JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12556

Research Article

Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences

Chao-Nan Cai[1,2], Hui Ma[1], Xiu-Qin Ci[1,3], John G. Conran[4], and Jie Li[1,3]*

[1]Plant Phylogenetics and Conservation Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla 666303, China

[4]Australian Centre for Evolutionary Biology and Biodiversity (ACEBB), and Sprigg Geobiology Centre (SGC), School of Biological Sciences, The University of Adelaide, Adelaide 5005, South Australia, Australia

*Author for correspondence. E-mail: jieli@xtbg.ac.cn

Received 28 June 2019; Accepted 1 December 2019; Article first published online 12 December 2019

Abstract The biologist's ruler for biodiversity is the species; accurate species identification is fundamental to the conservation of endangered species and in-depth biological scientific exploration. However, the delimitation and affinities of *Horsfieldia* in China has been controversial, owing in part to very low levels of molecular divergence within the family Myristicaceae. Because species boundaries and phylogenetic relationships within *Horsfieldia* are also unclear, 13 samples were collected across its distribution in China and their genomes were subjected to shotgun sequencing using Illumina platforms. A total of 40 487 994–84 801 416 pair-end clean reads were obtained and, after assembly, the complete chloroplast genome was recovered for all samples. Annotation analysis revealed a total of 112 genes, including 78 protein-coding genes, 30 transfer RNA, and 4 ribosomal RNA genes. Six variable loci (*petN-psbM*, *trnH-psbA*, *ndhC-trnV*, *psbJ-psbL*, *ndhF*, and *rrn5-rrn23*) were identified. Phylogenetic analyses strongly support the presence of four distinct species of *Horsfieldia* in China. In addition, samples that had been identified previously as *Horsfieldia kingii* (Hook. f.) Warb. were indistinguishable from those of *H. prainii* (King) Warb., suggesting that if *H. kingii* does occur in China, it was not collected in this study. Similarly, the chloroplast genome of one *H. hainanensis* Merr. sample from Guangxi province was identical to *H. tetratepala* C. Y. Wu, suggesting that the distribution range of *H. hainanensis* might be narrower than assumed previously. The phylogenetic relationships between the Chinese *Horsfieldia* species based on the whole chloroplast genomes was supported strongly, indicating the potential for using entire chloroplast genomes as super-barcodes for further resolution of the phylogeny of the genus *Horsfieldia*.

Key words: chloroplast genome, *Horsfieldia*, Myristicaceae, phylogenetics, species delimitation, super-barcodes.

1 Introduction

The core of taxonomy is to classify, identify, and describe organisms (Simpson, 2010; Steeves, 2011); as such, species delimitation is a crucial issue for biodiversity conservation (Niu et al., 2018). Traditional identification mainly relies on morphology, but this is often complicated by phenotypic plasticity, or the presence of morphologically cryptic taxa leading to identification errors, especially in the field (Hebert et al., 2003). Due to these limitations and the increasing scarcity of trained taxonomists, a new approach for species identification is therefore desirable. The implementation of DNA-based identification systems (such as DNA barcoding) has alleviated some of those problems to an extent (Brown et al., 1999; Bucklin et al., 1999), as DNA barcodes from unknown samples can be compared against authenticated reference sequences (Wilkinson et al., 2017). For example, the core barcode loci *rbcL* and *matK* are applied widely for plant species delimitation and identification of unknown samples (CBOL Plant Working Group, 2009). However, although DNA barcoding could potentially revolutionize species identification, poor sequence quality, hybridization, and incomplete lineage sorting resulting from relatively recent taxon origin could make targeted gene barcoding insufficient for delimitating some species (Steeves, 2011). Fortunately, recent studies have revealed that whole chloroplast genome sequences can serve as super-barcodes for species identification, providing useful information due to their relatively small size, non-recombinant, highly conserved nature, and maternal inheritance (Sato et al., 1999; Nock et al., 2011; Song et al., 2017; Chen et al., 2018; Wang et al.,

2018). In angiosperms, most chloroplast genomes have highly conserved quadripartite structures and organization of content consisting of a large (LSC) and small (SSC) single-copy region, connected end-to-end by two copies of an inverted repeat (IR) (Palmer, 1985; Daniell et al., 2016). The advent of next-generation sequencing techniques has also made it possible for rapid, efficient, and affordable high-quality sequencing of entire chloroplast genomes.

The genus *Horsfieldia* Willd. (Myristicaceae) has approximately 100 species distributed from Sri Lanka and southern China east to the Solomon Islands and northern Australia (Sinclair, 1975; De Wilde, 1984, 2000; Jessup, 2007). *Horsfieldia* species can have high economic value, as their trunks are used for redwood furniture and high-end decorative woodwork (Jiang et al., 2016) and the oily seeds are an ideal source for biodiesel (Xu et al., 2012); however, this economic importance, combined with habitat destruction and fragmentation, has seen a steady decline of the genus in China. At present, three Chinese species of *Horsfieldia* are listed as endangered at the national level, primarily as a result of deforestation and habitat destruction (Wang & Xie, 2004) and urgent measures are needed to conserve them. However, to achieve this, accurate identification is required at the species level. Accordingly, a thorough reinvestigation of species delimitation in Chinese *Horsfieldia* was undertaken.

The *Flora of China* (FOC) account of Li & Wilson (2008) recognized three *Horsfieldia* species in China: *H. kingii* (Hook. f.) Warb., *H. amygdalina* (Wall.) Warb., and *H. prainii* (King) Warb., but the species-level taxonomy of the genus in China is controversial, especially for *H. kingii*. In the FOC, *H. tetratepala* C. Y. Wu, *H. hainanensis* Merr. are considered synonyms of *H. kingii*, based on the shared possession of smooth twigs with small lenticels, papery glabrous leaf blades with an occasionally pubescent midrib, and ovoid seeds with a slightly pointed apex. In contrast, Li (1979) in the *Flora Reipublicae Popularis Sinicae* (FRPS) separated *H. kingii*, *H. hainanensis*, and *H. tetratepala* from each other on characters of the twigs, lenticels, leaf blade morphology, and the length of the infructescence, although these taxa can be discriminated further by the length of the male inflorescences and features of the aril (Li, 1979; Li & Wilson, 2008), in addition to geographic and ecological characteristics.

For example, *H. kingii* and *H. tetratepala* have longer male inflorescences than *H. hainanensis* and *H. tetratepala* also has longer fruit infructescences than *H. hainanensis*. Similarly, *H. kingii* has bright yellow arils (Fig. S1A), whereas those of *H. hainanensis* and *H. tetrapetala* (Fig. S1B) are red (Li, 1979; Datta & Rane, 2013). In addition, Chinese *Horsfieldia* occur mainly in Guangxi, Yunnan, and Hainan provinces near the borders of Myanmar and Vietnam, where they grow in ravines with dense primary rainforest at altitudes of 100–1200 m, with *H. hainanensis* distributed mainly in southern Guangxi and Hainan provinces, growing in the shady, wet forests of hills and valleys at altitudes of 400–450 m. *Horsfieldia kingii* is mainly recorded from south to southwest Yunnan from dense forests in ravines, whereas *H. tetratepala* is mainly distributed in southeast Yunnan, growing in dense forests in hills and valleys.

De Wilde (1984), Ye (2004), and Wu et al. (2015) studied the taxonomy of *H. tetratepala*, *H. hainanensis*, and *H. kingii* and supported Li & Wilson's (2008) FOC concept of a single species. However, as *Horsfieldia* species were identified traditionally in part on their geographical origin and only a small number of morphological characters (Merrill, 1932; Wu & Wang, 1957), sample identification is problematic, especially where species ranges overlap and/or critical characters for differentiation are missing. Because of this taxonomic confusion, as well as issues with intraspecies variability, molecular studies are needed to address *Horsfieldia* species boundaries and taxon identities in order to resolve species limits for the Chinese members of the genus and thus improve their conservation.

Previous studies on *Horsfieldia* have focused mainly on resolving issues of breeding technology (He et al., 2013), volatile oil chemistry (Dang et al., 2009), genetic diversity and population structure (Jiang et al., 2016, 2018), taxonomic position of the genus (Wu et al., 2015), phytochemistry and biological activity (Ma et al., 2014), and geographical distribution and habitat requirements (Zhong et al., 2018). Nevertheless, sequence data for Myristicaceae are limited, so the current study samples complete chloroplast genome data for Chinese *Horsfieldia* species across their geographic range to help resolve species limits. Specifically, this study aims to: (i) develop complete chloroplast genome sequences for seven Myristicaceae species using next-generation sequencing; (ii) identify any hypervariable regions in these chloroplast genomes; and (iii) use these data to define and differentiate Chinese *Horsfieldia* species and identify hypervariable genome regions suitable for future phylogenetic and biogeographic studies.

2 Material and Methods

2.1 Sampling and DNA extraction

Fresh samples were collected from Hainan, Guangxi, and Yunnan provinces during the 2017 and 2018, and dried immediately in the field using silica gel. Voucher specimens were deposited in the Herbarium of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (HITBC). Thirteen individuals of Chinese *Horsfieldia* were studied, with two *Myristica* species (*Myristica yunnanensis* Y. H. Li and *M. fragrans* Houtt.) and one *Knema* species (*Knema tenuinervia* W. J. de Wilde) used as outgroups (Table S1) (Sauquet et al., 2003).

The total genomic DNA was extracted from young leaves using a modified CTAB protocol adapted from Doyle & Doyle (1987) but using 4% CTAB. The DNA quality was assessed by 1.0% agarose gel electrophoresis and the solution was stored at −80 °C. A previously reported chloroplast genome for *H. prainii* (Mao et al., 2019) was also downloaded from GenBank (MH445411.1; https://www.ncbi.nlm.nih.gov/nuccore/1615640357).

2.2 Analyses of chloroplast genome sequence data

2.2.1 Illumina sequencing and genome assembly

Approximately 1.5 ng high-quality DNA from each of the 16 samples was used to build shotgun libraries with insert sizes of 300 bp. Sequencing was carried out with an Illumina HiSeq 2500 System (Illumina, San Diego, CA, USA) in the paired-end sequencing mode (150 bases each) at the Beijing Genomics Institute (Shenzhen, China). Raw reads were filtered first to

obtain the high-quality clean data by removing adaptor sequences and low-quality reads with Q-value ≤20. The quality-filtered reads were then subjected to de novo assembling with GetOrganelle (Jin et al., 2018).

2.2.2 Genome annotation, alignment, and visualization

The chloroplast genes were annotated with default parameters using an online Dual Organellar Genome Annotator tool (Wyman et al., 2004), with default parameters to predict protein-coding genes, transfer RNA (tRNA) genes, and ribosome RNA (rRNA) genes. A circular diagram for the chloroplast genome was generated using the OrganellarGenome DRAW tool (Lohse et al., 2013).

2.2.3 Microsatellite detection analysis simple sequence repeats

Simple sequence repeats (SSRs) were predicted by the microsatellites identification tool MISA (http://pgrc.ipk-gatersleben.de/misa/) with the parameters set to 10 repeat units ≥10 for mononucleotide SSRs, six repeat units ≥6 for dinucleotide, five repeat units ≥5 for trinucleotide, four repeat units ≥4 for tetranucleotide, and three repeat units ≥3 for pentanucleotide and hexanucleotide.

2.2.4 Identification of the hypervariable regions

All the chloroplast genome sequences were aligned using MAFFT version 7 (Katoh & Standley, 2013) and adjusted manually when necessary. Ambiguously aligned loci (e.g., N) were excluded from the analysis and a sliding window analysis was carried out using VariScan 2.0.3 software (Hutter et al., 2006) to analyze nucleotide diversity (π) across the chloroplast genome, window length set to 600 bp, and the step size set as 200 bp. In order to compare and analyze the chloroplast genomes of the *Horsfieldia* species, full alignments with annotations were visualized with Shuffle-LAGAN mode in mVISTA software (Mayor et al., 2000; Brudno et al., 2003) with the default settings and *H. hainanensis* (sample collected from Hainan province) used as a reference sequence.

2.2.5 Phylogenetic analysis

Phylogenetic analyses were undertaken on the 16 complete chloroplast genomes sequences using maximum likelihood (ML) and Bayesian inference analysis, with *M. yunnanensis*, *M. fragrans,* and *K. tenuinervia* as outgroups. The ML analysis was carried out using RAxML 8.2.10 following Stamatakis (2014), with 1000 bootstrap replicates carried out to obtain support values. jModelTest 2.1.10 (Darriba et al., 2012) was used to select the best-fit evolutionary model and gamma rate heterogeneity using the Akaike information criterion for each partition. Bayesian inference analysis was also undertaken based on the best-fit model using MrBayes 3.2.6 (Ronquist & Huelsenbeck, 2003). To allow adequate time for convergence, four Markov chains were run for 2 million generations, with sampling every 1000 generations. The first 25% trees were discarded as burn-in and the remaining trees were then used to estimate the 50% majority-rule consensus tree and Bayesian posterior probabilities.

Table 1 Summary of chloroplast genome characteristics of *Horsfieldia* species and outgroups *Myristica yunnanensis*, *M. fragrans*, and *Knema tenuinervia*

Attribute	*H. hainanensis_HN*	*H. tetratepala*	*H. prainii*	*H. amygdalina*	*M. yunnanensis*	*M. fragrans*	*K. tenuinervia*
Genome size (bp)	155 774	155 770	155 696	155 682	155 919	155 868	155 775
GC content (%)	39.2	39.2	39.2	39.2	39.2	39.2	39.2
LSC length (bp)	87 032	87 026	86 868	86 929	87 084	87 062	86 909
SSC length (bp)	20 610	20 612	20 666	20 587	20 693	20 664	20 761
IR length (bp)	24 066	24 066	24 081	24 083	24 071	24 071	24 075
Total number of genes	112	112	112	112	112	112	112
No. protein encoding	78	78	78	78	78	78	78
No. tRNA	30	30	30	30	30	30	30
No. rRNA	4	4	4	4	4	4	4

IR, inverted repeat; LSC, large single-copy; rRNA, ribosome RNA; SSC, small single-copy; tRNA, transfer RNA.

3 Results

3.1 Analyses of chloroplast genome sequence data

3.1.1 Genome features

All complete *Horsfieldia* chloroplast genomes were composed of single circular double-stranded DNA molecules and displayed the typical quadripartite structure of most angiosperms, including the SSC region and LSC region, separated by two IR regions (Fig. S1). The GC content of all the chloroplast genomes was relatively conserved at around 39.2% (Table 1) and genome size ranged from 155 682 to 155 774 bp. The length varied from 86 868 to 87 032 bp in the LSC region, 20 587–20 666 bp in the SSC region, and 24 066–24 083 bp in the IR region. These genome sequences were deposited into the GenBank under the accession numbers MN486685–MN486686 and MN495958–MN495971 (Table S1).

Each chloroplast genome contains 112 genes, including 78 protein-coding genes, 30 tRNA genes, and 4 rRNA genes (Table 1). Seven protein-coding genes (*rps19*, *rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, and *rps12*), five tRNAs (*trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, and *trnA-UGC*), and all of the rRNAs (*rrn16*, *rrn5*, *rrn4.5*, and *rrn23*) are duplicated in the IR region. The LSC region contained 61 protein-coding genes and 22 tRNA genes. The SSC region contained 11 protein-coding genes and three tRNA genes. There are 17 intron-containing genes, including five tRNA genes and 12 protein-coding genes, almost all of which are single-intron genes except for *ycf3* and *clpP*, each having two introns (Table S2). All *Horsfieldia* chloroplast genomes sizes and overall genomic structures are well conserved (Fig. S1).

3.1.2 Microsatellite detection analysis SSRs

Using the microsatellites identification tool MISA, 58 SSRs detected in the *H. hainanensis_HN* (from Hainan province) chloroplast genome (Table 2), including 50 homopolymers and eight tetrapolymers. Forty-eight of the 58 SSRs were composed solely of A or T bases and this AT-rich SSR phenomenon was also identified in the chloroplast genomes of other *Horsfieldia* species. In all chloroplast genomes, mononucleotide repeats were the most abundant repeat type followed by the tetranucleotide (Fig. 1). Dinucleotide and trinucleotide repeats were not detected, but pentanucleotide and hexanucleotide repeats were present in *H. prainii.* These SSRs therefore have potential to be used as molecular markers for population and phylogenetic studies in Myristicaceae.

3.1.3 Identification of hypervariable regions

In order to find sequence divergence among Myristicaceae chloroplast genomes, we undertook a sliding window analysis to calculate the nucleotide diversity (π) using VariScan 2.0.3 software (Hutter et al., 2006). The values ranged from 0 to 0.03880 (mean, 0.004837), revealing the slight differences among the chloroplast genomes. However, six highly variable loci with higher π values ($\pi > 0.02$) were able to be located precisely (Fig. 2). These regions included *petN-psbM*, *trnH-psbA*, *ndhC-trnV*, *psbJ-psbL*, *ndhF*, and *rrn5-rrn23*, four of which lie in the LSC region (Fig. 2). As expected, the LSC and SSC regions showed higher sequence divergence than IR regions, and coding regions are more conserved than non-coding regions, with the most divergent non-coding regions located in the intergenic spacers. The divergent Myristicaceae chloroplast genome regions were represented by five intergenic spacers (*petN-psbM*, *trnH-psbA*, *ndhC-trnV*, *psbJ-psbL*, and *rrn5-rrn23*) and one coding region (*ndhF*) (Fig. 2).

The sequence identities of *Horsfieldia* chloroplast genomes were plotted using mVISTA with *H. hainanensis_HN* used as a reference (Fig. 3). The results revealed high sequence similarity among all sequences, suggesting that *Horsfieldia* chloroplast genomes are rather conserved, with the *H. hainanensis_GX* chloroplast genome similar to that of *H. tetratepala*.

3.1.4 Comparison of IR/single copy boundaries

Inverted repeats of vascular plants often extend into neighboring single copy (SC) regions, causing differences at or near the border regions of the IR/SC (Ravi et al., 2008) and size variation of the chloroplast genome can be due partly to the expansion and contraction of the IR into or out of adjacent SC regions (Plunkett & Downie, 2000; Yang et al., 2016; Zhang et al., 2016; Xu et al., 2017). We therefore analyzed the gene contents at the border areas in detail in order to find potential evolutionary events (Fig. 4). Common genes adjacent to the border regions in eudicots are *trnH*, *rps19*, *ycf1*, and *ndhF*, but our results also found that *rps19* crossed the LSC/IRA region and that

Table 2 Simple sequence repeats in the *Horsfieldia hainanensis_HN* chloroplast genome

	Number of repeats														
Sequences	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
A/T	–	–	–	–	–	–	–	24	14	6	2	1	–	1	48
C/G	–	–	–	–	–	–	–	1	–	1	–	–	–	–	2
AAAC/GTTT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AAAG/CTTT	2	–	–	–	–	–	–	–	–	–	–	–	–	–	2
AACT/AGTT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AATG/ATTC	2	–	–	–	–	–	–	–	–	–	–	–	–	–	2
ACAT/ATGT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AGAT/ATCT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	1

–, none.

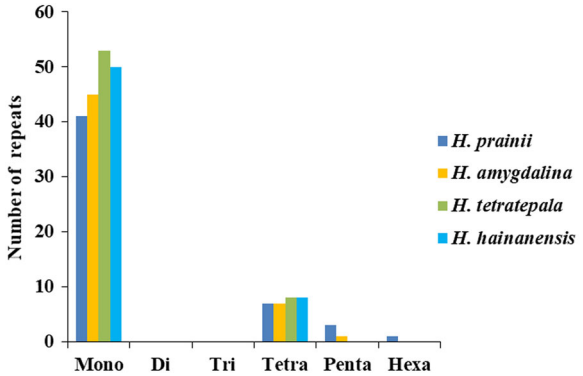

Fig. 1. Simple sequence repeats in four *Horsfieldia* chloroplast genomes. *Horsfieldia hainanensis* was collected from Hainan province, China.

81 bp of ψ*rps19* was located in the IRB region. In addition, *ycf1* was located at the SSC region, the pseudogene ψ*ycf1* fragment was lost, and because of the IR contraction, the position of the genes located at the borders have changed. For example, the *rrn5* and *trnR* genes at the IRB/SSC border were separated by 242–260 bp, whereas the gene *rrn5*/*ndhF* separation at the IRA/SSC border was 89–131 bp.

3.1.5 Phylogenetic analysis

Phylogenetic analysis of the entire chloroplast genome sequences using ML and Bayesian inference was undertaken for individuals from the *Horsfieldia* species and three outgroup species sampled here (Figs. 5, S2). Both analysis methods showed that *H. tetratepala* and *H. hainanensis_GX* (from Guangxi province) formed a clade with bootstrap values of 100%/1.0 and very low plastid genome sequence divergence, with the *H. tetratepala* genome only 4 bp shorter than *H. hainanensis_GX*. In addition, only one single nucleotide polymorphism and two indels were detected in the chloroplast genomes of *H. tetratepala* and *H. hainanensis_GX* (Table S3), strongly supporting their inclusion into a single species. In contrast, these samples were separated from *H. hainanensis_HN*, indicating some of the *H. hainanensis* samples used here were misidentified. Furthermore, the two samples designated for this study as *H. kingii* formed a clade with *H. prainii* (BS = 100%/1.0) and their chloroplast genomes were virtually identical. This implies that true samples of Chinese *H. kingii* are either too rare to be collected easily, are locally extinct, or previous records represent misapplications of the name to other taxa.

4 Discussion

Analysis of the *Horsfieldia* chloroplast genomes supports the recognition of at least four species present in China: *H. amygdalina*, *H. hainanensis*, *H. prainii*, and *H. tetratepala*. *Horsfieldia amygdalina* can be distinguished easily by its tiny leaves and glabrescent male perianth, whereas *H. prainii* is recognizable by its monoecious habit and obovate-oblong to panduriform leaf blades. Nevertheless, the results of the present study reject both the species concepts of Li's (1979) FRPS treatment and those of Li & Wilson (2008) in the FOC.

Despite being sterile, our *H. hainanensis* and *H. tetratepala* samples were largely separated in the phylogenies as two clades (both with 100%/1.0 support), with only one sample (*H. hainanensis_GX*) apparently misidentified. This separation was also supported by ecological differences, as *H. tetratepala* mainly grows in dense limestone karst forests of Yunnan and Guangxi provinces, whereas *H. hainanensis* occurs in the shady, wet forests of Hainan province on latosol soils. The study also strongly supported the placement of the *H. hainanensis_GX* sample with *H. tetratepala*, suggesting that the distributions of *H. hainanensis* and *H. tetratepala* need further study.

Chloroplast capture resulting from hybridization and introgression (Rieseberg & Soltis, 1991; Tsitrone et al., 2003)

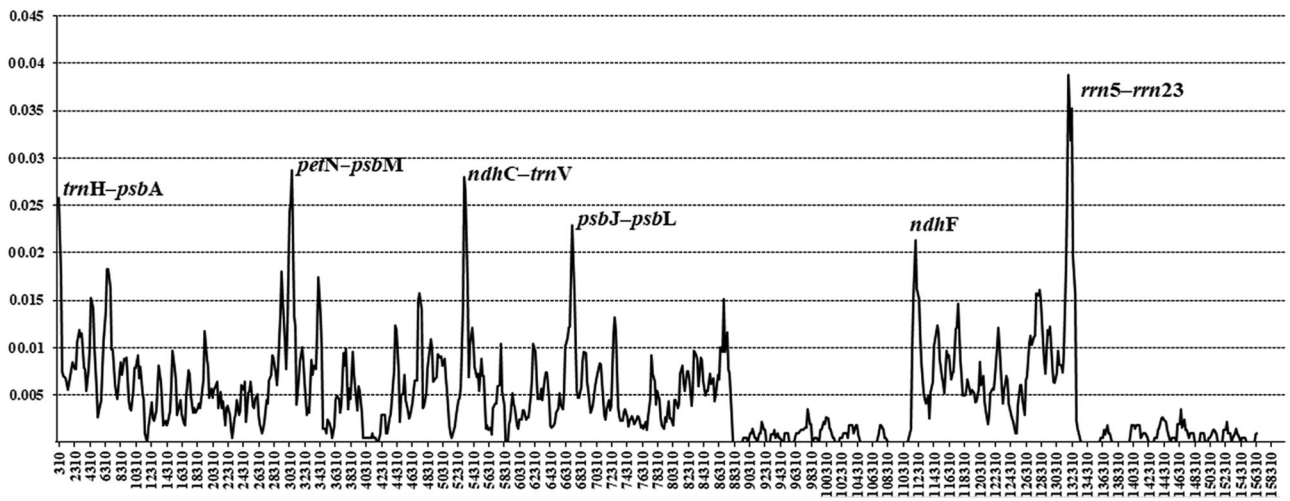

Fig. 2. Sliding window analysis of the whole chloroplast genome of seven Myristicaceae species, *Horsfieldia prainii*, *H. tetratepala*, *H. hainanensis*, *H. amygdalina*, *Myristica yunnanensis*, *M. fragrans*, and *Knema tenuinervia* (window length, 600 bp; step size, 200 bp). *X*-axis, position of the midpoint of a window; *Y*-axis, nucleotide diversity of each window (π). The four mutation hotspot regions (π > 0.02) are annotated.

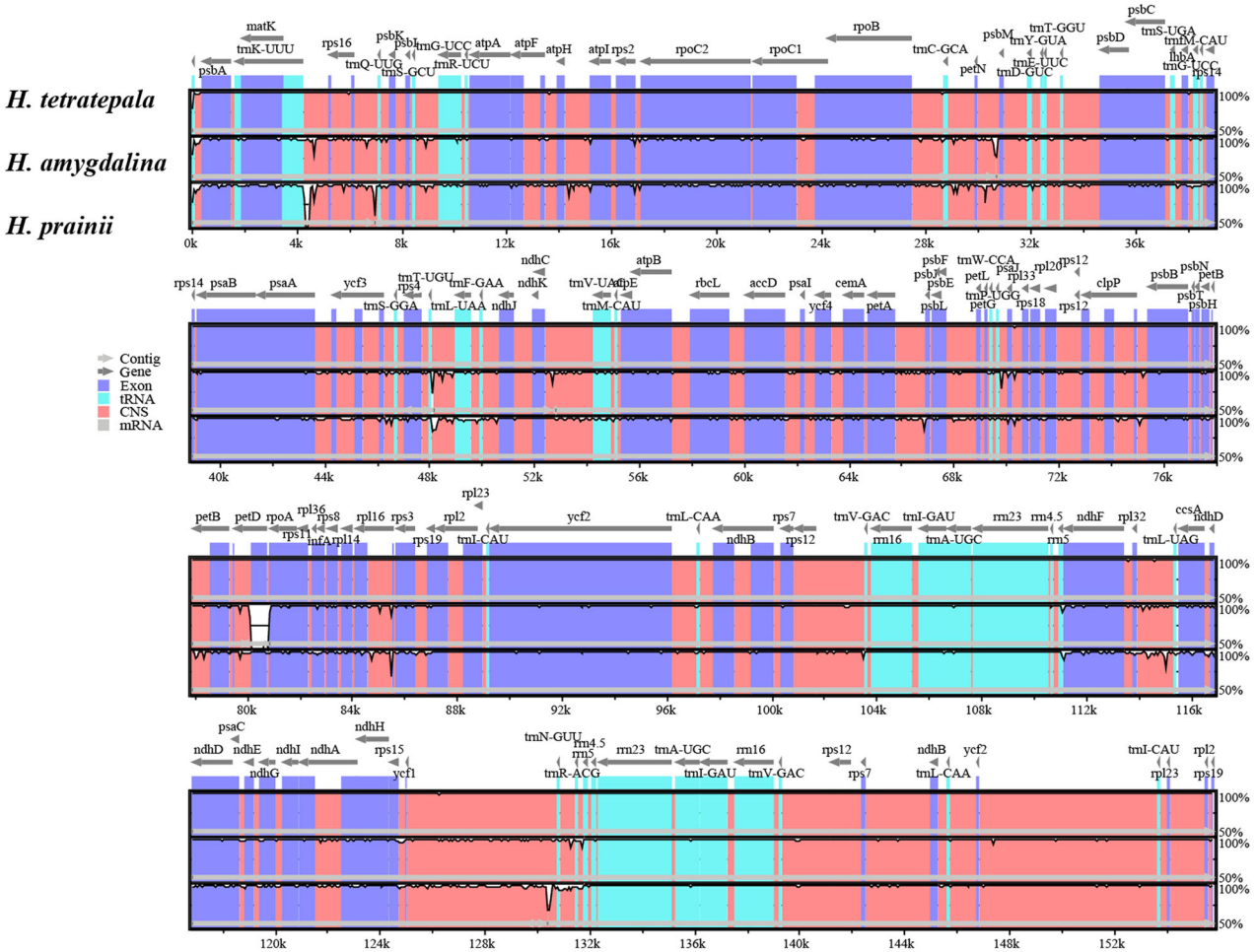

Fig. 3. Visualization alignment of four *Horsfieldia* chloroplast genome sequences, with the chloroplast genome of *H. hainanensis_HN* used as a reference. Vertical scale indicates the percentage of identity, ranging from 50% to 100%; the horizontal axis indicates the coordinates within the chloroplast genomes. Genome regions are color-coded as exon, transfer RNA (tRNA), messenger RNA (mRNA), and CNS (conserved non-coding sequences).

could be another explanation why the chloroplast genome analysis does not appear to reflect the species phylogeny. However, the *H. hainanensis_GX* and *H. tetratepala* samples have many similar morphological characteristics (e.g., the color of inner fruit wall and arils, leaf, and the number of lateral veins) and they grow in ecologically similar dense limestone karst forests of Guangxi and Yunnan provinces, respectively. The non-overlapping distributions between the two species also reduce the opportunities for hybridization. Accordingly, chloroplast capture seems unlikely as an explanation for the results seen in this study and the *H. hainanensis_GX* sample might simply represent a case of misidentification. Nevertheless, further research combining multiple molecular tools (e.g., nuclear DNA sequences) together with more comprehensive sampling is needed to determine if chloroplast capture does occur in this genus.

Although unequivocal Chinese samples of *H. kingii* were not available for our study, the reported aril and inner fruit wall colors of *H. kingii* differ from those of both *H. tetratepala* and *H. hainanensis* (Hou & Huang, 1964), as *H. kingii* has bright yellow arils and pinkish inner fruit walls (Fig. S3A; Datta & Rane, 2013), whereas *H. tetratepala* (Fig. S3B) and *H. hainanensis* (Hou & Huang, 1964) have yellow inner fruit walls and red arils. This supports the idea that *H. kingii* is a separate species, but its taxonomic status and presence in China requires further investigation. Such a study would also require access to verified samples of the *H. kingii* complex across its wider geographic range, extending into eastern Nepal and India.

Because of imprecise locality descriptions for historical herbarium collections, the danger of mudslides, and a lack of permits to some areas, only six of the putative Chinese *H. kingii* locations were visited, with plants found at only one of the sites, where two different samples were collected. These two field-identified *H. kingii* samples were found to have identical chloroplast genomes to *H. prainii*. However, as key reproductive characters were unavailable at the time of investigation, more work is required to clarify their identification. Nevertheless, our preliminary results imply that *H. kingii* might be too rare in China to be located in the wild easily, is locally extinct, or possibly represents a Nepalese and Indian taxon that has been confused in the past with Chinese species.

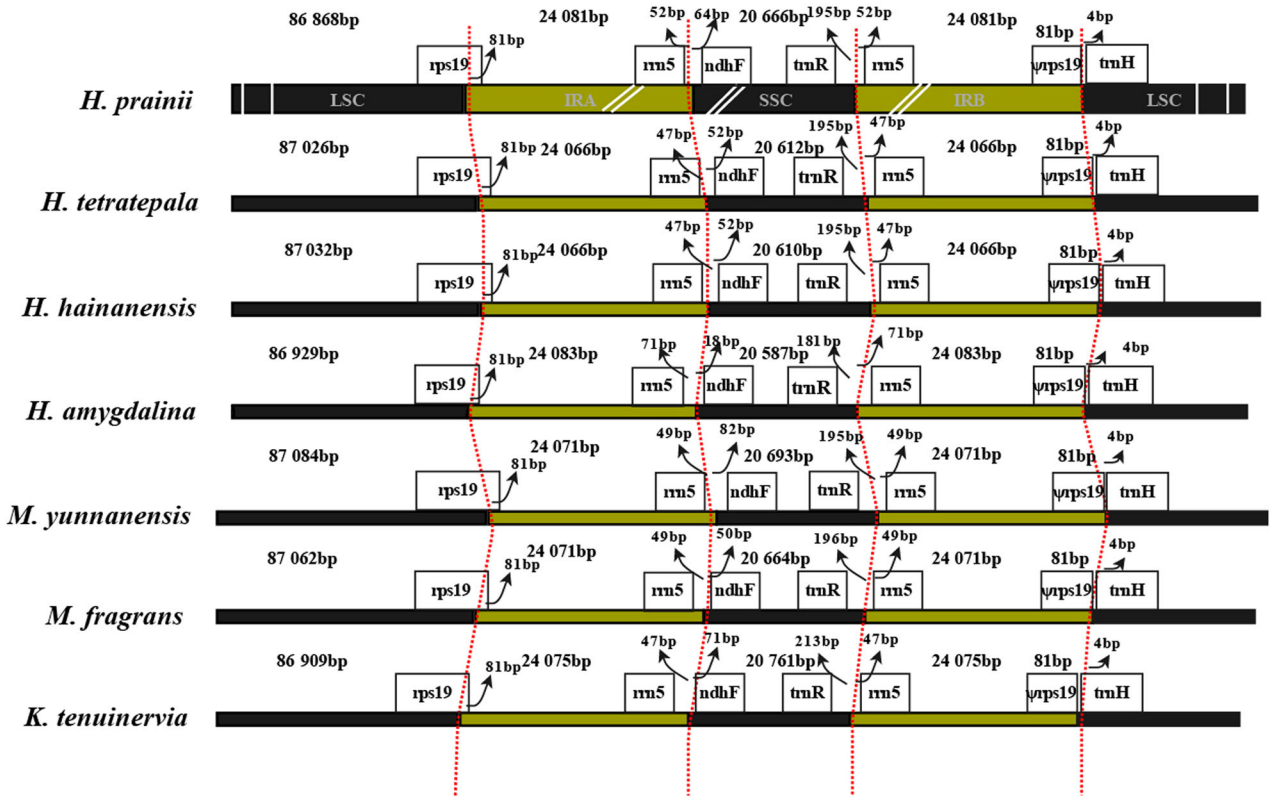

Fig. 4. Comparison of the inverted repeat/single copy boundary regions among the seven sequenced Myristicaceae chloroplast genomes. Numbers above the gene features refer to the distance between the ends of genes and the border sites. Ψ, pseudogene (*Horsfieldia hainanensis* collected from Hainan province, China). Figure is not to scale.

The analyses also show that although *Horsfieldia* chloroplast genomes share many similarities in terms of structure, gene order, gene content, and genome size (Fig. S1), there are slight size discrepancies between the species. The total length of the *Horsfieldia* chloroplast genome was 155 682–155 775 bp, making it 4310–4403 bp shorter than the previously reported value for *Parakmeria yunnanensis* Hu (Yang et al., 2014), albeit with the IR 2502–2519 bp

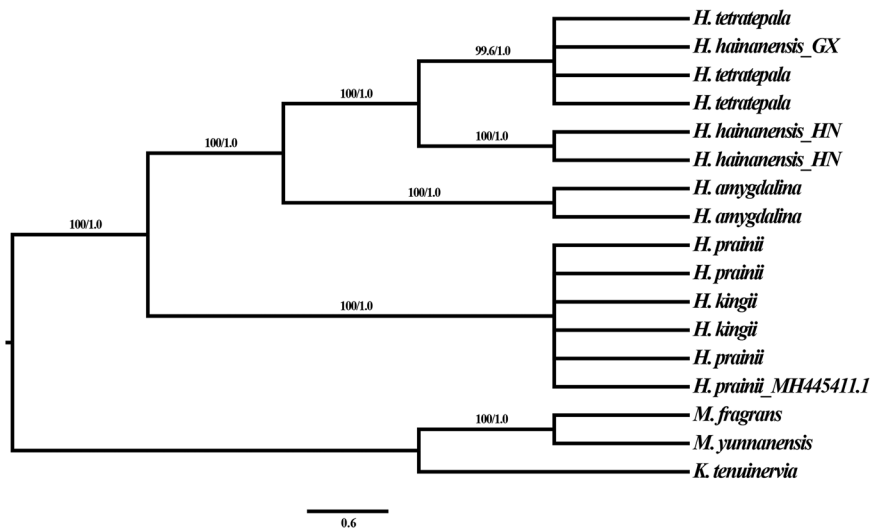

Fig. 5. Phylogenetic relationships based on complete chloroplast genome sequences of 16 Myristicaceae species (*Horsfieldia prainii*, *H. tetratepala*, *H. hainanensis*, *H. amygdalina*, *Myristica yunnanensis*, *M. fragrans*, and *Knema tenuinervia*) with maximum likelihood and Bayesian inference. Support values are bootstrap values (>50%, before slash) and posterior probability (>0.5, after slash), respectively. Sampling sites: *GX*, Guangxi province; *HN*, Hainan province.

shorter, but SSC 1832–1911 bp longer, giving four distinct chloroplast contractions relative to the IR of *P. yunnanensis* (IR = 26 585 bp). Inverted repeats can fluctuate in size by duplicating genes or single-copy DNA segments, or by losing duplicated sequences (Plunkett & Downie, 2000). As a result, the observed IR/SSC boundary regions are quite variable for the gene locations in the *Horsfieldia* species (Fig. 4) and the genes of the IR/SSC boundary regions are different from chloroplast genomes of other species. For example, the boundaries between the SSC regions and IR regions in the reported chloroplast genomes were located in gene *ycf1* (Raubeson et al., 2007). Our results show that gene *rrn5* nearest to the SSC/IR boundary and very short intergenic sequences were located at these boundaries. However, as it is uncertain whether these features are limited to the species in this study or represent a general pattern for *Horsfieldia*, sequencing of more species is needed to determine the nature of IR length variation within this genus.

Myristicaceae are an important family of pantropical rainforest trees, but relatively little is known about their phylogenetic relationships, partly due to the low levels of molecular divergence seen for traditional phylogenetic sequence markers (Sauquet et al., 2003; Doyle et al., 2004). As a result, phylogenetic inference within Myristicaceae using markers has proved unsuccessful, with 12 regions (*ndhF*, *trnK* 5′-intron, *mat*K, *trnK* 3′-intron, *trnT-trnL*, *trnL*, *trnL-trnF*, *atpB*, *rbcL*, 18S rDNA, *atp1*, and *matR*) showing only poor phylogenetic signal (Sauquet et al., 2003; Doyle et al., 2004). Similarly, Newmaster et al. (2010) evaluated seven plant barcode regions (UPA, *rpoB*, *rpoC1*, *accD*, *rbcL, matK,* and *trnH-psbA*) for the family, finding only the *trnH-psbA* intergenic spacer to be useful as a chloroplast barcode locus. The development of high-throughput sequencing technology and the ease of chloroplast genome sequencing means it is now possible to screen for taxon-specific DNA barcodes by searching the whole chloroplast genomes (Dong et al., 2013, 2014). Accordingly, using the chloroplast genomes recovered here, we were able to identify six variable loci (*petN-psbM*, *trnH-psbA*, *ndhC-trnV*, *psbJ-psbL*, *ndhF*, and *rrn5-rrn23*) which appear to be useful for the development of taxon-specific markers for future phylogenetic analyses within Myristicaceae.

Relatively little is known about the phylogeographic and biogeographic history of Myristicaceae, partly because there is very poor representation of the family in the fossil record (Doyle et al., 2004) and although fossil Myristicaceae wood (Boureau, 1950), leaves (Wolfe, 1977), pollen (Frederiksen, 1973; Jan du Chene et al., 1978), flowers (Poinar & Steeves, 2013), and fruits and seeds (Berry, 1929; Doyle et al., 2008) have been described, many of these cannot be assigned to the family with confidence (Doyle et al., 2008). The oldest diagnostic fossils of Myristicaceae are mid-Cenozoic-aged flowers from Dominican amber (Poinar & Steeves, 2013), but the suggested much earlier divergence ages for other magnoliid families (e.g., Zeng et al., 2014; Massoni et al., 2015b) and presence of well-supported Cretaceous fossils in sister lineages (Massoni et al., 2015a), combined with the relatively large, animal-dispersed diaspores of Myristicaceae that make long-distance dispersal less likely (Doyle et al., 2004), imply that the family are probably much older. Nevertheless, there are issues with apparently very low rates of evolution in Myristicaceae, suggesting that they could have diverged as late as the early Eocene (Massoni et al., 2015b). Thus, although this is an obvious area for future research once more members of the family have been investigated, at the present time we cannot determine reliable divergence dates for Chinese *Horsfieldia* species based on the whole chloroplast genomes presented here.

The taxonomic status of *Horsfieldia* species in China is unresolved and the specimens definitively assignable to *H. kingii* were not found in our field survey. Out of the 20 available Chinese herbarium records identified as *H. kingii*, 16 were more than 30 years old, one was 18 years old, and three were 5 years old, indicating its apparent rarity (or at least undercollection). Similarly, fieldwork investigations found that *H. hainanensis* and *H. tetratepala* were also rare and only encountered occasionally in remote areas and natural reserves. This situation is largely due to ongoing excessive logging and transition of forests into farmlands, resulting in wild Chinese *Horsfieldia* populations declining dramatically over the past few decades, with *H. kingii*, *H. hainanensis*, and *H. tetratepala* all now listed in the *China Species Red List* (Wang & Xie, 2004) and the *Conservation Program for Wild Plants with Extremely Small Populations in China* (State Forestry Administration of China, 2011), indicating that strict protective actions are required urgently if these taxa are to be conserved effectively.

Recent studies have shown that whole chloroplast genome sequences can serve as super-barcodes to provide suitable information for species identification, even in taxonomically difficult groups (e.g., Sato et al., 1999; Nock et al., 2011; Song et al., 2017; Chen et al., 2018; Wang et al., 2018). Comparing the sharply falling sequencing cost for a complete chloroplast genome against amplification of possibly dozens of gene fragments and their sequencing, the choice of applying the whole chloroplast genome as a super-barcode to species delimitation and phylogenetic studies in *Horsfieldia* looks to be highly cost-effective. Because complete chloroplast genome sequences can provide vastly increased sequence information for identifying different *Horsfieldia* species and their phylogenetic relationships, chloroplast genome data could serve as good candidates for high resolution DNA barcoding.

In addition, SSR markers combined with the highly variable chloroplast loci could be used to assess the genetic diversity within and between populations of endangered species for the formulation of more effective conservation strategies. As there are more than 100 species of *Horsfieldia*, the current study only represents a very limited sample and many more chloroplast genomes are required. Nevertheless, our results indicated that chloroplast genome sequences can help to increase species resolution and delimitation in the genus.

5 Conclusion

This study reports a chloroplast whole genome-level phylogenetic study of *Horsfieldia* and details the chloroplast genomes for seven species of Myristicaceae. The analyses support the recognition of at least four *Horsfieldia* species in China: *H. amygdalina*, *H. prainii*, *H. hainanensis*, and *H. tetratepala*. As unequivocal material of the rare *H. kingii*

was not collected for this study, its status in China is still unresolved and although it might still be found in remote areas, it might also have become extinct in recent decades, or possibly represents a misapplied name in the region. However, as key reproductive morphological characters were unavailable for samples attributed to this taxon at the time of the current investigation, more extensive fieldwork is required to clarify whether Chinese material of this species is still (or ever was) present.

The complete chloroplast sequences obtained in this study will also provide valuable data for wider studies into the phylogeny, taxonomy, and species delimitation of *Horsfieldia* across its range, using entire chloroplast genomes as a plant super-barcode. In particular, the six identified highly variable gene regions (*petN-psbM*, *trnH-psbA*, *ndhC-trnV*, *psbJ-psbL*, *ndhF*, and *rrn5-rrn23*) show good potential for use as DNA barcodes for phylogenetic and biogeographic studies in Myristicaceae, as well as for future investigations into the conservation genetics of these endangered Chinese rainforest trees.

Acknowledgements

We thank the forestry departments and nature reserves from the provinces and municipalities of Guangxi, Yunnan, and Hainan for their support and coordination of the fieldwork. We also thank H. H. Meng, C. Y. Zhang, Q. X. Hou, H. Ma, J. L. Hu, Z. Y. Liu, H. S. Bai, Q. Chen, H. Q. Chen, Y. W. Liang, C. H. Fu, H. J. Fan, D. Li, Z. W. Yang, Z. L. Liang, Z. Q. Nong, H. X. Mi, and J. G. Zhang for fieldwork assistance. This study was supported by the Science and Technology Basic Resources Investigation Program of China: Survey and Germplasm Conservation of Plant Species with Extremely Small Populations in South-west China (2017FY100100), Biodiversity Conservation Program of Chinese Academy of Sciences (ZSSD-013), and 135 Program of Chinese Academy of Sciences (No. 2017XTBG-T03).

References

Berry E. 1929. Early tertiary fruits and seeds from Belen, Peru. *Johns Hopkins University Studies in Geology* 10: 137–180.

Boureau E. 1950. Étude paléoxylologique du Sahara (IX). Sur un Myristicoxylon princeps n. gen., n. sp., du Danien d'Asselar (Sahara soudanais). *Bulletin du Muséum National d'Histoire Naturelle* 2: 523–528.

Brown B, Emberson RM, Paterson AM. 1999. Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification. *Bulletin of Entomological Research* 89: 287–294.

Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. 2003. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* 19: i54–i62.

Bucklin A, Guarnieri M, Hill RS, Bentley AM, Kaartvedt S. 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *Hydrobiology* 401: 239–254.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA* 106: 12794–12797.

Chen XL, Zhou JG, Cui YX, Wang Y, Duan BZ, Yao H. 2018. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. *Frontiers in Pharmacology* 9: 695.

Dang JL, Yang XB, Huang YF, Ye F, Luo T, Chen SL, Liu B, Bai YF. 2009. GC-MS analysis on the chemical constituents of essential oil from bark of *Horsfieldia hainanensis* Merr. *Journal of Chinese Medicinal Materials* 325: 714–716. (in Chinese)

Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology* 17: 134.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Datta A, Rane A. 2013. Phenology, seed dispersal and regeneration patterns of *Horsfieldia kingii*, a rare wild nutmeg. *Tropical Conservation Science* 6: 674–689.

De Wilde WJJO. 1984. A new account of the genus *Horsfieldia* (Myristicaceae). *The Gardens' Bulletin Singapore* 37: 115–179.

De Wilde WJJO. 2000. Myristicaceae. In: Stevens PF ed. *Flora Malesiana, Ser. I.* Leiden: National Herbarium Netherlands. 14: 1–632.

Dong WP, Liu H, Xu C, Zuo YJ, Chen ZJ, Zhou SL. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genetics* 15: 138.

Dong WP, Xu C, Cheng T, Lin K, Zhou SL. 2013. Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biology and Evolution* 5: 989–997.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Doyle JA, Manchester SR, Sauquet H. 2008. A seed related to Myristicaceae in the early Eocene of southern England. *Systematic Botany* 33: 636–646.

Doyle JA, Sauquet H, Scharaschkin T, Le Thomas A. 2004. Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences* 165: S55–S67.

Frederiksen N. 1973. New mid-Tertiary spores and pollen grains from Mississippi and Alabama. *Tulane Studies in Geology and Paleontology* 10: 65–86.

He GZ, Cai L, Liang G, Shi Q, Chen LW. 2013. Sowing and breeding technique of *Horsfieldia hainanensis* Merr. *Practical Forestry Technology* 6: 35–37. (in Chinese)

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

Hou KZ, Huang MX. 1964. *Flora of Hainan*. Beijing: Science Press. 1: 304. (in Chinese)

Hutter S, Vilella AJ, Rozas J. 2006. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7: 409.

Jan du Chene R, Onyike M, Sowunmi M. 1978. Some new Eocene pollen of the Ogwashi-Asabe formation, South-Eastern Nigeria. *Revista Espanola de Micropalontologia* 10: 285–322.

Jessup LW. 2007. Myristicaceae. In: Wilson AJG ed. *Flora of Australia*. Canberra: Australian Biological Resources Study. 2: 57–62.

Jiang YH, Liu XS, Xiang WH, Jiang Y, He YH. 2018. Genetic diversity and structure analysis of the endangered plant species *Horsfieldia hainanensis* Merr. in China. *Biotechnology & Biotechnological Equipment* 32: 95–101.

Jiang YH, Xiang WH, Jiang Y, He YH, Lin JY. 2016. Floristic composition, structure and phytogeographic characteristics of

Horsfieldia hainanensis Merr. community in Guangxi. *Journal of Beijing Forestry University* 381: 74–82. (in Chinese)

Jin JJ, Yu WB, Yang JB, Song Y, Li DZ, Yi TS. 2018. GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv*. doi: 10.1101/256479.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Li YH. 1979. Myristicaceae. In: *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press. 30: 176–205.

Li BT, Wilson TK. 2008. Myristicaceae. In: Wu Z-Y, Raven PH, Hong D-Y eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 7: 99–101.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW — A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41: W575–W581.

Ma Q, Zhan R, Chen YG. 2014. Chemical constituents and biological activity of *Horsfieldia* plants. *Guangzhou Chemical Industry* 42: 11–13. (in Chinese)

Mao CL, Zhang FL, Li XQ, Yang T, Liu J, Wu Y. 2019. The complete chloroplast genome sequence of *Horsfieldia pandurifolia* (Myristicaceae). *Mitochondrial DNA Part B Resources* 4: 949–950.

Massoni J, Couvreur TLP, Sauquet H. 2015a. Five major shifts of diversification through the long evolutionary history of Magnoliidae (angiosperms). *BMC Evolutionary Biology* 15: 49.

Massoni J, Doyle JA, Sauquet H. 2015b. Fossil calibration of Magnoliidae, an ancient lineage of angiosperms. *Palaeontologia Electronica* 18.1.2FC: 1–25.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16: 1046–1047.

Merrill ED. 1932. A fourth supplementary list of Hainan plants. *Lingnan Science Journal* 111: 37–61.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. 2010. Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* 8: 480–490.

Niu YT, Jabbour F, Barrett RL, Ye JF, Zhang ZZ, Lu KQ, Lu LM, Chen ZD. 2018. Combining complete chloroplast genome sequences with target loci data and morphology to resolve species limits in *Triplostegia* (Caprifoliaceae). *Molecular Phylogenetics and Evolution* 129: 15–26.

Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* 9: 328–333.

Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* 19: 325–354.

Plunkett GM, Downie SR. 2000. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Systematic Botany* 25: 648–667.

Poinar G, Steeves R. 2013. *Virola dominicana* sp. nov. (Myristicaceae) from Dominican amber. *Botany* 91: 530–534.

Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade MH, Boore JL, Jansen RK. 2007. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174.

Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. *Plant Systematics and Evolution* 271: 101–122.

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

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Research* 6: 283–290.

Sauquet H, Doyle JA, Scharaschkin T, Borsch T, Hilu KW, Chatrou LW, Thomas AL. 2003. Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: Implications for character evolution. *Botanical Journal of the Linnean Society* 142: 125–186.

Simpson MG. 2010. Plant systematics: An overview. In: Simpson MG ed. *Plant systematics*. 2nd edn. Elsevier: Academic Press. 1: 3–16.

Sinclair J. 1975. The genus *Horsfieldia* (Myristicaceae) in and outside Malaysia II. *The Gardens' Bulletin Singapore* 28: 1–181.

Song Y, Wang SJ, Ding YM, Xu J, Li MF, Zhu SF, Chen NZ. 2017. Chloroplast genomic resource of *Paris* for species discrimination. *Scientific Reports* 7: 3427.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

State Forestry Administration of China. 2011. *The saving and conservation program on extremely small populations in China*. Beijing: State Forestry Administration of China. (non-published materials for government departments, in Chinese)

Steeves RAD. 2011. *An intrageneric and intraspecific study of morphological and genetic variation in the neotropical* Compsoneura *and* Virola (*Myristicaceae*). Ph.D. Dissertation. Guelph: University of Guelph.

Tsitrone A, Kirkpatrick M, Levin DA, Morgan M. 2003. A model for chloroplast capture. *Evolution* 57: 1776–1782.

Wang AS, Wu HW, Zhu XC, Lin JM. 2018. Species identification of *Conyza bonariensis* assisted by chloroplast genome sequencing. *Frontiers in Genetics* 9: 374.

Wang S, Xie Y. 2004. China Species Red List Vol. I Red List. Beijing: Higher Education Press. (in Chinese)

Wilkinson MJ, Szabo C, Ford CS, Yarom Y, Croxford AE, Camp A, Gooding P. 2017. Replacing Sanger with next generation sequencing to improve coverage and quality of reference DNA barcodes for plants. *Scientific Reports* 7: 46040.

Wolfe J. 1977. Paleogene floras from the Gulf of Alaska region. *US Geological Survey Professional Paper* 997: 1–108.

Wu CY, Wang WT. 1957. Preliminary report of flora study on tropical and subtropical areas, Yunnan, China. I. *Acta Phytotaxonomica Sinica* 6: 183–254. (in Chinese)

Wu Y, Mao CL, Zhang FL, Yang XL, Zeng JS, Duan AA. 2015. Taxonomic position of *Horsfieldia pandurifolia* Hu (Myristicaceae). *Bulletin of Botanical Research* 35: 652–659. (in Chinese)

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Xu C, Dong WP, Li WQ, Lu YZ, Xie XM, Jin XB, Shi JP, He KH, Suo ZL. 2017. Comparative analysis of six Lagerstroemia complete chloroplast genomes. *Frontiers in Plant Science* 8: 15.

Xu YL, Cai NH, Wu Y, Duan AA. 2012. Fatty acid composition of several plants of *Horsfieldia*. *China Oils and Fats* 37: 80–82. (in Chinese)

Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources* 14: 1024–1031.

Yang YC, Zhou T, Duan D, Yang J, Feng L, Zhao GF. 2016. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science* 7: 959.

Ye M. 2004. *Systematics of Myristicaceae from China*. Ph.D. Dissertation. Guangzhou: South China Agricultural University. (in Chinese)

Zeng LP, Zhang Q, Sun RR, Kong HZ, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* 5: 4956.

Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, Zhang W, Kim K, Lee S, Yang T, Wang Y. 2016. The complete chloroplast genome sequences of five *Epimedium* species: Lights into phylogenetic and taxonomic analyses. *Frontiers in Plant Science* 7: 306.

Zhong SY, Chen GD, Qiu MH, Liu MH, Lin L, Lin ZW. 2018. Investigation on the geographical distribution and habitat characteristics of *Horsfieldia hainanensis* in Hainan island. *Journal of Fujian Forestry Science & Technology* 45: 82–86+106. (in Chinese)

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12556/suppinfo:

Table S1. Sample collection information of chloroplast genome sequences for the 16 samples from Myristicaceae.
Table S2. Gene contents in *H. hainanensis_HN* chloroplast genome.
Table S3. Differences between the *H. hainanensis_GX* and *H. tetratepala* chloroplast genomes.
Fig. S1. Gene map of the four *Horsfieldia* chloroplast genomes. The genes shown outside of the circle are transcribed clockwise, those inside are transcribed counter-clockwise (*H. hainanensis* collected from Hainan province).
Fig. S2. Phylogenetic relationships with branch lengths based on complete chloroplast genome sequences of 16 Myristicaceae species with maximum likelihood and Bayesian inference.
Fig. S3. Fruits of two *Horsfieldia* species in the *H. kingii* species complex showing differences in internal fruit wall and aril color. (**A**) Indian sample of *H. kingii* with pink inner fruit walls and yellow arils (Source: Nandini Velho CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=5524809); (**B**) *H. tetratepala* with yellow inner fruit walls and red arils (Source: Jinping County, Yunnan Province, Chao-Nan Cai).