

Accepted Manuscript

Spatial Phylogenetics of the Vascular Flora of Chile

Rosa A. Scherson, Andrew H. Thornhill, Rafael Urbina-Casanova, William A. Freyman, Patricio A. Pliscoff, Brent D. Mishler

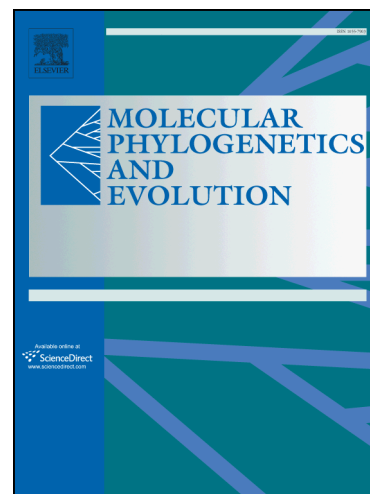
PII: S1055-7903(17)30137-9
DOI: <http://dx.doi.org/10.1016/j.ympev.2017.04.021>
Reference: YMPEV 5807

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 13 February 2017
Revised Date: 12 April 2017
Accepted Date: 23 April 2017

Please cite this article as: Scherson, R.A., Thornhill, A.H., Urbina-Casanova, R., Freyman, W.A., Pliscoff, P.A., Mishler, B.D., Spatial Phylogenetics of the Vascular Flora of Chile, *Molecular Phylogenetics and Evolution* (2017), doi: <http://dx.doi.org/10.1016/j.ympev.2017.04.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Spatial Phylogenetics of the Vascular Flora of Chile

Rosa A. Scherson^{1*}, Andrew H. Thornhill², Rafael Urbina-Casanova¹, William A. Freyman², Patricio A. Pliscoff³, Brent D. Mishler²

¹Laboratorio de Sistemática y Evolución de Plantas, Departamento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile. Santiago 9206, Chile.

²University and Jepson Herbaria, and Department of Integrative Biology, University of California, Berkeley, CA, 94720-2465, USA.

³Instituto de Geografía y Departamento de Ecología, Pontificia Universidad Católica de Chile, Av. Libertador Bernardo O'Higgins 340, Santiago 8331150, Chile.

*Corresponding author: Rosa A. Scherson

Laboratorio de Sistemática y Evolución de Plantas, Departamento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile. Santiago 9206, Chile

rositascherson@uchile.cl

Abstract

Current geographic patterns of biodiversity are a consequence of the evolutionary history of the lineages that comprise them. This study was aimed at exploring how evolutionary features of the vascular flora of Chile are distributed across the landscape. Using a phylogeny at the genus level for 87% of the Chilean vascular flora, and a geographic database of sample localities, we calculated phylogenetic diversity (PD), phylogenetic endemism (PE), relative PD (RPD), and relative PE (RPE). Categorical Analyses of Neo- and Paleo- Endemism (CANAPE) were also performed, using a spatial randomization to assess statistical significance. A cluster analysis using range-weighted phylogenetic turnover was used to compare among grid cells, and with known Chilean biotopes. PD patterns were concordant with known centers of high taxon richness and the Chilean biodiversity hotspot. In addition, several other interesting areas of concentration of evolutionary history were revealed as potential conservation targets. The south of the country shows areas of significantly high RPD and a concentration of paleo-endemism, and the north shows areas of significantly low PD and RPD, and a concentration of neo-endemism. Range-weighted phylogenetic turnover shows high congruence with the main biotopes of Chile. Even though the study was done at the genus level, the outcome provides an accurate outline of phylogenetic patterns that can be filled in as more fine-scaled information becomes available.

Keywords Chilean vascular flora; conservation; paleo and neo endemism; phylogenetic diversity; phylogenetic endemism; spatial phylogenetics.

1. Introduction

The human population is expanding at unprecedented rates. Impacts of this increase include dramatic changes in land use, pollution, invasive species, and climate change related phenomena (Millennium Ecosystem Assessment, 2005). Recently, these plus other human-induced phenomena were globally assessed and a set of nine planetary boundaries was defined as a "safe operating space" for humanity, providing quantitative thresholds beyond which human welfare and the resilience of ecosystems would be affected (Rockström et al., 2009a, 2009b; Steffen et al., 2015). Three of these boundaries have already been exceeded beyond the safe operating space, with biodiversity loss the most seriously affected (Rockström et al., 2009a).

Despite the unquestioned importance of biodiversity and the urgent need for conserving it, the assessment of which portion of it requires the highest priority attention has proved extremely complex, especially when resources for conservation are limited (Wilson et al., 2006). In that sense, the concept of "hotspots" has been useful for identifying places that need the most attention (Myers et al., 2000; Sechrest et al., 2002). However, recent studies have shown that hotspots of highest species richness are not always concordant with hotspots of greatest endemism or threat (Daru et al., 2014; Marchese, 2015).

Globally, with approximately 2 million described species and many more yet to be described (Stork, 1993; Mora et al., 2011), we lack much information regarding natural history, ecologically important traits, physiology, and taxonomy in many important groups (Cardoso et al. 2011; Whittaker et al. 2005). However, common ancestry makes it likely that more closely related species tend to be similar in their ecologies and life histories, suggesting that phylogeny-based measures of biodiversity can be used as a surrogate for feature diversity (Faith, 1992a; Forest et al., 2007; Collen et al., 2011; Thuiller et al., 2011).

Phylogenetic diversity (PD), the most commonly used metric based on evolutionary history, was initially defined by Faith (1992b) as the sum of branch lengths in a phylogeny connecting a set of taxa, and provides a measure of how much

evolutionary history (and subsequent feature diversity) would be lost if the area that contains those taxa was not conserved. PD is relatively robust to changes in taxonomic status and, importantly, allows direct comparisons among areas and different taxonomic groups.

In the last decade, phylogeny-based methods to measure biodiversity have been rapidly developing. One important area of development relates to endemism. Endemicity has always been a major consideration in historical biogeography and one of the most important criteria when defining conservation priorities (Linder, 2001). The concept of endemicity is implicitly and tightly correlated with Margules and Pressey's (2000) idea of irreplaceability, and consequently with how vulnerable a place is. However, the traditional definition of endemism has been narrowly construed to mean complete restriction of a taxon to a certain area (i.e., *absolute endemism*). An important advance was the addition of a broader concept of *relative endemism*, which refers to the *degree* of restriction of the range of a taxon along a quantitative scale from 1 (absolute endemism) to 0 (ubiquitous) (Crisp et al., 2001). In order to incorporate evolutionary information, the concept of relative endemism has recently been extended to the geographic restriction of clades at any level, a metric called *phylogenetic endemism* (PE; Rosauer et al. 2009; Mishler et al. 2014).

Further developments in PD and PE calculation have incorporated two derived metrics that compare PD (or PE) on the original phylogeny with that measured on a comparison phylogeny wherein all branches are equal (average) length: *relative phylogenetic diversity* (RPD) and *relative phylogenetic endemism* (RPE), respectively (Mishler et al. 2014). A spatial randomization approach has also been designed to allow the assessment of statistical significance of these metrics (Mishler et al. 2014).

Categorical Analysis of Neo- And Paleo-Endemism (CANAPE) is capable of distinguishing areas containing significant concentrations of rare long branches (paleo-endemism), rare short branches (neo-endemism), or mixed endemism (Mishler et al., 2014; Schmidt-Lebuhn et al., 2015; González-Orozco et al., 2016; Thornhill et al., 2016).

Phylogenetic diversity and endemism are most often used to characterize alpha diversity (localized PD or PE). Yet it is also important to look at beta-diversity, turnover in PD *among* local areas or "phylobetadiversity" (Graham & Fine, 2008), analogous to traditional measures of species turnover except that shared and unshared branches of the phylogeny are measured instead of shared and unshared species. Turnover can also be measured in PE, that is by weighting range-restricted branches more heavily than common branches, and can be called "phylogenetic range-weighted turnover" (Laffan et al., 2016). These measures allow better understanding of biotic regions and their boundaries, and can also be used to help determine complementarity as part of a conservation assessment (Thornhill et al. 2016).

Chile is a country of high endemism, and is thus a good study system for application of these new phylogenetic methods. It has been referred to as a "biogeographic island", due to the isolation caused by a succession of geological events. The breakup of Gondwana, the separation from Antarctica that generated the Humboldt current, the onset of aridity in the north, and the uplift of the Andes Mountains generated a space surrounded by natural barriers: the driest desert in the world to the north, frozen Antarctica to the south, the Andes Mountains to the east, and the Pacific Ocean to the west (Villagran & Hinojosa, 1997; Villagrán & Armesto, 2005; Moreira-Muñoz, 2011). The isolation of Chile has resulted in the largest number of endemic plant genera of any country in South America, and four endemic families of angiosperms (Moreira-Muñoz 2011; Urbina-Casanova et al. 2015).

Previous studies have shown that species richness concentrates in the central area of Chile, coincident with the location of the "Chilean winter rainfall-Valdivian forest" biodiversity hotspot ecoregion that has high priority for conservation (Arroyo et al. 2008; Arroyo et al. 2004; Myers et al. 2000). This is worrying given that Chile is a very centralized country; 80% of the population is concentrated in this hotspot, as well as the main centers for agricultural, industrial, and services activities (CAPP, 2008). Despite this, only 5.7% of terrestrial ecosystems are well represented within the officially protected areas in central Chile (Pliscoff & Fuentes-Castillo 2011). However, as discussed

above, hotspots should not necessarily be defined only for areas that harbor large numbers of species, but can also be areas that have unusual concentrations of evolutionary history (PD or PE; Marchese 2015; Thornhill et al. 2016).

The identification of areas of concentrated endemism is very important for several reasons. Interpreting the compositional and geographic patterning of those areas helps develop a better understanding of the biogeographic origin of the flora. In addition, different categories of endemism have different intrinsic values for conservation. Areas of neo-endemism imply localized processes of recent diversification, which could be ongoing. These are areas that are evidently producing new diversity and therefore possess a high value for conservation (Kraft et al., 2010). Areas of paleo-endemism imply concentrations of relictual taxa of high historic value that at the same time have high genetic and functional diversity, providing sources of resilience when facing uncertain future scenarios (Faith et al., 2010), and also possesses a high value for conservation.

Using only the endemic Chilean of vascular plant genera, Scherson et al. (2014) showed that the south of the country (below 35° S harbored many areas in which PD was higher than expected by chance, and conversely that the north of the country showed the opposite pattern. While these patterns were clearly defined, the present study was undertaken to see whether they hold when the entire flora is considered, and to add consideration of patterns of PE. Our general goal was to understand the distribution patterns of the full flora of Chile from an evolutionary perspective, focusing especially on patterns of endemism. Specific goals included: (1) developing a phylogenetic framework to discover patterns of PD and PE on the landscape, (2) studying biotic transitions among floristic regions using phylogenetic measures of beta-diversity, and (3) visualizing conservation priorities in the light of these measures.

2. Methods

2.1. DNA sequence sampling

We sampled 756 genera, representing 87% of all accepted Chilean native vascular genera. The list of genera was obtained from the Catalogue of the Vascular Plants of the Southern Cone of South America (Zuloaga et al., 2008), which is being constantly updated in its online version at the Darwinion Institute in Argentina (<http://www2.darwin.edu.ar/Proyectos/FloraArgentina/>; accessed July 2, 2015). Each genus was represented as one operational taxonomic unit (OTU) in the phylogeny, thus assuming all genera were monophyletic. For each native Chilean genus, DNA sequences were preferentially sampled from native Chilean species, however sequences from species native to other parts of the Southern Cone of South America were used to fill in missing data. Most OTUs, therefore, are chimeric with sequence data combined from multiple species from Chile and elsewhere in the Southern Cone.

2.2. Assembly of molecular data set and phylogenetic analyses

We used a concatenation of five chloroplast loci (*atpB*, *trnL-F*, *matK*, *ndhF*, and *rbcl*) and the nuclear loci 18S and ITS (see Appendix A in Supplementary materials for accession details). The majority of sequences were mined from GenBank and then aligned into matrices using Matrix Maker (Freyman and Thornhill, 2016 <https://github.com/wf8/matrixmaker>) a Python utility utilizing Biopython (Cock et al., 2009). For 17 genera, 56 new sequences were obtained in the lab. Sequences were aligned using MAFFT (Katoh & Standley, 2013) with the --adjustdirection flag to ensure correct sequence polarity.

The final concatenated molecular alignment contained 19,242 aligned nucleotides. A maximum likelihood analysis was performed in the CIPRES Portal (www.phylo.org) using the RAXML HPC2 on XSEDE analysis with a partitioned model. FigTree (Rambaut, 2009 <http://tree.bio.ed.ac.uk/software/figtree/>) was used to view the resulting bipartition tree and generate a nexus formatted file suitable for PD analyses. The topology was compared with angiosperm orders according to APG III 2009 (Stevens

2001 onwards) to check that each genus displayed correct ordinal membership.

2.3. Spatial data

Locality information was compiled from Chilean herbaria (CONC, SGO) and our own fieldwork. In addition, plant records were downloaded from the Global Biodiversity Information Facility (www.gbif.org) by filtering for the kingdom Plantae and the country Chile. Distributional maps of each genus were screened and corrections were made to records that fell outside of the Chilean territory using Google Refine v.2.5 (<https://code.google.com/p/google-refine/downloads/>); when a record could not be corrected it was excluded from the study. The coordinate information of the final dataset (141,223 records) was reprojected into an Albers Equal Area Coordinate system (EPSG: 5362).

2.4. Biodiversity analyses and associated statistical tests

Biodiverse v.1 (Laffan et al., 2010) was used to calculate taxon richness (TR), Phylogenetic Diversity (PD), Phylogenetic Endemism (PE), Relative Phylogenetic Diversity (RPD), and Relative Phylogenetic Endemism (RPE), as well as to conduct a spatial randomization using its "rand_structured" model. The choice of grid cell size was based on redundancy values. Redundancy is calculated as $1 - [\text{richness}/(\text{number of samples})]$. A value close to one represents a good overall sampling of taxa, while a value of zero means that there is only one sample per taxon, therefore poor sampling. We found that a 75 km x75 km grid cell was the optimal size for our dataset, since we had acceptable redundancy and it was still a fine enough scale to recognize spatial patterns. Observed PD and PE, as well as randomized PD, RPD, and CANAPE were calculated on the dataset using the methods and reasoning described in Mishler et al. (2014) and Thornhill et al. (2016).

2.5. Phylogenetic turnover measures to identify biotic transitions

Cluster analyses in Biodiverse were used to compare all grid cells with each other in terms of shared branches of the phylogeny they contain, using Unweighted Pair Group Method with Arithmetic Mean Averaging (UPGMA). Two phylogenetic turnover metrics were used: PhyloSørensen calculated phylogenetic difference using the branch lengths from the original phylogenetic tree, while phylogenetic range-weighted turnover (Laffan et al. 2016) calculated phylogenetic difference using the branch lengths from the range-weighted tree, which gives greater precedence to narrow-ranging taxa.

2.6. Phylogenetic turnover-Climatic zones comparisons

The results of the phylogenetic turnover analyses were compared with the main climatic patterns of continental Chile (Luebert & Pliscoff 2006), using the Kappa statistic (Cohen, 1960) as an indicator of concordance between the climatic and PD spatial patterns. Kappa was applied as a measure of agreement between two categorical outputs (main climatic zones and clusters) using Map Comparison Kit 3.2.3 (Visser & de Nus, 2006). Values range between 1 (complete agreement) and 0 (complete disagreement). To allow spatial comparisons, the climatic zones of Chile were converted to raster format at the same spatial resolution of the cluster analysis (75km). Deep clusters in the dendrogram containing a small number of grid cells were not included in this analysis.

3. Results

3.1. Observed biodiversity indices

Observed TR, WE, PD, and PE are shown in Fig. 1. As expected, the distribution of PD and taxon richness were very similar. Richness and endemism concentrated in the center

of Chile and decreased with latitude, to the north and to the south. One cluster of grid cells had high values of TR, WE, PD, and PE in the north of the country. This is probably an artifact of sampling, due to a northern-focused collection effort that produced a denser geographic database for that area (see discussion for details).

3.2. Statistically significant areas for PD and RPD

Significant areas of PD and RPD (Fig. 2) showed different latitudinal patterns. For PD, a few areas in the south of the country showed significantly high PD (phylogenetic over-dispersion), whereas numerous grid cells, especially in the north, showed significantly low PD (phylogenetic clustering). There was a striking north-south distinction for RPD, with the south having a concentration of branches that were significantly longer than expected by chance and the north having a concentration of branches that were significantly shorter.

3.3. Areas of significant PE

Significant areas of phylogenetic endemism are shown in Fig. 3; these areas appeared mostly in the north and the south of the country, with only a few significant grid cells in central Chile. The CANAPE analysis for all vascular plants showed a pattern in which areas dominated by paleo-endemism were present in the temperate forest and Austral areas of the country, and areas dominated by neo-endemism were abundant in the north along with areas of mixed endemism (Fig. 3a). The analyses with seed plants alone (Fig. 3b) and with angiosperms alone (Fig. 3c) showed no concentrations of paleo-endemism in the south-center of the country but a few in the Austral area, and both showed concentrations of neo-endemism in the north of the country.

3.4. Phylogenetic turnover as a measure of biotic transitions

The phylogenetic range-weighted turnover analysis (Fig. 4c) discriminated areas of phylogenetic similarity on the landscape that coincided more closely with the main climatic zones of Chile (Luebert & Pliscoff, 2006) than did the PhyloSørensen analysis (see Supporting Information). Further, the similarity relationships depicted by the two turnover analyses differ in that the Mediterranean bioclimate forms a larger cluster with the northern Tropical bioclimate in the RW analysis, whereas the PhyloSørensen analysis shows a larger cluster between the Mediterranean and the southern temperate bioclimate. This must be due to a difference in how wide-ranging versus range-restricted genera are distributed on the landscape and could have many explanations such as differential climatic tolerance or soil preference between the two categories.

The phylogenetic range-weighted turnover analysis showed four major clusters, with good geographic structure: (1) northern area, coincident with the Tropical bioclimate; (2) central area, coincident with the Mediterranean bioclimate; (3) southern area, coincident with the temperate bioclimate; and (4) areas along the Andes and into Patagonia. The results of the map comparisons showed high agreement especially between Tropical and Mediterranean climatic zones and clusters 1 and 2 respectively (Kappa index of agreement 0.92 and 0.83 respectively). The Temperate climatic zone and cluster 3 had a lower Kappa index of agreement of 0.68. Maps are shown in Fig. 4 (a and b).

4. Discussion

4.1. Spatial data and taxon sampling

Despite using many sources of information, we were only able to gather a total of 141,223 occurrence records for the native vascular plants of Chile. The paucity of

georeferenced data -“the Wallacean shortfall” - is a common problem (Lomolino, 2004; Cardoso et al., 2011), especially prevalent in developing countries (Dunn et al., 1997; Mennecke & West, 2001; Hortal et al., 2007); many herbaria in South America lack geographic coordinates for their collections, and most do not have electronic databases available for download. This poor sampling can be an issue for analyses of spatial patterns, especially when deciding on a grid size to work with. For our dataset, a 75 x 75 km grid cell was a good compromise between redundancy and resolution of spatial patterns.

Another problem encountered is the lack of homogeneity of the spatial dataset. For example, there was a small group of grid cells in the north of Chile in which an extensive government-funded effort was carried out to map biodiversity in detail (Hernández et al., 2013), providing a much denser sampling of localities. Those grid cells appeared to have exceptionally high TR and PD (see Fig. 1), which is caused by this sampling bias. However, the randomization seems to be relatively robust to this sampling heterogeneity; the patterns of significance detected in the well-sampled grid cells were the same as seen in the poorly sampled grid cells surrounding them. This is important, since heterogeneity in georeferenced data density for biodiversity studies is a common problem (Hortal et al., 2007). This study suggests that global patterns can still be studied if proper statistical methods are used.

Collection bias is another potential problem, even for densely sampled areas. In this sense, the use of species distributions models could be a way of increasing redundancy with fewer occurrence records. PD and related indices have been used with estimated occurrences using models of species distributions (Pio et al., 2011, 2014; Thuiller et al., 2011), however RPD, RPE, and CANAPE analyses have yet to be thoroughly tested using these types of range estimations. There could be species whose distributions are controlled by other factors instead of (or in addition to) the macro-environment, meaning that caution must be taken when choosing the taxa to model (Patsiou et al., 2014).

Taxon sampling is ideally made using lineages at as fine a scale as possible.

However, the lack of availability of DNA data and locality information for many species, combined with poor understanding of fine-scale taxonomy in some Chilean taxa, makes it more feasible and likely more taxonomically accurate to work at higher levels such as genera, at least until we have more complete databases. Working at these higher levels has proven very useful in outlining general spatial patterns of biodiversity in several previous studies (Forest et al., 2007; Thuiller et al., 2011; Thornhill et al., 2016). Adding finer scale OTUs in the future will fill out details in these broad patterns but is not likely to overturn them given that most of the phylogeny is being taken into account using the higher-level OTUs.

4.2. Spatial patterns of phylogenetic diversity and endemism in the vascular flora of Chile

Taxon richness in Chile has a known geographical pattern, concentrating in the Mediterranean climatic region of Central Chile, and decreasing both to the north and to the south (Arroyo et al., 2008; Moreira-Muñoz, 2011). Observed PD showed an almost identical pattern, which is to be expected, given that PD is tightly correlated with taxon richness (Faith, 1992b; Forest et al., 2007), especially if taxa are randomly sampled from the phylogeny. However, when considering statistically significant departures from the expected correlation, different patterns are apparent. For example, the south of the country (south of 35° latitude) showed some cells containing taxa representing more of the phylogeny than expected by chance, as well as many cells with a concentration of longer branches than expected by chance. The north of the country showed nearly the opposite patterns, with many cells containing a significant under-representation of the phylogeny and shorter branches than expected. The CANAPE analysis showed that most significant neo-endemism is found in the north of the country whereas the south contains most of the paleo-endemism.

Recent geological history of the country can help in understanding these patterns.

During glaciations, many lineages were restricted to a few refuges mostly located in southern areas, which constituted suitable places for the long-term survival of lineages that expanded after the ice melted (Taberlet & Cheddadi, 2002) Armesto et al., 1994; Villagrán et al., 1995; Villagrán, 2001). Phylogeographic patterns in plants and vertebrates also suggest the existence of areas that were stable glacial refuges in coastal regions between 36 and 41° south (Sérsic et al., 2011).

Patterns of endemism can also be understood by considering the paleofloras that contribute to modern formations, and the current conditions that subtend certain portions of them. For example, the south of Chile (34-45° south) contains temperate rainforest with many unique taxa of different origins, such as Laurifolius forest of neotropical origin, mixed with deciduous and evergreen forest of sub-antarctic origin (Moreira-Muñoz, 2011; Tecklin et al., 2011). In the Southern hemisphere, there is evidence of an earlier origin for floras that inhabit southern areas, than the ones further north. In woody plants for example, older families were found in southern temperate forests, whereas younger families were found to be distributed in mediterranean and arid climates (Segovia et al., 2013). This pattern is mainly due to the Australasian and Austral-Antarctic components of the southern Gondwanan flora, and is proposed as the *Austral niche conservatism hypothesis* (Segovia & Armesto, 2015). In fact, Hinojosa (2005) describes an Eocene to early Miocene (55–20 Mya) mixed flora with elements of the Austral-Antarctic floristic element for the south of Chile. The Neogene subtropical paleoflora is the likely precursor of the sclerophyllous forests in central Chile at 32–33° south, dated 20–15 Mya (Hinojosa et al., 2006). In central-northern Chile, by the mid-Miocene (15 Mya), the Andes Mountains reached about half their current altitude, possibly already exerting a rain-shadow effect (Gregory-Wodzicki, 2000; Schlunegger et al., 2010). Semi-arid to arid climate, based on sedimentological evidence, prevailed in the central Andes from 15–4 Mya. In the northernmost area of Chile, hyperaridity is attributed to a pre-existing arid condition and the onset of the Humboldt current 3.5–3 Mya (Hartley, 2003). Hence, modern floral elements in this area are more recent in origin. These different origins may explain why we see more significant paleo-endemism

to the south and neo-endemism to the north of the country.

Interesting differences were observed when different subsets of the flora were analyzed using CANAPE. Centers of paleo-endemism in the south were evident when all vascular plants were taken into account, but many disappeared when analyzing seed plants alone and angiosperms alone meaning that the monilophytes and gymnosperms are likely the main contributors to paleo-endemism in that part of the country. However, patterns of neo-endemism in the north of Chile remained consistent for the three subsets of the flora, suggesting that these patterns are due to the more recent evolution of angiosperm genera.

A spatial phylogenetic study of Australian angiosperm genera found a similar general pattern, with desert areas concentrating neo-endemism, and significant paleo-endemism concentrated in regions that still contain Gondwanan elements (such as western Tasmania) (Thornhill et al., 2016). In fact, a recent study in Tasmania correlates paleo-endemic clades with humid climates lacking extreme temperatures, and open vegetation with rare or no fire (Jordan et al., 2016). These predictive elements coincide with the conditions observed in the south of Chile.

4.3. The main regional patterns encountered in this study

Northern area (Tropical bioclimate): This arid and semiarid area showed low observed richness and significantly low PD, suggesting phylogenetic clustering, an indication of phylogenetic conservatism in niche preference. RPD was also significantly low, indicating that shorter phylogenetic branches are present than expected, and suggesting recent radiations have occurred in the area. CANAPE results showed a concentration of neo-endemism, supporting the idea that the area hosts recently derived genera that are also range-restricted taxa. This is coincident with previous studies showing recent radiations in the area. Guerrero et al. (2013) for example, studied two prevalent arid adapted plant clades (*Chaethanthera* and *Malesherbia*), and suggested that their invasion of desert areas occurred around the last 10 million years; however, they also showed an important time-lag between invasion and radiation of those clades, suggesting that

adaptation to arid conditions might take a long time to occur. Our determination of neo-endemism is even more significant considering that the analysis was made at the genus level and most proposals by researchers for recent radiation are judged at the species level.

Central area (Mediterranean bioclimate): The central area has been long recognized as the most species-rich (Arroyo et al. 2008; Arroyo et al. 2004; Myers et al. 2000). We found that it likewise is high in PD and PE. However, it does not exhibit significant patterns in PD, RPD, or CANAPE (except for a few cells in the Andes region showing significant mixed endemism in CANAPE). This means that even though taxon richness is high, the genera inhabiting this area are geographically widespread and come from a relatively even sampling of the phylogeny, with phylogenetic branches that are neither particularly long nor short.

Southern area (Temperate bioclimate): The southern area showed low observed TR and PD. There were a few areas of significantly high PD but many areas of significantly high RPD, meaning that in general the south has an over-representation of long branches, which could have arisen by extinction of close relatives. This is possible given the presence of glacial refuges, especially in the coastal mountains (Hinojosa 2005). CANAPE results are consistent with the idea of range restricted Gondwanan components of the flora, showing a concentration of paleo-endemic cells. Much of the paleo-endemism in this area appears to be in the non-angiosperm component (ferns and gymnosperms), which is evident when these taxa are removed from the analysis. Examples of gymnosperm taxa that inhabit the area that are probably relicts of previously more lineage-rich groups include the genus *Podocarpus*, the southern distribution of the monotypic genera *Austrocedrus*, restricted to the Andes mountain, and *Fitzroya* in small patches along both mountain ranges, and the only representative of the genus *Araucaria* in Chile, a classic example of a relict Gondwanan endemic (Veblen & Schlegel, 1982; Smith-Ramirez et al., 2005; Donoso, 2006).

Austral area (Boreal bioclimate): This area has a medium level of observed richness and PD, but many areas of significantly high RPD, indicating longer than expected branches, probably Gondwanan relicts. The CANAPE analysis also shows some centers of paleo-endemism, located on the coast, which has been recognized as a glacial refuge. Examples of taxa representing Gondwanan relicts include the genus *Nothofagus*, with the southern species distributed in coastal areas in the Patagonia of Chile and Argentina, and *Embothrium*, distributed in both mountain ranges and in Patagonia (Donoso 2006). This area is also rich in ferns, including genera such as *Ophioglossum*, *Hymenophyllum*, *Lophosoria*, and *Hyplolepis* with a southern distribution (Gunckel, 1984; Marticorena & Rodríguez, 1995).

4.4. Phylogenetic turnover in relation to biomes of Chile

The phylogenetic turnover clusters are highly geographically structured which suggests that there is a deep phylogenetic signal to the way that the flora has been formed in Chile. Range-weighted phylogenetic turnover shows transitions that reflect previously recognized biomes of Chile more closely than does the PhyloSørensen metric, which does not consider range restriction. As previously discussed by Laffan et al., (2016), when range restriction is not taken into account, biotic transitions can be obscured due to the effect of widespread taxa dominating the turnover patterns, and more so if these taxa are on long phylogenetic branches. This may be more prevalent in areas of high endemism, such as Chile. However, caution must be taken when interpreting these analyses because of the potential biases due to the level of terminals used in a phylogenetic tree. For example, in instances in which an OTU is widespread but has subclades that are range-restricted, phylogenetic turnover analyses are not be able to consider those range-restricted taxa. This could be relevant to the present study that includes genera that are widespread but show range-restricted species. One example is the genus *Astragalus*, distributed all along Chile, but containing extremely range-restricted species, especially in

the north of the country.

4.5. Conservation considerations

The flora of Chile shows clear patterns in the spatial distribution of its evolutionary heritage. The central part of the country has long been recognized as a biodiversity hotspot, concentrating the highest observed taxon richness, and our study confirms this using PD. However, our analyses indicate that there are other areas of the country that emerge as highly significant concentrations of diversity and endemism that are also worth preserving, in addition to this hotspot. In view of these results, future research can ask whether our currently protected areas are in fact covering these new zones that are interesting from an evolutionary point of view. Are we sufficiently protecting areas with concentrations of restricted long branches, relicts of paleofloras, or areas that harbor unique range restricted taxa of recent radiation?

Acknowledgements

We thank the Center for Latin American Studies at UC Berkeley and CONICYT for a seed money grant to begin this collaboration (fund 77447). CONICYT and the University of Chile also provided major funding under grants PII20150091 and Enlace Fondecyt Grant ENL035/16 respectively. The US NSF grant DEB-1354552 provided partial support for the UC Berkeley portion of this project. Nunzio Knerr from the National Research Collections of Australia – CSIRO, Australia provided technical assistance towards making the phylogenetic diversity and CANAPE figures. We thank the curators of the CONC, SGO, ULS and EIF herbaria in Chile who provided locality information for building the geographic database.

Supplementary material

Supplementary information may be found in the online version of this article, including the geographic database, alignments, and phylogeny used.

Appendix A GenBank accession data

Appendix B PhyloSorensen analysis, Figure S1.

References

- Armesto J., Villagrán C., & Donoso C. 1994. La historia del bosque templado chileno. *Amb. Des.* 3, 66–72.
- Arroyo M.T.K., Marquet P., Marticorena C., Simonetti J., Cavieres L.A., Squeo F., & Rozzi R. 2004. Chilean winter rainfall-Valdivian forest. Hotspots revisited: Earth's biologically wealthiest and most threatened ecosystems (Ed.) R.A. Mittermeier, P.R. Gil, M. Hoffmann, J. Pilgrim, C.G. Mittermeier, J. Lamoreux, and G. da Fonseca, CEMEX, Mexico DF.
- Arroyo M.T.K., Marquet P., Marticorena C., Simonetti J., Cavieres L.A., Squeo F., Rozzi R., & Massardo F. 2008. El hotspot chileno, prioridad mundial para la conservación. *Biodiversidad de Chile, patrimonio y desafíos* (Ed.) CONAMA, Ocho Libros, Santiago.
- CAPP 2008. Informe país: estado del medio ambiente en Chile. Instituto de asuntos públicos (Ed.) C. de A. de P.P. UdeC, Santiago.
- Cardoso P., Erwin T.L., Borges P., New T.R. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conserv.* 144, 2647–2655.
- Cohen J. (1960) A coefficient of agreement of nominal scales. *Educ. Psychol. Meas.* 20, 37–46.
- Collen B., Turvey S.T., Waterman C., Meredith H.M.R., Kuhn T.S., Baillie J.E.M., & Isaac N.J.B. 2011. Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 2611–22.
- Crisp M.D., Laffan S., Linder H.P., Munro A. 2001. Endemism in the Australian Flora. *J. Biogeogr.* 28, 183–198.
- Daru B.H., van der Bank M., Davies T.J. 2015. Spatial incongruence among hotspots and

complementary areas of tree diversity in southern Africa. *Divers. Distrib.* 7, 769-780.

Donoso C. .2006. Las especies arbóreas de los bosques templados de Chile y Argentina. Autoecología. Marisa Cuneo Ediciones, Valdivia, Chile.

Dunn C.E., Atkins P.J., Townsend J.G. 1997. GIS for development : a contradiction in terms? *Area* 29, 151–159.

Faith D. 1992a. Systematics and conservation: On predicting the feature odiversity of subsets of taxa. *Cladistics-the Int. J. Willi Hennig Soc.* 8, 361–373.

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

Faith D.P., Magallón S., Hendry A.P., Conti E., Yahara T., Donoghue M.J. 2010. Ecosystem services: An evolutionary perspective on the links between biodiversity and human well-being. *Curr. Opin. Environ. Sustain.* 2, 66–74.

Forest F., Grenyer R., Rouget M., Davies T.J., Cowling R.M., Faith D.P., Balmford A., Manning J.C., Procheş S., van der Bank M., Reeves G., Hedderson T. a J., Savolainen V. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, 445, 757–60.

González-Orozco C.E., Pollock L.J., Thornhill A.H., Mishler B.D., Knerr N.J., Laffan S.W., Miller J.T., Rosauer D.F., Faith D.P., Nipperess D.A., Kujala H., Linke S., Butt N., Külheim C., Crisp M.D., Gruber B. 2016. Phylogenetic approaches reveal biodiversity threats under climate change. *Nat. Clim. Chang.* 6, 1110-1114.

Graham C.H., Fine P.V. 2008. Phylogenetic beta diversity : linking ecological and evolutionary processes across space in time. *Ecol. Lett.* 11, 1265–1277.

Gregory-Wodzicki K.M. 2000. Uplift history of the Central and Northern Andes: A review. *Geol. Soc. Am. Bull.* 112, 1091–1105.

- Guerrero P.C., Rosas M., Arroyo M.T.K., Wiens J.J. 2013. Evolutionary lag times and recent origin of the biota of an ancient desert (Atacama-Sechura). *Proc. Natl. Acad. Sci. U. S. A.* 110, 11469–74.
- Gunckel H. 1984. *Helechos de Chile. Monografías Anexadas a los Anales de la Universidad de Chile.* Ediciones de la Universidad de Chile, Santiago.
- Hartley A. 2003. Andean uplift and climate change. *J. Geol. Soc. London.* 160, 7–10.
- Hernández H.J., Estades C., Faúndez L. 2013. Diagnóstico levantamiento Biodiversidad Región Arica y Parinacota BIP 30104625-0. Gobierno Regional de Arica y Parinacota, Universidad de Chile.
- Hinojosa L.F. 2005. Climatic and vegetational changes inferred from Cenozoic Southern South America paleoflora. *Rev. Geol. Chile* 32, 95–115.
- Hinojosa L.F., Armesto J.J., Villagran C. 2006. Are Chilean coastal forests pre-Pleistocene relicts? Evidence from foliar physiognomy, palaeoclimate, and phytogeography. *J. Biogeogr.* 33, 331–341.
- Hortal J., Lobo J.M., Jiménez-Valverde A. 2007. Limitations of Biodiversity Databases : Case Study on Seed-Plant Diversity in Tenerife , Canary Islands. *Conserv. Biol.* 21, 853–863.
- Jordan G.J., Harrison P.A., Worth J.R.P., Williamson G.J., Kirkpatrick J.B. 2016. Palaeoendemic plants provide evidence for persistence of open, well-watered vegetation since the Cretaceous. *Glob. Ecol. Biogeogr.* 25, 127–140.
- Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kraft N.J.B., Baldwin B.G., Ackerly D.D. 2010. Range size, taxon age and hotspots of neoendemism in the California flora. *Divers. Distrib.* 16, 403–413.
- 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.

Laffan S.W., Rosauer D.F., Virgilio G. Di, Miller C.E., González-Orozco C.E., Knerr N.J., Thornhill A.H., Mishler B.D. 2016. Range-weighted metrics of species and phylogenetic turnover can better resolve biogeographic transition zones. *Methods Ecol. Evol.* 7, 580–588.

Linder H.P. 2001. On areas of endemism, with an example from the African Restionaceae. *Syst. Biol.* 50, 892–912.

Lomolino M. 2004. Conservation biogeography. *Frontiers of Biogeography: New Directions in the Geography of Nature* (Ed.) M. Lomolino and L.R. Heaney), pp. 293–296. Sinauer Associates, Sunderland, Massachusetts.

Luebert F. & Pliscoff P. 2006. Sinopsis biotómica y vegetal de Chile. Editorial Universitaria, Santiago.

Marchese C. 2015. Biodiversity hotspots: A shortcut for a more complicated concept. *Glob. Ecol. Conserv.* 3, 297–309.

Margules C.R., Pressey R.L. 2000. Systematic conservation planning. *Nature* 405, 243–53.

Martcorena C., Rodríguez R. 1995. Flora de Chile Vol.I. Pteridophyta - Gymnospermae. Editora Anibal Pinto S.A., Concepción.

Mennecke B.E., West L.A.J. 2001. Geographic Information Systems in Developing Countries : Issues in Data Collection , Implementation and Management. *J. Glob. Inf. Manag.* 9, 45–55.

Millennium Ecosystem Assessment. 2005. Ecosystems and Human Well-being: Synthesis. Island Press, Washington DC.

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. Nat. Commun. 5, 1–10.

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., Worm B. 2011. How many species are there on earth and in the ocean? PLoS Biol. 9, 1–8.

Moreira-Muñoz A. 2011. Plant geography of Chile. Springer, Santiago.

Myers N., Mittermeier R.A., Mittermeier C.G., Fonseca G.A.B., Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.

Patsiou T.S., Conti E., Zimmermann N.E., Theodoridis S., Randin C.F. 2014. Topo-climatic microrefugia explain the persistence of a rare endemic plant in the Alps during the last 21 millennia. Glob. Chang. Biol. 20, 2286–2300.

Pio D. V., Engler R., Linder H.P., Monadjem A., Cotterill F.P.D., Taylor P.J., Schoeman M.C., Price B.W., Villet M.H., Eick G., Salamin N., Guisan A. 2014. Climate change effects on animal and plant phylogenetic diversity in southern Africa. Glob. Chang. Biol. 20, 1538–1549.

Pio D. V., Broennimann O., Barraclough T.G., Salamin N. 2011. Spatial Predictions of Phylogenetic Diversity in Conservation Decision Making. Conserv. Biol. 25, 1229–1239.

Rockström J., Steffen W., Noone K., Persson Å., Chapin F.S., Lambin E., Lenton T.M., Scheffer M., Folke C., Schellnhuber H.J., Nykvist B., de Wit C. a., Hughes T., van der Leeuw S., Rodhe H., Sörlin S., Snyder P.K., Constanza R., Svedin U., Falkenmark M., Karlberg L., Corell R.W., Fabry V.J., Hansen J., Walker B., Liverman D., Richardson K., Crutzen P., Foley J. 2009a. A safe operating space for humanity. Nature 461, 472–475.

Rockström J., Steffen W., Noone K., Persson Å., Chapin F.S., Lambin E., Lenton T.M., Scheffer M., Folke C., Schellnhuber H.J., Nykvist B., de Wit C. a., Hughes T., van der Leeuw S., Rodhe H., Sörlin S., Snyder P.K., Costanza R., Svedin U., Falkenmark

- M., Karlberg L., Corell R.W., Fabry V.J., Hansen J., Walker B., Liverman D., Richardson K., Crutzen P., Foley J. 2009b. Planetary boundaries: Exploring the safe operating space for humanity. *Ecol. Soc.* 14, 32.
- Rosauer D., Laffan S.W., Crisp M.D., Donnellan S.C., Cook L.G. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* 18, 4061–72.
- Scherson R. a., Albornoz A. a., Moreira-Muñoz A.S., Urbina-Casanova R. 2014. Endemicity and evolutionary value: a study of Chilean endemic vascular plant genera. *Ecol. Evol.* 4, 806–816.
- Schlunegger F., Kober F., Zeilinger G., Rotz R. 2010. Sedimentology-based reconstructions of paleoclimate changes in the Central Andes in response to the uplift of the Andes, Arica region between 19 and 21°S latitude, northern Chile. *Int. J. Earth Sci.* 99, 123–137.
- Schmidt-Lebuhn A.N., Knerr N.J., Miller J.T., Mishler B.D. 2015. Phylogenetic diversity and endemism of Australian daisies (Asteraceae). *J. Biogeogr.* 42 (6), 1114–1122.
- Sechrest W., Brooks T.M., Fonseca G.A.B., Konstant W.R., Mittermeier R.A., Purvis A., Rylands A.B., Gittleman J.L. 2002. Hotspots and the conservation of evolutionary history. *PNAS* 99, 17–21.
- Segovia R.A., Armesto J.J. 2015. The Gondwanan legacy in South American biogeography. *J. Biogeogr.* 42, 209–217.
- Segovia R. a., Hinojosa L.F., Pérez M.F., Hawkins B. a. 2013. Biogeographic anomalies in the species richness of Chilean forests: Incorporating evolution into a climatic - historic scenario. *Austral Ecol.* 38, 905–914.
- Sérsic A.N., Cosacov A., Cocucci A. a., Johnson L. a., Pozner R., Avila L.J., SITES Jr. J.W., Morando M. 2011. Emerging phylogeographical patterns of plants and

terrestrial vertebrates from Patagonia. *Biol. J. Linn. Soc.* 103, 475–494.

Smith-Ramirez C., Pliscoff P., Teillier S., Barrera E. 2005. Patrones de riqueza y distribución de la flora vascular en la Cordillera de la Costa de Valdivia, Osorno y Llanquihue, Chile. *Historia, biogeografía y ecología de los bosques costeros de Chile* (ed. by C. Smith-Ramirez, J.J. Armesto, and C. Valdovinos), pp. 253–278. Editorial Universitaria, Santiago.

Steffen W., Richardson K., Rockström J., Cornell S., Fetzer I., Bennett E., Biggs R., Carpenter S.R., de Wit C. a., Folke C., Mace G., Persson L.M., Veerabhadran R., Meyers B., Sörlin S. 2015. Planetary Boundaries: Guiding human development on a changing planet. *Science* 347, 1259855.

Stork N.E. 1993. How many species are there? *Biodivers. Conserv.* 2, 215–232.

Taberlet P., Cheddadi R. 2002. Ecology. Quaternary refugia and persistence of biodiversity. *Science* 297, 2009–10.

Tecklin D., DellaSala D.A., Luebert F., & Pliscoff P. 2011. Valdivian temperate rainforests of Chile and Argentina. *Temperate and Boreal Rainforests of the World: Ecology and Conservation* (Ed.) D. DellaSala, pp. 132–153. Island Press, Washington DC.

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

Thuiller W., Lavergne S., Roquet C., Boulangeat I., Lafourcade B., Araujo M.B. 2011. Consequences of climate change on the tree of life in Europe. *Nature* 470, 531–4.

Veblen T.T., Schlegel F.F. 1982. *Reseña ecológica de los bosques del sur de Chile*. *Bosque* 4, 73–115.

Villagrán C. 2001. Un modelo de la historia de la vegetación de la Cordillera de La Costa

de Chile central-sur: la hipótesis glacial de Darwin. *Rev. Chil. Hist. Nat.* 74, 793–803.

Villagrán C., Armesto J.J. 2005. Fitogeografía histórica de la Cordilera de la Costa de Chile. *Historia, biogeografía y ecología de los bosques costeros de Chile* (Ed.) C. Smith-Ramirez, J.J. Armesto, and C. Valdovinos, pp. 99–115. Editorial Universitaria, Santiago.

Villagran C., Hinojosa F. 1997. Historia de los bosques del sur de Sudamérica , II : Análisis fitogeográfico History of the forests of southern South America , II : phytogeographical analysis. *Rev. Chil. Hist. Nat.* 70, 241–267.

Villagrán C., Moreno P., Villa R. 1995. Antecedentes palinológicos acerca de la historia cuaternaria de los bosques chilenos. *Ecología de los Bosques Nativos de Chile* pp. 69. Editorial Universitaria, Santiago.

Visser H., de Nus T. 2006. The map comparison kit. *Environ. Model. Softw.* 21, 346–358.

Whittaker R.J., Araujo M.B., Jepson P., Ladle R.J., Watson J.E.M., Willis, K. J. R.J. 2005. Conservation biogeography: assessment and prospect. *Divers. Distrib.* 11, 3–23.

Wilson K. a, McBride M.F., Bode M., Possingham H.P. 2006. Prioritizing global conservation efforts. *Nature* 440, 337–40.

Zuloaga F., Morrone O., M.J. B. 2008. Catálogo de las plantas vasculares del cono Sur (Argentina, Sur de Brasil, Paraguay y Uruguay). Volumen 2. Dicotyledoeae: Acanthaceae-fabaceae (Abarema-Schizolobium). *Monographs in Systematic Botany from the Missouri Botanical Garden.* pp. 1–2286. St. Louis Missouri.

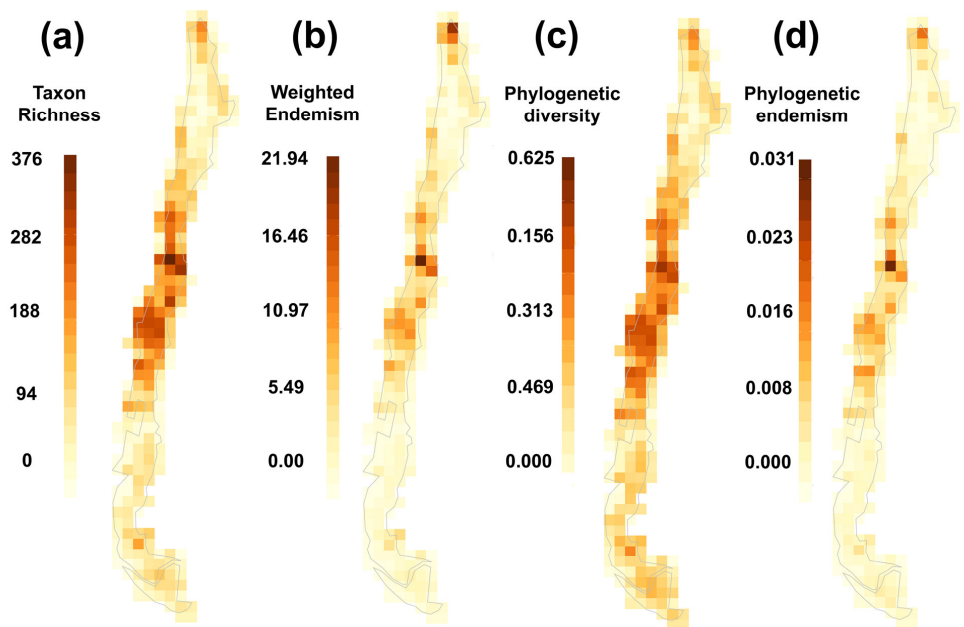
Figure Captions

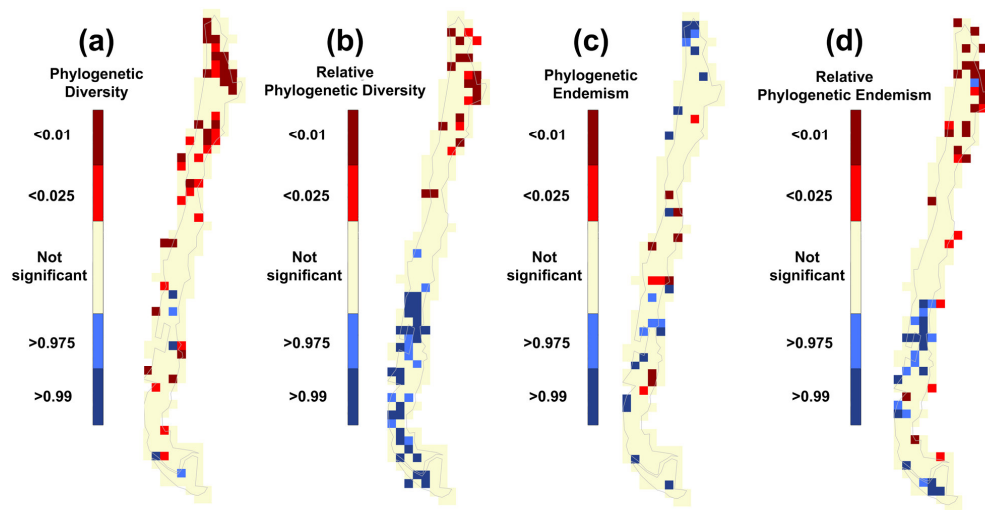
Fig. 1. Observed biodiversity patterns of Chilean vascular flora. (a) Taxon (genus) richness (TR); (b) weighted endemism (WE); (c) phylogenetic diversity (PD); and (d) phylogenetic endemism (PE)

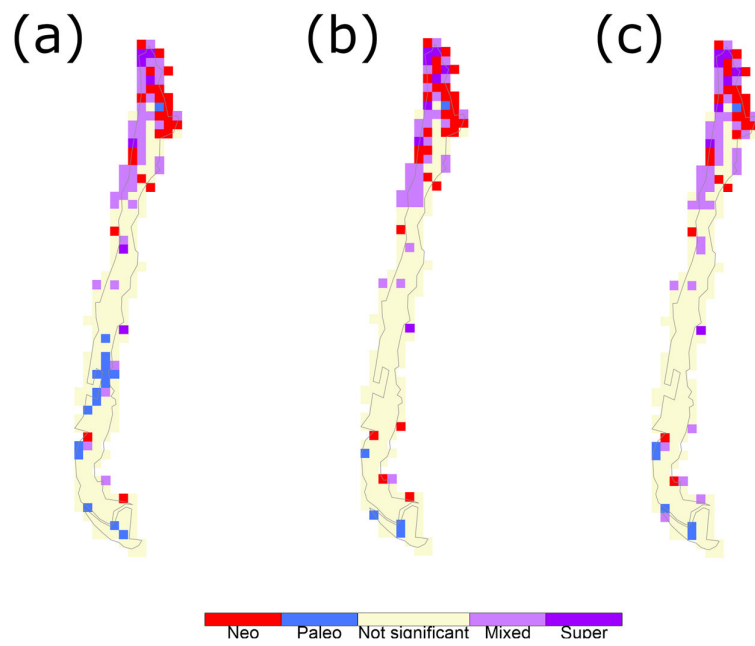
Fig. 2. (a) Significance results for phylogenetic diversity, from two-tailed randomization tests (see methods) of Chilean vascular plant genera. Beige cells are not significant, red cells contain significantly less PD than expected; blue cells contain significantly more PD than expected. (b) Significance results for relative phylogenetic diversity (RPD), from the same randomization tests. Beige cells are not significant; red cells contain significantly low RPD, meaning that the phylogenetic branches present in that grid cell are shorter than expected. Blue grid cells contain significantly high RPD, meaning that the branches present in that area are longer than expected.

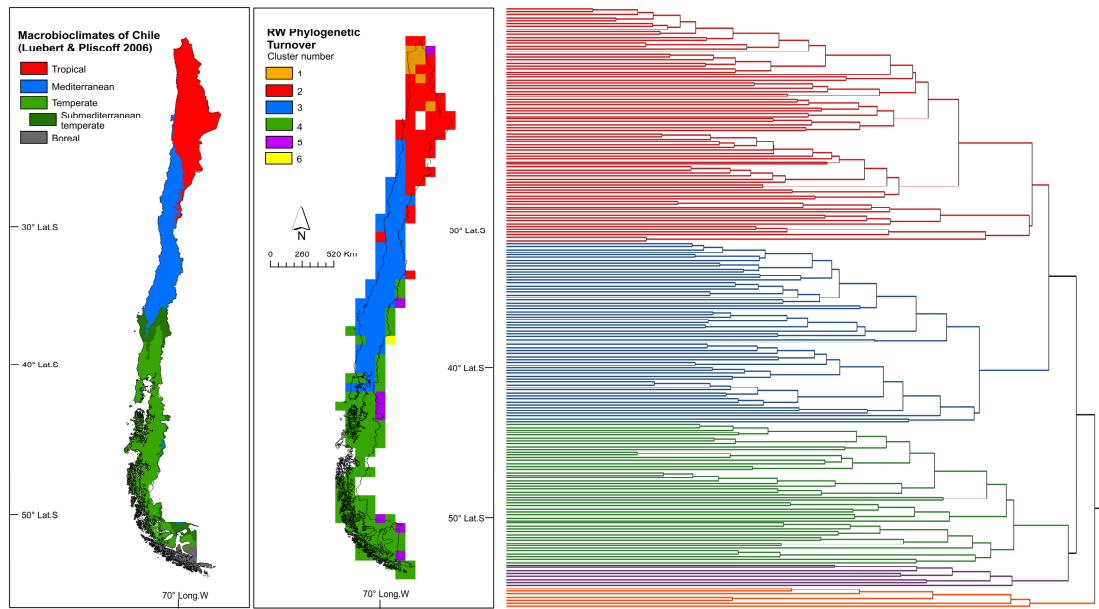
Fig. 3. Map of centers of phylogenetic endemism derived from the CANAPE analyses. Beige cells are not significant; red cells indicate centers of neo-endemism; blue cells indicate centers of paleo-endemism. Purple cells are a mix of neo and paleo-endemism and darker purple cells indicate centers of super-endemism. (a) analysis done with all vascular plants; (b) analysis including only spermatophytes (seed plants); (c) analysis including only angiosperms (flowering plants).

Fig. 4. Range-weighted phylogenetic turnover analysis and its relationship with the macrobioclimates of Chile. (a) map of the microbioclimates of Chile according to Luebert & Pliscoff (2006), (b) RW Phylogenetic turnover analysis, indicating the geographic position of the clusters shown in (c).









Highlights

- Adds evolutionary information to standard assays of biodiversity that use species
- Interesting centers with significant paleo-endemism were found in the south of Chile
- The north of Chile showed high phylogenetic endemism, including much neo-endemism
- Range-weighted phylogenetic turnover proved useful in detecting biotic transitions
- The results provide important aids for conservation assessment

Graphical abstract

