



ORIGINAL RESEARCH

The use of DNA barcodes to estimate phylogenetic diversity in forest communities of southern China

Jiajia Liu¹ | Juan Liu² | You-Xia Shan¹ | Xue-Jun Ge¹ | Kevin S. Burgess³

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou, China

²Collaborative Innovation Center of Jiangxi Typical Trees Cultivation and Utilization, Jiangxi Agriculture University, Nanchang, China

³Department of Biology, College of Letters and Sciences, Columbus State University, University System of Georgia, Columbus, Georgia

Correspondence

Xue-Jun Ge, Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou, China.
Email: xjge@scbg.ac.cn

and
Kevin S. Burgess, Department of Biology, College of Letters and Sciences, Columbus State University, University System of Georgia, Columbus, Georgia.
Email: burgess_kevin@columbusstate.edu

Funding information

CAS Strategic Priority Research Program, Grant/Award Number: XDB3103; National Natural Science Foundation of China, Grant/Award Number: 31500335, 31500341 and 31570210; Ministry of Science and Technology of the People's Republic of China, Grant/Award Number: 2014CB954100

Abstract

To elucidate potential ecological and evolutionary processes associated with the assembly of plant communities, there is now widespread use of estimates of phylogenetic diversity that are based on a variety of DNA barcode regions and phylogenetic construction methods. However, relatively few studies consider how estimates of phylogenetic diversity may be influenced by single DNA barcodes incorporated into a sequence matrix (conservative regions vs. hypervariable regions) and the use of a backbone family-level phylogeny. Here, we use general linear mixed-effects models to examine the influence of different combinations of core DNA barcodes (*rbcl*, *matK*, ITS, and ITS2) and phylogeny construction methods on a series of estimates of community phylogenetic diversity for two subtropical forest plots in Guangdong, southern China. We ask: (a) What are the relative influences of single DNA barcodes on estimates phylogenetic diversity metrics? and (b) What is the effect of using a backbone family-level phylogeny to estimate topology-based phylogenetic diversity metrics? The combination of more than one barcode (i.e., *rbcl* + *matK* + ITS) and the use of a backbone family-level phylogeny provided the most parsimonious explanation of variation in estimates of phylogenetic diversity. The use of a backbone family-level phylogeny showed a stronger effect on phylogenetic diversity metrics that are based on tree topology compared to those that are based on branch lengths. In addition, the variation in the estimates of phylogenetic diversity that was explained by the top-rank models ranged from 0.1% to 31% and was dependent on the type of phylogenetic community structure metric. Our study underscores the importance of incorporating a multilocus DNA barcode and the use of a backbone family-level phylogeny to infer phylogenetic diversity, where the type of DNA barcode employed and the phylogenetic construction method used can serve as a significant source of variation in estimates of phylogenetic community structure.

KEYWORDS

Bayesian tree, ITS, *matK*, phylogenetic inference, *rbcl*, subtropical forest

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Plant DNA barcodes, based on either single or multilocus regions of the chloroplast and/or nuclear genomes, have been applied to questions in community ecology (Kress et al., 2009; Valentini, Pompanon, & Taberlet, 2009). Estimates of phylogenetic genetic diversity can be used to quantify the evolutionary and ecological processes associated with community assembly, composition, and structure at different spatiotemporal scales (Cavender-Bares, Kozak, Fine, & Kembel, 2009; Helmus, Bland, Williams, & Ives, 2007; Mouquet et al., 2012; Webb, 2000). The branch lengths and topology of community phylogenies can influence estimates of phylogenetic diversity in different ways (Boyle & Adamowicz, 2015; Mazel et al., 2016; Swenson, 2009). For estimates based on DNA barcodes, the metric used to assess phylogenetic diversity may be influenced by the evolutionary rate of the barcode(s) employed. For example, DNA barcode regions that are phylogenetically conservative or hypervariable may under- or overestimate phylogenetic diversity, respectively. The effect of barcode region (and their combinations) on estimates of phylogenetic diversity metrics has not been empirically tested and may be a potential source of variation that requires consideration when assessing community phylogenetic structure.

Estimates of phylogenetic diversity may be also influenced by the type of phylogenetic construction method employed. Typically, tree topologies at deep phylogenetic nodes (e.g., family level) that have been inferred with a limited set of barcodes are largely incongruent with broadly accepted patterns of taxonomic relationships (e.g., APG IV; Byng et al., 2016). To constrain deep phylogenetic nodes and follow broadly accepted phylogenetic patterns, supertree methods (Bininda-Emonds & Sanderson, 2001; Webb & Donoghue, 2005) can be combined with DNA barcode sequence data (Erickson et al., 2014; Kress et al., 2010) to provide more accurate depictions of topology. Furthermore, the incorporation of a backbone phylogeny can provide more accurate estimates of the branch lengths (Boyle & Adamowicz, 2015) and potentially affect the metrics of phylogenetic community diversity (Swenson, 2009). Since ecological and evolutionary processes might operate at different phylogenetic depths (Mazel et al., 2016), it seems reasonable that phylogenetic diversity metrics that are sensitive to processes operating at deep phylogenetic depths may be strongly influenced by combining supertree methods with DNA barcode sequence data, whereas those estimates that are largely capturing diversity at the tips of the phylogeny may be less influenced, although this remains to be tested.

In contrast to the branch length-based metrics, several phylogenetic diversity metrics (e.g., PAE, the relationship between species evolutionary distinctiveness and abundance; IAC, the imbalance of abundances at higher clades) have been developed to capture information on both the topology and branch lengths of phylogenies connecting the species of a community (Cadotte et al., 2010; Krajewski, 1994; Vanewright, Humphries, & Williams, 1991). These topology-based metrics have also been shown to be valuable for

predicting patterns of abundance, community composition, and ecosystem functioning (Cadotte et al., 2010; Liu et al., 2018; Liu, Zhang, et al., 2015), but are seldomly evaluated in terms of how the branch lengths and topologies of community phylogenies may affect estimates of community phylogenetic diversity. Here, we predict that the use of a backbone phylogeny will have a strong influence on topology-based metrics.

To assess the potential variance in estimates of phylogenetic diversity associated with DNA barcodes and phylogeny construction methods, we first constructed a series of phylogenies, using Bayesian tree inference, for two distinct tropical forest communities that vary in elevation in the Dinghushan National Nature Reserve, Guangzhou, China. Specifically, we sampled two plastid gene regions (*rbcl* + *matK*) and the nuclear ribosomal internal transcribed spacers (ITS and ITS2 as part of the ITS region but with considerable power in species identification and resolution, see Chen et al., 2010) for all trees in each plot and constructed a series of phylogenies using different barcode combinations. To investigate the effects of supertree methods on estimates of phylogenetic diversity, we constructed another series of phylogenies with backbone family-level phylogenies based on APG IV (Byng et al., 2016). Taking a multi-model comparative approach, we assessed the relative contribution of single and multilocus barcodes, family-level backbone, and their combinations to predict the variance in estimates of phylogenetic diversity metrics. We address the following questions: (a) What are the relative influences of single DNA barcodes on estimates on phylogenetic diversity metrics? (b) What is the effect of using a backbone family-level phylogeny to estimate phylogenetic diversity metrics?

2 | MATERIALS AND METHODS

2.1 | Study sites

Both study plots were located at the Dinghushan National Nature Reserve, Guangdong province, South China (Figure 1: 23°10'N, 112°31'E; 23°10'N, 112°32'E), where the mean annual temperature is 21.0°C (range: -0.2°C to 38.1°C) and mean annual rainfall is 1,927 mm (Liu, Yan, et al., 2015). One plot is located in a subtropical mountain evergreen forest (600 m a.s.l.), while the other plot is located in a subtropical valley rain forest (100 m a.s.l.). Both plots have the same sampling area (1 ha) and similar arboreal species richness. There were a total of 114 trees with the abundance of each species being calculated by counting the number of individuals at breast height >10 cm in both plots. The mountain evergreen forest plot had 75 species with 41 unique species, and the valley rain forest plot had 73 species with 39 unique species. We list the detailed species information in Supporting Information Table S1.

2.2 | Community phylogenies

An exhaustive description of the methods for DNA extraction, PCR amplification, and sequencing can be found in Liu, Yan, et al. (2015). Here, we briefly describe the methods for phylogenetic construction. For the 114 species across both plots, we aligned *rbcl* and *matK* using

MAFFT (Katoh & Standley, 2013) and then eliminated divergent regions using Gblocks (Castresana, 2000). We aligned ITS and ITS2 using SATé (Liu et al., 2012). We then concatenated subsets of the *rbcl*, *matK*, ITS, and ITS2 sequences to generate a total of seven super matrices: (a) *rbcl* + *matK*, (b) *rbcl* + ITS, (c) *rbcl* + ITS2, (d) *matK* + ITS, (e) *matK* + ITS2, (f) *rbcl* + *matK* + ITS, (g) *rbcl* + *matK* + ITS2.

To assess the influence of a constrained family-level backbone on community phylogenetic diversity metrics, we constructed a total of fourteen species-level phylogenies based on the seven super matrices: one set based on Bayesian phylogenies and a second set based on Bayesian phylogenies with a constrained backbone topology at the family level based on the APG IV system (Byng et al., 2016). We then selected the best model of nucleotide substitution based on the lowest Akaike information's criterion (AIC) for each barcode region using the function "modelTest" in the *phangorn* library (Schliep, 2011) in R (R Core Team, 2016). For all barcode combinations, modelTest found that the best model was the generalized time reversible (GTR) model with a gamma distribution parameter describing among-site rate variation and a proportion of invariant sites parameter. We constructed all Bayesian phylogenies in MrBayes 3.2.5 (Ronquist et al., 2012) using four chains with 1,000,000 generations, a sampling and diagnostic frequency of 100, and a 20% burn in. We chose one representative of an early diverging gymnosperm lineage, *Cunninghamia lanceolata*, as the root for the Bayesian phylogenies. We then used a semi-parametric rate-smoothing method to transform the phylogeny to an ultrametric tree using the "chronopl" function with λ value 1,000 in the R *ape* library (Paradis, Claude, & Strimmer, 2004). For Bayesian phylogenies without family-level backbone constraint, we ranked all the post topologies by the symmetric distance with the backbone topology at the family level based on APG IV system using the function "treedist" in the R *phangorn* library (Schliep, 2011). Then we selected the top ranking 500 topologies for further analysis. We also randomly selected 500 topologies for Bayesian phylogenies with backbone for comparison. We used these selected topologies to estimate the posterior probabilities of the nodes for the Bayesian phylogenies and the Bayesian phylogenies with backbone, respectively (Figure S1–S7).

2.3 | Phylogenetic diversity metrics

For each of the fourteen Bayesian phylogenies (7 supermatrices with and 7 supermatrices without the backbone), we calculated several measures of phylogenetic diversity for all plants in the data set as well as at the plot level: Faith's PD, which sums all phylogenetic branch lengths (Faith, 1992); mean pairwise distance (MPD), which is the average distance separating all pairs of species of a community on the phylogenetic tree (Webb, Ackerly, McPeck, & Donoghue, 2002); and mean nearest taxon distance (MNTD), which is the average of the shortest phylogenetic distance for each species to its closest relative in the assemblage (Webb et al., 2002). We calculated MPD and MNTD using a species presence/absence matrix as well as a species abundance matrix. We denoted MPD_{ed} , $MNND_{ed}$ for the abundance-weighted versions of the metrics, respectively. In addition, we calculated (a) a metric of phylogenetic-abundance

evenness (PAE), which evaluates the relationship between the abundance and the distribution of terminal branch lengths (Cadotte et al., 2010) and (b) the imbalance of abundances at higher clades (IAC), which encapsulates the distribution of individuals across the nodes in the phylogeny (Cadotte et al., 2010). These diversity measures were chosen because of their wide use in ecology and conservation and because they represent measures of diversity that are based upon either branch lengths or tree topology.

2.4 | Linear mixed-effects models

To determine the effects of single and multilocus DNA barcodes, family-level backbone, and their combinations on each measure of phylogenetic diversity, we constructed a series of linear mixed-effects models using the "lme" function in the *nlme* library in R (Jose Pinheiro, Bates, DebRoy, Sarkar, & Team, 2016). The general form of the GLMM is as follows:

$$\log(\text{phylogenetic diversity}) = \alpha + \delta_{\text{plot}} + \text{rbcl}\beta_1 + \text{matK}\beta_2 + (\text{ITS}/\text{ITS2})\beta_3 + \text{backbone}\beta_4.$$

where *rbcl*, *matK*, ITS, ITS2, and the use of a family-level backbone phylogeny were set as fixed effects (not including the global intercept, α), and the plots (100 and 600 m) are random effects. We modeled the plots as random intercepts (δ_{plot}) to account for plot-level differences in measures of phylogenetic diversity that were unrelated to the particular barcodes and the use of a family-level backbone phylogeny. To meet the assumptions of normality, we log-transformed all measures of phylogenetic diversity. We evaluated model support using Akaike's Information Criterion corrected for small sample sizes (AIC_c ; Burnham & Anderson, 2002, 2004). To describe the proportion of variance explained by just the fixed factors and by the fixed and random factors together, we used the function "r.squaredGLMM" in the library *MuMIn* in R to calculate marginal R^2 and conditional R^2 , respectively (Barton, 2018; Nakagawa & Schielzeth, 2013). To check the robustness of multi-model inferences according to random sampling, we randomly resampled the suites of estimates of phylogenetic diversity 100 times and reran the multi-model inference for random datasets each time. The model ranks were consistent among random samples. Here, we only present the multi-model inference results based on original measures of phylogenetic diversity. To provide a relative rank of the importance of main predictors, we calculated standardized coefficients (β_n/SE_n) for each n term in the models featured in each subset, averaged these across all models based on AIC_c weights ($wAIC_c$) (re-calculating $\sum wAIC_c = 1$ over the models in which each term appeared), and then calculated the mean and confidence intervals (95%) of standardized coefficients for each term.

3 | RESULTS

Of the 24 general linear mixed-effect models that were constructed (including the intercept-only model), estimates of phylogenetic diversity based on models that included multi-locus barcodes had

TABLE 1 General linear mixed-effect model (GLMM) results for phylogenetic diversity metrics as a function of several fixed factors and hierarchical random factors

Metric	Model	AIC _c	ΔAIC _c	wAIC _c	R _m ²	R _c ²
PD	~B + M + R + I	-21,519	0	0.999	9.92	72.98
	~M + R + I	-21,483	36	<0.001	9.84	72.90
	~B + M + I	-20,854	665	<0.001	8.61	71.62
MPD	~B + M + R + I	-20,165	0	0.999	31.34	32.18
	~M + R + I	-20,105	60	<0.001	31.04	31.88
	~B + M + R + I2	-19,618	547	<0.001	28.65	29.47
MPD _{ed}	~B + M + R + I	-20,835	0	0.999	25.03	45.34
	~M + R + I	-20,806	29	<0.001	24.91	45.22
	~B + M + R + I2	-20,103	732	<0.001	22.12	42.34
MNTD	~B + M + R + I	-6,867	0	0.999	11.88	78.80
	~M + R + I	-6,848	20	<0.001	11.85	78.76
	~B + M + I	-4,935	1933	<0.001	9.09	74.69
MNTD _{ed}	~B + M + R + I	-11,181	0	0.999	15.11	63.78
	~M + R + I	-11,150	31	<0.001	15.03	63.69
	~B + M + R + I2	-10,583	598	<0.001	13.52	62.25
PAE	~B + M + R + I	-35,388	0	0.999	21.14	46.87
	~B + R + I	-34,775	613	<0.001	18.75	44.48
	~M + R + I	-34,729	660	<0.001	18.57	44.29
IAC	~B + M + R + I2	-53,553	0	0.999	0.034	99.893
	~B + M + R + I	-53,460	94	<0.001	0.033	99.892
	~B + M + R	-53,434	120	<0.001	0.033	99.892

Notes. Fixed factors are single plant barcodes (M = *matK*, R = *rbcl*, I = ITS, I2 = ITS2) and family-level backbone (B). Random factors are plots (100 and 600 m). Metrics are shown for seven phylogenetic diversity metrics (PD: phylogenetic diversity, MPD: mean pairwise distance, MPD_{ed}: abundance-weighted MPD, MNTD: mean nearest taxon distance, MNTD_{ed}: abundance-weighted MNTD, PAE: phylogenetic-abundance evenness, IAC: imbalance of abundance among clades). Values are shown for the information-theoretic Akaike's information criterion corrected for small samples (AIC_c), change in AIC_c relative to the top-ranked model (ΔAIC_c), AIC_c weight (wAIC_c, model probability), and the marginal and total variance explained (R_m², R_c²) as a measure of the model's goodness-of-fit. The top 3 models are listed; the full table is shown in Supporting Information Table S2–S8.

higher rankings than those based on single barcodes (Table 1, Supporting Information Table S2–S8). The top-rank models for all metrics except IAC included *rbcl*, *matK*, ITS, and family-level backbone (wAIC_c = 0.999 for PD, MPD, MPD_{ed}, MNTD, MNTD_{ed}, and PAE) (Table 1 and Supporting Information Tables S2–S7), which

accounted for >9% of the variances explained for each estimate. By contrast, ITS2 instead of ITS was included in the most parsimonious models for IAC (wAIC_c = 0.999; Table 1 and Supporting Information Table S8). However, IAC estimates were much less dependent on the combination of DNA sequence data and phylogeny

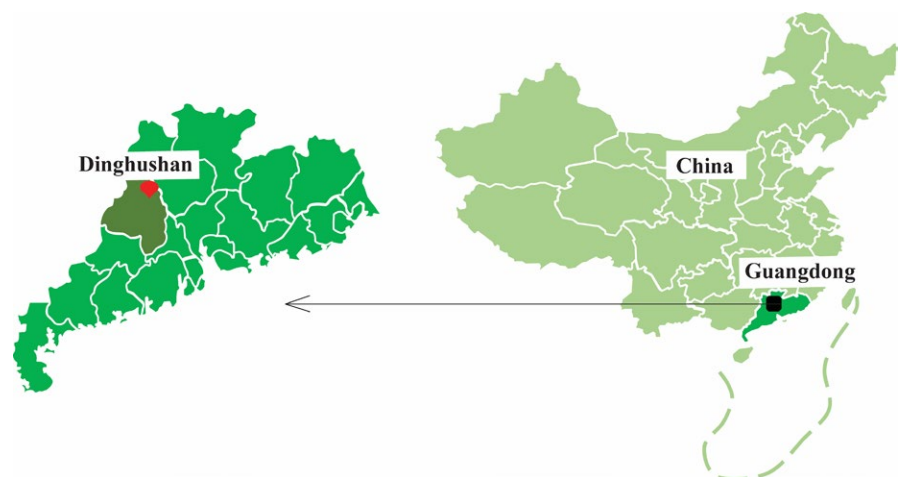


FIGURE 1 Map of the study sites on Dinghu Mountain, Guangzhou, Guangdong Province, China

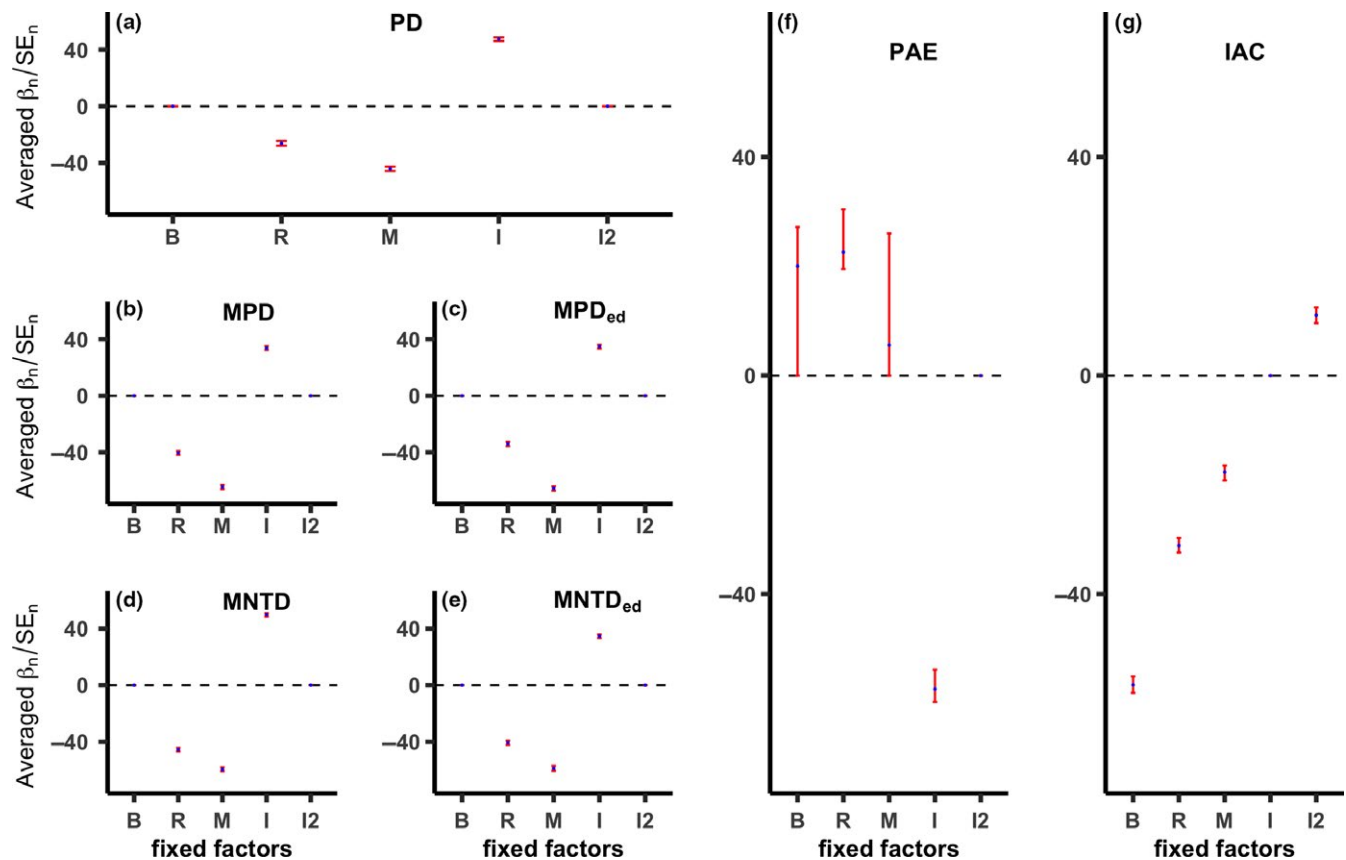


FIGURE 2 Averaged model standardized coefficients for each term considered in the general linear mixed-effect model sets to model the variances in measures of phylogenetic diversity of subtropical forest communities in China. Negative values indicate a negative relationship to estimates of phylogenetic diversity. β_n = estimated model term (n) coefficient, SE_n = term standard error, B = family-level backbone, R = *rbcl*, M = *matK*, I = ITS, I2 = ITS2. Analyses include a series of estimates of phylogenetic diversity (PD = phylogenetic diversity, MPD = mean pairwise distance, MPD_{ed} = abundance-weighted MPD, MNTD = mean nearest taxon distance, MNTD_{ed} = abundance-weighted MNTD, PAE = phylogenetic-abundance evenness, IAC = imbalance of abundance among clades). Shown in blue points and red error bars are the mean and confidence interval (95%) of the bootstrapping values of averaged model standardized coefficients for each term and each metric

construction method compared to estimates for other metrics ($R^2_m < 0.1\%$; Table 1 and Supporting Information Table S8).

For phylogenetic diversity metrics based on branch length, *rbcl*, *matK*, and ITS showed stronger effects on estimates of phylogenetic diversity than ITS2 and the backbone family-level phylogeny (Table 1, Figure 2a–e); *matK* had the strongest effect across the branch length-based metrics (Figure 2a–e). Family-level backbone had profound effects on measures of phylogenetic diversity that were based on tree topology (Figure 2f,g), but the direction of its effect depended on the topology-based metric. ITS was the most influential factor for estimates of PAE (Figure 2f), whereas family-level backbone was the most influential factor for IAC (Figure 2g).

4 | DISCUSSION

Our results reveal that multilocus barcodes outperform single-locus barcodes in explaining maximum variation in estimates of

phylogenetic diversity regardless of the phylogenetic reconstruction methods used, both in terms of model ranking and model-averaged, standardized effects. This result is in line with previous meta-analytical and experimental evidence that suggests a combination of more than one DNA barcode locus, including a phylogenetically conservative coding locus and one or more rapidly evolving barcode regions, are essential for inferring robust phylogenetic relationships among plants (Burgess et al., 2011; Fazekas et al., 2008; Hollingsworth, Forrest, et al., 2009; Kress & Erickson, 2007; Kress et al., 2009; Li et al., 2011; Liu, Yan, et al., 2015). Here, the reason for the complementary influence of DNA barcodes with different rates of evolution might be due to different ecological and evolutionary processes operating at different evolutionary time scales, which contribute differently to plant community phylogenetic structure (Mazel et al., 2016). For example, conserved DNA barcodes might provide important insight into the processes acting at long evolutionary time scales, whereas rapidly evolving barcodes might signal more recent speciation events (Webster, Payne, & Pagel, 2003).

Of the universal barcodes that were used in our models, the effects of both chloroplast DNA regions (*rbcl* & *matK*) were evident across all phylogenetic diversity metrics. Notably, *matK* tended to be a more important factor for inferring phylogenetic diversity metrics that are based on branch length methods, while *rbcl* had a greater influence on topology-based metrics. This result suggests that *rbcl* and *matK* might be useful for estimates phylogenetic diversity for subtropical plant communities in China by establishing deep phylogenetic branches and terminal branches, respectively. Indeed, there is increasing evidence that *matK* is the most variable coding region of the angiosperm plastome and as such, in most genera, *matK* has higher species discriminatory power compared to *rbcl* (Hilu et al., 2003; Hollingsworth, Clark, et al., 2009; Liu, Yan, et al., 2015). However, our results also suggest that the influences of DNA barcodes on measures of phylogenetic diversity might be inconsistent with their discriminatory success and more dependent on the methods and metrics used, both sources of variation that will likely have broad implications for future studies.

In this study, ITS had a stronger effect than ITS2 on estimates of community phylogenetic diversity except for IAC. This result is not surprising given that ITS2 is only one of three partitions in the ITS gene region (ITS1, 5.8S, ITS2; Coleman, 2003). Collectively, the effects of ITS on estimates of phylogenetic diversity that were based on branch length methods were comparable to those of regions of the plastid genome (i.e., *matK*) using model-averaged, standardized coefficients. Although this finding implies that ITS does not estimate branch lengths better than *matK*, ITS does show a much stronger influence on estimates of PAE than barcode regions of the plastid genome. Because PAE stresses the phylogenetic-abundance distributions among terminal branches (Cadotte et al., 2010), our results indicate that ITS may be a better estimator, over the other three DNA barcodes, of phylogenetic relationships at the tips of the community phylogeny. Of the four DNA barcode markers used in this study, ITS has been shown to have the highest species discriminatory power due to its ability to differentiate closely related, congeneric species (Li et al., 2011). Meanwhile, ITS2 outperformed ITS in predicting the variation in IAC, which stresses the topology at deep nodes of the phylogeny. This result suggests that ITS2 may be better tool for estimating phylogenetic relationships at deep clades (Yao et al., 2010).

We found that models that included a backbone family-level phylogeny had the highest support but only showed a strong effect for measures of phylogenetic diversity that are based on the topology of the community phylogeny. In our study system, the use of a backbone family-level phylogeny was required to explain maximum variation in topology-based metrics, which is consistent with our expectation that a limited number of DNA barcodes might generate inconsistent relationships deep within the phylogeny compared to broadly accepted patterns (Erickson et al., 2014). Such inconsistency might have a significant influence on measures of phylogenetic diversity, which are more sensitive to the basal topology of phylogenies. Indeed, we found evidence that the use of a backbone family-level phylogeny was more effective on estimates of IAC than that of PAE, given PAE measures the phylogenetic-abundance distribution

among terminal branches and IAC quantifies the imbalance of abundances at deeper clades (Cadotte et al., 2010). However, the effects of enforcing a backbone for deeper relationships in the phylogeny were negligible for estimates of branch length-based metrics.

Although the optimal combination of DNA barcodes (e.g., *rbcl* + *matK* + ITS) and the use of a backbone family-level phylogeny served as a consistent and accurate predictor for the metrics of phylogenetic diversity considered here, the explanatory power of the top-rank models varied depending on how phylogenetic diversity was measured. For example, mean pairwise distance (MPD) attained the highest proportion (>31%) of the explained variation, whereas the imbalance of abundances at higher clades (IAC) attained the lowest proportion (0.1%). It is generally agreed that different ecological processes (i.e., environmental filtering and limiting similarity) and evolutionary processes (i.e., local adaptation, speciation, extinction) operating at different spatiotemporal scales can contribute to community structure (Cavender-Bares et al., 2009; Swenson, 2011). However, determining the relative contribution of ecological versus evolutionary processes contributing to community patterns can be difficult (Cavender-Bares et al., 2009). Among the metrics of phylogenetic diversity considered, MPD was “best” explained by the combination of DNA barcodes, which is consistent with previous studies. For example, Mazel et al. (2016) showed that MPD is more sensitive to long-term evolutionary processes compared to PD and MNTD. However, our results for IAC, which is independent of the DNA barcodes and phylogeny construction methods, suggest that ecological processes are mainly generating community phylogenetic patterns at our sites. Given that combinations of DNA barcodes, the use of a backbone family-level phylogeny, and the plots together accounted for the substantial proportions ($R^2_c > 32\%$) of variation in a series of estimates of phylogenetic diversity, this result is in line with the view that both ecological and evolutionary processes are influencing biodiversity at our sites. Future studies should consider the inclusion of functional, environmental, and demographic data to further elucidate the underlining ecological and evolutionary mechanisms contributing to community structure.

This study examined the influence of different combinations of four core DNA barcodes and community phylogeny reconstruction methods on a series of estimates of community phylogenetic diversity metrics for two subtropical forest plots in Guangdong, southern China. There are, however, a number of additional DNA fragments including coding regions (i.e., *rpoB* & *rpoC1*) and noncoding spacer regions (i.e., *atpF-atpH*, *trnH-psbA*, and *psbK-psbI*) that have been proposed as candidates for universal plant DNA barcodes (Hollingsworth, Forrest, et al., 2009). Furthermore, our power to detect influences of plant DNA barcodes and tree construction methods on estimates of phylogenetic diversity was limited by a small set of phylogenetic diversity metrics. Although these factors should be considered in future studies, our study provides insight into the magnitude of the influence of single barcodes, or their combination, and phylogeny reconstruction methods on community phylogenetic patterns. Notably, our study underscores the complexity of explaining community phylogenetic patterns, where future studies should

evaluate the sensitivity of phylogenetic diversity metrics to the methods employed.

ACKNOWLEDGMENTS

The authors thank Dr. WH Ye for the data of the field experiment. Funded by the National Natural Science Foundation of China (31500335, 31570210, 31500341), CAS Strategic Priority Research Program (XDB3103), and Ministry of Science and Technology of the People's Republic of China (2014CB954100).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

JJL conceived the idea, JL collected the DNA data, JJL and YXS analyzed the data, JJL and KSB led the writing of the manuscript, and all authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The essential data in the data analysis of this paper will be archived in a public accessible repository upon acceptance.

ORCID

Jiajia Liu  <https://orcid.org/0000-0002-9778-060X>

You-Xia Shan  <https://orcid.org/0000-0002-6417-4777>

Xue-Jun Ge  <https://orcid.org/0000-0002-5008-9475>

Kevin S. Burgess  <https://orcid.org/0000-0002-9663-9428>

REFERENCES

- Barton, K. (2018). MuMIn: Multi-model inference. Retrieved from <https://CRAN.R-project.org/package=MuMIn>
- Bininda-Emonds, O. R. P., & Sanderson, M. J. (2001). Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Systematic Biology*, 50, 565–579. <https://doi.org/10.1080/10635150120358>
- Boyle, E. E., & Adamowicz, S. J. (2015). Community phylogenetics: Assessing tree reconstruction methods and the utility of DNA barcodes. *Plos One*, 10, e0126662.
- Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., Newmaster, S. G., ... Barrett, S. C. H. (2011). Discriminating plant species in a local temperate flora using the rbcL plus matK DNA barcode. *Methods in Ecology and Evolution*, 2, 333–340.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.
- Burnham, K. P., & Anderson, D. R. (2004). Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33, 261–304.
- Byng, J. W., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Judd, W. S., Mabberley, D. J., ... 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, 1–20.
- Cadotte, M. W., Davies, T. J., Regetz, J., Kembel, S. W., Cleland, E., & Oakley, T. H. (2010). Phylogenetic diversity metrics for ecological communities: Integrating species richness, abundance and evolutionary history. *Ecology Letters*, 13, 96–105. <https://doi.org/10.1111/j.1461-0248.2009.01405.x>
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17, 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- 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, 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., ... Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *Plos One*, 5, e8613.
- Coleman, A. W. (2003). ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics*, 19, 370–375. [https://doi.org/10.1016/S0168-9525\(03\)00118-5](https://doi.org/10.1016/S0168-9525(03)00118-5)
- Erickson, D. L., Jones, F. A., Swenson, N. G., Pei, N., Bourg, N. A., Chen, W., ... Kress, W. J. (2014). Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: A mega-phylogeny approach. *Frontiers in Genetics*, 5, 358.
- Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61, 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
- Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., ... Barrett, S. C. H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *Plos One*, 3, e2802. <https://doi.org/10.1371/journal.pone.0002802>
- Helmus, M. R., Bland, T. J., Williams, C. K., & Ives, A. R. (2007). Phylogenetic measures of biodiversity. *American Naturalist*, 169, E68–E83.
- Hilu, K. W., Borsch, T., Muller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., ... Chatrou, L. W. (2003). Angiosperm phylogeny based on matK sequence information. *American Journal of Botany*, 90, 1758–1776. <https://doi.org/10.3732/ajb.90.12.1758>
- Hollingsworth, M. L., Clark, A. A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., ... Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, 9, 439–457. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>
- Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., ... Grp, C. P. W. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12794–12797.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Krajewski, C. (1994). Phylogenetic measures of biodiversity – a comparison and critique. *Biological Conservation*, 69, 33–39. [https://doi.org/10.1016/0006-3207\(94\)90326-3](https://doi.org/10.1016/0006-3207(94)90326-3)
- Kress, W. J., & Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. *Plos One*, 2, e508.
- Kress, W. J., Erickson, D. L., Andrew Jones, F., Swenson, N. G., Perez, R., Sanjurjo, 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 of the United States of America*, 106, 12794–12797.

- States of America*, 106, 18621–18626. <https://doi.org/10.1073/pnas.0909820106>
- Kress, W. J., Erickson, D. L., Swenson, N. G., Thompson, J., Uriarte, M., & Zimmerman, J. K. (2010). Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican Forest Dynamics plot. *Plos One*, 5, e15409. <https://doi.org/10.1371/journal.pone.0015409>
- Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., ... China Plant, B. O. L. G. (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. <https://doi.org/10.1073/pnas.1104551108>
- Liu, J., Liu, D., Xu, K., Gao, L.-M., Ge, X.-J., Burgess, K. S., & Cadotte, M. W. (2018). Biodiversity explains maximum variation in productivity under experimental warming, nitrogen addition, and grazing in mountain grasslands. *Ecology and Evolution*, 1–19. <https://doi.org/10.1002/ece3.4483>
- 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, 188–199.
- Liu, J., Zhang, X., Song, F., Zhou, S., Cadotte, M. W., & Bradshaw, C. J. A. (2015). Explaining maximum variation in productivity requires phylogenetic diversity and single functional traits. *Ecology*, 96, 176–183.
- Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P., & Linder, C. R. (2012). SATé-II: Very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology*, 61, 90–106. <https://doi.org/10.1093/sysbio/syr095>
- Mazel, F., Davies, T. J., Gallien, L., Renaud, J., Groussin, M., Munkemüller, T., & Thuiller, W. (2016). Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography*, 39, 913–920. <https://doi.org/10.1111/ecog.01694>
- Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., ... Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. <https://doi.org/10.1093/bioinformatics/btg412>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2016). nlme: Linear and nonlinear mixed effects models. Retrieved from <https://CRAN.R-project.org/package=nlme>
- R Core Team. (2016). R: A language and environment for statistical computing. In R. F. S. Computing, editor.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, 539–542.
- Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27, 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Swenson, N. G. (2009). Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *Plos One*, 4, e4390.
- Swenson, N. G. (2011). The role of evolutionary processes in producing biodiversity patterns, and the interrelationships between taxonomic, functional and phylogenetic biodiversity. *American Journal of Botany*, 98, 472–480. <https://doi.org/10.3732/ajb.1000289>
- Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, 24, 110–117. <https://doi.org/10.1016/j.tree.2008.09.011>
- Vanewright, R. I., Humphries, C. J., & Williams, P. H. (1991). What to protect – systematics and the agony of choice. *Biological Conservation*, 55, 235–254. [https://doi.org/10.1016/0006-3207\(91\)90030-D](https://doi.org/10.1016/0006-3207(91)90030-D)
- Webb, C. O. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *American Naturalist*, 156, 145–155. <https://doi.org/10.1086/303378>
- Webb, C. O., Ackerly, D. D., McPeck, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>
- Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology Notes*, 5, 181–183. <https://doi.org/10.1111/j.1471-8286.2004.00829.x>
- Webster, A. J., Payne, R. J. H., & Pagel, M. (2003). Molecular phylogenies link rates of evolution and speciation. *Science*, 301, 478–478. <https://doi.org/10.1126/science.1083202>
- Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., ... Chen, S. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *Plos One*, 5, e13102.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Liu J, Liu J, Shan Y, Ge X-J, Burgess KS. The use of DNA barcodes to estimate phylogenetic diversity in forest communities of southern China. *Ecol Evol.* 2019;9:5372–5379. <https://doi.org/10.1002/ece3.5128>