Phylogenetic diversity and endemism: metrics for identifying critical regions of conifer conservation in Australia

Annasophie C. Lee

ABSTRACT

Accurately and sufficiently quantifying biodiversity is integral for conservation. Traditional metrics for measuring biodiversity, species richness (SR) and weighted endemism (WE), do not take into account the evolutionary history of organisms. Phylogenetic diversity (PD) addresses the shortcomings of SR by quantifying the evolutionary connections among the species present in an area. Phylogenetic endemism (PE) addresses the shortcomings of WE and represents the ranges of the branches of the evolutionary tree connecting the species in an area. Australia, with its advanced digitization of spatial reference data is the best model system for quantitative studies of biodiversity at present. I created a phylogeny for the 39 indigenous Australian conifer species using *matK* and *rbcL* sequences from *GenBank* and sequencing the 4 species for which there were no existing data. I used spatial data from Australia's Virtual Herbarium and removed records of conifers that were not naturally-occurring. I then used *Biodiverse v 0.17* to calculate PD, PE and two derived metrics, Relative Phylogenetic Diversity (RPD) and Relative Phylogenetic Endemism (RPE). These metrics identify regions with statistically significantly high or low levels of PD and PE based on randomization tests. The results show that conifer RPD is significantly low on the Northeast coast of Australia, although conifer PD is high in the same region. RPE is significantly high along the Northeast coast, in the same region that RPD is low. Most of these regions are currently encompassed by legally protected Australian Reserves. More precise estimates of biodiversity can be used by conservation policy-makers.

KEYWORDS

Australian endemics, biodiversity conservation, phylogeny, *Biodiverse, ArcMap*

INTRODUCTION

Conserving global biodiversity, the variability between organisms, species, or ecosystems (Jensen et al. 1990), is integral for conservation efforts (Pimm et al. 1995, Reaka-Kudla 1997). However, prioritization of critical species or regions for biodiversity conservation is a major challenge for conservation policy-makers from a number of perspectives. Historically, conservation efforts have often been focused on either conserving key species or regions (Lombard et al. 2003). To identify key regions and species for conservation, measures of endemism have played a central role to quantify how restricted a species is to a given region. The degree to which species are restricted or widely dispersed is a strong predictor of extinction risk (Gaston and Fuller 2009). Identifying these species at risk for extinction can be based on evolutionary history, geographic location, or a combination of the two. Geographically rare species are at greater risk of extinction (Gaston and Fuller 2009), and phylogenetically rare species (Crozier 1997) contain disparate genetic information and contribute heavily to biodiversity, thus it is critical to examine the intersection of these subjects.

The quantification of biodiversity has historically been problematic and current metrics are problematic because they do not include an evolutionary perspective. For example, enumeration of species is hindered by the lack of a universal agreed-upon species concept across researchers reflecting an arbitrarily decided level of genetic and morphological variation which leads to inconsistency in taxonomic ranking or hierarchy (Nixon and Wheeler 1990, Mishler 2009). Additionally, inconsistencies in identification and discovery of species lead to false classifications that both under and overestimate biodiversity._ More importantly, these issues with naming and identifying species are compounded when traditional biodiversity metrics are calculated without considering evolutionarily relatedness between species, and their dispersal from their geographic origins. Species richness (SR), the absolute number of species in a region, was developed to quantify the number of species in a region and weighted endemism (WE) quantified their level of endemism (Crisp et al. 2001, Chao 2005). However, SR and WE as measures of biodiversity consider only the terminal taxa of a phylogenetic tree, without considering the evolutionary relationships among them (Rosauer et al. 2009). Species vary in their evolutionary isolation and genetic diversity, and these differences give insight into how species may have evolved and which are most important for biodiversity conservation (Mooers

2007). SR and WE do not include information about how closely related species are, excluding relationships between sister groups given by the phylogeny. Consequently, these metrics are limited in their ability to describe biodiversity patterns as they are a more surface-level analysis of biodiversity as compared to one that incorporates the evolutionary perspective (Mooers 2007, Faith 1997, Rosauer 2009).

Diversity measures based on phylogeny, or the evolutionary relationships between species, have since been developed to address the shortcomings of descriptors such as species richness and species endemism. Phylogenies are derived from shared, homologous characters, or characteristics shared by all the descendants of a common ancestor, and are an indication of recently shared ancestry (Eldridge and Cracraft, 1980). Phylogenetic diversity (PD) and phylogenetic endemism (PE) are metrics that provide a more comprehensive view of diversity within and between species (Faith and Baker 2006, Meiri and Mace 2009, Rosauer et al. 2009, Davies and Buckley 2011). PD calculates the shared evolutionary history of specified taxa (Davies and Buckley 2011, Rosauer et al. 2009) and is largely resistant to taxonomic uncertainty, or the discrepancies in the identification of species, because it relies on robust hypothesis of evolution, derived from the shared homologous characters between species (Mace et al. 2003). PD has been utilized to understand global patterns of biodiversity, and is especially useful when the taxonomy of a clade is poorly understood (Meiri and Mace 2009). PE is a measure of the amount of shared evolutionary history between a set of branches on a phylogenetic tree in relation to how widespread the branches are geographically (Rosauer et al 2009). WE is the sum of the inverse of the species' range found over a fixed area (Crisp et al. 2001, Knapp 2002). PE, unlike WE, incorporates the ranges of all the branches of the tree connecting the species, not just the terminal branches (Roasuer et al. 2009). This weighted phylogenetic endemism provides a more comprehensive measure of the distribution of rarity than weighted endemism of species alone. PD and PE are more robust to changes in taxonomic classification than SR and WE, and PE analyzes endemism across a consistent spatial scale, regardless of previously defined geographic boundaries (Rosauer et al. 2009). These metrics provide evolutionary and genetic information necessary for making informed conservation-policy.

Calculating PD and PE requires a high resolution of spatial distribution information along with a highly resolved phylogeny. Australia is the best model system, at present, for this type of study due to the advanced state of digitization of herbaria voucher specimens and spatial

reference data (Nagalingum et al. in prep, Mishler et al. in review). Australia's Virtual Herbarium (AVH), contains millions of records of spatial flora collections from Australia's major Herbaria. Additionally, Australia is important for global biodiversity conservation as it is rich with endemic species, resulting from its geographic isolation (Ingelby 2009, Crisp et al. 2001). Australia has a high diversity of conifers compared to many regions in the world, especially in contrast with Northern Hemisphere where conifers are abundant but not diverse (Leslie et al. 2012). Conifers are also largely confined to either the Northern or Southern hemisphere, specifically extant species of Araucariaceae, Podocarpaceae, and the Callitroideae (the sister group to Cupressoideae), fossil records also indicate that these trends have persisted throughout time (Leslie et al. 2012). Although metrics such as species richness and species endemism have been calculated for many conifer species in Australia (Pimm et al. 1995, Austin et al. 1996) calculation of diversity metrics from an evolutionary perspective using PD and PE remains to be accomplished.

The main objective of this study was to calculate and visually display diversity metrics that couple phylogenetic and spatial information. I calculated PD and PE to identify regions of Australia most densely populated with phylogenetically rare conifers and compared these results with Australian natural reserves to identify regions of phylogenetic rarity that are not currently being protected. I hypothesized that species which are evolutionarily distant will be more geographically distant and closely-related species will be spatially clustered (Forest et al. 2007). Additionally, I evaluated the relationships between PD, PE and traditional diversity metrics such as SE and WE. I expected PE to be correlated with WE; however I expected WE to fail at consistently predicting areas of high PE (Rosauer et al. 2009). These results will prove valuable to informing conservation-policy makers regarding critical regions of conifer conservation.

METHODS

Spatial data acquisition

I studied the 39 indigenous species of conifers in Australia (Appendix A, List 1). To obtain specimen locations, I used data from AVH (http://avh.ala.org.au/). The AVH is a digital database containing 75% of the 6 million specimens of plants, algae and fungi that have been collected by Herbaria in Australia. I downloaded a total of 12,300 Australian endemic conifer species datapoints and then used Google Refine version 2.5 (http://code.google.com/p/google-refine/), to clean the dataset and remove non-conifer records, foreign collections (as well as Norfolk and Macquarie Islands), and any naturalized specimens grown in a botanic garden or otherwise. I reconciled the taxonomy against a classification for extant conifers with the Australian Plant Census (APC) (http://www.anbg.gov.au/chah/apc/index.html) and corrected any misspellings. I trimmed records without geographic coordinates from the dataset. I then transformed the latitude and longitude values of the remaining records into xy meter coordinates using the Albers projection which corrects for inconsistencies in grid size of latitude and longitude near the earth's poles. This cleaned dataset contained 7300 spatial records (Fig.1)

Fig.1 Spatial location of individual conifer specimens. Specimens collected using AVH database.

Molecular data acquisition: Phylogeny

I used two genes, *matK* and *rbcL* to create a phylogeny, using both existing and new sequence data. *RbcL* is commonly referred to as the "universal barcode" for plants; however using two genes, both *matK* and *rbcL*, is more informative and created a more complete and accurate phylogeny (Quinn et al. 2002). I searched the online database *GenBank* (http://www.ncbi.nlm.nih.gov/genbank/) using scientific names of each of the 40 species in my study. I noted which sequences were unavailable in *GenBank* and saved the accession numbers of the available sequences (Appendix A). Once I identified which species were missing, I collected plant tissue for *Callitris baileyi*, *Callitris monticola*, *Callitris oblonga*, *Callitris columellaris*, *Actinostrobus acuminatus* and *Microstrobos niphophilus* at the Royal Botanic Garden, Sydney. I located the species I needed in the botanic garden and cut a piece of fresh leaf from which to extract DNA.

After I cataloged the plant tissue in the Royal Botanic Garden Herbarium's collection, I prepped my tissue samples for DNA extraction by sealing them in a silica gel filled box to desiccate them. I then performed DNA extractions using a Qiagen DNEasy Kit (Germany,www.qiagen.com) with minor modifications. These modifications were: using 1 zirconia bead and 5 mg sand instead of 50 µL small zirconia beads, not using any liquid nitrogen, using the lyser (written "bead-beater" in kit) for 25 seconds, incubating at 65°C for 40 minutes, and incubating the products of buffer AE and DNA for 10 minutes .

Once I extracted DNA from the leaf tissue, I amplified the regions *matK* and *rbcL* using PCR. I performed the standard procedure using the primers Forward TX2 and Reverse TX4 to amplify *matK* regions and the primers Forward *rbcL*_1 and Reverse *rbcL*_635. I ran a program called Immolase 50°C on the Thermocycler (Corbett Life Science, Palm-Cycler) for 2.5 hours. Then I loaded the product into wells on gels and ran electrophoresis on the gel with indicator and gel red at 300 W for approximately 10 minutes, checking to see the movement of the bands periodically. I then transferred the plates to a UV hood and visualized the plates. After taking note of which trials were successful, I collected the PCR products for sequencing. I then sent the PCR products to the Genetic Sequencing Lab on the UC Berkeley campus. The sequenced products were then sent back to me as a data file.

Phylogeny Construction

To create the Australian conifer phylogeny, I acquired DNA sequences from the processes outlined above and used the default settings for the MUSCLE alignment in Geneious (http://www.geneious.com/) to align the sequences for each gene region, *matK* and *rbcL*. Once I aligned the genetic sequences, I deleted any unreliable end pieces that were unlikely to represent *rbcL* or *matK* gene regions. I chose one *matK* and one *rbcL* sequence to represent each species using the following criteria, known as taxon priming: longest sequence, a sequence which withstands a cluster analysis, and Australian in origin. I then created a concatenated matrix including *rbcL* and *matK*, a total of 2783 base pairs, and used the default parameters in GARLI (Genetic Algorithm for Rapid Likelihood Inference) version 0.951 (https://code.google.com/p/garli/) to create a Maximum Likelihood phylogeny (Fig 2). I then compared the relationships in the phylogeny I created with previously published conifer phylogenies (e.g., Leslie et al. 2012).

Biodiverse: Spatial Location and Phylogeny

Biodiverse v 0.17 (http://code.google.com/p/biodiverse/) is a program that uses a phylogeny and specimen level spatial data to create a map of the occurrence of species across a region and calculates SR, WE and phylogenetic metrics PD and PE. SR and WE require only spatial data, whereas PD and PE require spatial and phylogenetic data. I loaded the cleaned spatial data I acquired from AVH into *Biodiverse* which displayed a map each species' occurrence (Fig. 1) and the phylogeny I created from the gene regions *matK* and *rbcL (*Fig 2).

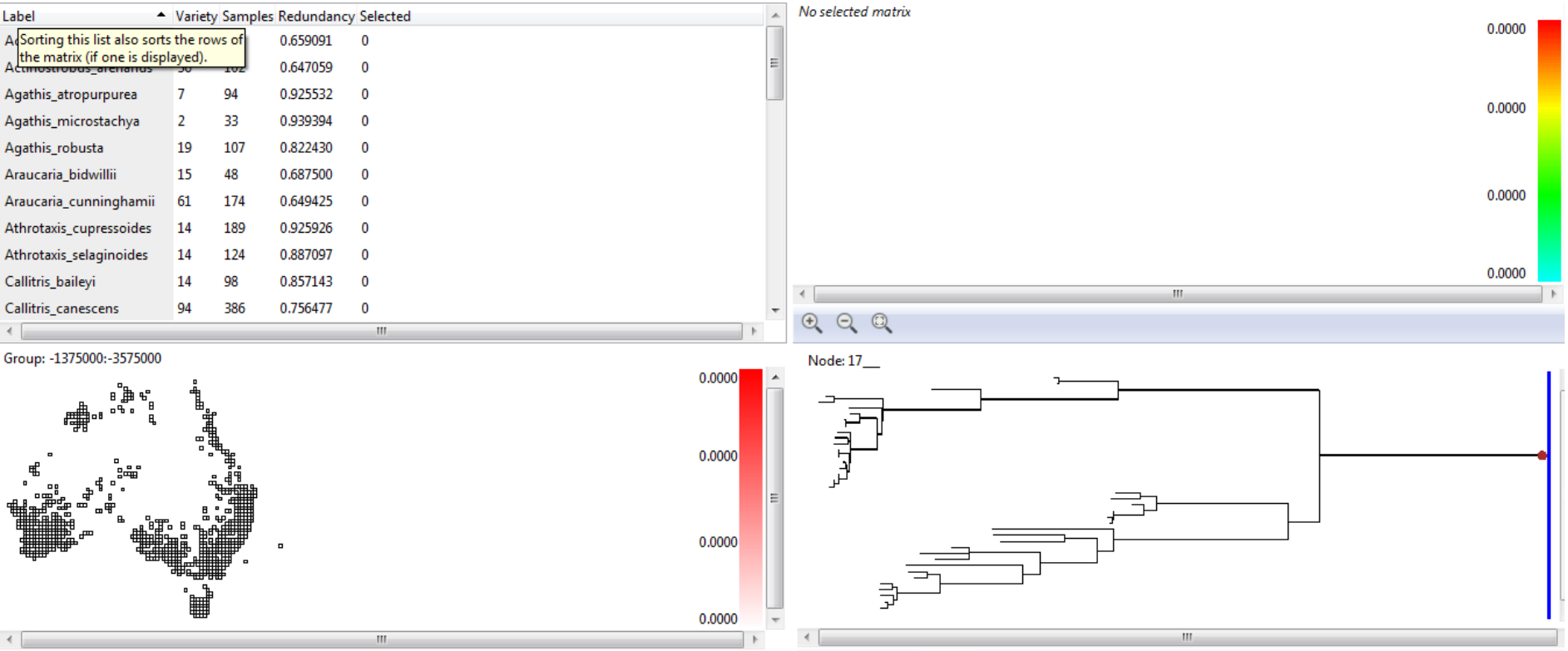


Fig 2. Screenshot of Biodiverse interface. The lower left displays the 50,000m^2 grid cells with species' occurrence data. The lower right is the phylogeny without *Ginkgo biloba*, because Gingko does not occur in Australia. The upper left displays the species names, number of occurrences in each grid cell and the redundancy value (how many samples were downloaded from AVH).

First, I calculated species richness—defined as the number of species in an area (here represented by 50,000 m^2 grids). Second, I calculated PD (Eq 1, Rosauer et al 2009), which is calculated by summing the branch lengths on the phylogenetic subtree connected the species in a particular gird. Third I calculated PE, defined as PD weighted by the inverse of the branchlength's ranges. PE incorporates the spatial range of the phylogenetic branch lengths down to the root of the phylogeny (Rosauer et al. 2009). For example, if a widely distributed taxon is sister to a narrowly distributed (highly endemic) species, the highly endemic species will be negatively weighted by its sister and the PE score of the pair will be lowered.

$$\mathbf{PD} = \sum_{c \in C} L_c \qquad \text{(Eq. 1)}$$

where L_c is the length of branch c and C is the set of branches in the minimum spanning path connecting the species (Rosauer et al. 2009).

$$\mathrm{PE} = \sum_{\{c \in C\}} \frac{L_c}{R_c} \qquad \text{(Eq. 2)}$$

Where variables are defined as above, and R_c is clade range, the combined ranges of the descendant taxa of branch *c*, so that overlapping areas are considered only once (Rosauer et al. 2009).

To discern any correlations between SR and PD, I created a scatterplot of SR as a percent of total number of species against PD. I performed the same calculation for WE versus PE, to graphically display any correlation between the two metrics. I also calculated the correlation coefficient for each relationship (r^2).

Using PD and PE alone is insufficient, because they are biased by the number of taxa in a cell. For example, PD is expected to be greater when there are more taxa present since more of the phylogenetic tree is accounted for in that cell (Faith and Baker 2006). Thus, PD may be greater as a result of collecting effort compared to regions with fewer collections. I therefore derived more informative metrics, Relative Phylogenetic Diversity (RPD) and Relative Phylogenetic Endemism (RPE) from PD and PE through a standardization and randomization process (Mishler et al. in review). First, I standardized the PD value for each grid cell by the number of taxa (PD/SR), calculating RPD. To

calculate RPE, I weighted PE by WE (PE/WE). To test for significance of the RPD and RPE results, I ran a randomization in *Biodiverse*. This randomization process keeps the number of taxa present in each grid cell constant, based on the number present in the real data set, and randomly draws that number of taxa from the phylogeny. The width of the distribution of the taxa is also held constant, for instance, if a taxon is present in 50 grid cells in the real data set, it will only be present in 50 grid cells in the randomization. This process is repeated 999 times and yields a two-tailed hypothesis test that asks: is RPD (or RPE) for a grid cell significantly higher or lower than one would expect given a random distribution of that number of species? If the RPD or RPE of the real data falls in the top or bottom 2.5% of the 1000 values, the RPD/RPE of that cell is statistically significant. A final metric was necessary a result of the inability of the randomization of RPE to identify regions that were significantly high in both PE *and* WE. In other words, some grids cells contained a significantly high level of species endemism *and* a high level of phylogenetic endemism, and the high WE essentially cancelled out the high PE in the RPE metric (PE/WE). Thus, I needed to calculate a metric called "super-endemism" (Mishler et al. in review) for grid cells significantly high in both WE and PE.

Spatial Analysis

To determine whether areas of significant RPD and RPE were correlated with protected regions in Australia, I overlaid map layers of Natural Parks and Reserves in Australia using *ArcMap v 10.1* (GISESRI). I gathered the data layers from the Atlas of living Australia (http://spatial.ala.org.au/) and loaded the data layers into *ArcMap*, projected them, if they were not already in the projection GDA94 / Australian Albers**.** I then projected the *Biodiverse*-exported ASCII grid files into the same projection to visualize properly. I clipped the data if it contained more data points than the continent of Australia. Then I symbolized the data to display the Australian Protected regions shapefile (CAPAD 2010) and overlaid an outline of the shape of Australia in GDA94/Australian Albers Projection to display the continent's bounds. I used this visualization process to discern any correlations or patterns in the data layers over the randomization maps of RPD, RPE and super-endemism.

RESULTS

Study organisms and study site

The phylogeny I created is fully resolved, and provides a robust hypothesis of the evolutionary relationships between Australian endemic clades. However, it probably includes an incorrect relationship: *Microstrobos niphophilus* probably belongs in the same clade as *Microstrobos fitzgeraldii* (Leslie et al. 2012). For the purposes of these calculations it does not make a difference, because both PD and PE take into account branch lengths, and the erroneous branch is very short. Fig. 2 is the result of a maximum likelihood phylogenetic tree for the 39 conifer species, rooted on the outgroup, *Gingko biloba*.

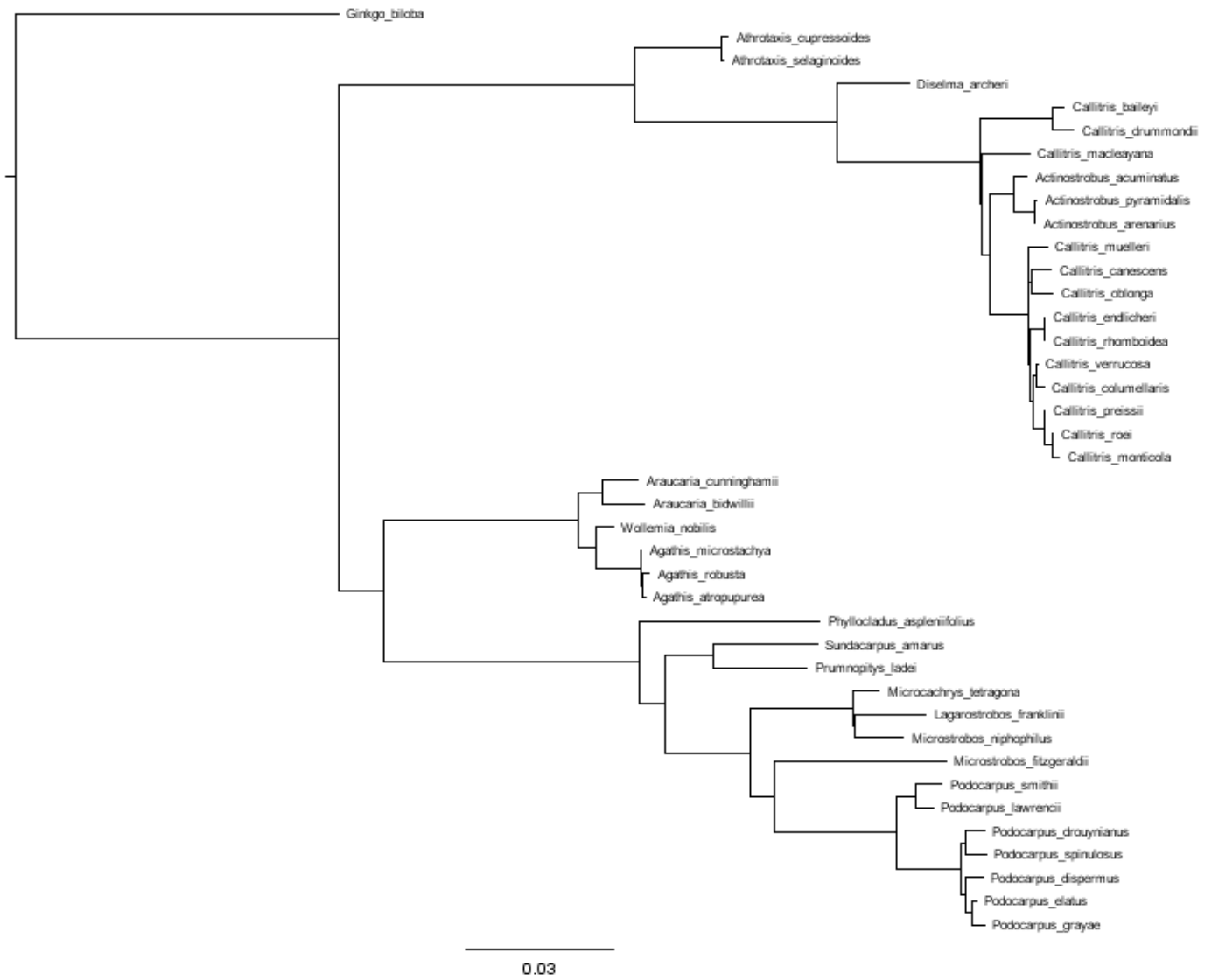
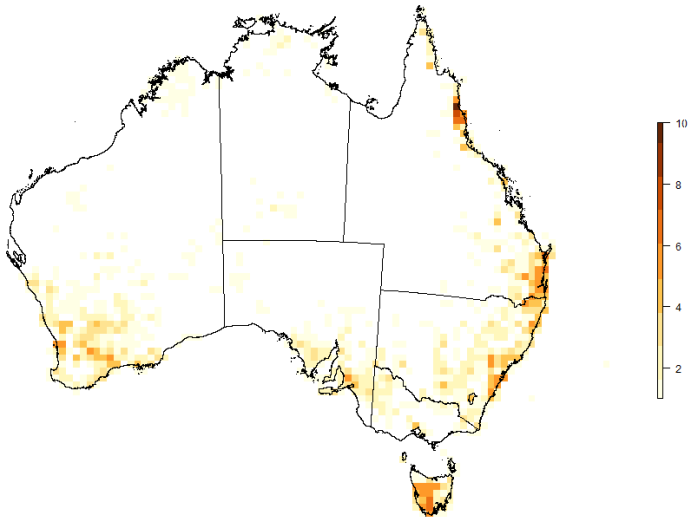


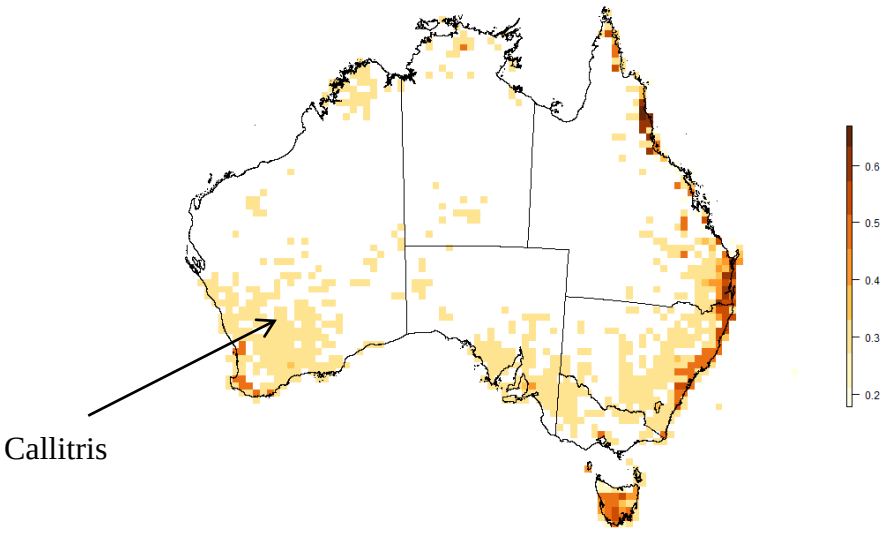
Fig 2. Maximum likelihood phylogeny of endemic Australian conifers. Derived from *matK* and *rbcL* gene regions and calculated using *GARLI v. 0.951*.

***Biodiverse*: Geographic Location and Phylogeny**

I found that species richness was highest in Tasmania and on the Northeast coast of Australia (Fig 3a). PD was more scattered than SR, but also clumped in Tasmania and on the East Coast (Fig 3b). SR was fairly strongly correlated with PD (r^2=0.7544) Tasmania has an especially high PD score and contains species that are distantly-related, *Athrotaxis, Diselma, Lagarostobos, Microstrobos, Phyllocladus* and *Podocarpus*.



3a



3b

Fig 3a Species Richness Species richness of endemic conifers species in Australia. Red regions represent high species diversity.
Fig 3b. PD of conifers Phylogenetic Diversity of conifer species in Australia. The dark red regions, primarily on the East Coast and Tasmania, represent high levels of PD. The genus *Callitris* was widely distributed, especially on the West coast of Australia.

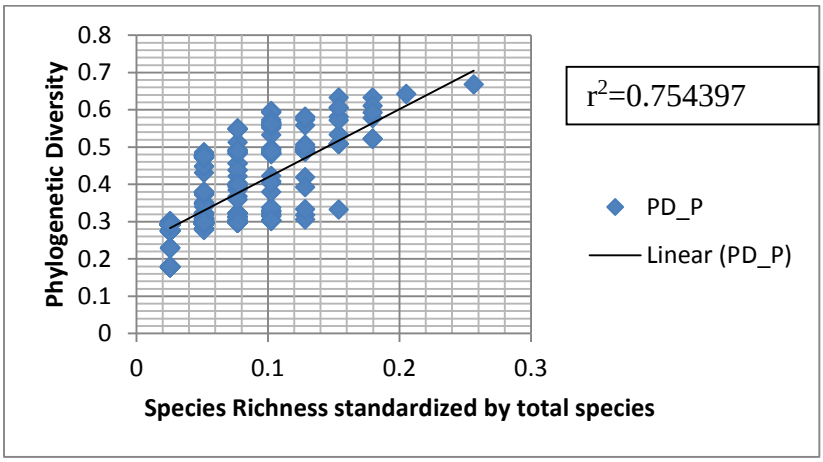
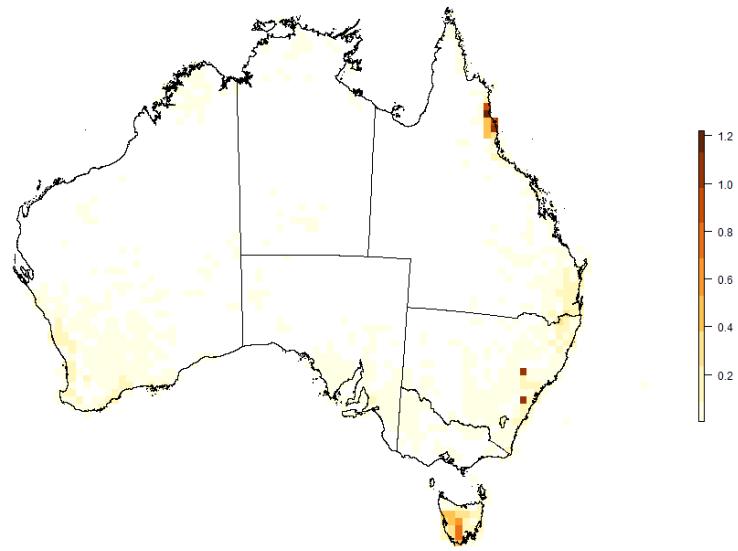
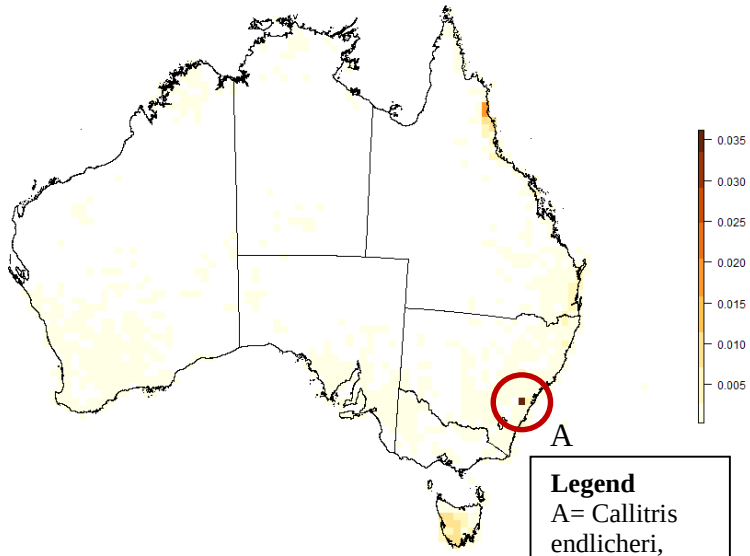


Fig 3c. Scatterplot of species richness (%) against phylogenetic diversity weighted by branch lengths.

WE was concentrated primarily in Tasmania and along the Northeast coast (Fig 4a). PE was not as high in Tasmania, but also was concentrated along the Northeast (Fig 4b). WE was highly correlated with PE, but underestimated some regions of high PE. For example, the grid cell which contained the highest PE value, 0.0361, was underestimated by WE (Fig 4c). This grid cell contained *Callitris*, *Microstrobos*, and *Podocarpus*, who are not sister terminal taxa on the phylogeny.



4a



4b

Legend
A= Callitris endlicheri, Callitris muelleri, Microstrobos fitzgeraldii, Podocarpus spinulosus

Fig.4a Weighted endemism of conifers Weighted endemism of endemic Australian conifers.
Fig 4b Phylogenetic Endemism of Australian conifers. Dark regions represent high PE (PE>0.035). The grid cell labeled A contains the genii *Callitris*, *Microstrobos* and *Podocarpus*. PE is also relatively high on the Northeast coast.

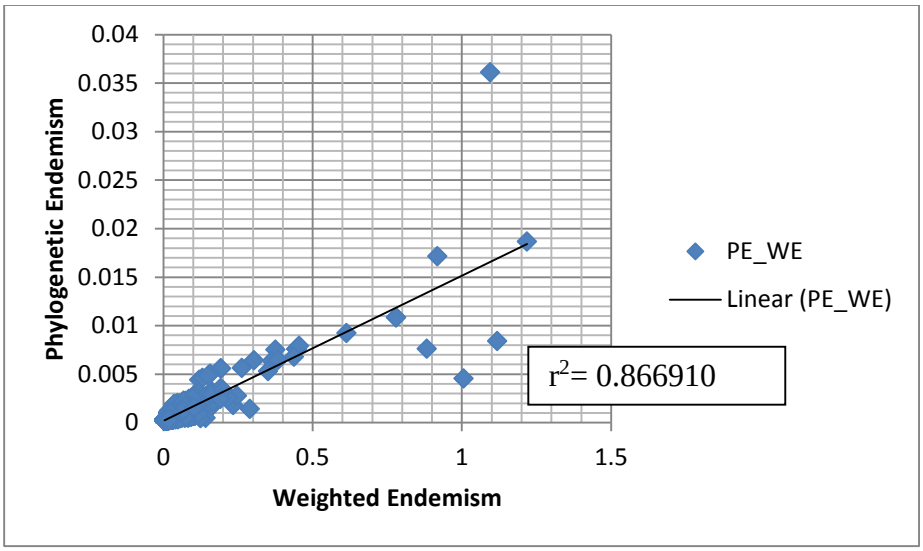


Fig 4c. Scatterplot of weighted endemism against phylogenetic endemism. PE is overall strongly correlated with WE, but this correlation does not hold for some values of high PE or high WE.

Using the standardization and randomization process for PE and PE , I found that significantly low regions of RPD were scattered throughout the country, and significantly high levels of RPD were concentrated along the North and Central East coast (Fig. 5). RPE was statistically significantly high on the central East Coast of Australia and low in the southern coastal regions of the country and in Tasmania (Fig. 6).

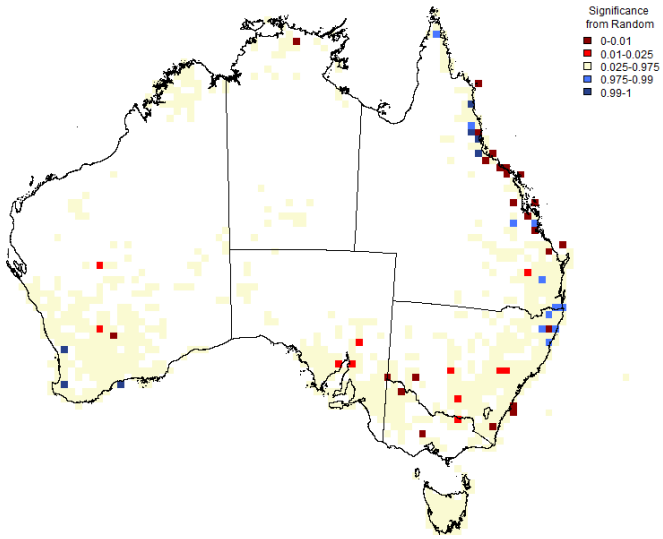


Fig. 5. RPD of conifers. Significantly low RPD (red) is scattered throughout the south and on the Northeast coast. There are fewer regions of significantly high RPD (blue), and they are concentrated primarily on the Northeast coast.

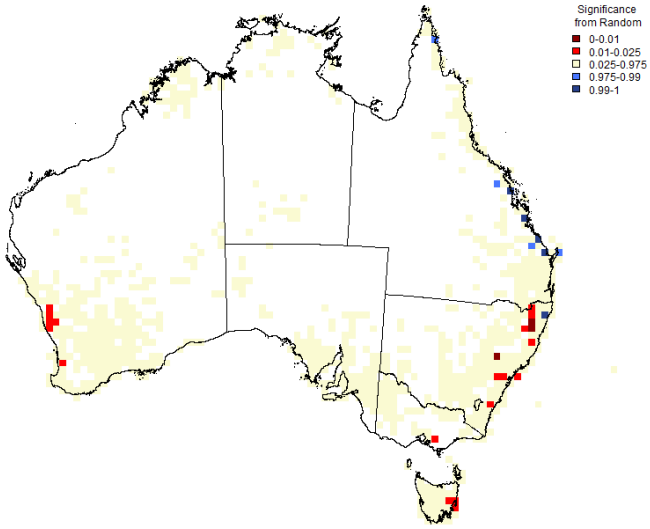


Fig 6 RPE of conifers Significantly high RPE (blue) is concentrated along the Northeast coast of Australia. Significantly low RPE (red) is clustered in the Southeast and Southwest.

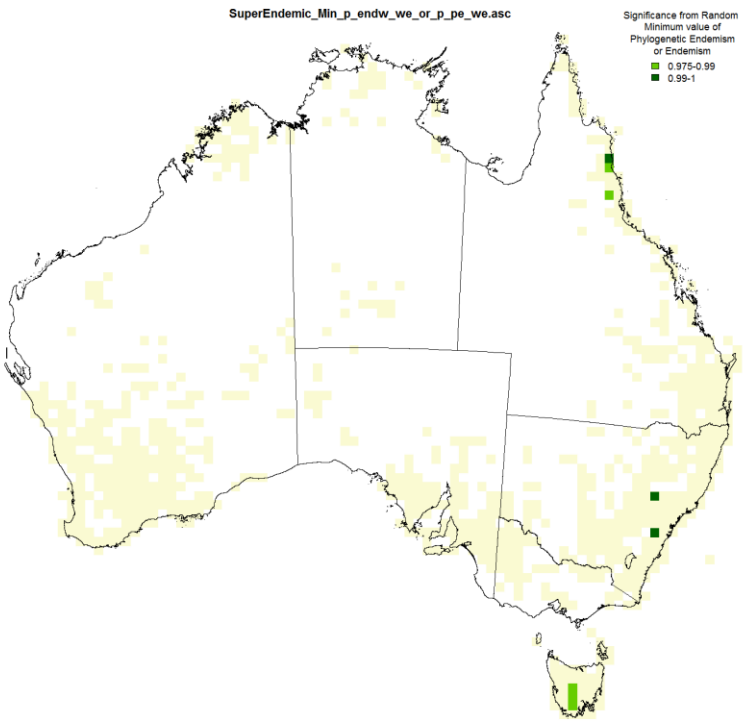
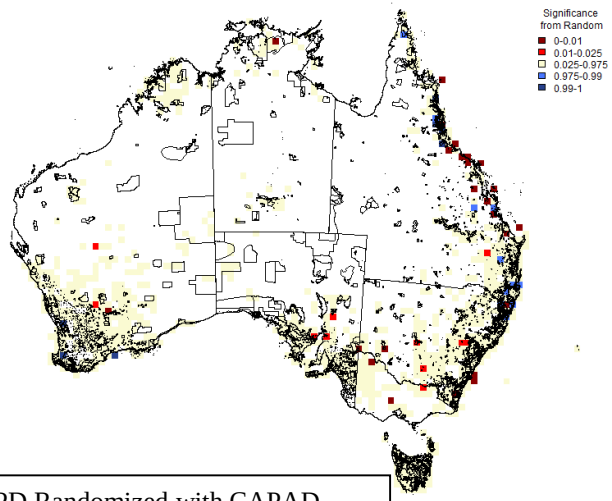


Fig. 7 Sites of Superendmism. Green areas indicate areas heavily concentrated with both high geographic endemism (WE) and high phylogenetic endemism (PE). Darker green areas are more statistically significant.

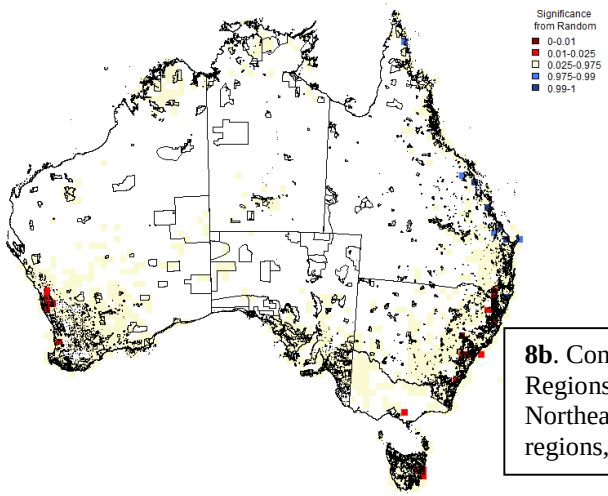
ArcGIS Analysis

After calculating PE, PD RPE and RPE and super-endemism and visualizing with CAPAD 2010 Protected Regions, I found that the majority of regions with significantly high RPD or RPE were protected (Fig 8a and Fig 8b), at least partially. I also found that areas of super-endemism were largely covered by Australian Protected Regions, especially those in Tasmania (Fig 9).

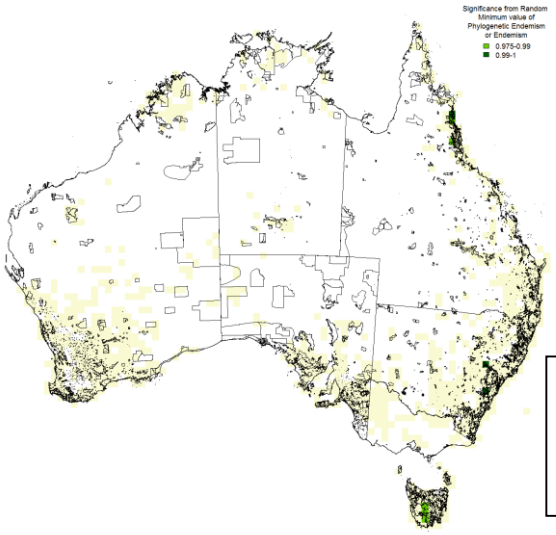
Fig 8a,8b,8c Randomizations of Relative Metrics overlaid with Australian Protected Regions (CAPAD 2010)



8a. Conifer RPD Randomized with CAPAD overlaid. Regions high in RPD (blue) are primarily protected, with a few exceptions along the central Eastcoast of Australia.



8b. Conifer RPE Randomized with CAPAD overlaid. Regions high in RPE (blue) are concentrated in the Northeast coast and are mostly covered by protected regions, except near Brisbane and the Sunshine Coast.



8c. Conifer super-endemic sites with CAPAD overlaid. Darker green regions are more super-endemic. These regions are noticeably covered, with some regions only partially covered in central Tasmania.

DISCUSSION

Accurately and sufficiently quantifying biodiversity is essential for conservation efforts. In this study, I explored biodiversity metrics which quantified the spatial distribution of evolutionary history of Australian endemic conifer species in comparison to tradition metrics which do not take evolutionary history into account. SR and PD were largely correlated, with some exceptions where SR did not predict PD values accurately. WE and PE were also largely correlated, but that correlation broke down for some high values of WE or PE. The spatial and phylogenetic analysis yielded that most regions, high or low with PD and PE, are currently being protected as reserves under Australian law.

Phylogenetic Metric Performance

Regional trends in species richness, endemism vs. PD and PE

As a whole, the continent of Australia had relatively low PD values, which could be due to biogeographic barriers to dispersal and diversification (Faith 2006). SR and PD were largely correlated, which one would expect (Forest et al 2007), given that the more terminal taxa that are sampled from a specific grid cell, the more of the phylogenetic tree is sampled. However, some regions had more or less PD than predicted by their SR (Fig 3c). This correlation was weak for intermediate levels of species richness and PD (Fig 3c). In most cases, SR underpredicted PD, meaning that there were more distantly –related taxa in that grid-cell than expected given SR count. Regions high in PD, which are characterized by many distantly related taxa, were concentrated on the Southeast coast of Australia and throughout Tasmania (Fig 3b). These regions have been found to have a high diversity of conifers in previous studies (Hill and Brodribb 1999, Jordan and Hill 2002). Fossils for 33 species of conifers have been found in north western Tasmania which indicates high conifer diversity relative to the size of the region (Jordan and Hill 2002). Tasmania and South Eastern Australia experienced a decline in conifer diversity after the early Oligocene (Hill and Brodribb 1999) and other evidence suggests that most of the endemic genera, *Athrotaxis*, *Lagarostrobos* and *Microcachrys,* represent the only surviving members of lineages extending back to at least the earliest Cretaceous (Gadek et al. 2000; Biffin

et al. 2011). These genera were also more geographically widespread in the past (Hill and Brodribb 1999). *Athrotaxis*, *Lagarostrobos*, *Michrocachrys*, *Dislema* and *Phyllocladus* are largely restricted geographically to Tasmania. These findings suggest that these clades' ranges may be restricted by an ecological factor that has changed through time. Regions low in PD, which are characterized by many closely-related species, were more prevalent and were concentrated inland of the coast and were primarily comprised of the genus *Callitris*, which is widespread throughout Australia(Fig 3b). Regions low in PD relative to their species richness estimate may be regions of isolated, large radiations (Fritz and Rahbeck 2012).

WE and PE were strongly correlated (r^2= 0.866910); however, WE underestimated the highest values of PE (Fig 4c). WE both overpredicted and underpredicted high PE scores (Fig 4c, due to the fact that closely related taxa may affect the result of PE if they contribute to the range of a clade with taxa in the study area (Roasuer et al 2009). There were few regions high in PE, and they were concentrated Tasmania and on the Northeast coast, potentially due to the aforementioned endemic history of Tasmania.

Interaction between Ecology and Evolution

RPD provides a significance factor for PD; RPD compared the observed PD values for each grid cell under a randomization and found regions of phylogenetic clumping and phylogenetic overdispersion. Phylogenetic clumping, or underdispersion, occurs when a geographic region has species that are more closely related phylogenetically to each other than one would expect by chance. Closely related species which occupy the same spatial region may have similar habitat preferences. If these regions are left unprotected, the chance of extinction of these species significantly increases (Wiens and Graham 2005).

Phylogenetic over-dispersed geographic regions contain species that are significantly more distantly related to each other than one would expect by chance. These regions have higher PD values than expected at random, and thus harbor more evolutionary history than expected by the number of taxa in the grid cell (Sechrest et al. 2002). These regions are especially important for conservation because of their high level of phylogenetic diversity. Regions significantly phylogenetically over-dispersed are found along the East coast (Fig 6). A possible explanation of this phylogenetic over-dispersion is that there is competition occurring in the region (Horner-

Devine and Bohannan 2006). Species that are distantly related to each other and are spatial proximate may have similar habitat requirements.

RPE allows us to distinguish between diversification patterns. Regions with significantly low RPE scores are dominated by rare short-branches. Such regions may represent a place of neoendemism, a region where diversification is happening. Region with significantly high RPE are dominated by rare long-branches, known as paleoendemics, perhaps indicated a refugium. It is critical that we make these distinctions as it informs how we make conservation policy regarding those regions and species. Traditional metrics are unable to distinguish centers of paleoendemism and neoendemism.

I was surprised to find that Tasmania did not have significantly high levels of RPD or RPE, and actually had significantly low levels of RPE. To investigate this further, I ran another standardization and randomization to identify areas of super-endemism (Mishler et al. in review), Tasmania, especially central Tasmania, contained many regions of super-endemism, meaning that it contained high levels of geographic weighted endemism *and* high phylogenetic endemism. A possible explanation for this is that there are both rare, short phylogenetic branches and rare, long phylogenetic branches which occupy these grid cells.

Protected Regions: Spatial Analysis

I found that regions of high RPD, high RPE and superendemic sites were largely within the bounds of Australia's protected regions (Fig 8a,b,c). There were a few regions which are not currently within the bounds of the Australian Protected Regions (Fig 8a,b,c), which should be further studied as they are critical for conifer diversity conservation.

Limitations and Future Directions

A key limitation of my study is the spatial scale at which I performed analyses. Ecologcial and evolutionary patterns may differ at different spatial scales. Thus, it is important to re-analyze the data at different spatial scales, for instance 100,000m^2 grids or 25,000m^2 grids to check for consistency among the spatial scales. For this study, we chose 50,000m^2 grids because they have been shown to display subtleties of the data, and roughly estimate community sizes

(Mishler et al. in review, Nagalingum et al. in prep). Another spatial limitation stems from my use of the CAPAD 2010 shapefile in its entirety. This shapefile included all parklands, not only major reserves or conifer-specific reserves, and the number of vectors in this data layer made it difficult to interpret how effectively regions of high PD and PE are being conserved. Additionally, I was unable to answer one of my original research questions, which was to identify and map biogeographic regions that could be potential environmental explanations of PD/PE trends. I plan to continue this analysis and overlay these factors in the future.

Phylogenetically, my study is limited in its robustness, because I focused on a subset of species inhabiting the continentand this is a monophyletic group in relation to Gingko biloba, but polyphylys may be nested in these lineages. The phylogeny used for this study probably contains an error, a *matK* sequence for *Microstrobos niphophilus* which needs to be re-sequenced. Due to time constraints, I was unable to re-sequence it in time for this paper. It, however, does not affect the calculation of PD and PE as all of the branch lengths are incorporated that join sister taxa which share a spatial grid cell (Roasuer et al 2009)

After calculating the metrics RPD and RPE, which are readily comparable between groups, I plan to compare my results to other studies currently underway (Mishler et al in review), (Nagalingum et al in prep). These comparisons will yield innovative and important results for conservation policy. The ultimate goal of these projects is to conduct PD and PE metrics on all of the flora and fauna of Australia to inform critical policy decisions.

Broader Implications and Conclusions

Examining the intersection of evolutionary history and spatial distribution of conifer species is a key method for properly informing conservation policy. Historically, approaches to biodiversity conservation have attempted to apply different concepts. Some have been more concerned with conserving rare species, while others have focused on key habitats. PD and PE are metrics which provide a way to account for both geography and evolutionary rarity. They are not in disagreement with SR and WE, instead they incorporate these metrics and provide more insight into the evolutionary and ecological processes that have occurred throughout time.

ACKNOWLEDGEMENTS

This project would not have been possible without the enthusiasm and support of Team ES196—especially Patina Mendez for her dedication, thorough editing, constant positivity and life advice. I cannot thank her enough for all of her support. I thank my loving family, especially my mother and father for their dedication throughout my life and especially through this project. Dr. Nathalie Nagalingum, Dr. Brent Mishler, Sonia Nosratinia and Nunzio Knerr were all integral to this project. They deserve so much more than a "thank you"; I could not have undertaken this thesis project without their continual dedication, time and assistance. Nathalie has become an incredible mentor and friend; I truly appreciate her patience and willingness to take me on as a student over the past year. Brent has spent countless hour helping me to unpack the subtleties of phylogenetics, the methods and the principles. His mentorship throughout this process has been invaluable. Sonia's help was critical to my understanding and practice of the methods. I cannot thank her enough for having spent so much time working with me. Nunzio is our resident data-processor, R- Biodiverse,-ArcMap-extraordinaire and I could not have processed this data without his help. I am so grateful to have gotten the chance to work with such a talented, intelligent group of people and I look forward to collaborating with them in the future.

REFERENCES

Chao, A. 2005. Species richness estimation. Encyclopedia of Statistical Sciences. 7909-7916.

Crisp, M.D., S. Laffan, H. P. Linder, and A. Monro. 2001. Endemism in the Australian flora. Journal of Biogeography 29: 183-198.

Corey S. J., T. A. Waite. 2008. Phylogenetic autocorrelation of extinction threat in globally imperiled amphibians. Diversity and Distributions 14:614-629.

Crozier R. H. 1997. Preserving the Information Content of Species: Genetic Diversity, Phylogeny, and Conservation Worth. Annual Review of Ecology and Systematics 28:243-268.

Davies, J. T. and L. B. Buckley. 2011. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present- day richness gradients for mammals. The Royal Society 366: 2414-2425.

Faeth, S.H., C. Bang, and S. Saari. 2011. Urban biodiversity: patterns and mechamisms. Annals of the New York Academy of Sciences 1223: 69-81.

Faith, D. P., and A. M. Baker. 2006. Phylogenetic diversity (PD) and biology conservation: some bioinformatics challenges. Evolutionary Bioinformatics 2: 70-77.

Forest, F. R. Greyner, M. Rouget, J. Davies, R.M. Cowling, D.P. Faith, A. Blamford, J.C. Manning, S. Proches, M. van der Bank, G. Reeves, T.A. J. Hedderson, and V. Savolanien. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445: 757-759

Fritz S. A., C. Rahbek. 2012. Global patterns of amphibian phylogenetic diversity. Journal of Biogeography 39:1373-1382.

Jordan, . G.J. and R.S. Hill (2002) Cenozoic plant macrofossil sites of Tasmania. Papers and Proceedings of the Royal Society of Tasmania, 136 .127-139

Leslie, A.B., J. M. Beaulieub, S.R. Hardeep, P.R. Cranea, M.J. Donoghueb, and S. Mathews. 2012. Hemisphere- scale differences in conifer evolutionary dynamics. Proceedings of the National Academy of Sciences of the United States of America. 109: 16217-16221.

Lombard A. T., R. M. Cowling, R. L. Pressey, and A. G. Rebelo. 2003. Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region. Biological Conservation 112:45-62.

Mishler, B. D., N. Knerr, C. E. González-Orozco, A.H. Thornhill, S.W. Laffan, and J.T. Miller. (in review) Phylogenetic approaches to biodiversity, endemism, and conservation. Nature Communications

Mishler B. D. 2009. Three Centuries of Paradigm Changes in Biological Classification: Is the End in Sight? Taxon 58:61-67.

Nixon K. C., Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. Cladistics 6:211-223.

Ingelby, S. 2009. Endemism in Australian mammals. Australian Museum. http://australianmuseum.net.au/Endemism-in-Australian-mammals (Version 5/7/2012).

Kalendar R, D. Lee, and A. H. Schulman. 2011. Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. Genomics, 98: 137-144.

Mace, G. M., J.L. Gittleman, and A. Purvis. 2003. Preserving the tree of life. Science. 300: 1707-1709.

Meiri, S. and G. Mace. 2009. New taxonomy and the origin of species. Public Library of Science Biology 5: 1385- 1387.

Mooers, A. O. 2007. The diversity of biodiversity. Nature 445: 717-718.

Pimm, S. L.; G.J. Russell.; J.L. Gittleman,., and T.M. Brooks. 1995. The future of biodiversity. Science 269: 347–350.

Quinn, C.J., R.A. Price, and P.A. Gadek. 2002. Familial concepts and relationships in the conifer based on rbcL and matK sequence comparisons. Kew Bulletin 57: 513- 531.

Reaka-Kudla, M.L. 1997. Biodiversity II: understanding and protecting our biological resources. National Academy of Sciences, USA.

Regional Population Growth, Australia, 2010-11. Australian Bureau of Statistics.

Rosauer, D.F. and S.W. Laffan. 2008. Linking phylogenetic trees, taxonomy and geography to map phylogeography using Biodiverse. Taxonomic Data Working Group 2008. Perth, Australia.

Rosauer, D., S. W. Laffan, M.D. Crisp., S.C. Donnellans, and L.G. Cook. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Molecular Ecology 18: 4061-4072.

Sechrest W., T. M. Brooks, G. A. B. da Fonseca, W. R. Konstant, R. A. Mittermeier, A. Purvis, A. B. Rylands, and J. L. Gittleman. 2002. Hotspots and the conservation of evolutionary history. Proceedings of the National Academy of Sciences 99:2067-2071.

Williams, P. H., C. J. Humphries, P.L. Forey, and R.I. Vane-Wright.1994. Biodiversity, taxonomic relatedness, and endemism in conservation. Systematics and Conservation Evaluation. Oxford University Press, Oxford, UK.

APPENDIX A: LIST OF INDIGENOUS CONIFER SPECIES

List 1: Indigenous conifer species list, including the outgroup used for this study, *Ginkgo biloba*

Actinostrobus arenarius
Callitris baileyi
Callitris columellaris
Callitris monticola
Callitris oblonga
Callitris roei
Microstrobos niphophilus
Agathis atropurpurea
Agathis microstachya
Agathis robusta
Araucaria bidwillii
Araucaria cunninghamii
Microcachrys tetragona
Actinostrobus acuminatus
Actinostrobus pyramidalis
Athrotaxis cupressoides
Athrotaxis selaginoides
Callitris canescens
Callitris drummondii
Callitris endlicheri
Callitris macleayana
Callitris muelleri
Callitris preissii
Callitris rhomboidea
Callitris verrucosa
Diselma archeri
Lagarostrobos franklinii
Microstrobos fitzgeraldii
Phyllocladus aspleniifolius
Podocarpus dispermus
Podocarpus drouynianus
Podocarpus elatus
Podocarpus grayae
Podocarpus lawrencei
Podocarpus smithii
Podocarpus spinulosus
Prumnopitys ladei
Sundacarpus amarus
Wollemia nobilis
Ginkgo biloba

APPENDIX B: GENBANK ACCESSION NUMBERS

Table 1: Genbank Accession numbers for matK and rbcL gene regions

rbcL	matK
JF725937 Actinostrobus arenarius	JF725837 Actinostrobus arenarius
EU161450 Actinostrobus pyramidalis	JF725831 Actinostrobus pyramidalis
AF502087 Agathis atropurpurea	EU025977 Agathis atropurpurea
AF508920 Agathis microstachya	EU025978 Agathis microstachya
EF490509 Agathis robusta	AF456371 Agathis robusta
AM920227 Araucaria bidwillii	EU025974 Araucaria bidwillii
EF490510 Araucaria cunninghamii	EU025975 Araucaria cunninghamii
JF725921 Athrotaxis cupressoides	JF725821 Athrotaxis cupressoides
JF725938 Athrotaxis selaginoides	JF725838 Athrotaxis selaginoides
JF725945 Callitris canescens	JF725845 Callitris canescens
JF725939 Callitris drummondii	JF725839 Callitris drummondii
JF725932 Callitris endlicheri	AY988331 Callitris endlicheri
JF725933 Callitris macleayana	JF725833 Callitris macleayana
JF725924 Callitris muelleri	JF725824 Callitris muelleri
JF725940 Callitris preissii	JF725840 Callitris preissii
L12537 Callitris rhomboidea	JF725825 Callitris rhomboidea
JF725942 Callitris verrucosa	JF725842 Callitris verrucosa
JF725926 Diselma archeri	JF725826 Diselma archeri
HM593609 Lagarostrobos franklinii	EU161486 Lagarostrobos franklinii
HM593611 Microcachrys tetragona	EU161483 Microcachrys tetragona
AF249646 Microstrobos fitzgeraldii	EU161484 Microstrobos fitzgeraldii
AF249647 Microstrobos niphophilus	
AF249651 Phyllocladus aspleniifolius	AY442147 Phyllocladus aspleniifolius
JF969685 Podocarpus dispermus	HM593741 Podocarpus dispermus
HM593639 Podocarpus drouynianus	HM593742 Podocarpus drouynianus
HM593641 Podocarpus elatus	HM593745 Podocarpus elatus
AF249608 Podocarpus grayae	HM593750 Podocarpus grayae
HM593651 Podocarpus lawrencii	HM593755 Podocarpus lawrencii
HM593675 Podocarpus smithii	HM593779 Podocarpus smithii
AF249630 Podocarpus spinulosus	HM593780 Podocarpus spinulosus
HM593620 Prumnopitys ladei	HM593723 Prumnopitys ladei
AF249663 Sundacarpus amarus	HM593788 Sundacarpus amarus
EF490508 Wollemia nobilis	AF456377 Wollemia nobilis