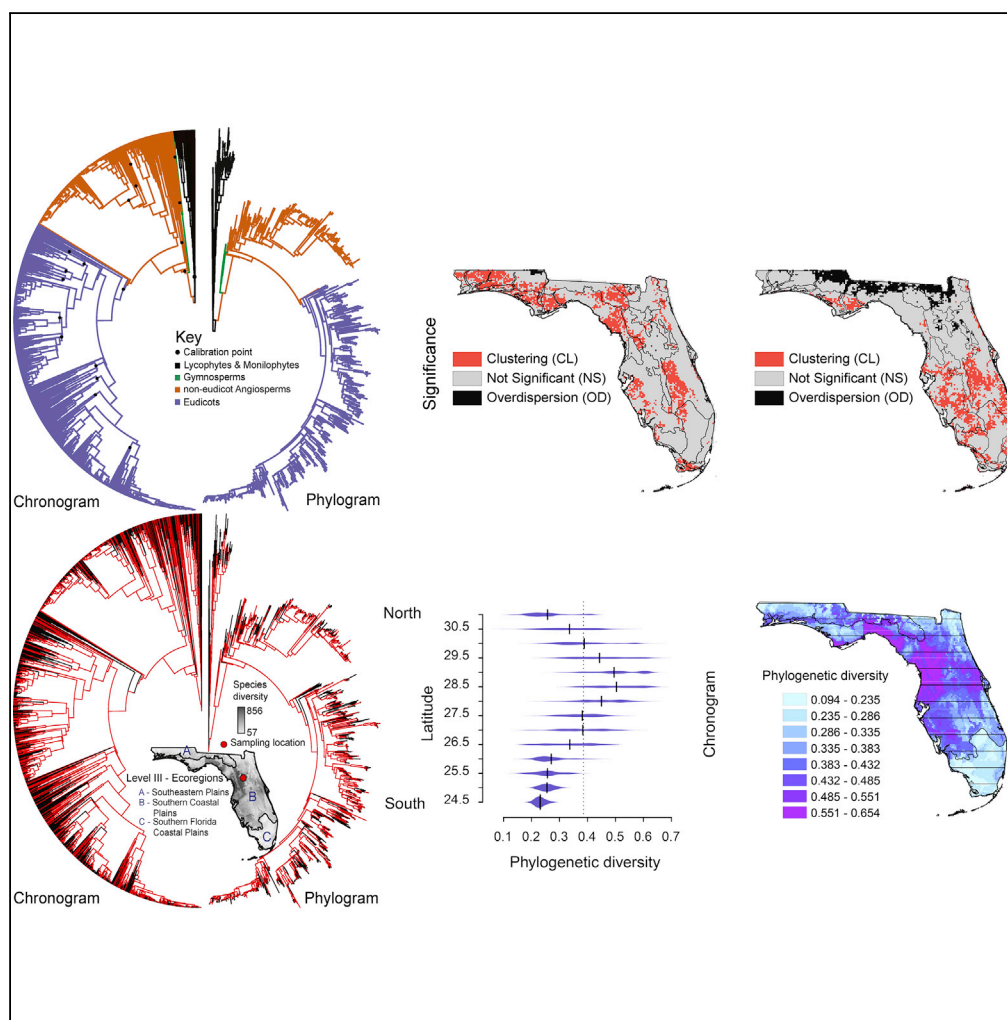**Article**

Spatial Phylogenetics of Florida Vascular Plants: The Effects of Calibration and Uncertainty on Diversity Estimates

Julie M. Allen,
Charlotte C.
Germain-Aubrey,
Narayani Barve, ...,
Douglas E. Soltis,
Robert Guralnick,
Pamela S. Soltis

jallen23@unr.edu

HIGHLIGHTS

Phylodiversity patterns of Florida show peaks of diversity in Northern peninsular areas

Different input trees and uncertainty had modest or little influence on phylodiversity

Chronogram or phylogram had a significant effect on the interpretation of phylodiversity

We highlight areas of high phylodiversity and how these results affect conservation

iScience

Article

Spatial Phylogenetics of Florida Vascular Plants: The Effects of Calibration and Uncertainty on Diversity Estimates

Julie M. Allen,[1,2,13,15,*] Charlotte C. Germain-Aubrey,[2,13] Narayani Barve,[2] Kurt M. Neubig,[3] Lucas C. Majure,[2,4] Shawn W. Laffan,[5] Brent D. Mishler,[6] Hannah L. Owens,[2,7] Stephen A. Smith,[8] W. Mark Whitten,[2] J. Richard Abbott,[2,9] Douglas E. Soltis,[2,10,11,12] Robert Guralnick,[2,12,14] and Pamela S. Soltis[2,10,12,14]

SUMMARY

Recent availability of biodiversity data resources has enabled an unprecedented ability to estimate phylogenetically based biodiversity metrics over broad scales. Such approaches elucidate ecological and evolutionary processes yielding a biota and help guide conservation efforts. However, the choice of appropriate phylogenetic resources and underlying input data uncertainties may affect interpretation. Here, we address how differences among phylogenetic source trees and levels of phylogenetic uncertainty affect these metrics and test existing hypotheses regarding geographic biodiversity patterns across the diverse vascular plant flora of Florida, US. Ecological niche models for 1,490 Florida species were combined with a "purpose-built" phylogenetic tree (phylogram and chronogram), as well as with trees derived from community resources (Phylomatic and Open Tree of Life). There were only modest differences in phylodiversity metrics given the phylogenetic source tree and taking into account the level of phylogenetic uncertainty; we identify similar areas of conservation interest across Florida regardless of the method used.

INTRODUCTION

The recent explosion of biodiversity data (spatial and genetic) along with environmental data (regarding climate, terrain, and vegetation), along with novel analytical methods and tools, has enabled an unprecedented capability to model species distributions and assemble those results into broad-scale diversity assessments (e.g., Tittensor et al., 2010; Ezard et al., 2011; Olalla-Tárraga et al., 2011; Nagalingum et al., 2015). Linking spatial ecological patterns to phylogenetic information is more powerful still (Mishler et al., 2014; Nagalingum et al., 2015) given that species assemblages encompassing deeper phylogenetic nodes and more evolutionary history are arguably more diverse than other areas with the same number of species connected via shallower nodes (Faith, 1992). Phylogenetic approaches extend diversity measurements from simplistic species counts to measures that also inform evolutionary pattern and process.

One of the key measures in spatial phylogenetics is phylogenetic diversity (PD; Faith, 1992). PD is calculated as the sum of branch lengths from a phylogenetic tree connecting the terminal taxa from a specific location, typically to the root of the tree. PD can be interpreted either as the amount of "feature diversity" contained within a region of interest when using a phylogram, i.e., the number of apomorphies present in an area, or as the amount of "evolutionary history" when using a time-calibrated chronogram (Davies and Buckley, 2012; Rosauer, 2010). Those regions with higher PD than others may be prioritized for conservation (i.e., as containing higher genetic diversity or a greater amount of evolutionary history), although there are obviously other potential criteria, such as threat status, that should be applied in conservation assessments (Jetz and Freckleton, 2015). PD is typically strongly correlated with species richness, because more terminal taxa in a sample means that a larger portion of the tree is expected to be sampled. Mishler et al. (2014) developed a compound spatial phylogenetic metric, Relative Phylogenetic Diversity (RPD), designed to examine whether unusually long or unusually short branches are present in a location. PD and RPD measures along with associated randomization tests can help elucidate the evolutionary processes that have generated biotas, which in turn support stronger assessments of conservation priorities.

The evolutionary trees used in spatial phylogenetic studies are often not built by the authors performing the study, and tree building is necessarily not afforded the same level of scrutiny as analysis of spatial

[1]Department of Biology, University of Nevada Reno, Reno, NV 89557, USA

[2]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

[3]Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901, USA

[4]Department of Research, Conservation and Collections, Desert Botanical Garden, Phoenix, AZ 85008, USA

[5]School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, Australia

[6]University and Jepson Herbaria, and Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720-2465, USA

[7]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, 2100 Copenhagen O, Denmark

[8]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48103, USA

[9]School of Mathematical and Natural Sciences, The University of Arkansas at Monticello, Monticello, AR 71655, USA

[10]Genetics Institute, University of Florida, Gainesville, FL 32608, USA

[11]Department of Biology, University of Florida, Gainesville, FL 32611, USA

[12]Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA

ecological data, despite the critical importance of trees for rigorous inference. Instead, ecologists have often relied on (1) converting a taxonomic hierarchy directly into a tree (e.g., Davies et al., 2007), (2) shortcut trees constructed for focal species via Phylomatic software (e.g., Webb and Donoghue, 2005; Webb et al., 2008; Wright et al., 2007; Liu et al., 2013) or the Open Tree of Life (OTL; Hinchliff et al., 2015), (3) automated assembly of published sequences such as PhyloGenerator (Pearse and Purvis, 2013), (4) literature-based trees (e.g., Beaulieu et al., 2012), or (5) framework trees vetted by the phylogenetics community (e.g., The Angiosperm Phylogeny Group IV, 2016; Soltis et al., 2011).

Despite the relative ease of acquiring such trees, their quality and inherent uncertainties are rarely examined, and the impact of these factors on PD assessment has not been well studied (but see Qian et al., 2015; Swenson, 2009; Molina-Venegas and Roquet, 2013; Rangel et al., 2015; Thornhill et al., 2017). There remains a need to better document how factors involved in constructing phylogenetic trees (e.g., phylogenetic uncertainty and taxon sampling) influence these metrics. Both tree topology and branch lengths are determined by the sampling of taxa and the gene sequences employed, and these factors must be considered when computing and interpreting PD measures. For example, limited taxonomic sampling from a tree will produce longer individual branches than are truly present, whereas limited sampling of genetic data may result in unrepresentative branch lengths. Likewise, the use of phylograms versus chronograms yields branch length differences and therefore different values for PD measures. The phylogenetic depth over which trees are computed will also affect the magnitude of these metrics: older clades have longer branches in a chronogram and therefore contribute to higher estimates of PD than younger clades. Finally, failure to account for tree uncertainty might inflate the confidence in a given result. This last issue is particularly underexplored, but crucial for interpreting PD values.

Here we provide a comprehensive examination of how the choice of input phylogenetic trees and inclusion of phylogenetic uncertainty affect the assessment of PD measures, utilizing Florida vascular plants as a case study. To test the importance of input trees, we developed phylogenetic trees for the specific purpose of estimating biodiversity through integration with distribution models. A key rationale for doing so was to determine if a more comprehensive, well-developed, and purpose-built phylogenetic tree would yield different estimates of PD relative to those built from easily available and existing trees obtained, for example, using Phylomatic (Webb and Donoghue, 2005), or by pruning a subtree from a pre-assembled supertree (i.e., the OTL; Hinchliff et al., 2015).

We chose Florida as the focus for study because it is home to approximately 4,300 species of native or naturalized vascular plants and a broad range of terrestrial and aquatic habitats (Wunderlin et al., 2017). Furthermore, Florida is part of the North American Coastal Plain biodiversity hotspot (Noss et al., 2015). Florida's flora ranges from temperate, eastern deciduous forest taxa in the north to tropical elements in central and southern Florida (Myers and Ewel, 1990); these unique floristic elements mix at transition zones, leading to novel communities that might be expected to have unusual phylogenetic affinities. At the same time, past climatic changes caused inundation of much of the state, forming ancient shorelines, such as the Lake Wales Ridge (LWR), that still harbor an unusual, highly endemic scrub flora and fauna (Dobson et al., 1997). The southern portion of Florida has a subtropical climate and includes unique ecoregions such as the Everglades, Big Cypress and Miami Ridge, and Pine Rocklands, each with characteristic floristic elements (Long and Lakela, 1971). Florida also supports the third highest concentration of federally sensitive, threatened, and endangered species in the United States (Ihlo et al., 2014), after California and Hawaii (Dobson et al., 1997). Furthermore, one-third of the flora of Florida is now composed of exotic species (either naturalized or invasive), and habitat loss due to human development is mounting (Gordon, 1998). Still, despite the magnitude of ecological and conservation concerns in this region, little is known about the overall geographic patterns of plant diversity in Florida.

The present study had both empirical and methodological goals. Our empirical goal was to test hypotheses regarding patterns of Florida biodiversity derived from previous studies of forest types, vertebrates, and butterflies. In particular, work by documented an overall decrease in diversity from north to south in Florida, although this pattern was only assessed qualitatively based on maps of richness from a variety of vertebrates and butterflies. These conversed patterns of diversity, when compared with general latitudinal diversity gradients (Wiens et al., 2009; Buckley et al., 2010), may relate more to the unique transitional zones from temperate to tropical floras in Florida and the underlying climate, soil, and terrain of the region than to temperature. Previous work has noted that transitional areas, such as the Southern Coastal

[13]These authors contributed equally

[14]Co-senior author

[15]Lead Contact

*Correspondence: jallen23@unr.edu

Plain ecoregion and northern peninsular Florida with southern hardwood forests and temperate broad-leaved evergreen forests, harbor particularly high diversity (Greller, 1980). Observed PD patterns in plants should be strongly concordant with previous hypotheses of diversity, but some geographic areas may harbor unexpectedly high areas of PD, such as the Miami Ridge ecoregion and its tropical hammock forest flora (Myers and Ewel, 1990). In such areas, where there may be mixing of floristic elements, we predicted concentrations of significantly overdispersed (e.g., even) lineages (based on PD) and concentrations of un-usually long branches (based on RPD). We further predicted concentrations of significant phylogenetic clustering in areas wherein habitat may select for specific community members (also called "habitat filtering"), as well as significant concentrations of shorter-than-expected branches in areas where lineages have potentially diversified in situ, such as the LWR. Finally, we attempted to contextualize these findings from a conservation perspective, given ongoing rapid anthropogenic changes to native landscapes in Florida.

Our methodological goals were to explore the effect of the choice of phylogenetic tree on spatial phylo-genetic metrics (PD and RPD) and to provide an approach to account more effectively for sources of uncer-tainty in phylogenetic trees. We generated PD and RPD using a variety of input phylogenetic trees and compared the results using multiple approaches to understand how to interpret differences and uncer-tainty in these assessments. We expected greater variation among branch lengths across the tree in chronograms than in phylograms, as branches can often be either greatly lengthened or shortened, reflect-ing constraints of evolutionary time. This difference was predicted to affect the distribution of observed PD, but the impact on significance tests is poorly characterized (but see Thornhill et al., 2017). We also examined how spatial phylogenetic metrics vary between trees pruned from existing supertrees and those inferred from curated analysis where stringent efforts have been made to close gaps in taxon sampling, using a strategic approach for gene sampling and branch length assessments. Finally, we used a Bayesian framework to generate a distribution of trees representing uncertainty in phylogenetic estimates to assess the impacts on PD.

RESULTS

Ecological Niche Models for Generating Species Lists per Pixel

Validation metrics across all models were high, with training Area Under the Curve (AUC) scores of >0.8 and test AUC scores within 0.15 of the training scores in nearly all cases. A small proportion of models had significantly worse performance, wherein the difference between training and testing was >0.5. Such met-rics were often due to low sample sizes; we removed species with outlier AUC scores >3 standard devia-tions away from the mean from our final analysis. Ultimately, we accepted 1,490 models (i.e., one per spe-cies), rejecting 12 species with poor model performance. Figure 1 shows species richness based on stacked models for Florida at a 4-km resolution, ranging from a low of 57 species to a high of 856. We do not focus on taxonomic measures of richness here, but instead utilize the lists of species per 4- ✕ 4-km pixel to mea-sure PD. For one cell in Figure 1 colored red (found in the central peninsula of Florida), we show all the phylogenetic branches linking taxa present in that cell. PD was highly correlated with species richness (Fig-ure S2). Observed PD measures across the state are summarized in Figure 2 for both the phylogram and chronogram.

Florida Plant Phylogeny Is Consistent with Previous Literature

The relationships among vascular plants based on the two plastid genes for the species sampled from Florida agree closely with the results of previous phylogenetic analyses based on more genes and taxa. For example, we recovered the major subclades of angiosperms, and relationships within and among those are also in agreement with broader analyses (Figure 1; Moore et al., 2007; Soltis et al., 2011; Ruhfel et al., 2014; Wickett et al., 2014). Long branches are pronounced in the lycophytes, monilophytes, and gymno-sperms, as well as in parasites, which tend to have increased substitution rates in plastid genes because of the loss of functionality, and in herbaceous lineages, wherein longer branches would be expected compared with woody relatives (Smith and Donoghue, 2008). Lineages known to have radiated rapidly (e.g., Asteraceae) also exhibit generally shorter branches as expected. Likewise, there are clades of very low phylogenetic resolution because of short branch lengths. This is particularly prevalent in Asteraceae and Poaceae, two of the most species-rich families of angiosperms in general and in Florida. The overall phylogenetic framework of the vascular plants of Florida is highly similar to the accepted framework based on broader geographic analyses, and relationships within subclades of vascular plants also reflect those found in other studies.
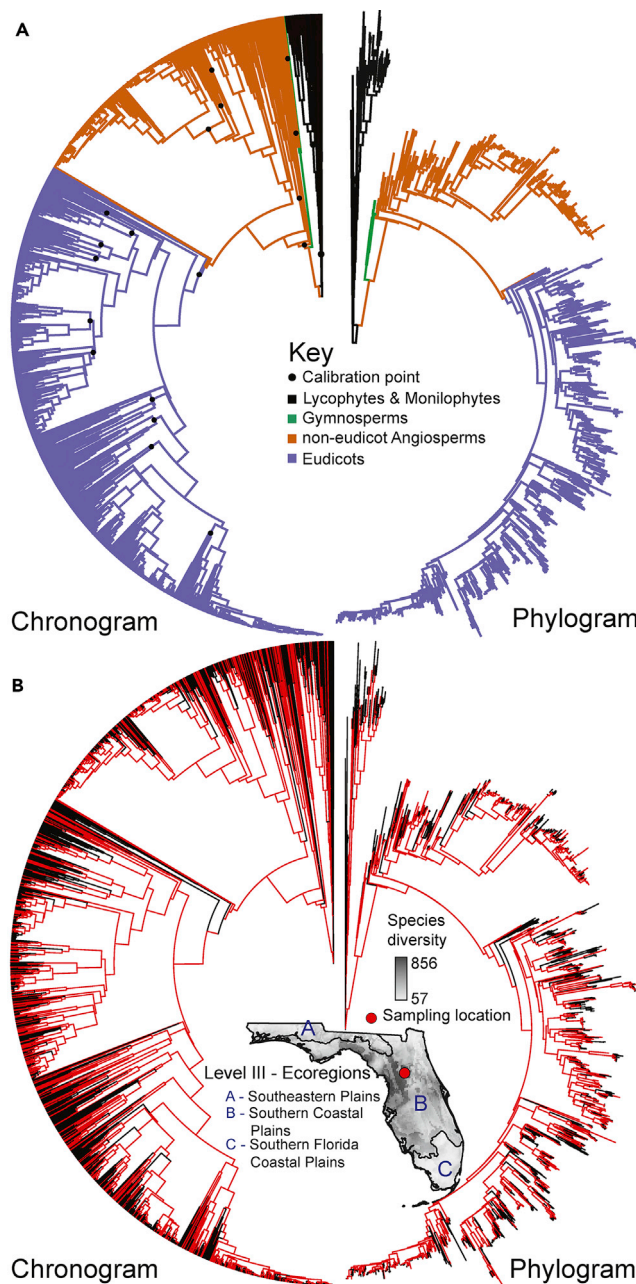
Figure 1. Phylogram, Chronogram and Species Richness of Vascular Plants in Florida

(A) Phylogeny of 1,490 vascular plants in Florida, shown as both chronogram (left) and phylogram (right). The black dots on the chronogram indicate the positions of the 17 calibration points.

(B) A map showing species richness, with PD from one grid cell highlighted in red on the phylogram and the chronogram.

The few instances in which the topology differs from published analyses are all minor deviations from expectations and result from the limited dataset (i.e., *matK, rbcL*; only Florida plants) employed here. These deviations (e.g., a genus with a single sampled species in Florida placed in a related genus with multiple species rather than as sister to that clade, or species of two closely related genera interdigitated) also likely reflect sampling issues and are surprisingly minor, given that the vascular plants of Florida are a small sub-set of global diversity. Furthermore, most inconsistencies between the Florida tree and more broadly sampled trees occur in regions of the phylogeny where relationships continue to be difficult to resolve (e.g., Lamiales and Asteraceae). Finally, uncertainty in a few very short branches is not likely to confound
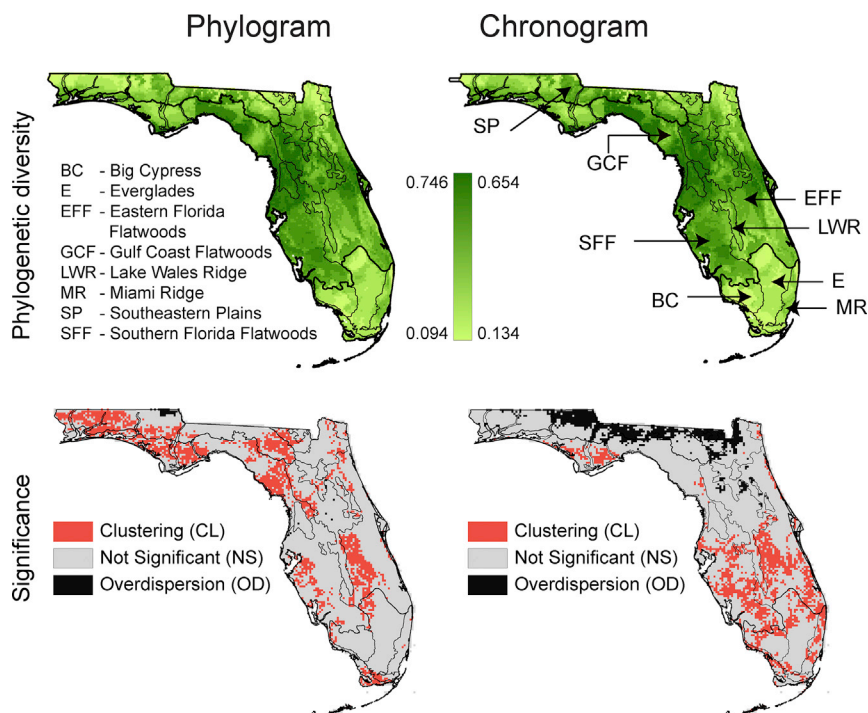
Figure 2. Phylogenetic Diversity of Vascular Plants in Florida

Observed (top panel) and significant (bottom panel) phylogenetic diversity measured from phylogram (left) and chronogram (right) for vascular plants. On the top panel, the Environmental Protection Agency Level III ecoregions are mapped with the darker lines, Level IV are mapped with the lighter lines, and areas of interest, e.g., Lake Wales Ridge and Miami Ridge, are identified.

PD analyses, because the better-supported long branches contribute the great majority of PD (González-Orozco et al., 2016). Calibrating this tree with the fossil constraints (Table S4) yielded a chronogram for comparison with the phylogram in downstream analyses (Figure 1); the chronogram has smoothed branch lengths relative to the phylogram (see Figure 1 for comparison).

Spatial Phylogenetic Patterns across Florida

Phylogenetic Diversity Patterns in PD

Plotting PD across Florida relative to the Level III and Level IV Environmental Protection Agency ecoregions (Figure 2) and to latitude (Figure 3) revealed both ecological and geographic patterns. The highest PD occurred from the northern parts of peninsular Florida south to near Orlando and St. Petersburg. For both the phylogram and chronogram, PD was higher in central Florida than in the northern and southern parts of the state (Figure 2). South Florida showed a mix of patterns longitudinally; the Everglades and Big Cypress had relatively low PD, whereas Miami Ridge, at the same latitude, had relatively high PD (visible as the long tail toward positive PD values in Figure 3 at latitude 26.5°N). The average PD values for the ecoregions also showed higher PD in the Southern Coastal Plain across peninsular Florida than elsewhere (Table 1).

Patterns of Significant Clustering or Evenness

Areas showing significantly high or low PD differed when measured on the phylogram versus the chronogram (Figure 2). The chronogram-derived values showed a strong pattern of evenness in the northern and central areas of Florida, especially in the Southeastern Coastal Plain ecoregion. The chronogram-based results suggest that significantly more evolutionary history than expected is assembled in the north central part of the state and significantly less than expected is assembled in the southern part of the state. In contrast, PD significance based on the phylogram, which more directly represents feature diversity, showed phylogenetic clustering in several regions of the state, particularly along the northwestern coast of Florida, with very little evenness anywhere.
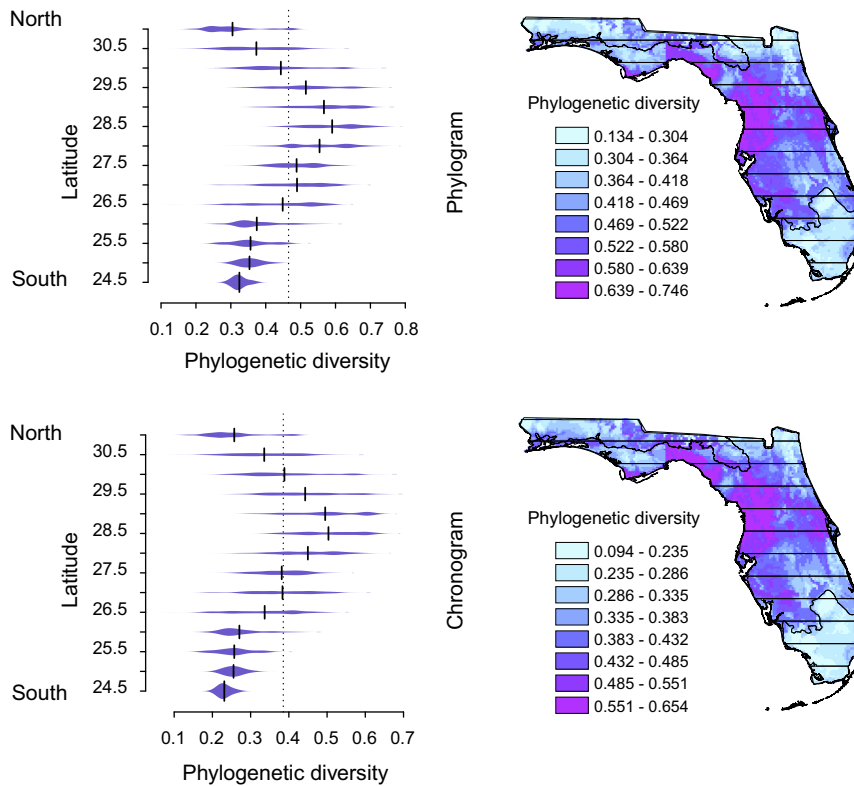
Phylogenetic Diversity by Latitude

Figure 3. Phylogenetic Diversity by Latitude

The study region was binned by 0.5° into 13 latitudinal sections represented by the lines on the maps on the right. Bean plots on the left represent the phylogenetic diversity values for the pixels within each section for both the phylogram (top) and chronogram (bottom).

Relative Phylogenetic Diversity Patterns

Patterns of RPD also differed substantially between the phylogram and chronogram. Geographic areas of major difference between the chronogram- and phylogram-derived values include (Figure 4): (1) northern Florida, where the chronogram yielded high observed RPD and a much more extensive concentration of significantly high RPD (i.e., longer branches than expected) than yielded by the phylogram; (2) central Florida, where the phylogram generally did not show significantly low RPD (i.e., shorter branches than expected), whereas the chronogram did; and (3) very southern Florida, including the Miami Ridge area

	Ecoregion	Mean	SD
Phylogram	Southeastern Plains	0.3997	0.0917
	Southern Coastal Plains	0.5071	0.0987
	Southern Florida Coastal Plains	0.3554	0.0512
Chronogram	Southeastern Plains	0.3621	0.0868
	Southern Coastal Plains	0.4214	0.0968
	Southern Florida Coastal Plains	0.2553	0.0426

Table 1. PD Calculations and SD for the Cells Contained in Three Ecoregions for the Phylogram and Chronogram
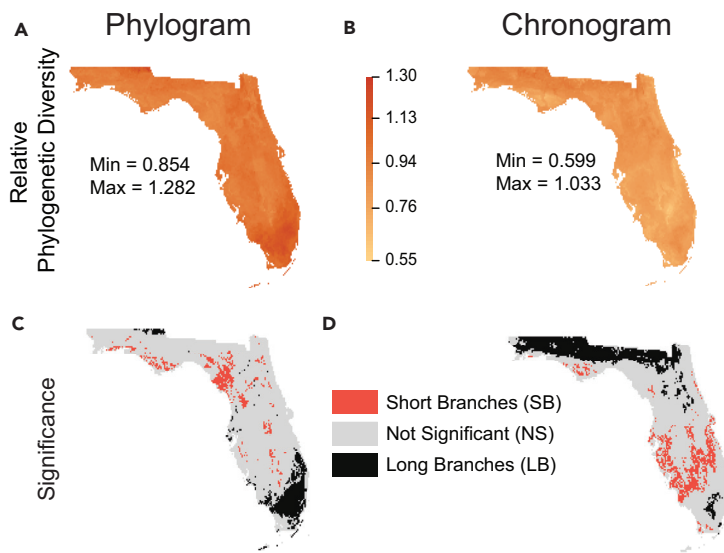
Figure 4. Relative Phylogenetic Diversity
Observed (top panel) and significant (bottom panel) relative phylogenetic diversity measured from (A and C) phylogram and (B and D) chronogram for vascular plants.

and the Everglades, where the phylogram resulted in high observed RPD and a much larger concentration of significantly high RPD than the chronogram.

Alternative Source Trees Yield Both Similarities and Differences in Patterns of PD

Similar patterns of PD emerged among the alternative source trees (from Phylomatic and the OTL) and the purpose-built tree (Figure 5). Similarities are expected between the purpose-built trees and those from the OTL as the branch lengths were determined from the same alignment. Each source phylogram produced PD values with significant clustering in the Florida panhandle and along peninsular Florida, whereas each source chronogram produced PD values showing evenness along the northern edge of the panhandle. We found that the OTL trees and purpose-built trees were the most similar, with fewer than 5% of the cells showing a difference in significance for phylogram-derived values and fewer than 20% of the cells showing a difference in significance for those based on the chronograms. Measures using the Phylomatic trees, despite having fewer deleted taxa, showed more differences from the purpose-built trees; > 25% of the cells showed a different significance result (Figure 5; Table 2). Taxon sampling differences among methods modestly affected this spatial phylogenetic metric (e.g., with fewer taxa, only the area east of LWR was prominent as an area of significant clustering in the phylogram; Figure 5B).

Tree Uncertainty Has Little Effect on Significance of Phylogenetic Diversity Scores

The standard deviations across the PD scores calculated from the 100 chronograms were larger in general than those calculated from the 100 phylograms, with the similarities between the two maps most prominent in the panhandle and far southern Florida (Figures 6A–6C). Significant clustering or evenness for the 100 Bayes trees is summarized in Figures 6D and 6E. As might be expected, no cells showed a change from significant clustering to significant evenness, or vice versa, across the 100 trees, yet some cells showed relatively high inconsistency in significance level in one direction or another. Key questions were whether phylogenetic uncertainty might lead to widespread errors in assessment of PD significance, particularly if changes due to uncertainty are spatially clustered, leading to geographic bias. Although there were some cells for which significance changed, the overall pattern does not suggest spatial structuring in uncertainty of PD significance. Approximately 80% of all pixels were consistent in significance for values derived from both the chronogram and the phylogram, whereas only 7% of the pixels showed the highest level of uncertainty and these were widely scattered (Table 3). Finally, general patterns of PD across latitude do not change when we include all the PD calculations across the 100 trees (Figure 7).
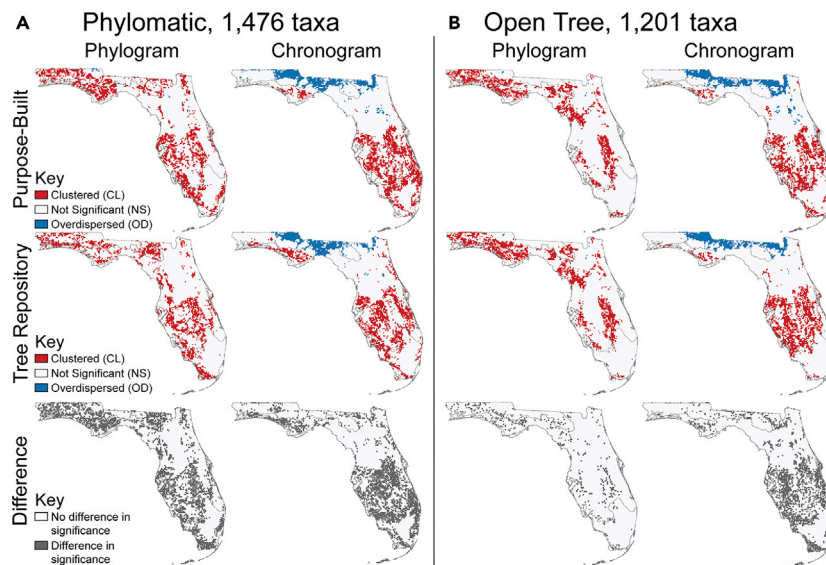
Figure 5. Diversity Hypothesis Tests Comparing Chronograms and Phylograms for Vascular Plants Built Using Either Our Purpose-Built Tree, Phylomatic, or Open Tree

In the top panel are phylograms and chronograms for the purpose-built tree pruned to the (A) Phylomatic and (B) Open Tree taxon dataset. In the middle panel are the (A) Phylomatic and (B) Open Tree trees. In the lower panel are the differences between the two maps. Gray pixels are those that changed in significance level between the Phylomatic and purpose-built tree and the Open Tree and purpose-built tree.

DISCUSSION

An Improved Understanding of Florida Floristic Diversity

Here we examined all vascular plant diversity that shapes vegetation definitions in Florida, instead of limiting our study to only the dominant vegetation. Importantly, we found peaks of plant diversity in northern peninsular Florida rather than in the panhandle. One major reason for putting effort into phylogenetic measures is that they connect to evolutionary and ecological processes that shape diversity patterns. For example, although PD may be the highest in northern peninsular Florida, rather than the panhandle, there is significantly more PD than expected in many panhandle areas, especially when considering chronograms rather than phylograms. Southeastern forests are composed of communities containing deep evolutionary branches, particularly in the time-calibrated phylogenies. These mixed forests are stable over long time periods, facilitating accumulation of a broad set of older lineages, as opposed to oscillations of more open oak savannah habitats and inundation during Pleistocene sea-level incursions in central and southern peninsular Florida.

In southern Florida, which was entirely submerged during the last interglacial (reviewed in Germain-Aubrey et al., 2014), we find an unusual pattern of phylodiversity, where only the phylogram shows a strong signal in RPD, i.e., significantly longer branches than expected given the null hypothesis. We argue that phylograms, often with relatively longer branches toward the tips, are likely to show stronger patterns in some cases than chronograms, which tend to redistribute branch length from terminal to deeper branches (demonstrated in Figure 1). In southern Florida, communities with taxa of Caribbean or Central/South American origin may be dominated by longer terminal branches. Further examination of both community composition and possible artifacts from methodological choices is warranted.

	Comparison with Purpose Built Tree	
	OpenTree	Phylomatic
Phylogram	373	2,126
Chronogram	1,517	1,945

Table 2. Number of Cells Showing Different Results between the Tree Resources
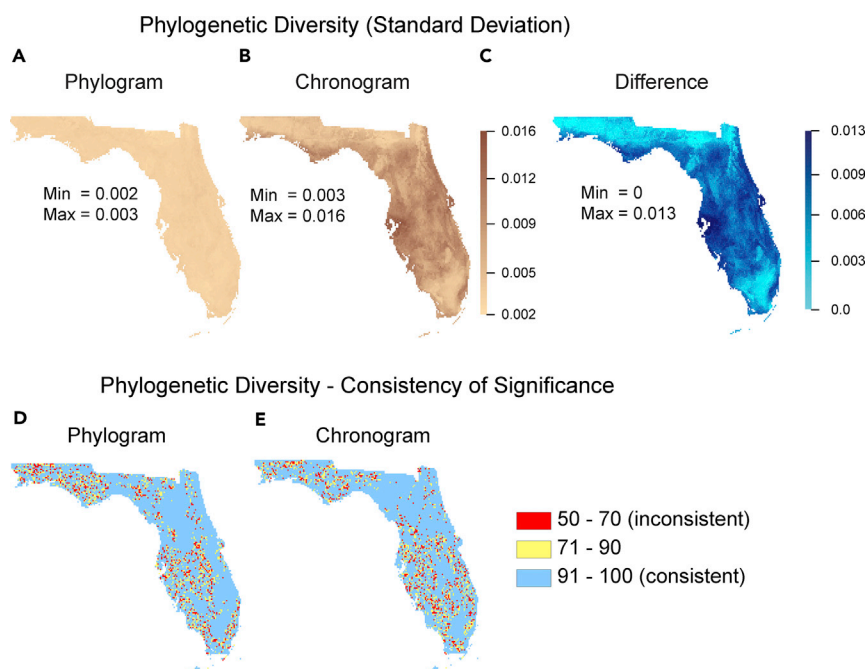
Figure 6. Uncertainty of Phylogenetic Diversity
Top: Standard deviation of observed Phylogenetic Diversity across all (A) 100 phylograms and (B) 100 chronograms selected from the post-burn-in distribution of trees in our Bayesian analysis. (C) Difference between the two on the right in blue. Bottom: Areas in light blue are those for which 91 (of 100) or more of the trees had the same level of significance. Areas in yellow are mostly consistent, with 71–90 (of 100) of the trees finding the same level of significance. Areas in red are the relatively inconsistent pixels, with only 50–70 of the trees having the same level of significance for (D) 100 phylograms and (E) 100 chronograms.

In the central peninsula of Florida, we find strong patterns of both phylogenetic clustering and shorter branches than expected using either the phylogram or the chronogram. Although central Florida is a floristically diverse area, it includes locations that were inundated during Pleistocene interglacial sea-level rise, as well as xeric scrub that was more persistent, but co-occurring taxa are likely filtered due to the evolutionarily conserved preferences of some lineages for the harsh environments of these areas (e.g., excessively drained soil and extreme heat). Alternatively, some of this pattern may be due to *in situ* differentiation, whereas some taxa may be more recent arrivals as many are derived from western North American lineages that dispersed eastward during more xeric interglacial periods (reviewed in Germain-Aubrey et al., 2014).

Our results are also consistent with those found in other, distantly related animal lineages. Used a Florida gap analysis to document high species richness in vertebrates and butterflies especially in the panhandle and extending into the core of central Florida. Although our methods differ from those used by these authors, especially given the focus in on just taxic measures, the results are broadly consistent, perhaps unsurprisingly because plant diversity may generally drive diversity in groups such as butterflies (Burkle et al., 2013).

Class[a]	Phylogram	Chronogram
50-70	642	645
71-90	1,013	997
91-100	6,508	6,521

Table 3. Number of Cells in Each Class of Uncertainty
[a]Class indicates the number of trees with a consistent level of significance. For example, 50-70 class indicates that for those pixels 50-70 of the trees were similarly significant meaning and 50-30 were not similarly significant. For the 71-90 class more of the trees were consistently significant, and for the 91-100 class the majority of the trees found the same level of significance, suggesting that those pixels are consistent when taking into account uncertainty in the phylogenetic estimates.
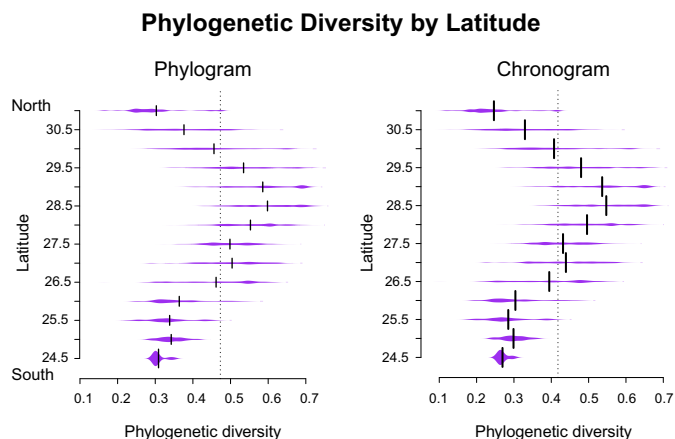
Phylogenetic Diversity by Latitude

Figure 7. Uncertainty of Phylogenetic Diversity by Latitude

The study region was binned by 0.5° into 13 latitudinal sections represented by the lines on the maps on the right. Bean plots represent the phylogenetic diversity values for all 100 trees for each of the pixels within each section for both the phylogram (left) and chronogram (right).

Finally, two particular areas of interest, given known endemicity and unusual floras, are the Miami Ridge/ Pine Rocklands and the LWR (location denoted in Figure 2, top left panel, and Figure S3). The Pine Rock-lands exhibits a diverse flora of hammock species and those common across the Bahamas and Greater Antilles (Myers and Ewel, 1990). In the Miami Ridge area, as expected, we found increased PD and signif-icant PD clustering for some pixels based on the chronogram and significantly high RPD in others for the phylogram. The LWR, in particular, is known to harbor high endemic species diversity (Myers and Ewel, 1990; Germain-Aubrey et al., 2014), which we hypothesize may show high neo-endemism when examined using phylogenetic endemism metrics in the future. It is beyond the scope of this study to investigate such patterns, especially given that we did not create full geographic range surfaces for some of the species examined, but we found that LWR is neither particularly high in PD and nor does it show significantly clustered or even lineages; however, areas immediately east of LWR show strong clustering. This region is a mosaic of habitats, including pine flatwoods, dry prairies, and marshes (Myers and Ewel, 1990), and the significant clustering in this area may indicate strong filtering for these habitats. It also suggests that conservation priorities should not only be concentrated in areas such as the LWR but also include those areas directly adjacent to it along zones of highly varying diversity. Zones of conservation priority are also found in areas such as the Miami Ridge/Pine Rocklands, which are under direct threat from rapid, continuing human development and provide a further strong justification for conservation actions to support these unique evolutionary assemblages. To obtain a complete picture of conservation priorities, future studies are needed of phylogenetic endemism and associated hypothesis tests (e.g., Cadotte and Davies, 2010; Rosauer et al., 2009; Tucker et al., 2012; Mishler et al., 2014) to complement the PD studies reported here. Future analyses of PD and RPD can compare the ecoregions noted above, as well as native habitats and protected areas, thereby informing conservation priorities for human managed habitats and regions as well as those areas experiencing rapid land conversion in Florida.

The Importance of Evaluating Input Trees for Phylogenetic Diversity

Quality of the Purpose-Built Tree and Community Tree Resources

Relationships within the major clades of ferns (monilophytes), gymnosperms, and angiosperms agree closely with broader phylogenetic analyses focused on those specific subclades (e.g., The Angiosperm Phy-logeny Group IV, 2016; Schuettpelz and Pryer, 2007; Smith et al., 2011; Soltis et al., 2011; Stevens, 2001). Some of the more difficult areas to resolve on the purpose-built Florida tree were appropriately resolved in the topology produced from the OTL. This is likely due to the continuously updated nature of the OTL, where the tree topology integrates previously estimated trees into the framework to produce a "synthesis" tree. Our results suggest that, in the future, the OTL may be an important resource for spatial phylogenetic analyses. Providing there is adequate sampling of the terminal taxa in a region represented in the OTL, researchers will be able to save numerous hours in building their own region-specific trees. Of course, for less well-studied regions of the world, many new sequences may need to be added; even for this

dataset, a large proportion of the species did not have existing sequence data, making it necessary to sequence many taxa to provide data for branch length estimation.

Although Phylomatic and the OTL may give relatively accurate topologies, calculating branch lengths for these trees remains problematic, particularly when using phylograms. To address this issue, we used our DNA sequence alignment to estimate branch lengths on the OTL topology, which likely explains why we found fewer differences between results based on our purpose-built tree and the OTL tree when compared with the Phylomatic tree (Figure 5). However, this method requires assembling an alignment for OTL phylogenies, which may defeat the purpose of using such resources. Current efforts already underway to add branch lengths estimates to the OTL method will further increase the strong utility of OTL as a source for spatial phylogenetics analysis.

Taxon sampling may be another issue with using trees from repositories. The Phylomatic tree contained almost all of our terminal taxa of interest (99%), and more taxa than the OTL tree (80%). It is unclear how many taxa would be available if we were to attempt an analysis of all vascular plant species in Florida (~4,300 species). In general, taxon sampling is a concern that is always difficult to overcome. In some studies, coarser-scale Operational Taxonomic Units (OTUs) such as genera are used to represent most of a flora when sequence data are limited at finer scales (e.g., Thornhill et al., 2016, 2017). In our case, we included only 35% of Florida's vascular plant species because the tree was built to match the species for which sufficient occurrence data were available for constructing distribution models. Although we do not know how these patterns might change if we were to add more taxa, pruning of species for comparison with community resources (e.g., OTL) provided a means to examine effects of reduced taxon sampling. We found that the number of cells with significant clustering decreased considerably in central Florida when ~250 taxa were removed (Figure 5). This result suggests that some power may be reduced with more limited taxonomic sampling. However, our current analyses likely reasonably capture general trends in PD in Florida, providing a much-needed, initial snapshot of diversity.

Tree Uncertainty

Using a different way to examine uncertainty, by comparing multiple outcomes of RaxML searches, Thornhill et al. (2017) found virtually no effect of tree uncertainty on spatial phylogenetics results. Here we used a Bayesian approach to examine uncertainty, which has the potential to yield trees with more differences, yet we hypothesized that tree uncertainty would have a minimal impact on measures of PD significance given that, in most cases, shorter branches are affected, especially for phylograms. However, with larger trees and limited character sampling (nucleotides), uncertainty could have an impact on assessments of PD. Figure 6 shows that approximately 20% of pixels showed moderate to high differences in significance among the 100 trees, flipping between non-significant and either significantly high or significantly low (but never from significantly high to significantly low). This pattern is found for both the phylogram and chronogram, although standard deviations per pixel are much higher for the chronograms. Tree uncertainty in chronograms has more impact on branch lengths due to time scaling—if nodes are uncertain then swapping of branches can result in more pronounced changes in branch lengths than seen at the same place in the corresponding phylogram wherein the uncertain branches tend to be quite short (see demonstration of this in González-Orozco et al., 2016). Our example is an empirical one, and more work using simulated trees could further elucidate expectations of the impacts of tree uncertainty as it relates to measures of phylodiversity.

Two key messages come from our analyses. First, although uncertainty is likely to affect judgments of significance in PD and RPD for certain grid cells, there appeared to be no geographic structuring of such cells, and thus the modest amounts of uncertainty seen here do not broadly affect conclusions at the landscape scale. For example, our general assessment of significant clustering in central Florida still appears to hold despite some differences in results among the 100 trees. Second, whereas changes between non-significant and significant clustering or evenness in one direction were seen occasionally due to uncertainty, no changes were seen between significantly high and significantly low. This suggests that, although uncertainty may affect our interpretations of significance or not, it will not change our interpretation of significant clustering to significant evenness. Still, we argue that phylogenetic uncertainty should be considered in phylodiversity analyses, which is currently often not the case.

Phylograms vs. Chronograms

The divergent results seen in significance tests inferred from the phylogram versus the chronogram were the largest differences observed in this analysis, much larger than differences due to tree uncertainty or

tree source. The tree topology is the same between the two analyses, whereas the branch lengths are different, and each approach has unique interpretations. An analogy would be travel directions for a route using either geographic distances or times; both indicators are informative in different ways. Generally, it is thought that evenness of lineages measured on the phylogram directly relates to the unexpectedly high genetic disparity and thus may indirectly relate to high functional trait disparity, if the change in those traits is correlated with genetic change in the markers employed. By contrast, evenness of lineages measured on the chronogram directly relates to unexpectedly high temporal disparity and may likewise indirectly relate to high functional trait disparity, if the change in those traits is correlated with time. Both correlations are quite plausible; the interpretation of differences in significance patterns may come down to tempo and mode of evolution. If anagenesis in functional traits is correlated with heterogeneous rates of genetic change on different branches, as, for example, due to generation time effects as commonly seen when comparing woody plants with herbaceous relatives (Smith and Donoghue, 2008) or major adaptive effects such as commonly seen when comparing parasitic plants with autotrophic relatives, then the phylogram will illuminate those processes with significance of PD. On the other hand, if anagenesis in functional traits is relatively uniform and generally correlated with the amount of time elapsed along a branch, then the chronogram will likely indicate that process.

The same distinction is important when using PD-related results to help set conservation priorities. Areas with high PD measured on the phylogram by definition have high genetic diversity, and this may be the better measurement if the goal is preserving genetic diversity, whereas high-PD areas measured on the chronogram contain an unusually large amount of evolutionary time, and this may be the better measurement if the goal is preserving evolutionary diversity. Which form of branch lengths is preferred for a proxy for functional trait diversity depends on the value placed on these processes along with the conservation priorities (Thornhill et al., 2017).

Limitations of the Study

Although this study includes 1,490 taxa of Florida plants, more than 4,000 vascular plant species are known to occur in the state, and it is possible that full inclusion of all species could affect the results presented here. Further efforts to assemble more complete distribution data, especially for range-restricted species, are ongoing, and those records can hopefully lead to further refined and accurate species distribution modeling. We note that although the phylogeny recovered using a small set of markers aligns with known relationships, further work to develop more robust phylogenetic hypotheses is a next step. Finally, further work correlating these patterns with areas of high population growth and encroaching sea-level rise will provide additional insights for conservation efforts and planning.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, five figures, and five tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.12.002.

AUTHOR CONTRIBUTIONS

J.M.A., C.C.G.-A., N.B., S.W.L., B.D.M., D.E.S., R.G., and P.S.S. contributed to the study conception and design. J.M.A., C.C.G.-A., K.M.N., L.C.M., W.M.W., J.R.A., D.E.S., and P.S.S. helped with the acquisition of data. J.M.A., C.C.G.-A., N.B., K.M.N., L.C.M., S.W.L., B.D.M., H.L.O., D.E.S., R.G., and P.S.S. helped with analysis and interpretation of data. All authors were involved in drafting of the manuscript, and S.W.L., B.D.M., J.R.A., H.L.O., D.E.S., R.G., and P.S.S. aided in critical revisions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Beaulieu, J.M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B.C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66, 2369–2383.

Buckley, L.B., Davies, T.J., Ackerly, D.D., Kraft, N.J.B., Harrison, S.P., Anacker, B.L., and Wiens, J.J. (2010). Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. Proc. Biol. Sci. 277, 2131–2138.

Burkle, L.A., Marlin, J.C., and Knight, T.M. (2013). Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science 339, 1611–1615.

Cadotte, M.W., and Davies, T.J. (2010). Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. Divers. Distrib. https://doi.org/10.1111/j.1472-4642.2010.00650.x.

Davies, R.G., Orme, C.D.L., Webster, A.J., Jones, K.E., Blackburn, T.M., and Gaston, K.J. (2007). Environmental predictors of global parrot (Aves: Psittaciformes) species richness and phylogenetic diversity. Glob. Ecol. Biogeogr. 16, 220–233.

Davies, T.J., and Buckley, L.B. (2012). Exploring the phylogenetic history of mammal species richness: a phylogenetic history of mammal richness. Glob. Ecol. Biogeogr. 21, 1096–1105.

Dobson, A.P., Rodriguez, J.P., Roberts, W.M., and Wilcove, D.S. (1997). Geographic distribution of endangered species in the United States. Science 275, 550–553.

Ezard, T.H.G., Aze, T., Pearson, P.N., and Purvis, A. (2011). Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332, 349–351.

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10.

Germain-Aubrey, C.C., Soltis, P.S., Neubig, K.M., Thurston, T., Soltis, D.E., and Gitzendanner, M.A. (2014). Using comparative biogeography to retrace the origins of an ecosystem: the case of four plants endemic to the central Florida scrub. Int. J. Plant Sci. 175, 418–431.

González-Orozco, C.E., Pollock, L.J., Thornhill, A.H., Mishler, B.D., Knerr, N., Laffan, S.W., and Gruber, B. (2016). Phylogenetic approaches reveal biodiversity threats under climate change. Nat. Clim. Change 6, 1110–1114.

Gordon, D.R. (1998). Effects of invasive, non-indigenous plant species on ecosystem processes: Lessons from Florida. Ecol. Appl. 8, 975–989.

Greller, A.M. (1980). Correlation of some climate statistics with distribution of broadleaved forest zones in Florida, U. S. A. Bull. Torrey Bot. Club 107, 189–219.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., and Cranston, K.A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. U S A 112, 12764–12769.

Ihlo, C., Urban, D., Olander, L., and Galik, C. (2014). Spatial distribution and characteristics of at-risk species in the southeast U.S. In Nicholas Institute for Environmental Policy Solutions Scoping Document: Part 2, Exploratory Analysis of Characteristics and Trends of At-risk Species in the Southeast U.S., Christy Ihlo, Dean Urban, Lydia Olander, and Christopher Galik, eds. (Duke University), pp. 1–11.

Jetz, W., and Freckleton, R.P. (2015). Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370, https://doi.org/10.1098/rstb.2014.0016.

Liu, X., Swenson, N.G., Zhang, J., and Ma, K. (2013). The environment and space, not phylogeny, determine trait dispersion in a subtropical forest. Funct. Ecol. 27, 264–272.

Long, R.W., and Lakela, O. (1971). A Flora of Tropical Florida: A Manual of the Seed Plants and Ferns of Southern Peninsular Florida (University of Miami Press).

Mishler, B.D., Knerr, N., González-Orozco, C.E., Thornhill, A.H., Laffan, S.W., and Miller, J.T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. Nat. Commun. 5, 4473.

Molina-Venegas, R., and Roquet, C. (2013). Directional biases in phylogenetic structure quantification: a Mediterranean case study. Ecography (Cop.) 37, 572–580.

Moore, M.J., Bell, C.D., Soltis, P.S., and Soltis, D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U S A 104, 19363–19368.

Myers, R.L., and Ewel, J.J. (1990). Ecosystems of Florida (University of Central Florida Press).

Nagalingum, N.S., Knerr, N., Laffan, S.W., González-Orozco, C.E., Thornhill, A.H., Miller, J.T., and Mishler, B.D. (2015). Continental scale patterns and predictors of fern richness and phylogenetic diversity. Front. Genet. 6, 132.

Noss, R.F., Platt, W.J., Sorrie, B.A., Weakley, A.S., Means, D.B., Costanza, J., and Peet, R.K. (2015). How global biodiversity hotspots may go unrecognized: Lessons from the North American coastal plain. Divers. Distrib. 21, 236–244.

Olalla-Tárraga, M.Á., McInnes, L., Bini, L.M., Diniz-Filho, J.A.F., Fritz, S.A., Hawkins, B.A., and Purvis, A. (2011). Climatic niche conservatism and the evolutionary dynamics in species range boundaries: Global congruence across mammals and amphibians. J. Biogeogr. 38, 2237–2247.

Pearse, W.D., and Purvis, A. (2013). PhyloGenerator: an automated phylogeny generation tool for ecologists. Methods Ecol. Evol. 4, 692–698.

Qian, H., Wiens, J.J., Zhang, J., and Zhang, Y. (2015). Evolutionary and ecological causes of species richness patterns in North American angiosperm trees. Ecography 38, 241–250.

Rangel, T.F., Colwell, R.K., Graves, G.R., Fučíková, K., Rahbek, C., and Diniz-Filho, J.A.F. (2015). Phylogenetic uncertainty revisited: Implications for ecological analyses. Evolution 69, 1301–1312.

Rosauer, D., Laffan, S.W., Crisp, M.D., Donnellan, S.C., and Cook, L.G. (2009). Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Mol. Ecol. 18, 4061–4072.

Rosauer, D. (2010). Techniques for Spatial Analysis of Phylogenetic Diversity, PhD thesis (School of Biological, Earth and Environmental Sciences, University of New South Wales).

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. (2014). From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14, 23.

Schuettpelz, E., and Pryer, K.M. (2007). Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56, 1037–1050.

Smith, S.A., and Donoghue, M.J. (2008). Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86–89.

Smith, S.A., Beaulieu, J.M., Stamatakis, A., and Donoghue, M.J. (2011). Understanding angiosperm diversification using small and large phylogenetic trees. Am. J. Bot. 98, 404–414.

Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., and Soltis, P.S.

(2011). Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. *98*, 704–730.

Stevens, P. (2001). Angiosperm Phylogeny website. www.mobot.org/MOBOT/research/APweb/.

Swenson, N.G. (2009). Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. PLoS One *4*, e4390.

The Angiosperm Phylogeny Group IV (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. *181*, 1–20.

Thornhill, A.H., Mishler, B.D., Knerr, N., Gonzalez-Orozco, C.E., Costion, C.M., Crayn, D.M., Laffan, S.W., and Miller, J.T. (2016). Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. J. Biogeogr. *43*, 2085–2098.

Thornhill, A.H., Baldwin, B.G., Freyman, W.A., Nosratinia, S., Kling, M.M., Morueta-Holme, N., Madsen, T.P., Ackerly, D.D., and Mishler, B.D. (2017). Spatial phylogenetics of the native California flora. BMC Biol. *15*, 96.

Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., and Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. Nature *466*, 1098–1101.

Tucker, C.M., Cadotte, M.W., Davies, T.J., and Rebelo, T.G. (2012). Incorporating geographical and evolutionary rarity into conservation prioritization. Conserv. Biol. *26*, 593–601.

Webb, C.O., and Donoghue, M.J. (2005). Phylomatic: Tree assembly for applied phylogenetics. Mol. Ecol. Notes *5*, 181–183.

Webb, C.O., Ackerly, D.D., and Kembel, S.W. (2008). Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics *24*, 2098–2100.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., and Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U S A *111*, E4859–E4868.

Wiens, J.J., Sukumaran, J., Pyron, R.A., and Brown, R.M. (2009). Evolutionary and biogeographic origins of high tropical diversity in old world frogs (Ranidae). Evolution *63*, 1217–1231.

Wright, I.J., Ackerly, D.D., Bongers, F., Harms, K.E., Ibarra-Manriquez, G., Martinez-Ramos, M., Mazer, S.J., Muller-Landau, H.C., Paz, H., and Wright, S.J. (2007). Relationships among ecologically important dimensions of plant trait variation in seven neotropical forests. Ann. Bot. *99*, 1003–1015.

Wunderlin, R.P., Hansen, B.F., Franck, A.R., and Essig, F.B. (2017). Atlas of Florida Plants. In (Application Development), USF Water Institute, S.M. Landry and K.N. Campbell, eds. (Institute for Systematic Botany, University of South Florida). http://florida.plantatlas.usf.edu/.

Supplemental Information

Spatial Phylogenetics of Florida Vascular

Plants: The Effects of Calibration

and Uncertainty on Diversity Estimates

Julie M. Allen, Charlotte C. Germain-Aubrey, Narayani Barve, Kurt M. Neubig, Lucas C. Majure, Shawn W. Laffan, Brent D. Mishler, Hannah L. Owens, Stephen A. Smith, W. Mark Whitten, J. Richard Abbott, Douglas E. Soltis, Robert Guralnick, and Pamela S. Soltis

TRANSPARENT METHODS

Ecoregions in Florida

The U.S. Environmental Protection Agency (EPA) has defined a nested set of North American ecoregions, areas that are similar in terms of both biotic and abiotic variables (Omernik and Griffith 2014). The EPA Level III ecoregions include 105 areas in North America, three of which are found in Florida (Figure S1). Many plant species from northern Florida have distributions that occur across the southeastern U.S.A.; therefore, we used distribution points across the three Level III ecoregions to generate species distribution models, as discussed below. The EPA's more subdivided Level IV ecoregions include 20 within Florida that define a number of important vegetation regions, including the Everglades and the Miami Ridge. We interpret results in light of the Level IV ecoregions and specific ecogeographic areas of interest.

Species Distribution Models and Predictions of Community Composition

To calculate PD across Florida, we developed species distribution models for 1,490 species of vascular plants. To generate the models, we collected specimen locality information of vascular plants of Florida based on the statewide flora by Wunderlin and Hansen (2011). These data came from seven herbaria in the southeastern U.S.A. (FLAS - University of Florida Herbarium, FSU - Florida State University Robert K. Godfrey Herbarium, LSU - Louisiana State University Herbarium, MISS - Tomas M. Pullen Herbarium Mississippi State, UNA - University of Alabama Herbarium, USCH - A.C. Moore Herbarium University of South Carolina, USF - University of South Florida Herbarium) and the Global Biodiversity Information Facility (GBIF), and inventory data from the Florida Natural Areas Inventory (FNAI). All non-florida taxa were removed before georeferencing. We used *Biogeomancer* (Guralnick et al. 2006) and Geolocate (http://www.museum.tulane.edu/geolocate/) to help retrieve computer-readable geospatial coordinates for specimen records that were not associated with GPS coordinates. This was especially important for older collections (i.e. pre-GPS collections). Taxonomic components of specimen records were input into the Taxonomic Name Resolution Service tool (Boyle et al. 2013) of the iPlant Collaborative (now Cyverse).

We limited the extent of our assembly of occurrences to three EPA ecoregions across Florida and the Southeastern U.S.: the Southeastern Plains (in the northwestern panhandle), the Southern Coastal Plains (covering the southern and eastern panhandle and most of the peninsula), and the Southern Florida Coastal

Plains (covering the southern tip of the peninsula and the Florida Keys; Figure S1). The species occurrences included in the model came from the entire extent of their ranges within the three ecoregions. Many of the species for which we developed models have wider ranges, but we chose to limit population occurrences and the extent of modeling to these three ecoregions. Our rationale is that populations adapting to local and regional conditions are more the norm than the exception, thus reducing the commission errors (Valladares et al. 2014) likely in the context of ecoregions and their strong ties to edaphic and terrain characteristics of those regions.

We set minimum requirements for modeling as follows: for widely distributed taxa, we retained only those with 30 or more unique occurrences; for narrowly distributed taxa, only those with 20 or more occurrences; and for Florida endemic plants, at least 10 occurrences were required to retain the species for modeling (Table S5). We coded each species either as narrow (if a species was restricted to 10 or fewer adjacent counties) or widespread (if a species was found in more than 10 counties, especially if not adjacent) according to the Florida Plant Atlas (http://florida.plantatlas.usf.edu/). In total, 1,502 species of vascular plants had sufficient data for modeling.

To incorporate the background environmental layers for building the models, we calibrated models utilizing bioclimatic, topographic, and edaphic variables. We downloaded bioclimatic variables at 30 arc-second (~1-km) spatial resolution from Worldclim (http://www.worldclim.org; Hijmans et al. 2005) and generated a correlation matrix among these 19 variables. We removed highly correlated bioclimatic variables (those with a correlation coefficient > 0.7) within the maximum model calibration region, resulting in eight non-correlated bioclimatic variables to build the models; these variables were Mean Diurnal Range (Mean of monthly (max temp - min temp)), Max Temperature of Warmest Month, Min Temperature of Coldest Month, Mean Temperature of Wettest Quarter, Mean Temperature of Driest Quarter, Annual Precipitation, Precipitation of Wettest Month, and Precipitation of Driest Month. We also downloaded a digital elevation model at ~1-km spatial resolution (GTOPO30, supplied by USGS; https://lta.cr.usgs.gov/GTOPO30) and a soil layer from USGS (USGS.gov), resulting in a total of 10 background variables to calibrate the models.

To fit the models to migration potential, we developed an automated methodology to select appropriate accessible areas for calibrating each species model (Barve et al. 2011). We used county-level presence information from the Atlas of Florida Plants (Wunderlin et al. 2017) in addition to the occurrence

records. We buffered the occurrences by the average distance between each occurrence using the county centroid as an occurrence. Thus, buffered distance varied depending upon available occurrences for each species. This buffering distance, while automated, ensures that accessible area was not overly broad, which typically negatively impacts model predictions and the ability to assess model performance accurately (Barve et al. 2011). It is still an open empirical question if such approaches are generally equivalent to hand-defined accessible areas. We clipped the environmental background data layers to accessible area shapefiles using the function CropRaster from the R package ENMGadgets (Barve and Barve 2014).

For all taxa that met data requirements, we constructed niche models using the software package MaxEnt 3.3.3k (Phillips et al. 2006), which employs a maximum entropy algorithm to predict the potential distribution of species. We used default parameters except the regularization parameter was set to 1.5 based on initial testing of model performance. We used 10 bootstrap replicates to calibrate the model and used the averaged model. We projected the model in the ecoregion(s) where the species is known to occur. Once species distribution models were assembled for all species, the resulting potential distribution outputs were converted into presence-absence maps by thresholding using two threshold methods. First, we used the equal test sensitivity and specificity threshold (ESS) values, because this approach optimized the omission and commission errors between the presences and background in the calibration region. In the second threshold method, we extracted the probability values from the generated surfaces by omitting the least suitable 10% of occurrences. Both threshold methods generated very similar distribution surfaces. Thus, we used the distributions generated by the ESS method as this approach equalizes the commission and omission error. To perform thresholding, we used the function ModelThreshold from the R package ENMGadgets (Barve and Barve 2014). We calculated model validation metrics for all models, including, in particular, AUC values generated from ROC curves. The training AUC was always >0.8 and the difference between training and testing was always <0.15 otherwise the model was discarded (Figure S5). We removed 12 species where niche models performed particularly poorly based on AUC values, resulting in a final set of 1,490 species distribution models.

These models extended across the three ecoregions then they were then clipped to the state borders of Florida, including all natural land-sea boundaries except Florida's northern border. Both for computational reasons and given the challenges with predictions of richness in thresholded and stacked species distribution

models (Calabrese et al. 2014), we resampled these potential distributions from 1-km to 4-km resolution, using the nearest neighbor algorithm in ArcMap, overlaid thresholded maps, and created species lists for each pixel at a resolution of 4-km. This reduced the number of pixels for randomization tests from ~600,000 to ~30,000 across the three ecoregions and was meant to provide a more accurate list of species per pixel, trading that benefit with the cost of lower spatial grain. Species lists are a critical component for phylodiversity analyses and are needed for linking to phylogenies.

Genetic Data Collection and Phylogenetic Tree Reconstruction

For the 1,490 species with distribution models, we produced phylogenetic trees using both published sequence data and newly generated sequences. We selected two genes commonly used in plant phylogenetics, *rbcL* and *matK* of the plastid genome (e.g., Hollingsworth et al. 2011), and queried GenBank for our taxa; 463 of these species had sequences available from Florida collections. These DNA sequences were checked throughout the alignment and tree-building process, and problematic sequences were removed.

A total of 1,027 species lacked DNA sequence data on GenBank. We therefore collected wild and cultivated material of these species from throughout Florida; voucher specimens were deposited at the University of Florida Herbarium (FLAS). In addition, herbarium specimens from FLAS were sampled for rare or hard-to-access taxa.

DNA extraction

Genomic DNA was extracted from silica-dried (Chase and Hills, 1991) and herbarium material using a modified CTAB method (Doyle and Doyle, 1987) scaled to a 1 mL volume reaction. Approximately 10 mg of dried tissue was ground in 1 mL of CTAB 2X buffer and 10 μL of proteinase-K. Problematic DNA isolates (i.e., samples that failed to provide high-quality amplicons through PCR) were cleaned using Axygen AxyPrepTM, Qiagen QIAquickTM, or Promega WizardTM cleaning kits according to the manufacturer's protocols. For further detail, see Neubig et al. (2014).

PCR, sequencing, and sequence editing

PCR amplification of *rbc*L was performed in several labs, so reaction components differ slightly. Either Biometra Tgradient or an Eppendorf Mastercycler EP Gradient S thermocycler was used. SigmaTM brand reagents were used to amplify rbcL in 25 μL volumes with the following reaction components: 0.5-1.0 μL of

template DNA (~10-100 ng), 18 μL of water, 2.5 μL of 10X buffer, 3 μL of 25mM MgCl2, 0.5 μL of 10 mM

dNTPs, 0.5 μL each of 10 μM primers, and 0.5 μL of Taq or with 1 μL of template DNA (~10-100 ng), 9.4

μL H2O, 5.0 μL of 5X buffer, 2.5 μL of 25 mM MgCl2, 1.0 μL of 2.5 mM dNTPs, 2.0 μL saturated betaine

solution, 2 μL each 5 μM primer, and 0.1 μL Taq. Three different primer combinations were used for *rbc*L

(Table S1). These were used in succession if the first combination failed to amplify: Z1 and 3' (Zurawski,

Clegg & Brown, 1984), rbcLaF and rbcLaR (Shokralla et al., 2010), and NY35 or NY149 (Cameron, 2004).

PCRs were amplified using 94°C, 3 min; 33X (94°C, 30 sec; 55°C, 30 sec; 72°C, 2 min); 72°C, 4 min, or as

described in Clayton et al. (2007).

The plastid gene *mat*K was amplified using SigmaTM brand reagents in 25 μL volumes with the

following reaction components: 0.5-1.0 μL of template DNA (~10-100 ng), 17.5 μL of water, 2.5 μL of 10X

buffer, 2.5 μL of 25mM MgCl2, 0.5 μL of 10 mM dNTPs, 0.5 μL each of 10 μM primers, and 0.5 μL of Taq.

For *mat*K, many primer combinations were required: 1) 390F and1326R (Cuénoud et al., 2002); 2) F

(Equisetum) + R (Equisetum) (Nicolalde-Morejón et al., 2011); 3) XF and Malp_R1 (Dunning and

Savolainen, 2010; Sun, McLewin & Fay, 2001), 3F; 4) 1R (CBOL Plant Working Group, 2009); 5) 472F and

1248R (Yu et al. 2011). PCR cycling conditions for the 390F and 1326R primer set were as follows: 94°C, 3

min; 33X (94°C, 30 sec; 51°C, 30 sec; 72°C, 1 min 30 sec); 72°C for 3 min. PCR cycling conditions for the

other primer combinations were equivalent to the previous specifications except the annealing temperature

was increased from 51°C to 53°C at each cycle. PCR products for both genes were sequenced via Sanger

sequencing using the ABI 3130 at the Interdisciplinary Center for Biotechnology Research at UF. Sequences

were edited in Sequencher 4.9 (Gene Codes, Corp., Ann Arbor, MI). Sequences were deposited in GenBank

(Table S5). Three outgroup taxa (*Physcomitrella patens* (Funariaceae), *Syntrichia ruralis* (Pottiaceae), and

Mastigophora woodsii (Mastigophoraceae)) were added using data from GenBank (Table S5).

Alignment and Tree Assembly

The final two-gene, 1,490-species matrix was nearly complete, with only 6.9% of taxa missing sequence

data. *matK* and *rbcL* DNA sequences were aligned according to a reference protein coding sequence using

Pal2Nal (Suyama et al., 2006) which builds a multiple codon alignment and were then checked by eye.

Initially trees for each gene were built to look for major issues within each genetic dataset and then combined.

PartitionFinder (Lanfear et al. 2012) was used to select the appropriate model and partition for the

concatenated dataset. All models of molecular evolution were tested using a BIC selection criterion, and four different partitioning schemes tested (Table S2). The best scheme as determined by PartitionFinder included each codon position for each gene as a separate partition and a GTR +I + G model of molecular evolution. RAxML (Stamatakis 2014, v. 7.03) was run with a partitioned dataset and the GTR + G model of molecular evolution for each partition. The invariant sites parameter (I) was not included in the run to prevent over-parameterization based on recommendations by A. Stamatakis (RAxML manual). Finally, 100 fast bootstrap replicates (Stamatakis et al. 2008) were used to determine level of internal support for initial tree evaluation.

The resulting phylogenetic tree was examined for topological issues, including possible misidentifications of species, contaminants, and misplaced taxa. When these issues were encountered, the alignment was examined and either edited or the problematic taxon removed.   For this final phylogenetic tree we added a single constraint to ensure that the three lycophytes sampled formed a clade (*Selaginella arenicola*, (*Lycopodiella appressa*, *Lycopodiella alopecuroides*)); because of their long branches and absence of close relatives, these taxa tended to float around the tree. Finally, the same partitioning scheme and model were used as before, with 200 bootstrap replicates.

The tree was then rebuilt using the methods above and reexamined to produce a 'purpose-built' phylogram of Florida plants. To assess tree uncertainty, the best tree from RAxML was used as a starting tree in a Bayesian analysis using MrBayes (version 3.2.2; Ronquist et al., 2012), with the same constraint and model for 20×10^6 generations with twenty independent runs. Each run was checked for stationarity and the top 10 runs found to have reached stationarity were selected. The posterior distribution of trees was pruned to the last 50% of the generations to remove burnin. The top 10 trees were sampled from each of the top 10 runs, based on likelihood score. To check that our 100 sampled trees represented much of the variation in the full posterior distribution of trees, a consensus tree was built from the full distribution of trees and support values estimated on each node. The support values were then estimated on the topology using only the 100 sampled trees with DendroPy v. 4.0.0 (Sakumaran and Holder 2010). There was a high correlation in the support values for the sample of trees to the support values of the full distribution (linear model; $p < 0.001$, $r2 = 0.98$; Figure S4). Therefore, these trees were considered to be a good representation of the posterior distribution of trees and used in downstream analyses to examine the effects of topological uncertainty on phylogenetic diversity metrics PD and RPD.

Phylogenetic Dating

The best maximum likelihood tree and the sampled bayesian trees were then rooted to the outgroups and modified into chronograms using 17 calibration points (Table S4). These points have been well curated as valid calibration points for Angiosperms (Bell et al., 2010). We added one root calibration to constrain the base of the tree (max age = 377 Million years; Soltis et al., 2002). The 17 calibration points were then used with penalized likelihood in the program r8s (Sanderson et al., 2003). To do this, a smoothing parameter was first estimated using the cross validation approach with a tree pruned to one taxon per genus to reduce the computational time necessary for each cross-validation. Smoothing parameters from -13 to 7 were tested at increments of one integer and the parameter with the lowest chi-squared error, 5.0, was selected as the smoothing parameter. The smoothing parameter was then used on the full tree with the same calibrations for the maximum likelihood tree as well as the 100 bayesian trees. In sum, we had a set of 101 phylograms and 101 chronograms. The NEXUS file with aligned sequence data and the 202 trees will be deposited in DRYAD upon acceptance.

Adding Missing Taxa

A total of 98 taxa, for which we built ecological niche models for, could not be included due to unavailability of tissue or poor sequencing results. These missing species are often rare and narrowly distributed, but strongly associated to local habitats and thus valuable to include. Rarity also means that quality tissue with the potential to yield adequate sequencing results for these ecologically important taxa is in most cases lacking. While adding missing taxa can add uncertainty to a phylogenetic analysis, these taxa only represent 6.5% of the full dataset and thus imputation error is likely to be smaller than in cases where the available data is sparse. In order to do so, we determined the placement of each species by finding an appropriate sister taxon within the most up to date phylogenetic studies. Taxa for which only a single representative of the genus was represented on the tree were placed sister to those taxa, and for taxa for which no known generic representative was available we placed the sister to a closely related genus based on the literature (Table S3). Once the placement of each taxon was determined within this classification, we developed a custom perl script that located the sister taxon, added the missing taxon forming an unresolved node, and the two sister taxa were considered to have the same branch lengths.

Evaluating Phylogenetic Diversity (PD) and Relative Phylogenetic Diversity (RPD)

There are a number of ways to calculate PD, in this analysis PD is expressed relative to the total length of the tree, so PD for a cell is the proportion of the total tree length present in that place (index PD_P in *Biodiverse*; Laffan et al. 2010). Both tails of the distribution are considered, providing a means to determine (using PD) if a region contains taxa that are more closely related to each other, or are more distantly related to each other, than expected, i.e., phylogenetic clustering or evenness, respectively (Webb et al. 2002).

Relative Phylogenetic Diversity (RPD) is the ratio of PD measured on the original tree to PD measured on a comparison tree that has the same topology but all branches of equal length (Mishler et al. 2014). The RPD ratio was evaluated for statistical significance using the same spatial randomizations described below and a two-tailed test, which determines whether either long or short branches are over-represented in a particular geographic region (Mishler et al. 2014).

Spatial randomization tests

Comparing observed patterns of richness and PD with results of the separate statistical tests provides the best basis for insights into ecological, evolutionary, and biogeographic processes, e.g., those processes that either allow older lineages to persist, new lineages to diversify, or biotic interactions that limit distribution of closely related taxa, as well as conservation decision-making (Thornhill et al. 2016).

The statistical significances of PD and RPD were assessed using the 'rand_structured' randomization option in *Biodiverse* (Laffan and Crisp 2003; Laffan et al. 2010). For each iteration, taxon occurrences were randomly reassigned to grid cells without replacement; thus, richness in each grid cell (i.e., the number of taxa present) and the range size of each taxon (i.e., the number of grid cells where it occurs) were held constant. The analysis was run with 999 iterations, re-calculating PD and RPD at each iteration using the randomly generated assemblages, and finally calculating the rank relative position of the observed index value for each cell versus the randomized values.

Comparing observed patterns of richness and PD with results of the separate statistical tests provides the best basis for insights into ecological, evolutionary, and biogeographic processes, e.g., those processes that either allow older lineages to persist, new lineages to diversify, or biotic interactions that limit distribution of closely related taxa, as well as conservation decision-making (Thornhill et al. 2016).

Both PD and RPD were assessed using a two-tailed test to determine. For PD this enables an assessment of whether a region contains taxa that are more closely, or distantly, related to each other than expected, i.e., phylogenetic clustering or evenness, respectively (Webb et al. 2002). For RPD, the test determines whether a location contains an over-representation of longer or shorter branches than expected (Mishler et al. 2014). In the case of the chronograms, this means that a location contains a greater number of older or younger branches than expected.

Spatial Arrangement of Phylogenetic Diversity

We examined patterns of diversity along latitudinal gradients and across Florida ecoregions. To examine latitudinal trends, we binned the state into 13 non-overlapping latitudinal bins of 0.5 degree and examined the trend in PD by latitude. We also compared PD and RPD among the larger EPA Level III and Level IV ecoregions to test whether different ecoregions showed differing patterns, with particular attention to transitional ecoregions (e.g., Southern Coastal Plain). We overlaid maps of significant PD in both phylograms and chronograms and counted the number of significant pixels within each of the three ecoregions described above. We produced maps of observed RPD and RPD significance to make visual comparisons among the ecoregions.

Purpose-Built Trees Compared to Generated Trees

To compare our purpose-built tree to an automatically generated tree, we built a tree using Phylomatic (Webb and Donoghue 2005). Our query with the 1,490 species for which we had distribution models (above) to Phylomatic resulted in 1,476 species found. We used two versions of this tree. The first was simply the Phylomatic tree, without age calibration, as is sometimes but now less frequently used in the literature for PD calculation. The second was a chronogram version constructed using the bladj algorithm in Phylocom (Webb et al. 2008) using 76 calibration points (Wikström et al. 2001). This method is commonly used for assessing PD when sequence data for some or all species are lacking (Montesinos-Navarro et al. 2017; Qian et al. 2017; Burkle et al. 2013; Athayde et al. 2015).

The Open Tree of Life is a recent synthesis product and has not been used extensively in the literature for spatial phylogenetic analysis; thus, there are not yet public methods, such as the bladj algorithm

in Phylomatic, for obtaining branch lengths and calibrating the tree. Therefore, we downloaded sequence data for the available species from our taxon list (*n*=1,201, approximately 80% of the full taxon list) and added branch lengths with RAxML, using the Open Tree of Life (Hinchliff et al. 2015) tree as a constraint and our alignment with the same molecular model as above. Finally, we calculated an Open Tree of Life chronogram using the same method as for our purpose-built tree and the same calibration points. To compare our purpose-built tree to the trees resulting from each of these methods, we pruned our tree to the taxa available in each of these repositories and conducted the following analyses with the pruned trees.

Impact of Tree Uncertainty on Phylodiversity Metrics

Estimates and significance of PD and RPD were calculated for each of the 100 Bayes phylograms and chronograms, as described above. To determine the level of variation in each of these metrics across the trees, the standard deviation of the observed values for each pixel was calculated and mapped. Finally, the number of times each pixel was found to be significantly high, significantly low, or not significant was tallied, and the most frequent of those categories was mapped. A score of 100 for a pixel means that all trees produced the same category of significance for that pixel. In contrast, a score of 33 means that values for a pixel are significantly high, low, and not significant an equal number of times, suggesting that phylogenetic uncertainty is impacting the spatial phylogenetic data and potential interpretations thereof.

REFERENCES

Athayde, E. A., Cancian, L. F., Verdade, L. M., & Morellato, L. P. C. (2015). Functional and phylogenetic diversity of scattered trees in an agricultural landscape: Implications for conservation. Agriculture, Ecosystems & Environment, 199, 272–281.

Barve, N., & Barve, V. (2014). ENMGadgets: tools for pre and post processing in ENM workflows. github.com/narayanibarve/ENMGadgets

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., … Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling, 222, 1810–1819.

Bell, C. D., Soltis, D. E., & Soltis, P. S. (2010). The age and diversification of the angiosperms re-revisited. American Journal of Botany, 97, 1296–1303.

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., … Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for automated standardization of plant names. BMC Bioinformatics, 14, 16. DOI: 10.1186/1471-2105-14-16

Burkle, L. A., Marlin, J. C., & Knight, T. M. (2013). Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science, 339, 1611–1615.

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. Global Ecology and Biogeography: A Journal of Macroecology, 23, 99–112.

Cameron, K. M. (2004). Utility of plastid *psaB* gene sequences for investigating intrafamilial relationships within Orchidaceae. Molecular Phylogenetics and Evolution 31,1157-1180.

CBOL Plant Working Group. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106,12794-12797.

Chase, M. W, & Hills, H. G. (1991). Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40, 215-220.

Clayton, J.W., Fernando, E.S., Soltis, P.S. & Soltis, D.E. (2007). Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. International Journal of Plant Science 168, 1325–1339. doi:10.1086/521796

Cuénoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J., & Chase, M.W. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. American Journal of Botany 89, 132-144.

Doyle, J. J., Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19, 11-15.

Dunning, L. T., & Savolainen, V. (2010). Broad-scale amplification of matK for DNA barcoding plants, a technical noteboj_1071 1..9. Botanical Journal fo the Linnean Society 164, 1-9.

Guralnick, R. P., Wieczoreck, J., Beaman, R., & Hijman, R. J. (2006). BioGeomancer: Automated georeferencing to map the worlds biodiversity data. PLoS Biology, https://doi.org/10.1371/journal.pbio.0040381

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965–1978.

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., … Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences of the United States of America, 112, 12764–12769.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. PloS One, 6(5), e19254. doi:10.1371/journal.pone.0019254

Laffan, S. W., & Crisp, M. D. (2003). Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. Journal of Biogeography, 30, 511–520.

Laffan, S. W., Lubarsky, E., & Rosauer, D. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. Ecography, 33, 643–647.

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution, 29, 1695–1701.

Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. Nature Communications, 5, 4473. DOI: 10.1038/ncomms5473.

Montesinos-Navarro, A., Verdú, M., Querejeta, J. I., & Valiente-Banuet, A. (2017). Nurse plants transfer more nitrogen to distantly related species. Ecology, 98 (5), 1300-1310.

Neubig, K. M., Whitten, W. M., Abbott, J. R., Elliott, S., Soltis, D. E., & Soltis, P. S. (2014). Variables affecting DNA preservation in archival plant specimens. In: Applequist, WL, Campbell, LM eds. *DNA Banking for the 21st Century: Proceedings of the U.S. Workshop on DNA Banking.* St. Louis, MO: The William L. Brown Center at the Missouri Botanical Garden.

Nicolalde-Morejón, F., Vergara-Silva, F., González-Astorga, J., Stevenson, D. W., Vovides, A. P., & Sosa, V. (2011). A character-based approach in the Mexican cycads supports diverse multigene combinations for DNA barcoding. Cladistics 27, 150-164.

Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. Environmental Management, 54, 1249–1266.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259.

Qian, H., Jin, Y., & Ricklefs, R. E. (2017). Patterns of phylogenetic relatedness of angiosperm woody plants across biomes and life-history stages. Journal of Biogeography, 44, 6. DOI: 10.1111/jbi.12936

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.

Sakumaran, J. & Holder, M. T. (2010) DendroPy: a Python library for phylogenetic computing. Bioinformatics, 15, 1569-1571.

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.

Shokralla, S., Singer, G. A. C., & Hajibabei, M. (2010). Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. Biotechniques 48, 233-234.

Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P. R., & Barraclough, T. G. (2002). Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proceedings of the National Academy of Sciences of the United States of America, 99, 4430–4435.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. Systematic Biology, 57, 758–771.

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

Sun, H., McLewin, W., & Fay, M.F. (2001). Molecular phylogeny of *Helleborus* (Ranunculaceae), with an emphasis on the East Asia–Mediterranean disjunction. Taxon 50, 1001-1018.

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research, *34*, W609–12.

Thornhill, A. H., Mishler, B. D., Knerr, N., Gonzalez-Orozco, C. E. , Costion, C. M., Crayn, D. M., Laffan, S. W., & J. T. Miller. (2016). Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. Journal of Biogeography. 43, 2085–2098.

Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M. B., Balaguer, L., Benito-Garzón, M., …& Zavala, M. A. (2014). The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecology Letters, *17*, 1351–1364.

Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for applied phylogenetics. Molecular Ecology Notes, 5, 181–183.

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. Annual Review of Ecology and Systematics, 33, 475–505.

Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics, 24, 2098–2100.

Wikström, N., Savolainen, V., & Chase, M. W. (2001). Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society of London B, 268, 2211–2220.

Wunderlin, R. P. & Hansen, B. F. (2011). Guide to the vascular plants of Florida. University Press of Florida.

Wunderlin, R. P., Hansen, B. F., Franck, A. R., & Essig, F. B. (2017). Atlas of Florida Plants (http://florida.plantatlas.usf.edu/). [Landry, S. M., & Campbell, K. N. (application development), USF Water Institute.] Institute for Systematic Botany, University of South Florida, Tampa.

Yu, J., Xue, J. H., & Zhou, S. L. (2011). New universal matK primers for DNA barcoding angiosperms. Journal of Systematics and Evolution 49, 176-181.

Zurawski, G., Clegg, M. T., & Brown, A. H. D. (1984). The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106, 735-749.

Supplementary Items Figure S1 – Table S4

Figure S1: Three ecoregions in Florida. Related to figures 1, 2, and 3.

Figure S2: Phylogenetic diversity correlated with species richness. Related to Figure 1

Figure S3: Map of Florida endembics. Related to Figure 2

Figure S4: Posterior probabilities of sample of trees vs. full distribution. Related to Figure 1

Figure S5: Values for the training AUC, testing AUC and ratio. Related to Figures 1, 2

Table S1: Primer names, sequences, and references. Related to Figure 1

Table S2: Models tested in Partition Finder. Related to Figure 1

Table S3: Placement of missing taxa. Related to Figure 1

Table S4: Calibration points and references. Related to Figure 1

Table S5: Locality data and taxonomic information of all species. Related to Figures 1, 2, 3, 4, 5, 6, 7

Figure S1: Three ecoregions in Florida. Related to figures 1, 2, and 3.

EPA Ecoregions of Florida level III (legend) and level IV (shades of brown, green and purple). The smaller map shows how the Southeastern Plain and Southern Coastal Plain regions extend beyond the Florida border. Maps modified from http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm.
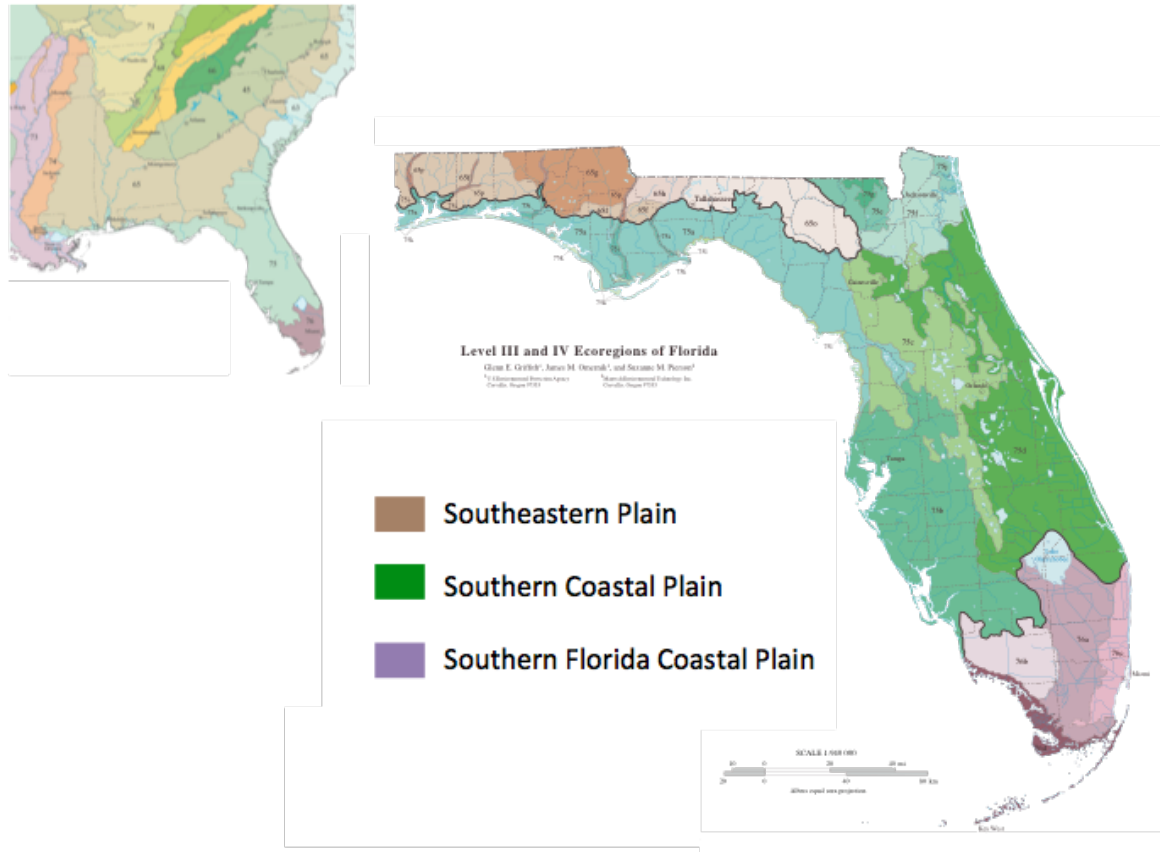
Figure S2: Phylogenetic diversity correlated with species richness. Related to Figure 1

Correlations of Phylogenetic Diversity and Species Richness for all 8,045 points in Florida for A) Phylogram B) Ultrametric Tree. Correlation Coefficient: Phylogram 0.9807995, Chronogram 0.9651092.
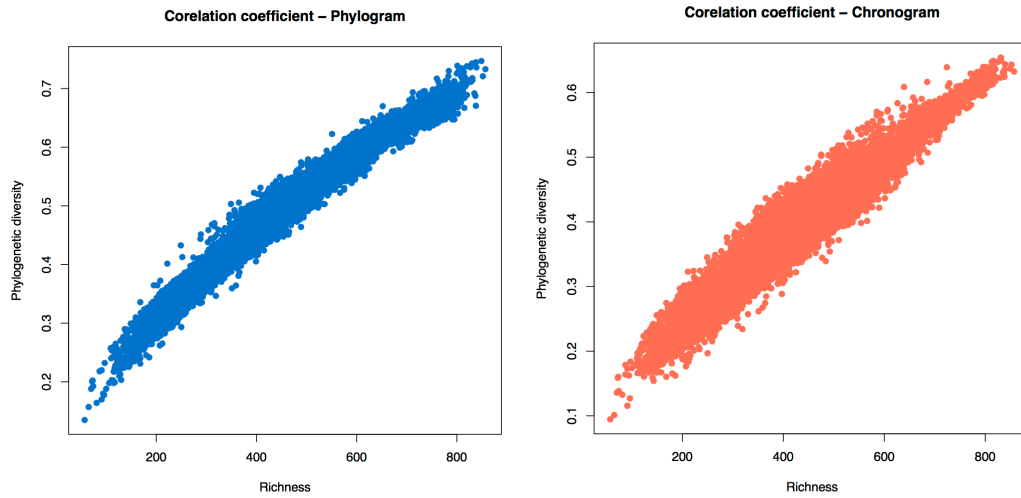

Figure S3: Map of Florida endembis. Related to Figure 2

Map of Florida endemics only in Florida overlaid with the Level 4 ecoregions - The majority of the endemic species are found along the Lake Wales Ridge area.
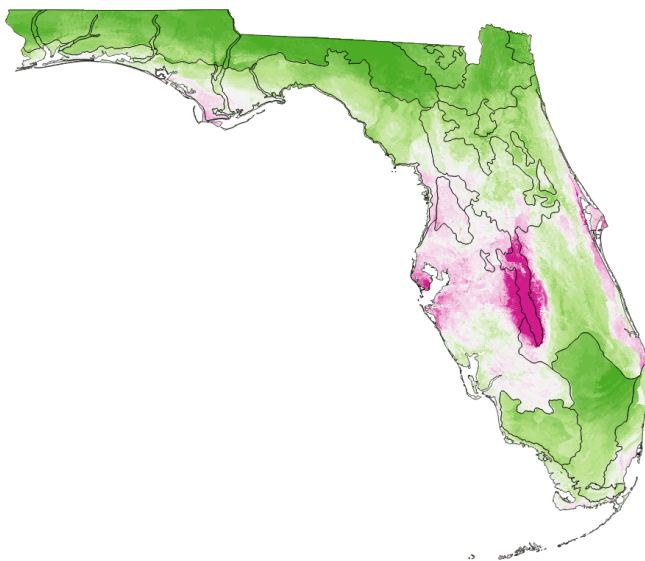Endemic Map of Florida Max = 38.964 Min = 1.989.

Figure S4: Posterior probabilities of sample of trees vs. full distribution. Related to Figure 1

The x-axis represents the posterior probabilities calculated from the entire post-burnin distribution of trees. The y-axis is the posterior probabilities calculated from our 100 sampled trees. The graph indicates a tight correlation (linear model; $p < 0.001$, $r^2 = 0.98$) between the full distribution and our sample suggesting that the sampled trees are a good representation of the variation found among the full distribution of trees.
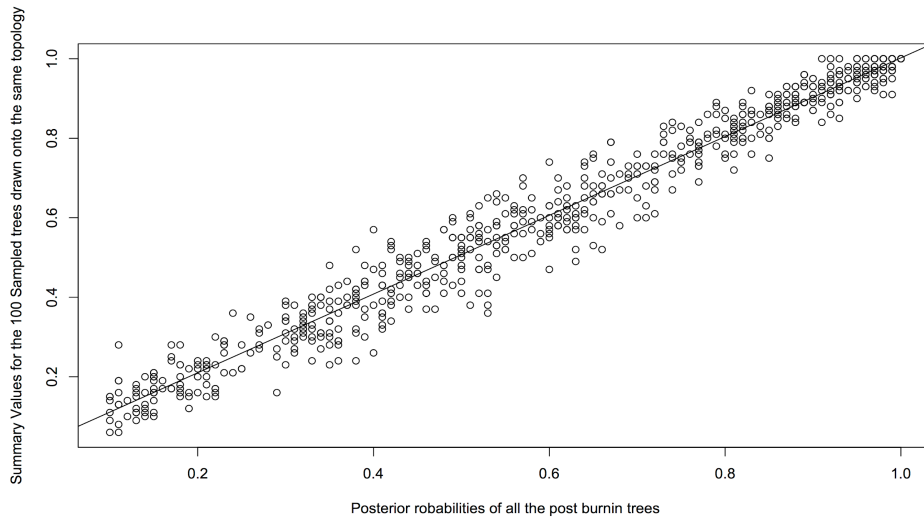

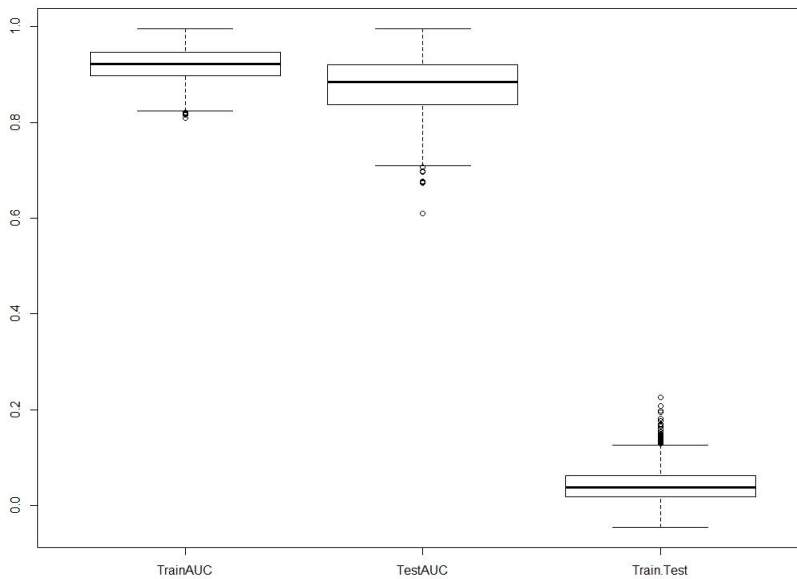**Figure S5:** Values for the training AUC, testing AUC and ratio. Related to Figures 1, 2

Table S1: Primer names, sequences, and references: Related to Figure 1

Primer names, sequences, and references used for both *matK* and *rbcL* amplifications along with the original references for the sequences.

Primer Name	Sequence	Reference
matK		
390F	CGATCTATTCATTCAATATTTC	Cuénoud et al. (2002)
1326R	TCTAGCACACGAAAGTCGAAGT	Cuénoud et al. (2002)
F (Equisetum)	ATACCCCATTTTATTCATCC	Nicolalde-Morejónet al., (2011)
R (Equisetum)	GTACTTTTATGTTTACGAGC	Nicolalde-Morejónet al., (2011)
XF	TAATTTACGATCAATTCATTC	Sun et al. (2001)
Malp_R1	ACAAGAAAGTCGAAGTAT	Dunning & Savolainen (2010)
3F	CGTACAGTACTTTTGTGTTTACGAG	CBOL working group (2009)
1R	ACCCAGTCCATCTGGAAATCTTGGTTC	CBOL working group (2009)
427F	CCCRTYCATCTGGAAATCTTGGTTC	Yu et al. (2011)
1248R	GCTRTRATAATGAGAAAGATTTCTGC	Yu et al. (2011)
rbcL		
Z1	ATGTCACCACAAACAGAAACTAAAGCAAGT	Zurawski et al. (1984)
3'	CTCGGAGCTCCTTTTAGTAAAAGATTGGGCCGA	Zurawski et al. (1984)
rbcLaF	ATGTCACCACAAACAGAGACTAAAGC	Shokralla et al. (2010)
rbcLaR	GTAAAATCAAGTCCACCRCG	Shokralla et al. (2010)
NY35	CTTCACAAGCAGCAGCTAGTTC	Cameron (2004)
NY149	ATGTCACCACAAACAGAAAC	Cameron (2004)

Table S2: Models tested in Partition Finder. Related to Figure 1.

Partition Finder Models:
allsame = (Gene1_pos1,Gene1_pos2,Gene1_pos3,Gene2_pos1,Gene2_pos2,Gene2_pos3);
alldiff = (Gene1_pos1) (Gene1_pos2) (Gene1_pos3) (Gene2_pos1) (Gene2_pos2) (Gene2_pos3);
12_3 = (Gene1_pos1, Gene1_pos2, Gene2_pos1, Gene2_pos2) (Gene1_pos3,Gene2_pos3);
genes = (Gene1_pos1, Gene1_pos2, Gene1_pos3) (Gene2_pos1, Gene2_pos2, Gene2_pos3);

Table S3: Placement of missing taxa. Related to Figure 1

Taxa were placed next to known congener if available. For taxa for which no congener was represented on the phylogeny the literature was searched to identify their closest relative on the phylogeny and that taxon is identified and the literature cited.

Missing Taxon	Where to Place	Source*
Aeschynomene_pratensis	Aeschynomene_indica	
Agave_decipiens	Agave_sisalana	
Ageratina_altissima	Ageratina_aromatica	
Aletris_aurea	Aletris_lutea	
Aletris_farinosa	Aletris_lutea	
Andropogon_arctatus	Andropogon_brachystachyus	
Anemia_adiantifolia	Lygodium_japonicum	Pryer et al., 2004
Arnoglossum_diversifolium	Arnoglossum_floridanum	
Asplenium_dentatum	Asplenium_platyneuron	
Asplenium_pumilum	Asplenium_platyneuron	
Bletia_purpurea	Epidendrum_nocturnum	Cameron et al., 1999
Carex_nigromarginata	Carex_baltzellii	
Catopsis_berteroniana	Tillandsia_bartramii	Givnish et al., 2007
Catopsis_floribunda	Tillandsia_bartramii	Givnish et al., 2007
Centrosema_arenicola	Centrosema_virginianum	
Chamaesyce_deltoidea	Chamaesyce_maculata	

Chamaesyce_garberi	Chamaesyce_maculata	
Chrysopsis_delaneyi	Chrysopsis_lanuginosa	
Ctenitis_sloanei	Dryopteris_ludoviciana	Liu et al., 2007
Delphinium_carolinianum	Clematis_terniflora	Emadzade et al., 2010
Digitaria_pauciflora	Digitaria_ciliaris	
Eragrostis_amabilis	Eragrostis_elliottii	
Erythronium_umbilicatum	Lilium_superbum	Allen et al., 2003; Fay et al., 2006; Clennett et al., 2012
Euphorbia_polyphylla	Euphorbia_exserta	
Euphorbia_rosescens	Euphorbia_exserta	
Galium_obtusum	Galium_hispidulum	
Gratiola_virginiana	Gratiola_pilosa	
Habenaria_floribunda	Habenaria_quinqueseta	
Harrisia_aboriginum	Harrisia_fragrans	
Harrisia_simpsonii	Harrisia_fragrans	
Helenium_flexuosum	Helenium_amarum	
Helianthus_resinosus	Helianthus_angustifolius	
Hymenocallis_palmeri	Hymenocallis_henryae	
Hypericum_drummondii	Hypericum_brachyphyllum	
Hypericum_edisonianum	Hypericum_brachyphyllum	
Hypericum_exile	Hypericum_brachyphyllum	
Ipomoea_microdactyla	Ipomoea_cordatotriloba	
Jacquemontia_pentanthos	Jacquemontia_tamnifolia	
Jacquemontia_reclinata	Jacquemontia_tamnifolia	
Lechea_divaricata	Lechea_deckertii	
Liatris_gholsonii	Liatris_chapmanii	
Lilium_catesbaei	Lilium_superbum	
Lilium_iridollae	Lilium_superbum	
Linum_westii	Linum_medium	
Ludwigia_decurrens	Ludwigia_microcarpa	
Lysiloma_latisiliquum	Acacia_auriculiformis	Miller et al., 2003

Macranthera_flammea	Seymeria_pectinata	Bennett & Mathews, 2006
Malaxis_unifolia	Epidendrum_nocturnum	Cameron et al., 1999
Manilkara_jaimiqui	Manilkara_zapota	
Marshallia_graminifolia	Balduina_atropurpurea	Watson et al., 1991
Matelea_alabamensis	Matelea_floridana	
Micranthemum_glomeratum	Micranthemum_umbrosum	
Microgramma_heterophylla	Phlebodium_aureum	Schneider et al., 2004
Monotropa_uniflora	Ceratiola_ericoides	Kron et al., 2002
Nymphoides_cordata	Nymphoides_aquatica	
Okenia_hypogaea	Boerhavia_diffusa	Douglas & Manos, 2007
Oldenlandia_salzmannii	Oldenlandia_corymbosa	
Ophioglossum_palmatum	Botrychium_lunarioides	Pryer et al., 2004
Panicum_repens	Panicum_dichotomiflorum	
Parnassia_caroliniana	Parnassia_grandifolia	
Pecluma_dispersa	Phlebodium_aureum	Schneider et al., 2004
Pecluma_plumula	Phlebodium_aureum	Schneider et al., 2004
Phegopteris_hexagonoptera	Macrothelypteris_torresiana	He & Zhang, 2012
Pinguicula_ionantha	Pinguicula_lutea	
Pinguicula_planifolia	Pinguicula_lutea	
Piptochaetium_avenacioides	Piptochaetium_avenaceum	
Platanthera_clavellata	Platanthera_ciliaris	
Platanthera_cristata	Platanthera_ciliaris	
Platanthera_integra	Platanthera_ciliaris	
Polygala_smallii	Polygala_lutea	
Potamogeton_floridanus	Potamogeton_diversifolius	
Ranunculus_pusillus	Clematis_terniflora	Emadzade et al., 2010
Sabatia_decandra	Sabatia_grandiflora	
Salvia_urticifolia	Salvia_azurea	
Schoenolirion_croceum	Schoenolirion_albiflorum	
Sida_ulmifolia	Sida_cordifolia	
Solanum_donianum	Solanum_americanum	

Spermacoce_terminalis	Spermacoce_remota	
Spigelia_gentianoides	Spigelia_loganioides	
Stachys_hyssopifolia	Stachys_floridana	
Stenanthium_densum	Schoenocaulon_dubium	Zomlefer et al., 2001
Stylisma_humistrata	Stylisma_abdita	
Tectaria_fimbriata	Tectaria_heracleifolia	
Tephrosia_angustissima	Tephrosia_rugelii	
Tephrosia_spicata	Tephrosia_rugelii	
Tillandsia_flexuosa	Tillandsia_bartramii	
Tillandsia_simulata	Tillandsia_bartramii	
Tragia_urticifolia	Tragia_saxicola	
Trichomanes_punctatum	Thelypteris_interrupta	Dubuisson et al., 2003; Pryer et al., 2004
Triphora_craigheadii	Epidendrum_nocturnum	Cameron et al., 1999
Utricularia_cornuta	Utricularia_foliosa	
Utricularia_resupinata	Utricularia_foliosa	
Utricularia_simulans	Utricularia_foliosa	
Verbena_simplex	Verbena_officinalis	
Vicia_ocalensis	Vicia_acutifolia	
Viola_bicolor	Viola_lanceolata	
Xyris_laxifolia	Xyris_difformis	
Ziziphus_celata	Ceanothus_americanus	Richardson et al., 2004; Islam & Simmons, 2006

Table S4: Calibration points and references. Related to Figure 1

Node	Calibration	Reference
Fabales	min_age=59.9	Bell et al., 2010
Arecales	min_age=65	Bell et al., 2010
Sapindales	min_age=65	Bell et al., 2010
Malvales	min_age=65.5	Bell et al., 2010; based on Wheeler et al., 1987; 1994
Poales	min_age=68.1	Bell et al., 2010
Myrtales	min_age=88.2	Bell et al., 2010
Caryophyllales	min_age=83.5	Bell et al., 2010
Cornales	min_age=85.8	Bell et al., 2010
Ericales	min_age=91.2	Bell et al., 2010
Fagales	min_age=87.5	Takahashi et al. 2008; Magallon et al. 2015
Magnoliales	min_age=108	Doyle and Endress 2010; Magallon et al. 2015
Polypods	min_age=121	Schneider et al., 2004
Eudicots	min_age=125	Bell et al., 2010
Gymnosperms	min_age=290	Bell et al., 2010
Monilophytes	age=354	Schneider et al., 2004
Euphyllophyta	min_age=380	Schneider et al. 2004
Tracheophyta	min_age=416	Clarke et al., 2011