Explore the evolutionary cause of phylogenetic diversity between East Asian and East North American disjucted forest communities

Hanyang Lin, Miao Sun, Daijiang Li, Doug Soltis, Yunpeng Zhao, Pam Soltis

2019-06-14 14:13:37

## Warning: package 'knitr' was built under R version 3.5.2

**Running headline**: Evolutionary history of EA and ENA disjuction forest communities

**Abstract**:

# Introduction

The deciduous forests of eastern Asia (EA) and eastern North America (ENA) represent relics of once-widespread mixed mesophytic forests of the Northern Hemisphere. Despite over a century of observation and study, many questions remain regarding the origin, evolution, and assembly of these forests. Disjunct EA and ENA forests provide an excellent natural system to examine the link between species diversity and ecological function within the context of a shared phylogenetic history.

Here is the second paragraph of the introduction.

# Materials and Methods

## species list

For 5 Chinese site, Shengnongjia Moutain (Mt.), Gutian Mt., Tianmu Mt., Changbai Mt., Dongling Mt., we collected species checklist from their nature reserves (See Table S1). [**Citation?**] Since not all sites have complete species checklist, for 6 sites in Unites States, we collected species list for each site using two approaches. Intially, we consulted from species list if exist, from local flora, and vouchers for our samplings. Second approch, we extract seed plant species of each site from public databases of GBIF, and iDigBio with a defined polygon. The polygon was created by centered coordinates of each site with 10 km radius. All these procedures were conducted using R packages (“rinat”, “sp”, “spocc”, “tidyverse”, and “wicket”).  
All the resultant species lists of each site were validated and reconsicled with The Plant List using R packages “Taxonstand”. We also generated a whole comprehensive species list by concatenating the species lists from all sites, and remove the duplicates.

## Community phylogeny

Based on the comprehensive species lists generated above, we queried the broadly inclusive seed plant phylogeny (Smith and Brown [2018](#ref-Smith_Brown_2018)) via functions from Phyx (Brown et al. [2017](#ref-Brown_et_al_2017)) and python scripts from OpenTree PY Toys (<https://github.com/blackrim/opentree_pytoys>). The final cleaned tree is ultrametric and contained 5,608 tips. Smaller trees for each site were also generated via the same approch for future comparision.

## Phylogenetic diversity (PD)

A series of supermatrixs was generated by scoring 5,608 tips with presents [“1”] and absents [“0”] of each of 11 sites, Angiosperms vs. Gymnosperms, woody vs. herbaceous, and disjunct vs. non-disjunct.

Then Standardized effect size of phylogenetic diversity (Faith’s PD), Standardized effect size of mean pairwise distances (MPD), Standardized effect size of mean nearest taxon distances (MNTD) in Alpha diversity were calculated based on phylogenetic trees and matrixes generated above, using R package (“picante”; Kembel et al. [2010](#ref-Kembel_et_al_2010)). Likewise, inter-community mean pairwise distance (MPD), inter-community mean nearest taxon distance, and Unweighted UniFrac distance between communities in Beta diversity were also calculated using the same package. While Phylogenetic Community Dissimilarity (PCD) in Beta diversity was calculated by cpp version of R package (“phyr”; <https://github.com/daijiang/phyr>). The same functionality, but 20 times faster than it in “picante”.

## Diversification Rate

To investiagte the cause of PD pattern between forest communities in EA and ENA, we also employed the DR statistic (Jetz et al. [2012](#ref-Jetz2012)), one of the most widely used semiparametric approaches to estimate diversification rate. The DR statistic quantifies the “splitting rate” from each extant species to the tree root as a model-free estimate of diversification rate. Methods followed those described in Jetz et al. ([2012](#ref-Jetz2012)) and Harvey et al. ([2017](#ref-Harvey2017)).

# Results

Q1: Which floristic region, i.e., EAS or ENA, harbors greater species diversity, including species richness (SR) and phylogenetic diversity (PD)?

A1: Collectively, EA sites show greater species diversity than ENA. In term of species richness, sites in EA harbor over one and half times of species number as ENA sites do (3667 vs. 2194). Meanwhile, both observed and standard-effective-size (SES) PD are determined to be greater in EAS sites than ENA (observed PD: 66034.38 vs. 44617.83, SES-PD: -2.71 vs. -6.40).

Q2: Among 11 EAS-ENA sites, where does seed plant biodiversity hotspot locate at? Do diversity changes along any gradients or follow any trends? And to what extent could diversity anomaly be observed?

A2: SES-PD of EAS sites tends to decrease along latitude, which has not been observed in ENA. Among 11 EAS-ENA sites, northernmost EAS site, CBS, shows the lowest SES-PD (-8.80), while southernmost EAS site, GTS, shows the greatest SES-PD (1.87). It is worth noticing that two northernmost sites in each continent show most significant asymmetry between species richness and SES-PD. Showing no gap in species richness (1140 spp.) comparing to the average diversity of species among five EAS sites (1174 spp.), CBS site embodies lowest SES-PD (-8.92). In ENA, The White Mountain site have only 254 seed plants species inhabited, which is far below the average level of species richness in ENA (ca. 600 spp.), yet harboring the second highest phylogenetic diversity in ENA. In EAS, two northern sites, CBS and DLS, show high similarity and two southern sites, GTS and TMS show close affinity too. In ENA, dissimilarity sites among sites tends to aggravate with larger gap in latitude. From a global perspective, heterogeneities of phylogenetic diversity among sites are examined to be greater in EAS than ENA.

Q3: What will the diversity scenarios be when seed plants are categorized to angiosperms vs. gymnosperms, or woody vs. herbaceous plants, or genera exhibiting EAS-ENA disjunction distribution vs. the remains?

A3: Globally, herbaceous species have nearly twice members as woody species do (3750 vs. 1858 spp.), yet showing lower SES-PD (-22.02 vs. -9.29). Comparing to herbaceous plants, woody lineages show higher values on SES-PD, MPD and MNTD in all 11 sites. Angiosperms lineages show over 70 times in species richness with gymnosperms (5532 vs. 76 spp.), yet showing lower SES-PD (-12.26 vs. -8.91). When pooling together, lineages exhibiting EAS-ENA disjuct distribution show lower SES-PD value than the remains (-14.68 vs. -4.25). However, on individual site level, those EAS-ENA disjunction lineages show greater values on SES-PD, MPD and MNTD than both total communities and plant communities comprising of those not showing EAS-ENA disjunct distribution (referred to non-disjunct hereafter). Moreover, the total communities and non-disjunct lineages share large similarity on diversity curves while the EAS-ENA disjunctions show hardly identical trends with both of them.

Q4: What pattern of community structure and putative community assemblage process show at regional and local scale respectively?

A4: Note that NRI and NTI are two oft-used indexes in estimating community structure. Somehow, these two indexes could show opposite polarities. Basically, comparing to NTI, NRI could be used to better demonstrate community structure globally since it is generated from MPD. Hence, preferences towards using NRI on elucidating the community structure have been made in this study. Collectively, both EAS and ENA mixed mesophytic seed plant communities are determined to show no significant trend toward either phylogenetic over-dispersion or clustering at continental level based on NRI (-0.57 vs. 1.28). Specifically, majority of all 11 sites show the same community structure with the global scenario. In other words, most communities are examined to be assembled from nearly neutral process. However, some exceptions have been found. The White Mountain site show significant over-dispersion (NRI = -2.64), suggesting seed plant community there could be assembled by competitive exclusion or density effect. The Coweeta site show significant clustering (NRI = 2.05) which could suggest that habitat filtering had played a vital role in community assemblage. Moreover, based on NTI values, nine out of 11 sites show significant clustering (NTI >= 1.96).

Q5: Similar to Q3, considering different categories, what community structure and community assemblage process could be inferred?

A5: Despite communities from both the Old and the New Worlds show largely no significant phylogenetic clustering or over-dispersion, seed plants with different lifeforms serve as strikingly different entities considering community structure and community assemblage process. On any scales, with mere exceptions, woody seed plants show significant over-dispersion while herbaceous lineages significant shows the opposite, providing insights that woody communities have been shaped by competitive exclusion or density effect whilst the herbaceous components have undergone habitat filtering. Speaking of EAS-ENA disjunctions, those lineages show significant phylogenetic over-dispersion when taking all mixed forests from EAS and ENA as one mega community (NRT = -6.41) while the counterpart show significant clustering (NRI = 10.08). Within each continent, identical pattern could also be found. However, these overwhelming trends seem to diminish at finer scale.

# Discussion

# Conclusion

# Acknowledgement

This work was supported by the National Science Foundation Dimensions of Biodiversity US-China (DEB-1442280 to P.S.S. and D.E.S.), and National Natural Science Foundation of China (Grant no. xxx to xxx). We thank Dr. Matthew A. Gitzendanner for bioinformatics trouble shooting. We thank the HiPerGator cluster at the University of Florida and CIPRES for providing us extensive computational resources.

# Author contributions

The authors declare no conflict of financial interests. H.Y.L., M.S and P.S.S. designed the study; M.S. and H.Y.L. conducted the analyses; D.J.L. advised with data interpretation; M.S. and H.Y.L. drafted the manuscript; D.J.L., Y.P.Z., D.E.S., and P.S.S. revised the manuscript; D.E.S., P.S.S., and Y.P.Z. supervised the work. All authors contributed, and approved the final manuscript.

Mark Whitten?

#### Tables

Table. ??:

?? Geo-informatiion of US-CN 11 sites.

|  |  |  |
| --- | --- | --- |
| Site | Area (km²) | Elevation |
| col 3 is | right-aligned | $1600 |
| col 2 is | centered | $12 |
| zebra stripes | are neat | $1 |

|  |  |  |  |
| --- | --- | --- | --- |
| Col A | Col B | Col C | Col D |
| row 1 | 190 |  |  |
|  | 0.13 | 0.12 | 0.12 |
|  | 0.14 | 0.13 | 0.50 |
|  | 0.15 | 0.31 | 0.52 |

#### Figures

Insert figure by code chunk. And cross-ref it back as Figure ??.

Or if you already have the figure: And cite it as Figure 1.

![Figure 1: Caption here.](data:application/pdf;base64,)

Figure 1: Caption here.

More details can be found at [here](https://bookdown.org/yihui/bookdown/).

# References

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33:1886–1888.

Harvey, M. G., G. F. Seeholzer, B. T. Smith, D. L. Rabosky, A. M. Cuervo, and R. T. Brumfield. 2017. Positive association between population genetic differentiation and speciation rates in new world birds. Proceedings of the National Academy of Sciences 114:6328–6333.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444 EP.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany 105:302–314.