

**Supporting Information.** Daijiang Li, Lauren Trotta, Hannah E. Marx, Julie M. Allen, Miao Sun, Douglas E. Soltis, Pamela S. Soltis, Robert P. Guralnick, and Benjamin Baiser. 2019. For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*.

## Appendix S1

### **Section S1: details about the three purpose-built phylogenies used**

#### **Pine Rockland community phylogeny**

The pine rockland community phylogeny was constructed from a combination of field and herbarium based collections and supplemented with pre-existing sequences from the Flora of Florida project at the University of Florida. The Floristic Inventory of South Florida (FISF), which includes a comprehensive list of pine rockland species (n = 592, Gann GD 2018), guided our sampling efforts. Specimens were collected in fruit or flower over the course of all seasons and across multiple pine rockland fragments in Miami Dade County from 2014 to 2016. We collected material from 331 new field collections, 17 herbarium collections at Fairchild Tropical Botanic Garden (FTBG; Miami, FL, USA) or the University of Florida (FLAS; Gainesville, FL, USA), and 58

field-collected plants currently in cultivation at Fairchild Tropical Botanic Garden. For field and herbarium collections, we extracted total genomic DNA and amplified three commonly used plastid barcodes (Kress et al. 2009): *rbcL*, *matK*, and *psbA-trnH*. Amplification success was confirmed with gel electrophoresis and Sanger sequencing was performed by Beckman Coulter Genomics (Cambridge, Massachusetts, USA), Genewiz (South Plainfield, New Jersey, USA), or Eurofins Genomics (Louisville, Kentucky, USA). In all, we were able to include 540 taxa in our community phylogeny, 90.88% of all vascular plants that occur on the FISF species list.

Both newly generated and pre-existing sequences were edited and assembled using Geneious R9 (Biomatters, Auckland, New Zealand), and resulting alignments were constructed using the MAFFT v1.3.5 (Kato and Standley 2013) plugin in Geneious. We visually inspected individual alignments before concatenating all three loci into a final alignment. We used PartitionFinder v1.1.1 (Lanfear et al. 2012) and the Akaike information criterion (AIC) to determine the optimal model of evolution for each barcode as well as the best overall partitioning scheme, which were used for all down stream Maximum likelihood (ML) and Bayesian Inference (BI) analyses. We conducted all analyses on the HiPerGator 2.0 supercomputing cluster at the University of Florida. ML analysis was performed in RAxML v8.2.8 (Stamatakis 2014) with an ordinal level topological constraint, following the Angiosperm Phylogeny Group IV (APG IV et al. 2016) for flowering plants and the Pteridophyte Phylogeny Group (Schneider et al. 2016) for ferns and lycophytes. One thousand bootstrap replicates were used to determine clade support. We used a single lycophyte species (*Selaginella eatonii*) as the outgroup. Dating analysis was performed in BEAST 2.4.6 (Bouckaert et al. 2014). Our BEAST analysis included nine fossil constraints selected from previous molecular dating studies in addition to ordinal and superclade constraints sensu APG IV (Schuettpelz and Pryer 2009, Bell et al. 2010, Magallón et al. 2013, APG IV et al. 2016). The tree topology that resulted from our ML search was used as the starting tree for this analysis after it was calibrated to match the fossil constraints using the *chronos* function in the R package *ape* (Paradis et al. 2004). More details about all included analyses can be found in Trotta et al. (2018).

## Alpine community phylogeny

For each of the species occurring in the Écrins National Park, France ( $n = 1,345$ ), we used the PHLAWD pipeline (Smith et al. 2009) to retrieve sequence data for five gene regions (*atpB*, *rbcL*, *matK*, *trnTLF*, and *ITS*) from GenBank (release 209). Sites that were missing across >50% of the taxa in the output alignment were removed from each gene region using PHYUTILITY (Smith and Dunn 2008), and maximum likelihood (ML) inference implemented in RAxML (Stamatakis 2014) was used to estimate gene trees. Outlier taxa (i.e. those falling outside of clades defined by the APG III taxonomy) were identified by visual inspection of each gene tree and were removed before concatenation into a final alignment using PHYUTILITY (Smith and Dunn 2008). The resulting super-matrix consisted of 79% ( $n = 1,065$ ) of the species surveyed and was used to estimate a ML community phylogeny of the Écrins flora in RAxML v8.0.4, under a GTR-CAT model of nucleotide evolution with simultaneous rapid bootstrap and ML search using 999 replicates, which is ideal for large nucleotide alignments (Stamatakis 2014). We used the ‘congruification’ approach (Eastman et al. 2013) to map divergence times from a reference time-tree (Soltis et al. 2011, Zanne et al. 2014) with concordant nodes on the best ML estimate of tree topology in the R package *geiger* v2.0.6 (Pennell et al. 2014), and the phylogeny was scaled to time using *treePL* (Smith and O’Meara 2012). Further details can be found in Marx et al. (2017).

## Florida Flora phylogeny

We constructed a phylogeny for the vascular flora of Florida based on sequence data of two plastid genes commonly used in plant phylogenetics: *rbcL* and *matK*. We first leveraged existing published sequence data on GenBank. We obtained sequence data from GenBank for 463 species. We checked these sequences throughout the alignment and tree-building process and removed problematic sequences. The remaining species lacked published DNA sequence data. We thus collected fresh materials of these species and sequenced them and deposited voucher specimens at

the University of Florida herbarium (FLAS). For rare and hard-to-access species, we sampled specimens from FLAS. We aligned sequences of *rbcL* and *matK* according to a reference protein-coding sequence using Pal2Nal (Suyama et al. 2006). We used PartitionFinder v1.1.1 and AIC (Lanfear et al. 2012) to determine the optimal model of evolution. We then ran RAxML v7.03 (Stamatakis 2014) on the partitioned dataset. We used 100 bootstrap replicates to determine the internal support for the initial tree evaluation. We used three outgroup taxa using data from GenBank: *Physcomitrella patens* (Funariaceae), *Syntrichia ruralis* (Pottiaceae), and *Mastigophora woodsii* (Mastigophoraceae). We examined all phylogenetic trees for issues including possible misidentifications of species, contaminants, and misplaced taxa. We used the best tree from RAxML as a starting tree in a Bayesian analysis using MrBayes v3.2.2 (Ronquist et al. 2012) for  $10^6$  generations with 20 independent runs. The top 10 trees were sampled based on likelihood score as representations of the posterior distribution of trees. We used the tree with the highest likelihood in this study. We dated the tree, producing a chronogram, using r8s (Sanderson 2003) and 17 calibration points with a maximum age of 377 million years. More details about the phylogeny building process can be found in Allen et al. (2019).



Median correlation based on 100 simulations; dataset pine rockland

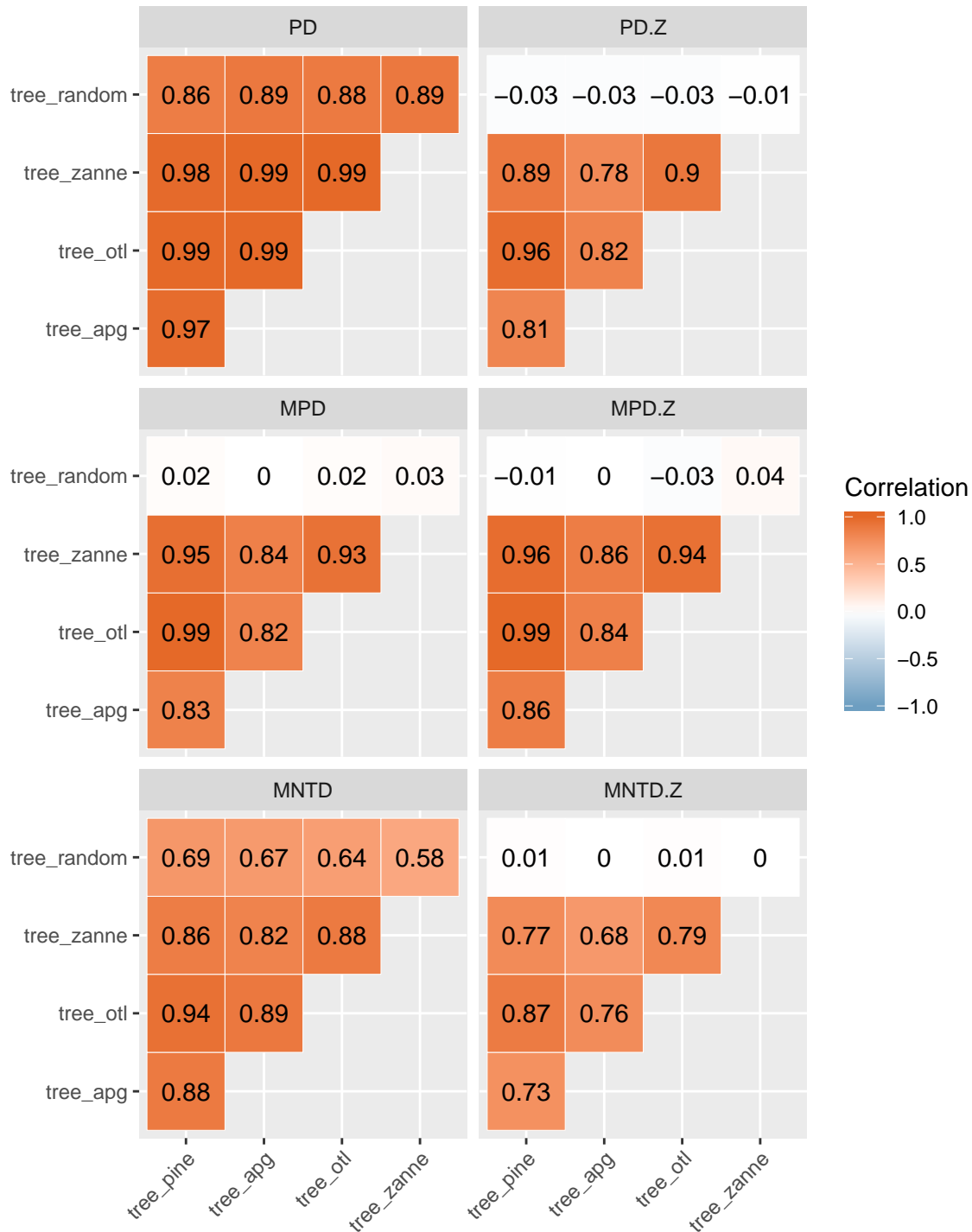


FIGURE S1: Median correlations of phylogenetic alpha diversity values based on 100 simulations of the pine rockland dataset. Within each simulation, different sites have *different* species richness. Results are similar to those in the main text, where species richness was the same across all sites within each simulation to remove effects of species richness. Note that PD and MNTD (but not MPD) are correlated with species richness, thus the high correlations between tree\_random and other phylogenies. Null models successfully removed these correlations (PD.Z and MNTD.Z).

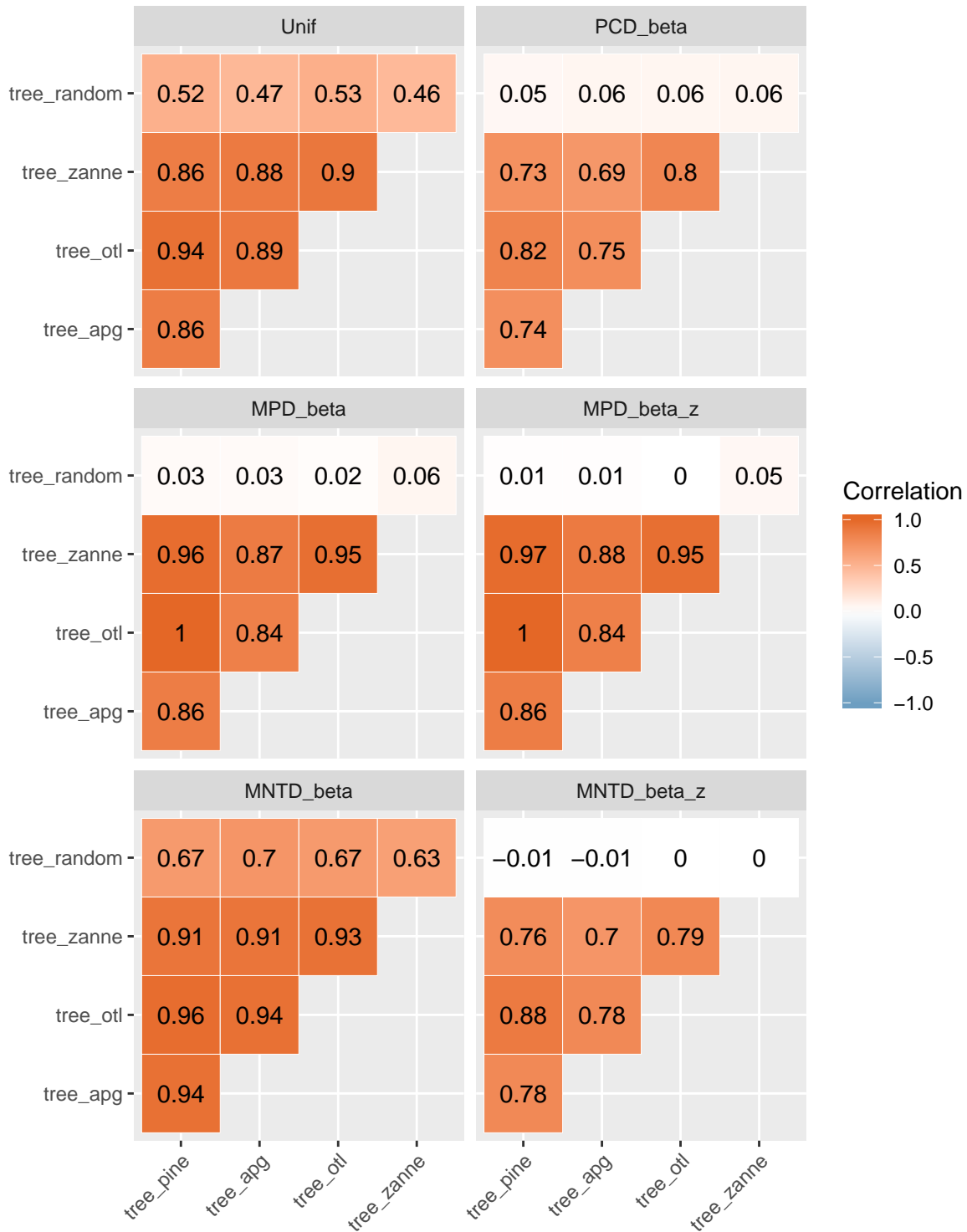


FIGURE S2: Median correlations of phylogenetic beta diversity values based on 100 simulations of the pine rockland dataset. Within each simulation, different sites have *different* species richness. Results are similar to those in the main text, where species richness was the same across all sites within each simulation to remove effects of species richness. Note that Unif and MNTD\_beta (but not MPD\_beta and PCD\_beta) are correlated with species beta diversity, thus the high correlations between tree\_random and other phylogenies. Null models successfully removed these correlations (MPD\_beta\_z and MNTD\_beta\_z).

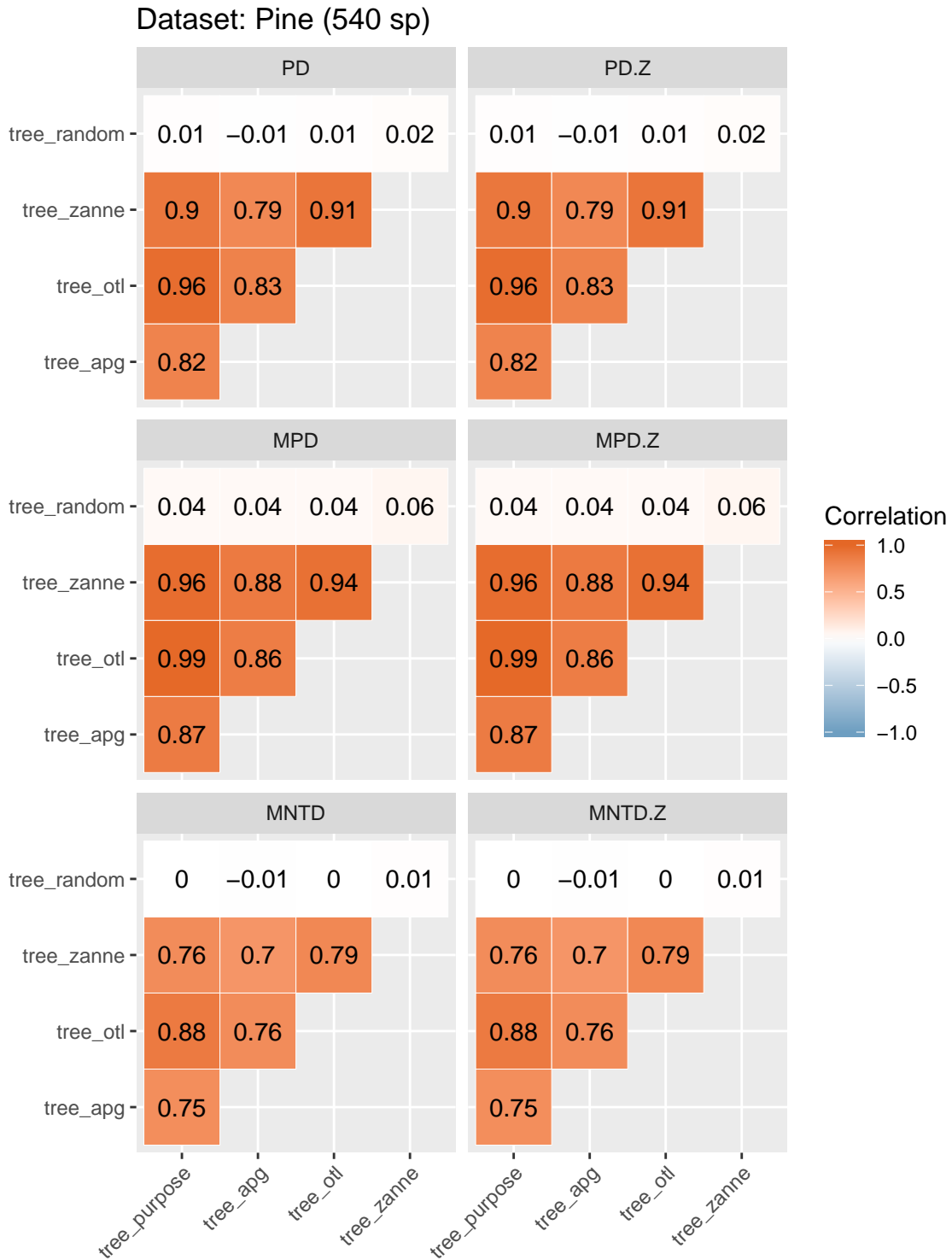


FIGURE S3: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the pine rockland dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

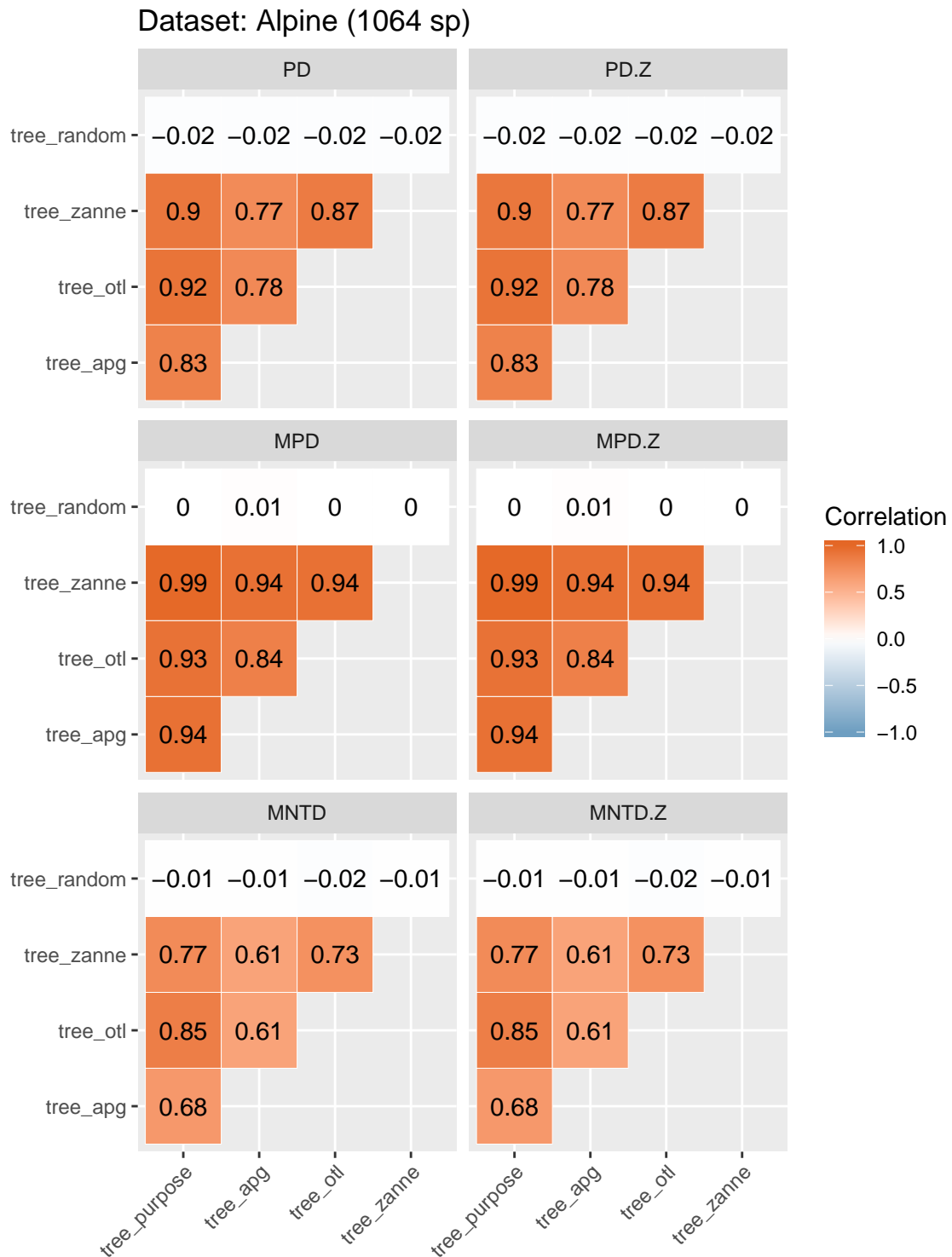


FIGURE S4: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the alpine dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

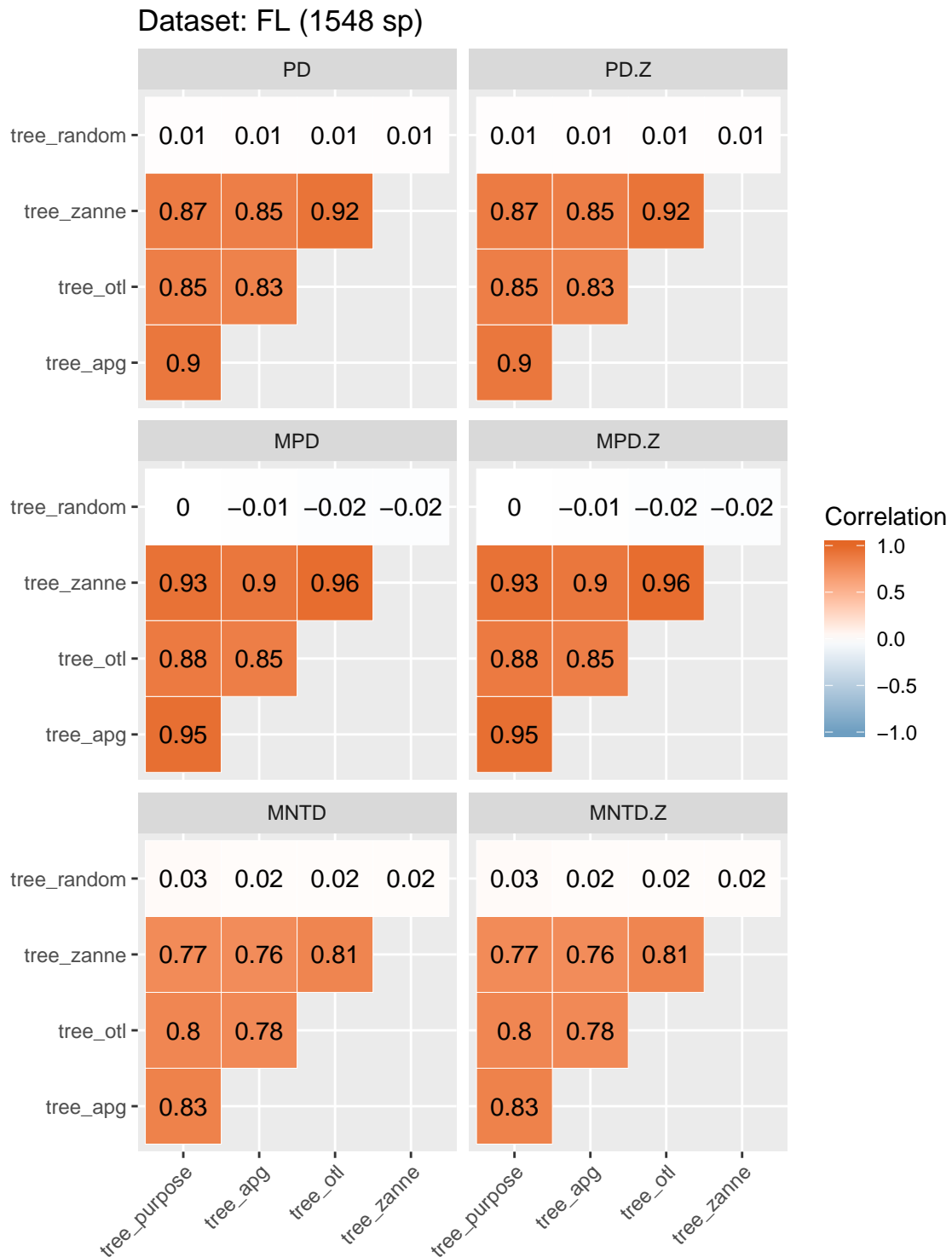


FIGURE S5: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the Florida flora dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

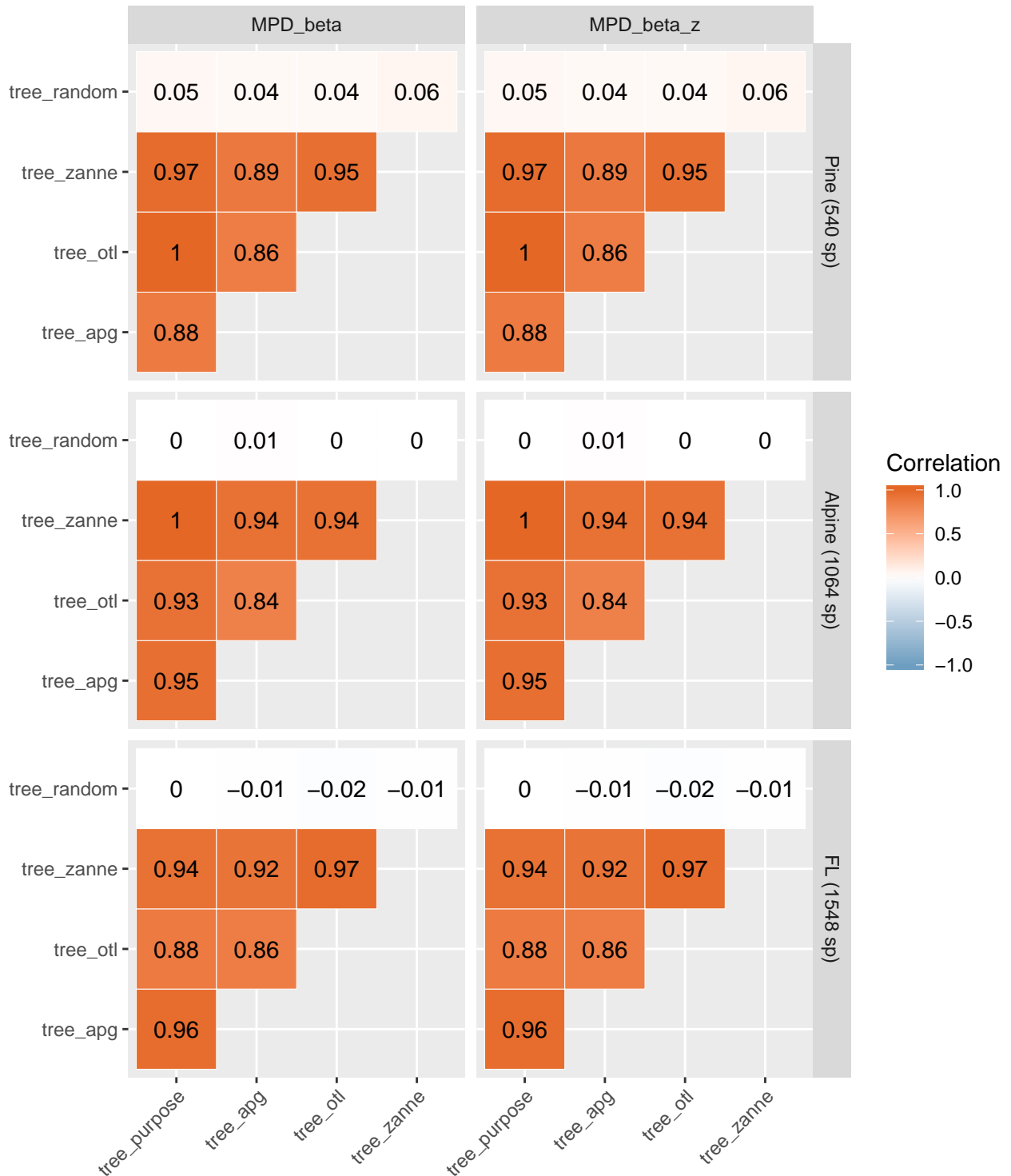


FIGURE S6: Correlations among observed phylogenetic beta diversity (MPD\_beta) values calculated from different phylogenies for all datasets are the same as correlations among their corresponding standardized effect size (SES, MPD\_beta\_z) calculated from different phylogenies.

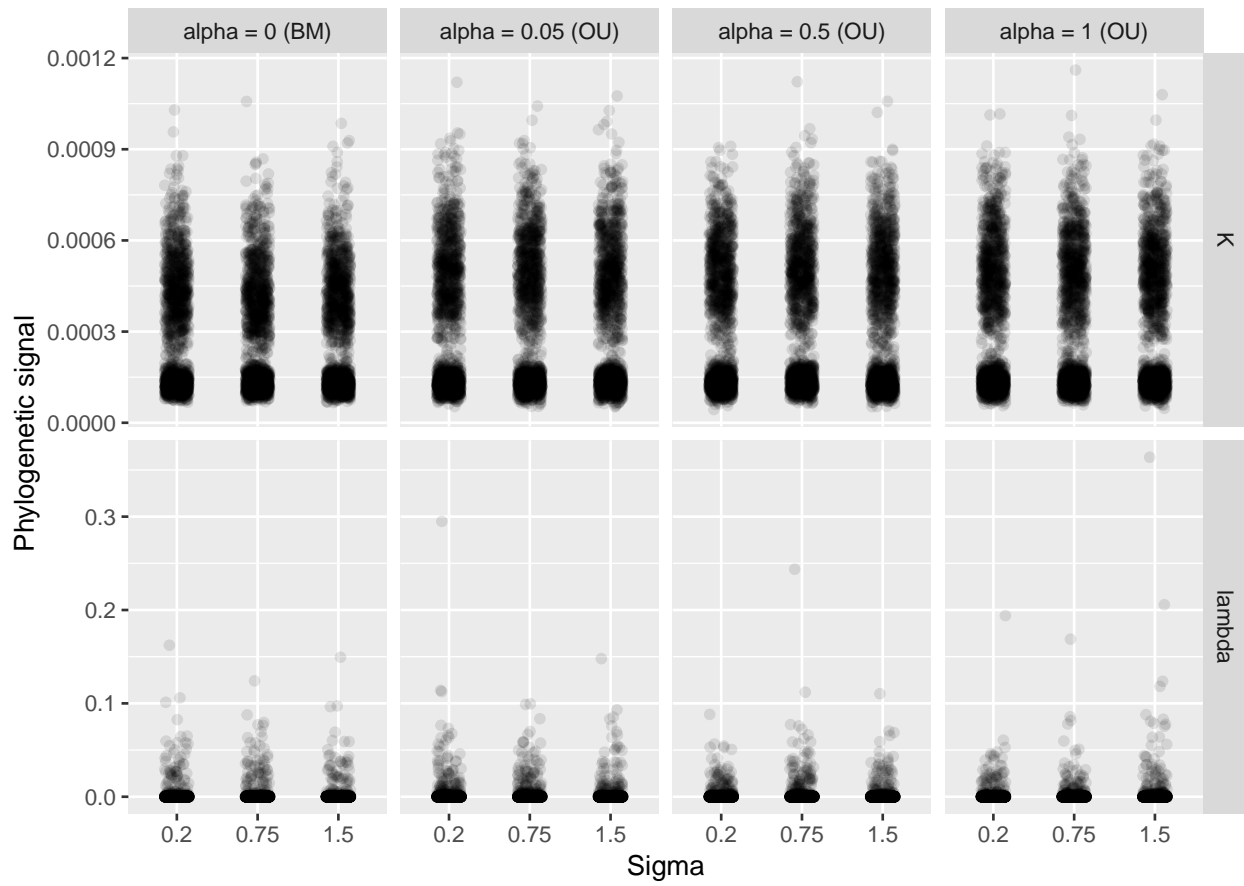


FIGURE S7: Estimated phylogenetic signal values based on tree\_random were all around zero as expected.

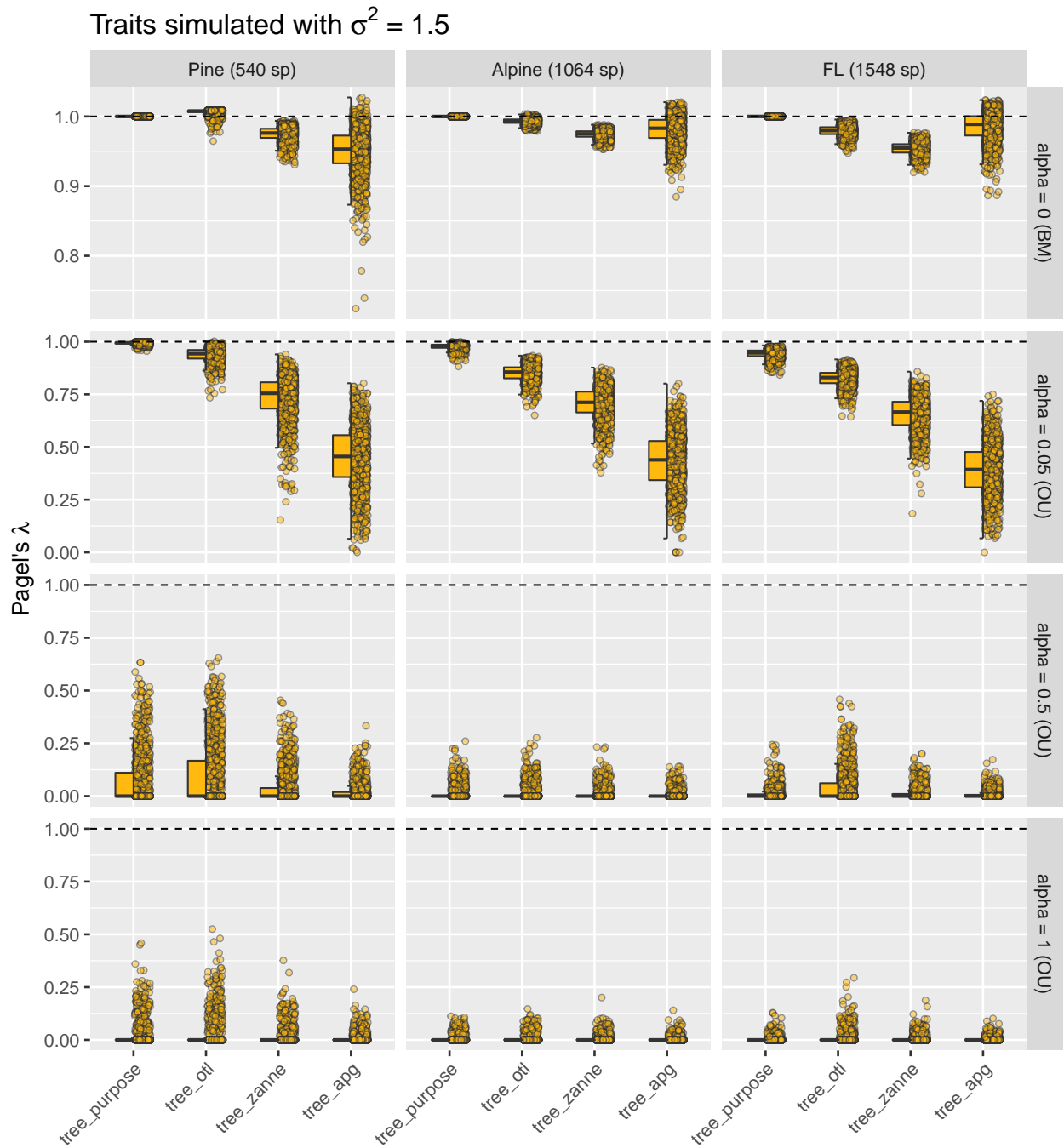


FIGURE S8: Estimated Pagel's  $\lambda$  when traits were simulated with divergence rate  $\sigma^2 = 1.5$ . The patterns here were similar as those in Fig. 4 in the main text, suggesting that sigma did not affect the estimates of Pagel's  $\lambda$ .



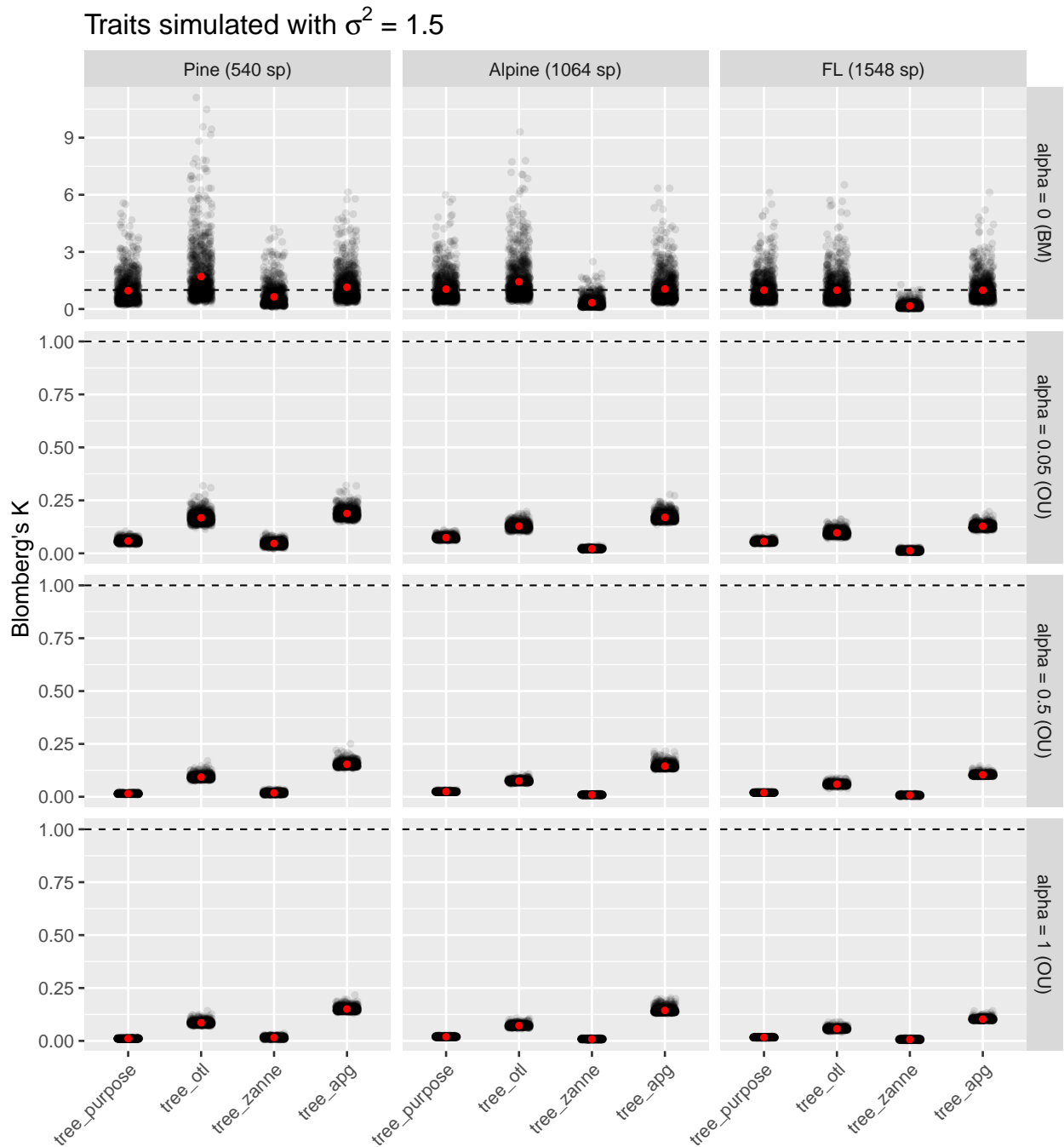


FIGURE S9: Estimated Blomberg's K when traits were simulated with divergence rate  $\sigma^2 = 1.5$ . The patterns here were similar as those in Fig. 5 in the main text, suggesting that sigma did not affect the estimates of Blomberg's K.

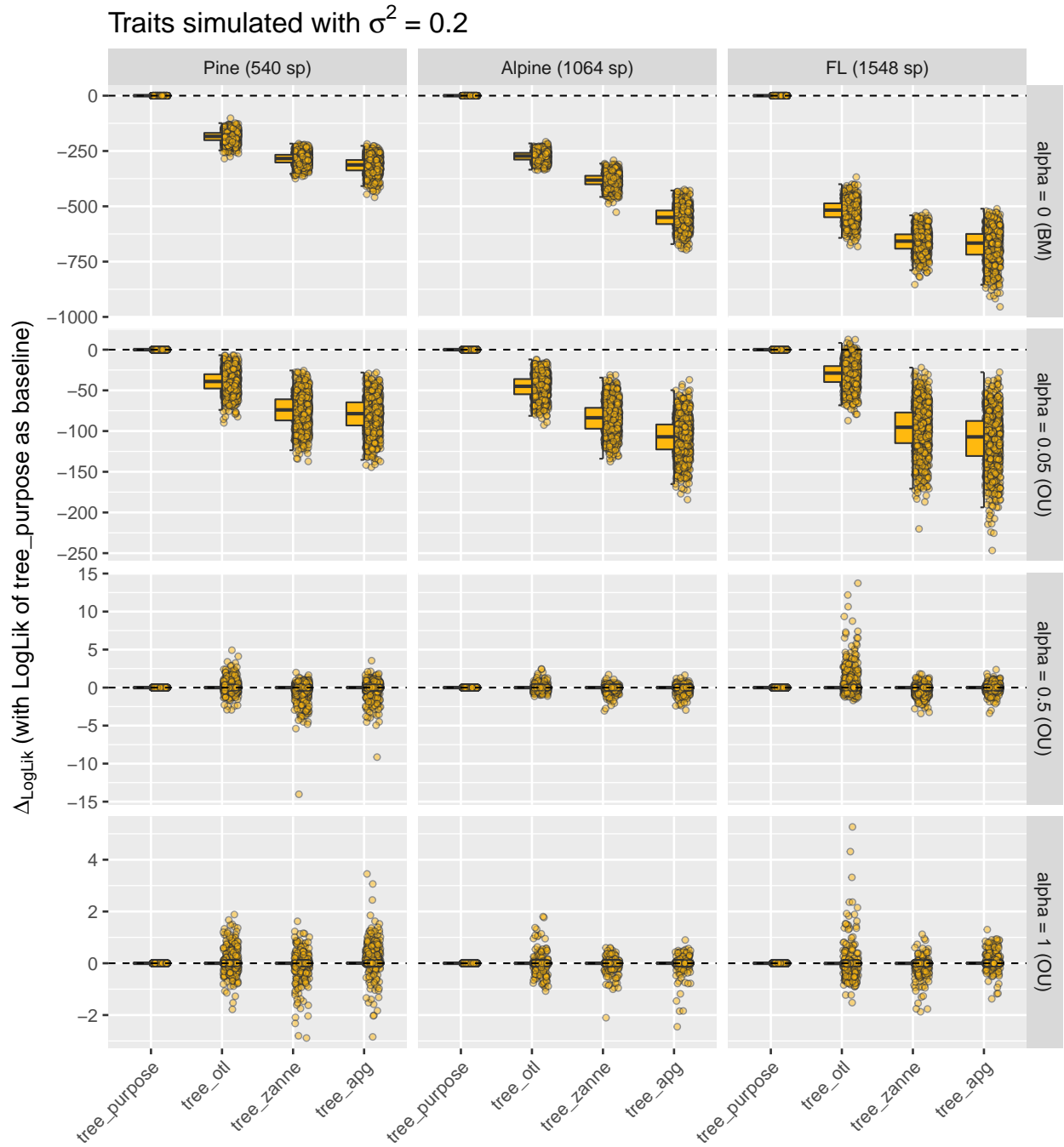


FIGURE S10: Differences in Log likelihood (LogLik) between tree\_purpose and other synthesis phylogenies when estimating phylogenetic signal with Pagel's  $\lambda$ . LogLik values of tree\_purpose were used as the baselines. tree\_otl has the highest LogLik among all synthesis phylogenies when traits were simulated under BM ( $\alpha = 0$ ) or weak OU ( $\alpha = 0.05$ ). When traits were simulated under strong OU ( $\alpha = 0.5$  and  $1$ ), all phylogenies had similar LogLik values that were not different from those of null models (i.e., no phylogenetic signal). Different divergence rates ( $\sigma^2$ ) had the pattern thus we only plotted  $\sigma^2 = 0.2$  here.

TABLE S1: The proportion of significant estimates ( $p < 0.05$ ) of phylogenetic signal (Pagel’s  $\lambda$  and Blomberg’s K) for traits simulated under different models of evolution with  $\sigma^2 = 0.2$ : BM (alpha = 0), OU (alpha = 0.05, 0.5, and 1), and random (i.e., alpha = Infinity). For traits simulated under BM, the proportion of significant results quantifies statistical power. For traits simulated under OU, the proportion of significant results quantifies statistical power in detecting less-than-BM signals (note the much higher power of K than  $\lambda$  when alpha = 0.5 and 1 for tree\_purpose). For traits simulated randomly, the proportion of significant results quantifies type I error rate (i.e., false positive). Abbreviations: apg, tree\_apg; otl, tree\_otl; purpose, tree\_purpose; zanne, tree\_zanne.

data_set	true_alpha	$\lambda$				K			
		apg	otl	purpose	zanne	apg	otl	purpose	zanne
Pine (540 sp)	0.00	1	1	1	1	1	1	1	1
	0.05	0.977	1	1	0.991	0.876	1	1	0.999
	0.50	0.070	0.115	0.100	0.046	0.073	0.342	1	0.196
	1.00	0.025	0.029	0.027	0.016	0.064	0.152	0.994	0.105
	Inf	0.004	0.002	0.003	0.003	0.038	0.045	0.036	0.042
Alpine (1064 sp)	0.00	1	1	1	1	1	1	1	1
	0.05	0.988	1	1	1	0.486	1	1	1
	0.50	0.010	0.022	0.020	0.009	0.053	0.136	1	0.213
	1.00	0.001	0.010	0.006	0.005	0.056	0.073	1	0.084
	Inf	0.002	0.000	0.001	0.001	0.043	0.042	0.048	0.049
FL (1548 sp)	0.00	1	1	1	1	1	1	1	1
	0.05	0.998	1	1	1	0.965	0.998	1	0.585
	0.50	0.069	0.123	0.077	0.059	0.071	0.107	0.998	0.066
	1.00	0.016	0.023	0.017	0.009	0.051	0.053	0.631	0.074
	Inf	0.005	0.002	0.005	0.004	0.031	0.044	0.051	0.051

## References

Allen, J. M., C. C. Germain-Aubrey, N. Barve, K. M. Neubig, L. C. Majure, S. W. Laffan, B. D. Mishler, H. L. Owens, S. A. Smith, W. M. Whitten, J. R. Abbott, D. E. Soltis, R. Guralnick, and P. S. Soltis. 2019. Spatial phylogenetics of florida vascular plants: The effects of calibration and uncertainty on diversity estimates. *iScience* 11:57–70.

APG IV, M. W. Chase, M. J. M. Christenhusz, M. F. Fay, J. W. Byng, W. S. Judd, D. E. Soltis, D. J. Mabberley, A. N. Sennikov, P. S. Soltis, and P. F. Stevens. 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.

Bell, C. D., D. E. Soltis, and P. S. Soltis. 2010. The age and diversification of the angiosperms re-revisited. *American journal of botany* 97:1296–1303.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS computational biology* 10:e1003537.

Eastman, J. M., L. J. Harmon, and D. C. Tank. 2013. Congruification: Support for time scaling large phylogenetic trees. *Methods in Ecology and Evolution* 4:688–691.

Gann GD, S. C. 2018. Floristic inventory of the florida keys database online.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution* 30:772–780.

Kress, W. J., D. L. Erickson, F. A. Jones, N. G. Swenson, R. Perez, O. Sanjur, and E. Bermingham. 2009. Plant dna barcodes and a community phylogeny of a tropical forest dynamics plot in panama. *Proceedings of the National Academy of Sciences* 106:18621–18626.

Lanfear, R., B. Calcott, S. Y. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of

partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution* 29:1695–1701.

Magallón, S., K. W. Hilu, and D. Quandt. 2013. Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* 100:556–573.

Marx, H. E., C. Dentant, J. Renaud, R. Delunel, D. C. Tank, and S. Lavergne. 2017. Riders in the sky (islands): Using a mega-phylogenetic approach to understand plant species distribution and coexistence at the altitudinal limits of angiosperm plant life. *Journal of biogeography* 44:2618–2630.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 61:539–542.

Sanderson, M. J. 2003. R8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.

Schneider, H., A. R. Smith, P. Hovenkamp, J. Prado, G. Rouhan, A. Salino, M. Sundue, T. E. Almeida, B. Parris<sup>10</sup>, E. B. Sessa<sup>11</sup>, and others. 2016. A community-derived classification for extant lycophytes and ferns 54:563–603.

Schuettpelz, E., and K. M. Pryer. 2009. Evidence for a cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proceedings of the National Academy of Sciences* 106:11200–11205.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. *BMC evolutionary biology* 9:37.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24:715–716.

Smith, S. A., and B. C. O’Meara. 2012. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28:2689–2690.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, J. B. Walker, M. J. Moore, B. S. Carlsward, and others. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American journal of botany* 98:704–730.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research* 34:W609–W612.

Trotta, L., B. Baiser, J. Possley, D. Li, J. Lange, S. Martin, and E. Sessa. 2018. Community phylogeny of the globally critically imperiled pine rockland ecosystem. *American journal of botany* 105:1735–1747.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlenn, B. C. O’Meara, A. T. Moles, P. B. Reich, and others. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506:89–92.