When tropical and subtropical congeners met

When tropical and subtropical congeners met: Multiple ancient hybridization events within *Eriobotrya* were detected in the Yunnan-Guizhou Plateau, a tropical-subtropical transition area in China

Sufang Chen[1*], Richard Milne[2*], Renchao Zhou[1*], Kaikai Meng[1], Qianyi Yin[1], Wei Guo[3], Yongpeng Ma[4], Kangshan Mao[5], Kewang Xu[6], Young-Dong Kim[7], Truong Van Do[8], Wenbo Liao[1], Qiang Fan[1]

[1] State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China
[2] Institute of Molecular Plant Sciences, University of Edinburgh, Edinburgh EH9 3JH, UK
[3] Department of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, People's Republic of China
[4] Kunming Botanical Garden, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China
[5] Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China
[6] Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China
[7] Department of Life Science, Multidisciplinary Genome Institute, Hallym University, Chuncheon City, South Korea
[8] Vietnam National Museum of Nature, Vietnam Academy of Science and Technology 18 Hoang Quoc Viet, Cau Giay, Ha Noi, Vietnam

[*] These authors contributed equally to this work
Authors for correspondence:
Qiang Fan, Email: fanqiang@mail.sysu.edu.cn
Wenbo Liao, Email: lsslwb@mail.sysu.edu.cn
Yongpeng MA, Email: mayongpeng@mail.kib.ac.cn
Kangshan Mao, Email: maokangshan@scu.edu.cn

Running Head: Extensive hybridization in *Eriobotrya*

ABSTRACT:

Global climate changes during the Miocene could have created ample opportunities for hybridization between members of tropical and subtropical biomes at the boundary between those zones, yet very few studies have tested this hypothesis. The Yunnan-Guizhou Plateau (YGP) in Southwest China is a biodiversity hotspot of vascular plants, located at a transitional area between the floristic regions of tropical Southeast Asia and subtropical East Asia. The genus *Eriobotrya* (Rosaceae) comprises both tropical and subtropical taxa, and has 12 species recorded from the YGP, making it a good system to test the hypothesis of between-biome hybridization. Therefore, we surveyed the evolutionary history of *Eriobotrya* by examining three chloroplast regions plus five nuclear genes for 817 individuals (47 populations) of 23 *Eriobotrya* species (including 19 populations of 12 species in the YGP), plus genome re-sequencing of 33 representative samples. We concluded that: 1) phylogenetic positions for 16 species (lineages) exhibited strong cyto-nuclear conflicts, most likely due to ancient hybridization; 2) The YGP is a hotspot for hybridization, where 11 species show clear evidence of chloroplast capture; 3) *Eriobotrya* likely originated in tropical Asia during the Eocene. From the Miocene onwards the intensification of the Eastern Asia Monsoon and global cooling may have caused shifts of the tropical-subtropical boundary and secondary contact between species, and thus provided ample opportunity to hybridization and diversification of *Eriobotrya* species, especially in the YGP. Our study highlights the significant role that paleo-climate changes likely played in driving hybridization and generating rich species diversity in climate transition zones.

Keywords: *Eriobotrya*; multiple ancient hybridization; chloroplast capture; genome resequencing; tropical and subtropical zones; global cooling

INTRODUCTION

Hybridization, defined as crosses between genetically distinct populations (Abbott et al., 2013) or distinct taxa (Rieseberg, 1997), has attracted rising interest among evolutionary biologists as an important process in the speciation of plants, animals, and fungi (Mallet, 2007; Soltis & Soltis, 2009). It has been estimated that as many as 25% of plant species hybridize naturally, though its prevalence in plant species can vary dramatically among regions and families (Ellstrand et al., 1996; Whitney et al., 2010). So far, research on hybridization has mainly focused on cataloging its frequency and exploring its evolutionary consequences (Goulet, Roda, & Hopkins, 2017), with much less attention to hot-spot areas where hybridization may be unusually common.

Since the advent of the Eastern Asian monsoon (EAM, approximately 22–25 million years ago [Ma]), the climate in southeastern China has shifted dramatically from arid to humid (Sun & Wang, 2005; Guo et al., 2008). Xerophilous plants dominant in the Paleogene were consequently replaced by subtropical evergreen broadleaved forests, neighboring the tropical evergreen broadleaved forests in the south (Tao, 2000). Subsequent global cooling from the late Miocene onwards caused

these subtropical forests to migrate progressively southwards, while maintaining frequent contact with tropical forests to the south (Jin et al., 2003, Zhu, 2013). This would have created ample opportunities for hybridization between closely related species across the tropical and subtropical zones. However, the consequences of such contact have rarely been explored, and hence little is known about whether hybridization occurred across this ecological boundary, how often it might have occurred, and what evolutionary consequences there might have been.

The Yunnan-Guizhou Plateau (YGP) in southwestern China is climatically and biogeographically located at a transitional area between tropical Southeast Asia and subtropical East Asia. Strongly affected by the uplift of the Himalayas and the formation of the eastern monsoon climate, the YGP harbors a great number of vascular plant species, and has been identified as one of the top biodiversity hotspots in the world (Barthlott, Lauer, & Placke, 1996). The YGP hence serves as an ideal region to investigate the role of hybridization in speciation and species diversity in areas where tropical and subtropical zones meet.

The genus *Eriobotrya* Lindl. (Rosaceae) is widely distributed across tropical and subtropical East Asia. Along with its sister genus *Rhaphiolepis* (which might not be distinct; Liu et al. 2020), *Eriobotrya* diverged from *Cotoneaster* and *Heteromeles* approximately 37 million years ago (Zhang et al., 2017), indicating a relatively old origin. *Eriobotrya* comprises 29 accepted species, four varieties, and one form (POWO, www.ipni.orgpowo.science.kew.org, including one species that was recently described by Kang et al. (2021). Most *Eriobotrya* species are diploid, there is no known apomixis in the genus, and all are easily separable by morphology. The YGP is the most important distribution center of *Eriobotrya*, with at least 12 *Eriobotrya* species recorded there (Gu & Spongberg, 2003; Yang et al., 2005; Li et al., 2013). Therefore, *Eriobotrya* serves as a good system to study the roles that hybridization between species of tropical and subtropical zones has played in the evolutionary history of the YGP flora.

Within *Eriobotrya*, spontaneous hybrids (mostly F_1 generation) have been reported between *E. japonica* (Thunb.) Lindl. and *E. prinoides* Rehder & E.H.Wilson in Sichuan (China) (Fan et al. 2014), whereas Liu et al. (2020) detected strong cyto-nuclear incongruence in phylogenetic relationships among nine *Eriobotrya* species. Therefore, hybridization might be a significant evolutionary force within this genus, but a systematic examination is required to determine its extent, and evolutionary consequences.

In this study, we collected 817 individuals from 47 populations of 23 *Eriobotrya* species distributed in China, Vietnam, Myanmar, Nepal, and Malaysia, including 19 populations of 12 *Eriobotrya* species in the YGP. Based on Sanger sequencing of three chloroplast regions and five nuclear genes of these samples, haplotype network and Bayesian clustering analyses were performed to resolve genetic relationships among these species. Furthermore, shallow whole-genome Illumina sequencing was performed for 33 representative *Eriobotrya* samples and one *Rhaphiolepis* species to extract sequences of complete chloroplast genomes and 197 single-copy nuclear genes, which were used to reconstruct chloroplast and nuclear phylogenetic trees with

better resolution. By further integration of geographical distribution, ancient climate and geological changes—we aimed to (i) infer the origin of *Eriobotrya* and its subsequent dispersal routes, (ii) identify instances of past hybridization and thereby infer its roles in the evolutionary history of *Eriobotrya*, and hence (iii) determine to what extent hybridization occurred at the boundary of tropical and subtropical zones in the YGP and elsewhere, and what effect this might have had on the evolution of *Eriobotrya* within either biome.

MATERIAL AND METHODS

Sample collection

A total of 817 individuals were sampled from 47 populations of 23 *Eriobotrya* species distributed in China, Vietnam, Myanmar, Malaysia, and Nepal (Fig. 1, Fig. 2 and Table S1), plus 41 individuals of *Rhaphiolepis*, *Cotoneaster*, *Photinia*, and *Sorbus* (Table S2). Each population was ascribed to one of three geographical areas based on its position relative the boundary between tropical and subtropical zones (Institute of Geography & Chinese Academy of Sciences, 1959; Zhu, 2013) (Table S1): subtropical China (SUB), and tropical Asia (TRO), and the intermediate area between them (INT).

From each individual, fresh leaf material was dried and stored with silica gel. The geographic location of each population was recorded using a Garmin GPS unit (GPSMAP 62sc, Taiwan) within a margin of 10 m. The voucher specimens for each population were deposited in the Herbarium of Sun Yat-sen University (SYS). Total genomic DNA was extracted using a modified CTAB method (Huang, Ge, & Sun, 2000).

Sequence analysis of the population samples

Three chloroplast regions (*rbc*L, psbB, and *rpl*20-*rps*12) and five nuclear genes (*PXMP2*, *C23H*, *PEPC*, *GDSL2*, and *TPP2*) were sequenced for the 817 *Eriobotrya* and 41 outgroup samples, using primers designed according to previous studies (Fay, Swensen, & Chase, 1997; Heinze, 2007; Fan et al., 2014; Li et al., 2017) (Table S3, q.v. for gene ontology annotations). PCR amplification and Sanger sequencing were performed according to Fan et al. (2014).

The original sequences were assembled and edited using SeqMan$^{TM\ II}$ (DNASTAR, Inc., Madison, WI). Indels were identified by sequencing and checking the sequence chromatogram from both strands. Haplotype inference was implemented with DNAsp v5.10.01 (Librado & Rozas, 2009). Haplotype networks were constructed for the concatenated chloroplast regions, and for each of the five nuclear genes, using Network 4.6.1.2 (www.fluxus-engineering.com) with the median-joining algorithm (Bandelt, Forster, & Rohl, 1999).

To determine whether each *Eriobotrya* species maintains a distinct gene pool, or shows the signal of genetic admixture, Bayesian cluster analyses (Fuchs et al., 2015) were conducted using STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000). Individuals with >2 missing genes were excluded and then polymorphic sites for the five nuclear genes were extracted with DNAsp. To minimize the effect of linkage disequilibrium (LD), sites with $r > 0.2$ (calculated with DNAsp) were removed (Xu et

al., 2019). The remaining polymorphic sites from all five genes were concatenated and transformed into STRUCTURE input data format by replacing A, T, G, and C with 0, 1, 2, and 3 respectively, using a custom Perl script. Next, STRUCTURE analysis was performed on 285 polymorphic sites across 723 individuals of 23 *Eriobotrya* species to investigate species delimitation and population structure, setting an admixture model with correlated allele frequency, a 100,000 burn-in period, and 6 iterations of 100,000 Markov chain Monte Carlo replicates per *K* (2–18). STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/, Earl, & Vonholdt, 2012) was then used to determine the optimum *K*. From this analysis, any gene pool to which more than one species was assigned was termed a species complex, and in all such cases, STRUCTURE analysis was further performed on that species complex alone to see whether its components can be discriminated from one another.

Genome re-sequencing and chloroplast genome assembly

For genome re-sequencing analysis, at least one sample was selected for each species except for three species that had already been sequenced, i. e. *E. japonica* (Jiang et al., 2020), *E. laoshanica* W.B.Liao, Q.Fan & S.F.Chen (Chen et al., 2020) and *E. shanense* D.H. Kang, H.G. Ong & Y.D. Kim (Kang et al., 2021). For any species whose chloroplast haplotypes had been attributed to >1 different clades, between 2 and 5 individuals were resequenced. A total of 33 *Eriobotrya* individuals from 20 species, plus one *Rhaphiolepis* individual (Table S1) were selected for genome re-sequencing. High-quality total DNA was extracted using the TIANamp Genomic DNA Kit (TIANGEN Biotech Co. Ltd, Beijing), and then sent to Novogene Co., Ltd. (Beijing, China) for genome re-sequencing using an Illumina Hiseq X Ten platform. Approximately 6 Gb of clean data were obtained for each individual and then assembled into a circled chloroplast genome with NOVOPlasty 2.7.2 (Dierckxsens, Mardulyn, & Smits, 2017), using the chloroplast genome and *rbcL* gene of *E. japonica* (GenBank accession: KY085905) as the reference and seed, respectively.

Phylogenetic reconstruction based on chloroplast genomes

The 34 chloroplast genomes assembled in this study, together with 21 chloroplast genomes of *Eriobotrya*, *Rhaphiolepis,* and 28 other Rosaceae species downloaded from the NCBI website (Table S4), were aligned using MAFFT v7.407 (Katoh & Standley, 2013) and then manually checked and adjusted using Mega-X (Kumar et al., 2018). After removing gaps, the aligned chloroplast data were used to reconstruct phylogenetic trees based on maximum likelihood (ML) using IQ-TREE v1.6.8 (Nguyen et al., 2015) and Bayesian inference (BI) using MrBayes 3.2 (Ronquist et al., 2012). The best-fit model of DNA substitution was determined using IQ-TREE by calculating the Bayes information criteria. For Bayesian inference, one cold and three incrementally heated Markov chain Monte Carlo (MCMC) chains were run for 10,000,000 generations with 25% as burn-in. Four *Amelanchier* spp. were set as the outgroup species based on Zhang et al., (2017).

Identification of single-copy nuclear genes

Coding-region sequences of 959 single copy nuclear genes shared by *Arabidopsis*,

Populus, *Vitis*, and *Oryza* (Duarte et al., 2010) were extracted from the *Arabidopsis thaliana* cDNA library (http://www.plantgdb.org/HKDB/phplib/download.php?GDB=At) using a custom Perl script. A BLASTn search was then performed against the cDNA library of *E. japonica* (Jiang et al., 2020), and 676 top BLAST hits were obtained. Complete sequences of these 676 genes (including introns and UTRs) were then extracted from the whole genome sequences of *E. japonica* according to gene annotations of their cDNA. Using these genomic sequences as the query, a BLASTn search was performed against the whole genome sequences of *E. japonica* to see whether additional copies existed in the genome. Each gene was considered to be a single-copy gene in *Eriobotrya*, if no additional copy was detected with a score > 400 and an identity > 90. Manual checks were further performed to ensure that these BLAST hits were located on the 17 chromosomes of *E. japonica*. As a result of this, 197 nuclear genes were treated as single-copy in the genome of *Eriobotrya*.

Genome-wide SNPs calling

Illumina sequencing reads were obtained from 42 individuals in total, comprising 33 *Eriobotrya* individuals and one *Rhaphiolepis philippinensis* (S.Vidal) Kalkman individual sequenced in this study, plus eight others: three *Cotoneaster* species (Meng et al., 2021), two samples of *E. laoshanica* (Chen et al., 2020), and one accession each of *Crataegus mollis* (Torr. & A.Gray) Scheele (NCBI SRA accession: SRR3157040) and *Malus domestica* Borkh. (NCBI SRA accession: DRR183255), and *E. japonica* (Jiang et al., 2020). These were cleaned using fastp (Chen et al., 2018) by removing reads containing five or more unknown "N" bases or those having more than 10% of bases with a Q value < 20.

Next, using the whole genomic sequences of *E. japonica* as references, we used BWA 0.7.12-r1039 (Li, 2013) to map the short reads, SAMtools v1.9 (Li, 2011) to transform the resulting SAM files into BAM files and remove the duplicate reads, BCFtools v1.9 (http://samtools.github.io/bcftools/) to call SNPs, and also to filter sites with low quality (<20), low depth (<5), or those neighboring gaps (<5 bp), then to merge all outputs, and finally to extract SNPs across all samples. Gaps were not considered in the following analyses.

Phylogenetic reconstruction based on SNPs in the 197 nuclear genes

Based on information from gene annotations, BCFtools was used to extract SNPs from the 197 single-copy nuclear genes across the 42 samples. After removing sites shared by fewer than 10 samples, bases in each SNP site were extracted for each gene with a custom Perl script, generating 197 FASTA files for phylogenetic tree reconstruction.

Several methods that model ILS using multi-locus sequence data have been widely applied to infer species trees, and here we choose ASTRAL (Zhang et al., 2018) to construct a species tree. A total of 197 phylogenetic trees were first reconstructed, one for each gene, using the ML method implemented in IQ-TREE. From these the species tree was then reconstructed using ASTRAL v5.7.5, with local posterior probability (PP) and quartet score (QS) calculated for each node.

At the same time, a phylogenetic tree was reconstructed using the ML method based

on the concatenated SNPs from the 197 single-copy genes. Best models for each nuclear gene, and for the concatenated 197 genes, were auto detected with IQ-TREE, setting "-m MFP+ASC".

Testing for incongruence between chloroplast and nuclear phylogenetic trees

We first identified potential conflict signals by visually comparing support for any incongruence between the topology based on the chloroplast genome, and that based on the 197 nuclear genes. To simplify the analyses, we constructed simplified chloroplast and nuclear databases (SM datasets) by removal of all the species showing strong incongruence, and constructed ML trees from each such dataset. For each species that showed incongruence between chloroplast and nuclear data, its sequences were added to the SM datasets, and two Netwick trees were constructed based on chloroplast or nuclear data. This was done independently for each such species. Tree incongruence was then assessed using AU and weighted HW tests implemented in IQTREE.

Incongruence among gene trees could be explained by introgressive hybridization, or incomplete lineage sorting (ILS) (Maddison, 1997). To test whether ILS alone can explain incongruence between chloroplast and nuclear phylogenetic trees, DendroPy 4.4.0 (Sukumaran & Holder, 2010) was applied to simulate 20,000 chloroplast trees under the coalescent scenario, using the species tree as a guide tree with branch lengths scaled by four to account for the fact that any uni-parentally inherited organelle genome has one fourth of the effective population size of the bi-parentally inherited autosomal nuclear genes (Folk, Mandel, & Freudenstein, 2017, Li et al. 2021). Clade frequency of the simulated chloroplast genes was summarized on the empirical chloroplast tree with Phyparts (Smith et al., 2015). It was assumed that clades from the chloroplast tree should be present in the simulated gene trees with a high frequency under a scenario of ILS alone, whereas under a scenario of hybridization they would be unique to the empirical chloroplast tree and be absent or at low frequency in the simulated gene trees.

Gene flow detection

The program Dsuite (Malinsky, Matschiner, & Svardal, 2021) facilitates rapid genome scale estimations of gene flow across tens or even hundreds of populations or species directly based on the four-taxon D-statistic. We set as inputs the VCF file generated from genome-wide SNPs calling, plus the ML tree produced from the concatenated 197 nuclear genes, with *R. philippinensis* designated as outgroup, then Dsuite was used to calculate the D statistic, *f* (the admixture fraction), and test whether D is significantly different from zero. The D-statistic estimates were then visualized with the Ruby script "plot_d.rb" (https://github.com/mmatschiner), in which each square represented the highest estimate of D for each combination of P2 and P3 taxa.

Species network analysis

Due to computer limitations, the 33 *Eriobotrya* samples sequenced in this study and two samples of *E. laoshanica* sequenced previously (Chen et al., 2020) were divided into small datasets containing 10-15 samples to detect possible reticulation events using PhyloNet 3.8.0 (Wen et al., 2018). In addition, 197 individual gene trees

were generated for each dataset based on the ML method implemented in IQTREE. Finally, species networks that model ILS and gene flow were inferred with PhyloNet. Network searches were performed with the command "InferNetwork_MPL" allowing nodes with support > 0.8 (-b 0.8), between 0 and 4 hybridization events, and 50 runs of the search. Whether diversification in *Eriobotrya* was tree-like or reticulate was indicated by the log-pseudolikelihood values.

Based on the above analyses, six possible reticulation events were revealed: one was detected between species in tropical Asia, one in subtropical China, and four in the intermediate area. Three datasets were then assembled to show all possible reticulation events according to the geographic distributions of sampled material (Table S1) and their phylogenetic relationships. Dataset TRO comprised all populations collected from the geographical area TRO, plus two populations from the geographical area INT that fell within the same clade, with *Cotoneaster* Medikus set as outgroup for this clade. Dataset INT comprised all remaining populations collected from the geographical area INT, with *E. bengalensis* (Roxb.) Hook. f. was set as the outgroup. Finally, dataset SUB comprised nine populations collected from the geographical area SUB, belonging to three species all of which also have populations in TRO or INT, i.e. *E. cavaleriei* (H. Léveillé) Rehder, *E. fragrans* Champion ex Bentham and *E. deflexa* (Hemsl.) Nakai., with *E. bengalensis* again set as outgroup.

Divergence time estimation

Complete chloroplast genomes of the 33 *Eriobotrya* samples, one *Rhaphiolepis* sample, and 49 chloroplast genomes downloaded from the NCBI website were aligned with MAFFT. After removing gaps, Bayesian analyses were performed using BEAST version 1.8.4 (Drummond et al., 2012), setting an uncorrelated lognomal relaxed clock, the GTR + I + F + G4 nucleotide substitution model, and a Yule speciation process. Markov chain Monte Carlo (MCMC) analyses were performed for 20 million generations, and parameters were sampled every 1000th generation. Convergence of the MCMC chain to its stationary distribution was assessed in Tracer v1.5 (beast.community/tracer), and then a consensus tree was produced using TreeAnnotator after burn-in of the first 5000 trees.

Based on previous research on tribe Pyreae, Rosaceae (Xiang et al., 2016, Zhang et al., 2017), the genera *Amelanchier, Photinia* and *Cotoneaster* are closely related to *Eriobotrya* and *Rhaphiolepis*. Three constraint points were therefore set according to the fossil records of *Amelanchier*, *Photinia*, and *Cotoneaster* (Wolfe & Wehr, 1987, 1988; Yang & Yang, 2010), setting the minimum age of the stem node of the most recent common ancestor (MRCA) of each genus to 40.4, 40.4, and 15.5 Ma respectively. For the three constraint points, we used a log-normal distribution with a mean of 1.5, and a standard deviation of 1.

For the tree based on 197 concatenated nuclear genes, *Crataegus*, *Cotoneaster*, and *Malus* were selected as outgroup. The TN93+I+G4 nucleotide model was selected to perform the BEAST analysis, as mentioned above. As *Crataegus* formed a sister relationship with *Amelanchier*, the minimum root age of the tree and the minimum divergence time between *Cotoneaster* and other genera were set to be 40.4 and 15.5 Ma, respectively.

Inferring the ancient distribution of *Eriobotrya*

All the 33 *Eriobotrya* samples that were re-sequenced in this study, plus *E. laoshanica* (re-sequenced in our previous study, Chen et al. 2020), *R. indica* (Linnaeus) Lindley (common in subtropical southeastern China) and *R. philippinensis* (re-sequenced in this study), were ascribed to one of the three geographical areas defined above (TRO, SUB or INT, Table S1). The online software PastML (https://pastml.pasteur.fr/, Ishikawa et al., 2019) was then used to infer the ancestral distribution of every inner nodes within the rooted chloroplast ML phylogenetic tree.

RESULTS

Haplotype network based on population data

For the concatenated three chloroplast regions, reliable sequences were obtained for 640 individuals and the aligned length was 2,918 bp. After removing gaps, a total of 37 haplotypes were identified across all the sampled *Eriobotrya* and *Rhaphiolepis* species. According to the chloroplast haplotype network (Fig. 3a), these 37 haplotypes can be divided into five clades, each of which occupies a distinct geographical area (Fig. 2, Table S5). These clades are: 1) cSA, comprising ten haplotypes sampled from the geographical area TRO, plus two sampled from the geographical area INT; 2) cYN consisting of two haplotypes of *Rhaphiolepis* and nine *Eriobotrya* haplotypes sampled from INT (all in the YGP); 3) cSC, including one haplotype sampled from TRO, four from INT, and all three haplotypes sampled from the geographical area SUB; 4) cRH comprising five haplotypes of *Rhaphiolepis*; and 5) cHK, comprising only one haplotype sampled from INT.

For the five nuclear genes, reliable sequences were obtained for between 744 and 786 individuals, and the aligned length ranged from 379 to 835 bp. The number of polymorphic sites ranged from 65 to 188, which defined between 87 and 202 haplotypes. These nuclear genes harbored very high haplotype diversity, ranging from 0.829 to 0.971 (Table 1). According to the haplotype network of the gene *PXMP2* (Fig. 3b), three clades could be identified: 1) nRH including all the *Rhaphiolepis* species; 2) nSA containing five species sampled from TRO and two species (*E. seguinii* and *E. henryi*) from the YGP; and 3) nSC, including all other *Eriobotrya* species. In the haplotype network of the gene *TPP2*, the clade nSA was subdivided into two subclades, nSAI and nSAII; also a new clade nYN was formed comprising multiple haplotypes of seven species mostly from the YGP, however all these seven species also had haplotypes in the clade nSC (Fig. 3c). In the haplotype networks of *PEPC* and *C23H*, the clade nSA was again subdivided, but this time into four subclades (Fig. 3d and Fig. S1a). However, the last nuclear gene *GSDL2* did not provide clear resolution (Fig. S1b).

Structure analysis of the concatenated five nuclear genes

Analyses from Structure Harvester showed that the highest ΔK value was at $K = 4$ (Fig. 4a and Fig. 4b). The assignment results for $K = 4$ divided the 723 individuals of *Eriobotrya* species examined into 4 gene pools (G1-4), each of which demonstrated a small amount of admixture with other gene pools (Fig. 4c). The second highest ΔK value was at $K = 12$, at which gene pool G2 was unaffected but G1, G3 and G4 subdivided into two, five and four categories, respectively. Within G3, there was

admixture between the Yunnan taxa *E. obovata* W. W. Smith, *E. salwinensis* Handel-Mazzetti, *E. elliptica* Lindley and *E. bengalensis* Hook. f. var. *angustifolia* Cardot; within G4, *E. cavaleriei* was divided across two categories, one of them shared with *E. fulvicoma* W. Y. Chun ex W. B. Liao, F. F. Li & D. F. Cui. Among the 12 gene pools at K=12, only four gene pools comprised a single species, four showed a degree of admixture between species, and four comprised two or more species and were hence defined as species complexes (C1-C4, Fig. 4d). STRUCTURE analyses were then performed independently on each of these four complexes and these showed that most *Eriobotrya* species were distinct (Fig. 4e-h), with two exceptions: *E.* sp2 comprised a mixture of germplasm from *E. japonica* and *E. malipoensis* K. C. Kuan (Fig. 4g), whereas *E. deflexa* (from Hainan) shared same gene pool with *E. fragrans* (Fig. 4h).

Phylogenetic analysis based on chloroplast genomes

Complete chloroplast genome sequences were obtained for all the 33 *Eriobotrya* individuals and one *R. philippinensis* individual. The assembled chloroplast size varied from 159,853 bp (*R. philippinensis*) to 161,460 bp (*E. prinoides* Rehder & E. H. Wilson), with an average of 160,153 bp. In combination with 49 chloroplast genome sequences downloaded from the NCBI nucleotide database, the aligned 83 chloroplast genomes were 148,931 bp in length after removing gaps. With *Amelanchier* as the outgroup, the chloroplast phylogenetic trees based on ML and Bayes methods had identical topology, in which all the ingroup species fell into the same five clades as for the chloroplast haplotype network analysis, with high support values for each clade (Fig. 5a).

Phylogenetic analysis based on 197 nuclear genes

The species tree constructed from 197 single-copy nuclear genes revealed that *Rhaphiolepis* diverged first from all the *Eriobotrya* species, forming an independent clade nRH. All the *Eriobotrya* species could be divided into three clades (nSAI, nSAII, and nSC, Fig. 5b) with high support values (QS > 40 and PP = 1.00). Within the genus, 25 nodes had high support values (QS > 40 and PP ≥0.99), while 11 nodes in the subclade nSC had low support values (QS < 40 and PP < 0.9). The QS value for the species tree was estimated to be 0.70.

The ML and Bayesian trees constructed from the concatenated SNPs from the 197 nuclear genes were identical to one another, and also very similar to the species tree (Fig. 5b), except for of minor differences around three nodes (Fig. S2).

Species in conflict and tests of incongruence

Careful comparison of the haplotype networks (Fig. 3) and phylogenetic trees (Fig. 5) constructed from the chloroplast and nuclear datasets showed that phylogenetic positions for most *Eriobotrya* species were in strong conflict. To clarify their relationships, phylogenetic trees were constructed from only ten *Eriobotrya* species for which no strong conflict was detected, plus one *Cotoneaster* outgroup, based on chloroplast genomes and the concatenated nuclear genes (Fig. 5c). However, even within this reduced tree phylogenetic positions for three species were slightly different between the chloroplast and nuclear trees. However support values on the nodes involved in each of these conflicts were low in chloroplast and/or nuclear trees

(BS⩽90), so the incongruence was only weakly supported. However, the AU and weighted SH tests for the three species with conflicting positions (Table S6) showed that for two of them (*E. angustissima* Hook.f. and *E. deflexa*), the nuclear data significantly rejected the phylogenetic trees constructed from chloroplast data, whereas chloroplast data could not reject the phylogenetic trees constructed from nuclear data. For the third conflicted species (*E. laoshanica*), the reverse was true: i.e. nuclear data could not reject the chloroplast tree, while the chloroplast data rejected the nuclear tree significantly (Table S6). Hence, the phylogenetic relationships among the 11 species were defined as shown in Table S6.

R. philippinensis and the remaining 25 *Eriobotrya* individuals with strongly conflicted positions were each added individually to the simplified dataset and then AU and weighted SH tests (Table S7) were performed. For every one of these individuals, the nuclear data significantly rejected trees from chloroplast genome data, and vice versa.

In addition, a total of 20,000 chloroplast trees were simulated using the software Dendropy and mapped to the empirical chloroplast ML tree (Fig. S3). It was apparent that most clades were unique to the empirical chloroplast tree and absent in the simulated trees, supporting a scenario of hybridization.

Species networks construction

The PhyloNet network analyses predicted one hybridization event each in the TRO and SUB datasets, and four hybridization events within the INT dataset (Fig. 6) based on the log-pseudolikelihood values (Table S8). PhyloNet analyses were also performed with many other combinations of samples, and no additional hybridization events were predicted.

ABBA-BABA tests for gene flow among species and populations

D-statistics resolved a species complex not resolved by other methods, here termed the S-O-T-B complex, comprising *E. salwinensis*, *E. obovata, E. tengyuehensis* W. W. Smith, and *E. bengalensis* var. *angustifolia*. Furthermore, D-statistics showed multiple significant reticulation events between species, of which three were clearly resolved (Fig. 7): 1) between *E. cavaleriei* (including *E. fulvicoma*) and the S-O-T-B complex; 2) between *E. prinoides* (together with its two close relatives *E. grabrescens* J.E.Vidal, Adansonia sér. and *E. elliptica*) and the S-O-T-B complex; 3) between *E. serrata* J. E. Vidal and *E. malipoensis* (together with its two close relatives *E. japonica* and *E.* sp2 in Vietnam).

Divergence time estimation

Based on the chloroplast genome data, the stem age of *Eriobotrya* was dated to 41.07 Ma (95% HPD: 38.95–45.10 Ma, Fig. 8a). *Rhaphiolepis* was nested within *Eriobotrya* and together they formed a monophyletic group, whose TRMCA or crown age was dated to 31.80 Ma (95% HPD: 29.49–35.33 Ma), at which time clade cSA diverged from other *Eriobotrya* and the eight *Rhaphiolepis* species examined. Clade cSC diverged next, ~14.37 Ma (95% HPD: 12.82–16.11 Ma), followed by cHK (comprising only *E. fragrans* from Hong Kong) ~13.46 Ma (95% HPD:12.02–15.12 Ma). Finally, clade cRH (formed by the *Rhaphiolepis* species) diverged from cYN ~13.11 Ma (95% HPD: 11.74-14.77 Ma).

Based on the concatenated 197 nuclear genes, the stem age of all the *Eriobotrya* species plus *R. philippinensis* was dated to 40.52 Ma (95% HPD: 38.21–45.78 Ma, Fig. 8b), and *R. philippinensis* diverged from the monophyletic group containing all the *Eriobotrya* species at about 25.21 Ma (95% HPD: 23.65–28.48). Within *Eriobotrya*, clades nSAI and nSAII diverged from clade nSC successively at 21.97 Ma (95% HPD: 20.65–24.85 Ma) and 19.55 Ma (95% HPD: 18.33–22.08 Ma), respectively. Within clade nSC, subclades nSCI, nSCII, nSCIII, and nSCIV diverged from subclade nSCV at 14.01 Ma (95% HPD: 13.14–15.85 Ma), 12.66 Ma (95% HPD: 11.85–14.31 Ma), 10.97 Ma (95% HPD: 10.27–12.40 Ma), and 10.07 Ma (59% HPD: 9.44–11.40 Ma), respectively.

Ancestral distribution of the *Eriobotrya* species

PastML analyses assigned the ancestor of all the *Eriobotrya* species, and all the inner nodes of clade cSA, to the tropical area. Within clade cSC, the two inner nodes of the three species in subtropical China (*E. cavaleriei* in Jixangxi, *E. fragrans* in Guangdong, and *E. fulvicoma* in Guangdong) were assigned to the subtropical area, while all the other inner nodes were assigned to the intermediate area (Fig. S4).

DISCUSSION

Here, we provide multiple forms of evidence to support a central hypothesis that global climate changes during the Miocene could have created ample opportunities for hybridization between members of tropical and subtropical biomes at the boundary between those zones. We focused our research on the genus *Eriobotrya* and the Yunnan-Guizhou Plateau (YGP) in Southwest China, which is a transitional area between the tropical Southeast Asia and the subtropical East Asia floristic regions, and found that numerous hybridization events in *Eriobotrya* are likely to have taken place and contributed to the rich species diversity in the genus. A recent phylogenetic delineation of biogeographic regions for Chinese flora (Ye et al., 2019) suggests different affinities for the YGP depending on methodology: based on taxonomic regionalization, it is a part of the tropical Southeast Asia floristic region, yet according to phylogenetic regionalization it is part of the subtropical East Asia floristic region. This reflects the complex evolutionary history of vascular in this biodiversity hotspot (Ye et al., 2019), but rampant hybridization might be a contributor to this anomaly. Hence we advocate that hybridization during diversification, as detected here in *Eriobotrya*, was a key process during the evolution of regional flora at the transitional area between the tropical and subtropical floristic regions of East Asia.

Extensive cyto-nuclear conflicts in *Eriobotrya*

Recent studies using high throughput sequencing found that extensive interspecific gene flow commonly occurs in temperate and tropical trees (Larson et al., 2021; Linan et al., 2020; Schley et al., 2020; Wang et al., 2020), and our phylogenetic survey also revealed multiple instances of interspecific gene flow in *Eriobotrya*. The phylogenetic positions of *R. philippinensis* and 16 *Eriobotrya* species exhibited strong cyto-nuclear conflicts between a phylogenetic tree constructed from chloroplast genomes (Fig. 5a) and that based on 197 nuclear genes (Fig. 5b). Possible explanations for such incongruences include ILS, gene tree estimation errors, or interspecific gene flow

(hybridization) (Morales-Briones, Liston, & Tank, 2018). The mapping of the 20,000 simulated chloroplast trees with the empirical chloroplast tree (Fig. S3) indicated that hybridization rather than ILS caused most of these strong cyto-nuclear conflicts. Further, ILS would lead to random retention patterns for ancestral chloroplast genomes (Escobar et al., 2011; Ferrer Obiol et al., 2021). Under ILS, therefore, the 11 YGP species that exhibit cyto-nuclear conflict would be expected to be randomly distributed across the cYN, cSC, or cSA clades of the chloroplast phylogenetic tree, but in fact all of them retained chloroplast genomes from the common ancestor of the clade cYN.

Artificial hybridization experiments among 12 *Eriobotrya* species showed that nearly 70% of inter-specific combinations successfully set fruits (Li et al., 2016), suggesting that post-zygotic reproductive isolation in *Eriobotrya* species is incomplete. Furthermore, hybridization events commonly leave only one cytotype within the resulting lineage (Seehausen, 2004). Taken together, this indicates that the occurrence of multiple hybridization events is the most probable explanation for most of these cyto-nuclear conflicts within *Eriobotrya*. Specific hybridization events that would explain the observed discordance in each of the 17 species that exhibit it are summarized and listed in Table 2.

Regarding the consequences of these hybridization events, haplotype networks of four nuclear genes examined (Fig. 3b, 3d, Fig. S1) supported that nuclear haplotypes for 14 *Eriobotrya* species involved in hybridization could all come from their male parents, as their putative parents located in different clades and all these nuclear haplotypes were found in the clade of their male parents. For the last two species *E. fragrans* in Guangdong and *E. bengalensis* var. *angustifolia*, both parents located in the same clades, and nuclear haplotype network did not provide clear clues about whether they came from male or female parents. However, in the haplotype network of the last nuclear gene TPP2 (Fig. 3c), seven of the species involved in hybridization in the YGP each possessed two categories of haplotypes: in each case, one category of haplotypes clustered in the clade nSC representing the paternal genome, and the other category of haplotypes formed the clade nYN representing the maternal genome. These data provided clear clues that hybridization has occurred during the evolutionary history of these species, however it appears that only a small part of the maternal genomes have been retained in the genomes of putative hybrid lineages. In such cases, the detection of hybridization events based on nuclear data alone may be difficult. In this study, although Bayesian clustering analysis (Fig. 4), species network analysis (Fig. 6), and gene flow analysis (Fig. 7) were performed on nuclear data, none of them successfully identified the above predicted hybridization events. This phenomenon, termed chloroplast capture, probably results from hybridization between two species, followed by repeated backcrossing with the pollen donor, yielding a hybrid entity with the nuclear genome nearly the same as male parent and the chloroplast genome of the female parent (Seehausen, 2004, Soltis, 2013; Acosta & Premoli, 2010; Hojjati et al., 2019).

Of particular interest is that certain clades present in the chloroplast tree (cYN, cRH, and cHK; Fig. 5a) were absent from the nuclear tree (Fig. 5b), with the species concerned distributed across other clades. This pattern could also be seen in haplotype

559 network analyses with population data (Fig. 3 and Fig. S1). This means that none of the
560 23 sampled *Eriobotrya* species could have served as chloroplast donors to species
561 within the missing clades (cYN, cRH, or cHK), and because our study sampled all 16
562 *Eriobotrya* species recorded from China, including all 12 from the YGP, it is very
563 unlikely that any extant local species could be the donor in these cases. Instead, it
564 appears that the donor species are either extinct or might in some cases have migrated
565 southwards into tropical Asia.

566 **Hybridization hotspot in the YGP, a tropical-subtropical transition zone**

567 Although hybridizations and hybrid zones between plant species pairs have
568 occasionally been reported in transitional zones along various environmental gradients
569 (e.g. latitudinal or altitudinal gradients, and vegetation transition zones; Abbott, 2017),
570 there have been very few reports of extensive hybridization events within a single genus
571 within a particular biome transition zone. Nevertheless, in our study, 11 of the 12
572 *Eriobotrya* species collected from the YGP had been involved in hybridization events,
573 suggesting a pattern of multiple hybridization events facilitating species diversification.
574 These 11 species formed a monophyletic clade (cYN) in the chloroplast tree (Fig. 5a),
575 whereas in the nuclear tree, two of them clustered with the widespread species *E.*
576 *bengalensis* in the clade nSA, while the remaining nine were scattered across clade nSC
577 (Fig. 5b). A plausible explanation is that at least four species of the clade nSC in
578 subtropical China, and also *E. bengalensis* from tropical Asia, had invaded the YGP,
579 hybridized with and ultimately replaced local species. These processes could have
580 occurred simultaneously through pollen swamping and repeated backcrossing to the
581 invader, leading to eventual chloroplast capture.

582 Chloroplast capture events observed in tropical Asia could also have resulted from
583 the southward dispersal of *Eriobotrya* species currently in the YGP, such as *E. serrata*
584 in Vietnam and *E. bengalensis* var. *angustifolia* in Myanmar (Table 2). *E. serrata* might
585 have further spread to tropical areas such as Vietnam and Myanmar, possibly invading
586 the range of a congener by hybridizing with standing stock of another species, then
587 gradually replacing it by repeated crossing, especially if changing conditions caused
588 selection to favour the invader. This could lead to chloroplast capture by the invader,
589 for example, *Betula pubescens* captured the plastid of *B. nana* (Eidesen, Alsos &
590 Brochmann 2015). This might also have occurred in *E. elliptica* sampled from Tibet,
591 China and *E. glabrescens* from Myanmar (Table 2). It is also possible that *E. prinoides*
592 in the YGP captured cpDNA of local species during further spread to tropical areas of
593 Tibet then Myanmar in this way.

594 Furthermore, both PhyloNet and STRUCTURE analysis revealed reticulate
595 evolution among the S-O-T-B complex, which comprises four taxa endemic to the YGP:
596 *E. salwinensis*, *E. obovata*, *E. tengyuehensis* and *E. bengalensis* var. *angustifolia*;
597 however these analyses showed different relationships among them. Furthermore, D-
598 statistics revealed significant gene flow between the S-O-T-B complex and both *E.*
599 *cavaleriei* and *E. prinoides* (Fig. 7). The quartet score and local posterior probability
600 for the phylogenetic position of the S-O-T-B complex were very low (QS=34 and
601 PP=0.39), also indicating strong conflict among gene trees (Fig. 5b). Therefore,
602 hybridization might have been involved in the origin of the S-O-T-B complex,

diversification within it, and/or its subsequent interaction with species outside the complex.

In addition, PhyloNet analysis supported gene flow from the ancestor of *E. cavaleriei* to the ancestor of *E. malipoensis* and from the ancestor of *E. malipoensis* to *E. serrata* (Fig. 6c), with the latter further supported by D-statistics (Fig. 7) and the low values of QS and PP for the phylogenetic position of the subclade nSCII (QS=39 and PP=0.8, Fig. 5b). In the STRUCTURE analysis across 727 individuals of 23 *Eriobotrya* species, instances of genetic admixture occurred sporadically across many *Eriobotrya* species, suggesting interspecific gene flow involving numerous *Eriobotrya* species (Fig. 4c-d). Within the clade nSC in the nuclear species tree (Fig. 5b), QS and PP values for 11 nodes were very low (QS<40 and PP<0.90), showing conflicts in gene trees possibly resulting from ILS, tree estimation errors or hybridization. Hence, our study revealed an extraordinarily complicated history of reticulate evolution among *Eriobotrya* species, especially in the YGP. Nevertheless, the STRUCTURE analysis based on unlinked SNPs in five nuclear genes showed that most species can be assigned to independent gene pools (Fig. 4d-h) indicating that these species are independent evolutionary units that may have experienced, and in many cases been altered or even formed by, ancient hybridization events.

Paleoclimate changes drove diversification and hybridization of *Eriobotrya*

The reconstructed biogeographic history suggests that *Eriobotrya* most likely expanded northward to southern China from tropical areas, and diverged into two clades, one from tropical Asia and the other southern China. Ancestral area reconstruction indicated that *Eriobotrya* probably originated from tropical areas (Fig. S4), and that it diverged (along with its sister genus *Rhaphiolepis*) from other related genera at 41.07 Ma based on chloroplast genome analysis (Fig. 8a), or 40.52 Ma based on nuclear data (Fig. 8b), slightly earlier than previous estimates based on chloroplast genomes (~36 Ma (Zhang et al., 2017) or transcriptome analyses (~34 Ma; Xiang et al., 2016). This period corresponded to the warm Eocene, when most genera of Maleae first arose (Zhang et al. 2017). Around 9 Ma later, ~31.80 Ma, *Eriobotrya* diversified into a tropical Asian clade (cSA) and a predominantly southern Chinese clade (comprising the four subclades cHK, cRH, cSC, and cYN) (Fig. 8a), indicating that *Eriobotrya* had spread to southern China around the early Oligocene.

The establishment of the East Asian Monsoon (EAM) (22–25 Ma) greatly changed the climate in subtropical China, creating many suitable habitats for the development of subtropical evergreen forests (Tao, 2000). The global temperature reached a climatic optimum at 14–16 Ma, coincident with the intensification of the EAM at ~14 Ma (Sun & Wang, 2005, Zachos et al., 2001). This intensification of the EAM might have supported a high speciation rate in southern China (Kong et al., 2017), such as the early diversification and subsequent radiation of the fern genus *Lepisorus* in the Miocene and Pliocene (Wang et al., 2012). In this study, crown diversification within the subtropical Chinese clade (cHK+cRH+cSC+cYN) was dated to 14.37 Ma (Fig. 8a), while the TMRCA of the nuclear clade nSC was also dated to a similar age (14.01 Ma, Fig. 8b). Hence the intensification of EAM and the climatic optimum during the Miocene might have provided a suitable environment for the spread of these *Eriobotrya* species in

subtropical China, which further diversified into four clades (cYN, cRH, cSC, and cHK in Fig. 5 & 8).

Among these, Clade cYN comprises 11 *Eriobotrya* species, all of which were involved in ancient hybridization events. Based on the TRMCA of this clade (8.35 Ma) and the ages of its internal nodes (mostly 6-8 Ma), it appears that most of those hybridization events occurred in the YGP around 6-9 Ma and later, during or after the intensification of the EAM around 7-9 Ma (Lu & Guo, 2014). Kadereit (2015) proposed that climate-induced changes in distribution ranges could have greatly promoted hybrid speciation. Climate-driven range expansion can cause lineages to meet and hybridise, whereas contraction of parental lineages could result in geographical isolation and the stabilization of hybrid lineages newly isolated from their parents, such as the speciation of *Ostryopsis intermedia* (Liu et al., 2014, Wang et al., 2021) and *Pinus densata* (Wang et al., 2011).

The co-occurrence of the intensification of the EAM and continuous global cooling during the late Miocene could have greatly changed local habitats within the YGP and adjacent regions, possibly favouring certain hybridised lineages that combined beneficial germplasm from original local stock with that from invaders. This cooling might also have caused southward migration of many lineages into tropical areas such as Vietnam and Myanmar, leading to further instances of hybridization, only this time with cooler-adapted lineages invading the range of tropical ones. Notably, the chloroplast phylogeny contains three clades (cYN, cRH, and cHK) that are not in the nuclear tree, suggesting that the members of each of these clades donated chloroplasts to invading lineages before going extinct, perhaps due to climate change and/or competition from invading and hybrid lineages.

Overall, the history of *Eriobotyra* appears to be one in which range shifts and invasions seem to have occurred at the genomic level, rather than the species level, in a lot of its lineages.

CONCLUSION

Our work demonstrated multiple and frequent hybridization events in *Eriobotrya*, mostly centered in the YGP, which bordered the tropical and subtropical zones. Based on our data, we postulate that climatic warming in the mid-Miocene (ca. 14 Ma) allowed tropical and subtropical *Eriobotrya* species to invade the YGP, possibly leading to multiple hybridization events with local species as part of the invasion process. Later, a comparable process occurred in the opposite direction as global cooling allowed the resultant lineages to invade more tropical areas to the south and interact with lineages there. At the same time, the intensification of the EAM and the continuous global cooling during the late Miocene (approximately 8 Ma) could have favored hybrid lineages over local species in the YGP, promoting their establishment and extinction, respectively.

In addition to these climate change-driven processes, further hybridization events occurred among lineages within the YGP leading to the formation of the S-O-T-B complex. Therefore the YGP, serving as the distribution and biodiversity center of *Eriobotrya*, has greatly contributed to the speciation and diversification of *Eriobotrya*. All in all, our case study using *Eriobotrya* as a model system highlights the potential

contribution of paleoclimate change to driving hybridization, and consequent range invasions at a genomic level, promoting rapid species diversification in climate transition zones.

DATA ACCESSIBILITY

Chloroplast and nuclear haplotypes were uploaded to GenBank under accessions: MT740924-MT741083, MT746200-MT747025, and MW788222-MW788322, assembled chloroplast genomes for *Eriobotrya* and *Rhaphiolepis* species were uploaded to GenBank under accessions: MT872352, MT872353, MT876388-MT876406 and MT890250-MT890262.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (31670189 and 31800175), and 2021 Guangdong Special Fund for Natural Resources Management and Ecological Forestry Construction (2021GJGY001). We would like to thank Jialiang Li for his assistance in Dendropy simulations and editor and the two anonymous reviewers for providing comments that improved the manuscript.

AUTHOR CONTRIBUTIONS

SC, RM, RZ, WL, QF, YM and KM designed the project; WG, YK, WL, QF collected the samples; QY and SC ran the experiment; SC, RM, RZ, KM, QY analyzed the results; SC, RM, RZ, WL, QF, YM and KM wrote the manuscript. SC, MR, RZ contributed equally to this work.

REFERENCES

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. Journal of Systematics and Evolution 55(4): 238-258. https://doi.org/ 10.1111/jse.12267

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229-246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Acosta, M. C., & Premoli, A. C. (2010). Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution*, 54(1), 235-242. https://doi.org/10.1016/j.ympev.2009.08.008

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37-48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barthlott, W. B., Lauer, W. P. A., & Placke, A. J. E. (1996). Global distribution of species diversity in vascular plants: towards a world map of phytodiversity. *Erdkunde*, 50(4), 317-327. https://doi.org/10.3112/erdkunde.1996.04.03

Chen, S. F., Meng, K. K., Guo, X. B., Zhao, W. Y., Liao, W. B., & Fan, Q. (2020). A new species of *Eriobotrya* (Rosaceae) from Yunnan Province, China. Phytokeys, 146, 61-69. https://doi.org/ 10.3897/phytokeys.146.50728

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884-i890. https://doi.org/10.1093/bioinformatics/bty560

735 Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of
736 organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4),
737 e18. https://doi.org/10.1093/nar/gkw955
738 Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian
739 Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and*
740 *Evolution*, 29(8), 1969-1973. https://doi.org/10.1093/molbev/mss075
741 Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., … & de
742 Pamphilis, C. W. (2010). Identification of shared single copy nuclear genes in
743 *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across
744 various taxonomic levels. *BMC Evolutionary Biology*, 10, 61.
745 https://doi.org/10.1186/1471-2148-10-61
746 Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and
747 program for visualizing STRUCTURE output and implementing the Evanno
748 method. *Conservation Genetics Resources*, 4(2), 359-361.
749 https://doi.org/10.1007/s12686-011-9548-7
750 Eidesen, P. B., Alsos, I. G., Brochmann, C. (2015). Comparative analyses of plastid and
751 AFLP data suggest different colonization history and asymmetric hybridization
752 between *Betula pubescens* and *B.nana*. *Molecular Ecology* 24(15): 3993-4009.
753 Ellstrand, N. C., Whitkus, R. & Rieseberg, L. H. (1996). Distribution of spontaneous
754 plant hybrids. *Proceedings of the National Academy of Sciences of the United*
755 *States of America*, 93: 5090-5093. https://doi.org/10.1073/pnas.93.10.5090
756 Escobar, J. S., Scornavacca, C., Cenci, A., Guilhaumon, C., Santoni, S., Douzery, E. J.
757 P., Ranwez, V., Glemin, S., David, J. (2011). Multigenic phylogeny and analysis
758 of tree incongruences in Triticeae (Poaceae). BMC Evolutionary Biology 11:
759 181. https://doi.org/10.1186/1471-2148-11-181
760 Fan, Q., Chen, S. F., Li, M. W., Guo, W., Jing, H. J., Wu, W., … & Liao, W. B. (2014).
761 Molecular evidence for natural hybridization between wild loquat (*Eriobotrya*
762 *japonica*) and its relative *E. prinoides*. *BMC Plant Biology*, 14, 275.
763 https://doi.org/10.1186/s12870-014-0275-6
764 Fay, M., Swensen, S., & Chase, M. (1997). Taxonomic affinities of *Medusagyne*
765 *oppositifolia* (Medusagynaceae). *Kew Bulletin*, 52(1), 111-120.
766 https://doi.org/10.2307/4117844
767 Ferrer Obiol J., James H. F., Chesser R. T., Bretagnolle V., González-Solís J., Rozas J.,
768 Riutort M., Welch A. J., (2021). Integrating sequence capture and restriction
769 site-associated DNA sequencing to resolve recent radiations of pelagic seabirds.
770 Systematic Biology 70(5): 976-996. https://doi.org/10.1093/sysbio/syaa101
771 Folk, R. A., Mandel, J. R., & Freudenstein, J. V. (2017). Ancestral gene flow and
772 parallel organellar genome capture result in extreme phylogenomic discord in a
773 lineage of angiosperms. *Systematic Biology*, 66(3), 320–337.
774 https://doi.org/10.1093/sysbio/syw083
775 Fuchs, J., Ericson, P. G. P., Bonillo, C., Couloux, A., & Pasquet, E. (2015). The complex
776 phylogeography of the Indo-Malayan *Alophoixus bulbuls* with the description
777 of a putative new ring species complex. *Molecular Ecology*, 24(21), 5460-5474.
778 https://doi.org/10.1016/j.scienta.2008.03.012

779 Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in plants: old Ideas, new
780 techniques. *Plant Physiology*, 173(1), 65-78.
781 https://doi.org/10.1104/pp.16.01340
782 Gu, C. Z., & Spongberg, S. A. (2003). *Eriobotrya* Lindley. In: Wu CY, Raven PH, Hong
783 DY eds. *Flora of China* 9. Science Press, Beijing & Missouri Botanical Garden
784 Press, St. Louis, 140-143.
785 Guo, Z. T., Sun, B., Zhang, Z. S., Peng, S. Z., Xiao, G. Q., Ge, J. Y., … & Wei, J. J.
786 (2008). A major reorganization of Asian climate by the early Miocene. *Climate*
787 *of the Past*, 4, 153-174. https://doi.org/10.5194/cp-4-153-2008
788 Heinze, B. (2007). A database of PCR primers for the chloroplast genomes of higher
789 plants. *Plant Methods*, 3, 4. https://doi.org/10.1186/1746-4811-3-4
790 Hojjati, F., Adams, R. P., & Terry, R. G. (2019). Discovery of chloroplast capture in
791 *Juniperus excelsa* complex by multi-locus phylogeny. *Phytotaxa*, 413(1), 11-26.
792 https://doi.org/10.11646/phytotaxa.413.1.2
793 Huang, J. C., Ge, X. J., & Sun, M. (2000). Modified CTAB protocol using a silica
794 matrix for isolation of plant genomic DNA. *Biotechniques*, 28(3), 432-434.
795 https://doi.org/10.1006/biol.1999.0237
796 Ishikawa, S. A., Zhukova, A., Iwasaki, W., & Gascuel, O. (2019). A fast likelihood
797 method to reconstruct and visualize ancestral scenarios. *Molecular Biology and*
798 *Evolution*, 36(9), 2069-2085. https://doi.org/10.1093/molbev/msz131
799 Jiang, S., An, H. S., Xu, F. J., & Zhang, X. Y. (2020). Chromosome-level genome
800 assembly and annotation of the loquat (*Eriobotrya japonica*) genome.
801 *Gigascience*, 9(3), 1-9. https://doi.org/10.1093/gigascience/giaa015
802 Jin, J. H., Liao, W. B., Wang, B. X., & Peng, S. L. (2003). Global change in Cenozoic
803 and evolution of flora in China. *Guihaia*, 23(3), 217-225.
804 Kadereit, J. W. (2015). The geography of hybrid speciation in plants. *Taxon*, 64(4), 673-
805 687. https://doi.org/10.12705/644.1
806 Kang, D. H., Ong, H. G., Lee, J. H., Jung, E. K., Kyaw, N. O., Fan, Q., & Kim, Y. D.
807 (2021). A new broad-leaved species of loquat from eastern Myanmar and its
808 phylogenetic affinity in the genus *Eriobotrya* (Rosaceae). *Phytotaxa*, 482(3),
809 279-290. https://doi.org/10.11646/phytotaxa.482.3.6
810 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software
811 version 7: improvements in performance and usability. *Molecular Biology and*
812 *Evolution*, 30(4), 772-780. https://doi.org/10.1093/molbev/mst010
813 Kong, H. H., Condamine, F. L., Harris, A. J., Chen, J. L., Pan, B., Mller, M., … & Kang,
814 M. (2017). Both temperature fluctuations and East Asian monsoons have driven
815 plant diversification in the karst ecosystems from southern China. *Molecular*
816 *Ecology*, 26(22), 6414-6429. https://doi.org/10.1111/mec.14367
817 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular
818 evolutionary genetics analysis across computing platforms. *Molecular Biology*
819 *and Evolution*, 35(6), 1547-1549. https://doi.org/10.1093/molbev/msy096
820 Institute of Geography, Chinese Academy of Sciences, 1959. Integrated
821 Physiogeographical Regionalization of China (Preliminary draft). Science Press,
822 Beijing.

Larson, D. A., Vargas, O. M., Vicentini, A., Dick, C. W. (2021). Admixture may be extensive among hyperdominant Amazon rainforest tree species. New Phytologist. https://doi.org/10.1111/nph.17675.

Li, G. F., Liang, G. L., & Lin, S. Q. (2013). A study on karyotype analysis of genus *Eriobotrya* and its application to systematic taxonomy. *Acta Horticulturae Sinica*, 40(8), 1465-1474.

Li, G. F., Yang, X. H., Qiao, Y. C., Gao, Y. S., Jiang, Y. Y., & Lin, S. Q. (2016). Study on interspecific and intergeneric hybridization compatibility of *Eriobotrya* and related genera. *Acta Horticulturae Sinica*, 43(6), 1069-1078. https://doi.org/10.16420/j.issn.0513-353x.2016-0215

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXivorg*. https://arxiv.org/abs/1303.3997

Li, J. L., Zhang, Y. J., Ruhsam, M., Milne, R. I., Wang, Y., Wu, D., … & Mao, K. S. (2021). Seeing through the hedeg: phylogenomics of Thuja (Cupressaceae) reveals prominent incomplete lineage sorting and ancient introgression for Tertiary relict flora. *Cladistics*, 0: 1-17. https://doi.org/10.1111/cla.12491.

Li, M. W., Chen, S. F., Zhou, R. C., Fan, Q., Li, F. F., & Liao, W. B. (2017). Molecular evidence for natural hybridization between *Cotoneaster dielsianus* and *C. glaucophyllus*. *Frontiers in Plant Science*, 8: 704. https://doi.org/10.3389/fpls.2017.00704

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452. https://doi.org/10.1093/bioinformatics/btp187

Linan, A. G., Lowry, P. P., Miller, A., Schatz, G. E., Sevathian, J. C., & Edwards, C. E. (2021). Rad-seq reveals patterns of diversification, hybridization, and the accumulation of reproductive isolation in a clade of partially sympatric, tropical island trees. Molecular Ecology. https://doi.org/10.1111/MEC.15736.

Liu, B. B., Abbott, R. J., Lu, Z. Q., Tian, B., & Liu, J. Q. (2014). Diploid hybrid origin of *Ostryopsis intermedia* (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change. *Molecular Ecology*, 23(12), 3013-3027. https://doi.org/10.1111/mec.12783

Liu, B. B., Liu, G. N., Hong, D. Y., & Wen, J. (2020). *Eriobotrya* belongs to *Rhaphiolepis* (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data. *Frontiers in Plant Science*, 10, 1731. https://doi.org/10.3389/fpls.2019.01731

Lu, H. Y., & Guo, Z. T. (2014). Evolution of the monsoon and dry climate in East Asia during late Cenozoic: A review. *Science China (Earth Sciences)*, 57(1), 70-79. https://doi.org/10.1007/s11430-013-4790-3

Maddison, W. B. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523-536. https://doi.org/10.2307/2413694

867 Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and
868 related admixture evidence from VCF files. *Molecular Ecology Resources*, 21,
869 584-595. https://doi.org/10.1111/1755-0998.13265
870 Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279-283.
871 https://doi.org/10.1038/nature05706
872 Meng, K. K., Chen, S. F., Xu, K. W., Zhou, R. C., Li, M. W., Dhamala, M. K., … &
873 Fan, Q. 2021. Phylogenomic analyses based on genome-skimming data reveal
874 cyto-nuclear discordance in the evolutionary history of *Cotoneaster* (Rosaceae).
875 *Molecular Phylogenetics and Evolution*, 107083.
876 https://doi.org/10.1016/j.ympev.2021.107083
877 Morales-Briones, D. F., Liston, A., & Tank, D. C. (2018). Phylogenomic analyses reveal
878 a deep history of hybridization and polyploidy in the Neotropical genus
879 *Lachemilla* (Rosaceae). *New Phytologist*, 218(4), 1668-1684.
880 https://doi.org/10.1111/nph.15099
881 Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a
882 fast and effective stochastic algorithm for estimating maximum-likelihood
883 phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.
884 https://doi.org/10.1093/molbev/msu300
885 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure
886 using multilocus genotype data. *Genetics*, 155(2), 945-959. PMCID:
887 PMC1461096
888 Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and*
889 *Systematics* 28(1): 359-389.
890 Ronquist, F., Teslenko, M., van der Mark. P, Ayres, D. L., Darling, A., Hohna, S., … &
891 Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic
892 inference and model choice across a large model space. *Systematic Biology*,
893 61(3), 539-542. https://doi.org/10.1093/sysbio/sys029
894 Schley, R. J., Pennington, R. T., Pérez-Escobar, O. A., Helmstetter, A. J., Estrella, M.
895 de la, Larridon, I., … Forest, F. (2020). Introgression across evolutionary scales
896 suggests reticulation contributes to Amazonian tree diversity. Molecular
897 Ecology, 2020, 29: 4170-4185. https://doi.org/10.1111/mec.15616.
898 Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology and*
899 *Evolution*, 19(4), 198-207. https://doi.org/10.1016/j.tree.2004.01.003
900 Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015). Analysis of phylogenomic
901 datasets reveals conflict, concordance, and gene duplications with examples
902 from animals and plants. *BMC Evolutionary Biology*, 15, 150.
903 https://doi.org/10.1186/s12862-015-0423-0
904 Soltis, P. S. (2013). Hybridization in Plants. *Encyclopedia of Biodiversity*, 167(2), 166-
905 176. https://doi.org/10.1016/B978-0-12-384719-5.00202-1
906 Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual*
907 *Review of Plant Biology*, 60, 561-588.
908 https://doi.org/10.1146/annurev.arplant.043008.092039
909 Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python library for phylogenetic
910 computing. *Bioinformatics*, 26(12), 1569-1571.

https://doi.org/10.1093/bioinformatics/btq228
Sun, X. J., & Wang, P. X. (2005). How old is the Asian monsoon system? Palaeobotanical records from China. *Palaeogeography Palaeoclimatology Palaeoecology*, 222, 181-222. https://doi.org/10.1016/j.palaeo.2005.03.005
Tao, J. R. (2000). The evolution of the Late Cretaceous –Cenozoic floras in China. Science Press, Beijing.
Wang, B. S., Mao, J. F., Gao, J., Zhao, W., & Wang, X. R. (2011). Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Molecular Ecology*, 20(18), 3796-3811. https://doi.org/10.1111/j.1365-294X.2011.05157.x
Wang, M. C., Zhang, L., Zhang, Z. Y., Li, M. M., Wang, D. Y., Zhang, X., … & Ma, T. (2020), Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. New Phytologist, 225: 1370-1382. https://doi.org/10.1111/nph.16215
Wang, L., Schneider, H., Zhang, X. C., & Xiang, Q. P. (2012). The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biology*, 12, 210. https://doi.org/10.1186/1471-2229-12-210
Wang, Z. F., Jiang, Y. Z., Bi, H, Lu, Z. Q., Ma, Y. Z., Yang, X. Y.,… & Liu, J. Q. (2021). Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Molecular Plant*, 14(2), 208-222. https://doi.org/10.1016/j.molp.2020.11.008
Wen, D. Q., Yu, Y., Zhu, J. F., & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Systematic Biology*, 67(4), 735-740. https://doi.org/10.1093/sysbio/syy015
Whitney, K. D., Ahern, J. R., Campbell, L. G., Albert, L. P., & King, M. S. (2010). Patterns of hybridization in plants. *Perspectives in Plant Ecology Evolution and Systematics*, 12(3), 175-182. https://doi.org/10.1016/j.ppees.2010.02.002
Wolfe, J. A., & Wehr, W. (1987). Middle Eocene dicotyledonous plants from Republic, Northeastrn Washington. U.S. *Geological Surver Bulletin*, 1597, 1-25
Wolfe, J. A., & Wehr, W. (1988). Rosaceous Chamaebatiaria-Like Foliage from the Paleogene of Western North America. *Aliso*, 12(1), 177-200. https://doi.org/10.5642/aliso.19881201.14
Xiang, Y. Z., Huang, C. H., Hu, Y., Wen, J., Li, S. S., Yi, T. S., … & Ma, H. (2016). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution*, 34, 262-281. https://doi.org/10.1093/molbev/msw242
Yang, H., & Yang, S. (2010). The Shanwang fossil biota in eastern China: a Miocene Konservat-Lagersttte in lacustrine deposits. *Lethaia*, 27(4), 345-354. https://doi.org/10.1111/j.1502-3931.1994.tb01585.x
Yang, X. H., Glakpe, K., Lin, S. Q., Hu, Y. L., He, Y. H., Nguyen, T., … & Liu, C. M. (2005). Taxa of plants of genus *Eriobotrya* around the world and native to South-eastern Asia. *Journal of Fruit Science*, 22(1), 55-59.
Ye, J. F., Lu, L. M., L, B., Yang, T., Zhang, J. L., Hu, H. H., Li, R., … & Chen, Z., D. (2019). Phylogenetic delineation of regional biota: a case study of the Chinese flora. Molecular Phylogenetics and Evolution, 135: 222-229

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, 292(5517), 686-693. https://doi.org/10.1126/science.1059412

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). Astral-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(S6), 153. https://doi.org/10.1186/s12859-018-2129-y

Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., … & Yi, T. S. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytologist,* 214(3), 1355-1367. https://doi.org/10.1111/nph.14461

Zhu, H. (2013). Geographical elements of seed plants suggest the boundary of the tropical zone in China. *Palaeogeography Palaeoclimatology Palaeoecology*, 386, 16-22. https://doi.org/10.1016/j.palaeo.2013.04.007

Tables

Table 1 Genetic diversity of the five nuclear genes

Loci	Number of individuals (*N*)	Aligned length	polymorphic sites (*S*)	Number of Haplotypes (*H*)	Haplotype diversity H_d	Nucleotide diversity (π)
PMP2	744	835	161	175	0.971	0.011
C23H	787	684	188	202	0.946	0.010
PEPC	765	676	116	139	0.934	0.007
GDSL2	780	415	65	87	0.857	0.010
TPP2	786	379	100	101	0.829	0.016

973 Tables 2 Putative hybrid species and their parent species predicted from cyto-nuclear conflicts in
974 phylogenetic trees

Putative hybrid species	Putative Female parent	Putative Male parent	Locations
Rhaphiolepis species	Ancestor of the RH clade	*Rhaphiolepis sp.*	?
E. henryi	Ancestor of the YN clade	*E. bengalensis*	China: Yunnan
E. seguinii	Ancestor of the YN clade	*E. bengalensis*	China: Yunnan
E. prinoides	Ancestor of the YN clade	Ancestor of *E. prinoides*	China: Yunnan
E. salwinensis	Ancestor of the YN clade	Ancestor of *E. prinoides*	China: Yunnan
E. bengalensis var. *angustifolia* in Yunnan	Ancestor of the YN clade	Ancestor of *E. prinoides*	China: Yunnan
E. obovata	Ancestor of the YN clade	Ancestor of *E. prinoides*	China: Yunnan
E. tengyuehensis	Ancestor of the YN clade	Ancestor of *E. prinoides*	China: Yunnan
E. malipoensis	Ancestor of the YN clade	Ancestor of *E. japonica*	China: Yunnan
	Ancestor of *E. cavaleriei*	Ancestor of *E. japonica*	China: Yunnan
	Ancestor of *E. laoshanica*	Ancestor of *E. japonica*	China: Yunnan
E. serrata in Yunnan	Ancestor of the YN clade	Ancestor of *E. japonica*	China: Yunnan
E. sp1	Ancestor of the YN clade	Ancestor of *E. deflexa*	China: Yunnan
E. cavaleriei in Yunnan	Ancestor of the YN clade	*E. cavaleriei*	China: Yunnan
E. fragrans in Guangdong	*E. cavaleriei*	*E. deflexa*	China: Guangdong
E. fragrans in Hongkong	The HK clade	*E. deflexa*	China: Hongkong
E. deflexa in Taiwan	Ancestor of *E. stipularis*	*E. deflexa*	China: Tawan
E. elliptica	*E. bengalensis*	Ancestor of *E. prinoides*	China: Tibet
E. sp2	Ancestor of *E. stipularis*	Ancestor of *E. japonica*	North Vietnam
E. glabrescens	*E. bengalensis*	Ancestor of *E. prinoides*	Myanmar
E. bengalensis var. *angustifolia* in Myanmar	*E. bengalensis*	Ancestor of *E. japonica*	Myanmar
E. serrata in Vietnam	*E. bengalensis*	Ancestor of *E. japonica*	South Vietnam
E. bengalensis var. *angustifolia* in Vietnam	*E. angustissima*	*E. stipularis*	South Vietnam

975

Supplementary tables

Table S1 The geographic information of the sampled populations of *Eriobotrya*.

Table S2 The geographic information of closely related species of *Eriobotrya* sampled in this study

Table S3 Primer information of the chloroplast regions and the nuclear genes

Table S4 GenBank accession numbers for chloroplast genomes of species used in this study

Table S5 Geographical distribution of the 37 chloro-types sampled from the 47 populations of 23 *Eriobotrya* species

Table S6 The AU and weighted SH tests for species in weak conflicts

Table S7 AU and weighted SH tests for species demonstrating strong cyto-nuclear conflicts in their phylogenetic positions

Table S8 The -log pseudolikelihood values in PhyloNet analyses

Figures

Fig. 1 The studied species *Eriobotrya*. a. *E. seguinii*; b. *E. henryi*; c. *E. salwinensis*; d. *E. prinoides*; e. *E. obovata*; f. *E. japonica*; g. *E. malipoensis*; h. *E. cavaleriei*; i. *E. fragrans*; j. *E. deflexa*. Photo of *E. japonica* was taken by Shaoping Chen, *E. deflexa* was taken by Mutan Hsieh, and others by Q. Fan. The border line between tropical and subtropical zones was drawn according to Institute of Geography & Chinese Academy of Sciences (1959).

Fig. 2 Map depicting the distribution of 47 sampled populations of *Eriobotrya*, four clades of chloroplast haplotypes (cSA, cYN, cSC and cHK), and three clades of nuclear haplotypes (nSAI, nSAII and nSC).

Fig. 3 Haplotype networks of the combined four chloroplast regions (a) and three nuclear genes *PXMP2* (b), *TPP2* (c) and *PEPC* (d) constructed for 817 *Eriobotrya* individuals. Clades divided in chloroplast haplotype: cSA, cYN, cSC and cHK. Clades divided in nuclear haplotypes: nSA and nSC in the gene *PXMP2*; nSAI, nSAII, and nSC in the gene *TPP2*; and nNV, nDU, nBE, and nSH in the gene *PEPC*.

Fig. 4 The best *K* values based on mean L(*K*) (a) and Δ*K* (b), assignment results for *K*=4 (c) and *K*=12 (d) across 723 individual based on unlinked SNPs from five nuclear genes, assignment results for optimal *K* value across three species in the clade nSAI (e), three species in the clade nSAII (f) , five species in the subclades nSC1 and nSCII (g), three species in subtropical China including *E. deflexa*, *E. fragrans*, and *E. cavaleriei* (h). The two clades nSAI and nSAII and the five subclades (nSC1, nSCII, nSCIII, nSCIV, and nSCV) was divided according to nuclear species tree. C1-C4: species complexes in which >1 species were assigned to the same gene pool.

Fig. 5 The ML phylogenetic tree constructed from chloroplast genomesof 83 *Eriobotrya* and other related species (a), the species tree constructed from 197 single-copy nuclear genes of 43 *Eriobotrya* and other related species (b), and the ML phylogenetic trees constructed from chloroplast genomes (left) and 197 nuclear genes (right) of the 11 *Eriobotrya* species after removing those in strong cyto-nuclear conflicts (c). Above the nodes of chloroplast tree (a), Maximum likelihood ultrafast bootstrap support values (BS>50) and Bayesian posterior probabilities (PP>0.50) are shown: "*": BS≥ 95% or PP≥0.99. "-": BS<50 or PP<0.50, nodes in red were different between ML and Bayes trees. Above the nodes of the nuclear species tree (b), quartet score (QS) are shown on right and local posterior probabilities (Astral-PP) are shown on the left, nodes in red were different between species tree and phylogenetic trees constructed from the concatenated SNPs of the 197 nuclear genes. BS values were shown above the nodes (c). Species in strong cyto-nuclear conflicts were marked with "#"(a, b).

Fig. 6 Reticulated phylogenetic network as inferred by PhyloNet based on TRO dataset (a), SUB dataset (b), and INT dataset (c). Light blue horizontal branches indicate inferred hybridization events, and numbers next to the branches show the estimated proportion of genes contributed by each lineage in the hybridization event.

Fig. 7 Heatmap summarizing the D-statistic estimates and their P-values from 3741 tests ($P < 1 \times 10^{-6}$). Taxa P2 and P3 were displayed on the x- and y-axes and each square represented the highest estimate of each combination of P2 and P3 taxa.

Fig. 8 Divergence time estimation of *Eriobotrya* based on chloroplast genomes (a) and the 197 nuclear genes (b), and climate changes during the recent geological history in East Asia (c, Zachos et al., 2001). Red star indicated fossil calibration points based on fossil records. Median ages of the nodes are shown above the branches, with blue bars indicating the 95% highest posterior density intervals.

Fig. S1 Haplotype networks of the two nuclear genes *C23H* (a) and *GSDL2* (b) constructed from 817 *Eriobotrya* individuals.

Fig. S2 The ML phylogenetic tree constructed from the concatenated SNPs of the 197 single-copy nuclear genes. Maximum likelihood ultrafast bootstrap support values (BS>50) and Bayesian posterior probabilities (PP>0.50) are shown: "*": BS≥95% or PP ≥0.99.

Fig. S3 The map of the 20,000 simulated chloroplast trees on the empirical chloroplast tree. The percentage of simulated trees be present in the clades of the empirical chloroplast tree was shown within the node.

Fig. S4 The inferred ancestral distribution of the inner nodes based on the rooted chloroplast ML phylogenetic tree using maximum likelihood method.