

RESEARCH

Open Access



# Testing plastomes and nuclear ribosomal DNA sequences as the next-generation DNA barcodes for species identification and phylogenetic analysis in *Acer*

Ning Fu<sup>1,2</sup> , Yong Xu<sup>3</sup>, Lu Jin<sup>1</sup>, Tian-Wen Xiao<sup>1</sup>, Feng Song<sup>1</sup>, Hai-Fei Yan<sup>1</sup>, You-Sheng Chen<sup>1\*</sup> and Xue-Jun Ge<sup>1\*</sup>

## Abstract

**Background** *Acer* is a taxonomically intractable and speciose genus that contains over 150 species. It is challenging to distinguish *Acer* species only by morphological method due to their abundant variations. Plastome and nuclear ribosomal DNA (nrDNA) sequences are recommended as powerful next-generation DNA barcodes for species discrimination. However, their efficacies were still poorly studied. The current study will evaluate the application of plastome and nrDNA in species identification and perform phylogenetic analyses for *Acer*.

**Result** Based on a collection of 83 individuals representing 55 species (c. 55% of Chinese species) from 13 sections, our barcoding analyses demonstrated that plastomes exhibited the highest (90.47%) species discriminatory power among all plastid DNA markers, such as the standard plastid barcodes *matK* + *rbcl* + *trnH-psbA* (61.90%) and *ycf1* (76.19%). And the nrDNA (80.95%) revealed higher species resolution than ITS (71.43%). *Acer* plastomes show abundant interspecific variations, however, species identification failure may be due to the incomplete lineage sorting (ILS) and chloroplast capture resulting from hybridization. We found that the usage of nrDNA contributed to identifying those species that were unidentified by plastomes, implying its capability to some extent to mitigate the impact of hybridization and ILS on species discrimination. However, combining plastome and nrDNA is not recommended given the cytonuclear conflict caused by potential hybridization. Our phylogenetic analysis covering 19 sections (95% sections of *Acer*) and 128 species (over 80% species of this genus) revealed pervasive inter- and intra-section cytonuclear discordances, hinting that hybridization has played an important role in the evolution of *Acer*.

**Conclusion** Plastomes and nrDNA can significantly improve the species resolution in *Acer*. Our phylogenetic analysis uncovered the scope and depth of cytonuclear conflict in *Acer*, providing important insights into its evolution.

**Keywords** *Acer*, Plastome, nrDNA, Next-generation DNA barcodes, Phylogeny

\*Correspondence:  
You-Sheng Chen  
yschen@scbg.ac.cn  
Xue-Jun Ge  
xjge@scbg.ac.cn

<sup>1</sup>Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Conghua Middle School, Guangzhou 510920, China



## Introduction

The accurate identification and description of species is a fundamental task in biology. Despite an estimated 10 million eukaryotic species globally, fewer than 3 million have been scientifically described [1, 2]. The discovery and description of these species require significant resources, including trained personnel and substantial investments of time and money. Even for species with scientific descriptions, traditional morphological methods for identifying unknown specimens can be challenging due to factors such as incomplete specimens, a shortage of taxonomists, or a lack of distinguishing features between species [3–5].

DNA barcoding, an approach to identifying species based on short DNA sequences, offers a solution to the challenges of traditional morphological classification. This approach has been widely studied and applied in animals due to its convenience and efficiency, with the mitochondrial sequence cytochrome oxidase I (*COI*) proving particularly useful as a DNA barcode [6–13]. However, the standard DNA barcodes used in plants, such as ITS, *rbcl*, *matK*, and *trnH-psbA*, do not consistently provide satisfactory species discrimination, especially for recently differentiated species [14–20].

The complete plastome and nuclear ribosomal DNA (nrDNA), which possess much more variable characters, have been recommended as next-generation barcodes (super barcodes/barcodes 2.0) [21–24]. Plastome and nrDNA, which also have multiple copies in each cell of plants, thus can be easily assembled from genome skimming data [15, 16, 25, 26]. With the ever-decreasing cost of genome skimming, more and more barcodes 2.0 have been generated from different plants [3, 27–33]. However, many of these studies only sampled one individual per species [28, 31, 32]. This approach is unable to reveal species boundaries because it fails to test species-level monophyly [3, 29]. Low species resolution from plastomes was sometimes reported, i.e., 27.27% in *Schima* [34], 28.6% in *Fargesia* [33], and c. 50% in *Rhododendron* [3], and chloroplast capture resulting from hybridization may be one of the main reasons for DNA barcoding failure in plants. The efficacy of barcodes 2.0 in more plant taxa, especially for those taxonomic challenging taxa, needs to be further assessed. Moreover, it is worth noting whether the addition of nrDNA can provide different insights from plastome, given the differences between their hereditary processes.

*Acer* L., also known as maple, is an economically important and species-rich genus with over 150 species globally [35, 36]. According to the widely accepted classification by de Jong [35], *Acer* species worldwide were divided into 19 sections. *Acer* is a taxonomic difficult genus, exhibiting abundant morphological variations due to the frequent interspecific/intraspecific hybridization/

introgression [35, 37–47]. The morphological characteristics of inflorescence, leaf shape, bud scale, and fruit shape are highly variable among *Acer* species, and even among the conspecific individuals, there are significant differences in the morphology of vegetative organs [35, 37–40, 42, 44, 45]. An efficient DNA barcode is needed for precise species identification for *Acer* species.

Low species resolution was observed when utilizing several DNA barcodes, including *rbcl*, *matK*, *psbA-trnH*, *trnL-trnF*, *trnS-trnG*, ITS2, and ITS [37, 39, 48]. Lin et al. [37] reported a relatively high species resolution using ITS (73.09%); however, their sample size was limited to 52 individuals of 41 species, supplemented by 119 downloaded ITS sequences from only 10 species. Furthermore, they found ITS ineffective in discriminating species within sect. *Palmata* due to share identical sequences, indicating a shortage in interspecific variations. Similarly, Han et al. [39] reported a peak species resolution of 90.47% when combining four traditional barcodes (ITS+*rbcl*+*matK*+*trnS-trnG*); nevertheless, their study included only 18 *Acer* species (averaging 2 species per section), resulting in inadequate sampling representation within each section.

In recent years, several phylogenetic studies have acquired substantial progress by using plastomes or genome-wide data in *Acer* [49–52]. These studies both obtained highly supported phylogenies and revealed the phylogenetic relationships between *Acer* sections. Most notably, Li et al. [49] uncovered the phylogenetic relationships between 16 *Acer* sections based on 500 nuclear loci. Nevertheless, to our knowledge, no study has extensively compared the phylogenies generated from plastomes and large-scale nuclear sequences and visualized the comparison results for *Acer* so far. This hinders our further understanding of the evolution of this genus.

In this study, we applied a genome skimming approach to obtain whole plastomes and nrDNA of 83 individuals representing 55 *Acer* species. By evaluating the usefulness of plastome and nrDNA as barcodes 2.0 for this taxonomic difficult genus, we aim to address the following issue: (1) Compared to standard/taxon-specific DNA markers, can plastomes and nrDNA improve species discriminatory power in the genus *Acer*? (2) If so, to what extent and how do they enhance the discriminatory power? (3) What insights can plastomes provide into the evolution of *Acer*?

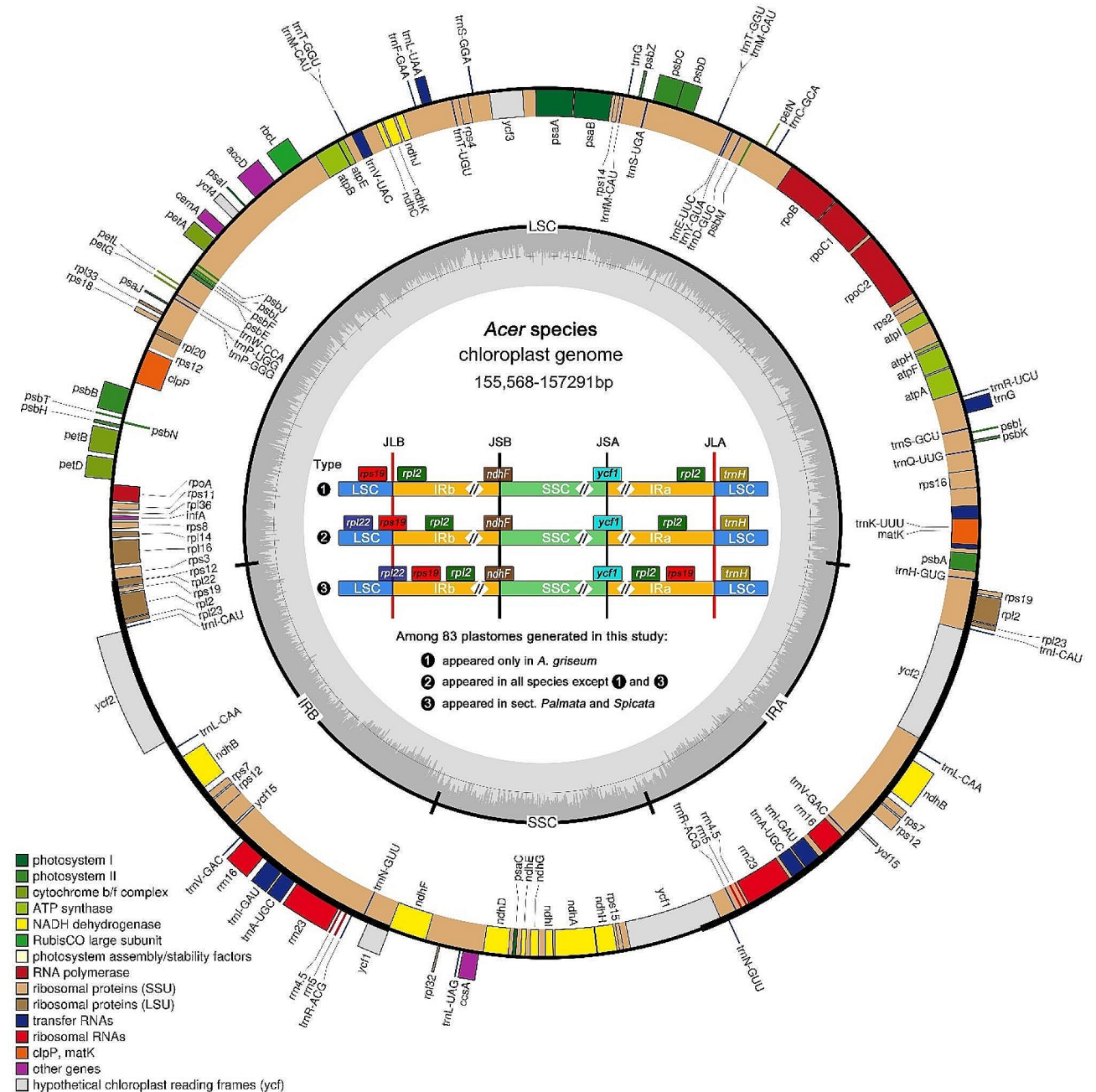
## Results

### Characteristics of *Acer* plastome

Complete plastomes of 83 accessions were successfully obtained without a gap. The size ranges from 155,568 bp (*A. carpinifolium* NJ216) to 157,291 bp (*A. confertifolium* GN100) (Table S1). All sequenced plastomes exhibited the typical quadripartite structure, consisting of a

large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted-repeat (IR) regions (IRa and IRb) (Fig. 1). The overall GC content of these new sequences range from 37.9 to 38% (Table S1). Due to the presence of GC-rich rRNA, IR regions have the highest GC content (42.7–43%), which is higher than the LSC (36–36.2%) and the SSC (32.1–32.4%). All plastomes contain 82 protein-coding genes, 31 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes (Table S2).

The comparative analysis of IR boundaries among 83 plastomes generated in this study uncovered three types of IR boundaries (Fig. 1). Type 1 only appears in *A. griseum*, while type 3 only exists in sect. *Palmata* and sect. *Spicata*; all the remaining *Acer* species exhibit type 2. From type 1 to type 3, a gradual expansion of the IRb region into the LSC region was observed. Previous studies reported that the expansion/contraction of IR borders could result in gene duplication/loss [53–55]. In this



**Fig. 1** Plastome map of *Acer* species and three types of IR boundary identified in this study. Genes inside the outer circle are transcribed clockwise while those outside are transcribed counterclockwise. Genes are color-coded according to their function. Darker gray columns in the inner circle represent the GC content and the lighter gray columns accordingly correspond to the AT content

study, plastomes with a type 3 IR boundary harbor one more copy of gene *rps19* than the other two types due to the expansion of the IRb region into the LSC region, congruent with the results of previous studies [51, 56, 57]. In the study by Xia et al. [51], it was also found that the IR boundary of *A. griseum* is type 1. We also validated the boundary region of this species by aligning the NGS data against its plastome, confirming its existence (Figure S1). This type 1 boundary has also been reported in other species, such as *A. maximowiczii* in Areces-Berazain et al. [57], and *A. amplum* and *A. sterculiaceum* in Wang et al. [56]. However, in our study, these three species did not exhibit a type 1 IR boundary, and they have all been validated (Figure S1).

### Divergence hotspots

The five most variable regions were identified as divergent hotspots in the sliding window analysis (Figure S2). The most variable marker is *ndhC-trnV* (Pi=0.02339), followed by *ndhF-trnL* (Pi=0.02265), *trnK-rps16* (Pi=0.01933), *trnS-trnfM* (Pi=0.01889), *ycf1* (Pi=0.01331) (Table S3). *Ycf1* had the highest percentage of variable sites (11.77%) and contained the most variable sites (513), as well as parsimony informative (PI) sites (291), while *ndhF-trnL* exhibited the highest percentage of PI sites (7.52%). The four most variable markers (*ndhC-trnV*, *ndhF-trnL*, *trnK-rps16*, and *trnS-trnfM*) were combined as a dataset to assess their discriminatory power for the following barcoding analysis. *Ycf1* showed relatively higher individual variation, with haplotypes up to 63, which is much higher than 55 (the number of sampled species in this study), thus it was separately evaluated for the barcoding analysis.

### Characteristics of different barcoding datasets

The plastome dataset (dataset A) was the largest among plastid datasets (dataset A-E), with an aligned length of 138,552 bp (Table 1). The nrDNA dataset (dataset F) had an aligned length of 6,773 bp, which is much longer than the ITS dataset (dataset G, 734 bp). Dataset H was the

largest (145,325 bp) among all datasets as it combined the plastome dataset and nrDNA dataset.

The plastome+nrDNA dataset (dataset H) had the largest number of variable sites (7,869) and PI sites (5,108) (Table 1). The plastome dataset (dataset A) contains 7,501 variable sites and 4,811 PI sites, much higher than that of the standard plastid barcodes (*matK+rbcL+trnH-psbA*, dataset E) (225 variable sites and 148 PI sites) and that of the taxon-specific hypervariable markers (dataset C and D). The nrDNA dataset (dataset F) had many more variable sites (368) and PI sites (297) than the ITS dataset (dataset G) (159 variable sites and 131 PI sites). Among all datasets, the ITS dataset (dataset G) (with 21.66% variable and 17.85% PI sites) exhibited the highest percentage of variable sites as well as PI sites, followed by *ycf1* (dataset D), then the combination of the four most variable markers (dataset C).

### Species discrimination

#### Species discrimination based on phylogenetic tree

In the tree-based method, a species with all conspecific individuals resolved as monophyletic (with a support value  $\geq 50\%$ ) was considered to be successfully identified. The plastome-wide datasets (datasets A and B) exhibited higher resolution than the standard plant barcodes (*matK+rbcL+trnH-psbA*, dataset E) and taxon-specific hypervariable markers (datasets C and D) for the 21 species with multiple individuals sampled (Table 2; Figs. 2 and 3, Figure S3). The plastome, coding region, and plastome+nrDNA (dataset H) datasets all showed the highest resolution of 90.47% (19/21 species successfully discriminated), followed by the combination of the four most variable markers (80.95%) and nrDNA (80.95%), *ycf1* (76.19%), ITS (66.67%), and *matK+rbcL+trnH-psbA* (61.90%).

#### Species discrimination based on K2P distance

In the distance-based method, a species with multiple individuals was regarded as successfully identified when it had a distinct barcoding gap, which means that its minimum interspecific distance is larger than its maximum

**Table 1** Feature comparison of different datasets

Data set	Data set code	Aligned length (bp)	Variable sites		PI sites		Haplotypes
			Number	%	Number	%	
Plastome	A	138,552	7,501	5.41	4,811	3.47	75
Coding region	B	76,307	2,903	3.80	1,815	2.38	70
Combined four most variable markers	C	6,511	629	9.66	433	6.65	63
<i>ycf1</i> (SSC portion)	D	4,359	513	11.77	291	6.68	63
<i>matK+rbcL+trnH-psbA</i>	E	3,615	225	6.22	148	4.09	48
nrDNA	F	6,773	368	5.43	297	4.39	68
ITS	G	734	159	21.66	131	17.85	58
Plastome+nrDNA	H	145,325	7,869	5.41	5,108	3.51	80

Note: PI: Parsimony informative sites

**Table 2** Comparison of species discriminatory efficiency between two methods

Data set	Data set code	Tree-based method	Distance-based method	0K2P	0K2P <sup>55</sup>	AMID
Plastome	A	90.47% (19/21)	90.47% (19/21)	0	0	220 (20–1,004)
Coding region	B	90.47% (19/21)	85.71% (18/21)	0	0	62 (5–292)
Combined four most variable markers	C	80.95% (17/21)	71.43% (15/21)	2	3	17 (0–71)
<i>ycf1</i> (SSC portion)	D	76.19% (16/21)	76.19% (16/21)	2	3	12 (0–59)
<i>matK+rbcL+trnH-psbA</i>	E	61.90% (13/21)	61.90% (13/21)	7	35	5 (0–28)
nrDNA	F	80.95% (17/21)	76.19% (16/21)	3	7	17 (0–71)
ITS	G	66.67% (14/21)	71.43% (15/21)	3	9	11 (0–45)
Plastome+nrDNA	H	90.47% (19/21)	90.47% (19/21)	0	0	212 (30–942)

Note: 0K2P: the total number of species (with multiple individuals) that failed to be discriminated due to showing minimum interspecific K2P distance of zero with other species; 0K2P<sup>55</sup>: the total number of species pairs with 0 K2P distance based on all 83 samples representing 55 *Acer* species; AMID: the average of minimum interspecific differences calculated from 21 *Acer* species with multiple individuals

intraspecific distance [58, 59]. The total number of barcoding gaps in eight datasets ranged from 13 to 19 (Figure S4, Table 2). On the whole, the distance-based method exhibited a similar tendency to the tree-based method. Among the eight datasets, both the plastome and plastome+nrDNA datasets had the highest resolution of 90.47%, followed by the coding region dataset (dataset B) (85.71%), both *ycf1* and nrDNA datasets were 76.19%, both the combined four most variable markers and ITS datasets were 71.43%, finally the *matK+rbcL+trnH-psbA* dataset was 61.90% (Table 2).

Among the 21 species with multiple individuals, no species failed to be discriminated because none showed a minimum interspecific K2P distance of zero in the plastome, coding region, and plastome+nrDNA datasets (Table 2). Furthermore, even among all 83 samples representing 55 species, there were also no species pairs showing 0K2P distance in these three datasets. In contrast, both datasets C and D had 3 pairs of species exhibiting 0K2P distance. For other datasets (datasets E–G), 7 to 35 pairs of species were found with 0K2P distance.

#### Comparison of species discriminatory power between plastome and standard plant barcodes

The plastome dataset significantly improved the species resolution compared to the standard plant barcodes. In the tree-based method, six species were additionally identified by the plastome dataset compared to the standard plant barcodes *matK+rbcL+trnH-psbA* (Table 3). These six species include four species of sect. *Palmata* (i.e., *A. fabri*, *A. flabellatum*, *A. japonicum*, *A. tutcheri*), *A. maximowiczii* of sect. *Macrantha*, and *A. oblongum* of sect. *Oblonga*.

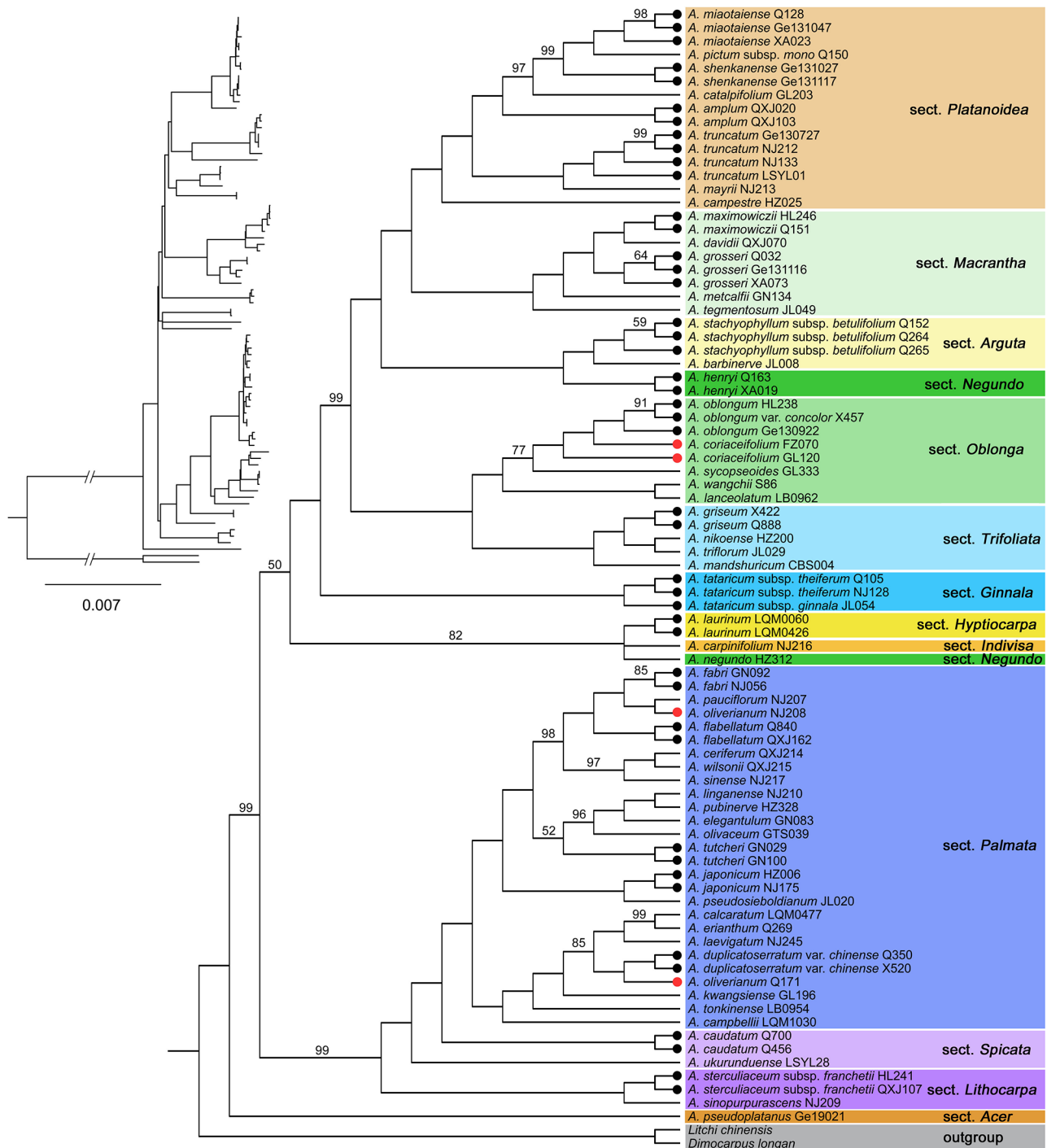
The plastome also increased the support value when species were discriminated (Table 3). Among the 19 species that were successfully discriminated by the plastome dataset, 18 species obtained 100% support value, and *A. fabri* was supported at 85%. However, among the 13 species that were successfully identified by the *matK+rbcL+trnH-psbA* dataset, only six species were supported at 100%, while the support values of five

species were below 90% (three species acquired support values below 65% when they were successfully identified).

#### Phylogenetic analysis of *Acer*

An ML tree containing 267 *Acer* plastomes (128 species and 19 sections) was first constructed (Figure S5). Based on this ML tree, we selected 128 representative accessions (one accession per species) for the following phylogenetic analysis. Using these 128 plastomes (128 species, c. 81% of *Acer* species), two datasets of 80 CDSs were constructed. For these two datasets, tree topologies generated from ML and BI analyses were consistent, and the partitioning strategy only had a slight effect on topology as well as the node support values of the phylogeny (Figure S6). We obtained a well-supported phylogenetic tree after integrating the results of these two datasets (i.e., retaining the higher supported clades) (Fig. 4a).

Comparing the resulting plastid phylogenetic tree with the phylogeny of Li et al. [49] based on 500 nuclear loci, we found many significant cytonuclear discordances between/within sections (see red branches in Fig. 4). Sect. *Platanoidea* and sect. *Macrantha* were 100% supported as sisters in our plastid phylogeny, however, they were quite distant in the nuclear phylogeny. Similar discordances also occurred in sects. *Indivisa* and *Parviflora*, sects. *Rubra* and *Parviflora*, sects. *Macrophylla* and *Negundo*, and sects. *Acer* and *Glabra*. In the nuclear phylogeny, sect. *Arguta* was closely related to sect. *Palmata*, but they were quite distantly related in the plastid phylogeny. And similar conflicts were also found between sects. *Parviflora* and *Glabra*, sects. *Indivisa*, *Lithocarpa* and *Ginnala*, sects. *Platanoidea* and *Macrophylla*. Moreover, we found that sects. *Negundo* and *Parviflora* were both monophyletic in the nuclear tree, however, they were both non-monophyletic with distantly related species in the plastid tree. In addition, although sect. *Acer* was non-monophyletic in both the plastid and nuclear trees, it also exhibited intra-section cytonuclear conflict.



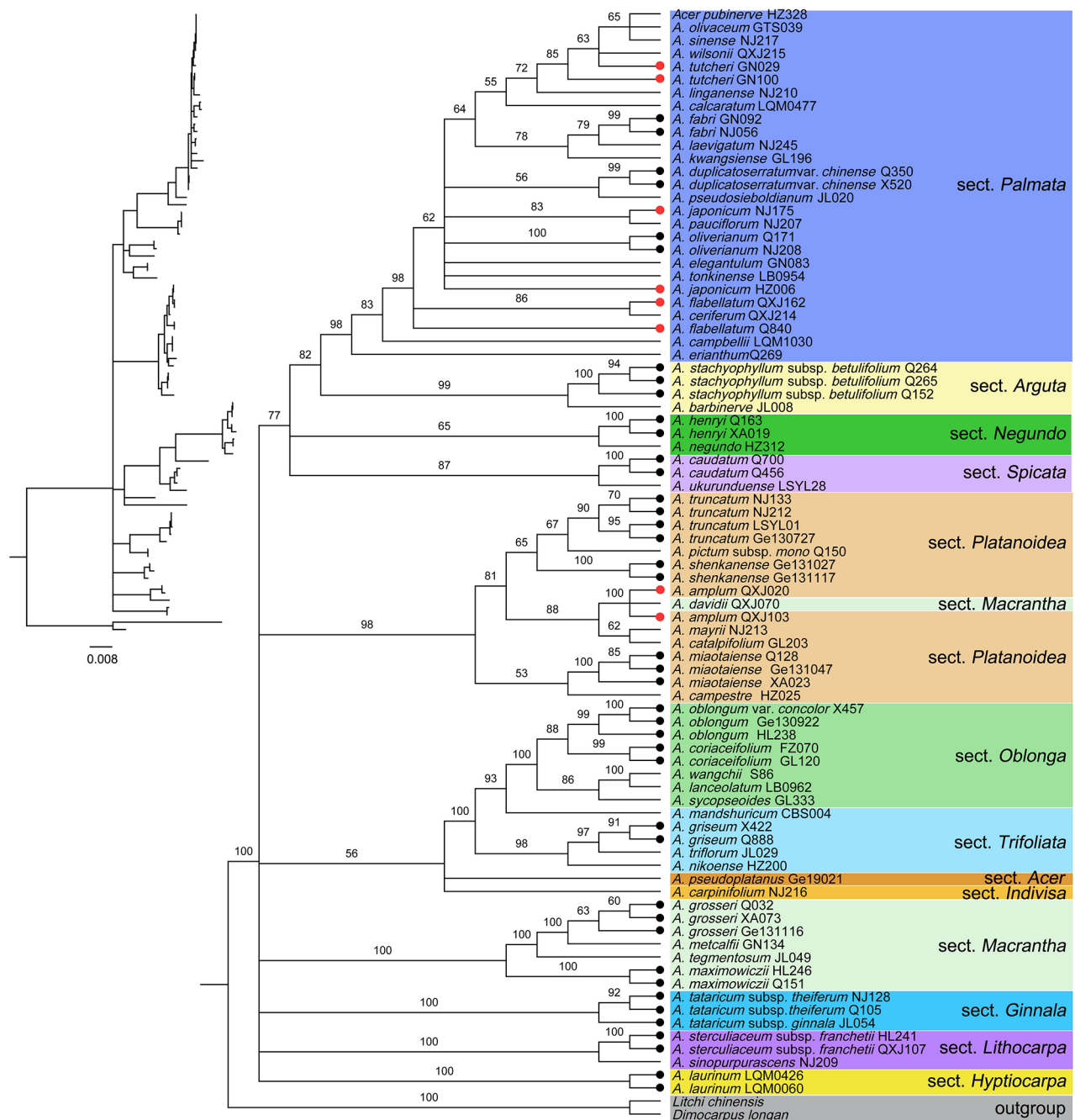
**Fig. 2** ML tree inferred from complete plastomes generated by this study. ML bootstrap support (BS) values are shown at nodes. Clades were set to polytomy when BS < 50%. Species with multiple individuals sampled were marked with dots at branch ends, with black indicating monophyly, while red indicating non-monophyly

## Discussion

### Comparison of species discriminatory power among different barcodes

Plastomes and nrDNA serving as barcodes 2.0 can effectively improve the species resolution compared to

standard DNA barcodes, as revealed by Ji et al. [29] and Fu et al. [3]. Likewise, our barcoding analyses, conducted on various datasets using two different species-identification methods (tree-based and the distance-based), demonstrated that plastomes exhibited the highest



**Fig. 3** ML tree inferred from nrDNA generated by this study. ML bootstrap support (BS) values are shown at nodes. Clades were set to polytomy when BS < 50%. Species with multiple individuals sampled were marked with dots at branch ends, with black indicating monophyly, while red indicating non-monophyly

species discriminatory power (90.47%). Furthermore, the plastome dataset revealed significantly higher species resolution than any other plastid DNA markers, including the standard plastid barcodes (*matK*+*rbcl*+*trnH-psbA*) and taxon-specific hypervariable DNA markers (Table 2). Additionally, nrDNA was found to be more preferable than ITS in our analyses (Tables 2 and 3). This

highlights the importance of considering nrDNA in DNA barcoding studies.

The species resolution of both single plastid sequences and their combinations revealed low species resolution in *Acer*. Han et al. [39], Lin et al. [37], and Lin et al. [48] found that each single plastid locus (such as *matK*, *rbcl*, *trnH-psbA*, *trnL-trnF*, and *trnS-trnG*) provided a species resolution of less than 50% in *Acer*, due to the





lack of genetic variations. Therefore, we constructed a concatenated dataset of standard plastid barcodes (*matK+rbcL+trnH-psbA*) to get more genetic variations. However, the species resolution of this dataset (61.90%) is still insufficient and is the lowest among all datasets (Table 2). Moreover, in this dataset (dataset E, Table 2), a total of 35 pairs of species exhibited 0 K2P distance, indicating a lack of interspecific variations and highlighting the challenge of DNA barcoding in *Acer*. The hypervariable regions in plastome were considered to be useful for species discrimination by Areces-Berazain et al. [57] and Dong et al. [52]. However, our results revealed that the two datasets with five hypervariable regions (dataset C and D; Table 2) showed significantly less resolution than that of the plastome dataset. Although *trnS-trnG* and *trnL-trnF* were previously used as taxon-specific markers in other studies [39, 60], our sliding window analysis did not support their designation as hypervariable regions in *Acer*.

ITS usually demonstrates a better performance than plastid DNA barcodes in most related studies [18] and *Acer* [37, 39]. Both Lin et al. [37] (73.09%) and our study revealed higher species resolution by ITS (66.67% in the tree-based method, and 71.43% in the distance-based method, respectively). However, ITS did not reveal interspecific variations for 9 pairs of species (0K2P<sup>55</sup>: 9, Table 2). Due to the longer sequence, nrDNA showed better performance (80.95% and 76.19% for the tree-based method and the distance-based method, respectively) than ITS.

#### Signal underlying the improvement of species discrimination efficiency of barcodes

The increase in species resolution comes from additional interspecific variation [3]. In our study, the ITS dataset contains fewer variable characters than the *matK+rbcL+trnH-psbA* dataset (Table 1), however, it showed higher species resolution than the *matK+rbcL+trnH-psbA* dataset both in the tree-based and distance-based method (Table 2). The higher resolution of the ITS dataset may benefit from its richer interspecific variations because there were fewer species failed to be discriminated due to showing a minimum interspecific K2P distance of zero in the ITS dataset compared to the *matK+rbcL+trnH-psbA* dataset (3 vs. 7, Table 2). Our regression analysis did show a significantly negative correlation between the species resolution and the total number of 0K2P (Figure S7). This indicates that the lack of interspecific variations is a significant factor hindering the performance of DNA barcodes. Thus, investigating whether barcodes can provide sufficient interspecific variations before their use should be a priority.

Based on all 55 species sampled, we found substantially more species pairs with 0K2P distance in the

*matK+rbcL+trnH-psbA* dataset (0K2P<sup>55</sup>: 35, Table 2), indicative of the lack of interspecific variations in this dataset. In contrast, the number of 0K2P species pairs in the plastome dataset is still zero, and plastomes were proved to have no shortage of interspecific variations because the range of minimum interspecific differences is 20–1,004, with an average of 220 (dataset A, Table 2). However, our undersampling of closely related species may lead to the current overestimation of interspecific variations in the plastome dataset.

Interspecific differences, which reflect the absolute number of interspecific variations, might be a more intuitive quantitative index than K2P distance. To eliminate the impact of undersampling of related species as much as possible, we downloaded some plastomes from NCBI to increase the sampled species to 128 (c. 81% of genus *Acer*) (Figure S6). We found plastomes can still provide abundant interspecific variations (Figure S8), with only 11 pairs of species exhibiting interspecific differences below 10, while 5 of them are subspecies pairs, and only one pair shows interspecific differences of zero (Table S4). It is worth noting that the potential hybridization may lead to underestimation of interspecific differences because hybridization could lead to the chloroplast capture between two species [3, 29, 34]. It follows that *Acer* plastomes could provide rich interspecific variations even in the case of underestimation.

#### Potential reasons for species discrimination failure of plastome

The lack of variations between recently diversified species was regarded as one reason for species discrimination failure of barcodes 2.0 [3, 29, 34]. A negative correlation between the species discriminatory efficiency (SDE) of barcodes and the number of 0K2P was found in this study (Figure S7). However, when the number of 0K2P reaches zero, the SDE will not be improved even if the dataset continues to be longer and contains more variations. For instance, the two plastome-wide datasets (dataset A and B) get the same SDE (90.47%) in the tree-based method, though dataset A is longer and shows a significantly higher average of minimum interspecific difference (AMID) than dataset B (Table 2). This implies that the interspecific variation may have reached saturation for distinguishing existing species. Hybridization and/or incomplete lineage sorting (ILS) may be more possible causes limiting the further improvement of SDE, with a premise that the possibility of misidentification was ruled out because we have identified the specimen carefully and repeatedly. Nevertheless, our inadequate sampling of closely related species may have contributed to this inference.

*Acer* is a speciose genus with extensive interspecific hybridization under natural conditions [37–44, 46,

47]. Due to the characteristics of maternal inheritance of plastomes, hybridization can lead to the sharing of identical or similar plastomes (i.e., chloroplast capture) between species [3, 16, 22, 29, 61]. *Acer* plastomes are maternally inherited [62], they may thus not reflect species boundaries. For instance, *A. oliverianum* was 100% supported as monophyletic in our nrDNA ML tree (Fig. 3), however, the two individuals of this species were relatively distant in our plastome ML tree (Fig. 2). This cytonuclear conflict, accompanied by the grouping of *A. oliverianum* plastomes with other species reflects geographical proximity rather than taxonomic affinity (Fig. 2, Table S5), implying the presence of hybridization.

In addition to hybridization, ILS may be another cause of barcode failure, especially for recently differentiated species [34, 63, 64]. Previous studies reported that the formation of reciprocal monophyly alleles could take millions of years following the speciation event under different practical demographic parameters [65, 66]. For trees, reaching full monophyly may take 50 million years [67]. Therefore, though related *Acer* species have accomplished morphological differentiation, ancestral polymorphism at molecular levels may remain. For example, *A. coriaceifolium* was strongly resolved as monophyletic in our nrDNA ML tree and as a sister to *A. oblongum* (Fig. 3). However, one sample (FZ070) of *A. coriaceifolium* was found to cluster with *A. oblongum* in the plastome ML tree (Fig. 2). Given the taxonomic affinity between *A. coriaceifolium* and *A. oblongum* [42], ILS could not be excluded as a possible cause. More nuclear sequences are needed to confirm whether hybridization or ILS is responsible for this cytonuclear discordance.

#### Suggestion for the usage of barcodes 2.0

Fu et al. [3] demonstrated that the concatenation of plastome and nrDNA can marginally improve the SDE in *Rhododendron*. Nevertheless, our result showed that the SDE was not enhanced when the plastome was combined with nrDNA (Table 2). Although combining them had increased the total number of variable sites (Table 1), the AMID of this dataset was lower than that of the plastome dataset (Table 2). This suggested that concatenating plastome and nrDNA had led to a reduction in the average minimum inter-species genetic variations available, which may be detrimental to species identification. Furthermore, the resulting ML tree inferred from the plastome+nrDNA dataset contained more polytomies than that of the plastome dataset (Fig. 2, Figure S3), illustrating the phylogenetic signal conflict between plastome and nrDNA. Given that the potential hybridization could blur inter-species genetic variations and what we mentioned above, combining plastome and nrDNA is not suggested for species identification in taxa with extensive hybridization similar to *Acer*.

We proved that plastomes can provide much richer interspecific variations and are therefore superior to standard barcodes and taxon-specific hypervariable plastid makers. However, due to the chloroplast capture resulting from hybridization [62], plastomes may not track species boundaries [16, 61]. Biparentally inherited nuclear sequences may be a better choice under this circumstance. For example, we found that two species that failed to be identified by plastomes were precisely successfully discriminated by nrDNA (Table 3). Given this outcome, nrDNA may compensate for the shortcomings of the plastome in species resolution when facing hybridization or ILS, and thus should be included in barcodes 2.0.

Notably, previous barcoding studies did not include ETS (external transcribed spacer) when using nrDNA (Figure S9), i.e., only used the 18 S–5.8 S–26 S cistron including ITS1 and 2 [3, 29, 34]. In our study, we additionally used a portion of ETS (with an aligned length of 834 bp), and this practice is conducive to improving the SDE (Table S6, Figure S10). We suggest incorporating the ETS sequence when using nrDNA in future studies.

Because of the significantly higher SDE of the barcodes 2.0 and the ever-decreasing cost of genome skimming, accompanied by the convenience of assembling plastomes and nrDNA, barcodes 2.0 will be a superior alternative compared to the combination of standard barcodes or any other plastid makers. However, for some more complex taxa, such as *Rhododendron* [3], *Fargesia* [33], and *Schima* [34], the SDE of barcodes 2.0 is unsatisfactory because lower than 60%. Hybridization, recent divergence, ILS, and taxonomic over-splitting are all suggested to be potential causes for the species discrimination failure of barcodes 2.0, and the addition of more nuclear sequences is recommended for these intractable genera [3, 29, 33, 34]. Nevertheless, not all taxa will be as complex as the above-mentioned genera. The situation of different genera still needs to be further studied, and there is still a lack of research on barcodes 2.0 so far.

#### Insights into the phylogenetics of *Acer*

Previous studies on plastid phylogenetics mainly sampled only one species per Sects. [52, 56, 57], however, the phylogenetic position of a single species may not represent the systematic position of a given section if that section is non-monophyletic. Insufficient taxon sampling can lead to strong systematic bias [68], and the increase in taxon sampling can be highly conducive to improving phylogenetic analyses [69]. Thus, it is necessary to sample as many species as possible for a given section to confirm its plastid systematic position.

In our plastid phylogenetic analysis, we sampled over 80% of *Acer* species according to de Jong [35] (Fig. 4, Figure S5–S6). This contributed to confirming the

plastid phylogenetic position of various sections. Notably, we found many prominent cytonuclear discordances between sections and within sections after comparing our plastid phylogeny with the phylogeny of Li et al. [49] based on 500 nuclear loci (Fig. 4). The causes of cytonuclear conflict include hybridization (especially organellar capture) and ILS [70–73]. ILS could apply to rapidly diverged species/lineages [74], i.e., for closely related species/lineages, which means that the affinity will be shown in both the plastid tree and nuclear tree, as revealed by Li et al. [73] in *Thuja*. However, most of the inter- and intra-section cytonuclear discordances illustrated in Fig. 4 merely reflect the closeness in one tree, while showing a quite distant relationship in another tree. ILS may not be the major factor accounting for these cytonuclear conflicts because the affinities were not shown in both the plastid tree and nuclear tree. And the most typical examples of this are the relationships between sects. *Platanoides* and *Macrantha*, sects. *Arguta* and *Palmata*. It may follow that hybridizations are widely present between sections and have played a significant role in the evolution history of *Acer*. Nevertheless, to our knowledge, there is currently no research that details the extensive inter-section hybridization process of this genus. Further studies on gene flow using comprehensive nuclear genome-wide data and extensive species sampling are needed to explore this matter thoroughly in the future.

## Conclusion

Here we sequenced and assembled the plastomes as well as nrDNA of 83 individuals from 55 *Acer* species, and then assessed and compared the species discriminatory power of different barcoding datasets in *Acer*. Our results illustrated that both plastomes and nrDNA can effectively improve the species resolution in *Acer*, and plastomes exhibited the highest species resolution and most abundant interspecific variations. The use of nrDNA helps discriminate species that cannot be identified by plastomes. The plastid phylogenetic framework generated here enriched our understanding of the evolution of *Acer*, especially highlighting the role of hybridization in it.

## Methods

### Taxon sampling

83 individuals of 55 *Acer* species were sampled in this study (Table S5). Healthy leaves were collected and dried with silica gel. Voucher specimens were deposited at the herbarium of South China Botanical Garden (IBSC), Chinese Academy of Sciences, China. These 55 *Acer* species represent 13 major sections currently recognized in *Acer* [35, 42], 21 species were sampled with multiple (2–4) individuals, and the remaining 34 species with a single individual. All samples were identified by Dr. You-Sheng

Chen. We also downloaded 184 *Acer* plastomes (Table S7) from GenBank. In total, 267 *Acer* plastomes (83+184) representing 128 species and 19 sections were used in our phylogenetic analysis and only sect. *Wardiana* (a monotypic section with only one species *A. wardii* W.W. Sm.) was not included, according to Xu et al. [42] and de Jong [35] (we adopted the treatment that sect. *Pentaphylla* was split into sect. *Oblonga* and *Pentaphylla* by Xu et al. [42]). In addition, the nrDNA (MW0702 and MW070204) and plastomes of two individuals, *Dimocarpus longan* and *Litchi chinensis*, were downloaded as outgroups (Table S7).

### DNA extraction, sequencing, assembly and annotation

Total genomic DNA was extracted from silica gel-dried leaves using the modified CTAB method [75]. Pair-end (PE) libraries with an average insert size of 270 base-pair (bp) were constructed at Beijing Genomics Institute (BGI, Shenzhen, China). Then, the libraries were sequenced on an Illumina X ten platform (San Diego, California) to generate 150 bp PE reads. Raw reads were subjected to quality check using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Clean reads were obtained after raw reads were trimmed and adaptors were removed by using Trimmomatic v0.36 [76]. Finally, each sample generated approximately 2–4 Gb of clean data. We assembled clean reads into plastome and nrDNA using the toolkit GetOrganelle v1.7.5 [77]. This toolkit extracts plastome reads and nuclear reads from total genomic reads for the following assembly by spades v3.10 [78]. As in rare cases, GetOrganelle generated some non-overlapping contigs instead of a complete plastome. Therefore, we mapped reads against these non-overlapping contigs to extend their ends to close the gap in Geneious, performing with medium-low sensitivity for 100 iterations.

Two independent approaches were applied to annotate 83 plastomes generated in this study. Firstly, the annotation of the plastome sequences was performed with GeSeq [79] choosing the plastome of *Acer miaotaiense* P. C. Tsoong (GenBank accession No.: NC\_030343) as the reference genome. In the meantime, ARAGORN was selected as a third party to annotate tRNA. Secondly, we used MAFFT v7.388 [80] to align and annotate these plastome sequences by using the “Annotation Transfer” option with *Acer platanoides* L. (GenBank accession No.: MN864507) as reference in Geneious v2019.2.1. The annotation results from GeSeq and Geneious were subsequently compared and integrated. The annotation of nrDNA was conducted in Geneious with *Acer pentaphyllum* (GenBank accession number: MW070163) as the reference. The plastome map was drawn by using OGDRAW within GeSeq. Newly generated plastomes and nrDNA here were finally uploaded to GenBank

(accession numbers in Table S8). Bwa v0.7.17-r1188 [81] and SAMtools v1.5 [82] were used to map the NGS data against corresponding plastome for validation of IR boundary, and the outputs were visualized in Geneious.

### Plastome analyses

The borders between the four plastome regions, i.e., LSC/IRb (JLB), SSC/IRb (JSB), SSC/IRa (JSA), and LSC/IRa (JLA), were visualized using the online program IRscope (<https://irscope.shinyapps.io/irapp/>). A sliding window analysis was performed in DnaSP v6.12.03 [83] to locate hypervariable genomic regions. The 83 *Acer* plastomes were aligned using MAFFT v7.388 [80] with default settings and used as the input file. The window length and step size were set to 600 bp and 100 bp, respectively. Those genomic regions with crest Pi (nucleotide diversity) values exceeding 0.020 and aligned lengths longer than 600 bp were identified as hypervariable genomic regions, and they were subsequently extracted from the plastome alignment using Geneious and analyzed separately to evaluate their characteristics. In addition, the analysis of indel polymorphism was also conducted in DnaSP.

### Data analyses for species discrimination

We constructed the following eight datasets based on our 83 samples of 55 *Acer* species: (A) the whole plastome with one IR removed, (B) the concatenation of the coding regions of protein-coding genes (PCG), rRNA genes and tRNA genes, (C) the combination of the four most variable markers identified by sliding window analysis in this study (*trnK-rps16+trnS-trnfM+ndhC-trnV+ndhF-trnL*), (D) *ycf1* (SSC portion), (E) the combination of three standard plastid barcodes (*matK+rbcL+trnH-psbA*) (F) the nrDNA sequence (ETS+18 S+ITS1+5.8 S+ITS2+26 S), (G) ITS (ITS1+5.8 S+ITS2), (H) the combination of plastome and nrDNA.

All the coding sequences in annotated plastomes, including the coding sequences of protein, rRNA, and tRNA, were individually extracted by applying a Python script ([https://github.com/Kinggerm/PersonalUtilities/blob/master/get\\_annotated\\_regions\\_from\\_gb.py](https://github.com/Kinggerm/PersonalUtilities/blob/master/get_annotated_regions_from_gb.py)). The ITS sequences were extracted from the annotated nrDNA assemblies in Geneious. For each dataset, the alignment was generated by MAFFT v7.388 [80] and then checked and manually modified in Geneious.

We accessed the species resolution of the above datasets using tree-based and distance-based methods. In the tree-based method, phylogenetic analyses were performed using maximum likelihood (ML) analysis in RAxML v8.2.12 [84] with GTR+ $\Gamma$  model, and 1,000 rapid bootstrap replicates were generated to evaluate the support values for each node. In the distance-based method, the pairwise distance was calculated using the Kimura

2-parameter (K2P) model [85] in the software MEGA7 [86]. The scatter plot of the minimum interspecific distance versus maximum intraspecific distance was generated to illustrate the barcoding gaps for each dataset. For comparing the richness of interspecific variations among different datasets, the pairwise differences (use No. of differences as a model when calculating pairwise distance) were also estimated in MEGA7.

In addition, a dataset containing 267 *Acer* plastomes (184 downloaded and 83 generated in this study) representing 128 species was constructed, and the ML analysis was performed on this dataset. Based on the resulting ML tree, 128 representative individuals (one individual per species) were selected for calculating interspecific differences and the following phylogenetic analysis. When situations where individuals of species from different sections nest with each other occur, our sampling principle is as follows: (1) retain the monophyletic and only-one-sample species; (2) prioritize our own samples; (3) retain individuals within their correct section while excluding those strays. This approach aims to mitigate potential identification errors and the impacts of hybridization, thus focusing more on inter-section relationships.

### Phylogenetic analysis

In total, 128 plastomes representing 128 *Acer* species (c. 81% of this genus) and 19 (95%) sections were sampled for the phylogenetic reconstruction. The 80 protein-coding sequences (CDSs) in annotated plastomes were individually extracted applying the aforesaid Python script and aligned using MAFFT with default settings. Two datasets were constructed based on these 80 CDSs using two partitioning strategies. For the first dataset, the alignments of the 80 CDSs were concatenated and regarded as a whole (i.e., unpartitioned strategy). For the second one, the alignments of the 80 CDSs were concatenated but partitioned (i.e., partitioned strategy). The ML and Bayesian inference (BI) analyses were both performed on these two datasets.

PartitionFinder2 [87] was used to select the best partitioning scheme and best-fit substitution models for the partitioned dataset. The model of evolution was set as 'all' and other parameters were kept as default. The 80 data blocks were consolidated into 31 subsets in the best-fit scheme (Table S9). These subsets and their corresponding substitution models were specified in both ML and BI analyses. For the unpartitioned dataset, GTR+I+G was selected as the best-fit substitution model using ModelTest-NG [88] under the corrected Akaike Information Criterion (AICc).

All ML analyses were performed using IQ-TREE [89] with 1000 ultrafast bootstraps [90]. All BI analyses were conducted in MrBayes v3.2.6 [91], and two MCMC runs were performed with 5 million generations and four

chains, sampling every 1000 generations and discarding the 25% as burnin. LogCombiner within Beast v2.6.4 [92] was then applied to combine log files of the two MCMC runs. Tracer v1.7.2 [93] was finally used to confirm that the effective sample size (ESS) for each parameter was larger than 200 to ensure the convergence of MCMC run.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05073-w>.

Supplementary Material 1

## Acknowledgements

We acknowledge Tong-Jian Liu, Jia-Jia Liu, Lu Liu, Lian-Sheng Xu, Nan Zhao, and Yu-Ying Zhou for their help in analyses and experiments. We also thank Yun-Fei Deng, Xiu-Juan Qiao, Qiao-Ming Li, Shuan-Lu Dong, Feng Jiang, Ji Ye, Feng-Lin Chen, and Yi-Hua Tong for providing samples.

## Author contributions

XJG conceived the idea and designed the experiments. YX collected the samples. YSC, YX, and NF identified the samples. NF, YX, LJ, TWX, and FS analyzed the sequence data. NF drafted the manuscript. XJG, YSC, and HFY revised the manuscript. All authors read and approved the final manuscript.

## Funding

This study was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB31000000).

## Data availability

All complete plastomes and nrDNA sequences used in this study are available from the National Center for Biotechnology Information (NCBI) (see Table S7, S8) and the Science Data Bank at <https://doi.org/10.57760/sciencedb.18484>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 21 February 2024 / Accepted: 26 April 2024

Published online: 23 May 2024

## References

1. Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, et al. Counting animal species with DNA barcodes: Canadian insects. *Phil Trans R Soc B*. 2016;371(1702):1–10. <https://doi.org/10.1098/rstb.2015.0333>.
2. Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on Earth and in the ocean? *PLoS Biol*. 2011;9(8):e1001127. <https://doi.org/10.1371/journal.pbio.1001127>.
3. 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*. 2021;00:1–11. <https://doi.org/10.1111/1755-0998.13479>.
4. Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, et al. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J*. 2015;14(1):8–21. <https://doi.org/10.1111/pbi.12419>.
5. Vohra P, Khera KS. DNA barcoding: current advances and future prospects—a review. *Asian J Biol Life Sci*. 2013;3(3):185–9.
6. deWaard JR, Ratnasingham S, Zakharov EV, Borisenko AV, Steinke D, Telfer AC, et al. A reference library for Canadian invertebrates with 1.5 million barcodes, voucher specimens, and DNA samples. *Sci Data*. 2019;6(1):308. <https://doi.org/10.1038/s41597-019-0320-2>.
7. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc R Soc Lond B*. 2003;270(1512):313–21. <https://doi.org/10.1098/rspb.2002.2218>.
8. Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou JM, Chacon I, et al. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour*. 2009;9(Suppl 1):1–26. <https://doi.org/10.1111/j.1755-0998.2009.02628.x>.
9. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proc Natl Acad Sci U S A*. 2008;105(12):6350–5. <https://doi.org/10.1073/pnas.0712181105>.
10. Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN. Comprehensive DNA barcode coverage of north American birds. *Mol Ecol Notes*. 2007;7:535–43. <https://doi.org/10.1111/j.1471-8286.2006.01670.x>.
11. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. *PLoS Biol*. 2004;2(10):e312. <https://doi.org/10.1371/journal.pbio.0020312>.
12. Hebert PDN, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci*. 2003;270(Suppl):S96–9. <https://doi.org/10.1098/rsbl.2003.0025>.
13. Gregory TR. DNA barcoding does not compete with taxonomy. *Nature*. 2005;434:1067. <https://doi.org/10.1038/4341067b>.
14. Kress WJ. Plant DNA barcodes: applications today and in the future. *J Syst Evol*. 2017;55(4):291–307. <https://doi.org/10.1111/jse.12254>.
15. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol Ecol*. 2016;25:1423–8. <https://doi.org/10.1111/mec.13549>.
16. Hollingsworth PM, Li DZ, van der Bank M, Twyford AD. Telling plant species apart with DNA: from barcodes to genomes. *Phil Trans R Soc B*. 2016;371(1702):1–9. <https://doi.org/10.1098/rstb.2015.0338>.
17. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLoS ONE*. 2011;6(5):e19254. <https://doi.org/10.1371/journal.pone.0019254>.
18. Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, et al. 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>.
19. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2009;106(31):12794–7. <https://doi.org/10.1073/pnas.0905845106>.
20. Song F, Li T, Burgess KS, Feng Y, Ge XJ. Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. *BMC Plant Biol*. 2020;20(1):261. <https://doi.org/10.1186/s12870-020-02466-5>.
21. Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes? *Plant J*. 2017;90(4):808–18. <https://doi.org/10.1111/tpj.13491>.
22. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour*. 2015;15(5):1067–78. <https://doi.org/10.1111/1755-0998.12375>.
23. Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JM, et al. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot*. 2012;99(2):320–9. <https://doi.org/10.3732/ajb.1100570>.
24. 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>.
25. Zeng CX, Hollingsworth PM, Yang J, He ZS, Zhang ZR, Li DZ, et al. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods*. 2018;14:43. <https://doi.org/10.1186/s13007-018-0300-0>.
26. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am J Bot*. 2012;99(2):349–64. <https://doi.org/10.3732/ajb.1100335>.

27. Zhang W, Sun Y, Liu J, Xu C, Zou X, Chen X, et al. DNA barcoding of *Oryza*: conventional, specific, and super barcodes. *Plant Mol Biol.* 2020;105:215–28. <https://doi.org/10.1007/s11103-020-01054-3>.
28. Zhang Z, Zhang Y, Song M, Guan Y, Ma X. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. *Front Pharmacol.* 2019;10:1441. <https://doi.org/10.3389/fphar.2019.01441>.
29. Ji Y, Liu C, Yang Z, Yang L, He Z, Wang H, et al. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Mol Ecol Resour.* 2019;19(5):1333–45. <https://doi.org/10.1111/1755-0998.13050>.
30. Fu CN, Wu CS, Ye LJ, Mo ZQ, Liu J, Chang YW, et al. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Sci Rep.* 2019;9(1):2773. <https://doi.org/10.1038/s41598-019-39161-x>.
31. Chen X, Zhou J, Cui Y, Wang Y, Duan B, Yao H. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. *Front Pharmacol.* 2018;9:1–11. <https://doi.org/10.3389/fphar.2018.00695>.
32. Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang X. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Sci Rep.* 2018;8(1):1184. <https://doi.org/10.1038/s41598-018-19591-9>.
33. Lv SY, Ye XY, Li ZH, Ma PF, Li DZ. Testing complete plastomes and nuclear ribosomal DNA sequences for species identification in a taxonomically difficult bamboo genus *Fargesia*. *Plant Divers.* 2023;45(2):147–55. <https://doi.org/10.1016/j.pld.2022.04.002>.
34. Yu XQ, Jiang YZ, Folk RA, Zhao JL, Fu CN, Fang L, et al. Species discrimination in *Schima* (Theaceae): next-generation super-barcodes meet evolutionary complexity. *Mol Ecol Resour.* 2022;00:1–15. <https://doi.org/10.1111/1755-0998.13683>.
35. de Jong PC. Worldwide maple diversity. In: *Proc Int Maple Symposium*: 2002; 2002: 1–12.
36. Crowley D, Barstow M, Rivers M, Harvey-Brown Y. The Red List of *Acer*: revised and extended. *Descanso House*, 199 Kew Road, Richmond, Surrey, TW9 3BW, UK: Botanic Gardens Conservation International; 2020.
37. Lin L, Zhu Z, Lin L, Kuai B, Ding Y, Du T. Implications of nrDNA and cpDNA region in *Acer* (Aceraceae): DNA barcoding and phylogeny. *Int J Agric Biol.* 2019;21:1073–82. <https://doi.org/10.17957/IJAB/15.0996>.
38. Gao J, Liao PC, Meng WH, Du FK, Li JQ. Application of DNA barcodes for testing hypotheses on the role of trait conservatism and adaptive plasticity in *Acer* L. section *Palmata* Pax (Sapindaceae). *Braz J Bot.* 2017;40(4):993–1005. <https://doi.org/10.1007/s40415-017-0404-1>.
39. Han YW, Duan D, Ma XF, Jia Y, Liu ZL, Zhao GF, et al. Efficient identification of the forest tree species in Aceraceae using DNA barcodes. *Front Plant Sci.* 2016;7:1707. <https://doi.org/10.3389/fpls.2016.01707>.
40. Li J. Phylogenetic evaluation of series delimitations in section *Palmata* (*Acer*, Aceroidae, Sapindaceae) based on sequences of nuclear and chloroplast genes. *Aliso.* 2011;29(1):43–9. <https://doi.org/10.5642/aliso.20112901.05>.
41. Liao PC, Shih HC, Yen TB, Lu SY, Cheng YP, Chiang YC. Molecular evaluation of interspecific hybrids between *Acer albopurpurascens* and *A. buergerianum* var. *formosanum*. *Bot Stud.* 2010; 51:413–420.
42. Xu TZ, Chen YS, de Jong PC, Oterdoom HJ, Chang CS. *Aceraceae*. *Flora of China*. Volume 11. Beijing, China: Science; 2008.
43. Grimm GW, Denk T, Hemleben V. Evolutionary history and systematics of *Acer* section *Acer* – a case study of low-level phylogenetics. *Plant Syst Evol.* 2007;267:215–53. <https://doi.org/10.1007/s00606-007-0572-8>.
44. Li J, Yue J, Shoup S. Phylogenetics of *Acer* (Aceroidae, Sapindaceae) based on nucleotide sequences of two chloroplast non-coding regions. *Harv Papers Bot.* 2006;11(1):101–15. [https://doi.org/10.3100/1043-4534\(2006\)11\[101:Poasb\]2.0.Co;2](https://doi.org/10.3100/1043-4534(2006)11[101:Poasb]2.0.Co;2).
45. de Jong PC. *Maples of the world*. Portland: Timber; 1994.
46. Gao J, Meng WH, Fang D, Li JQ. DNA barcoding of *Acer palmatum* (Aceraceae). *Plant Sci J.* 2015;33(6):734–43. <https://doi.org/10.11913/PSJ.2095-0837.2015.60734>.
47. Lin L, Lin LJ, Zhu ZY, Ding YL, Kuai BK. Studies on the taxonomy and molecular phylogeny of *Acer* in China. *Acta Horticulturae Sinica.* 2017;44(8):1535–47. <https://doi.org/10.16420/j.issn.0513-353x.2016-0912>.
48. Lin L, Zhu ZY, Lin LJ, Liu F, Zhou Y, Li W, et al. Application of ITS2 sequences for species identification and phylogeny of Genus *Acer* (Aceraceae). *Int J Agric Biol.* 2020;24:1582–90. <https://doi.org/10.17957/IJAB/15.1598>.
49. Li J, Stukel M, Bussies P, Skinner K, Lemmon AR, Lemmon EM, et al. Maple phylogeny and biogeography inferred from phylogenomic data. *J Syst Evol.* 2019;57(6):594–606. <https://doi.org/10.1111/jse.12535>.
50. Areces-Berazain F, Hingsinger DD, Strijk JS. Genome-wide supermatrix analyses of maples (*Acer*, Sapindaceae) reveal recurring inter-continental migration, mass extinction, and rapid lineage divergence. *Genomics.* 2021;113(2):681–92. <https://doi.org/10.1016/j.ygeno.2021.01.014>.
51. Xia X, Yu X, Fu Q, Zhao Y, Zheng Y, Wu Y, et al. Comparison of chloroplast genomes of compound-leaved maples and phylogenetic inference with other *Acer* species. *Tree Genet Genomes.* 2022;18(2):1–12. <https://doi.org/10.1007/s11295-022-01541-2>.
52. Dong PB, Wang RN, Afzal N, Liu ML, Yue M, Liu JN, et al. Phylogenetic relationships and molecular evolution of woody forest tree family Aceraceae based on plastid phylogenomics and nuclear gene variations. *Genomics.* 2021;113(4):2365–76. <https://doi.org/10.1016/j.ygeno.2021.03.037>.
53. Wang W, Chen S, Zhang X. Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (bambusoideae: Olyreae). *Molecules.* 2018;23(7):1–20. <https://doi.org/10.3390/molecules23071537>.
54. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76:273–97. <https://doi.org/10.1007/s11103-011-9762-4>.
55. Fu N, Ji M, Rouard M, Yan HF, Ge XJ. Comparative plastome analysis of Musaceae and new insights into phylogenetic relationships. *BMC Genomics.* 2022;23(1):223. <https://doi.org/10.1186/s12864-022-08454-3>.
56. Wang W, Chen S, Zhang X. Complete plastomes of 17 species of maples (Sapindaceae: *Acer*): comparative analyses and phylogenomic implications. *Plant Syst Evol.* 2020;306:61. <https://doi.org/10.1007/s00606-020-01690-8>.
57. Areces-Berazain F, Wang Y, Hingsinger DD, Strijk JS. Plastome comparative genomics in maples resolves the infrageneric backbone relationships. *PeerJ.* 2020;8:e9483. <https://doi.org/10.7717/peerj.9483>.
58. Collins RA, Cruickshank RH. The seven deadly sins of DNA barcoding. *Mol Ecol Resour.* 2013;13(6):969–75. <https://doi.org/10.1111/1755-0998.12046>.
59. Liu J, Shi L, Han J, Li G, Lu H, Hou J, et al. Identification of species in the angiosperm family Apiaceae using DNA barcodes. *Mol Ecol Resour.* 2014;14(6):1231–8. <https://doi.org/10.1111/1755-0998.12262>.
60. Tian X, Guo ZH, Li DZ. Phylogeny of Aceraceae based on ITS and *trnL-F* data sets. *Acta Bot Sin.* 2002;44:714–24. <https://doi.org/10.3321/j.issn:16729072.2002.06.015>.
61. Petit RJ, Excoffier L. Gene flow and species delimitation. *Trends Ecol Evol.* 2009;24(7):386–93. <https://doi.org/10.1016/j.tree.2009.02.011>.
62. Du FK, Petit RJ, Liu J. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea Asperata* complex in China, and comparison with other conifers. *Mol Ecol.* 2009;18(7):1396–407. <https://doi.org/10.1111/j.1365-294X.2009.04107.x>.
63. Nichols R. Gene trees and species trees are not the same. *Trends Ecol Evol.* 2001;16(7):358–64. [https://doi.org/10.1016/s0169-5347\(01\)02203-0](https://doi.org/10.1016/s0169-5347(01)02203-0).
64. Woolfit M. Effective population size and the rate and pattern of nucleotide substitutions. *Biol Lett.* 2009;5(3):417–20. <https://doi.org/10.1098/rsbl.2009.0155>.
65. Hudson RR, Coyne JA. Mathematical consequences of the genealogical species concept. *Evolution.* 2002;56(8):1557–65. <https://doi.org/10.1111/j.0014-3820.2002.tb01467.x>.
66. Knowles LL, Carstens BC, Weins J. Delimiting species without Monophyletic Gene Trees. *Syst Biol.* 2007;56(6):887–95. <https://doi.org/10.1080/10635150701701091>.
67. Naciri Y, Linder HP. Species delimitation and relationships: the dance of the seven veils. *Taxon.* 2015;64(1):3–16. <https://doi.org/10.12705/641.24>.
68. Heath TA, Hedtke SM, Hillis DM. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol.* 2008;46(3):239–57. <https://doi.org/10.3724/SPJ.1002.2008.08016>.
69. Zwickl DJ, Hillis DM, Crandall K. Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol.* 2002;51(4):588–98. <https://doi.org/10.1080/10635150290102339>.
70. Dalquen DA, Zhu T, Yang Z. Maximum likelihood implementation of an isolation-with-migration model for three species. *Syst Biol.* 2016;66(3):379–98. <https://doi.org/10.1093/sysbio/syw063>.
71. Morales-Briones DF, Liston A, Tank DC. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the neotropical genus *Lachemilla* (Rosaceae). *New Phytol.* 2018;218(4):1668–84. <https://doi.org/10.1111/nph.15099>.
72. Olave M, Avila LJ, Sites JW, Morando M, Freckleton R. Detecting hybridization by likelihood calculation of gene tree extra lineages given

- explicit models. *Methods Ecol Evol.* 2017;9(1):121–33. <https://doi.org/10.1111/2041-210x.12846>.
73. Li JL, Zhang YJ, Ruhsam M, Milne RI, Wang Y, Wu DY, et al. Seeing through the hedge: Phylogenomics of Thuja (Cupressaceae) reveals prominent incomplete lineage sorting and ancient introgression for Tertiary relict flora. *Cladistics.* 2021;1–17. <https://doi.org/10.1111/cla.12491>.
  74. Flouri T, Jiao X, Rannala B, Yang Z, Yoder AD. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol Biol Evol.* 2018;35(10):2585–93. <https://doi.org/10.1093/molbev/msy147>.
  75. Doyle J, Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19:11–5.
  76. 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>.
  77. 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>.
  78. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
  79. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45:W6–11. <https://doi.org/10.1093/nar/gkx391>.
  80. 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>.
  81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
  82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
  83. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34(12):3299–302. <https://doi.org/10.1093/molbev/msx248>.
  84. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
  85. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16:111–20. <https://doi.org/10.1007/BF01731581>.
  86. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
  87. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 2016;34(3):772–3. <https://doi.org/10.1093/molbev/msw260>.
  88. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *bioRxiv.* 2019. <https://doi.org/10.1101/612903>.
  89. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>.
  90. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 2013;30(5):1188–95. <https://doi.org/10.1093/molbev/mst024>.
  91. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna 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>.
  92. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10(4):e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>.
  93. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67(5):901–4. <https://doi.org/10.1093/sysbio/syy032>.

## Publisher's Note

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