Can plastid genome sequencing be used for species identification in Lauraceae?

ZHI-FANG LIU[1,2,13], HUI MA[1], XIU-QIN CI[1,3], LANG LI[1,3], YU SONG[4], BING LIU[5,6], HSI-WEN LI[7], SHU-LI WANG[1,2,8], XIAO-JIAN QU[9], JIAN-LIN HU[1,2], XIAO-YAN ZHANG[1,2], JOHN G. CONRAN[10], ALEX D. TWYFORD[11,*], JUN-BO YANG[12,*], PETER M. HOLLINGSWORTH[13,*] and JIE LI[1,3,*]

[1]*Plant Phylogenetics and Conservation Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China*
[2]*University of Chinese Academy of Sciences, Beijing, China*
[3]*Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla, China*
[4]*Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China*
[5]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China*
[6]*Sino-African Joint Research Center, Chinese Academy of Sciences, Wuhan, China*
[7]*Herbarium (KUN), Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China*
[8]*Tibet Agriculture & Animal Husbandry University, Nyingchi, China*
[9]*Shandong Provincial Key Laboratory of Plant Stress Research, College of Life Sciences, Shandong Normal University, Ji'nan, China*
[10]*Australian Centre for Evolutionary Biology and Biodiversity & Sprigg Geobiology Centre, School of Biological Sciences, University of Adelaide, Adelaide, Australia*
[11]*Institute of Evolutionary Biology, Ashworth Laboratories, The University of Edinburgh, Edinburgh, UK*
[12]*Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China*
[13]*Genetics and Conservation Section, Royal Botanic Garden Edinburgh, Edinburgh, UK*

Using DNA barcoding for species identification remains challenging for many plant groups. New sequencing approaches such as complete plastid genome sequencing may provide some increased power and practical benefits for species identification beyond standard plant DNA barcodes. We undertook a case study comparing standard DNA barcoding to plastid genome sequencing for species discrimination in the ecologically and economically important family Lauraceae, using 191 plastid genomes for 131 species from 25 genera, representing the largest plastome data set for Lauraceae to date. We found that the plastome sequences were useful in correcting some identification errors and for finding new and cryptic species. However, plastome data overall were only able to discriminate *c.* 60% of the species in our sample, with this representing a modest improvement from 40 to 50% discrimination success with the standard plant DNA barcodes. Beyond species discrimination, the plastid genome sequences revealed complex relationships in the family, with 12/25 genera being non-monophyletic and with extensive incongruence relative to nuclear ribosomal DNA. These results highlight that although useful for improving phylogenetic resolution in the family and providing some species-level insights, plastome sequences only partially improve species discrimination, and this reinforces the need for large-scale nuclear data to improve discrimination among closely related species.

ADDITIONAL KEYWORDS: cytonuclear discordance – DNA barcoding – nrDNA – phylogenetics – plastomes.

*Corresponding authors. E-mails: jieli@xtbg.ac.cn; Alex. Twyford@ed.ac.uk; jbyang@mail.kib.ac.cn; PHollingsworth@ rbge.org.uk.

INTRODUCTION

The aim of DNA barcoding is to use standardized DNA sequences to aid in species identification (Hebert *et al.*, 2003; Hollingsworth, 2011). Various regions have been proposed for DNA barcoding in plants (Kress & Erickson, 2007; Lahaye *et al.*, 2008), with the plastid loci *matK*+*rbcL* adopted as core DNA barcodes (CBOL Plant Working Group, 2009), with these loci now widely used alongside the nuclear region ITS and other plastid loci such as *trnH–psbA* (Chase *et al.*, 2005; Kress *et al.*, 2005; China Plant BOL Group, 2011; Hollingsworth, Graham & Little, 2011). Despite many benefits of using these standardized loci for plant DNA barcoding, it has long been recognized that no single suite of loci will be suitable across all plant taxa. In groups where standard DNA barcoding 'fails' due to technical issues, such as mutations in primer-binding regions, or biological issues, such as rapid divergence, researchers must augment the standard loci with additional sequence data. The massive improvements in genomic sequencing technologies allow researchers to explore many possible options for 'genomic DNA barcodes' to improve plant species identification.

Whole plastid genomes (plastomes) have been proposed as suitable targets for the next wave of plant DNA barcoding approaches (Kane *et al.*, 2012; Ruhsam *et al.*, 2015; Hollingsworth *et al.*, 2016; Twyford & Ness, 2017; Krawczyk *et al.*, 2018), as they can be recovered using 'genome skimming' (low-coverage whole genome sequencing), a cost-effective and scalable sequencing approach that can be performed on a range of material (such as degraded herbarium samples) without prior sample optimization (Alsos *et al.*, 2020). As well as plastome sequences, genome skimming also typically recovers the nuclear ribosomal DNA (nrDNA) assembly, collectively extending the plant barcode from a few thousand to hundreds of thousands of bases. Genome skimming also helps to circumvent primer issues, polymerase chain reaction (PCR) failures from amplicon sequencing of degraded DNA and different loci being preferred for different taxonomic groups, because shotgun sequencing is effective in routinely recovering plastid loci and nuclear ribosomal sequences due to their high copy numbers in plant cells (Coissac *et al.*, 2016). However, there remain drawbacks to this approach for DNA barcoding of plants, as the plastid is a small organelle in which all loci are tightly linked and it therefore does not necessarily reflect the diverse history of the nuclear genome. This is compounded by the typically uniparental inheritance of the plastid genome and resulting sensitivity to demographic change. Similarly, the associated nrDNA, although a useful additional source of characters, still represents a fraction of the nuclear genome and can be problematic for species identification and phylogenetic analysis due

to a range of issues, such as high diversity affecting the reliability of sequence alignments, incomplete concerted evolution and frequent paralogy. These issues create a tension between the technical appeal of plastomes and nrDNA in terms of ease of use, vs. their suitability as representatives of the evolutionary history of a species. Although the discriminatory power of next-generation DNA barcoding in plants has been evaluated in some recent studies (e.g. Kane *et al.*, 2012; Li *et al.*, 2015; Ruhsam *et al.*, 2015; Ji *et al.*, 2019; Pang *et al.*, 2019), few studies have undertaken direct comparisons between the suitability of standard DNA barcodes and plastomes for genus and species identification using multiple individuals per species from a specific family.

Lauraceae were established by de Jussieu (1789) in his *Genera Plantarum*, based on the type genus *Laurus* L. (Linnaeus, 1753). Lauraceae are evergreen or sometimes deciduous shrubs or trees (except for *Cassytha* L., which is a twining, virtually leafless, parasitic perennial vine), often with aromatic bark and foliage (Chanderbali, van der Werff & Renner, 2001; Li *et al.*, 2008a). Lauraceae comprise *c*. 50 genera and 2500–3000 species from predominantly tropical and subtropical regions (van der Werff & Richter, 1996), and they are most diverse in tropical Asia, tropical America and Madagascar (Gentry, 1988; van der Werff & Richter 1996; Li *et al.*, 2008a). They are economically and ecologically important as sources of medicines, timber, fruits, spices and perfumes (Kostermans, 1957; van der Werff & Richter, 1996; Li *et al.*, 2008a), and are present in wet forests at any elevation and are frequently forest dominants (van der Werff & Richter, 1996). Nevertheless, despite their importance, the classification of Lauraceae is poorly known (van der Werff & Richter, 1996) and their broad-scale classification has depended traditionally on the morphology of inflorescences and flowers (Nees von Esenbeck, 1836; Rohwer, 1993; van der Werff & Richter, 1996), although many groups have species that show exceptions. For example, whereas most Lauraceae flowers are regular with three whorls, groups such as Laureae have flowers that are frequently irregular, sometimes with more than three whorls of fertile stamens (Rohwer, 1993). Vegetative morphological similarities between taxa and intra-taxon variability are also causes of taxonomic confusion.

Previous phylogenetic studies have used various plastid (*matK*, *trnK*, *trnL–trnF*, *psbA–trnH*, *trnT–trnL*, *rps16*) and nuclear regions (26S, *RPB2*, *LEAFY* and ITS) to study relationships in Lauraceae (Rohwer, 2000; Chanderbali *et al.*, 2001; Rohwer & Rudolph, 2005; Li *et al.*, 2008a, 2011; Rohwer *et al.*, 2009; Huang *et al.*, 2016; Mo *et al.*, 2017). Recently, Song *et al.* (2020) showed that expanding to whole plastome

sequences can produce better resolved evolutionary relationships than Sanger sequencing of a few key loci. Although overall relationships among Lauraceae are mostly well known, species relationships in genera are still poorly understood. Most studies to date have only sampled a single individual per taxon, and sampling multiple individuals per species across diverse taxa would allow us to test for species-level monophyly and discrimination (Ji *et al.*, 2019). Our previous study (Liu *et al.*, 2017) used standard DNA plant barcodes to resolve Lauraceae relationships and classification by sampling multiple individuals per species; however, the resolution was poor.

Accordingly, here we investigate whether the plastome can improve species discrimination relative to standard DNA barcodes in Lauraceae. Specifically, we have four aims: (1) to determine if plastome-based DNA barcodes improve taxonomic resolution and the potential for species identification compared to standard DNA barcodes; (2) to establish whether some proposed 'Lauraceae-specific' plastid barcodes provide useful information for species discrimination; (3) to relate genetic to morphological data to detect cases of mistaken identity and facilitate species discovery; and (4) to reconstruct phylogenetic relationships in Lauraceae and see if plastid genomes match the species boundaries determined from analysis of nrDNA sequences.

MATERIAL AND METHODS

SAMPLING AND SEQUENCING

Our data set consists of plastid genomes and nrDNA from 80 *de novo* genome skims, augmented with plastid genome-only data from GenBank and LCGDB (https://lcgdb.wordpress.com) for a further 111 individuals (last search 3 September 2019). The resulting 191 plastome samples represented 133 species, 131 of which represent 25 of the 55 currently recognized

genera of Lauraceae and two of which represent the outgroup family Calycanthaceae.

The 191 samples included 101 species with $N = 1$ sample, and a further 90 individuals from 34 species with $N > 1$ individual plastid genome sampled (mean three individuals per species, range two to five) from 16 genera. At the genus level, 25 genera were sampled, with 21 genera with more than one individual species sampled. A summary of the data set generated for this study is shown in Table 1 with details of the taxa sampled in the Supporting Information (Table S1). From the complete plastid genomes, we could also subsample regions for comparative analyses of species discrimination. For these analyses we compared (1) the complete plastid genome, (2) the gene regions from the plastid genomes, and (3) the standard DNA barcode matrix (*rbcL*+*matK*+*trnH–psbA*) and some barcode regions proposed as being useful for the family, e.g. Lauraceae-specific barcodes (*ycf1*+*ndhH–rps15*+*trnL–ycf2*; for further information see Note S1).

The 80 *de novo* plastid genomes (including one individual sequenced twice) were derived from samples collected from eight provinces in China (Guangdong, Guangxi, Hainan, Hubei, Jiangxi, Sichuan, Yunnan and Xizang) and sites in Japan and Myanmar (Supporting Information, Table S2). Leaf tissue for each taxon was dried with silica gel and vouchers were deposited at the Herbarium of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (HITBC), Yunnan, China. The specimens and vouchers were identified by morphological and molecular comparisons as described previously (Liu *et al.*, 2017).

Total genomic DNA was extracted using a modified CTAB method (Doyle & Doyle, 1987) with a Tiangen DNA secure Plant Kit (DP320). Yield and integrity of genomic DNA extracts were quantified by fluorometric quantification on the Qubit (Invitrogen, Carlsbad, CA, USA) using the dsDNA HS kit and by visual assessment on a 1% agarose gel. The extracted DNA was sheared

Table 1. Comparison of characteristics of different data sets in Lauraceae

Data sets	Subsets	Number of taxa	Number of species	Number of genera	Number of sites	Best fit model of ML analysis
191 plastomes data set	Plastomes	191	131	25	180 129	TVM+F+I+G4
	Concatenated genes	191	131	25	79 387	GTR+F+I+G4
	Extracted specific matrix	191	129	24	2704	TVM+F+G4
	Extracted standard matrix	191	131	25	2144	K3Pu+F+I+G4
Cytonuclear discordance data set	80 plastomes	79	71	21	172 022	TVM+F+I+G4
	80 rDNA	79	71	21	6281	TN+F+I+G4

into *c.* 500-bp fragments for library construction using the standard protocol for the NEBNext Ultra IITMDNA Library Prep Kit for Illumina. All samples were sequenced on the Illumina HiSeq 2000 and Illumina HiSeq X at BGI and Novogene (Supporting Information, Table S2).

ASSEMBLY, ANNOTATION AND GENE SUBSAMPLING

GetOrganelle (Jin *et al.*, 2018) was used for assembly of plastomes and nrDNA (18S–ITS1–5.8S–ITS2–26S). GetOrganelle uses baiting and iterative mapping to assemble plastomes with minimal manual intervention and integrates SPAdes (Bankevich *et al.*, 2012), Bowtie2 (Langmead & Salzberg, 2012), BLAST+ (Camacho *et al.*, 2009) and Bandage (Wick *et al.*, 2015). Comparison of the published plastomes for Lauraceae (Song *et al.*, 2015, 2016, 2017a, b, 2018, 2020; Hinsinger & Strijk, 2017; Wu *et al.*, 2017; Zhao *et al.*, 2018) led us to choose *Litsea glutinosa* (Lour.) C.B.Rob (KU382356) as the plastid genome reference for assembly, and the pipeline reference Embryophyta plant nuclear as the partial or complete nrDNA (18S–ITS1–5.8S–ITS2–26S) sequence assembly reference https://github.com/Kinggerm/GetOrganelle. All plastid genomes were checked manually for assembly quality, particularly at the inverted repeat boundaries. MAFFT (Kuraku *et al.*, 2013; Katoh, Rozewicki & Yamada, 2017) was used for sequence alignment, followed by a manual check using Mesquite (Maddison & Maddison, 2018) and Geneious11.1.4. Alignments in FASTA format were exported for each data set.

Plastid genomes were annotated using PGA (Qu *et al.*, 2019) and GeSeq (Tillich *et al.*, 2017). For plastome annotations, as well as *Litsea glutinosa*, two other well-annotated early-diverging species were also used as references [*Caryodaphnopsis henryi* Airy Shaw (MF939346) and a new unpublished species of *Beilschmiedia* Nees (C4011)]. To standardize annotation, we also re-annotated the previously published sequences with this workflow. After annotation, a manual check was undertaken and the reading frame was verified in Geneious11.1.4 (https://www.geneious.com) by visually inspecting the start and stop codons. The orientations of the inverted repeats (IRs) were checked by LASTZ (Harris, 2007). Transfer RNAs (tRNAs) were confirmed by their specific structure predicted by tRNAscan-SE 2.0 (Lowe & Chan, 2016; de Santana Lopes *et al.*, 2019) compared with other published annotated genomes (Shinozaki, 1986; Song *et al.*, 2017b; Qu *et al.*, 2019). Gene extraction was performed using the script 'get_annotated_regions_from_gb.py' of Jin (https://github.com/Kinggerm/PersonalUtilities) to obtain the annotated regions, then checked manually using Geneious11.1.4 with the matrix concatenated using AMAS (Borowiec, 2016). In addition, the specific and standard plastid barcode sequences from the 191 plastomes were extracted and concatenated using Geneious11.1.4.

SPECIES DISCRIMINATION AND PHYLOGENETIC ANALYSES

Our measures of species discrimination success for the genomic regions in question sampled multiple individuals per species to assess if these individuals were more closely related to each other than to other species. The species discrimination statistics only used cases where $N > 1$ individual sampled per species, but singleton species samples were included as decoys, as they occupy phylogenetic space and can 'disrupt' species-level monophyly, causing species recovery to 'fail', but they are not themselves included in the discrimination statistics. Overall, we recorded the proportion of species and genera that resolved as monophyletic following phylogenetic analysis.

The utility of different data sets for species and genera identification were investigated using the two tree-based approaches: ML (maximum likelihood) methods using IQTREE (Trifinopoulos *et al.*, 2016) and NJ (neighbour joining) methods using Geneious11.1.4. The best-fit model for each data set was determined using ModelFinder (Kalyaanamoorthy *et al.*, 2017), with the best-fit substitution model selected by the option –TEST and using a tree search with 1000 bootstrap replicates in a single run (Kalyaanamoorthy *et al.*, 2017). The number of species or genera with multiple accessions resolving as monophyletic was recorded, as was the branch support for each node with > 50% support.

To provide a best estimate of the phylogeny of Lauraceae, we also undertook phylogenetic analysis using ASTRAL-III (Zhang *et al.*, 2018), as a recent study verified that the multispecies coalescent method for determining phylogeny offered a high level of accuracy with plastid data (Gonçalves *et al.*, 2019). This approach considers variation in the phylogenetic signal across plastid genes, so we based our family-level phylogenetic tree on concatenated gene regions (excluding intergenic spacers) from the 191 plastome data set, summarized by the coalescent-based ASTRAL method. The concatenated matrix yielded 113 gene trees (absent genes were treated as missing data). Construction of the species tree was performed using the separate gene ML trees from the 191 plastomes as input for ASTRAL-III (nodes with < 10% bootstrap support were collapsed), with 100 bootstrap replicates generated to assess bootstrap support using the coalescent model. The R package phytools (Revell, 2012) was used to compare

ML trees between the 191 plastomes and concatenated genes trees.

To understand the relationships between plastomes and nrDNA better, a further analysis was conducted using only our 80 assembled whole plastid genomes and associated nrDNA (18S–ITS1–5.8S–ITS2–26S). This data set was used as the nrDNA and the plastome sequence data are derived from the same individuals. Cytonuclear discordance analysis of the 80 newly sequenced plastomes and nuclear DNA data sets was performed using ML and phytools was used for comparing the resulting ML trees.

RESULTS

Sequence characteristics

The aligned consensus length of the 191 complete plastomes was 180 129 bp, with the corresponding concatenated genes, the extracted 'Lauraceae-specific' and 'standard' barcode matrices being 79 387, 2704 and 2144 bp, respectively. For the newly sequenced samples used in the analysis of cytonuclear discordance, the nrDNA (18S–ITS1–5.8S–ITS2–26S) alignment was 6281 bp, with the corresponding plastome length 172 022 bp (Table 1). The plastid genome sizes were similar between accessions, except for the parasitic genus *Cassytha* (Supporting Information, Table S3), with *Cassytha filiformis* L. MH03, MH04 having the smallest plastid genome (114 705 bp). *Cassytha* has lost one IR region and most *ndh* genes, with remnants of some *ndh* regions as pseudogenes. The largest plastid genome is *Syndiclis* sp. ZF61 (Cryptocaryeae) with 158 639 bp, and Cryptocaryeae overall have larger plastome genomes (157 057–158 639 bp based on the unaligned sequences; Table S3), due in part to multiple large insertions in *Beilschmiedia*, *Cryptocarya* R.Br., *Endiandra* R.Br. and *Syndiclis* Hook.f. For example, in *Beilschmiedia*, there are insertions up to 723 bp relative to the related genus *Neocinnamomum* H.Liu. However, there are also large deletions, such as a 1657-bp deletion in *Beilschmiedia* compared with other early-diverging genera. For most species with more than one individual, the genome sizes were mostly the same, although some were variable, e.g. *Neocinnamomum delavayi* (Lecomte) H.Liu and *Phoebe bournei* (Hemsl.) Yen C.Yang (Table S3). These length variations mainly relate to indels located at the beginning of the large single copy (LSC) region.

Mistaken identification and species discovery

After combining DNA sequences and relating these sequences to existing morphological characters, a few putative species were divergent from most individuals sequenced for their assigned species or genus, and these were found to be nested in other taxa (labelled red in Fig. 1 and Supporting Information, Figs S1–S9). For the publicly available downloaded samples, we can only speculate about potential misidentifications, as even checking voucher specimens can leave uncertainty. For the individuals that we sampled and sequenced, the examination procedure of Liu *et al.* (2017) was followed; we rechecked our sequences together with the morphology of our vouchers, herbarium specimens from HITBC, KUN and living specimens in the XTBG and KIB botanic gardens.

Twelve individuals were found to be mislabelled or potentially so (Table 2), including ZF14, which was similar vegetatively to *Alseodaphnopsis petiolaris* (Meisn.) H.W.Li & J.Li and initially identified as such. However, the plastome sequence of ZF14 clustered with *Machilus* Nees. Recollecting the sample and repeating the experimental procedures and analyses resulted in ZF14b (Supporting Information, Tables S1–S3) still clustering with *Machilus* (Fig. 1; Figs S1–S9); however, as *Machilus* was found to be monophyletic in previous studies (Li *et al.*, 2011; Song *et al.*, 2020), we therefore assumed our initial identification of ZF14 as *Alseodaphnopsis petiolaris* was an error (Table 2).

An additional anomaly in *Cassytha* was deemed worthy of further investigation. As there is just one species of *Cassytha* (*C. filiformis*) described in China (Li *et al.*, 2008a), we treated all *Cassytha* samples as *C. filiformis*; however, our plastome data showed that MH01, 02 and SZ01 clustered with the GenBank plastome of *C. capillaris* Meisn. 1258175302 collected from Indonesia. A comparison of all plastomes of *Cassytha* spp. showed that the plastome sizes of the GenBank *C. capillaris* 1258175302 and our new sequences of MH01, 02 and SZ01 were much larger than those of the remaining samples assigned to *C. filiformis* (MH03, 04, 1258175251, 1243302039 and 1474379909), with clear insertion/deletion sites separating these sample groups. Morphological comparisons were also made and the sample fruits of SZ01 were found to be reddish and barely strigose. These features are recorded as distinguishing characteristics of *C. capillaris* (Weber, 1981, 2007). The specimens of MH01 and 02 were older, and less suited for comparisons, particularly as the colours have faded on their specimens.

In addition, Mo *et al.* (2017) described GLQ26 and GLQ33 as a new record for Yunnan Province, China, of *Alseodaphnopsis rugosa* (Merr. & Chun) H.W.Li & J.Li. However, our results placed these samples instead with the recently described *Alseodaphnopsis maguanensis* L.Li & J.Li (Li *et al.*, 2020).

Further comparisons of other samples using genetic data matched against morphological characters suggests that several of them represent potential
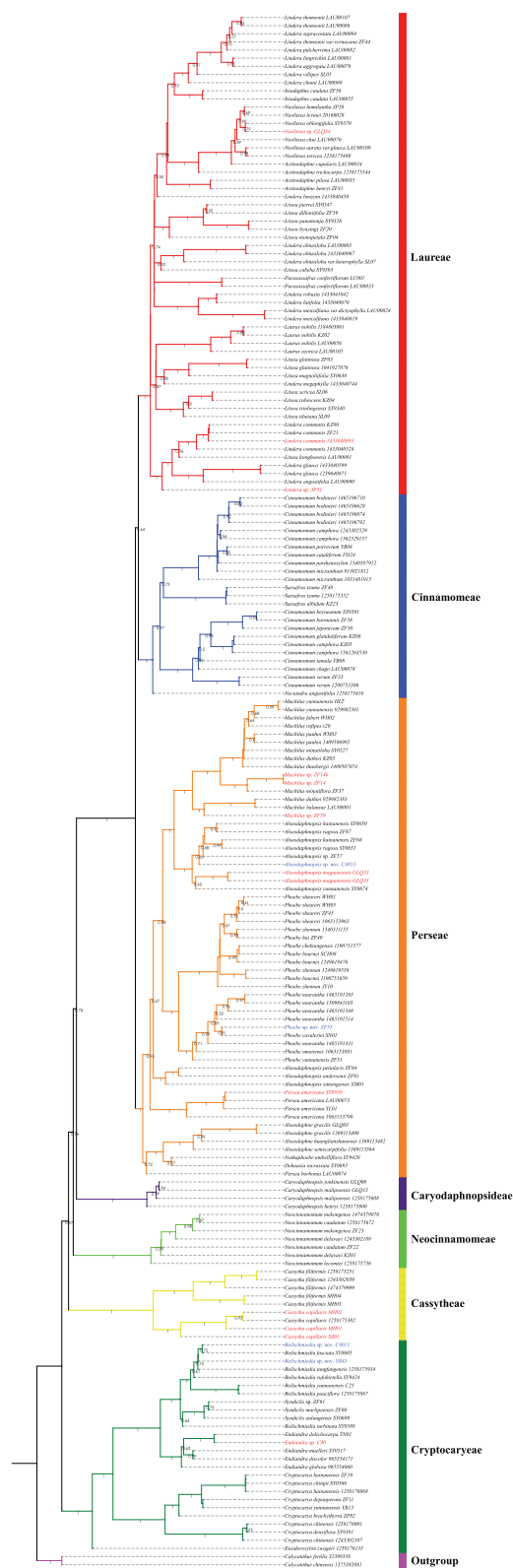
Figure 1. Phylogenetic relationships of Lauraceae species tree generated from concatenated genes of the plastid genomes bases on ASTRAL analysis.

Table 2. Original species determinations and correct species using DNA barcodes

Original species determination	Corrected species determination
Actinodaphne pilosa GLQ34	*Neolitsea* sp. *GLQ34*
Alseodaphnopsis rugosa GLQ26	*Alseodaphnopsis maguanensis GLQ26*
Alseodaphnopsis rugosa GLQ33	*Alseodaphnopsis maguanensis GLQ33*
Alseodaphnopsis petiolaris ZF14	*Machilus* sp. *ZF14*
Beilschmiedia robusta C40	*Endiandra* sp. *C40*
Cassytha filiformis MH01	*Cassytha capillaris MH01*
Cassytha filiformis MH02	*Cassytha capillaris MH02*
Cassytha filiformis SZ01	*Cassytha capillaris SZ01*
Cinnamomum caudiferum JP31	*Lindera* sp. *JP31*
Cinnamomum sp. *ZF59*	*Machilus* sp. *ZF59*
Lindera nacusua 1433040893	*Lindera communis 1433040893*
Persea americana var. *drymifolia SY9559*	*Persea americana SY9559*

new species of *Alseodaphnopsis*, *Beilschmiedia* and *Phoebe* Nees (labelled in blue in Fig. 1 and Supporting Information, Figs S1–S9). These samples warrant further investigation and are noted here as having either divergent sequences or atypical morphologies compared to their sister species in the phylogenetic tree.

COMPARISON OF DISCRIMINATION EFFICIENCY

The resolution rates of species (41.2–58.8%) and genera (52.4–66.7%) varied by data source and analytical method (Table 3; Fig. 2). The species and genera successfully distinguished are indicated with a circle and a star in Supporting Information Figures S1–S8. The two tree-based methods (ML and NJ) have the same discrimination ability at the species level, whereas ML performed better than NJ at the genus level (Fig. 2). The plastomes and concatenated genes gave the highest resolution rates (Table 3; Fig. 2; Figs S1–S4) and a comparison is shown in Figure S9, with lower resolution from the Lauraceae-specific barcodes, and the standard barcodes giving the lowest resolution of all (Table 3; Fig. 2; Figs S5–S8).
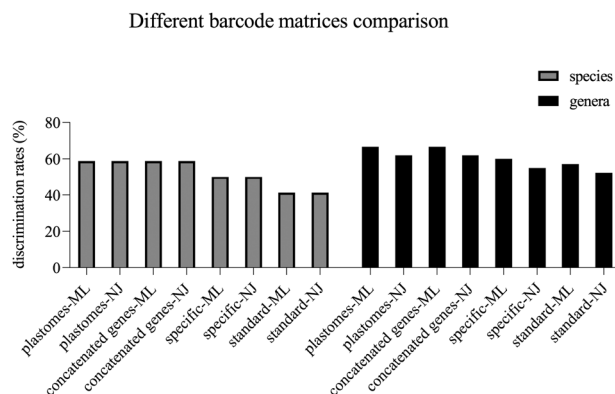
ANALYSES OF PHYLOGENY AND CYTONUCLEAR DISCORDANCE

Along with a comparison of the previous phylogenetic hypotheses and based on all previous studies, Figure 3

Table 3. Success in species discrimination based on data from the 191 plastomes data set (based on species with two or more sampled individuals per species)

Barcode regions	Plastomes-ML	Plastomes-NJ	Concatenated genes-ML	Concatenated genes-NJ	Specific-ML	Specific-NJ	Standard-ML	Standard-NJ
Successful species/ sampled species	20/34 (58.8%)	20/34 (58.8%)	20/34 (58.8%)	20/34 (58.8%)	16/32 (50%)	16/32 (50%)	14/34 (41.2%)	14/34 (41.2%)
Successful genera/ sampled genera	14/21 (66.7%)	13/21 (61.9%)	14/21 (66.7%)	13/21 (61.9%)	12/20 (60%)	11/20 (55%)	12/21 (57.1%)	11/21 (52.4%)

Figure 2. Discrimination efficiency comparison of 191 plastomes and sub-sampled regions.

shows the current most likely phylogenetic backbone of Lauraceae. The results of our study are shown in Figure 1. Our results strongly support Lauraceae as monophyletic [bootstrap support (BS) = 100%] (Fig. 1), sister to a clade containing the outgroup species [*Calycanthus chinensis* (W.C.Cheng & S.Y.Chang) P.T.Li and *Calycanthus floridus* var. *glaucus* (Willd.) Torr. & A.Gray; Calycanthaceae]. Lauraceae formed seven clades (Fig. 1), agreeing with previous phylogenetic analyses (Fig. 3). In the first-diverging clade I, Cryptocaryeae comprised five genera, with *Eusideroxylon* Teijsm. & Binn. sister to a clade with three monophyletic genera (*Cryptocarya*, *Endiandra* and *Syndiclis*) and *Beilschmiedia* paraphyletic (Fig. 1). Clade II (BS = 100%) consisted of Cassytheae with one genus, *Cassytha*, but displaying long branch lengths (Supporting Information, Fig. S9). Clade III (BS = 100%), Neocinnamomeae, included only *Neocinnamomum*. Clade IV (BS = 100%), Caryodaphnopsideae, similarly only contained *Caryodaphnopsis* Airy Shaw. Seven genera of Perseeae (*Persea* Mill., *Dehaasia* Blume, *Nothaphoebe* Blume, *Alseodaphne* Nees, *Alseodaphnopsis, Phoebe* and *Machilus*) were sister to a clade comprising taxa from Cinnamomeae and Laureae (Fig. 1). *Persea* was separated into two clades, with *P. borbonia* L. mixed with *Dehaasia*, *Nothaphoebe* and *Alseodaphne*, whereas *P. americana* Mill. was sister to the remaining Perseeae. Species of *Alseodaphnopsis* were found in two separate clades: one resolved at the base of the clade containing the *Phoebe–Machilus* lineage, and the other sister to *Machilus* (Fig. 1). The sister relationship between Cinnamomeae and Laureae was moderately supported (BS = 64%) (Fig. 1). For Cinnamomeae, *Nectandra angustifolia* Nees & Mart. ex Nees was sister to the remainder. *Sassafras* J.Presl was nested in *Cinnamomum* Schaeff., making the latter paraphyletic. In Laureae, *Iteadaphne* Blume (one species only), *Laurus* (two sampled species),
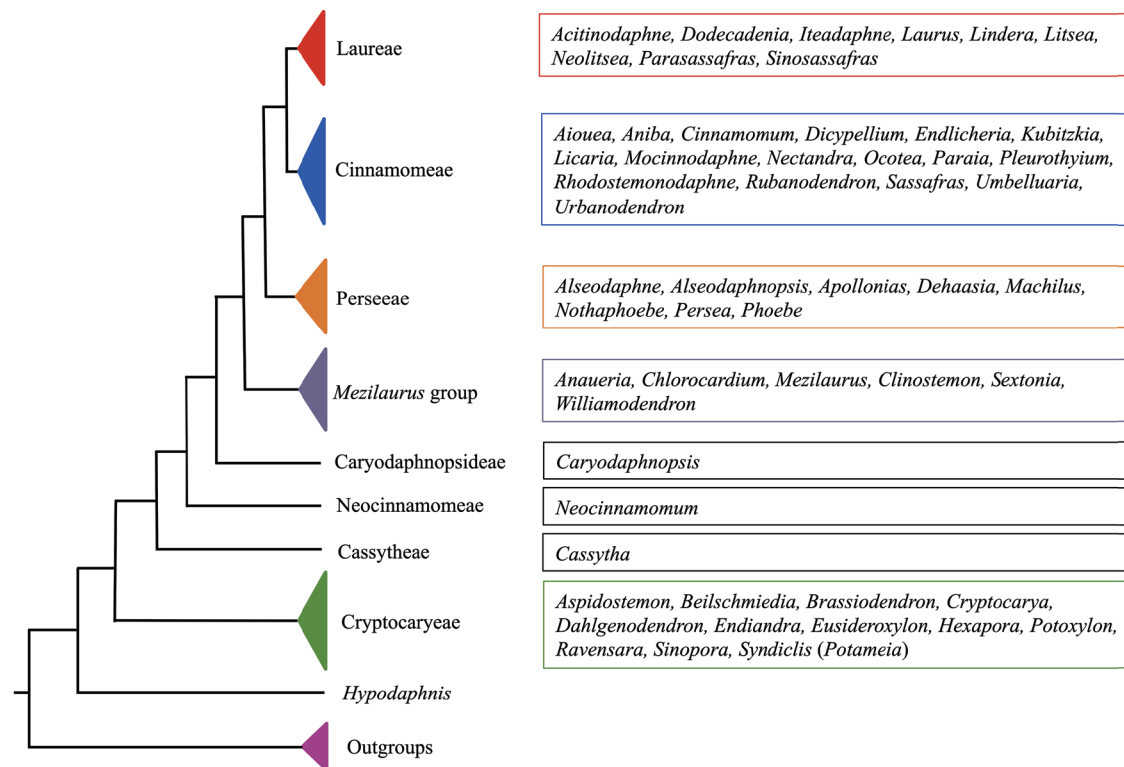
Figure 3. Review of current phylogenetic relationships of Lauraceae based on previous Sanger and plastid genome sequences.

Neolitsea (Benth.) Merr. and *Parasassafras* D.G.Long (one species only) were monophyletic ([Fig. 1](#)), whereas *Lindera* Thunb., *Litsea* Lam. and *Actinodaphne* Nees were either poly- or paraphyletic.

Alseodaphnopsis was monophyletic in a previous nrDNA study ([Mo *et al.*, 2017](#)), but our study using plastid genes did not support this. Phylogenetic incongruence of plastomes and nuclear DNA extends across the hierarchical taxonomic levels of Lauraceae, even though the information provided by nrDNA is limited ([Fig. 4](#)). First, although two early-diverging tribes, Cryptocaryeae and Cassytheae, gave consistent relationships, there were conflicting patterns on the placement of *Cryptocarya depauperata* H.W.Li and *Cryptocarya hainanensis* Merr. Second, Neocinnamomeae and Caryodaphnopsideae changed their phylogenetic positions in the nrDNA tree relative to the plastome tree, and this inconsistency was well supported (100%) in a bootstrap analysis ([Fig. 4](#)). Third, although Cinnamomeae, Laureae and Perseeae were consistent at the tribe level, cytonuclear discordance occurred within them ([Fig. 4.2](#)). In total, 19 individuals from our 80 samples were consistent across the plastome and nrDNA trees, with the remaining 61 individuals showing conflict between the nuclear and plastome phylogenomic analyses.

DISCUSSION

DNA BARCODING PERFORMANCE

There remain relatively few studies assessing the power of complete plastome sequences in plant barcoding ([Ji *et al.*, 2019](#)). The discriminatory power revealed by previous studies sampling multiple individuals from multiple congeneric species is variable. For instance, the plastome was shown to successfully distinguish some closely related species in *Quercus* L. ([Pang *et al.*, 2019](#)), *Stipa* L. ([Krawczyk *et al.*, 2018](#)) and *Taxus* L. ([Fu *et al.*, 2019](#)). However, plastomes failed to significantly improve species identification in *Panax* L. ([Ji *et al.*, 2019](#)) and New Caledonian *Araucaria* Juss. ([Ruhsam *et al.*, 2015](#)).

Our comparison in Lauraceae indicates that the discrimination rates of the plastome was higher than those of standard DNA barcodes or Lauraceae-specific barcodes. The Lauraceae-specific barcodes (i.e. loci selected as having potential for use in Lauraceae) gave a modest increase in resolution, but this was still only improved from 41% at the species level (for standard barcodes) to 50%. The highest species resolution in our study was *c*. 60% from the complete plastome sequences. Even here, however, with far from complete species-level
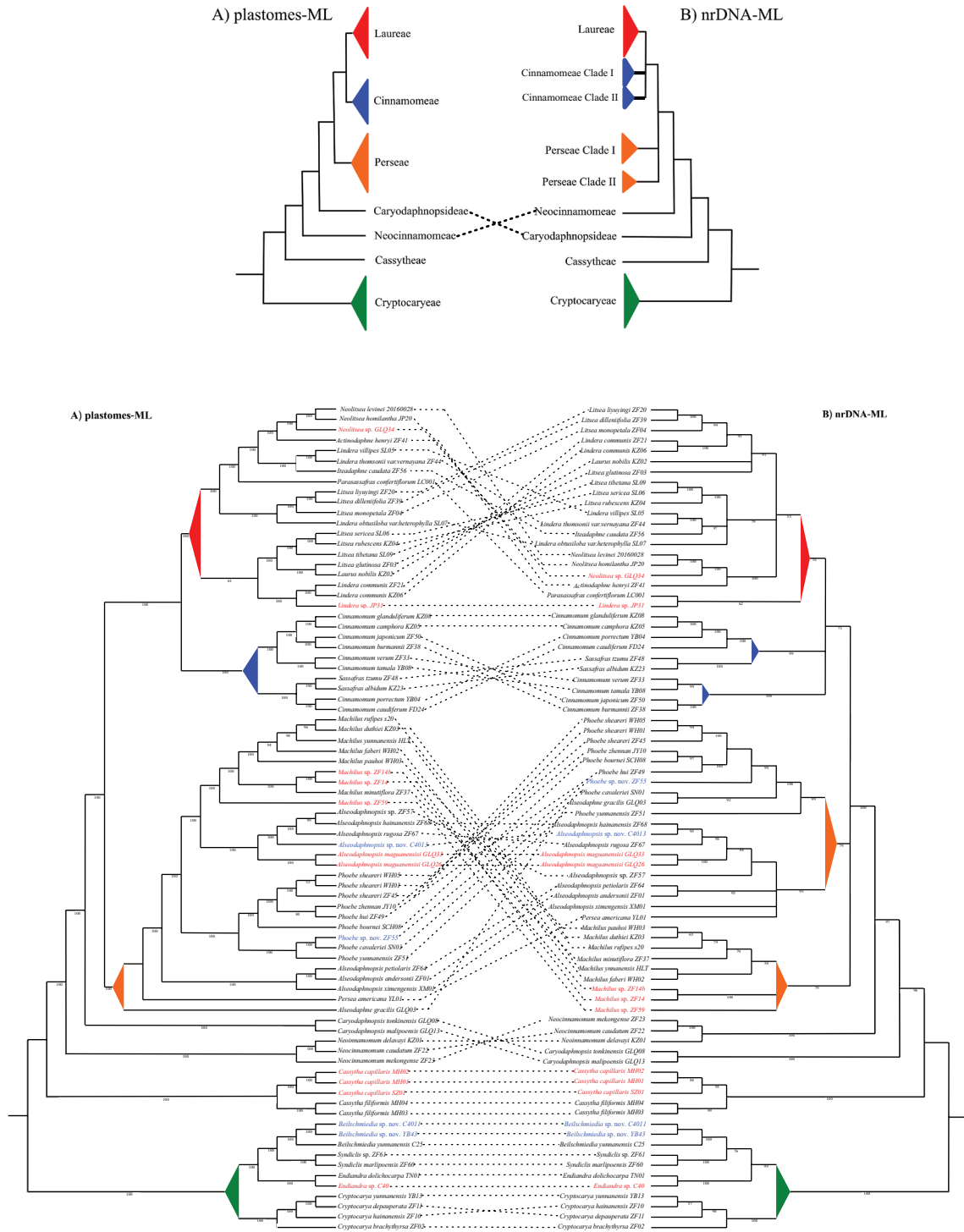
Figure 4. Discordance between 80 GetOrganelle-assembled plastomes and nrDNA sequences based on ML analyses.

sampling and 180 kb of sequence data, many species are not distinguishable with sequence data. Further sampling is likely to decrease this discrimination success, as the phylogenetic space becomes more densely occupied, and the situations where DNA barcodes did not discriminate between species are typically associated with higher sample density of co-occurring congeneric species. It is not clear what the primary driver(s) is for this discrimination failure in Lauraceae; plastid genomes being shared by hybridization, recent diversification or rapid radiations, slow sequence mutation rates and/or restricted infraspecific gene flow could be involved (Fazekas *et al.*, 2009; Hollingsworth *et al.* 2011, 2016; Ruhsam *et al.*, 2015).

Although we are focusing on the 'failure' of the DNA barcoding data to discriminate among taxa in Lauraceae, an additional factor to consider is whether the taxonomy of the family needs further revision. There are many species from different genera of Lauraceae with similar morphological characteristics. For example, some *Machilus* spp. are similar to *Phoebe*, as are some *Dehassia* to *Alseodaphne* and *Alseodaphnopsis*. Conversely, some taxa that were thought to be easy to discriminate using morphological characteristics were grouped closely on the phylogenetic tree. For example, the simple leaved taxa of *Cinnamomum* showed a close relationship with the lobed leaved taxa of *Sassafras*. Building on this example, the affinity of taxa in these genera is further revealed by reconsideration of morphological characters. Thus, simple leaves also exist in *Sassafras* which are also morphologically quite similar to *Cinnamomum* section *Camphora* Meissn. Likewise, the flowers of *Sassafras* in the Asian species show similarities to those of *Cinnamomum,* and the fruits of *Sassafras* species are similar to those of *Cinnamomum* section *Camphora*. Thus, part of the conflicting signal between the plastid data and morphology-based classifications may also be due to a complex history of the interpretation of morphological characters in the family.

Despite the data showing imperfect resolution at the species level, our results also enabled the detection of misidentifications as well as the identification of cryptic species and potential new species. Previously only *Cassytha filiformis* was known in China (Li *et al.*, 2008a). Here, through our barcode research, we can now confirm that there is another *Cassytha* sp., *Cassytha capillaris*, present in China, representing a new record for the country. Our study also highlighted other potential new taxa in *Alseodaphnopsis*, *Beilschmiedia* and *Phoebe* that warrant further investigation (Fig. 1; Supporting Information, Figs S1–S9).

PHYLOGENETIC RELATIONSHIPS AND INCONGRUENCE BETWEEN PLASTOMES AND NRDNA

The plastome sequence analysis recovered seven tribes using the multispecies coalescent method ASTRAL, with this approach giving the most detailed insight into the relationships of Lauraceae, although we lack samples from two remaining clades in Figure 3 for which previous molecular data exist: *Hypodaphnis* Stapf from Cameroon, Gabon and Nigeria (Rohwer, 1993); and *Mezilaurus* Taub. from South America (Rohwer, 1993; Chanderbali *et al.*, 2001; Rohwer & Rudolph, 2005).

Our results provide the most comprehensive plastome-based phylogenetic hypothesis for relationships in Lauraceae. However, there are still many non-monophyletic groups and taxonomic issues that need to be resolved. The paraphyletic relationships in *Actinodaphne*, *Beilschmiedia*, *Lindera*, *Litsea* and *Persea* have been documented in previous studies (Rohwer, 2000; Chanderbali *et al.*, 2001; Li, Li & Conran, 2007; Li *et al.*, 2008b, 2011; Rohwer *et al.*, 2009), whereas the relationships of *Alseodaphnopsis* seen here conflict with previous studies (Mo *et al.* 2017). Mo *et al.* (2017) published the new genus *Alseodaphnopsis* based on nuclear DNA regions, but the genus is poorly known; the limited availability of collections makes morphological diagnostic characters hard to find (van der Werff, 2019), and our study recovered the genus as polyphyletic. Similar to Rohde *et al.* (2017) and Trofimov & Rohwer (2020), we found that *Sassafras* was nested in *Cinnamomum*, with their morphological similarities discussed above.

Cytonuclear discordance has been observed in many plant groups (Rieseberg & Soltis, 1991; Soltis & Kuzoff, 1995; Folk *et al.*, 2018; Liu *et al.*, 2018, 2019; Ji *et al.*, 2019; Lee-Yaw *et al.*, 2019). The relationships in Lauraceae reconstructed using nrDNA sequences contrast with the phylogenetic tree derived from plastomes (Fig. 4). This incongruence between the maternally inherited plastid and biparentally inherited nuclear DNA has been reported in Lauraceae by Rohde *et al.* (2017) using *psbA–trnH* plus *trnG–trnS* plus ITS.

The results seen here show that is is feasible to use either plastomes or nrDNA for early-diverging groups such as Cryptocaryeae and Cassytheae, as the relationships were consistent with either data source, but the conflicts in the remaining tribes caution against the use of plastomes and nrDNA to infer relationships in isolation. Specifically, the well-supported incongruent sister relationship between Caryodaphnopsideae and Neocinnamomeae suggests that there may have been historical reticulation or other complex processes shaping the early evolutionary history of these groups. In addition, the incongruent phylogenetic relationships in Laureae, Perseeae and

Cinnamomeae emphasize the significance of hybrid origin and reticulate evolution inside these three large groups, with multiple examples of inter-specific and inter-generic incongruence in these tribes (Fig. 4.2).

A particularly difficult problem with the phylogenetics of Laureae is that relationships in the *Litsea* complex remain unresolved (Li *et al.*, 2008b). Analyses of complete plastomes and nrDNA did not resolve relationships in the *Litsea* complex, instead splitting it into several well-supported but incongruent subclades (Fig. 4.2). In contrast, the weakly supported lack of monophyly for Perseeae and Cinnamomeae in the nrDNA analyses is more likely to be due to the limited information content of our nrDNA data. Overall, to improve phylogenetic resolution the next step will be to generate data from a substantial number of nuclear markers or whole genomes for these complex groups of Lauraceae.

CONCLUSIONS

Our plastid genome data resulted in a modest increase in discriminatory power in Lauraceae compared to standard DNA barcodes or regions selected for the family. Using plastomes as genomic DNA barcodes was nevertheless useful in the correction of misidentified species, the discovery of cryptic species and in forming the foundation of the description of new species. The plastome data set also provided a useful phylogenetic framework for the family, but cytonuclear discordance suggests caution is needed in the interpretation of plastomes and/or nrDNA phylogenetic analyses of the family. This case study reiterates the value of accessing multi-locus information from the nuclear genome for species discrimination and understanding phylogenetic relationships, especially among closely related taxa.

DATA ACCESSIBILITY

GenBank numbers, plastid genomes MT621572–MT621650 and nrDNA MT628590–MT628656, MT669015–MT669026, noted in Table S2.

ACKNOWLEDGEMENTS

REFERENCES

Alsos IG, **Lavergne S**, **Merkel MKF**, **Boleda M**, **Lammers Y**, **Alberti A**, **Pouchon C**, **Denoeud F**, **Pitelkova I**, **Puşcaş M. 2020.** The treasure vault can be opened: large-scale genome skimming works well using herbarium and silica gel dried material. *Plants* **9:** 432.

Bankevich A, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin AV**, **Sirotkin AV**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology* **19:** 455–477.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4:** e1660.

Camacho C, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

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:** 12794–12797.

Chanderbali AS, **van der Werff H**, **Renner SS. 2001.** Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden* **88:** 104–134.

Chase MW, **Salamin N**, **Wilkinson M**, **Dunwell JM**, **Kesanakurthi RP**, **Haider N**, **Haidar N**, **Savolainen V.**

2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **360:** 1889–1895.

China Plant BOL Group, 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:** 19641–19646.

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. 2016. From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology* **25:** 1423–1428.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochemistry Bulletin, Botanical Society of America* **19:** 11–15.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, Newmaster SG, Hajibabaei M, Husband BC. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* **9 Suppl s1:** 130–139.

Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany* **105:** 364–375.

Fu CN, Wu CS, Ye LJ, Mo ZQ, Liu J, Chang YW, Li DZ, Chaw SM, Gao LM. 2019. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Scientific Reports* **9:** 2773.

Gentry AH. 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Annals of the Missouri Botanical Garden* **75:** 1–34.

Gonçalves DJP, Simpson BB, Ortiz EM, Shimizu GH, Jansen RK. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* **138:** 219–232.

Harris RS. 2007. *Improved pairwise alignment of genomic DNA*. Ph.D. Dissertation, Pennsylvania State University.

Hebert PD, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. B: Biological Sciences* **270:** 313–321.

Hinsinger DD, Strijk JS. 2017. Toward phylogenomics of Lauraceae: the complete chloroplast genome sequence of *Litsea glutinosa* (Lauraceae), an invasive tree species on Indian and Pacific Ocean islands. *Plant Gene* **9:** 71–79.

Hollingsworth PM. 2011. Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 19451–19452.

Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and using a plant DNA barcode. *PLoS One* **6:** e19254.

Hollingsworth PM, Li DZ, van der Bank M, Twyford AD. 2016. Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371:** 20150338.

Huang JF, Li L, van der Werff H, Li HW, Rohwer JG, Crayn DM, Meng HH, van der Merwe M, Conran JG, Li J. 2016. Origins and evolution of cinnamon and camphor: a phylogenetic and historical biogeographical analysis of the *Cinnamomum* group (Lauraceae). *Molecular Phylogenetics and Evolution* **96:** 33–44.

Ji YH, Liu CK, Yang ZY, Yang LF, He ZS, Wang HC, Yang JB, Yi TS. 2019. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Molecular Ecology Resources* **19:** 1333–1345.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2018. GetOrganelle: a simple and fast pipeline for *de novo* assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv* **2018:** 256479

de Jussieu AL. 1789. *Antonii Laurentii de Jussieu genera plantarum: secundum ordines naturales disposita, juxta methodum in horto regio parisiensi exaratam, anno M. DCC. LXXIV.* Paris: apud viduam Herissant et Theophilum Barrois.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14:** 587–589.

Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JM, Cronk Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany* **99:** 320–329.

Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20:** 1160–1166.

Kostermans AJGH. 1957. Lauraceae. *Pengumuman Balai Besar Penjelidikan Kehutanan Indonesia* **57:** 1–64.

Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J. 2018. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). *Scientific Reports* **8:** 1924.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* **2:** e508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 8369–8374.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research* **41:** W22–W28.

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 2923–2928.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9:** 357–359.

Lee-Yaw JA, **Grassa CJ**, **Joly S**, **Andrew RL**, **Rieseberg LH. 2019.** An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (Helianthus). *The New Phytologist* **221:** 515–526.

Li H-W, **Li J**, **Huang P-H**, **Wei F-N**, **Tsui H-P**, **van der Werff H. 2008a.** *Calycanthaceae–Schisandraceae, Flora of China vol. 7.* Beijing and St. Louis: Science Press and Missouri Botanical Garden Press, 102–254.

Li J, **Conran JG**, **Christophel DC**, **Li Z-M**, **Li L**, **Li H-W. 2008b.** Phylogenetic relationships of the *Litsea* complex and core Laureae (Lauraceae) using ITS and ETS sequences and morphology. *Novon* **18:** 4–8.

Li L, **Li J**, **Conran JG**, **Li X-W. 2007.** Phylogeny of *Neolitsea* (Lauraceae) inferred from Bayesian analysis of nrDNA ITS and ETS sequences. *Plant Systematics and Evolution* **269:** 203–221.

Li L, **Li J**, **Rohwer JG**, **van der Werff H**, **Wang ZH**, **Li HW. 2011.** Molecular phylogenetic analysis of the *Persea* group (Lauraceae) and its biogeographic implications on the evolution of tropical and subtropical Amphi-Pacific disjunctions. *American Journal of Botany* **98:** 1520–1536.

Li L, **Tan YH**, **Meng HH**, **Ma H**, **Li J. 2020.** Two new species of *Alseodaphnopsis* (Lauraceae) from southwestern China and northern Myanmar: evidence from morphological and molecular analyses. *PhytoKeys* **138:** 27–39.

Li X, **Yang Y**, **Henry RJ**, **Rossetto M**, **Wang Y**, **Chen S. 2015.** Plant DNA barcoding: from gene to genome. *Biological Reviews of the Cambridge Philosophical Society* **90:** 157–166.

Linnaeus CV. 1753. *Species plantarum*, Vol. **1**. Stockholm: L. Salvius.

Liu J, **Milne RI**, **Moller M**, **Zhu GF**, **Ye LJ**, **Luo YH**, **Yang JB**, **Wambulwa MC**, **Wang CN**, **Li DZ**, **Gao LM. 2018.** Integrating a comprehensive DNA barcode reference library with a global map of yews (*Taxus* L.) for forensic identification. *Molecular Ecology Resources* **18:** 1115–1131.

Liu Y, **Johnson MG**, **Cox CJ**, **Medina R**, **Devos N**, **Vanderpoorten A**, **Hedenäs L**, **Bell NE**, **Shevock JR**, **Aguero B**, **Quandt D**, **Wickett NJ**, **Shaw AJ**, **Goffinet B. 2019.** Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* **10:** 1485.

Liu ZF, **Ci XQ**, **Li L**, **Li HW**, **Conran JG**, **Li J. 2017.** DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. *PLoS One* **12:** e0175788.

Lowe TM, **Chan PP. 2016.** tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* **44:** W54–W57.

Maddison WP, **Maddison DR. 2018.** *Mesquite: a modular system for evolutionary analysis. Version 3.51.* Available at: http://www.mesquiteproject.org.

Mo YQ, **Li L**, **Li JW**, **Rohwer JG**, **Li HW**, **Li J. 2017.** Alseodaphnopsis: a new genus of Lauraceae based on molecular and morphological evidence. *PLoS One* **12:** e0186545.

Nees von Esenbeck CGD. 1836. *Systema laurinarum.* Berlin: Sumptibus Veitii et Sociorum.

Pang XB, **Liu HS**, **Wu SR**, **Yuan YC**, **Li HJ**, **Dong JS**, **Liu ZH**, **An CZ**, **Su ZH**, **Li B. 2019.** Species identification of oaks (*Quercus* L., Fagaceae) from gene to genome. *International Journal of Molecular Sciences* **20:** 5940.

Qu XJ, **Moore MJ**, **Li DZ**, **Yi TS. 2019.** PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15:** 50.

Revell LJ. 2012. *phytools*: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3:** 217–223.

Rieseberg LH, **Soltis DE. 1991.** Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5:** 65–84.

Rohde R, **Rudolph B**, **Ruthe K**, **Lorea-Hernández FG**, **de Moraes PLR**, **Li J**, **Rohwer JG. 2017.** Neither *Phoebe* nor *Cinnamomum*—the tetrasporangiate species of *Aiouea* (Lauraceae). *Taxon* **66:** 1085–1111.

Rohwer JG. 1993. Lauraceae. In: Kubitzki K, Rohwer JG, Bittrich V. eds. *The families and genera of vascular plants, Vol. 2. Flowering plants. Dicotyledons: magnoliid, hamamelid and caryophyllid families.* Berlin: Springer-Verlag, 366–391.

Rohwer JG. 2000. Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Systematic Botany* **25:** 60–71.

Rohwer JG, **Li J**, **Rudolph B**, **Schmidt SA**, **van der Werff H**, **Li H-W. 2009.** Is *Persea* (Lauraceae) monophyletic? Evidence from nuclear ribosomal ITS sequences. *Taxon* **58:** 1153–1167.

Rohwer JG, **Rudolph B. 2005.** Jumping genera: the phylogenetic positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) based on different analyses of *trnK* intron sequences. *Annals of the Missouri Botanical Garden* **92:** 153–178.

Ruhsam M, **Rai HS**, **Mathews S**, **Ross TG**, **Graham SW**, **Raubeson LA**, **Mei W**, **Thomas PI**, **Gardner MF**, **Ennos RA**, **Hollingsworth PM. 2015.** Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Molecular Ecology Resources* **15:** 1067–1078.

de Santana Lopes A, **Gomes Pacheco T**, **Nascimento da Silva O**, **Magalhães Cruz L**, **Balsanelli E**, **Maltempi de Souza E**, **de Oliveira Pedrosa F**, **Rogalski M. 2019.** The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* **250:** 1229–1246.

Shinozaki K, **Ohme M**, **Tanaka M**, **Wakasugi T**, **Hayashida N**, **Matsubayashi T**, **Zaita N**, **Chunwongse J**, **Obokata J**, **Yamaguchi-Shinozaki K**, **Ohto C**, **Torazawa K**, **Meng BY**, **Sugita M**, **Deno H**, **Kamogashira T**, **Yamada K**, **Kusuda J**, **Takaiwa F**, **Kato A**, **Tohdoh N**, **Shimada H**, **Sugiura M. 1986.** The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal* **5:** 2043–2049.

Soltis DE, **Kuzoff RK. 1995.** Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* **49:** 727–742.

Song Y, **Dong W**, **Liu B**, **Xu C**, **Yao X**, **Gao J**, **Corlett RT. 2015.** Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Frontiers in Plant Science* **6:** 662.

Song Y, **Yao X**, **Liu B**, **Tan Y**, **Corlett RT. 2018.** Complete plastid genome sequences of three tropical *Alseodaphne* trees in the family Lauraceae. *Holzforschung* **72:** 337–345.

Song Y, **Yao X**, **Liu B**, **Tan Y**, **Gan Y**, **Yang J**, **Corlett RT. 2017a.** Comparative analysis of complete chloroplast genome sequences of two subtropical trees, *Phoebe sheareri* and *Phoebe omeiensis* (Lauraceae). *Tree Genetics & Genomes* **13:** 120.

Song Y, **Yao X**, **Tan Y**, **Gan Y**, **Corlett RT. 2016.** Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Canadian Journal of Forest Research* **46:** 1293–1301.

Song Y, **Yu WB**, **Tan Y**, **Liu B**, **Yao X**, **Jin J**, **Padmanaba M**, **Yang JB**, **Corlett RT. 2017b.** Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the magnoliids. *Genome Biology and Evolution* **9:** 2354–2364.

Song Y, **Yu WB**, **Tan YH**, **Jin JJ**, **Wang B**, **Yang JB**, **Liu B**, **Corlett RT. 2020.** Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *Journal of Systematics and Evolution* **58:** 423–439.

Tillich M, **Lehwark P**, **Pellizzer T**, **Ulbricht-Jones ES**, **Fischer A**, **Bock R**, **Greiner S. 2017.** GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45:** W6–W11.

Trifinopoulos J, **Nguyen LT**, **von Haeseler A**, **Minh BQ. 2016.** W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44:** W232–W235.

Trofimov D, **Rohwer JG. 2020.** Towards a phylogenetic classification of the *Ocotea* complex (Lauraceae): an analysis with emphasis on the Old World taxa and description of the new genus *Kuloa*. *Botanical Journal of the Linnean Society* **192:** 510–535.

Twyford AD, **Ness RW. 2017.** Strategies for complete plastid genome sequencing. *Molecular Ecology Resources* **17:** 858–868.

van der Werff H. 2019. *Alseodaphnopsis* (Lauraceae) revisited. *Blumea* **64:** 186–189.

van der Werff H, **Richter HG. 1996.** Toward an improved classification of Lauraceae. *Annals of the Missouri Botanical Garden* **83:** 409–418.

Weber JZ. 1981. A taxonomic revision of *Cassytha* (Lauraceae) in Australia. *Journal of the Adelaide Botanic Gardens* **3:** 187–262.

Weber JZ. 2007. *Cassytha*. In: Wilson, AJG ed. *Flora of Australia, vol. 2: Winteraceae to Platanaceae.* Collingwood: ABRS/CSIRO, 117–136.

Wick RR, **Schultz MB**, **Zobel J**, **Holt KE. 2015.** Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics (Oxford, England)* **31:** 3350–3352.

Wu CS, **Wang TJ**, **Wu CW**, **Wang YN**, **Chaw SM. 2017.** Plastome evolution in the sole hemiparasitic genus laurel dodder (*Cassytha*) and insights into the plastid phylogenomics of lauraceae. *Genome Biology and Evolution* **9:** 2604–2614.

Zhang C, **Rabiee M**, **Sayyari E**, **Mirarab S. 2018.** ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19:** 153.

Zhao ML, **Song Y**, **Ni J**, **Yao X**, **Tan YH**, **Xu ZF. 2018.** Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Scientific Reports* **8:** 8844.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Note S1. Background and notes on Lauraceae-specific barcodes.
Table S1. A summary of all taxa compared for plastid sequence variation including *de novo* sequences generated for this study, and sequences retrieved from public sequence databases.
Table S2. Taxon sampling, material sources and GenBank accession numbers for *de novo* sequenced individuals.
Table S3. Plastid genome sizes of the 191 individuals.
Figure S1. ML tree of 191 whole plastid genomes.
Figure S2. NJ tree of 191 whole plastid genomes.
Figure S3. ML tree of concatenated genes extracted from 191 whole plastid genomes.
Figure S4. NJ tree of concatenated genes extracted from 191 whole plastid genomes.
Figure S5. ML tree of the Lauraceae-specific barcode matrix (*ycf1*+*ndhH*–*rps15*+*trnL*–*ycf2*) extracted from 191 whole plastid genomes.
Figure S6. NJ tree of the Lauraceae-specific barcode matrix (*ycf1*+*ndhH*–*rps15*+*trnL*–*ycf2*) extracted from 191 whole plastid genomes.
Figure S7. ML tree of the standard barcode matrix (*rbcL*+*matK*+*trnH*–*psbA*) extracted from 191 whole plastid genomes.
Figure S8. NJ tree of the standard barcode matrix (*rbcL*+*matK*+*trnH*–*psbA*) extracted from 191 whole plastid genomes.
Figure S9. ML tree comparisons of 191 whole plastid genomes and corresponding concatenated genes.