

Delimiting 33 *Carpinus* (Betulaceae) species with a further phylogenetic inference

Congcong Dong¹, Zhiqiang Lu³, Han Zhang¹, Jianquan Liu^{1,2}, Minjie Li^{1*}

¹State Key Laboratory of Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou, China, 730000, People's Republic of China

²Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, People's Republic of China

³CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China

*Corresponding author: Minjie Li

Emails: liminjie@lzu.edu.cn

Abstract

Carpinus (Betulaceae) has approximately 52 species distributed in the Northern Hemisphere, with many species of *Carpinus* found in China. However, the species boundaries and phylogenetic relationships remain poorly understood. This study reported ITS sequences for 225 individuals of 33 *Carpinus* species, mainly from China. We also included eight *Ostrya* species in our analyses, the closely related sister group of *Carpinus*. We aimed to delimit these species based on ITS sequences and clarify their phylogenetic relationships by constructing tree-like topology and networks at population level. We found that only 17 of 33 species could be delimited from the closely related ones based on species-specific mutations in ITS sequence variation, including all species of sect. *Distegocarpus*, and sect. *Carpinus* subsect. *Carpinus*. *Carpinus* subsect. *Carpinus* contained two endangered species, although one seemed to be a recently originated allopolyploid species with genetic additivity from two likely parents in the ITS sequence variation. 16 species of sect. *Carpinus* subsect. *Polyneurae* were classified into three species complexes, in each of which two or more could be not distinguished from each other. The closely related species of these complexes may still diverge at the early stage without genetic distinction in the nuclear ITS sequences because of too short of divergence time and frequent gene flow. Otherwise, some species may be established based on the intraspecific variations without genetic bases for an independently evolving unit.

Keywords *Carpinus*, ITS sequences, species delimitation, phylogeny

Introduction

Accurate species delimitation is critical in most biological disciplines because species generally is the basic element in many studies (Darwin 1859; Simpson 1951; Mayr 1982). For example, such a failure of accuracy may result in over-or under-estimation of the total number of the endangered species for conservation aims (Richardson 1978; Mayden 1999). However, defining a species and delimiting species boundaries remains highly disputed (Mayden 1997; De Queiroz 2007; Liu, 2016). Except for morphological gaps at the population level, numerous studies suggest that genetic distance at the nuclear orthologous genes or chloroplast (cp) DNAs should be used to delimit species (Yang and Rannala 2010; Liu 2016; Hu *et al.* 2015; Feng *et al.* 2020), which can be further referenced as a molecular marker to barcode and identify species (Besansky *et al.* 2003; Shaw *et al.* 2005; Kress *et al.* 2005). Several cpDNAs, such as *rbcL* and *matK*, *trnH-psbA*, and *trnL-F*, were first suggested to barcode and delimit plant species (Besansky *et al.* 2003; Shaw *et al.* 2005; Kress and Erickson 2007; CBOL Plant working 2009). The cpDNAs are uniparentally (mostly maternally) inherited without discernible recombination. Therefore, besides incomplete lineage sorting and low-divergence, these cpDNA markers may fail to distinguish the closely related species because of hybridization and organelle introgression (Degnan and Rosenberg 2009; Liu 2016; Chan *et al.* 2017). The effectiveness of ITS sequence variation has been widely tested in delimiting species in diverse groups of plants (e.g. Wang *et al.* 2011; Hu *et al.* 2015; Feng *et al.* 2021), although it still fails to distinguish two species with an extremely short divergence.

This study aimed to delimit the genus *Carpinus* L.(Betulaceae) using ITS sequences. This genus contains approximately 52 species, disjunctively distributed in Europe, North America, and East Asia as a typical monoecious genus with both pollen and fruits dispersed by the wind over long distances (Li and Skvortsov 1999). Approximately 35 species were found in East Asia, and 29 were endemic to China (Li and Skvortsov 1999; Fu 2000). Two sections (sects.) *Distegocarpus* and *Carpinus*, and three subsections (subsects. *Carpinus*, *Monbeigianae*, and *Polyneurae*) have been recognized within *Carpinus* (Li and Cheng 1979). The delimitation between subsections *Monbeigianae* and *Polyneurae* were distinct based on statistical analyses of morphological characters from leaves and fruit bracts (Jeong and Chang 1997). Congruently, intraspecific variations and morphological polymorphisms led to significant difficulties in delimiting some close *Carpinus* species (Hu 1964; Li and Cheng 1979; Li and Skvortsov 1999; Yoo and Wen 2007; Li, 2008). During the phylogenetic construction of *Carpinus*, the ITS sequences were reported for 13 species with only nine different sequences identified (Yoo and Wen 2007), suggesting that at least four species shared the same ITS sequence variation with others. Similarly, the combined cpDNAs (*matK*, *trnL-trnF*, and *psbA-trnH*), of which at least two (*trnL-trnF* and *psbA-trnH*) were used to discriminate the closely related species (Kress and Erickson 2007; CBOL Plant working 2009), showed the very low variation between sampled species. Only five different sequences were identified, with at least eight species sharing the same cpDNA sequence variations with other species (Yoo and Wen 2007). In several studies, new species were described for the genus *Carpinus* (Liu and Lin 1986; Liang and Zhao 1991; Yi 1992; Tong *et al.* 2014; Lu 2017; Lu *et al.* 2018), and totaling about 40 Chinese *Carpinus* species have been recorded, among which, two prominently endangered species were classified as *C. tientaiensis* Cheng and *C. putoensis* Cheng (Qin and Zhao 2017). It was necessary to scale interspecific genetic distinctness between these species using an ITS marker that designated the sister and closely related species in numerous groups (Li *et al.*, 2011; Wang *et al.* 2011; Hu *et al.* 2015; Feng *et al.* 2021). It remains unknown how many, especially the two endangered tree species, are 'good' species with evolutionary distinctness as independently evolving units.

This study collected 191 individuals of 22 species for the genus *Carpinus*. Multiple individuals and populations were collected for most species. We downloaded all available ITS sequences for *Carpinus* and then *Ostrya*, a sister group of *Carpinus* in Betulaceae. We aimed to examine whether ITS sequences showed variation between different individuals and populations of each species. We further explored whether these sequences could distinguish most species of the genus defined by morphological traits found in most plant groups (Li *et al.* 2011). In addition, we also hoped to explore whether such a delimitation showed species dia. Finally, we used the ITS sequences to construct phylogenetic relationships for this genus. These results are beneficial for improving our understanding of taxonomy, species divergence, and phylogeny of the genus.

Material and method

Species sampling, sequencing, and alignment

According to the Flora of China (33 species, Li and Skvortsov, 1999) and the recently announced new species of *Carpinus*, 191 samples of 22 *Carpinus* species were collected with more than two samples included for each species (see detail in Table S1). The detailed information of collection sites for these samples is shown in Table S1. All fresh leaves of the samples were dried in silica gel for total DNA extraction. The specimens of these samples were deposited in the Herbarium of Lanzhou University, China. We also downloaded 34 ITS sequences of 11 *Carpinus* species, 27 ITS sequences of eight *Ostrya* species, and two ITS sequences of two *Corylus* species from NCBI with the accession numbers listed in Table S2 for further experimentation.

The total genomic DNA (gDNA) for each target sample was extracted according to the modified CTAB method, with approximately 30 milligrams (mg) of dried leaves used (Li *et al.* 2013). The PCR amplification mixture had a total volume of 25 microliters (μL), made up of 10–40 nanograms per microliter ($\text{ng}/\mu\text{L}$) of plant gDNA, 2.5 μL of 10 $\times$ PCR buffer, 0.5 millimoles per liter (mmol/L) of dNTPs, 2 $\mu\text{L}/\text{L}$ of ITS forward and reverse sequence primers (Wen and Zimmer 1996), 0.2–0.3 μL rTaq enzyme [5 units per microliter ($\text{U}/\mu\text{L}$); Takara, Dalian, China] and ddH₂O. The PCR reaction was performed on a T1 PCR instrument (Biometra, Germany) using an optimized program in which the initial denaturation temperature was at 94 °C for 4 minutes (min), then 36 cycles at 94°C for 45 seconds (sec), 54 °C renaturations for 50 sec, 72 °C extensions for 90 sec, and one 72 °C extension for 10 min after the end of the 36th cycle. All PCR products were detected by 1% agarose gel electrophoresis, and then the products were then purified using the TIAN quick Midi Purification Kit according to the protocols (Beijing, China). All purified samples were sequenced using Sanger technology in Tsingke (China, Wuhan). The newly generated ITS sequences were stored in NCBI with the accession number MW928890–MW929080 and OK560470–OK560480 (see detail in Table S2).

We found that the ITS sequence (GenBank accession No. AF432051) of the prominently endangered species *C. putosensis* was clustered together with *C. mianningensis* T. P. Li (see the result section). Given the chromosome number of *C. putoensis* with $2n = 14x = 112$ (Meng *et al.* 2004), we phased the ITS sequences of *C. putoensis* to ascertain whether it is an allopolyploid or autopolyploid according to the following procedures. The purified PCR products of *C. putosensis* were recombined into pMD 19-T vectors according to the protocols of this vector kit (Baosheng, Dalian, China). The reaction system was prepared in 15 μL with 1.5 μL pMD 19-T, 1.5 μL of PCR product, 4.5 μL aseptic water (H_2O), and 7.5 μL of ligation

solution. The reaction was carried out in Multi Temp III (America, GE) at 16 °C for 1 hour. The recombinant pMD 19-T vectors were then transformed into DH5 α competent Cell (Dongsheng Biotech), and blue-white screening was used to select the positive clones for Sanger sequencing in Tsingke (China, Wuhan) by the *BcaBEST*TM sequencing primers (Baosheng, Dalian, China).

We employed SeqMan (DNASTar, Burland 2000) to edit the forward and reverse sequences to obtain a consensus sequence for each newly generated ITS sequence. All sequences were aligned using MEGA v7.0 (Kumar *et al.* 2016), employing fine manual adjustment. In the aligned sequence dataset, insertions and deletions were regarded as missing data in phylogenetic analyses.

Phylogenetic analyses

Bayesian inference (BI), maximum likelihood (ML), and maximum parsimony (MP) approaches were deployed to reconstruct the phylogeny of *Carpinus* based on the ITS dataset. The Akaike information criterion (AIC) was used to evaluate the best-fitting model in jModelTest v2.1.7 (Darriba *et al.* 2012). The ML analysis was implemented using RAxML version 8.2.8 (Stamatakis 2014) with 1000 bootstrap replicates using the GTRGAMMA nucleotide substitution model estimated from AIC. BI analysis was performed in MrBayes v3.2.0 (Ronquist and Huelsenbeck 2003) from a random starting tree with the best-fitting GTR model, and then four Markov Chain Monte Carlo (MCMC) chains consisting of three heated and one cold chain were run to estimate the posterior distribution of the model parameters. We ran 10,000,000 generations for the trees and drew a tree every 1000 generations. The first 20% of trees were discarded as burn-in, and the remaining trees were applied to infer the majority-consensus tree and the posterior probabilities. MP analysis was performed using PAUP v4.0 (Swofford 2002). Three analyses were carried out independently using four MCMC (three hot and one cold chain), with 10,000,000 generations for each run and sampling every 1000 generations. The first 25% of trees were discarded as burn-in, and the remaining trees were used to construct a 50% majority-rule consensus tree. FigTree version 1.4.4 (Rambaut 2018) visualized the final trees.

The neighbor-net graphic was analyzed in SplitsTree4 (Hudson and Bryant 2006) using an uncorrected P-distance method with 1000 replicates for bootstrapping to support and further ensure the phylogenetic relationships. DnaSP v5.10.01 (Librado and Rozas 2009) generated ribotypes among *Carpinus* species. Genealogical relationships of ribotypes were inferred from a median-joining method in Network 4.6.1.3 (Bandelt *et al.* 1999). The geographical distribution of ribotypes was recorded at a species-defined population level.

Results

Sequence sampling and variation

In this study, we downloaded two *Corylus* species as outgroups (*C. californica* (A. DC.) A. Heller and *C. heterophylla* Fisch. ex Trautv. and 27 ITS sequences for eight *Ostrya* species from the NCBI database, including two *O. carpinifolia* Scop., seven *O. japonica* Sarg., two *O. knowltonii* Sarg., two *O. multinervis* Rehder, five *O. rehderiana* Chun, four *O. trichocarpa* D. Fang & Y. S. Wang, three *O. virginiana* (Mill.) K. Koch, and two *O. yunnanensis* W. K. Hu (Table S2). We also retrieved 34 ITS sequences of 11 *Carpinus* species from the NCBI database, including six *C. betulus*, five *C. caroliniana*, five *C. japonica* Blume, two *C. kawakamii* Hayata, two *C. laxiflora*, two *C. mianningensis*, three *C. orientalis* Mill., one *C. putoensis*, two *C. rankanensis* Hayata, three *C. tibetana* Z. Qiang Lu

& J. Quan Liu, and three *C. tientaiensis* (Table S2). We newly generated 202 ITS sequences of 23 *Carpinus* species, including four *C. chuniana* Hu, seven *C. cordata*, two *C. fangiana*, nine *C. fargesiana* H. J. P. Winkl., ten *C. henryana* (H. J. P. Winkl.) H. J. P. Winkl., 22 *C. hupeana* Hu, five *C. langaoensis* Z. Qiang Lu & J. Quan Liu, two *C. londoniana*, two *C. mollicoma* Hu, two *C. monbeigiana*, two *C. omeiensis* Hu & W. P. Fang, eight *C. polyneura* Franch., six *C. pubescens*, five *C. purpurinervis* Hu, 11 ITS clones of *C. putoensis*, two *C. rupestris* A. Camus, three *C. shensiensis* Hu, three *C. stipulate* H. J. P. Winkl., 15 *C. sungpanensis* W. Y. Hsia, 11 *C. tsaiana* Hu, 36 *C. tschonoskii*, three *C. turczaninowii*, and 32 *C. viminea* (Table S1). This process obtained 236 ITS sequences of 33 *Carpinus*, including two European species (*C. betulus* and *C. orientalis*), one North American species (*C. caroliniana*), and 30 Asian species. In combination with 27 *Ostrya* sequences and two *Corylus* sequences, the length of these 265 ITS sequences varied between 599 base pairs (bp) to 603 bp with the alignment length of 613 bp. A total of 102 parsimony informative sites were identified among these ITS sequences.

Phylogenetic analyses

The resulting trees showed an almost identical tree topology among BI, ML, and MP analyses (Fig. 1). Our phylogenetic tree revealed that all *Carpinus* species were clustered into two groups, corresponding to the previous morphologically-defined sections [i.e. sect. *Distegocarpus* (PP/BS/MP = 0.96/83/90)], the support values successively represented as BI posterior probability (PP), ML bootstrap support (BS), and MP bootstrap value, and sect. *Carpinus* (PP/BS/MP = 1/76/58). However, *Carpinus* was considered non-monophyletic because of a tentative affinity between sections *Carpinus* and *Ostrya* (PP/BS/MP = 0.96/<50/<50) were observed, where both were sister to sect. *Distegocarpus* with high support values (PP/BS/MP = 1/100/100) (Fig. 1). Within sect. *Distegocarpus*, four species were well delimited, and *C. japonica* (PP/BS/MP = 1/98/98) and *C. fangiana* (PP/BS/MP = 0.98/97/93) were successively diverged from the closest related *C. rankanensis* (PP/BS/MP = 0.98/100/77) and *C. cordata* (PP/BS/MP = <0.5/<50/<50). However, just 13 of the 29 sect. *Carpinus* species were discriminated from other closely related species. In addition, three highly supported clades were identified in the sect. *Carpinus* and species delimitation in the first two clades were well solved but not in the third. We found that *C. caroliniana* (PP/BS/MP = 1/100/100), *C. laxiflora* (PP/BS/MP = 0.84/99/80), and *C. viminea* (samples were paraphyletic) joined in the first clade (PP/BS/MP = 1/97/92) with unresolved interspecific relationships (Fig. 1). Similarly, species were successfully divided into monophyletic lineages in the second clade, except the ITS clones of *C. putoensis*, which clustered together with *C. mianningensis* (PP/BS/MP = 0.61/93/<50) and *C. tschonoskii* var. *tschonoskii*, respectively (PP/BS/MP = 1/62/82, Fig. 1). Further, we found that the European *C. betulus* (PP/BS/MP = 1/97/100) and Asian *C. tientaiensis* (PP/BS/MP = 1/99/100) were successfully sister to other Asian species (PP/BS/MP = 1/94/70 and PP/BS/MP = 0.9/66/<50). *C. putoensis* - *C. mianningensis* showed a genetic affinity (PP/BS/MP = 0.72/56/83) with *C. langaoensis* (PP/BS/MP = 1/97/87), and all of them were sister to *C. putoensis* - *C. tschonoskii* (PP/BS/MP = 1/100/99) with low support values (PP/BS/MP = 0.58/<50/<50). Contrary to the first two clades, only four of the 20 species in the third clade (PP/BS/MP = 1/87/92) were identified from the closely related other, including *C. tibetana* (samples were paraphyletic), *C. orientalis* (PP/BS/MP = 0.85/60/<50), *C. monbeigiana* (PP/BS/MP = 0.66/100/95) and *C. kawakamii* (PP/BS/MP = 0.71/96/80). The rest of the species were clustered into three species complexes. *C. londoniana*, *C. mollicoma*, *C. omeiensis*, *C. polyneura*, and *C. rupestris* formed the basal species complex I; *C. chuniana* and *C. tsaiana* constituted species complex II and were sister to species complex III,

including *C. hupeana*, *C. fargesiana*, *C. henryana*, *C. pubescens*, *C. purpurinervis*, *C. shensiensis*, *C. stipulata*, *C. sungpanensis* and *C. turczaninowii*.

We inferred the neighbor-net graph using uncorrected P-distance to confirm species boundaries and phylogenetic affinities further. Compared to the tree-like topology, we found a clear genetic split between the genera *Carpinus* and *Ostrya*. The sects. *Distegocarpus* and *Carpinus* formed a separate group and were sisters (Fig. 2). However, species delimitation and interspecific relationships revealed by the neighbor-net graph were principally consistent with the ITS tree topology (Fig. 1 and Fig. 2). Within sect. *Distegocarpus*, the four species involved were well delimited. Within the first clade of sect. *Carpinus*, a closer relatedness between *C. laxiflora* and *C. viminea* was supported, and both were sisters to *C. caroliniana*. The interspecific relationships among well-clustered species in the second clade agreed with those revealed by the tree-like topology (Fig. 1 and Fig. 2). Like the tree-like topology, species falling into the third clade were difficult to delimitate and mostly linked with the narrowly-meshed networks (Fig. 2). These species were further clustered into three species complexes, and their relationships were congruent with the tree-like topology (Fig. 1).

Ribotype Network of *Carpinus* species

We reconstructed the ribotype network among the *Carpinus* species to examine interspecific relationships (Fig. 3B). The alignment length of the 236 ITS sequences of 33 species of *Carpinus* was 605 bp. According to the 69 parsimony informative sites, a total of 26 ribotypes (R1–R26) were detected for 33 population-defined species (Table S3), when all ribotypes (R12₁–R12₇) representing different ITS clones of *C. putoensis* were treated as R12. We found that 19 (R1–R15 and R19–R22) of 26 ribotypes were species-specific and capable of delimiting 17 species from each other. Consistent with the tree-like topology and neighbor-net graph, two sections of *Carpinus* were split in the network. Within sect. *Distegocarpus*, four species-specific ribotypes (R1–R4) were identified, and the ribotype connections corresponded to the interspecific relationships recovered by the tree-like topology and neighbor-net graph (Fig. 2 and Fig. 3B). Within the sect. *Carpinus*, three clades were discerned, and the species-specific ribotypes (R5–R15) were identified in the first two clades, while the interspecies-shared ribotypes were primarily detected in the third clade (e.g. R16, R18, and R26). This result was indicative of the three species complexes. In detail, four ribotypes (R5–R8) were identified for the three species of the first clade, and two ribotypes were observed in *C. caroliniana* (R7–R8) (Fig. 3). Seven ribotypes (R9–R15) were detected for the six species of the second clade, and two ribotypes (R13–R14) were detected in *C. tschonoskii* (Fig. 3). Although *C. putoensis* fell together with *C. mianningensis* and *C. tschonoskii* in the tree-like topology and neighbor-net graph (Fig. 1 and Fig. 2), the species-specific ribotypes were identified for *C. putoensis* (R12₁–R12₇). Within the species-specific ribotypes for *C. putoensis* (R12₁–R12₇), two were linked to R14 of *C. tschonoskii*, and the remaining five were connected with R11 of *C. mianningensis* (Fig. 3B). In the third clade, a ‘star-like’ network was observed, and the dominant R16 was placed at the center position. Furthermore, we found that R16 was mainly shared by species complex III, including *C. fargesiana*, *C. henryana*, *C. hupeana*, *C. pubescens*, *C. purpurinervis*, *C. shensiensis*, *C. stipulata*, *C. sungpanensis*, and *C. turczaninowii*. R18 was shared by *C. chuniana* and *C. tsaiana* within species complex II, and R26 was observed among *C. mollicoma*, *C. omeiensis*, *C. polyneura*, and *C. rupestris* belonging to the species complex I (Fig. 3B). In addition, R21 was shared by *C. tibetana* and *C. orientalis*, and R23 was shared by *C. londoniana* and *C. polyneura*. *C. hupeana* (R16–R17) and *C. orientalis* (R22–R23) were found to have two ribotypes, respectively. Multiple ribotypes (R23–R26) ribotypes were identified in *C.*

polyneura. A species-unique ribotype was also observed in *C. monbeigiana* (R19) and *C. kawakamii* (R20), respectively (Table S3).

When the identified ribotypes were projected to the geographical map at species-defined populations, we found that the Asian species in the phylogenetic analyses showed no shared ribotypes with the other species even when they were distributed at the same localities (Fig. 3A). However, the interspecies-shared R16 was widespread in the whole distribution range and identified among different species collected in the same localities. The shared R26 was found in different species (e.g. *C. mollicoma*, *C. polyneura* and *C. rupestris*) sampled at the same locality and found in a distinct locality where *C. omeiensis* was sampled. The shared R18 was observed in different localities for different species. Noticeably, R18 and R26 were mainly distributed in the southeast inland of China, specifically occurring at localities with closer geographical distances. The shared R21 occurred all across Eurasia.

Discussion

It is critical to determine whether one described species represents an independently evolving unit because some species are established based on intraspecific variants or interspecific hybrids (Liu 2016). The establishment of independently evolving units is crucial for conserving endangered species (Wiens 2007). In this study, we used the widely used genetic marker ITS for species identification (Li *et al.* 2011) to examine species distinctness of the genus *Carpinus*, containing two well-known endangered tree species (Qin and Zhao 2017). By sampling multiple individuals for each species, we aimed to examine whether a genetic gap could exist between two species to exclude intraspecific variations and interspecific hybrids at the utmost extent. We clustered all exemplified individuals into three groups, sect. *Distegocarpus*, subsect. *Carpinus* and subsect. *Polyneurae*. We found that only 17 of 33 species seemed to be ‘good’ species, which ITS sequence variations could diagnose. In addition, most species within the sect. *Distegocarpus*, and subsect. *Carpinus* were well delimited. However, some species of subsect. *Polyneurae* could not be discriminated by the ITS fragment from the closely related sister species. Recent divergence and frequent hybridization may have blurred interspecific boundaries in these species complexes, although further population-level statistical analyses are needed to examine whether morphological gaps exist between these ‘species’ without genetic distinctness.

‘Good’ species in two of three phylogenetic groups

The phylogeny-based method distinguished species because all individuals of one species were assumed to be derived from one common ancestor and therefore clustered together as one monophyletic group (Liu, 2016). In addition, such analyses can further construct phylogenetic relationships. Phylogenetic analyses of 236 ITS sequences from 33 *Carpinus* species revealed two distinct clades corresponding to the morphology-related sects. *Distegocarpus* and *Carpinus* (Li and Skvortsov 1999). This result is consistent with the previous phylogenetic results based on ITS and the nuclear low-copy nitrate reductase (*Nia*) gene (Chen *et al.* 1999; Kato *et al.* 1999; Yoo and Wen 2007; Li 2008; Grimm and Renner 2013; Lu *et al.* 2018).

No shared ribotypes were observed between sect. *Distegocarpus* and sect. *Carpinus*, suggesting two independent evolutionary lineages without gene flow (Fig. 3B). Four species in sect. *Distegocarpus* were delimited as independently evolving units (Fig. 3B). Within sect. *Carpinus*, three groups were identified but did not correspond to the three subsections (Fig. 1, Fig. 2, and Fig. 3B). According to bract characters (Li and Skvortsov 1999), the first two groups belong to subsect. *Carpinus* whose bracts were lobed at bases of inner and outer

margins, while the third one corresponded to subsects. *Monbeigianae* and *Polyneurae* with only an inflexed lobe or auricle at the base of the inner margins. This result suggests that these two subsections could be incorporated into one, i.e. subsect. *Polyneurae* (Jeong and Chang 1997) (Fig. 1). In addition, no ribotype was shared between nine species of subsect. *Carpinus*, of which, eight species including *C. betulus*, *C. caroliniana*, *C. mianningensis*, *C. langaoensis*, *C. laxiflora*, *C. tientaiensis*, *C. tschonoskii*, and *C. viminea* were well discriminated (Fig. 3B), with an exception that the obtained ITS clones of *C. putoensis* were mixed with *C. mianningensis* and *C. tschonoskii* (Fig. 1 and Fig. 2).

It should be noted that *C. tientaiensis* has a very small population size, with only 19 wild individuals recorded (Qin and Zhao 2017). The ITS sequence of *C. tientaiensis* differs from those of the closely related species with four species-specific mutations. Therefore, *C. tientaiensis* is a well-delimited ‘good’ species. The field investigations showed that *C. putoensis* is a critically endangered species, containing one individual that was more than 250 years old, occurring only on the Zhoushan Islands of Zhejiang Province, China (<http://www.iplant.cn/rep/prot/Carpinus%20putoensis>). The chromosome number of *C. putoensis* was found to be a polyploid species with $2n = 14x = 112$ (Meng *et al.* 2004). Our phased ITS sequences from *C. putoensis* suggested that one type of ITS from this species was derived from *C. mianningensis*, and other ITS sequences were derived from *C. tschonoskii*. Therefore, *C. putoensis* may originate from the hybrid polyploidization between *C. mianningensis* and *C. tschonoskii*. Therefore, *C. putoensis* remains an independently evolving lineage because the allopolyploidization immediately leads to direct postzygotic isolation (Li *et al.* 2021). However, because of the small population (one individual), rare sexual recombination, tree life longevity, and early speciation stage, multiple ITS copies from both parents have not experienced the concerted evolution into a single sequence and not accumulated the species-specific mutations.

Low species discrimination in subsect. *Polyneurae*

There are more described species (20) in the subsect. *Polyneurae* than other two groups: sect. *Distegocarpus* (4) and subsect. *Carpinus* (9). In addition, subsect. *Polyneurae* has a large distributional range than sect. *Distegocarpus* and subsect. *Carpinus* (i.e. the distribution of R16, Fig. 2). Most species of subsect. *Polyneurae* have large populations, and none of them have been listed as endangered. Surprisingly, only four out of 20 species were distinguished (i.e. *C. kawakamii*, *C. monbeigiana*, *C. orientalis*, and *C. tibetana*). However, the rest of the 16 species were clustered into three different species complexes due to the shared ribotypes within each complex. These species complexes can be distinguished from others and contain at least three ‘good’ species. Although *C. londoniana* has similar bracts with *C. tientaiensis* and *C. viminea* of subsect. *Carpinus*, *C. londoniana* was clustered with species of subsect. *Polyneurae* and shared R23 with *C. polyneura* in phylogenetic analyses (Fig. 1 and Fig. 2). Species complex I comprised *C. londoniana*, *C. mollicoma*, *C. omeiensis*, *C. polyneura*, and *C. rupestris*, with only the shared R26 detected and the rest of the ribotypes (R23–R25) mostly unique to *C. polyneura* (Fig. 3B). Species complex II comprised *C. chuniana* and *C. tsaiiana*, where R18 was detected between them (Fig. 3B). Compared to the first two, species complex III comprised nine species, with only the widespread R16 observed among these species, with R17 specific to *C. hupeana* (Fig. 3B).

Three explanations may account for low species discrimination in the three species complexes of subsect. *Polyneurae*. Firstly, species divergence within these complexes is still at the early stage of evolution. Although morphological distinctness can be detected, unique and specific mutations have not accumulated between these closely related species. Secondly,

the described species within some species complexes (for example, *C. mollicoma*, *C. polyneura*, and *C. rupestris*) are adjacently distributed. These adjacent distributions promote frequent gene flow through wind-mediated pollen and fruit dispersal (Li and Skvortsov 1999) but substantially reduce divergence for evolving units. The occurrence records suggest that most of these described species were distributed widely with large population sizes. The large population sizes and the long generation times of these tree species may further delay the speciation process and shorten the time scales for producing and accumulating unique mutations because of fewer intra-specific recombination and frequent inter-specific gene flow (Li and Skvortsov 1999). However, the endangered species with small population sizes (for example, *C. tientaiensis*) may accumulate specific mutations quickly because of the increased intra-specific recombination and reduced hybridization with the closely related species. Finally, it needs to be noted that some species in these species complexes are described by intraspecific variations rather than independently evolving units (Liu 2016). Within this scenario, morphological and genetic gaps could not be distinguished through population-level analyses even when based on sensitive molecular markers for species divergence, such as genome re-sequencing data (e.g. Yang *et al.* 2019; Li *et al.* 2019).

In the future, related studies should be conducted to confirm retaining or incorporating these species because the different taxonomic species should represent independently evolving lineages with both morphological and genetic gaps at the population level despite the early speciation stages (Liu 2016).

Appendix of 40 species and 8 varietas in the genus *Carpinus* L. in China and the infrageneric categories based on a combination of genetic and morphological differences:

1. Sect. *Distegocarpus* (Sieb. et Zucc.) Sarg.
 - 1.1. *Carpinus cordata* Blume
 - 1.1.a. *Carpinus cordata* var. *cordata*
 - 1.1.b. *Carpinus cordata* var. *mollis* (Rehd.) Cheng ex Chen in Y. Chen
 - 1.1.c. *Carpinus cordata* var. *chinensis* Franch.
 - 1.2. *Carpinus fangiana* Hu
 - 1.3. *Carpinus rankanensis* Hayata
 - 1.3.a. *Carpinus rankanensis* var. *rankanensis*
 - 1.3.b. *Carpinus rankanensis* var. *matsudae* Yamamoto
2. Sect. *Carpinus*——*Carpinus* sect. *Eucarpinus* Sarg.
 - 2.1. subsect. *Carpinus*
 - 2.1.1. *Carpinus langaoensis* Z. Qiang Lu & J. Quan Liu
 - 2.1.2. *Carpinus laxiflora* (Siebold & Zucc.) Blume
 - 2.1.3. *Carpinus londoniana* H. Winkler
 - 2.1.3.a. *Carpinus londoniana* var. *londoniana*
 - 2.1.3.b. *Carpinus londoniana* var. *xiphobracteata* P. C. Li
 - 2.1.3.c. *Carpinus londoniana* var. *lanceolata* (Hand.-Mazz.) P. C. Li
 - 2.1.3.d. *Carpinus londoniana* var. *latifolia* P. C. Li
 - 2.1.4. *Carpinus mianningensis* T. P. Li
 - 2.1.5. *Carpinus tientaiensis* W. C. Cheng
 - 2.1.6. *Carpinus tschonoskii* Maximowicz
 - 2.1.6.a. *Carpinus tschonoskii* var. *tschonoskii*
 - 2.1.6.b. *Carpinus tschonoskii* var. *falcatisbracteata* (Hu) P. C. Li
 - 2.1.7. *Carpinus viminea* Lindl.

- 2.1.7.a. *Carpinus viminea* var. *viminea*
 2.1.7.b. *Carpinus viminea* var. *chiukiangensis* Hu
 2.2. subsect. *Polyneurae*
 2.2.1. *Carpinus chuniana* Hu
 2.2.2. *Carpinus dayongina* K. W. Liu & Q. Z. Lin
 2.2.3. *Carpinus fargesiana* H. Winkler
 2.2.3.a. *Carpinus fargesiana* var. *fargesiana*
 2.2.3.b. *Carpinus fargesiana* var. *hwai* (Hu & W. C. Cheng) P. C. Li
 2.2.4. *Carpinus firmifolia* (H. Winkler) Hu
 2.1.5. *Carpinus gigabracteatus* Z. Qiang Lu
 2.2.6. *Carpinus hebestroma* Yamamoto
 2.2.7. *Carpinus henryana* (H. Winkler) H. Winkler
 2.2.8. *Carpinus hupeana* Hu
 2.2.9. *Carpinus insularis* N. H. Xia, K. S. Pang & Y. H. Tong
 2.2.10. *Carpinus kawakamii* Hayata
 2.2.11. *Carpinus kweichowensis* Hu
 2.1.12. *Carpinus mengshanensis* S. B. Liang & F. Z. Zhao
 2.2.13. *Carpinus microphylla* Z. C. Chen ex Y. S. Wang & J. P. Huang
 2.2.14. *Carpinus minutiserrata* Hayata
 2.2.15. *Carpinus mollicoma* Hu
 2.2.16. *Carpinus monbeigiana* Hand.-Mazz.
 2.2.17. *Carpinus oblongifolia* (Hu) Hu & W. C. Cheng
 2.2.18. *Carpinus omeiensis* Hu & Fang
 2.2.19. *Carpinus polyneura* Franch.
 2.2.20. *Carpinus pubescens* Burkill
 2.2.21. *Carpinus purpurinervis* Hu
 2.2.22. *Carpinus putoensis* W. C. Cheng
 2.2.23. *Carpinus rupestris* A. Camus
 2.2.24. *Carpinus shensiensis* Hu
 2.2.25. *Carpinus stipulata* H. Winkler
 2.2.26. *Carpinus sungpanensis* W. Y. Hsia
 2.2.27. *Carpinus tibetana* Z. Qiang Lu & J. Quan Liu
 2.2.28. *Carpinus tsaiiana* Hu
 2.2.29. *Carpinus tsunghienensis* Hu
 2.2.30. *Carpinus tureczaninowii* Hance

Data Availability

The newly generated data were submitted in NCBI (<https://www.ncbi.nlm.nih.gov/>) with the GenBank accession numbers shown in Table S1; the GenBank accession numbers of downloaded sequences from NCBI are shown in Table S2.

Supporting Information

Table S1. The collection information of 191 individuals in *Carpinus* and the haplotypes of all individuals in this study.

Table S2. All sequences downloaded from NCBI in this study.

Table S3. The haplotypes of all *Carpinus* samples in this study and their corresponding species.

Sources of Funding

We thank the Supercomputing Center of Lanzhou University for computing supporting. This work was supported by the National Natural Science Foundation of China (31901074 and 31590821), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31010300).

Authors' contributions

JQL conceived the study. ZQL collected the samples and carried out the experiments. CCD, HZ and MJL performed the analyses and wrote the manuscript. JQL provided technical and fee support in the experiments. All authors read and reviewed the manuscript.

Acknowledgements

The authors would like to thank Kelsey McWilliams at Texas A&M University for her assistance with English language and grammatical editing, and thank editor Jeremy Beaulieu, and the anonymous reviewers for their time and effort on the improvement of our manuscript.

Conflict of Interest

The authors declare no competing interests.

Accepted Manuscript

Literature Cited

- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37–48.
- Besansky NJ, Severson DW, Ferdig MT. 2003. DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. *Trends Parasitol* 19:545–546.
- CBOL Plant Working Group 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106:12794–1279.
- Chan KO, Alexander AM, Grismer LL, Su YC, Grismer JL, Quah ESH, Brown RM. 2017. Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Molecular Ecology* 26:5435–5450.
- Chen ZD, Manchester SR, Sun HY. 1999. Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. *American Journal of Botany* 86:1168–1181.
- Darriba D, Taboada G, Doallo R. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772.
- De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56:879–886.
- Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24:332–340.
- Feng XY, Wang XH, Chiang YC, Jian SG, Gong X. 2021. Species delimitation with distinct methods based on molecular data to elucidate species boundaries in the *Cycas taiwaniana* complex (Cycadaceae). *Taxon* 64:714–726.
- Fu LG. 2000. *Higher Plants of China*. China Shandong: Qingdao Press.
- Grimm GW, Renner SS. 2013. Harvesting Betulaceae sequences from GenBank to generate a new chronogram for the family. *Botanical Journal of the Linnean Society* 172:465–477.
- Hu H, Al-Shehbaz I, Sun Y, Hao G, Qian W, Liu JQ. 2015. Species delimitation in *Orychophragmus* (brassicaceae) based on chloroplast and nuclear DNA barcodes. *Taxon* 64:714–726.
- Hu HH. 1964. The materials on the monography of genus *Carpinus* Linn. of China. *Acta Phytotaxonomica Sinica* 9:281–298.
- Hudson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23:254–267.
- Jeong J, Chang CS. 1997. Reconsideration of *Carpinus* L. (Betulaceae) of Korea primary based on quantitative characters. *Korean Journal of Plant Taxonomy* 27:157–187.
- Kato H, Oginuma K, Gu Z, Hammel B, Tobe H. 1999. Phylogenetic relationships of Betulaceae based on matK sequences with particular reference to the position of *Ostryopsis*. *Acta Phytotaxonomica et Geobotanica* 49:89–97.

- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102:8369–8374.
- Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* 2:e508.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33:1870–1874.
- Li DZ, Gao LM, Li HT, Wang H, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* 108:19641–19646.
- Li J, Wang S, Jing Y, Wang L, Zhou SL. 2013. A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany* 48:72–78.
- Li JH. 2008. Sequence of low-copy nuclear gene support the monophyly of *Ostrya* and paraphyly of *Carpinus* (Betulaceae). *Journal of Systematics and Evolution* 46:333–340.
- Li MJ, Zheng ZY, Liu JC, Yang YX, Ren GP, Ru DF, Zhang SZ, Du X, Ma T, Richard M, Liu, JQ. 2021. Evolutionary origin of a tetraploid *Allium* species on the Qinghai–Tibet Plateau. *Molecular Ecology* 30:5780–5795.
- Li PC, Skvortsov AK. 1999. Betulaceae. In: Wu CY, Raven PH. eds. *Flora of China*. Beijing: Science Press and Missouri: Missouri Botanical Garden Press, 4:284–313.
- Li PC, Cheng SX. 1979. Betulaceae. in Kuang KZ, Lee PC, eds. *Flora republicae popularis sinicae*. Beijing: Science, 21:44–137.
- Li Y, Yang YZ, Yu L, Du X, Ren GP. 2018. Plastomes of nine hornbeams and phylogenetic implications. *Ecology and Evolution* 8:8770–8778.
- Liang SB, Zhao FZ. 1991. A new species of *Carpinus* from Shandong. *Bulletin of Botanical Research* 11:33–34.
- Librado P, and Rozas J. 2009. DnaSP v5, a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Liu JQ. 2016. The integrative species concept and species on the speciation way. *Biodiversity Science* 24:1004–1008.
- Liu KW, Lin QZ. 1986. A new species of *Carpinus* from hunan. *Bulletin of Botanical Research* 6:143–145.
- Lu ZQ, Liu SY, Yang XY, Liang QL, Yang YZ, Zhang D, Richard M, Liu JQ. 2017. *Carpinus langaoensis* (Betulaceae), a new hornbeam species from the Daba Mountains in Shaanxi, China. *Phytotaxa* 295:185–193.
- Lu ZQ, Li Y, Yang XY, Liu JQ. 2018. *Carpinus tibetana* (Betulaceae), a new species from southeast Tibet, China. *PhytoKeys* 98:1–13.
- Mayden RL. 1997. A hierarchy of species concepts: The denouement in the saga of the species

- problem. In: Michael. FC, Dawah HA, Wilson MR, eds. *Species: The Units of Biodiversity*. UK, London: Chapman and Hall, 54:381–424.
- Mayden R L. 1999. Consilience and a hierarchy of species concepts: advances toward closure on the species puzzle. *Journal of Nematology* 31:95–116.
- Mayr E. 1982. The growth of biological thought: Diversity, evolution, and inheritance. Cambridge, MA: Harvard University Press.
- Prantl K. 1894. Betulaceae. in A Engler, K Prantl, eds. *Die natu"rlichen Pflanzenfamilien*. Leipzig: Engelmann, 3:38–46.
- Qin HN, Zhao LN. 2017. Evaluating the threat status of higher plants in China. *Biodiversity Science*. 25:689–695.
- Rambaut A. 2018. FigTree–version 1.4.4, Tree figure drawing tool. Available from <http://tree.bio.ed.ac.uk/software/figtree>.
- Richardson IRK. 1978. Endemic taxa and the taxonomist. In: stress HE, ed. *Essays in Plant Taxonomy*. London: Academic Press, 245–262.
- Ronquist F, Huelsenbeck JP. 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92:142–166.
- Simpson GG. 1951. The Species Concept. *Evolution* 5:285–298.
- Swofford DL. 2002. PAUP: Phylogenetic Analysis Using Parsimony (and other methods), Version 4.0a149. Massachusetts: Sunderland Sinauer.
- Stamatakis A. 2014. RaxML version 8: A tool for phylogenetic analysis and post–analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Tong YH, Pang K, Xia N. 2014. *Carpinus insularis* (Betulaceae), A New Species from Hong Kong, China. *Journal of Tropical and Subtropical Botany* 22:121–124.
- Wang Q, Yu QS, Liu JQ. 2011. Are nuclear loci ideal for barcoding plants? A case study of genetic delimitation of two sister species using multiple loci and multiple intraspecific individuals. *Journal of Systematics and Evolution* 49:82–188.
- Wen J, EA Zimmer. 1996. Phylogeny and biogeography of *Panax* L. (the Ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution* 6:166–177.
- Wiens JJ. 2007. Species delimitation: New approaches for discovering diversity. *Systematic Biology* 56:875–878.
- Yang XY, Wang ZF, Luo WC, Guo XY, Zhang CH, Liu JQ, Ren GP. 2019. Plastomes of Betulaceae and phylogenetic implications. *Journal of Systematics and Evolution* 57:508–518.

- Yang ZH, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 107:9264–9269.
- Yi TP. 1992. A new species of genus *Carpinus* L. from Sichuan. *Bulletin of Botanical Research* 12(4):335–337.
- Yoo KO, Wen J. 2002. Phylogeny and biogeography of *Carpinus* and subfamily Coryloideae (Betulaceae). *International Journal of Plant Sciences* 163:641–650.
- Yoo KO, Wen J. 2007. Phylogeny of *Carpinus* and subfamily Coryloideae (Betulaceae) based on chloroplast and nuclear ribosomal sequence data. *Plant Systematics and Evolution*. 267:25–35.

Accepted Manuscript

Figures

Figure 1 Phylogenetic analysis of *Carpinus*, with *Ostrya* and *Corylus* as the outgroup based on ITS sequences. (A) Tree-like topology. Bayesian inference (BI) phylogenetic tree is shown due to the similar topologies with maximum likelihood (ML) and Maximum parsimony (MP). The numbers on the branches indicate PP/BS/MP support values from BI posterior probabilities, ML bootstrap supports, and MP bootstrap values. - indicates PP/BS/MP values less than 0.5/50/50. The infrageneric categories of *Carpinus* are represented by different colors: red for the sect. *Distegocarpus*, yellow and green for the sect. *Carpinus* subsect. *Carpinus*, and purple for the sect. *Carpinus* subsect. *Polyneurae*, and black for *Ostrya*. I, II, and III represent the three species complexes in subsect. *Polyneurae*.

Figure 2 Neighbor-net graph of *Carpinus*, *Ostrya*, and *Corylus* using ITS sequences. The infrageneric categories of *Carpinus* correspond to those in Figure 1. Within in subsect. *Polyneurae*, members of the first two species complexes are shown, but not for the members of species complex III due to the narrowly-meshed networks among those species, except some individuals of *C. hupeana*.

Figure 3 Ribotype analysis of *Carpinus* species based on ITS sequences. (A) Geographical distribution of ribotypes of Asian *Carpinus* species. The solid lines indicate the corresponding relationships between species and ribotypes. (B) Ribotype network of *Carpinus* species. Circle size indicates the ribotype frequency. Small red circles indicate extinct or un-sampled ribotypes.

Figure 1

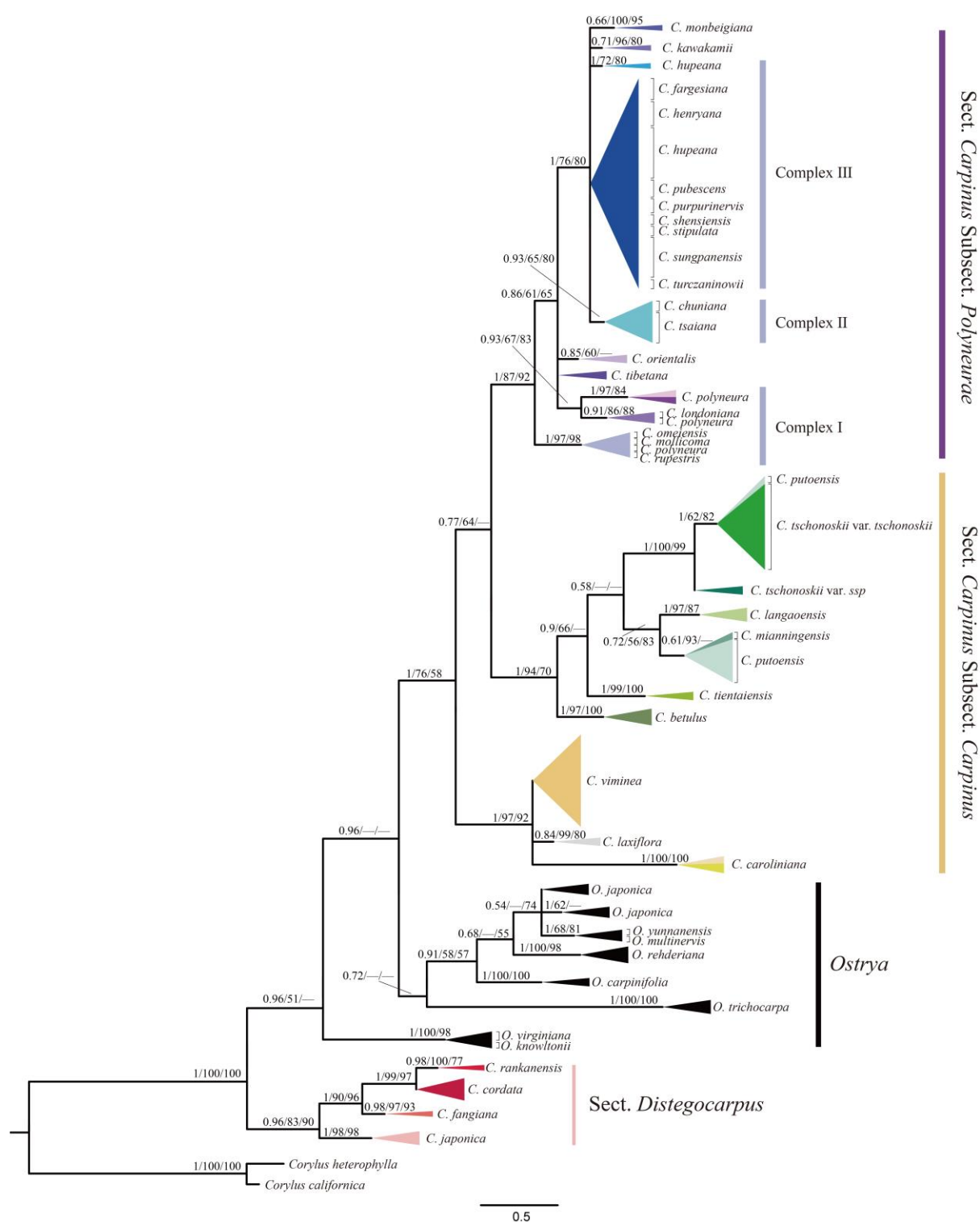


Figure 2

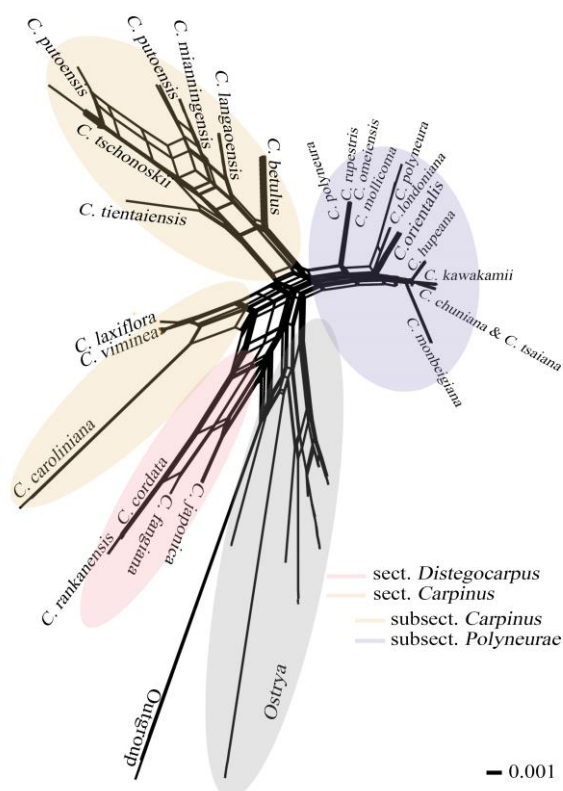


Figure 3

