# Community phylogenetics require phylogenies reconstructed from plastid genomes

Lu Jin[1,2,§], Jia-Jia Liu[2, §], Tian-Wen Xiao[2,4], Qiao-Ming Li[3], Lu-Xiang Lin[3], Xiao-Na Shao[3,4], Chen-Xin Ma[2,4], Bu-Hang Li[5], Xiang-Cheng Mi[6], Xiu-Juan Qiao[7,8], Ju-Yu Lian[9], Gang Hao[1*], Xue-Jun Ge[2*]

[1]*College of Life Sciences, South China Agricultural University, Guangzhou 510642, China*

[2]*Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*

[3]*CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650201, China*

[4]*University of Chinese Academy of Sciences, Beijing 100039, China*

[5]*School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China*

[6]*State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*

[7]*Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China*

[8]*Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan 430074, China*

[9]*Center of Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou 510650, China*

[§:] Lu Jin and Jia-Jia Lu contributed equally to this work.

***Corresponding authors:**

Xue-Jun Ge: xjge@scbg.ac.cn

Gang Hao: haogang@scau.edu.cn

**Running title**: phylogeny and community phylogenetics

## Abstract

Phylogenetic trees have been extensively used in community ecology. However, how the phylogenetic reconstruction affects ecological inferences is poorly understood. In this study, we reconstructed three different types of phylogenetic trees (a synthetic-tree generated using VPhylomaker, a barcode-tree generated using *rbcL+matK+trnH-psbA* and a genome-tree generated from plastid genomes) that represented an increasing level of phylogenetic resolution among 580 woody plant species from six dynamic plots in subtropical evergreen broadleaved forests of China. We then evaluated the performance of each phylogeny in estimations of community phylogenetic structure, turnover and phylogenetic signal in functional traits. As expected, the genome-tree was most resolved and most supported for relationships among species. For local phylogenetic structure, the three trees showed consistent results with Faith's PD and MPD; however, only the synthetic-tree produced significant clustering patterns using MNTD for some plots. For phylogenetic turnover, contrasting results between the molecular trees and the synthetic-tree occurred only with nearest neighbor distance. The barcode-tree agreed more with the genome-tree than the synthetic-tree for both phylogenetic structure and turnover. For functional traits, both the barcode-tree and genome-tree detected phylogenetic signal in maximum height, but only the genome-tree detected signal in leaf width. This is the first study that uses plastid genomes in large-scale community phylogenetics. Our results highlight the outperformance of genome-trees over barcode-trees and synthetic-trees for the analyses studied here. Our results also

point to the possibility of Type I and II errors in estimation of phylogenetic structure and turnover and detection of phylogenetic signal when using synthetic-trees.

**Keywords**: community phylogeny, DNA barcoding, phylogenetic alpha diversity, phylogenetic beta diversity, phylogenetic community structure, phylogenetic signal, plastid genome

# 1 INTRODUCTION

Over the last decade, phylogenetic trees have been increasingly used in community ecology to represent all the component species in a community according to their phylogenetic relationships (Cavender−Bares et al., 2009; Miller et al., 2017; Swenson, 2013; Webb et al., 2002). The underlying core assumption of community phylogenetics is phylogenetic niche conservatism, that is, phylogenetic relationships can be used as a proxy for ecological niche similarity among species, because closely-related species retain ancestral, niche-related, ecological traits over time (i.e., phylogenetic signal in functional traits) (Crisp & Cook, 2012). Based on this assumption, ecologists can infer the patterns of phylogenetic structure in communities by comparing metrics that measure phylogenetic diversity in communities between observed communities and random communities generated by null models (e.g., reshuffled taxa) (Cavender−Bares et al., 2009; Webb et al., 2002). A significant, non-random pattern means that species in a community are more or less similar than expected by chance, which may indicate the role of environmental filtering or competitive exclusion in local community assembly (Cavender−Bares et al., 2009; Webb et al., 2002; but see Mayfield &

Levine, 2010). Phylogenetic signal in functional traits is often quantified simultaneously to investigate the ecological similarity of related species and to determine if phylogenetic relatedness can be used as a proxy for ecological niche similarity (Losos, 2008; Yang et al., 2014). Further, phylogenetic turnover between communities (i.e., phylogenetic beta diversity) may be more informative for inferences of regional processes (e.g., dispersal and speciation) (Graham & Fine, 2008; Swenson, 2011). Despite the insight provided by phylogenetic relationships of species to species coexistence and community assembly processes, the robustness of inferences made from limitations in the reconstruction of phylogenetic trees is poorly known.

Since the pioneering work of Webb (2000) and Webb et al. (2002), synthetic-trees, which are generated by pruning and grafting taxa from an existing supertree (i.e., APG III tree, Angiosperm Phylogeny Group, 2009), often using the online software Phylomatic (Webb & Donoghue, 2005), have become popular among ecologists. Phylomatic is a convenient way to generate phylogenetic trees for species under study in the absence of sufficient, molecular data. Recently, Jin and Qian (2019) developed an R package V.PhyloMaker, which may outcompete Phylomatic by including far more species (74,533 species) and all families of extant vascular plants. Using a dated mega−tree derived primarily from Smith and Brown (2018) and Zanne et al. (2014), V.PhyloMaker can generate phylogenies for very large species lists in a short amount of time. However, synthetic-trees are limited by unresolved nodes (i.e., polytomies), especially at terminal branches (e.g., generic and species level), and inaccurate branch lengths estimation (Beaulieu et al., 2012; Kress et al., 2009; Molina-Venegas & Rodríguez, 2017), which may lead to biased results (i.e., Type I or II error) in community phylogenetic structure, turnover, and trait evolution (Boyle & Adamowicz, 2015; Jantzen et al., 2019; Mazel et al., 2016; Molina-Venegas & Rodríguez, 2017; Swenson, 2009).

In contrast, barcode-trees, which are reconstructed using either single or multilocus regions from the chloroplast and/or nuclear genomes (i.e., DNA barcodes), provide better resolution at the generic or species level and allow direct estimates of branch lengths, resulting in increasing application in community phylogenetics (Erickson et al., 2014a; Kress et al., 2009; Liu et al., 2019; Pei et al., 2011). However, barcode-trees are constrained by a limited number of DNA barcodes (e.g., *rbcL*, *matK*, and *trnH-psbA*), which may lack sufficient genetic variation to fully resolve community phylogenetic relationships, especially for species-rich tropical and subtropical regions (Hollingsworth et al., 2016; Li et al., 2015; Liu et al., 2015). Although several studies have recently indicated that synthetic-trees, especially those from V.PhyloMaker, and barcode-trees may have comparable performance for inferring community phylogenetic structure and phylogenetic signal in traits (Jantzen et al., 2019; Li et al., 2019; Qian & Jin, 2021; Xu et al., 2020), it is unclear whether these methods are robust for communities of species-rich genera.

To better resolve phylogenetic relationships among species from species-rich genera, biologists have recently proposed 'ultra-barcoding' (Kane et al., 2012) or 'super-barcoding' (Li et al., 2015), which represents whole plastid genomes. Genome-trees, which are often based on plastid genomes, have been shown to be more powerful than barcode-trees in resolving intrageneric phylogenetic relationships (e.g., *Taxus,* Fu et al., 2019; *Diospyros*, Li et al., 2018; *Calligonum*, Song et al., 2020; and *Ilex*, Yao et al., 2021) because the plastid genome provides more informative and variable loci (Hollingsworth et al., 2016; Tonti‐Filippini, Nevill, Dixon, & Small, 2017). Furthermore, recent studies (Ahrendsen, Aust, & Roxanne Kellar, 2016; Pischl, Burke, Bach, & Duvall, 2020) have indicated that more robust phylogenies, which have resulted from genomic methods, may provide more reliable phylogenetic diversity (PD) metrics. Nonetheless, genome-trees have rarely been assessed in

the context of community phylogenetics.

Subtropical evergreen broad-leaved forest (EBLF) is a major type of forest across the world (Corlett & Hughes, 2015). Developed under the influence of the East Asian monsoon, which is caused by the uplift of the Qinghai-Tibetan Plateau, subtropical EBLFs inhabit approximately 1/4 of China and have high seed plant diversity (Song, 1988). In these biodiversity hotspots, limited species resolution has been found for closely related species from species-rich genera, such as *Ilex*, *Ficus*, *Viburnum*, and *Rhododendron*, using standard DNA barcodes (Liu et al., 2019, 2015). Plastid genomes may have the power to resolve relationships among species within these complex genera and to provide a robust community phylogeny, which may contribute to more accurate estimates of community phylogenetic structure, turnover, and phylogenetic signal in traits.

Here, we hypothesize that (1) genome-trees outperform barcode-trees and synthetic-trees in resolving closely related species and in detecting phylogenetic signal in community structure, turnover, and functional traits, and (2) barcode-trees will agree more with genome-trees than synthetic-trees in inferences of community phylogenetic structure, turnover, and phylogenetic signal in traits. We tested these hypotheses in six dynamic plots from subtropical EBLFs in China. We reconstructed three different types of phylogenetic trees, representing an increasing level of species resolution, for 580 woody plant species from the dynamic plots. A synthetic-tree was assembled using V.PhyloMaker. A barcode-tree was generated using three standard DNA barcodes: *rbcL*, *matK*, and *trnH-psbA*. A genome-tree was generated using 79 protein-coding genes from the chloroplast genome. For each phylogenetic tree, we estimated local community phylogenetic structure for each plot using three different PD metrics: Faith's PD, mean pairwise distance (MPD), and mean nearest taxon distance (MNTD). We also assessed intercommunity phylogenetic turnover using the mean phylogenetic dissimilarity

between communities based on pairwise species distance (Dpw) and nearest neighbor distance (Dnn). Three functional traits were compiled for target species and analyzed for phylogenetic signal: maximum height (m), leaf length (cm), and leaf width (cm). Non-random patterns were identified for each of the analyses above and then compared among the three different trees. Finally, we regressed phylogenetic structure and turnover resulting from the genome-tree to that resulting from the barcode-tree and the synthetic-tree, respectively, and compared their model performance to determine whether the barcode-tree was more similar to the genome-tree than the synthetic-tree.

## 2 MATERIALS AND METHODS

### 2.1 **Data compilation and sampling**

The community composition data used in this study were compiled from six dynamic plots, which covered the main regions of subtropical EBLFs in China (Fig. 1). These plots were permanent and established following the standard protocol of Center for Tropical Forest Science-Forest Global Earth Observatory (CTFS-ForestGEO) (Condit, 1998). They ranged in area from 20 to 25 ha, in elevation from 350 to 2500 m, and in number of species from 88 to 232 (Table 1). Within each plot, all free-standing, woody individuals with stems ≥ 1 cm diameter at breast height (DBH, 1.3 m) were tagged, mapped, and identified to species. Species names in the dataset were standardized according to The Plant List (www.theplantlist.org). For each species in a plot, voucher specimens were sampled and deposited in the herbarium of South China Botanical Garden (IBSC), and fresh leaf material was sampled from one to two tagged individuals and dried in silica gel for subsequent DNA extraction. For each species, we first searched the National Center for Biotechnology

Information (NCBI) (www.ncbi.nlm.nih.gov) for plastid genome data and downloaded data for 101 species belonging to 66 genera and 38 families (Supplemental File 1). For the remaining species, we extracted, sequenced, assembled, and annotated their genomic DNA (see below). As a result, incorporating with 479 newly generated plastid genomes (GenBank accession MW800906-MW801384), a total of 580 assembled plastid genome sequences representing 580 woody species, 35 orders, 82 families, and 229 genera (Angiosperm Phylogeny Group, 2016; Christenhusz et al., 2011) were included in our final dataset (Supplemental File 1).

**2.2 DNA extraction, sequencing, assembly, and annotation**

Total genomic DNA was extracted from approximately 100 mg of silica gel-dried leaves following the hexadecyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987). The extracted DNA was subsequently sheared to 300-500-bp fragments for library construction. Paired-end (PE) reads (2 × 150 bp) were sequenced on a Illumina HiSeq X Ten platform in Beijing Genomics Institute (BGI, Shenzhen, China). Trimmomatic v0.35 (Bolger, Lohse, & Usadel, 2014) was used to trim adapters and to remove low-quality reads, generating approximately 3 Gb of clean, PE reads per sample. Plastid genomes were assembled from the filtered data using NOVOPlasty v1.1 (Dierckxsens, Mardulyn, & Smits, 2017) and GetOrganelle v1.7.0 (Jin et al., 2020) and were annotated using DOGMA (Wyman, Jansen, & Boore, 2004) and GeSeq (Tillich et al., 2017). Start and stop codons were manually adjusted in Geneious v11.0.2 (Ripma, Simpson, & Hasenstab-Lehman, 2014) when necessary.

**2.3 Phylogenetic tree reconstruction**

Three different types of mega-phylogenies were reconstructed for a dataset of 580 total species sampled in this study. A synthetic-tree was generated using the 'mega-tree' function and scenario 3, as recommended by Qian & Jin (2021), in the R package V.PhyloMaker (Jin & Qian, 2019). This mega-tree is currently the largest, dated phylogenetic tree for vascular plants and includes 74,533 species and all families of extant vascular plants. A barcode-tree was generated using three standard DNA barcode markers (*rbcL*, *matK*, and *trnH-psbA*), which were extracted from assembled plastid genomes using a python script (Jin, 2020). A genome-tree was generated using 79 protein-coding genes (coding sequences [CDS]), which were also extracted from the same assembled plastid genomes as DNA barcodes. For both the barcode-tree and the genome-tree, sequence alignment for each locus was performed using MAFFT v7 (Katoh & Standley, 2013) and then concatenated to build a DNA supermatrix (Supplemental File 2 & 3). After model selection using jModelTest v2.0 (Darriba, Taboada, Doallo, & Posada, 2012), the maximum likelihood (ML) tree was reconstructed for each supermatrix using RAxML v8.2.12 (Stamatakis, 2014) in the CIPRES Science Gateway (Miller et al., 2010) under the GTRGAMMA model. Node support was estimated using bootstrap (BS) values with 1,000 replicates. In addition, divergence time analyses using penalized likelihood method were conducted to infer ultrametric chronograms for both barcode-tree and genome-tree in treePL (Smith & O'Meara, 2012). Secondary calibration points were taken from a dated phylogeny published by Magallón (2015) (Supplemental File 4 & 5).

**2.4 Community phylogenetic structure and turnover**

For each of the three phylogenetic trees reconstructed above, we quantified local community

phylogenetic structure and intercommunity phylogenetic turnover using several metrics that are widely implemented in community phylogenetics. To evaluate community phylogenetic structure for each plot, we used (1) Faith's PD, which sums all branch lengths in a phylogeny that connect species in a community (Faith, 1992); (2) MPD, which is the average pairwise phylogenetic distance for all species in a community (Webb et al., 2002); and (3) MNTD, which is the average shortest phylogenetic distance for each species to its closest relative in an assemblage (Webb et al., 2002). Given that our plots varied markedly in the number of species, we calculated standardized effect size (SES) for each metric (SES.PD, SES.MPD, and SES.MNTD) using a null model, where the tip labels in a phylogenetic tree were randomly shuffled (Swenson, 2014). Here, we reshuffled the taxa 1,000 times and calculated each metric for each random permutation. The SES for each metric was calculated as follows:

$$SES.Metric = \frac{Metric_{observed} - mean(Metric_{rand})}{sd\,Metric_{rand}},$$

where $Metric_{observed}$ is the observed value of a metric, and $mean(Metric_{rand})$ and $sdMetric_{rand}$ are the mean and standard deviation of the random permutations for a metric, respectively. A negative *SES.Metric* indicates phylogenetic clustering, whereas, a positive value indicates phylogenetic overdispersion (Webb et al., 2002).

To assess phylogenetic turnover among plots, we applied two distance-based measurements for phylogenetic dissimilarity between communities: Dpw, which is sensitive to variation at shallow phylogenetic levels, and Dnn, which evaluates deeper turnover in the phylogeny (Swenson, 2011; Webb, Ackerly, & Kembel, 2008). The SES of the two metrics was also calculated to evaluate whether the phylogenetic dissimilarity between two plots differed from randomly expected. Here, the null model and the calculation of the SES were the same as above. MPD, MNTD, Dnn, and Dpw were weighted by abundance. All

phylogenetic metrics were calculated with the R package Picante (Kembel et al., 2010).

### 2.5 Phylogenetic signal in functional traits

To test whether phylogenetic tree reconstruction influences the detection of phylogenetic signal in functional traits, we compiled data for three fundamental traits that are involved in plant ecological strategies for species in our plots from the Flora Reipublicae Popularis Sinicae (Flora Reipublicae Popularis Sinicae Editorial Committee, 1959–2013) and the Flora of China (Wu & Raven, 1994–2011): maximum height (m), leaf length (cm), and leaf width (cm) (Pérez-Harguindeguy et al., 2013; Westoby, Falster, Moles, Vesk, & Wright, 2002). The degree of phylogenetic signal in functional traits was tested with each phylogenetic tree using Blomberg's K (Blomberg, Garland, & Ives, 2003) and Pagel's λ (Pagel, 1999), which are based on the assumption of a Brownian model of trait evolution. For both Blomberg's K and Pagel's λ, values close to 0 indicate no phylogenetic signal, and values close to 1 indicate that closely related species tend to share similar trait values. The statistical significance of both Blomberg's K and Pagel's λ was assessed by randomly shuffling trait distributions across phylogenetic tips 999 times using the 'phylosig' function in the R package phytools (Revell, 2012).

### 2.6 Statistical analyses

To statistically test whether the barcode-tree agreed more with the genome-tree than the synthetic-tree for inferences of community phylogenetic structure and turnover, we regressed the SES of each metric calculated above from the genome-tree to the corresponding value from the synthetic-tree or the barcode-tree with the 'lm' function in the R package stats (R

Core Team, 2020). All regression lines were forced through the origin. Therefore, a regression slope higher or lower than one indicates that the predictor variable tends to under- or over-estimate the response variable (Swenson, 2009). Two linear regression models for each metric were compared and ranked according to the corrected Akaike Information Criterion ($AIC_c$; Burnham & Anderson, 2002, 2004) using the 'model.sel' function in the R package MuMIn (Bartoń, 2020). We also used the percentage of deviance explained in the response variable (DE) as a measure of the goodness-of-fit for each model.

### 2.7 Assumption checks

To test whether our results for the whole community were robust among major clades in the phylogeny, we chose three orders (Aquifoliales, Ericales, and Rosales) containing species-rich genera that are common in subtropical EBLFs (e.g., *Ilex*, *Rhododendron,* and *Ficus*, respectively) and three major clades of vascular plants (angiosperms, core eudicots, and magnoliids). Then, we constructed a series of linear models for each metric, similar as above, except that clade was an additional variable that was considered alone or additively/interactively with the SES of each metric using the synthetic-tree or the barcode-tree. Models were ranked according to the $AIC_c$, and DE was calculated as a model's goodness-of-fit as above.

## 3 RESULTS

### 3.1 Species resolution

We reconstructed three different phylogenetic trees using different types of data (supertrees,

DNA barcodes, and plastid genomes) to assess their performance on species resolution and support. The three phylogenetic trees, the synthetic-tree, the barcode-tree, and the genome-tree, respectively, are shown in Figure 2. At the ordinal level, the topologies of the synthetic-tree and the barcode-tree were nearly identical with that of the genome-tree, except for the position of Crossosomatales in the barcode-tree and Boraginales, Lamiales, Celastrales, and Malpighiales in the synthetic-tree (Fig. S1). In the synthetic-tree, 10.02% of the internal nodes, most of which belonged to complex genera (e.g., *Ilex*, *Acer*, and *Magnolia*), were unresolved, whereas, both the barcode-tree and the genome-tree had well-resolved phylogenetic relationships among all species. The genome-tree was more supported by higher ML bootstrap values than the barcode-tree, especially in species-rich genera (e.g., *Prunus* and *Ficus*) (Fig. 2, Table 2). Specifically, in the genome-tree, 4.32% of nodes had moderate bootstrap support values ($70 < BS \leq 85$), and 83.94% of nodes showed had high bootstrap support values ($BS > 85$); whereas, in the barcode-tree, 6.74% of nodes had moderate bootstrap support values, and 66.49% had high bootstrap support values (Table 2).

### 3.2 Phylogenetic structure, turnover, and signal in functional traits

We calculated a variety of metrics for inferences of community phylogenetic structure, turnover, and phylogenetic signal in functional traits from each tree reconstruction above to assess the performance of tree reconstructions on these methods. We conducted three tests for community phylogenetic structure across all three trees: SES.PD, SES.MPD, and SES.MNTD. For both SES.PD and SES.MPD, all three phylogenetic trees showed generally consistent patterns of local community phylogenetic structure with three plots being significantly clustered. However, only two plots (ALS and BDGS) showed consistent patterns

of phylogenetic structure between the two metrics in all three trees (Fig. 3). For SES.MNTD, only three plots (ALS, BDGS, and GTS) showed consistent patterns of phylogenetic structure in all three trees, and only the synthetic-tree showed significant phylogenetic clustering for the remaining plots (Fig. 3). In the linear models constructed for each metric (including the intercept-only model), the SES of each metric from the barcode-tree vastly outperformed the corresponding metric from the synthetic-tree (Fig. 4, Table S1), implying that the barcode-tree is more similar to the genome-tree than the synthetic-tree for inferences of phylogenetic structure. This result was robust regardless of the major clades used in our study (Table S2). Further, for some clades containing species-rich genera (e.g., Aquifoliales and Rosales), the deviation between the synthetic-tree and the genome-tree was evident, e.g., the slopes and $R^2$ of the regressions were both close to 0 (Fig. S2, Table S3).

We conducted two tests for community phylogenetic turnover across all three trees: SES.Dpw and SES.Dnn. Similar results to the patterns of phylogenetic structure were found for phylogenetic turnover. Generally, all three phylogenetic trees showed consistent patterns of phylogenetic turnover for both SES.Dpw and SES.Dnn. For SES.Dpw with all the three trees reconstructed, four plots (GTS, CBL, BDGS, and ALS) showed significant dissimilarities between each other whereas DHS and BSZ showed no significant dissimilarity between other plots (Fig. 5). For SES.Dnn with all three trees, there was also no significant dissimilarity between DHS and other plots, while significant dissimilarities were found between ALS and the other four plots (GTS, CBL, BDGS, and BSZ), and between BDGS and BSZ. However, only the synthetic-tree showed significant phylogenetic turnover between CBL and BDGS with SES.Dnn (Fig. 5). Meanwhile, the patterns of phylogenetic turnover derived using the barcode-tree were more congruent with those using the genome-tree than those using the synthetic-tree . For both SES.Dpw and SES.Dnn, the barcode-tree displayed

higher $R^2$ (0.99 and 1, respectively) and DE (98.6 and 99.6, respectively) than the synthetic-tree ($R^2$: 0.95 and 0.93, respectively; DE: 95.3 and 93.2, respectively) when regressing against the genome-tree (Fig. 6, Table S4).

We conducted two tests for phylogenetic signal in maximum height, leaf length, and leaf width across all three trees: Blomberg's K and Pagel's λ. For maximum height, Blomberg's K detected significant phylogenetic signal in the barcode-tree and the genome-tree, but not in the synthetic-tree (Table 3). For leaf length, no phylogenetic signal was detected in any of the trees using either Blomberg's K or Pagel's λ (Table 3). For leaf width, significant phylogenetic signal was detected only in the genome-tree using Blomberg's K (Table 3). For all of the three functional traits, the Barcode tree and Genome tree of the three trees reconstructed here obtained the most similar values of K and λ (Table 3).

# 4 DISCUSSION

## 4.1 Greatest species resolution with a genome-tree

Our results demonstrate that plastid genomes, used for super-barcoding, can reconstruct a phylogenetic tree that has greater species resolution than a barcode-tree or a synthetic-tree. It is not surprising that the synthetic-tree performed the worst in resolving closely related species because the genera/species that were present in our plots but absent from the mega-tree were attached to the synthetic-tree as polytomies, at the midpoint of the family/genus branch in scenario 3 of V.PhyloMaker (Qian & Jin, 2020). Meanwhile, the 79 protein-coding genes that were extracted from plastid genomes for the genome-tree may have provided more informative and variable sites than the limited number of standard DNA barcodes used in the barcode-tree, resulting in greater species resolution in the genome-tree and in species-rich

genera. Further, our genome-tree agreed with a mega-phylogeny that was reconstructed from 80 genes from 2,881 plastid genomes (Li et al., 2019), representing 85% of extant angiosperm families and all orders in China; the phylogenetic relationships of all angiosperm orders shared by both trees were nearly identical (Fig. S1). We also noted that the ML bootstrap support values of our barcode-tree (BS > 70: 73.23% and BS > 85: 66.49%) were slightly higher than those of previous studies (BS > 70: 65.3% and BS > 85: 54%, Kress et al., 2009; BS > 70: 60.17% and BS > 85: 53.53%, Liu et al., 2019). This implies that the direct extraction of DNA barcodes from assembled plastid genomes can produce more complete barcode markers than experimental extraction with PCR.

### 4.2 Consistent patterns of phylogenetic structure and turnover among three different phylogenetic trees

The estimation of phylogenetic diversity within and between species assemblages has been increasingly used by ecologists to infer processes in shaping community structure and by conservation biologists to prioritize areas for conservation (Cadotte & Davies, 2010; Miller et al., 2017; Webb et al., 2002). The results of these analyses may be affected by the choice of method used in phylogenetic reconstruction as phylogenetic diversity metrics are calculated based on the topology and branch lengths of phylogenies (Faith, 1992; Webb et al., 2008, 2002). Two common approaches for generating phylogenies for a set of target species are pruning and grafting from a supertree or using DNA barcodes for tree reconstruction (Erickson et al., 2014; Li et al., 2019; Lu et al., 2020). The effects of these approaches on estimating patterns of phylogenetic structure and turnover have been assessed in several studies, and most have indicated that strong correlations of phylogenetic diversity estimates

exist between synthetic-trees and barcode-trees (Boyle & Adamowicz, 2015; Jantzen et al., 2019; Li et al., 2019; Liu et al., 2019; Xu et al., 2021). However, there are still debates about the usage of both approaches because the deviation between the patterns of phylogenetic structure and turnover from a synthetic-tree or a barcode-tree and the corresponding true values is still poorly understood.

This is the first study that explicitly quantifies the deviation between phylogenetic diversity estimates using synthetic-trees and barcode-trees and the most realistic values using a large community phylogeny reconstructed from plastid genomes. First, our results found generally consistent patterns of phylogenetic structure and turnover among the three different phylogenetic reconstructions, especially for SES.MPD and SES.Dpw. Both metrics were calculated based on pairwise phylogenetic distance in a community (Webb et al., 2008, 2002), and Mazel et al. (2016) indicate that MPD may be highly sensitive to deep branching structure. Our results seem to agree with Mazel et al. (2016) since our three phylogenetic trees shared similar topologies at the ordinal level. Second, for SES.MNTD and SES.Dnn, which were calculated based on nearest neighbor distance of species in a community (Webb et al., 2008, 2002), the synthetic-tree rejected null expectations more often than the barcode-tree and the genome-tree, although the three trees generally shared similar patterns. This could be due to MNTD being highly sensitive to terminal branching (Mazel et al., 2016; Tucker et al., 2017). Different scenarios of community phylogenetic structure (clustering or overdispersion) in deep versus terminal nodes in phylogenies may reflect the temporal hierarchy of biogeographic processes (Bose, Ramesh, Pélissier, & Munoz, 2019). Similarly, higher basal turnover would suggest that entire clades track environmental conditions; whereas, higher terminal turnover would suggest that selective pressures promote divergence into habitats, and modern species likely occupy different niches than their ancestors (Jin et al., 2015). Thus,

inaccurate estimates of phylogenetic structure and turnover at shallow evolutionary time-scales in synthetic-trees may lead to erroneous interpretations of ecological and evolutionary processes. Third, our results show that barcode-trees vastly outperform synthetic-trees when using genome-trees as references, although synthetic-trees provide a reasonable approximation of genome-trees for phylogenetic diversity measurements.

These findings may have important implications for selecting tree reconstruction methods in community phylogenetics and biodiversity conservation. On one hand, the use of a synthesic-tree may be sufficient when researchers just focus on the metric values of community phylogenetic diversity or turnover, e.g., the relationship between phylogenetic properties and the environment for macroecological patterns (Qian & Jin, 2020) or the identification of key areas with high PD scores for biodiversity conservation (Forest et al., 2007), rather than their significance. On the other hand, the use of a synthetic-tree may be valid in regions with less congeneric species, where more consistent results can be obtained from phylogenies that are resolved at the species or genus level (Lehtonen et al., 2015; Qian & Jin, 2020). However, we still recommend the use of barcode-trees when genomic data are unavailable, not only because the synthetic-tree often rejected the null expectation (i.e., Type I error), but also because the estimates of community phylogenetic structure and turnover from the barcode-tree were nearly identical to those from the genome-tree for different metrics and clades. Moreover, barcode-trees will be produced with less effort and cost as sequence databases (e.g., NCBI and Barcode of Life Data System [BOLD]) grow with increasingly cheaper and faster sequencing technologies (DeSalle & Goldstein, 2019; Li et al., 2015).

Despite the variation in patterns of phylogenetic structure due to different phylogenetic trees and different metrics, we found consistent patterns of clustering in local community phylogenetic structure for two plots, ALS and BDGS, which were further away from the

418 coastline than other plots and distributed at relatively higher elevations, implying that more
419 environmental stress existed in these plots due to drought and cold (Shepherd & Wynne
420 Griffiths, 2006). However, for patterns of phylogenetic turnover, it is surprising that DHS, a
421 plot closest to the tropics (Fig. 1), showed no significant dissimilarity from other plots using
422 both Dpw and Dnn. A possible explanation may be that the tropics are acting as both a cradle
423 and a museum for biodiversity, with the majority of clades originating and persisting (Bowen,
424 Rocha, Toonen, & Karl, 2013; Cavender−Bares, Gonzalez−Rodriguez, Pahlich, Koehler, &
425 Deacon, 2011; Jablonski, Roy, & Valentine, 2006; Mittelbach et al., 2007); more in-depth
426 explanations will be the object of future study.

427

428 **4.3 Genome-trees have the most potential to detect phylogenetic signal in functional**
429 **traits**

430 The presence of phylogenetic signal in functional traits, which suggests an association
431 between phylogenetic relatedness and interspecific similarities (Blomberg et al., 2003),
432 provides important insights into the ecological and evolutionary processes underlying
433 community structure (Mouquet et al., 2012). Our results showed that the genome-tree had the
434 highest number of functional traits exhibiting phylogenetic signal. This implies that the
435 genome-tree has the most power to detect phylogenetic signal in functional traits among the
436 three trees. However, the limited number of functional traits used in this study may constrain
437 the plausibility of the hypothesis that functional traits show phylogenetic signal when species
438 resolution is better supported. More functional traits are required to confirm this assumption.
439 Moreover, contrasting results for phylogenetic signal in functional traits, i.e., maximum
440 height and leaf width, among the three phylogenetic trees were found only with Blomberg's

K. This suggests that Pagel's λ may be more insensitive to variation in tree topology and branch lengths than Blomberg's K (Boyle & Adamowicz, 2015; Molina-Venegas & Rodríguez, 2017; Münkemüller et al., 2012), and that Blomberg's K may have higher statistical power in identifying phylogenetic signal than Pagel's λ (Li et al., 2019). Davies et al. (2012) concluded that Blomberg's K can inflate estimates of phylogenetic conservatism in highly unresolved trees. Our results partially agree with Davies et al. (2012) because we found that the highest values of both statistics for all traits occurred in the synthetic-tree. Therefore, our study suggests that both Pagel's λ and Blomberg's K should be evaluated together when estimating phylogenetic signal. Second, caution should be taken when using a pruned synthetic-tree as a proxy for phylogenetic relationships to estimate phylogenetic signal in traits because synthetic-trees often falsely accept null expectations (i.e., Type II error).

## 5 CONCLUSIONS

Our study demonstrates the outperformance of genome-trees over synthetic-trees and barcode-trees in systematics and community phylogenetics. Although the patterns of phylogenetic structure and turnover were generally consistent among the three phylogenetic tree reconstructions, our results point to the possibility of Type I error associated with synthetic-trees, even though the mega-tree used was the current, largest dated phylogenetic tree for vascular plants. Second, our results validate the use of barcode-trees when genomic data are not available. On one hand, cost and time can be reduced with a limited number of DNA markers. On the other hand, the results from barcode-trees can be nearly identical to those from genome-trees. Third, the inference of trait evolution strongly depends on phylogenetic reconstruction and test statistic. Caution should be taken when interpreting the

phylogenetic signal of functional traits when using highly unresolved trees and/or a single test statistic.

## ACKNOWLEDGMENTS

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences, Grant No. XDB31000000; the National Natural Science Foundation of China, Grant No. 31500335. The authors acknowledge the Chinese Forest Biodiversity Monitoring Net-work (CForBio) for installing and supporting the forest dynamic plots in China, and all the field technicians who have helped census the plots. They also thank Yu-Ying Zhou, Tong-Jian Liu, Yong Xu, Feng Song and Zhao Nan for their kind help in molecular experiments and data analysis. The authors would like to thank TopEdit for linguistic assistance of this manuscript.

## AUTHORS' CONTRIBUTIONS

XJG and GH conceived and designed the research; QML, LXL, BHL, XCM, XJQ and JYL collected the voucher specimens and fresh leaf materials at the individual forest dynamic plots; JL, TWX, XNS and CXM performed the experiments; JL and JJL analyzed the data; JL and JJL wrote the manuscript, with significant contributions from XJG and GH.

## DATA ACCESSIBILITY

The essential data in the data analysis of this paper is appended.

# REFERENCES


Ahrendsen, D. L., Aust, S. K., & Roxanne Kellar, P. (2016). Biodiversity assessment using next-generation sequencing: Comparison of phylogenetic and functional diversity between Nebraska grasslands. *Plant Systematics and Evolution*, *302*(1), 89–108. doi: 10.1007/s00606-015-1246-6

Angiosperm Phylogeny Group. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, *161*(2), 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Angiosperm Phylogeny Group. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1–20. doi: 10.1111/boj.12385

Bartoń, K. (2020). MuMIn: Multi-Model Inference (Version 1.43.17). Retrieved from https://CRAN.R-project.org/package=MuMIn

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bose, R., Ramesh, B. R., Pélissier, R., & Munoz, F. (2019). Phylogenetic diversity in the Western Ghats biodiversity hotspot reflects environmental filtering and past niche diversification of trees. *Journal of Biogeography*, *46*(1), 145–157. doi: 10.1111/jbi.13464

Bowen, B. W., Rocha, L. A., Toonen, R. J., & Karl, S. A. (2013). The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, *28*(6), 359–366. doi: 10.1016/j.tree.2013.01.018

Boyle, E. E., & Adamowicz, S. J. (2015). Community phylogenetics: Assessing tree reconstruction methods and the utility of DNA barcodes. *PLoS ONE*, *10*(6), e0126662. doi: 10.1371/journal.pone.0126662

Burnham, K. P., & Anderson, D. R. (Eds.). (2002). Model selection and multimodel inference. In *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (pp. 1–48). New York, NY: Springer. doi: 10.1007/978-0-387-22456-5_1

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. doi: 10.1177/0049124104268644

Cadotte, M. W., & Davies, T. J. (2010). Rarest of the rare: Advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Diversity and Distributions*, *16*(3), 376–385. doi: https://doi.org/10.1111/j.1472-4642.2010.00650.x

Cavender−Bares, J., Gonzalez−Rodriguez, A., Pahlich, A., Koehler, K., & Deacon, N. (2011). Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. *Journal of Biogeography*, *38*(5), 962–981. doi: https://doi.org/10.1111/j.1365-2699.2010.02451.x

Cavender−Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, *12*(7), 693–715. doi: 10.1111/j.1461-0248.2009.01314.x

Christenhusz, M. J. M., Reveal, J. L., Farjon, A., Gardner, M. F., Mill, R. R., & Chase, M. W. (2011). A new classification and linear sequence of extant gymnosperms. *Phytotaxa*, *19*(1), 55–70. doi: 10.11646/phytotaxa.19.1.3

Condit, R. (1998). *Tropical forest census plots*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-03664-8

Corlett, R. T., & Hughes, A. C. (2015). *Subtropical Forests*. Routledge Handbooks Online. doi:

10.4324/9781315818290.ch4
Crisp, M. D., & Cook, L. G. (2012). Phylogenetic niche conservatism: What are the underlying evolutionary and ecological causes? *New Phytologist*, *196*(3), 681–694. doi: 10.1111/j.1469-8137.2012.04298.x
Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772–772. doi: 10.1038/nmeth.2109
Davies, T. J., Kraft, N. J. B., Salamin, N., & Wolkovich, E. M. (2012). Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology*, *93*(2), 242–247. doi: 10.1890/11-1360.1
DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, *7*, 302. doi: 10.3389/fevo.2019.00302
Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. doi: 10.1093/nar/gkw955
Doyle, Jeff. J., & Doyle, Jane. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.
Erickson, D. L., Jones, F. A., Swenson, N. G., Pei, N., Bourg, N. A., Chen, W., … Kress, W. J. (2014a). Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: A mega-phylogeny approach. *Frontiers in Genetics*, *5*. doi: 10.3389/fgene.2014.00358
Erickson, D. L., Jones, F. A., Swenson, N. G., Pei, N., Bourg, N. A., Chen, W., … Kress, W. J. (2014b). Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: A mega-phylogeny approach. *Frontiers in Genetics*, *5*, 358. doi: 10.3389/fgene.2014.00358
Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. doi: 10.1016/0006-3207(92)91201-3
Flora Reipublicae Popularis Sinicae Editorial Committee. (1959–2013). *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press.
Forest, F., Grenyer, R., Rouget, M., Davies, T. J., Cowling, R. M., Faith, D. P., … Savolainen, V. (2007). Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, *445*(7129), 757–760. doi: 10.1038/nature05587
Fu, C.-N., Wu, C.-S., Ye, L.-J., Mo, Z.-Q., Liu, J., Chang, Y.-W., … Gao, L.-M. (2019). Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Scientific Reports*, *9*(1), 2773. doi: 10.1038/s41598-019-39161-x
Graham, C. H., & Fine, P. V. A. (2008). Phylogenetic beta diversity: Linking ecological and evolutionary processes across space in time. *Ecology Letters*, *11*(12), 1265–1277. doi: 10.1111/j.1461-0248.2008.01256.x
Hollingsworth, P. M., Li, D.-Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: From barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 20150338. doi: 10.1098/rstb.2015.0338
Jablonski, D., Roy, K., & Valentine, J. W. (2006). Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science (New York, N.Y.)*, *314*(5796), 102–106. doi: 10.1126/science.1130880
Jantzen, J. R., Whitten, W. M., Neubig, K. M., Majure, L. C., Soltis, D. E., & Soltis, P. S. (2019). Effects of taxon sampling and tree reconstruction methods on phylodiversity metrics. *Ecology and Evolution*, *9*(17), 9479–9499. doi: 10.1002/ece3.5425
Jin, J.-J. (2020). *Kinggerm/PersonalUtilities* [Python]. Retrieved from https://github.com/Kinggerm/PersonalUtilities (Original work published 2016)
Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., & Li, D.-Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*(1), 241. doi: 10.1186/s13059-020-02154-5
Jin, L. S., Cadotte, M. W., & Fortin, M.-J. (2015). Phylogenetic turnover patterns consistent with niche conservatism in montane plant species. *Journal of Ecology*, *103*(3), 742–749. doi:

https://doi.org/10.1111/1365-2745.12385
Jin, Y., & Qian, H. (2019). V.PhyloMaker: An R package that can generate very large phylogenies for vascular plants. *Ecography*, *42*(8), 1353–1359. doi: 10.1111/ecog.04434
Kane, N., Sveinsson, S., Dempewolf, H., Yang, J. Y., Zhang, D., Engels, J. M. M., & Cronk, Q. (2012). Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, *99*(2), 320–329. doi: 10.3732/ajb.1100570
Katoh, K., & Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. doi: 10.1093/molbev/mst010
Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., … Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. doi: 10.1093/bioinformatics/btq166
Kress, W. J., Erickson, D. L., Jones, F. A., Swenson, N. G., Perez, R., Sanjur, O., & Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*, *106*(44), 18621–18626. doi: 10.1073/pnas.0909820106
Li, D., Trotta, L., Marx, H. E., Allen, J. M., Sun, M., Soltis, D. E., … Baiser, B. (2019). For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*, *100*(9), e02788. doi: 10.1002/ecy.2788
Li, H.-T., Yi, T.-S., Gao, L.-M., Ma, P.-F., Zhang, T., Yang, J.-B., … Li, D.-Z. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*, *5*(5), 461–470. doi: 10.1038/s41477-019-0421-0
Li, W., Liu, Y., Yang, Y., Xie, X., Lu, Y., Yang, Z., … Suo, Z. (2018). Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. *BMC Plant Biology*, *18*(1), 210. doi: 10.1186/s12870-018-1421-3
Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. (2015). Plant DNA barcoding: From gene to genome. *Biological Reviews of the Cambridge Philosophical Society*, *90*(1), 157–166. doi: 10.1111/brv.12104
Liu, J., Liu, J.-J., Ge, X.-J., Huang, G.-M., Zhou, Z.-L., & Wan, S.-Z. (2019). The use of DNA barcoding to assess phylogenetic β-diversity in mid-subtropical evergreen broad-leaved forests of china. *Forests*, *10*(10), 923. doi: 10.3390/f10100923
Liu, J., Yan, H.-F., Newmaster, S. G., Pei, N., Ragupathy, S., & Ge, X.-J. (2015). The use of DNA barcoding as a tool for the conservation biogeography of subtropical forests in China. *Diversity and Distributions*, *21*(2), 188–199. doi: 10.1111/ddi.12276
Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, *11*(10), 995–1003. doi: 10.1111/j.1461-0248.2008.01229.x
Lu, L., Hu, H., Peng, D., Liu, B., Ye, J., Yang, T., … Chen, Z. (2020). Noise does not equal bias in assessing the evolutionary history of the angiosperm flora of China: A response to Qian (2019). *Journal of Biogeography*, jbi.13947. doi: 10.1111/jbi.13947
Magallón, S., Gómez−Acevedo, S., Sánchez−Reyes, L. L., & Hernández−Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, *207*(2), 437–453. doi: 10.1111/nph.13264
Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, *13*(9), 1085–1093. doi: 10.1111/j.1461-0248.2010.01509.x
Mazel, F., Davies, T. J., Gallien, L., Renaud, J., Groussin, M., Münkemüller, T., & Thuiller, W. (2016). Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography*, *39*(10), 913–920. doi: 10.1111/ecog.01694
Miller, E. T., Farine, D. R., & Trisos, C. H. (2017). Phylogenetic community structure metrics and null models: A review with new methods and software. *Ecography*, *40*(4), 461–477. doi: 10.1111/ecog.02070

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. New Orleans, LA, USA: IEEE. doi: 10.1109/GCE.2010.5676129

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., … Turelli, M. (2007). Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters*, *10*(4), 315–331. doi: https://doi.org/10.1111/j.1461-0248.2007.01020.x

Molina-Venegas, R., & Rodríguez, M. Á. (2017). Revisiting phylogenetic signal; strong or negligible impacts of polytomies and branch length information? *BMC Evolutionary Biology*, *17*. doi: 10.1186/s12862-017-0898-y

Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., … Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, *87*(4), 769–785. doi: 10.1111/j.1469-185X.2012.00224.x

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756. doi: 10.1111/j.2041-210X.2012.00196.x

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. doi: 10.1038/44766

Pei, N., Lian, J.-Y., Erickson, D. L., Swenson, N. G., Kress, W. J., Ye, W.-H., & Ge, X.-J. (2011). Exploring tree-habitat associations in a Chinese subtropical forest plot using a molecular phylogeny generated from dna barcode loci. *PLoS ONE*, *6*(6), e21273. doi: 10.1371/journal.pone.0021273

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., … Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, *61*(3), 167. doi: 10.1071/BT12225

Pischl, P. H., Burke, S. V., Bach, E. M., & Duvall, M. R. (2020). Plastome phylogenomics and phylogenetic diversity of endangered and threatened grassland species (Poaceae) in a North American tallgrass prairie. *Ecology and Evolution*, *10*(14), 7602–7615. doi: 10.1002/ece3.6484

Qian, H., & Jin, Y. (2021). Are phylogenies resolved at the genus level appropriate for studies on phylogenetic structure of species assemblages? *Plant Diversity*. doi: 10.1016/j.pld.2020.11.005

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/. *Available*.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Ripma, L. A., Simpson, M. G., & Hasenstab-Lehman, K. (2014). Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in Oreocarya (Boraginaceae)1. *Applications in Plant Sciences*, *2*(12). doi: 10.3732/apps.1400062

Shepherd, T., & Wynne Griffiths, D. (2006). The effects of stress on plant cuticular waxes. *The New Phytologist*, *171*(3), 469–499. doi: 10.1111/j.1469-8137.2006.01826.x

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314. doi: 10.1002/ajb2.1019

Smith, S. A., & O'Meara, B. C. (2012). treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. doi: 10.1093/bioinformatics/bts492

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

Song, Y. (1988). The essential characteristics and main types of the broad-leaved evergreen forest in China. *Phytocoenologia*, 105–123. doi: 10.1127/phyto/16/1988/105

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Swenson, N. G. (2009). Phylogenetic Resolution and Quantifying the Phylogenetic Diversity and

Dispersion of Communities. *PLoS ONE*, *4*(2). doi: 10.1371/journal.pone.0004390
Swenson, N. G. (2011). Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *PLoS ONE*, *6*(6), e21264. doi: 10.1371/journal.pone.0021264
Swenson, N. G. (2013). The assembly of tropical tree communities – the advances and shortcomings of phylogenetic and functional trait analyses. *Ecography*, *36*(3), 264–276. doi: 10.1111/j.1600-0587.2012.00121.x
Swenson, N. G. (2014). *Functional and phylogenetic ecology in R*. New York, NY: Springer New York. doi: 10.1007/978-1-4614-9542-0
Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*(W1), W6–W11. doi: 10.1093/nar/gkx391
Tonti−Filippini, J., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes? *The Plant Journal*, *90*(4), 808–818. doi: 10.1111/tpj.13491
Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., … Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews of the Cambridge Philosophical Society*, *92*(2), 698–715. doi: 10.1111/brv.12252
Webb, C. O. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *The American Naturalist*, *156*(2), 145–155. doi: 10.1086/303378
Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics (Oxford, England)*, *24*(18), 2098–2100. doi: 10.1093/bioinformatics/btn358
Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505. doi: 10.1146/annurev.ecolsys.33.010802.150448
Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology Notes*, *5*(1), 181–183. doi: 10.1111/j.1471-8286.2004.00829.x
Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A., & Wright, I. J. (2002). Plant ecological strategies: Some leading dimensions of variation between species. *Annual Review of Ecology and Systematics*, *33*(1), 125–159. doi: 10.1146/annurev.ecolsys.33.010802.150452
Wu, Z. Y., & Raven, P. H. (1994–2011). *Flora of China*. Beijing/St. Louis: Science Press/Missouri Botanical Garden Press.
Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, *20*(17), 3252–3255. doi: 10.1093/bioinformatics/bth352
Xu, Y., Liu, J.-J., Li, H.-N., Liu, J., Burgess, K. S., & Ge, X.-J. (2020). The effects of evolutionary and environmental variance on estimates of phylogenetic diversity in temperate forest plots. *Journal of Plant Ecology*, (rtaa078). doi: 10.1093/jpe/rtaa078
Yang, J., Zhang, G., Ci, X., Swenson, N. G., Cao, M., Sha, L., … Lin, L. (2014). Functional and phylogenetic assembly in a Chinese tropical tree community across size classes, spatial scales and habitats. *Functional Ecology*, *28*(2), 520–529. doi: 10.1111/1365-2435.12176
Yao, X., Song, Y., Yang, J.-B., Tan, Y.-H., & Corlett, R. T. (2021). Phylogeny and biogeography of the hollies (*Ilex* L., Aquifoliaceae). *Journal of Systematics and Evolution*, *59*(1), 73–82. doi: https://doi.org/10.1111/jse.12567
Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., … Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. doi: 10.1038/nature12872

## Tables and Figures

**Table 1.** Six dynamic plots in the subtropical evergreen broadleaved forests of China used in this study.

| Plot | Size (ha) | Elevation (m) | Species | Genera | Families | Coordinates |
|---|---|---|---|---|---|---|
| Ailaoshan (ALS)* | 20 | 2500 | 63 | 41 | 24 | 24.54N, 101.03E |
| Badagongshan (BDGS) | 25 | 1401 | 194 | 97 | 55 | 29.77N, 110.09E |
| Baishanzu (BSZ) | 24 | 1527 | 135 | 78 | 41 | 27.76N, 119.20E |
| Chebaling (CBL) | 25 | 793 | 176 | 109 | 59 | 24.43N, 114.15E |
| Dinghushan (DHS) | 20 | 350 | 131 | 85 | 47 | 23.17N, 112.51E |
| Gutianshan (GTS) | 20 | 581 | 136 | 89 | 52 | 29.25N, 118.12E |

*Consisted of 20 1-ha separated dynamics plots.

745 **Table 2.** Comparison of ML bootstrap support values between barcode-tree and genome-tree
746 reconstructed for 580 species of woody plants across six dynamic plots in the subtropical
747 evergreen broadleaved forests of China.

| Clades (n= ) | Phylogeny | Resolution | | | |
|---|---|---|---|---|---|
| | | 0~50% | 51%~70% | 71%~85% | 86%~100% |
| All (580) | barcode-tree | 16.23 | 10.19 | 6.74 | 66.49 |
| | genome-tree | 7.08 | 4.32 | 4.32 | 83.94 |
| *Ilex* (31) | barcode-tree | 43.33 | 6.67 | 3.33 | 46.67 |
| | genome-tree | 20 | 13.33 | 13.33 | 53.33 |
| *Acer* (15) | barcode-tree | 14.29 | 14.29 | 21.43 | 50 |
| | genome-tree | 0 | 21.43 | 14.29 | 64.29 |
| *Prunus* (14) | barcode-tree | 23.08 | 2 | 15.38 | 61.54 |
| | genome-tree | 0 | 0 | 0 | 100 |
| *Ficus* (12) | barcode-tree | 36.36 | 36.36 | 0 | 27.27 |
| | genome-tree | 9.09 | 9.09 | 27.27 | 54.55 |

748 *Notes:* the barcode-tree and genome-tree were reconstructed using three standard DNA
749 barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes,
750 respectively. Some species-rich genera (*Ilex*, *Acer*, *Pruns*, *Ficus*) are also shown in the table,
751 with the number of nodes per genus given in parentheses after the genus name. Values shown
752 are the percentage of nodes that exhibit a particular range of ML bootstrap level.

**Table 3.** Phylogenetic signal tests of functional traits for 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China with three phylogenetic trees.

| Phylogeny | Maximum Height (m) | | Leaf Length (cm) | | Leaf Width (cm) | |
|---|---|---|---|---|---|---|
| | Blomberg's K | Pagel's λ | Blomberg's K | Pagel's λ | Blomberg's K | Pagel's λ |
| genome-tree | 0.007** | 0.211* | 0.003 | 0.067 | 0.005* | 0.255*** |
| barcode-tree | 0.011** | 0.216* | 0.004 | 0.086 | 0.004 | 0.225*** |
| synthetic-tree | 0.010 | 0.281* | 0.018 | 0.122 | 0.021 | 0.307*** |

*Notes:* the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. For both Blomberg's K and Pagel's λ, value close to 0 indicates no phylogenetic signal, and value close to 1 indicates that closely-related species tend to share similar trait value. Significance levels are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Fig 1.** The distribution of the six dynamic plots in the subtropical evergreen broadleaved forests of China used in this study**.** The map was generated using ArcGIS 10.1.

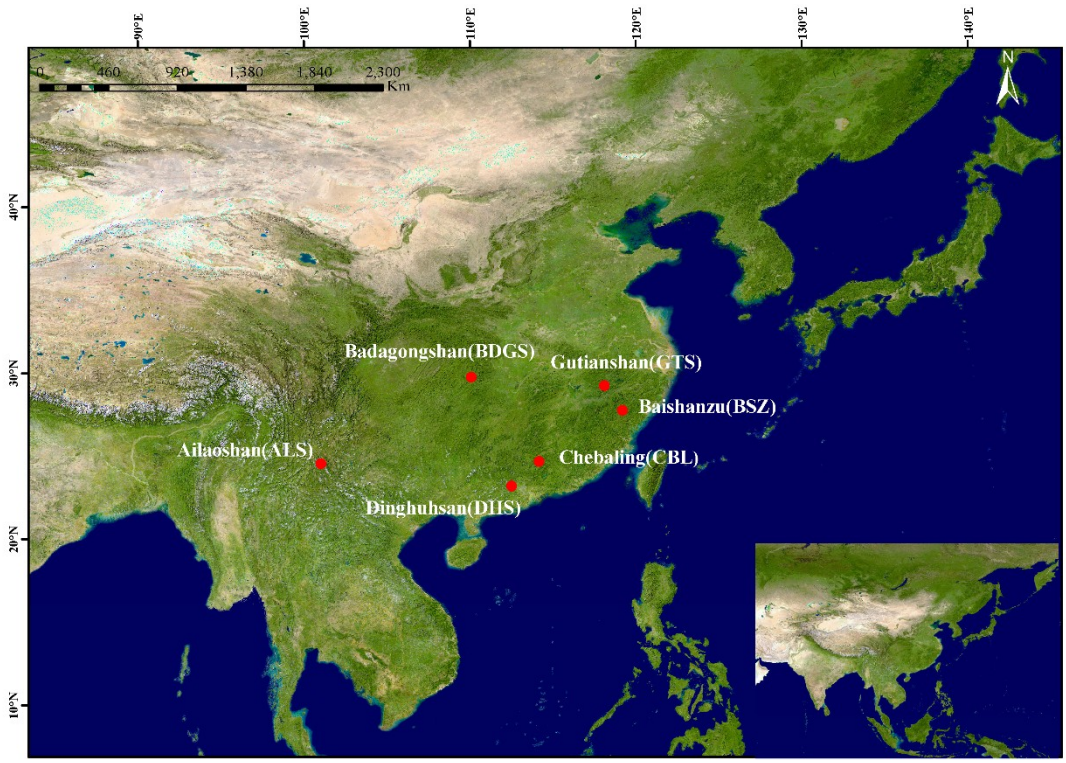


**Fig. 2.** Three different type of phylogenies reconstructed for 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China. (a) A synthetic-tree was generated using R package V.PhyloMaker, the barcode-tree (b) and genome-tree (c) were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Colored bars around a phylogeny represent seven major taxonomic groups (Gymnosperms, Basal angiosperms, Magnoliids, Monocots, Basal eudicots, Superrosids, and Superasterids). Different colors of the dots at the branches of the barcode-tree and genome-tree represent the different levels of bootstrap values.

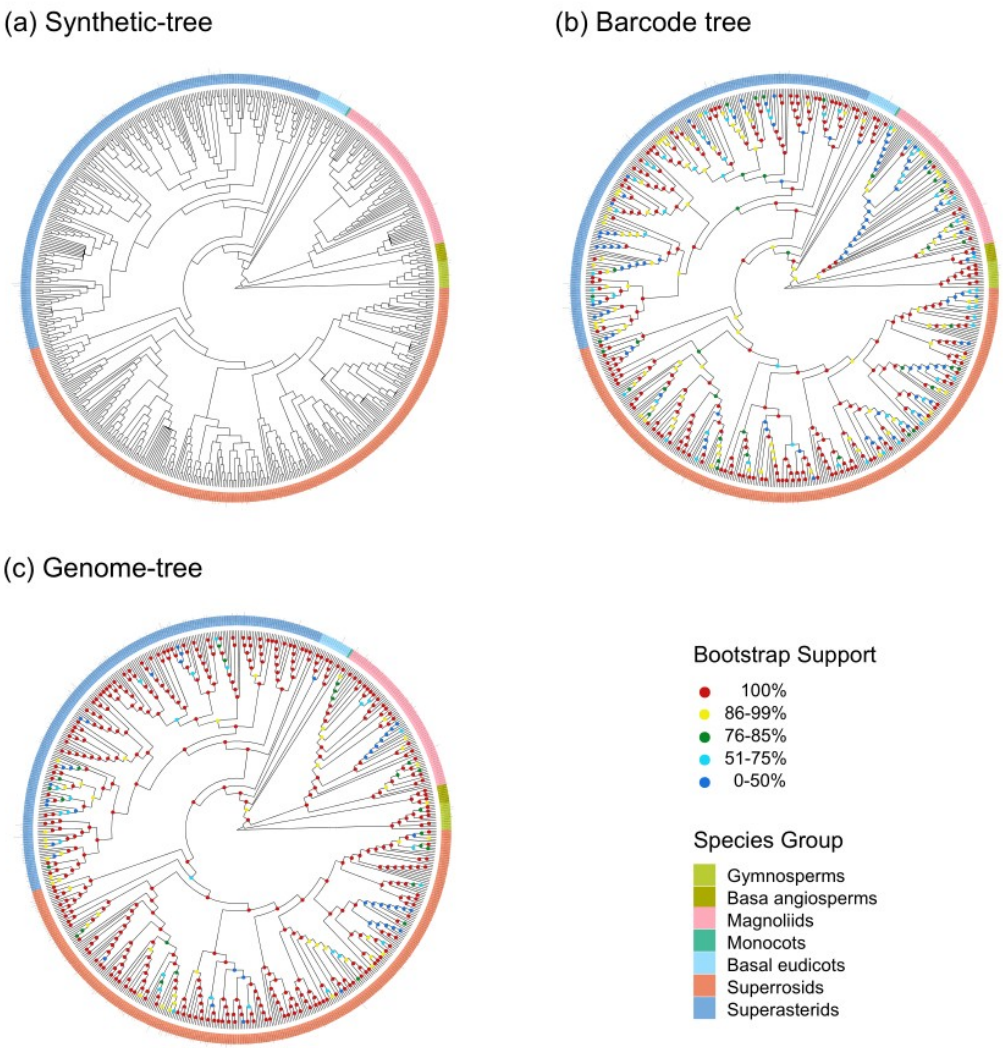


**Fig. 3.** Patterns of standardized phylogenetic community structure of each dynamic plot calculated using synthetic-tree, barcode-tree and genome-tree. For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Negative values indicate phylogenetic clustering, and positive values indicate phylogenetic overdispersion. Significance levels are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Detailed information on SES.PD, SES.MPD and SES.MNTD see Materials and Methods.

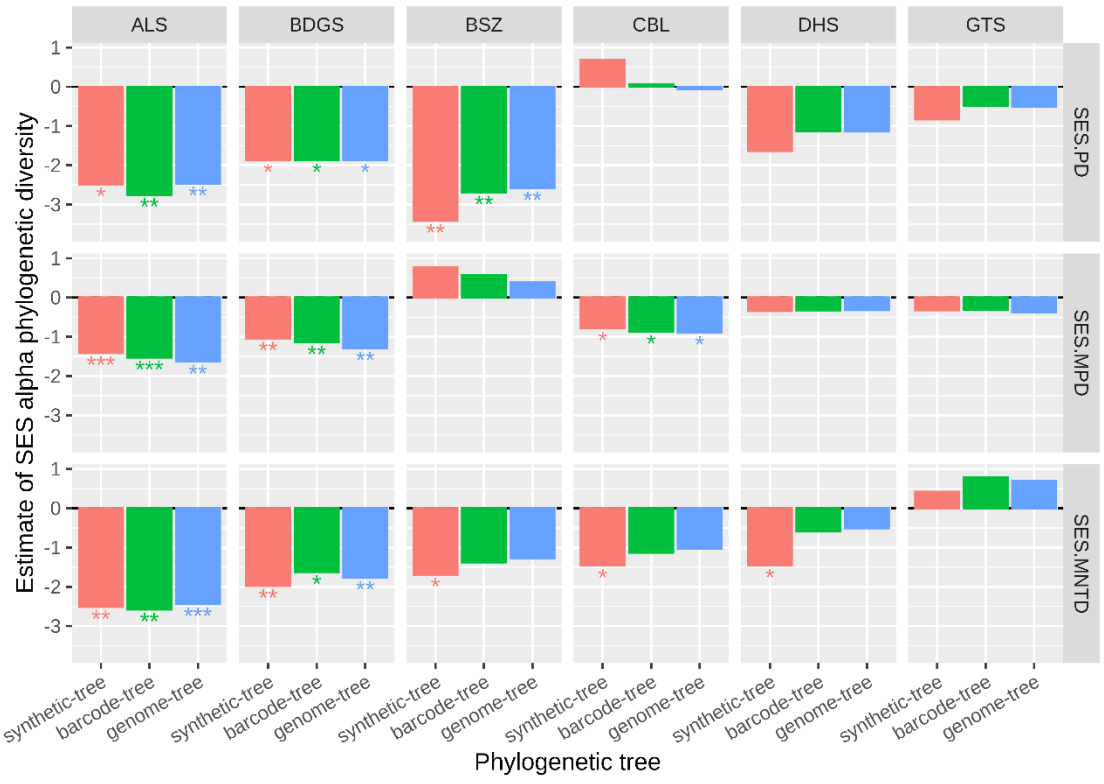


**Fig. 4.** Linear regression of standardized community phylogenetic structure calculated using genome-tree against those based on synthetic-tree (left) and barcode-tree (right). For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively.

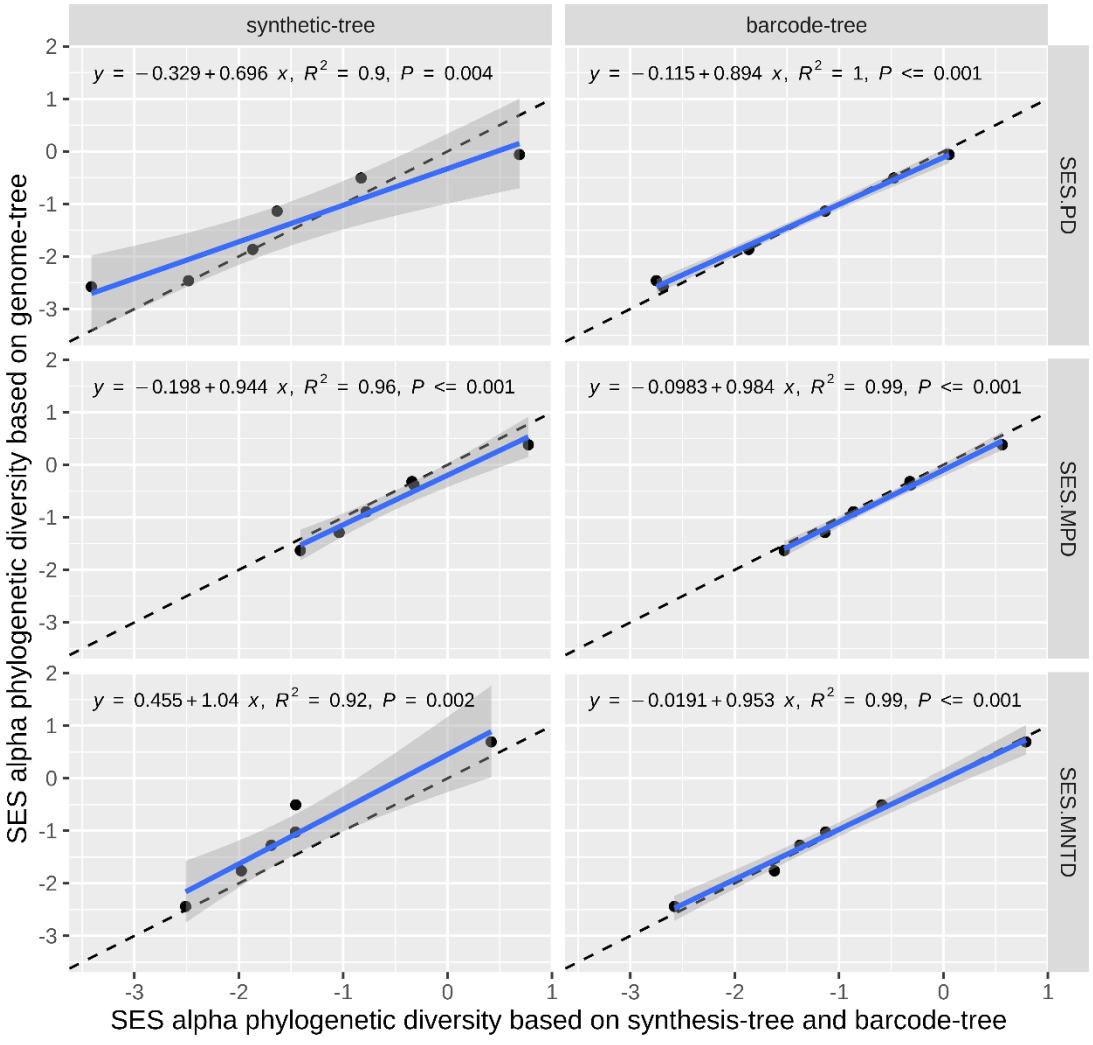


**Fig. 5.** Heatmaps and cluster dendrograms (UPGMA) based on standardized phylogenetic turnover (SES.Dpw and SES.Dnn) calculated with different phylogenies. For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Tiles with warm tones indicate high turnover between plots; cool tones indicate lowturnover between plots. Tiles with stars show pair of plots with higher or lower turnover than expected by chance (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

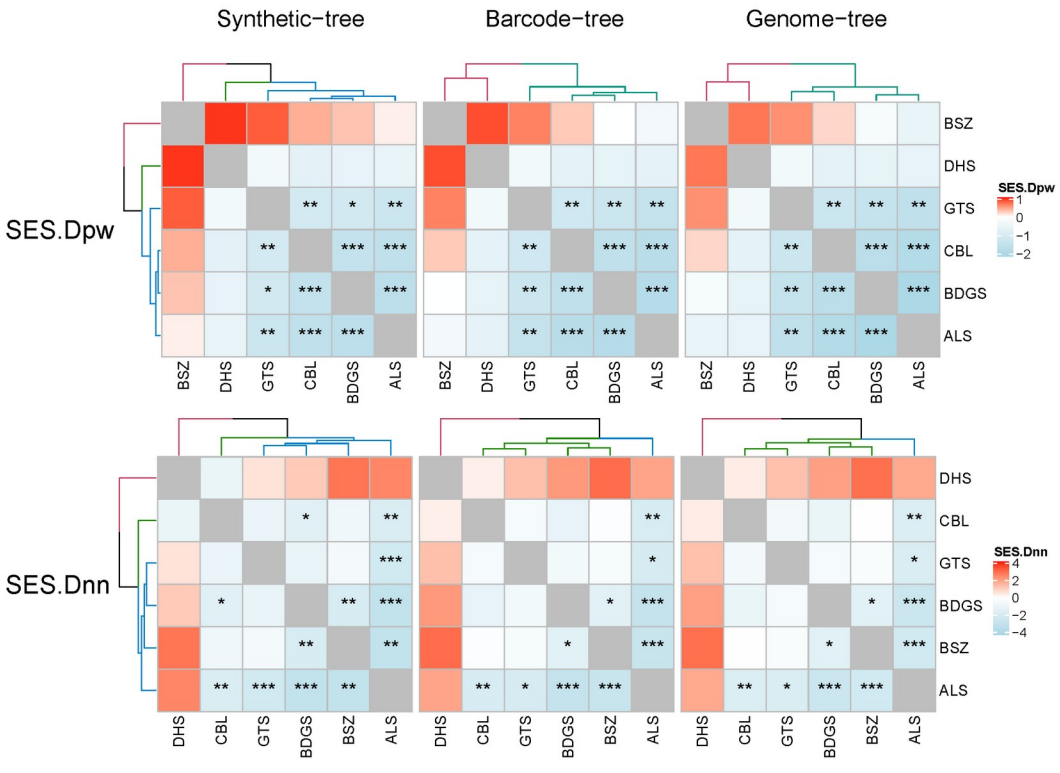


**Fig. 6.** Linear regression of standardized phylogenetic turnover (SES.Dpw and SES.Dnn) calculated using genome-tree against those based synthetic-tree (red) and barcode-tree (blue). For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively.

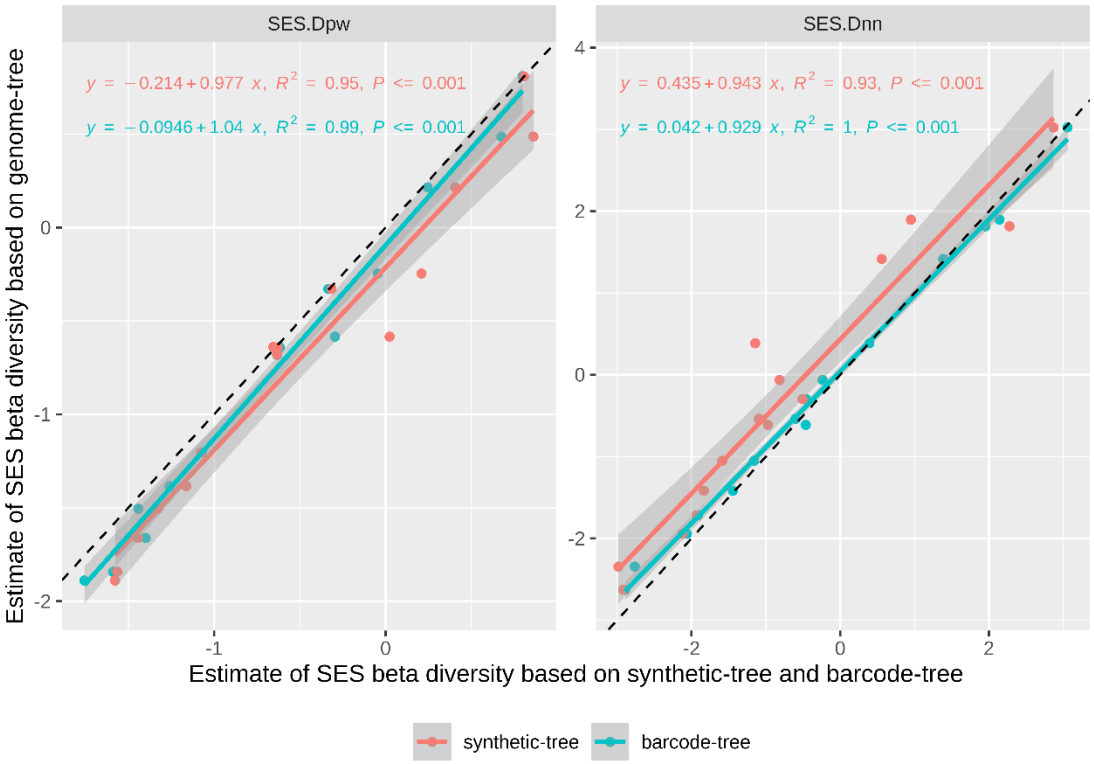


**Table S1.** Linear regression models of standardized phylogenetic structure calculated using genome-tree of a function of those based on synthetic-tree and barcode-tree**.**

| Model | df | logLik | AICc | wAICc | weight | DE |
|---|---|---|---|---|---|---|
| SES.PD | | | | | | |
| ~barcode-tree | 3 | 7.962 | 2.076 | 0.000 | 0.999 | 99.5 |
| ~synthetic-tree | 3 | -1.334 | 20.668 | 18.593 | <0.001 | 89.8 |
| ~1 | 2 | -8.189 | 24.378 | 22.303 | <0.001 | 0.0 |
| SES.MPD | | | | | | |
| ~barcode-tree | 3 | 8.020 | 1.961 | 0.000 | 0.986 | 99.1 |
| ~synthetic-tree | 3 | 3.734 | 10.531 | 8.570 | 0.014 | 96.2 |
| ~1 | 2 | -6.079 | 20.157 | 18.196 | <0.001 | 0.0 |
| SES.MNTD | | | | | | |
| ~barcode-tree | 3 | 5.464 | 7.072 | 0.000 | 0.998 | 99.0 |
| ~synthetic-tree | 3 | -0.767 | 19.534 | 12.463 | 0.002 | 92.2 |
| ~1 | 2 | -8.427 | 24.853 | 17.782 | <0.001 | 0.0 |

*Notes:* For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Detailed information on SES.PD, SES.MPD and SES.MNTD see Materials and Methods. Values are shown for the estimated number of model parameters (df), maximum log-likelihood (logLik), the information-theoretic Akaike's information criterion corrected for small samples (AICc), AICc weight (wAICc, model probability), and the percentage of deviance explained (DE) as a measure of the model's goodness-of-fit.

**Table S2.** Linear models of standardized community phylogenetic structure calculated using genome-tree as a function of those based on synthetic-tree and barcode-tree and major clades.

| Model | df | logLik | AICc | wAICc | weight | DE |
|---|---|---|---|---|---|---|
| SES.PD | | | | | | |
| ~ barcode-tree | 3 | -6.051 | 18.852 | 0.000 | 0.997 | 96.2 |
| ~ barcode-tree + clades | 8 | -5.03 | 31.393 | 12.541 | 0.002 | 96.4 |
| ~ barcode-tree * clades | 13 | 4.59 | 33.366 | 14.514 | 0.001 | 97.9 |
| ~ synthetic-tree | 3 | -34.94 | 76.629 | 57.777 | <0.001 | 81.2 |
| ~ 1 | 2 | -65.022 | 134.408 | 115.556 | <0.001 | 0.0 |
| ~ clades | 7 | -59.611 | 137.221 | 118.369 | <0.001 | 26.0 |
| SES.MPD | | | | | | |
| ~ barcode-tree * clades | 13 | 6.295 | 29.955 | 0.000 | 0.982 | 99.4 |
| ~ barcode-tree | 3 | -15.811 | 38.371 | 8.417 | 0.015 | 97.9 |
| ~ barcode-tree + clades | 8 | -9.947 | 41.228 | 11.273 | 0.003 | 98.5 |
| ~ synthetic-tree | 3 | -39.109 | 84.969 | 55.014 | <0.001 | 92.3 |
| ~ 1 | 7 | -77.973 | 173.946 | 143.992 | <0.001 | 33.6 |
| ~ clades | 2 | -85.343 | 175.049 | 145.094 | <0.001 | 0.0 |
| SES.MNTD | | | | | | |
| ~ barcode-tree | 3 | 4.023 | -1.296 | 0.000 | 0.975 | 95.4 |
| ~ barcode-tree + clades | 8 | 7.652 | 6.030 | 7.326 | 0.025 | 96.3 |
| ~ barcode-tree * clades | 13 | 8.510 | 25.526 | 26.821 | <0.001 | 96.5 |
| ~ synthetic-tree | 3 | -35.971 | 78.693 | 79.989 | <0.001 | 58.0 |
| ~ 1 | 2 | -51.585 | 107.534 | 108.83 | <0.001 | 0.0 |
| ~ clades | 7 | -46.961 | 111.921 | 113.217 | <0.001 | 22.7 |

*Notes:* For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Clades included three orders (Aquifoliales, Ericales and Rosales) containing species-rich genera that are dominated

in subtropical EBLFs (e.g., *Ilex*, *Rhododendron* and *Ficus*) and three major phylogenetic groups of vascular plants (Angiosperms, Core Eudicots and Magnoliids). Detailed information on SES.PD, SES.MPD and SES.MNTD see Materials and Methods. Values are shown for the estimated number of model parameters (df), maximum log-likelihood (logLik), the information-theoretic Akaike's information criterion corrected for small samples (AICc), AICc weight (wAICc, model probability), and the percentage of deviance explained (DE) as a measure of the model's goodness-of-fit.

**Table S3.** Linear regression of standardized community phylogenetic structure calculated using genome-tree against those based on synthetic-tree and barcode-tree across different clades.

| clades | model | SES.PD | | SES.MPD | | SES.MNTD | |
|---|---|---|---|---|---|---|---|
| | | Slope | $R^2$ | Slope | $R^2$ | Slope | $R^2$ |
| Angiosperms | ~ synthetic-tree | 0.876 | 0.947 | 1.196 | 0.998 | 1.026 | 0.893 |
| | ~ barcode-tree | 0.891 | 0.999 | 1.208 | 0.995 | 0.919 | 0.991 |
| CoreEudicots | ~ synthetic-tree | 0.939 | 0.987 | 1.102 | 0.947 | 1.065 | 0.855 |
| | ~ barcode-tree | 0.940 | 0.989 | 0.966 | 0.993 | 0.976 | 0.995 |
| Magnoliids | ~ synthetic-tree | 0.677 | 0.215 | 0.979 | 0.982 | 1.047 | 0.569 |
| | ~ barcode-tree | 0.482 | 0.507 | 0.917 | 0.977 | 0.980 | 0.919 |
| Ericales | ~ synthetic-tree | 0.784 | 0.914 | 1.232 | 0.888 | 1.268 | 0.958 |
| | ~ barcode-tree | 0.885 | 0.977 | 1.030 | 0.962 | 1.004 | 0.984 |
| Aquifoliales | ~ synthetic-tree | 0.338 | 0.153 | -0.190 | 0.242 | -0.026 | 0.023 |
| | ~ barcode-tree | 1.035 | 1.000 | 0.825 | 0.667 | 0.542 | 0.471 |
| Rosales | ~ synthetic-tree | 1.191 | 0.697 | 0.577 | 0.561 | -0.041 | 0.006 |
| | ~ barcode-tree | 0.980 | 0.996 | 0.878 | 0.986 | 1.132 | 0.984 |

*Notes:* For the clades, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Detailed information on SES.PD, SES.MPD and SES.MNTD see Materials and Methods. Values are shown for the slope of the linear regression and proportion of variance explained ($R^2$).

**Table S4.** Linear regression models of standardized phylogenetic turnover between plots calculated using genome-tree against those based on synthetic-tree and barcode-tree.

| Model | df | logLik | AICc | wAICc | weight | DE |
|---|---|---|---|---|---|---|
| SES.Dpw | | | | | | |
| ~ barcode-tree | 3 | 14.088 | -19.994 | 0.000 | 0.999 | 98.6 |
| ~ synthetic-tree | 3 | 4.825 | -1.468 | 18.526 | <0.001 | 95.3 |
| ~ 1 | 2 | -18.151 | 41.302 | 61.296 | <0.001 | 0.0 |
| SES.Dnn | | | | | | |
| ~ barcode-tree | 3 | 12.198 | -16.214 | 0.000 | 0.999 | 99.6 |
| ~ synthetic-tree | 3 | -8.444 | 25.071 | 41.285 | <0.001 | 93.2 |
| ~ 1 | 2 | -28.66 | 62.319 | 78.533 | <0.001 | 0.0 |

*Notes:* For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Detailed information on SES.Dpw, SES.Dnn see Materials and Methods. Values are shown for the estimated number of model parameters (df), maximum log-likelihood (logLik), the information-theoretic Akaike's information criterion corrected for small samples (AICc), AICc weight (wAICc, model probability), and the percentage of deviance explained (DE) as a measure of the model's goodness-of-fit.

## Figure legend

**Fig S1.** Comparison of ordinal-level phylogenetic relationships among (a) synthetic-tree (generated by R package V.PhyloMaker), (b) genome-tree (based on 79-CDS-gene matrix) and (c) barcode-tree (based on *rbcL* + *matK* + *trnH-psbA* matrix) generated from the ML analysis. Branch colors correspond to major groups, as in Fig. 2.

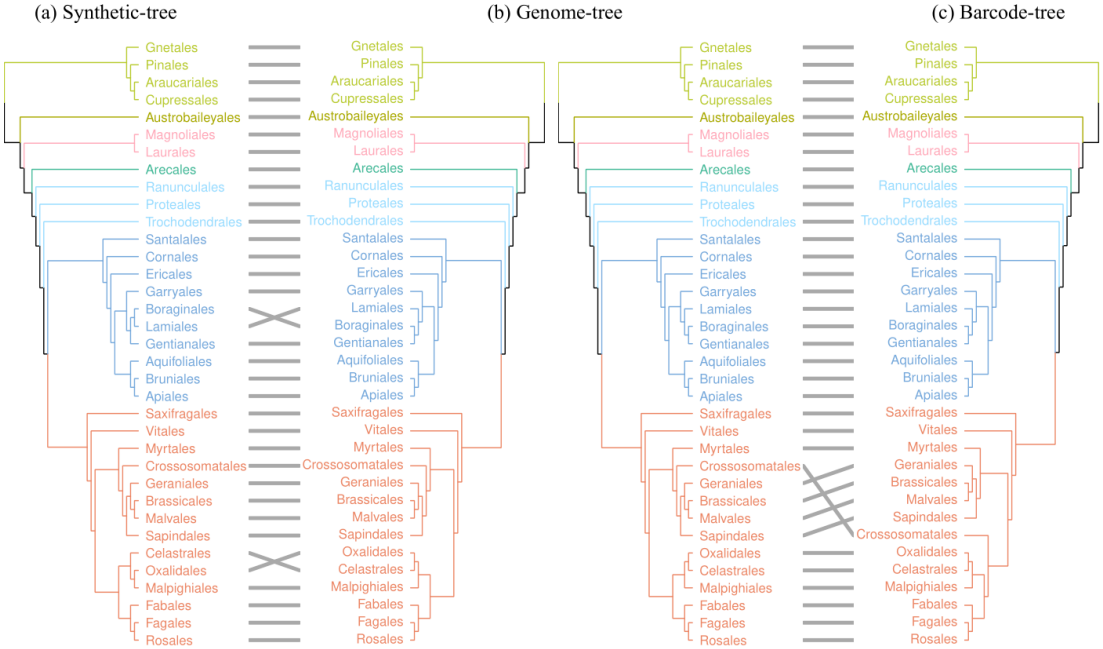
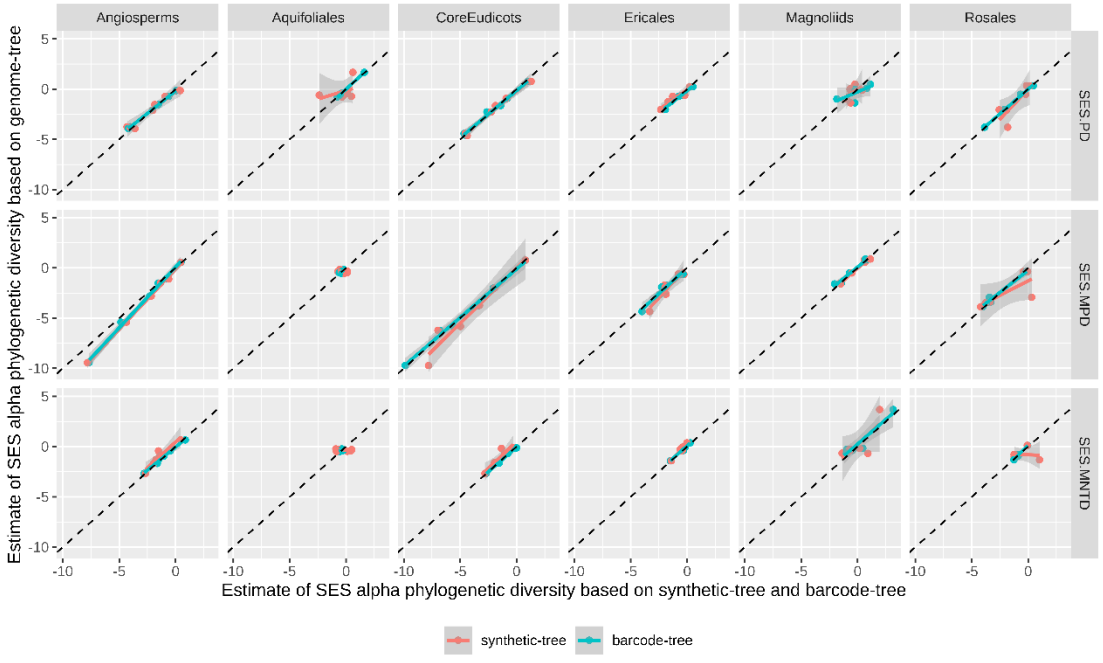


**Fig. S2.** Linear regression of standardized community phylogenetic structure calculated using genome-tree against those based on synthetic-tree (red) and barcode-tree (blue) across different clades. For 580 species of woody plants across six dynamic plots in the subtropical evergreen broadleaved forests of China, the synthetic-tree was reconstructed using R package V.PhyloMaker, the barcode-tree and genome-tree were reconstructed using three standard DNA barcodes (*rbcL* + *matK* + *trnH-psbA*) and 79 protein-coding genes from plastid genomes, respectively. Clades included three orders (Aquifoliales, Ericales and Rosales) containing species-rich genera that are dominated in subtropical EBLFs (e.g., *Ilex*, *Rhododendron* and *Ficus*) and three major phylogenetic groups of vascular plants (Angiosperms, Core Eudicots and Magnoliids). Detailed information on SES.PD, SES.MPD and SES.MNTD see Materials and Methods.



## Supplemental Files

**Supplemental File 1.** Species sampled in six dynamic plots for this study.

**Supplemental File 2.** Alignment file of the three standard DNA barcode markers (*rbcL*, *matK*, and *trnH-psbA*) used in this study.

**Supplemental File 3.** Alignment file of the 79 protein-coding genes (CDS genes) extracted from plastid genomes.

**Supplemental File 4.** Input file for treePL used to date barcode-tree.

**Supplemental File 5.** Input file for treePL used to date genome-tree