Received: 10 February 2021 | Revised: 22 June 2021 | Accepted: 5 July 2021

DOI: 10.1111/1755-0998.13479

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*

Chao-Nan Fu[1,2] | Zhi-Qiong Mo[1,3] | Jun-Bo Yang[2] | Jie Cai[2] | Lin-Jiang Ye[1,3] | Jia-Yun Zou[1,3] | Han-Tao Qin[1,3] | Wei Zheng[1,3] | Peter M. Hollingsworth[4] | De-Zhu Li[1,2,3] | Lian-Ming Gao[1,5]

[1]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

[2]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

[3]University of the Chinese Academy of Sciences, Beijing, China

[4]Royal Botanic Garden Edinburgh, Edinburgh, UK

[5]Lijiang Forest Ecosystem National Observation and Research Station, Kunming Institute of Botany, Chinese Academy of Sciences, Lijiang, Yunnan, China

Correspondence
Peter M. Hollingsworth, Royal Botanic Garden Edinburgh, Edinburgh, EH3 5LR, UK.
Email: PHollingsworth@rbge.org.uk

De-Zhu Li and Lian-Ming Gao, CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China.
Emails: dzl@mail.kib.ac.cn and gaolm@mail.kib.ac.cn

Funding information
Large-scale Scientific Facilities of the Chinese Academy of Sciences, Grant/Award Number: 2017-LSFGBOWS-02; National Natural Science Foundation of China, Grant/Award Number: 31670213 and 91631101; Program of Science and Technology Talents Training of Yunnan Province, China, Grant/Award Number: 2017HA014; Strategic Priority Research Program of Chinese Academy of Sciences, Grant/Award Number: XDB31000000; Special Research Assistant Funding Project of the Chinese Academy of Sciences, Grant/Award Number: E0295111Q1

Abstract

Standard plant DNA barcodes based on 2–3 plastid regions, and nrDNA ITS show variable levels of resolution, and fail to discriminate among species in many plant groups. Genome skimming to recover complete plastid genome sequences and nrDNA arrays has been proposed as a solution to address these resolution limitations. However, few studies have empirically tested what gains are achieved in practice. Of particular interest is whether adding substantially more plastid and nrDNA characters will lead to an increase in discriminatory power, or whether the resolution limitations of standard plant barcodes are fundamentally due to plastid genomes and nrDNA not tracking species boundaries. To address this, we used genome skimming to recover near-complete plastid genomes and nuclear ribosomal DNA from *Rhododendron* species and compared discrimination success with standard plant barcodes. We sampled 218 individuals representing 145 species of this species-rich and taxonomically difficult genus, focusing on the global biodiversity hotspots of the Himalaya-Hengduan Mountains. Only 33% of species were distinguished using ITS+*matK*+*rbcL*+*trnH*-*psbA*. In contrast, 55% of species were distinguished using plastid genome and nrDNA sequences. The vast majority of this increase is due to the additional plastid characters. Thus, despite previous studies showing an asymptote in discrimination success beyond 3–4 plastid regions, these results show that a demonstrable increase in discriminatory power is possible with extensive plastid genome data. However, despite these gains, many species remain unresolved, and these results also reinforce the need to access multiple unlinked nuclear loci to obtain transformative gains in species discrimination in plants.

Chao-Nan Fu and Zhi-Qiong Mo contributed equally to this study.

KEYWORDS
Himalaya-Hengduan Mountains, infrageneric phylogenetic resolution, next generation DNA barcoding, *Rhododendron*, species discrimination

1 | INTRODUCTION

Distinguishing among the world's species remains a pressing challenge in biology. Of the estimated 10 million eukaryotic species on earth, only c. 2 million have been scientifically described (Hebert et al., 2016; Mora et al., 2011). Even for species that have been described, identification of unknown specimens is often challenging where the differences between species are subtle and/or where there is a shortage of experts on the group in question (Mishra et al., 2016; Vohra & Khera, 2013). These problems are exacerbated if the material available is suboptimal in one way or another (e.g., fragmented tissue, processed tissue, juvenile/sterile specimens, or organism parts which lack diagnostic features).

DNA barcoding circumvents many of these challenges, and its deployment on a massive scale is greatly accelerating the characterization of species diversity (deWaard et al., 2019). In many animals, there is a surprisingly clear signature of interspecific differentiation detected by just a small portion of the mitochondrial genome, the 648 bp cytochrome oxidase 1 (CO1) barcode region (Hebert et al., 2003). Although there are cases where DNA barcodes do not provide species-level resolution in animals (Hebert et al., 2003; Vences et al., 2005)—overall there is remarkable concordance between data from the CO1 barcode and established species-limits in well studied groups such as butterflies (Burns et al., 2008; Janzen et al., 2009) and birds (Hebert et al., 2004; Kerr et al., 2007). In some other major groups (e.g., Hymenoptera and Diptera) there are more discrete sequence clusters from DNA barcodes than expected from the existing morphological classification, suggesting the presence of large numbers of previously undetected “cryptic” species (Hebert et al., 2016).

In plants the situation is more complicated, in part due to the slow substitution rates of plant organelle genomes and the high frequency of interspecific hybridization (Hollingsworth et al., 2011, 2016; Kress, 2017; Li et al., 2011). This results in the relatively frequent situation of plant species that are morphologically and ecologically distinct, being indistinguishable via standard DNA barcoding approaches. Attention is thus being given to design the next wave of plant DNA barcoding approaches, and suggestions range from mining the plastid genome for a series of taxon specific barcodes, genome skims recovering complete plastid genomes and nrDNA, through to targeted access to the nuclear genome via hybrid capture (Coissac et al., 2016; Dong et al., 2014; Hollingsworth et al., 2016; Li et al., 2015).

Desirable traits of any future barcoding approaches include cost-efficiency and scalability (to allow deployment on a massive scale), universality of approach (to enable synergies of different projects building and being able to query a common reference library), and a significant improvement in resolving power compared to the existing standard plant barcodes (Hollingsworth et al., 2016).

Genome skimming (i.e., shallow pass shotgun sequencing of c. 1–2 Gbp per sample) is an appealing approach given the straightforward and universal nature of its application (Coissac et al., 2016; Kane et al., 2012; Straub et al., 2012). Genome skimming works well with degraded DNAs, and benefits from the ever decreasing costs and improved efficiency of short-read sequencing technologies (Zeng et al., 2018). A shallow pass shotgun sequence routinely recovers complete (or near complete) plastid genomes, complete nuclear ribosomal DNA assemblages, and patchy coverage of varying depth of other fractions of the nuclear genome (Kane et al., 2012; Nock et al., 2011). There is a steady growth in the use of genome skimming in plants, and an associated development of bioinformatic tools and pipelines for data management and analyses (Fu et al., 2019; Ji et al., 2019; Jin et al., 2020; Song et al., 2020; Tonti-Filippini et al., 2017).

Despite this growth in genome skimming studies, there remains few studies that have empirically tested the resolution gains of this approach in terms of species discrimination (Fu et al., 2019; Ji et al., 2019). Most studies to-date have focused on a single individual per species (Bi et al., 2018). These studies reveal the extent of variation in plastid and ribosomal sequences, but not the extent to which this variation tracks species boundaries.

The degree to which plastid genome and nrDNA barcodes from genome skimming will make a material improvement in resolution is dependent on how often standard DNA barcoding fails due to a shortage of variable characters versus failure due to transpecific sharing of plastid and ribosomal haplotypes. The mechanisms underlying cytoplasmic introgression in plants are well characterized, and inter-specific transfer of "barcode containing" genomic regions is beyond refute (Rieseberg & Soltis, 1991). The impacts of introgressive hybridization on the sharing of barcodes among related species can be further exacerbated by selective sweeps, as in the case of *Salix* where 53 species from three subgenera share an identical barcode haplotype (Percy et al., 2014). However—a simple lack of variation between species is not uncommon in standard barcoding studies—and it remains possible that material gains in species resolution may still be generated, by simply obtaining more variable characters from complete plastid genomes and nrDNA assemblages.

To test the potential gains in species resolution in plant barcoding studies from genome skimming, we have used the study system of *Rhododendron* species in the Himalaya-Hengduan Mountains, to explore the gains in discriminatory power and phylogenetic resolution from plastid genome sequences and nrDNA assemblages, compared to standard plant barcodes.

Rhododendron (Ericaceae) is a species-rich and taxonomically difficult genus with over 1000 species globally (Chamberlain et al., 1996). It is the largest genus of flowering plants in China including ~590 species (Fang et al., 2005). The genus is divided into eight subgenera consisting of 12 sections and 59 subsections (24 subsections of *R.* subg. *Hymenanthes* and 35 of *R.* subg. *Rhododendron*, respectively) (Chamberlain et al., 1996) but some of the subgenera, sections and subsections are not supported as monophyletic in molecular phylogenetic studies (Brown et al., 2006; Gao et al., 2002; Goetsch et al., 2005; Kurashige et al., 2001; Shrestha et al., 2018). The Himalaya-Hengduan Mountains is a centre of diversity and diversification in the genus, with more than 320 species, of which about 66% are endemic to the region (Fang et al., 2005; Shrestha et al., 2018; Yan et al., 2015). Species from five of the eight subgenera are present, with most species (over 90%) belonging to the subgenera *Hymenanthes* and *Rhododendron* (Chamberlain et al., 1996; Fang et al., 2005).

More than one third species of *R.* sect. *Rhododendron* are polyploids (Ammal, 1950; Atkinson et al., 2000) and hybridization/introgression among sympatric species is considered frequent (Milne et al., 2010). Diversification in the *Rhododendron* species in the Himalaya-Hengduan mountains is associated with the relatively recent uplift of the Tibetan plateau and climate change during Neogene (Ding et al., 2020; Shrestha et al., 2018). Combined these attributes of frequent polyploidy, hybridization/introgression and recent speciation create a challenge for DNA barcoding (Hollingsworth et al., 2016; Percy et al., 2014; Yan et al., 2015). This is reflected in a recent study in which only 42% of *Rhododendron* species in the Himalaya-Hengduan Mountains were distinguished based on the combination of standard DNA barcodes (e.g. *matK*+*trnH-psbA*+ITS or *rbcL*+*matK*+*trnH-psbA* +ITS).

In this study we address the following questions: (1) Compared to standard DNA barcodes, do plastomes and nrDNA sequences improve intrageneric phylogenetic resolution and species identification in the species-rich and taxonomically difficult genus *Rhododendron*? (2) If so, what are the levels of increase in discriminatory power, and what insights does this provide into the nature of species differences in *Rhododendron*?

2 | MATERIALS AND METHODS

2.1 | Taxa sampling

A total of 218 individuals representing 145 *Rhododendron* species of four subgenera (*Hymenanthes*, *Rhododendron*, *Tsutsusi* and *Azaleastrum*), and eight sections (*Azaleastrum*, *Choniastrum*, *Ponticum*, *Pogonanthum*, *Rhododendron*, *Vireya*, *Tsutsusi*, *Brachycalyx*), occurring in the global biodiversity hotspots of the Himalaya-Hengduan Mountains were sampled in this study. This covers ~45% species of *Rhododendron* species in the region. Two to nine individuals per species (including ranks of subspecies and variety) were sampled for 42 species and the remaining 103 species had a single individual sampled. Two individuals of *Diplarche multiflora* were included as outgroups. Detailed information of sampling, classification and vouchers are provided in Table S1.

Healthy and fresh leaves were collected and dried immediately in silica gel for total genomic DNA extraction. Voucher specimens collected of each individual/species were deposited at the Herbarium of Kunming Institute of Botany (KUN), Chinese Academy of Sciences. All the samples were examined and identified by *Rhododendron* taxonomists at KUN.

2.2 | DNA extraction, sequencing, assembly and annotation

Total genomic DNA was extracted using a modified CTAB method (Doyle & Doyle, 1987), in which 4% CTAB was used with incorporation of 0.1% DL-dithiothreitol (DTT). DNA extracts were quantified and sheared into about 500 bp fragments for library construction using standard protocols (NEBNext Ultra IITMDNA Library Prep Kit for Illumina). Paired-end sequencing from both ends of 150 bp fragments was performed on the Illumina HiSeq X Ten platform at the BGI company in Wuhan, China, to generate ~2 Gbp data for each individual.

The plastome and nrDNA were de novo assembled using GetOrganelle pipeline (Jin et al., 2020). In this pipeline, plastome reads and nuclear reads were separately extracted from total genomic reads and subsequently assembled by SPADES v 3.10 (Bankevich et al., 2012). As the plastid genome in *Rhododendron* and many other Ericaceae species contain long repeats (except for the inverted repeats regions) (Fajardo et al., 2013; Li et al., 2020; Shen et al., 2020), it is extremely difficult to assemble the complete plastid genome from genome skimming sequences. Therefore, we produced plastid genome scaffolds. These scaffolds were annotated and checked using GENEIOUS v8.1 (Kearse et al., 2012), with comparison to the published plastome of *Rhododendron delavayi* Franch. (GenBank accession: NC_047438) as a reference. This reference genome was sequenced and assembled using data from the PacBio Sequel platform and the Illumina HiSeq 4000 (Liu et al., 2019). The nrDNAs were annotated using Geneious with *Aralia elata* (GenBank accession: KT380919) as the reference.

2.3 | Data analyses

The protein coding genes, rRNA genes and intergenic regions in annotated plastid genomes were separately extracted using a python script (https://github.com/Kinggerm/PersonalUtilities/blob/master/get_annotated_regions_from_gb.py). Each region was aligned using MAFFT v7.22 (Katoh & Standley, 2013) and manually modified in GENEIOUS. The ITS regions were extracted from nrDNAs in Geneious, and the nrDNA and ITS were both aligned by MAFFT. In order to compare the discrimination power of the plastid and nrDNA genomic barcodes, with standard DNA barcodes (including the best combinations of standard DNA barcode evaluated in a previous *Rhododendron* study (Yan et al., 2015)), we constructed eight data sets by concatenating different aligned regions, including (A) the concatenated coding genes, rRNA genes and intergenic regions of the plastid genome using the maximum coverage for each individual, (B) the plastid genome data from (A) but using only data recovered from all samples (i.e., regions with missing data in A are removed in B), (C) *rbcL+matK+trnH-psbA* barcoding regions, (D) the 18S-5.8S-26S nrDNA cistron including ITS1 and 2, (E) ITS consisting of ITS1-5.8S-ITS2, (F) ITS+*matK+trnH-psbA*, (G) ITS+*rbcL+matK+trnH-psbA*, and (H) data set A+data set D.

Two widely used methods, tree-based and distance-based analyses, were performed with the above data sets to evaluate species discrimination success. Phylogenetic analyses were conducted using maximum likelihood (ML) analysis using RAXML v8.1.11 (Stamatakis, 2006). The analyses were conducted using the GTR+Γ model, with the option of rapid bootstrap of 1000 replicates. Pairwise distance was calculated in MEGA X (Kumar et al., 2018) using the Kimura 2-parameter (k2p) model. The minimum interspecific distance for all species and the maximum intraspecific distance for species with multiple individuals were calculated using a python script.

3 | RESULTS

3.1 | Characteristics of data sets

As expected, the plastid genomes of all *Rhododendron* species here failed to assemble to a complete circular structure. However, a large amount of plastome sequence was assembled. A total of 72 protein coding genes, 4 rRNA genes, and 64 intergenic regions were assembled and annotated in the scaffold sequences (Table S2). Seven genes (*clpP*, *petB*, *petD*, *rpl16*, *rps12*, *ycf1*, *ycf2*) present in the complete reference plastome of *R. delavayi* were not recovered from any samples in this study. Combining all the plastid DNA regions, the maximum plastid genome data set had a length of 107,970 bp (data set A). Data set B consists of those regions recovered from all 220 sampled individuals including 47 protein coding genes, 2 rRNA genes, and 32 intergenic regions with a combined length of 41,374 bp. The nrDNA (data set D) was 5,969 bp in length, and data set H is combination of data set A + data set D. Various combinations of standard barcodes range in length from 733 bp (ITS) to 2,722 bp (ITS+*rbcL+matK+trnH-psbA*) (data sets C, E, F, G) (Table 1).

The plastid genome data set (data set A) (10,119 bp variable and 7,336 bp parsimony informative (PI) sites) and nrDNA (235 bp variable and 181 bp PI sites) contain many more variable sites and PI sites than the combination of standard DNA barcodes from the plastid genome (*matK+rbcL+trnH-psbA*; 240 bp variable and 189 bp PI sites) and ITS (125 bp variable and 103 bp PI sites). However, the standard DNA barcodes have a higher percentage of variable and PI sites than the other data sets, with the difference being most marked between ITS (with 17.05% variable and 14.05% PI sites) and the nrDNA assembly (with 3.94% variable and 3.03% PI sites) (Table 1).

3.2 | Phylogenetic resolution

Based on the data sets with only plastid data (data sets A–C), subgenera *Rhododendron*, *Hymenanthes* and *Tsutsusi* were resolved as monophyletic, whereas *R.* subg. *Azaleastrum* was resolved as paraphyletic with subg. *Tsutsusi* nested within it (Figure 1, Figure S1A–C). Based on the data sets with nuclear data, and both nuclear and plastid data (data sets D–H), the overall topology of monophyletic subgenera *Rhododendron*, *Hymenanthes* and *Tsutsusi* was retained (with subg. *Tsutsusi* nested within a paraphyletic subg. *Azaleastrum*), but surprisingly, the sampled individual of a species of *R.* subg. *Hymenanthes* (*R. wardii*) was resolved within *R.* subg. *Rhododendron* (Figure S1D–G) or resolved as sister group to all other species of subgenera *Hymenanthes* and *Rhododendron* in the combined plastome and nrDNA data set H (Figure S1H).

All sections (*Azaleastrum* and *Choniastrum* of *R.* subg. *Azaleastrum*; *Tsutsusi* and *Brachycalyx* of *R.* subg. *Tsutsusi*; *Ponticum* of *R.* subg. *Hymenanthes*; *Pogonanthum* and *Vireya* of *R.* subg. *Rhododendron*) except *Rhododendron* of *R.* subg. *Rhododendron* were resolved as monophyletic based on the plastid genome data sets (data sets A, B). *R.* sect *Rhododendron* was resolved as paraphyletic with sections *Pogonanthum* and *Vireya* embedded within it (Figure 1; Figure S1A–B). In the phylogenetic analyses with other

TABLE 1 Comparison of the characteristics of the alignments of different data sets

Alignment	Data set code	Length (bp)	Variable sites (%)	Parsimony informative sites (%)
Plastome	A	107,970	10,119 (9.37%)	7,336 (6.79%)
Plastome with no missing data	B	41,374	3,441 (8.32%)	2,501 (6.04%)
matK+rbcL+psbA-trnH	C	1,989	240 (12.07%)	189 (9.50%)
nrDNA (3S region)	D	5,969	235 (3.94%)	181 (3.03%)
ITS	E	733	125 (17.05%)	103 (14.05%)
ITS+*matK+psbA-trnH*	F	2,021	321 (15.88%)	258 (12.77%)
ITS+*matK+rbcL+psbA-trnH*	G	2,722	365 (13.41%)	292 (10.73%)
Plastome+nrDNA	H	113,939	10,359 (9.09%)	7,517 (6.60%)

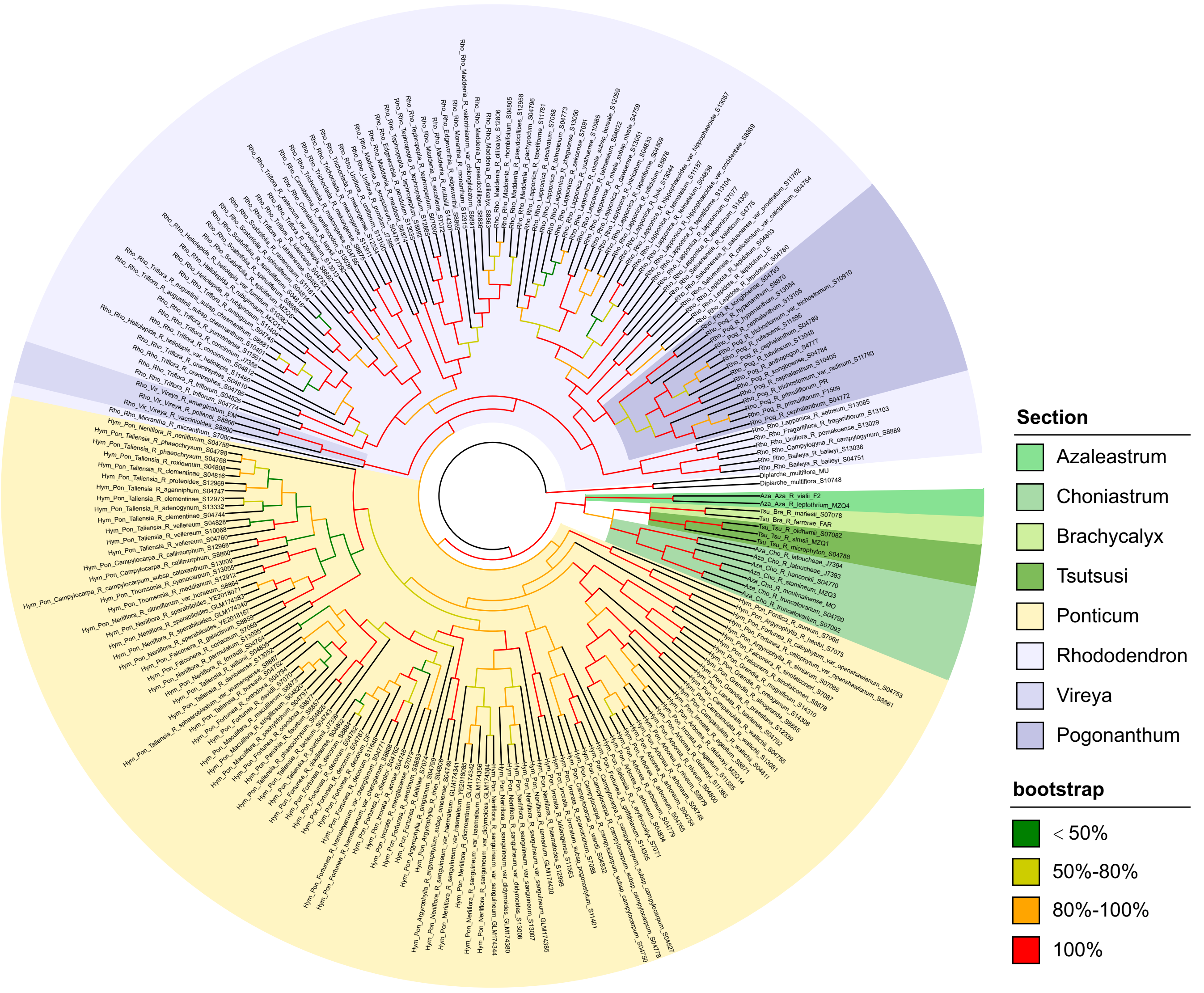

FIGURE 1 Maximum likelihood tree based on the plastid genome data set A. The background colours represent the eight sections, and the branch colours represent the bootstrap values

TABLE 2 Species discrimination success of eight data sets based on tree-based and distance-based methods

Data set	Data set code	Tree-based method[a]	Distance-based method[a]
Plastome	A	(22/42) 52.38%	(21/42) 50%
Plastome with no missing	B	(21/42) 50%	(22/42) 52.38%
matK+rbcL+*trnH-psbA*	C	(10/42) 23.81%	(10/42) 23.81%
nrDNA	D	(8/42) 19.05%	(10/42) 23.81%
ITS	E	(7/42) 16.67%	(7/42) 16.67%
ITS+matK+*trnH-psbA*	F	(12/42) 28.57%	(13/42) 30.95%
ITS+matK+*rbcL*+trnH-*psbA*	G	(14/42) 33.33%	(12/42) 30%
Plastome+nrDNA	H	(23/42) 54.76%	(23/42) 54.76%

[a]The data in the table reports only the discrimination success of species with >1 individual sampled per species. The success statistics in parentheses are the number of species discriminated out of the total tested.

data sets (data sets C–G), the results were similar but species of *R.* sect. *Pogonanthum* are mixed with species of the *R.* sect. *Rhododendron* clade (Figures S1C–F).

Only 10 of the 26 subsections of *R.* sect *Rhododendron* and *R.* sect *Ponticum* which had more than one sampled individual were resolved as monophyletic with the plastid genome data sets

(A and B). The success rate was much lower in other data sets (Figure 1; Figure S1C–G).

3.3 | Species discrimination

3.3.1 | Species discrimination based on phylogenetic analyses

In the phylogenetic analyses, a species was considered to be successfully identified when all conspecific individuals were resolved as monophyletic with a support value over 50%. The genomic barcodes showed much higher discriminatory power than different combinations of standard barcodes for the 42 species where multiple individuals were sampled. Among the eight data sets, the combined plastid genome and nrDNA (data set H) show the highest discriminatory power (55% of 42 species discriminated), followed by the plastid genome including missing data (data set A) (52%), then the plastid genome data set with no missing data (data set B) (50%), ITS+*matK*+*rbcL*+*trnH-psbA* (33%), ITS+*matK*+*trnH-psbA* (29%), *matK*+*rbcL*+*trnH-psbA* (24%), nrDNA (19%), and ITS alone showed the lowest resolution (17%) (Table 2). There is thus an increase in discriminatory power of 19%–22% with the genomic barcodes (data set H, 23 of 42 species resolved; data set A, 22 of 42 species resolved), compared to the best performing combination of standard barcodes (data set G, 14/42 species resolved).

It is noteworthy, that species with distinct named variants form a large portion of the "discrimination failures". Of the six species where our sampling included different named varieties and/or subspecies, all failed to resolve as a monophyletic clade in all data sets (Table S3).

3.3.2 | Species discrimination based on genetic distance

In the distance analyses, a species was considered to be successfully identified where its minimum interspecific distance is larger than its maximum intraspecific distance. The distance-based method showed a similar trend to the tree-based method. The combined plastid genome and nrDNA (data set H) show the highest discriminatory power (55% of 42 species discriminated), followed by the plastid genome data set B (no missing data) (52%), plastid genome data set A which included missing data (50%), ITS+*matK*+*trnH-psbA* (31%), ITS+*matK*+*rbcL*+*trnH-psbA* (30%), *matK*+*rbcL*+*trnH-psbA* (24%), nrDNA (24%), and finally the ITS data set (17%) (Table 2).

The interspecific genetic distance ranged from 0 to 0.0057 in combined plastid genome and nrDNA, and ranged from 0 to 0.0056 among all species in the plastid genome data sets (data set A). Two species pairs, *R. dichroanthum* versus *R. sanguineum*, and *R. maddenii* versus *R. scopulorum*, showed the minimum interspecific distance of zero in data set A, data set H and all the other data sets. A total of 14 species showed a minimum interspecific distance of zero in the plastid genome data set B, and 62 to 112 species showed the minimum interspecific distance of zero in data sets C–G (Table S4).

3.3.3 | Signal underlying the increase in species discrimination from the plastid genome data compared to standard plastid barcodes

Out of the 42 species with multiple individuals sampled, 12 species were successfully discriminated by the plastid genome (data set A), but not by the combination of standard plastid barcodes (*matK*+*rbcL*+*trnH-psbA*; data set C). This included six species from *R.* subg. *Hymenanthes*. For instance, all samples of *R. arboreum* and *R. niveum* were resolved as monophyletic, and distinguished from each other with 100% bootstrap support in data set A, whereas in the standard plastid barcode data set identical haplotypes were shared between all individuals of *R. arboreum* and *R. niveum*. For *R. hemsleyanum*, both sampled individuals of this species grouped as monophyletic with 100% bootstrap support in data set A, whereas this species shared an identical haplotype in the standard plastid barcode data set with some individuals of *R. decorum*, *R. discolor* and *R. qiaojiaense*. Similarly, all individuals of *R. sinofalconeri* formed a monophyletic clade in data set A (100% bootstrap support), whereas in the standard barcode data set the samples were resolved as paraphyletic with two species of *R.* subsect. *Grandia* (*R. oreogenum* and *R. praestans*). A similar pattern was also found for *R. vellereum* and *R. sperabiloides*. All samples of these species were resolved as monophyletic with data set A (100% bootstrap support), whereas they were resolved as paraphyletic in the standard plastid barcode data set.

Six species were also resolved in *R.* subg. *Rhododendron* with data set A but not by data set C. For *R. augustinii*, both individuals of this species were resolved as monophyletic in data set A (with 76% bootstrap support), whereas they shared identical haplotypes in the standard plastid barcode data set with those of *R. rubiginosum*, *R. concinnum* and *R. yunnanense*. All sampled individuals of *R. oreotrephes* and *R. rubiginosum* were resolved as monophyletic respectively in data set A (both with 100% bootstrap support), whereas with *matK*+*rbcL*+*trnH-psbA*, the individuals of the two species were resolved as paraphyletic with many species in *R.* subsections *Heliolepida* and *Triflora*. Similarly, all samples of *R. tephropeplum* formed a monophyletic clade with data set A (100% bootstrap support), whereas in the standard plastid barcode data set the samples were resolved as paraphyletic with *R. pendulum*. For *R. primuliflorum*, both sampled individuals of this species were resolved as monophyletic in data set A (with 93% bootstrap support), whereas with the standard plastid barcodes, they were mixed with species in *R.* sect. *Pogonanthum* (including samples with zero interspecific differences). All sampled individuals of *R. mekongense* were resolved as monophyletic with maximum support (100% support value) in data set A, but only 47% in data set C.

3.3.4 | Signal underlying the increase in species discrimination from the combined nuclear and plastid genome data

When the nrDNA data was combined with the plastome data (data set H), only one additional species was resolved (*R. oreodoxa*, 69% support value). This species was also identified (71% support value) in the plastid genome with no missing data (data set B), and weakly resolved by the ITS+*matK*+*trnH-psbA* with <50% bootstrap support (20%).

In contrast, two species (*R. delavayi* and *R. agastum*) were distinguishable with the combined plastid standard barcodes and ITS data sets (data set F and G), but not with the plastid genome data (data set A) or the combined data set (data set H). Furthermore, *R. delavayi* was resolved as monophyletic for ITS and the nrDNA data set, but not with any of the plastid data sets. However, *R. agastum* was not resolved as monophyletic for ITS, the wider nrDNA and any plastid data.

4 | DISCUSSION

4.1 | Discriminatory power of genome skim data versus standard barcodes

In the current study, data from standard barcode regions discriminated between 7 (ITS, 17%) and 14 (ITS+*rbcL*+*matK*+*trnH-psbA*, 33%) of the 42 species with >1 individual sampled. This range of 17%–33% of species being discriminated is similar to the findings in the *Rhododendron* barcoding study of Yan et al. (2015) where the maximum species discrimination from combinations of 3–4 barcode regions was 42%, and 12% from ITS.

In contrast, the discrimination from the genome skimming data was notably higher, with the plastid genome data set A discriminating 22 of the 42 species with >1 sampled individual, and bootstrap support for the monophyly of resolved species increased in all cases. The greater resolution of the standard plastid barcodes compared with ITS is mirrored by the genomic barcodes, with greater resolution of the plastid genome data set (22 species discriminated) compared to the nrDNA assembly (eight species discriminated). When the nrDNA data (data set D) was combined with the plastid genome data set (data set A), one additional species (*R. oreodoxa*) was resolved (Table S3).

Of the 12 species that were discriminated in data set A, but not with the standard plastid barcodes (data set C), the increased resolution comes from additional variable characters leading to species level monophyly in data set A, compared to the shortage of variable character with the standard barcoding regions. This increase in discrimination with the complete plastid genome data is noteworthy, in light of several studies which showed an asymptote in discriminatory power when small numbers of plastid regions are added (CBOL Plant Working Group, 2009; Fazekas et al., 2008; Li et al., 2011). Clearly there are likely to be different patterns of sequence variation among different taxonomic groups, but given the complexity of *Rhododendron* and its history of hybridization (Yan et al., 2015, 2017) and recent radiation (Milne et al., 2010; Shrestha et al., 2018), this is an encouraging result. It indicates material gains in species discrimination are possible by substantially increasing the number of variable plastid sites. It also suggests that aside from true plastid genome sharing among species (e.g., due to introgression), there is nevertheless a component of the discrimination challenge that is simply due to a shortage of variable sites.

In terms of the 20/42 species with multiple sampled individuals that failed to resolve with the plastome sequences in data set A, six were species where the sampled individuals consisted of different named varieties/subspecies. In one case (e.g., the different subspecies of *R. campylocarpum*), the individuals were resolved as phylogenetically disparate, suggesting that there may be an underlying taxonomic issue associated with how to treat these variants as opposed to a "lack of discriminatory power". However, the other species with named intraspecific variants did not show clear evidence of atypical disparate phylogenetic placements, and while an imperfect taxonomy could not be ruled out, the different intraspecific variants were no more phylogenetically dispersed than individuals in species without named intraspecific ranks.

Of the remaining 19 species with >1 sampled individual that were not resolved in data set A, one (*R. sanguineum*) showed some intraspecific variation, but some individuals of this species shared an identical haplotype with *R. dichroanthum*. One species (*R. oreodoxa*) showed minimum interspecific variation larger than the intraspecific variation but did not resolve as monophyletic. The other 17 species did not share identical haplotypes with other species, but their maximum intraspecific variation was larger than the minimum interspecific variation and they did not resolve as monophyletic (i.e., *R. agastum*, *R. cephalanthum*, *R. ciliicalyx*, *R. clementinae*, *R. R. concinnum*, *R. decorum*, *R. delavayi*, *R. heliolepis*, *R. hippophaeoides*, *R. hypenanthum*, *R. kongboense*, *R. nivale*, *R. phaeochrysum*, *R. pseudociliipes*, *R. tapetiforme*, *R. telmateium*, *R. trichostomum*).

4.2 | Potential reasons for species discrimination failure in *Rhododendron*

Rhododendron is one of the most taxonomically difficult and species-rich groups for species identification and phylogenetic inference due to its complex evolutionary history (Yan et al., 2015). Numerous species of *Rhododendron* occur sympatrically and undergo extensive interspecific hybridization and/or introgression (Ma et al., 2010; Milne et al., 2010; Yan et al., 2017, 2019; Zou et al., 2020). Hybridization/introgression can result in the sharing of maternally inherited plastid genomes between closely related species (Du et al., 2009), and thus the plastid genome may not track species boundaries (Hollingsworth et al., 2016; Petit & Excoffier, 2009). For example, *R. agastum* is verified as a natural hybrid between maternal *R. delavayi* and parental *R.*

irroratum (Zhang et al., 2007) or *R. decorum* (Zha et al., 2010), which likely explains why the plastid data fail to discriminate *R. agastum* and *R. delavayi* (Figures S1A, B). *R. dichroanthum* is another case of a species with hybrid origin with the inferred maternal parent being *R. sanguineum* (L. J. Ye, unpublished). As expected, these two species are indistinguishable using the plastid genome sequences. The plastid genome also failed to discriminate potential hybrid species in *Panax* (Ji et al., 2019). Interestingly, one sample of *R. wardii* of *R.* subg. *Hymenanthes* formed a monophyletic clade with species of this subgenus using the plastid genome data set (Figures S1A, B), but fell within the clade of *R.* subg. *Rhododendron* based on the nrDNA data sets (Figures S1D, E). It hints that interspecific hybridization may occur between the two subgenera. This is the first case of intersubgeneric natural hybridization in *Rhododendron* and warrants further investigation.

An additional indication of the importance of hybridization is the geographical signal in our data, with some samples grouping together by geographic proximity rather than taxonomic affinity, indicating gene flow among sympatric species. For example, one sample (S04822) of *R. telmateium*, grouped together with *R. nivale* subsp. *nivale* (S4759), but not with other samples of this species; with this grouping reflecting their shared location of Ganzhi, west Sichuan province. Similar geographical groupings of samples occurred with *R. nivale* subsp. *boreale* (S12059) and *R. yushuense* (S10985) from Yushu, Qinghai, and *R. intricatum* (S04833) and *R. dawuense* (S13051) from Ganzhi in Sichuan (Figure S1A).

In addition to hybridization, discriminating closely related species in recently radiated groups is an additional challenge for DNA barcoding (Coissac et al., 2016; Hollingsworth et al., 2016; Yan et al., 2015). Most species of subgenera *Hymenanthes* and *Rhododendron* are recently diversified in the Himalaya-Hengduan Mountains (Ding et al., 2020; Milne, 2004; Shrestha et al., 2018), and this is associated with a shortage of substitutions to distinguish closely related species. In this study, the phylogenetic trees of the species of *R.* subsect. *Lapponica* and *R.* sect. *Pogonanthum* displayed extremely short internal branch lengths among species (Figure S1, Table S4). Similarly, short branches also occur in subsections *Heliolepida*, *Triflora* of *R.* sect./subg. *Rhododendron*, and subsections *Fortunea*, *Neriiflora*, *Taliensia* of *R.* subg. *Hymenanthes*.

4.3 | Insights into subgeneric classification of *Rhododendron*

The genome skim data generated here also provides insights into the classification of *Rhododendron*. Three of the four subgenera were resolved as monophyletic, with *R.* subg. *Azaleastrum* resolving as two strongly supported separate monophyletic clades, corresponding to the two sections in the subgenus, supporting findings from previous studies (Gao et al., 2002, 2003; Goetsch et al., 2005; Kurashige et al., 2001; Shrestha et al., 2018). This supports the notion that these two sections (*Choniastrtum* and *Azaleatrum*) could be raised to the rank of subgenera (Gao et al., 2003; Kron & Judd, 1990). At the section level in the genus, all eight sections apart from *R.* sect. *Rhododendron* were resolved as monophyletic. *R.* sect. *Rhododendron* was paraphyletic with sect. *Pogonanthum* and *Vireya* nested within it, again matching results from previous studies (Goetsch et al., 2005; Kurashige et al., 2001; Shrestha et al., 2018; Yan et al., 2015). At the subsection level, the consistency with the existing classification was much lower. Only 10 out of the 34 sampled subsections (including eight subsections only sampled with a single species here) were resolved as monophyletic. The widespread non-monophyly of the subsections was also recovered by previous studies (Goetsch et al., 2005; Khan et al. 2020; Shrestha et al., 2018; Yan et al., 2015). The classification system of *Rhododendron* is mainly based on morphology (Chamberlain et al., 1996), and these studies and previous results suggest a revision of the subsections of *Rhododendron* may be warranted, although further evidence is required to establish a new stable classification at these lower taxonomic levels.

5 | CONCLUSIONS

Rhododendron represents a set of classical challenges for using DNA barcoding to discriminate among plant species, including a large number of closely related, co-occurring species, with evidence for a recent radiation and interspecific hybridization, resulting in associated taxonomic uncertainty. The current study has shown that a genome skimming approach to produce near-complete plastome and nrDNA sequence can provide more variation to discriminate *Rhododendron* species compared to standard barcodes. Although many species remain unresolved, there is a clear increase in discriminatory power, consistent with a shortage of variable characters being a rate limiting step for standard barcodes in the genus. With the decreasing costs and increasing ease of use of genome skimming, this illustrates the benefits to moving beyond standard barcodes in groups such as *Rhododendron* (Coissac et al., 2016; Hollingsworth et al., 2016; Li et al., 2015; Nevill et al., 2020; Tonti-Filippini et al., 2017; Yang et al., 2013). As noted elsewhere (Coissac et al., 2016), as the resulting plastome and nrDNA data are compatible with previous barcode data sets, this is a pragmatic augmentation of existing standard barcodes, rather than replacement. And given that genome skimming data also recovers the standard DNA barcode regions, this approach will continue to enrich the reference database of standard plant barcodes (Coissac et al., 2016).

Data from the nuclear genome will clearly be required to fully understand and resolve species limits in *Rhododendron*. Nuclear genomes of *R. delavayi*, *R. williamsianum* and *R. simsii* have recently been published (Soza et al., 2019; Yang et al., 2020; Zhang et al., 2017), and we have used these data alongside unpublished transcriptome data of *Rhododendron*, to design probes to assay variation in the nuclear genome using the target capture method (Nicholls et al., 2015; Senapathy et al., 2010). A future study will evaluate the efficacy of these newly developed nuclear markers in discriminating species in this most challenging genus of plants.

ACKNOWLEDGEMENTS

This study was supported by the Large-scale Scientific Facilities of the Chinese Academy of Sciences (2017-LSFGBOWS-02), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000), the National Natural Science Foundation of China (91631101, 31670213, 32000173), the Program of Science and Technology Talents Training of Yunnan Province, China (2017HA014), and the Special Research Assistant Funding Project of the Chinese Academy of Sciences (E0295111Q1). Molecular experiments and data analyses were performed at the Laboratory of Molecular Biology and iFlora High Performance Computing Centre of Germplasm Bank of Wild Species in Southwest China (GBOWS) respectively, Kunming Institute of Botany. PMH acknowledges funding support from the Scottish Government's Rural and Environment Science and Analytical Services Division.

DATA AVAILABILITY STATEMENT

Sequence alignments and all trees of Data set A-H are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.9s4mw6mgv.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

ORCID

Chao-Nan Fu https://orcid.org/0000-0001-7567-2686
Lian-Ming Gao https://orcid.org/0000-0001-9047-2658

REFERENCES

Ammal, E. (1950). Polyploidy in the genus *Rhododendron*. *The Rhododendron Year Book*, *5*, 92–98.

Atkinson, R., Jong, K., & Argent, G. (2000). Chromosome numbers of some tropical Rhododendrons (section *Vireya*). *Edinburgh Journal of Botany*, *57*(1), 1–7.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPADES: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bi, Y., Zhang, M. F., Xue, J., Dong, R., Du, Y. P., & Zhang, X. H. (2018). Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Scientific Reports*, *8*, 1184. https://doi.org/10.1038/s41598-018-19591-9

Brown, G. K., Craven, L. A., Udovicic, F., & Ladiges, P. Y. (2006). Phylogenetic relationships of *Rhododendron* section *Vireya* (Ericaceae) inferred from the ITS nrDNA region. *Australian Systematic Botany*, *19*(4), 329–342. https://doi.org/10.1071/SB05019

Burns, J. M., Janzen, D. H., Hajibabaei, M., Hallwachs, W., & Hebert, P. D. (2008). DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(17), 6350–6355. https://doi.org/10.1073/pnas.0712181105

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chamberlain, D., Hyam, R., Argent, G., Fairweather, G., & Walter, K. S. (1996). *The genus* Rhododendron*: its classification and synonymy*. :Royal Botanic Garden Edinburgh.

Coissac, E., Hollingsworth, P. M., Lavergne, S., & Taberlet, P. (2016). From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology*, *25*(7), 1423–1428. https://doi.org/10.1111/mec.13549

deWaard, J. R., Ratnasingham, S., Zakharov, E. V. et al (2019). A reference library for Canadian invertebrates with 1.5 million barcodes, voucher specimens, and DNA samples. *Scientific Data*, *6*(1), 308.

Ding, W. N., Ree, R. H., Spicer, R. A., & Xing, Y. W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581.

Dong, W., Liu, H., Xu, C., Zuo, Y., Chen, Z., & Zhou, S. (2014). A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genetics*, *15*, 138. https://doi.org/10.1186/s12863-014-0138-z

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, *19*, 11–15.

Du, F. K., Petit, R. J., & Liu, J. Q. (2009). More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Molecular Ecology*, *18*(7), 1396–1407.

Fajardo, D., Senalik, D., Ames, M., Zhu, H., Steffan, S. A., Harbut, R., Polashock, J., Vorsa, N., Gillespie, E., Kron, K., & Zalapa, J. E. (2013). Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics & Genomes*, *9*(2), 489–498. https://doi.org/10.1007/s11295-012-0573-9

Fang, M. Y., Fang, R. Z., He, M. Y., Hu, L. Z., Yang, H. B., & Chamberlain, D. F. (2005). *Rhododendron*, Vol. *14*. Science Press & Missouri Botanical Garden Press.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Percy, D. M., Hajibabaei, M., & Barrett, S. C. H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One*, *3*(7), e2802. https://doi.org/10.1371/journal.pone.0002802

Fu, C. N., Wu, C. S., Ye, L. J., Mo, Z. Q., Liu, J., Chang, Y. W., Li, D. Z., Chaw, S. M., & Gao, L. M. (2019). Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Scientific Reports*, *9*(1), 2773. https://doi.org/10.1038/s41598-019-39161-x

Gao, L. M., Li, D. Z., & Zhang, C. Q. (2003). Phylogenetic relationships of *Rhododendron* section *Azaleastrum* (Ericaceae) based on ITS sequences. *Acta Phytotaxonomica Sinica*, *41*(2), 173–179.

Gao, L. M., Li, D. Z., Zhang, C. Q., & Yang, J. B. (2002). Infrageneric and sectional relationships in the genus *Rhododendron* (Ericaceae) inferred from ITS sequence data. *Journal of Integrative Plant Biology*, *44*(11), 1351–1356.

Goetsch, L., Eckert, A. J., Hall, B. D., & Hoot, S. B. (2005). The molecular systematics of *Rhododendron* (Ericaceae): A phylogeny based upon RPB2 gene sequences. *Systematic Botany*, *30*(3), 616–626.

Hebert, P. D., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270 Suppl 1*, S96–99.

Hebert, P. D., Ratnasingham, S., Zakharov, E. V. et al (2016). Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 20150333.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, *2*(10), e312. https://doi.org/10.1371/journal.pbio.0020312

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One*, *6*(5), e19254. https://doi.org/10.1371/journal.pone.0019254

Hollingsworth, P. M., Li, D. Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 20150338.

Janzen, D. H., Hallwachs, W., Blandin, P., Burns, J. M., Cadiou, J.-M., Chacon, I., Dapkey, T., Deans, A. R., Epstein, M. E., Espinoza, B., Franclemont, J. G., Haber, W. A., Hajibabaei, M., Hall, J. P. W., Hebert, P. D. N., Gauld, I. D., Harvey, D. J., Hausmann, A., Kitching, I. J., … Wilson, J. J. (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, *9* (Suppl s1), 1–26. https://doi.org/10.1111/j.1755-0998.2009.02628.x

Ji, Y., Liu, C., Yang, Z. et al (2019). Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Molecular Ecology Resources*, *19*(5), 1333–1345.

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y. U., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02154-5

Kane, N., Sveinsson, S., Dempewolf, H. et al (2012). Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, *99*(2), 320–329.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). GENEIOUS BASIC: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kerr, K. C. R., Stoeckle, M. Y., Dove, C. J., Weigt, L. A., Francis, C. M., & Hebert, P. D. N. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, *7*(4), 535–543. https://doi.org/10.1111/j.1471-8286.2007.01670.x

Khan, G., Nolzen, J., Schepker, H., & Albach, D.C. (2020). Incongruent phylogenies and its implications for the study of diversification, taxonomy and genome size evolution of *Rhododendron* (Ericaceae). *bioRxiv*.

Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future. *Journal of Systematics and Evolution*, *55*(4), 291–307. https://doi.org/10.1111/jse.12254

Kron, K. A., & Judd, W. S. (1990). Phylogenetic relationships within the Rhodoreae (Ericaceae) with specific comments on the placement of *Ledum*. *Systematic Botany*, *15*(1), 57–68. https://doi.org/10.2307/2419016

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kurashige, Y., Etoh, J. I., Handa, T., Takayanagi, K., & Yukawa, T. (2001). Sectional relationships in the genus *Rhododendron* (Ericaceae): evidence from *matK* and *trnK* intron sequences. *Plant Systematics and Evolution*, *228*(1–2), 1–14. https://doi.org/10.1007/s006060170033

Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., Liu, J. Q., Chen, Z. D., Zhou, S. L., Chen, S. L., Yang, J. B., Fu, C. X., Zeng, C. X., Yan, H. F., Zhu, Y. J., Sun, Y. S., Chen, S. Y., Zhao, L., Wang, K., Yang, T., & Duan, G. W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(49), 19641–19646. https://doi.org/10.1073/pnas.1104551108

Li, H., Guo, Q., Li, Q., & Yang, L. (2020). Long-reads reveal that *Rhododendron* delavayi plastid genome contains extensive repeat sequences, and recombination exists among plastid genomes of photosynthetic Ericaceae. *PeerJ*, *8*, e9048.

Li, X. W., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y. T., & Chen, S. L. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*(1), 157–166. https://doi.org/10.1111/brv.12104

Liu, J., Chen, T., Zhang, Y., Li, Y., Gong, J., & Yi, Y. (2019). The complete chloroplast genome of *Rhododendron delavayi* (Ericaceae). *Mitochondrial DNA Part B*, *5*(1), 37–38.

Ma, Y., Milne, R. I., Zhang, C., & Yang, J. (2010). Unusual patterns of hybridization involving a narrow endemic *Rhododendron* species (Ericaceae) in Yunnan. *China. American Journal of Botany*, *97*(10), 1749–1757.

Milne, R. I. (2004). Phylogeny and biogeography of *Rhododendron* subsection *Pontica*, a group with a tertiary relict distribution. *Molecular Phylogenetics and Evolution*, *33*(2), 389–401. https://doi.org/10.1016/j.ympev.2004.06.009

Milne, R. I., Davies, C., Prickett, R., Inns, L. H., & Chamberlain, D. F. (2010). Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: between-lineage hybridisation during adaptive radiation? *Plant Systematics and Evolution*, *285*(3–4), 233–244. https://doi.org/10.1007/s00606-010-0269-2

Mishra, P., Kumar, A., Nagireddy, A. et al (2016). DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal*, *14*(1), 8–21.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biology*, *9*(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Nevill, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., van Leeuwen, S., Boykin, L. M., & Small, I. (2020). Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods*, *16*(1), 1. https://doi.org/10.1186/s13007-019-0534-5

Nicholls, J. A., Pennington, R. T., Koenen, E. J. M., Hughes, C. E., Hearn, J., Bunnefeld, L., Dexter, K. G., Stone, G. N., & Kidner, C. A. (2015). Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science*, *6*, 710. https://doi.org/10.3389/fpls.2015.00710

Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., & Henry, R. J. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, *9*(3), 328–333. https://doi.org/10.1111/j.1467-7652.2010.00558.x

Percy, D. M., Argus, G. W., Cronk, Q. C. et al (2014). Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? *Molecular Ecology*, *23*(19), 4737–4756.

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, *24*(7), 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Rieseberg, L. H., & Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, *5*(1), 65–84.

Senapathy, P., Bhasi, A., Mattox, J., Dhandapany, P. S., & Sadayappan, S. (2010). Targeted genome-wide enrichment of functional regions. *PLoS One*, *5*(6), e11138. https://doi.org/10.1371/journal.pone.0011138

Shen, J. S., Li, X. Q., Zhu, X. T., Huang, X. L., & Jin, S. H. (2020). The complete plastid genome of *Rhododendron pulchrum* and comparative genetic analysis of Ericaceae species. *Forests*, *11*(2), 158. https://doi.org/10.3390/f11020158

Shrestha, N., Wang, Z., Su, X. et al (2018). Global patterns of *Rhododendron* diversity: The role of evolutionary time and diversification rates. *Global Ecology and Biogeography*, *27*(8), 913–924.

Song, F., Li, T., Burgess, K. S., Feng, Y., & Ge, X. J. (2020). Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. *BMC Plant Biology*, *20*(1), 261. https://doi.org/10.1186/s12870-020-02466-5

Soza, V. L., Lindsley, D., Waalkes, A., Ramage, E., Patwardhan, R. P., Burton, J. N., Adey, A., Kumar, A., Qiu, R., Shendure, J., & Hall, B. (2019). The *Rhododendron* genome and chromosomal organization provide insight into shared whole-genome duplications across the heath family (Ericaceae). *Genome Biology and Evolution*, *11*(12), 3353–3371. https://doi.org/10.1093/gbe/evz245

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, *99*(2), 349–364. https://doi.org/10.3732/ajb.1100335

Tonti-Filippini, J., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes? *The Plant Journal*, *90*(4), 808–818. https://doi.org/10.1111/tpj.13491

Vences, M., Thomas, M., Bonett, R. M., & Vieites, D. R. (2005). Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1859–1868.

Vohra, P., & Khera, K. S. (2013). DNA barcoding: current advance and future prospects—a review. *Asian Journal of Biological and Life Sciences*, *2*(3), 185–189.

Yan, L. J., Burgess, K. S., Milne, R., Fu, C. N., Li, D. Z., & Gao, L. M. (2017). Asymmetrical natural hybridization varies among hybrid swarms between two diploid *Rhododendron* species. *Annals of Botany*, *120*(1), 51–61. https://doi.org/10.1093/aob/mcx039

Yan, L. J., Burgess, K. S., Zheng, W., Tao, Z. B., Li, D. Z., & Gao, L. M. (2019). Incomplete reproductive isolation between *Rhododendron* taxa enables hybrid formation and persistence. *Journal of Integrative Plant Biology*, *61*(4), 433–448.

Yan, L. J., Liu, J., Möller, M. et al. (2015). DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. *Molecular Ecology Resources*, *15*(4), 932–944.

Yang, F. S., Nie, S., Liu, H., Shi, T. L., Tian, X. C., Zhou, S. S., Bao, Y. T., Jia, K. H., Guo, J. F., Zhao, W., An, N. A., Zhang, R. G., Yun, Q. Z., Wang, X. Z., Mannapperuma, C., Porth, I., El-Kassaby, Y. A., Street, N. R., Wang, X. R., … Mao, J. F. (2020). Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nature Communications*, *11*(1), 5269. https://doi.org/10.1038/s41467-020-18771-4

Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., & Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology*, *13*, 84. https://doi.org/10.1186/1471-2148-13-84

Zeng, C. X., Hollingsworth, P. M., Yang, J., He, Z. S., Zhang, Z. R., Li, D. Z., & Yang, J. B. (2018). Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods*, *14*, 43. https://doi.org/10.1186/s13007-018-0300-0

Zha, H. G., Milne, R. I., & Sun, H. (2010). Asymmetric hybridization in *Rhododendron agastum*: a hybrid taxon comprising mainly F1s in Yunnan, China. *Annals of Botany*, *105*(1), 89–100. https://doi.org/10.1093/aob/mcp267

Zhang, J. L., Zhang, C. Q., Gao, L. M., Yang, J. B., & Li, H. T. (2007). Natural hybridization origin of *Rhododendron agastum* (Ericaceae) in Yunnan, China: inferred from morphological and molecular evidence. *Journal of Plant Research*, *120*(3), 457–463. https://doi.org/10.1007/s10265-007-0076-1

Zhang, L., Xu, P., Cai, Y. et al (2017). The draft genome assembly of *Rhododendron delavayi* Franch. var. delavayi. *Gigascience*, *6*(10), 1–11.

Zou, J. Y., Luo, Y. H., Burgess, K. S. et al (2020). Joint effect of phylogenetic relatedness and trait selection on the elevational distribution of *Rhododendron* species. *Journal of Systematics and Evolution*, https://doi.org/10.1111/jse.12690

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Fu, C.-N., Mo, Z.-Q., Yang, J.-B., Cai, J., Ye, L.-J., Zou, J.-Y., Qin, H.-T., Zheng, W., Hollingsworth, P. M., Li, D.-Z., & Gao, L.-M. (2021). Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*. *Molecular Ecology Resources*, 00, 1–11. https://doi.org/10.1111/1755-0998.13479