Appendix

27 September, 2018

Text A1: 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 rockalnd 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 (Katoh 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 determined 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. (*in revision*).

Figures

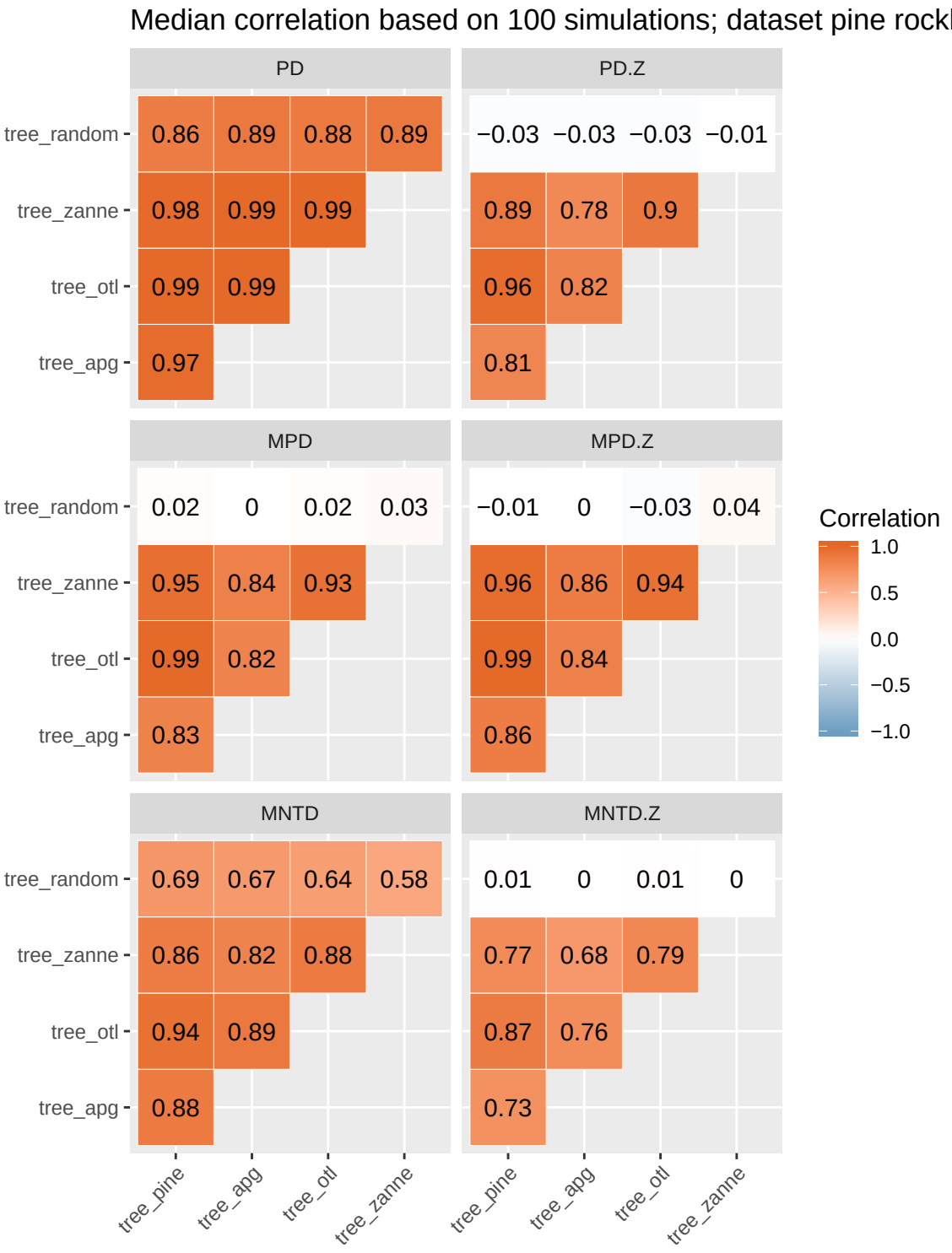

FIGURE **A1**: Median correlations of phylogenetic alpha diversity values based on 100 simulations of the pine rockland dataset. Within each simulation, different sites have *different* species richness. Results are similar to those in the main text, where species richness was the same across all sites 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`).

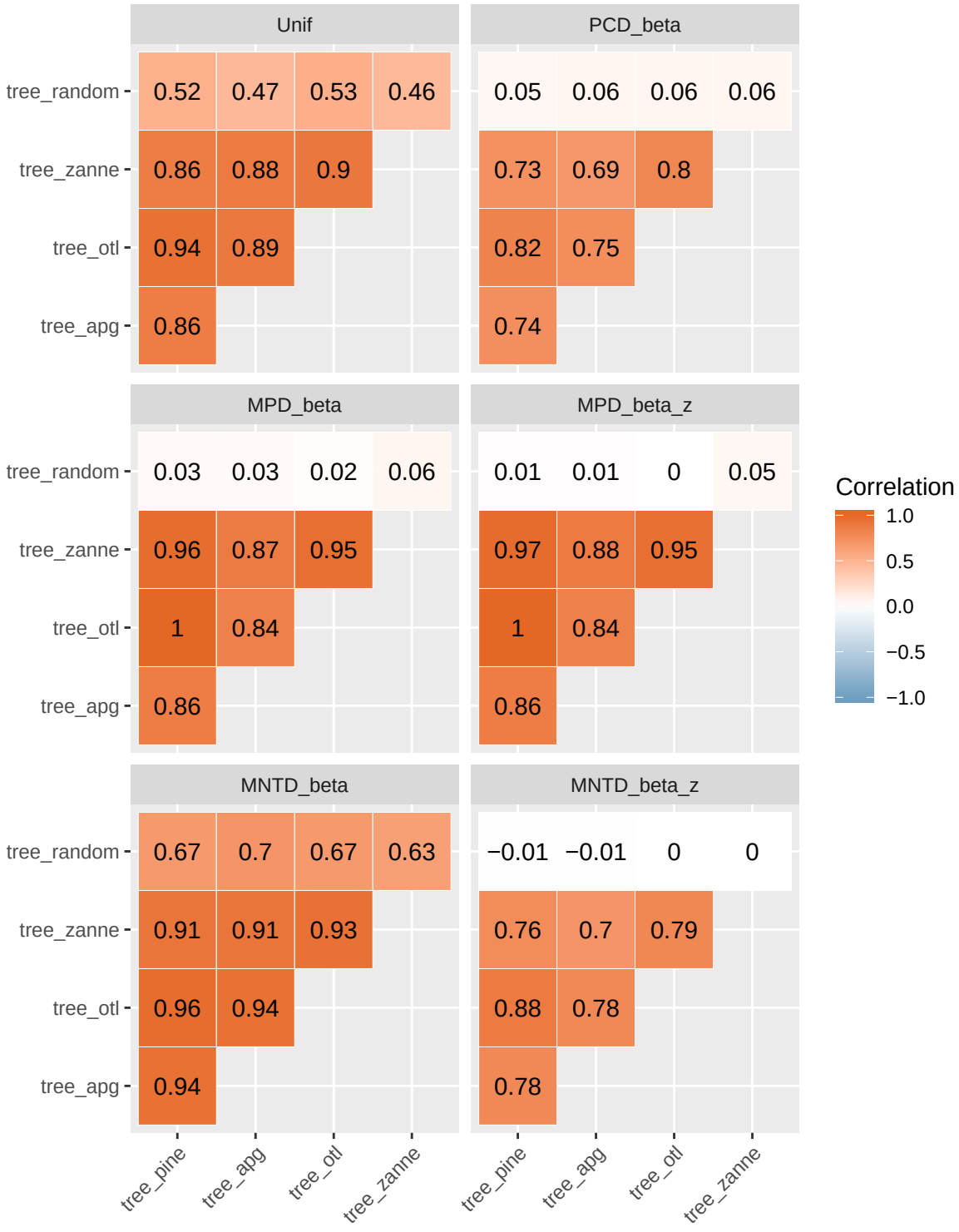

FIGURE **A2**: Median correlations of phylogenetic beta diversity values based on 100 simulations of the pine rockland dataset. Within each simulation, different sites have *different* species richness. Results are similar to those in the main text, where species richness was the same across all sites 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`).

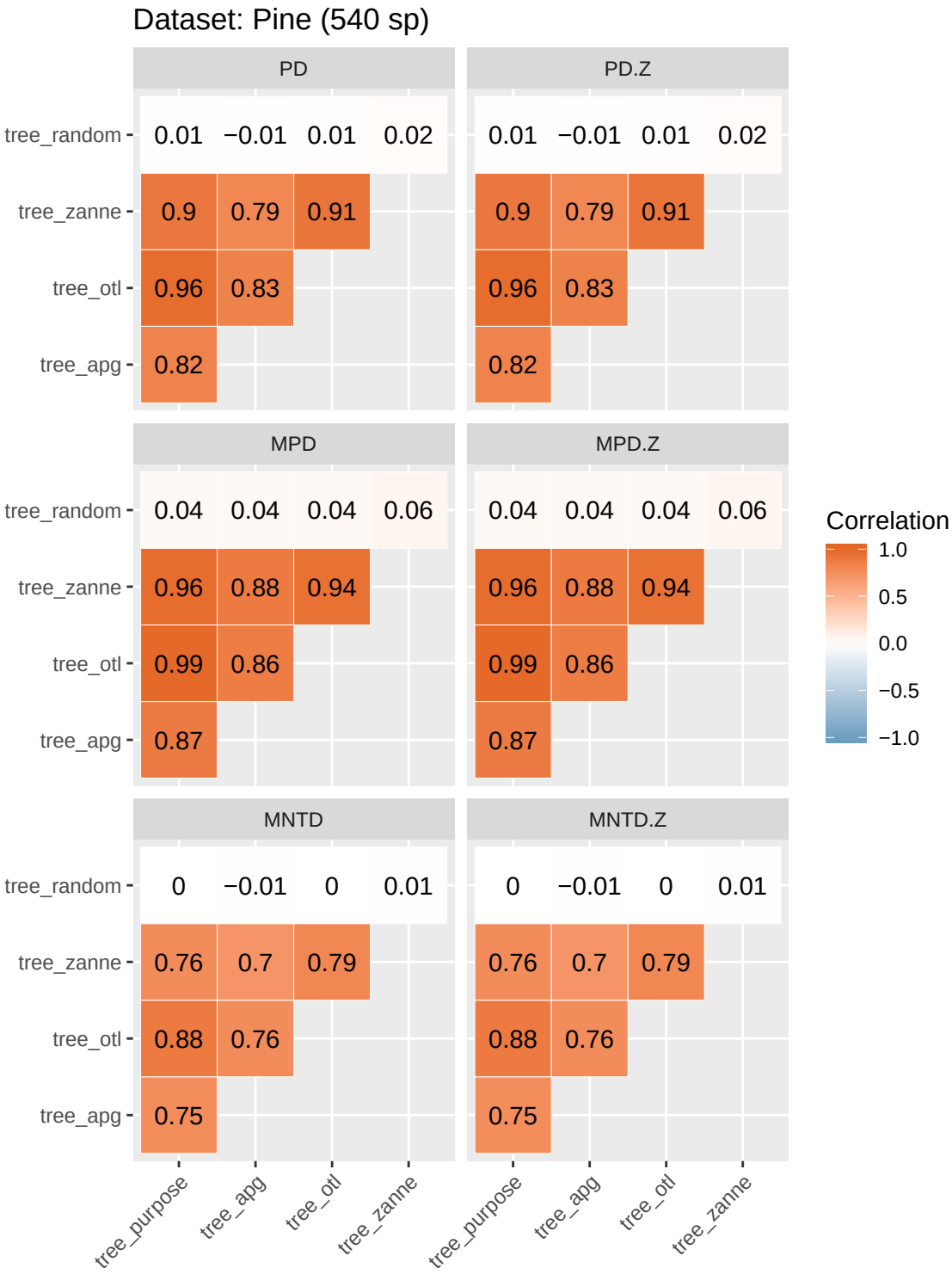

FIGURE **A3**: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the pine rockland dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

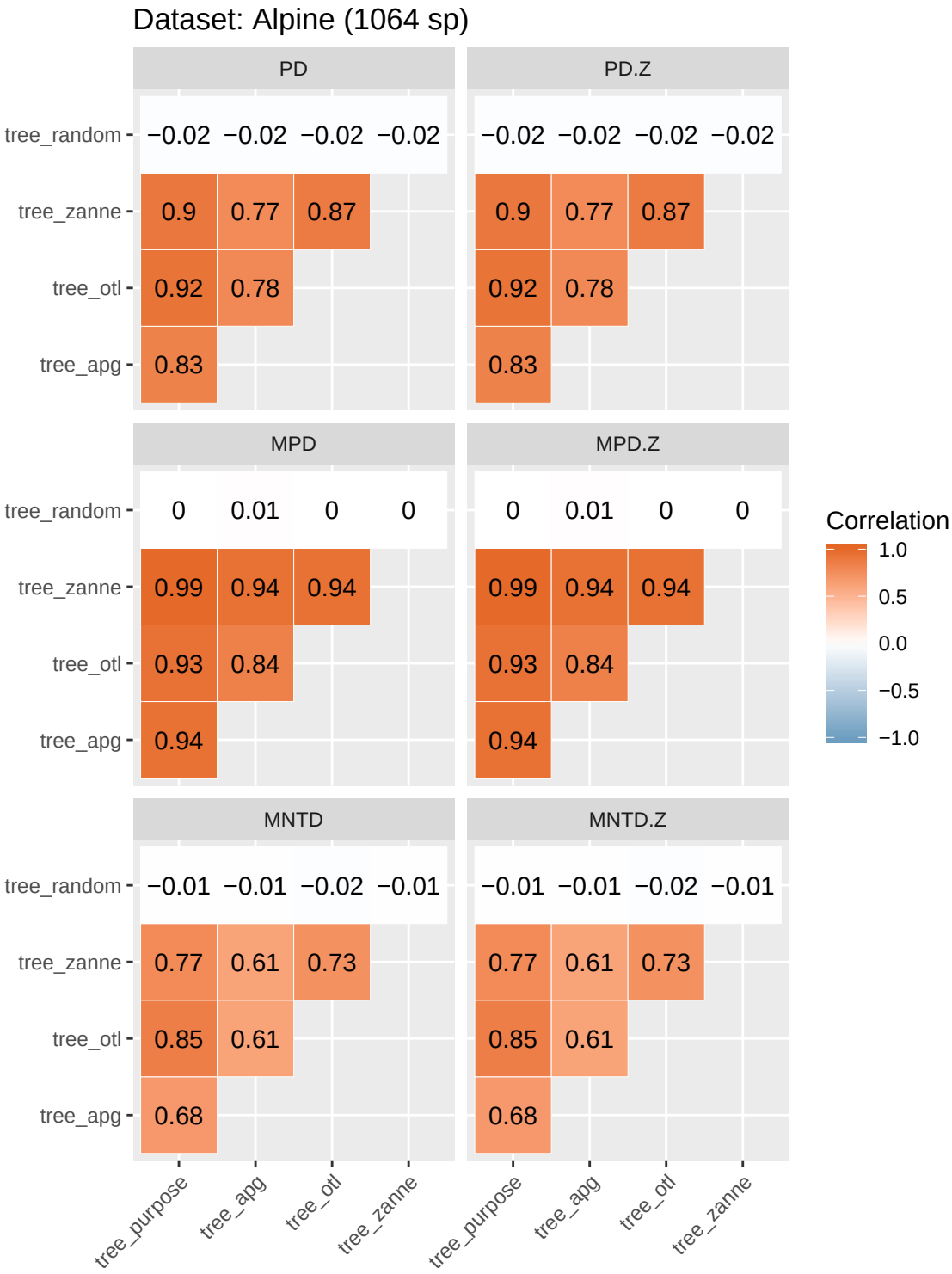

FIGURE A4: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the alpine dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

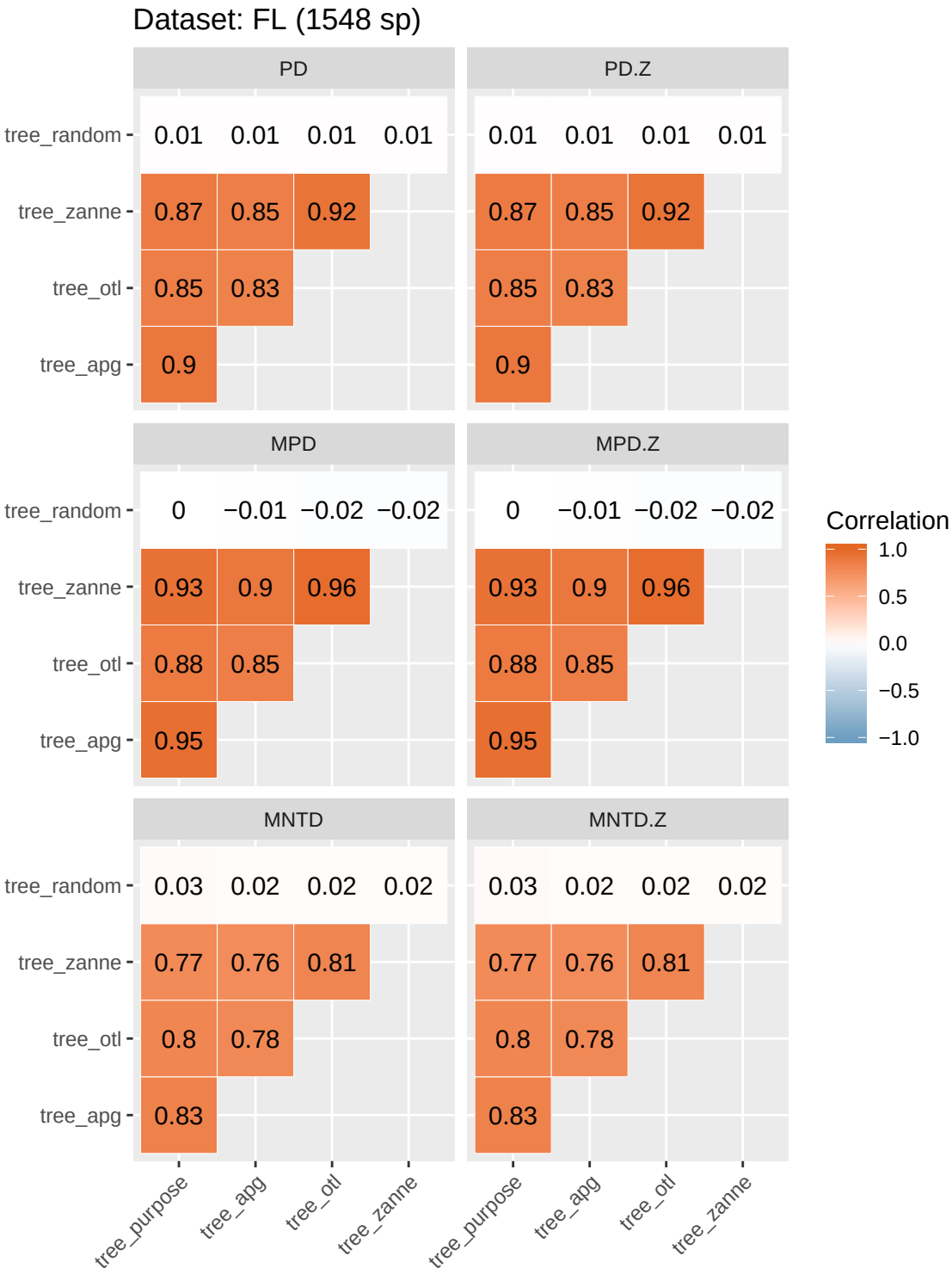

FIGURE **A5**: Correlations among observed phylogenetic alpha diversity values calculated from different phylogenies for the Florida flora dataset are the same as correlations among their corresponding standardized effect size (SES) calculated from different phylogenies.

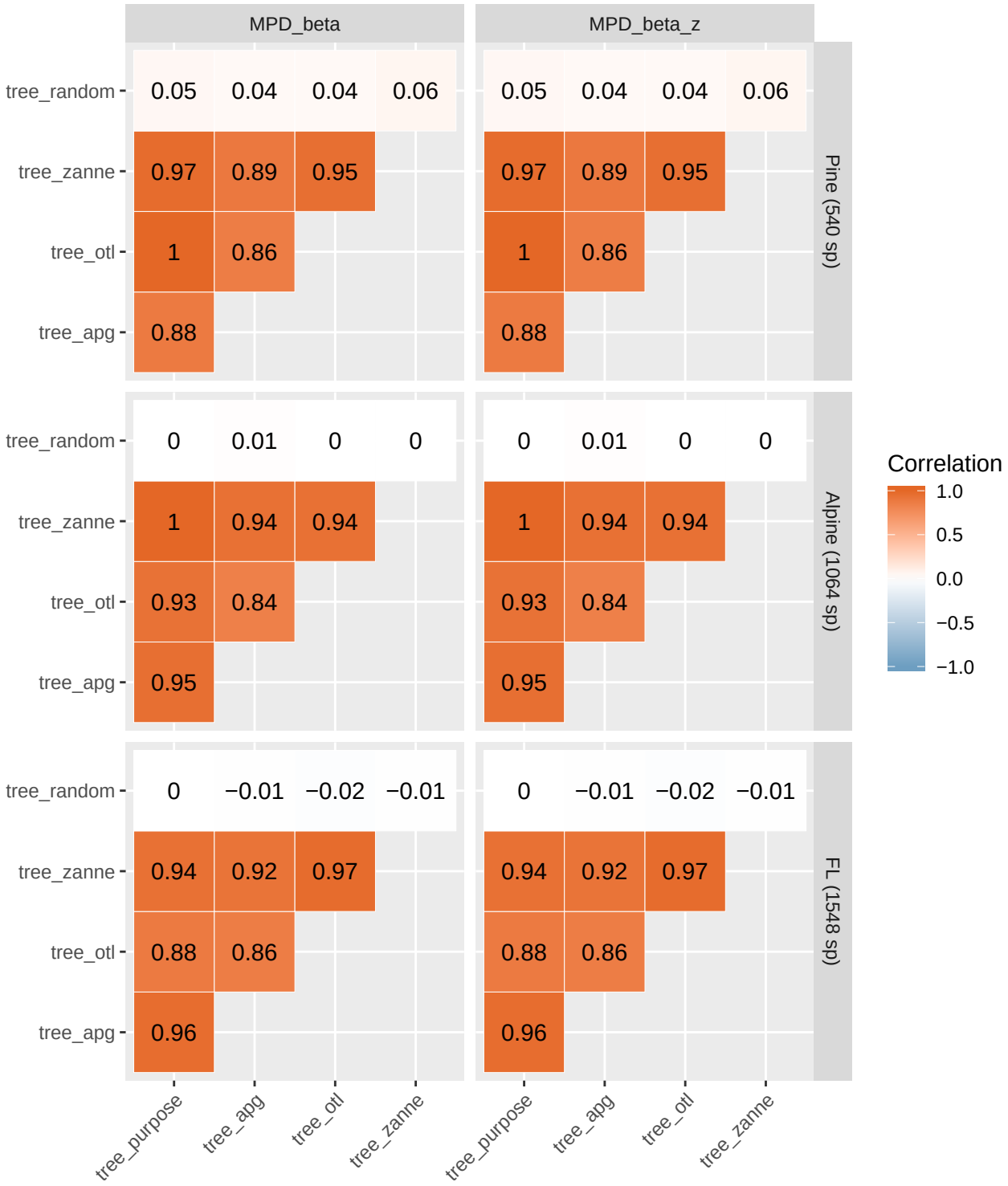

FIGURE **A6**: Correlations among observed phylogenetic beta diversity (`MPD_beta`) values calculated from different phylogenies for all datasets are the same as correlations among their corresponding standardized effect size (SES, `MPD_beta_z`) calculated from different phylogenies.

References

APG IV et al. 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. et al. 2010. The age and diversification of the angiosperms re-revisited. - American journal of botany 97: 1296–1303.

Bouckaert, R. et al. 2014. BEAST 2: A software platform for bayesian evolutionary analysis. - PLoS computational biology 10: e1003537.

Eastman, J. M. et al. 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. in press.

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

Kress, W. J. et al. 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. et al. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. - Molecular biology and evolution 29: 1695–1701.

Magallón, S. et al. 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. et al. 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. et al. 2004. APE: Analyses of phylogenetics and evolution in r language. - Bioinformatics 20: 289–290.

Pennell, M. W. et al. 2014. Geiger v2. 0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. - Bioinformatics 30: 2216–2218.

Ronquist, F. et al. 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. et al. 2016. A community-derived classification for extant lycophytes and ferns. 54: 563–603.

Schuettpelz, E. and Pryer, K. M. 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. and Dunn, C. W. 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. - Bioinformatics 24: 715–716.

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

Smith, S. A. et al. 2009. Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. - BMC evolutionary biology 9: 37.

Soltis, D. E. et al. 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.

125 Suyama, M. et al. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the
126 corresponding codon alignments. - Nucleic acids research 34: W609–W612.

127 Trotta, L. et al. 2018. Community phylogeny of the globally critically imperiled pine rockland
128 ecosystem. - American journal of botany in press.

129 Zanne, A. E. et al. 2014. Three keys to the radiation of angiosperms into freezing environments. -
130 Nature 506: 89–92.