc*. 2001;136(3):279–94. <https://doi.org/10.1111/j.1095-8339.2001.tb00573.x>.
61. Ren H, Pan KY, Chen ZD, Wang RQ. Structural characters of leaf epidermis and their systematic significance in Vitaceae. *Acta Phytotax Sin*. 2003;41(6):531–44. <https://doi.org/10.1023/A:1022289509702>.
62. Li QJ, Su N, Zhang L, Tong RC, Zhang XH, Wang JR, et al. Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae). *Sci Rep*. 2020;10(1):19781. <https://doi.org/10.1038/s41598-020-76699-7>.
63. Park I, Yang SY, Song JH, Moon BC. Dissection for floral micromorphology and plastid genome of valuable medicinal *Borages Arnebia* and *Lithospermum* (Boraginaceae). *Front Plant Sci*. 2020;11:606463. <https://doi.org/10.3389/fpls.2020.606463>.
64. Thomasson JR. Epidermal patterns of the lemma in some fossil and living grasses and their phylogenetic significance. *Science*. 1978;199(4332):975–7. <https://doi.org/10.1126/science.199.4332.975>.
65. Ortúñez E, de la Fuente V. Epidermal micromorphology of the genus *Festuca* L. in the Iberian Peninsula. *Plant Syst Evol*. 2010;284(3):201–18. <https://doi.org/10.1007/s00606-009-0248-7>.
66. Nobis M. Taxonomic revision of the *Stipa lipskyi* group (Poaceae: *Stipa* section *Smirnovia*) in the Pamir Alai and Tian-Shan Mountains. *Plant Syst Evol*. 2013;299(7):1307–54. <https://doi.org/10.1007/s00606-013-0799-5>.
67. Huang JF, Li L, van der Werff H, Li HW, Rohwer JG, Crayn DM, et al. Origins and evolution of cinnamon and camphor: a phylogenetic and historical biogeographical analysis of the *Cinnamomum* group (Lauraceae). *Mol Phylogenet Evol*. 2016;96:33–44. <https://doi.org/10.1016/j.ympev.2015.12.007>.
68. Gang Z, Liu B, Rohwer JG, Ferguson DK, Yang Y. Leaf epidermal micromorphology defining the clades in *Cinnamomum* (Lauraceae). *PhytoKeys*. 2021;182:125–48. <https://doi.org/10.3897/phytokeys.182.67289>.
69. Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, et al. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J*. 2011;9(3):328–33. <https://doi.org/10.1111/j.1467-7652.2010.00558.x>.
70. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol Ecol*. 2016;25(7):1423–8. <https://doi.org/10.1111/mec.13549>.

71. Li XW, Yang Y, Henry RJ, Rossetto M, Wang YT, Chen SL. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90(1):157–66. <https://doi.org/10.1111/brv.12104>.
72. Doyle J, Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry*. 1987;19(1):11–5.
73. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
74. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
75. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45(4):e18. <https://doi.org/10.1093/nar/gkw955>.
76. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15(1):50. <https://doi.org/10.1186/s13007-019-0435-7>.
77. Tillich M, Lehwark P, Pellizzer T, Ullbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(1):6–11. <https://doi.org/10.1093/nar/gkx391>.
78. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. 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>.
79. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 2016;41(1):575–81. <https://doi.org/10.1093/nar/gkt289>.
80. 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(1):575–81. <https://doi.org/10.1093/nar/gkt289>.
81. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
82. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. Vista: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(suppl 2):273–9. <https://doi.org/10.1093/nar/gkh458>.
83. 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>.
84. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25(11):1451–2. <https://doi.org/10.1093/bioinformatics/btp187>.
85. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106(6):411–22. <https://doi.org/10.1007/s00122-002-1031-0>.
86. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
87. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633>.
88. Stamatakis A. RaxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688–90. <https://doi.org/10.1093/bioinformatics/btl446>.
89. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE); 2010. New Orleans. p. 1–8.
90. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772. <https://doi.org/10.1038/nmeth.2109>.
91. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42. <https://doi.org/10.1093/sysbio/sys029>.

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

