Is the age of plant communities predicted by the age, stability and soil composition of the underlying landscapes? An investigation of OCBILs

MARIA BEATRIZ DE SOUZA CORTEZ^{1,2,*}, RYAN A. FOLK³, CHARLES J. GRADY⁴, JONATHAN P. SPOELHOF², STEPHEN A. SMITH⁵, DOUGLAS E. SOLTIS^{1,2,6,7} and PAMELA S. SOLTIS^{2,6,7}

Received 29 July 2020; revised 21 September 2020; accepted for publication 22 September 2020

Old, climatically buffered, infertile landscapes (OCBILs) have been hypothesized to harbour an elevated number of persistent plant lineages and are predicted to occur across different parts of the globe, interspersed with other types of landscapes. We tested whether the mean age of a plant community is associated with occurrence on OCBILs, as predicted by climatic stability and poor soil environments. Using digitized occurrence data for seed plants occurring in Australia (7033 species), sub-Saharan Africa (3990 species) and South America (44 482 species), regions that comprise commonly investigated OCBILs (Southwestern Australian Floristic Region, Greater Cape Floristic Region and campos rupestres), and phylogenies pruned to match the species occurrences, we tested for associations between environmental data (current climate, soil composition, elevation and climatic stability) and two novel metrics developed here that capture the age of a community (mean tip length and mean node height). Our results indicate that plant community ages are influenced by a combination of multiple environmental predictors that vary globally; we did not find statistically strong associations between the environments of OCBIL areas and community age, in contrast to the prediction for these landscapes. The Cape Floristic Region was the only OCBIL that showed a significant, although not strong, overlap with old communities.

ADDITIONAL KEYWORDS: Australia – $campos\ rupestres$ – climatic stability – GCFR – infertile soils – OCBIL theory – seed plant community age – South America – sub-Saharan Africa – SWAFR.

INTRODUCTION

The Earth's surface is a patchwork of ancient and recently formed landscapes, shaped by geological processes over millions of years that in turn shape today's biota. For example, the flora of North America reflects continent-wide landscape dynamics (Mishler *et al.*, 2020); geologically stable eastern North America supports forest communities that are phylogenetically overdispersed (with species more

distantly related than expected by chance), while relatively recent orogeny, volcanism and other large-scale changes in western North America have driven recent speciation in many clades, resulting in a flora that is phylogenetically clustered (with species more closely related than expected by chance). Although the impacts of landscape dynamics on patterns of species radiation emerge in individual areas (e.g. rapid floristic diversification accompanying recent orogeny in the Andes; Bell *et al.*, 2012; Lagomarsino *et al.*, 2016), a generalizable relationship between landscape age and biotic attributes remains unclear. OCBIL

¹Department of Biology, University of Florida, Gainesville, FL 32611, USA

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

³Department of Biological Sciences, Mississippi State University, Starkville, MS 39762, USA

⁴Biodiversity Institute & Natural History Museum, University of Kansas, Lawrence, KS 66045, USA

⁵Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48103, USA

⁶Genetics Institute, University of Florida, Gainesville, FL 32608, USA

⁷Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA

^{*}Corresponding author. E-mail: mariabiacortez@gmail.com

theory (Hopper, 2009) sets out seven predictions to characterize the biota of old, climatically buffered, and infertile landscapes (OCBILs): (1) reduced dispersibility, increased endemism and prevalent rarity; (2) accentuated persistence of old lineages and old individuals; (3) selection of heterozygosity; (4) prolonged speciation at the margins; (5) adaptations to saline soils; (6) nutritional and other biological specialization; and (7) particular vulnerability (e.g. to soil removal) and enhanced resilience (e.g. to fragmentation).

Whereas the ecological and environmental differences between OCBILs and young, often disturbed, fertile landscapes (YODFELs) are clearly defined conceptually, empirically establishing geographical limits between these landscapes is not straightforward (Mucina & Wardell-Johnson, 2011). OCBILs typically occur in elevated areas formed either by granite outcrops or by laterite, ironstone or sandstone hilltops, and are found interspersed with YODFELs and OCFELs (old, climatically buffered, fertile landscapes), forming complex and heterogeneous assemblies (Hopper et al., 2016).

Hopper (2009) highlighted three regions across the globe with landscapes he identified as OCBILs: the Southwestern Australian Floristic Region (SWAFR), the Greater Cape Floristic Region (GCFR) of South Africa [comprising the Cape Floristic Region (CFR) and the Succulent Karool and the Pantepuis of northeastern South America. Subsequently proposed regions where OCBILs could be commonly found include the campos rupestres in Brazil, the Hawkesbury basin sandstone vegetation of Sydney (Australia), Western Tasmania, New Caledonia, Madagascar and Indian Ocean Islands (Hopper et al., 2016). Not surprisingly, these areas are all situated in the Southern Hemisphere, where climatic conditions generally have long been more stable than in the Northern Hemisphere (Dynesius & Jansson, 2000) and where biodiversity hotspots prevail (Myers et al., 2000).

Together, the SWAFR, GCFR and campos rupestres occupy only ~0.4% of the Earth's land surface (Pidwirny, 2006), but they contain over 27 000 vascular plant species (Hilton-Taylor, 1996; Goldblatt et al., 2005; Silveira et al., 2016; Gioia & Hopper, 2017) corresponding to ~7.4% of the estimated 366 000 vascular plant species on Earth (Judd et al., 2008; Paton et al., 2008). Traditionally, productive warm and humid habitats, such as tropical forests, have received more attention from researchers interested in understanding patterns of high plant species diversity (Furley & Metcalfe, 2007). Recently, drier, nutrientimpoverished regions such as OCBILs have been recognized as important centres of floristic diversity, yet diversification patterns for these areas remain unclear (Werneck, 2011).

Beyond contributions to understanding the biological attributes of species in ancient landscapes, OCBIL theory could be useful to promote formulation of conservation strategies for these highly unusual areas (Hopper, 2009). To date, conservation efforts have instead largely focused on YODFELs. Understanding how OCBIL and YODFEL landscapes differ will result in more effective conservation strategies that consider the specific ecological and environmental conditions of each landscape. For example, conservation strategies to improve soil nutrition in young landscapes focus on nitrogen as the main limiting nutrient. whereas phosphorus levels are more important in old landscapes (Mucina & Wardell-Johnson, 2011). OCBILs and YODFELs also host very distinct biotas; while generalists tend to predominate in YODFELs (Hopper, 2009), OCBILs commonly contain localized specialists and endemics that face a higher risk of extinction (Myers et al., 2000). Human activities involving the exploitation of several metals (iron, cobalt and nickel are commonly mined in OCBILs) also increase extinction risk (Hopper et al., 2016).

According to the second prediction of OCBIL theory, the SWAFR, GCFR, campos rupestres and other OCBIL regions should contain both ancient lineages and old individual organisms (Hopper, 2009). A more stable climate would allow for persistence of ancient lineages that diversified deeper in time, particularly lineages surviving from the supercontinent Gondwana, and for individual genets to persist longer through selfing, clonality and underground resprouting structures (Hopper, 2009). Mucina & Wardell-Johnson (2011) further developed this hypothesis, proposing that OCBIL regions should host 'mature radiations' and preserve old lineages via low extinction rates.

Recent studies support this second prediction of OCBIL theory. A fossil dating from the Late Eocene, Banksia paleocrypta R.J.Carp., McLoughlin, R.S.Hill, McNamara & G.J.Jord (Proteaceae), was found in the SWAFR (Carpenter et al., 2014), where Banksia L. now contains 170 extant species (Cardillo & Pratt, 2013). This fossil presents anatomical structures suggesting adaptation to xeromorphic conditions, indicating that the Late Eocene climate was similar to contemporary climate in that area (Carpenter et al., 2014). In the GCFR, Mesembryanthemoideae, a clade within the flowering plant family Aizoaceae, is hypothesized to have radiated during the Plio-Pleistocene (Klak et al., 2014), and in the campos rupestres, the monocot Vellozia L. (Velloziaceae) has experienced higher diversification rates since the Early Miocene, around 23 Mya (Alcantara et al., 2018). Long-lived trees and herbs have also been reported in the *campos rupestres*; clonal Jacaranda decurrens Cham. (Bignoniaceae) has been conservatively estimated to be ~4000 years old (Alves et al., 2013), while a single plant of Vellozia

kolbekii Alves was estimated to be over 500 years old (Alves, 1994).

Lineage age varies significantly among OCBILs. In the GCFR, the Succulent Karoo is dominated by lineages no older than 17.5 Mya, whereas the fynbos presents a more complex flora, with age estimation for lineages ranging from 60 to 15 Mya (Verboom *et al.*, 2008). This difference is probably related to areas of refugia in the fynbos, where moister, cooler and higher elevation habitats enabled persistence of lineages, emphasizing complex and heterogeneous plant diversity patterns in OCBILs (Verboom *et al.*, 2008).

Although research supports the prediction that OCBILs will harbour both older lineages and long-lived individuals, most studies have focused on only a few taxonomic groups or on a specific geographical region (e.g. Alves et al., 2013; Cardillo & Pratt, 2013; Alcantara et al., 2018). A global approach would be valuable in testing the generalizability of OCBIL theory. The combination of digitized herbarium data and largescale phylogenies is an efficient tool for utilizing large quantities of botanical data to characterize spatial patterns from continental to regional scales (Thornhill et al., 2016, 2017; Zappi et al., 2017; Lu et al., 2018; Allen et al., 2019; Dagallier et al., 2020; Mishler et al., 2020) and to delimit phytochoria, geographical areas that each have a relatively uniform composition of plant species (Born et al., 2007).

Present biodiversity patterns cannot be understood as a product solely of current climatic conditions or biotic factors such as dispersal. To understand plant diversification patterns at any spatial scale, simultaneous consideration of contemporary and historical environmental factors is necessary. Speciation and extinction processes reflect changes occurring over millions of years (Eisenhardt et al., 2015), and present-day factors filter current community composition. According to Cowling & Lombard (2002), climatic stability during the Pleistocene is a hypothetical driver of plant diversity patterns observed across the eastern and western CFR. Climate stability during the Pleistocene potentially reduced extinction rates in the western area of the CFR (Cowling & Lombard, 2002) and increased speciation rates, probably through persistence and occupancy of diversified niches (Valente & Vargas, 2013). In the Cerrado of Brazil (one of the phytogeographical domains where *campos rupestres* occur), diversification patterns are also influenced by climatic stability; high levels of endemism are associated with stable and central high-elevation areas, which coincide with patches of *campos rupestres* (Vidal *et al.*, 2019).

In addition to climatic stability, precipitation seasonality may also influence plant diversification patterns (Mucina & Wardell-Johnson, 2011; Eisenhardt *et al.*, 2015). Seasonal precipitation

during winter is a defining characteristic of the GCFR (Born et al., 2007) and has probably contributed to greater species diversity by favouring non-sprouter plants (dominant in the CFR) (Cowling & Lombard, 2002). Other environmental factors, such as fire, are related to patterns of tree species distribution in the Atlantic forest of Brazil; patches of campos rupestres occurring in this phytogeographical domain are very prone to fire, which limits tree growth (Neves et al., 2017). Phytophysiognomic variation is another factor influencing plant diversity patterns in the campos rupestres with open vegetation areas presenting a pattern of phylogenetic clustering (Zappi et al., 2017).

The goal of the present study was to test a component of the second prediction of OCBIL theory, that OCBILs will be characterized by the persistence of old lineages. Geographically, we focused on Australia, sub-Saharan Africa and South America (SWAFR, the GCFR and the *campos rupestres*, respectively); these areas have been included in multiple studies testing OCBIL predictions (see Hopper et al., 2016, for a compilation of examples). We characterized these areas based on present and past environmental and climatic conditions to test whether ancient lineages are more common in regions identified as OCBILs compared to surrounding geographical regions. Unlike previous studies, which focused on developing OCBIL theory (e.g. Hopper et al., 2016), we provide for the first time a global quantitative framework to test the second hypothesis of OCBIL theory. We predict that an association between persistent seed plant lineages and OCBIL landscapes will be supported if environmental characteristics of OCBILs are significantly and consistently correlated with older community ages. Conversely, if these environmental characteristics are not significantly correlated with older community ages or results are dependent on the area, persistent seed plant lineages are not more likely to occur in OCBILs than in other landscapes.

MATERIAL AND METHODS

DESCRIPTION OF OCBIL REGIONS: SWAFR, GCFR AND CAMPOS RUPESTRES

The SWAFR and GCFR have served as background for numerous investigations conducted within the scope of OCBIL theory (Hopper *et al.*, 2016), and the *campos rupestres* has recently drawn attention from researchers interested in testing this theory's predictions (Silveira *et al.*, 2016; Morellato & Silveira, 2018). Moreover, all three areas have been classified, along with other regions, as 'pre-Pleistocene climatically stable landscapes' by Mucina & Wardell-Johnson (2011), in a critical appraisal of OCBIL theory.

Apart from exhibiting similar environmental characteristics, all three regions are known for their high species diversity with many endemics distributed in relatively small areas (Hilton-Taylor, 1996; Goldblatt et al., 2005; Gioia & Hopper, 2017). The unusually high levels of diversity and endemism found in these three regions are thought to be related to the occurrence of many microhabitats (Valente & Vargas, 2013; Marques et al., 2014). The floras of the SWAFR, GCFR and campos rupestres are all dominated by shrubs and grasses and unusual, species-rich families, such as Proteaceae (SWAFR and GCFR), Restionaceae (GCFR) and Velloziaceae (campos rupestres) (Born et al., 2007; Cardillo & Pratt, 2013; Silveira et al., 2016).

Among these three regions, the SWAFR and GCFR are more similar to each other than to the *campos rupestres* in terms of both floristic composition and climate, with both being considered Mediterraneantype ecosystems (Mucina & Wardell-Johnson, 2011; Linder, 2014). However, topography is less variable in the SWAFR than the GCFR (Cardillo & Pratt, 2013). The *campos rupestres* resembles the GCFR more than it does the SWAFR, with similar altitudinal ranges and a few shared species-rich plant lineages, such as Velloziaceae (Alcantara *et al.*, 2018).

The SWAFR occupies 300 000 km² and contains more than 8300 plant species, nearly half of which are endemic (Gioia & Hopper, 2017). It is the largest in area of these three OCBILs, yet not the most speciesrich. The GCFR comprises around 14 000 plant species; 9000 occur in the CFR (70% endemic) and nearly 5000 species occur in the Succulent Karoo (40% endemic) (Hilton-Taylor, 1996; Goldblatt et al., 2005). The GCFR occupies 190 250 km²; the CFR comprises 90 000 km² and the Succulent Karoo 100 250 km² (Hilton-Taylor, 1996; Goldblatt et al., 2005). According to Valente & Vargas (2013), the CFR is so disproportionately diverse that it contains 'double the number of species predicted by global models of biodiversity'. Lastly, the campos rupestres occupies ~65 000 km² distributed across disjunct mountain chains in central, southeastern and north-eastern Brazil, and is home to more than 5000 plant species (Silveira et al., 2016). The campos rupestres is the only region among the three that exhibits this disjunct pattern, and it is spread across a wider latitudinal range than the SWAFR and GCFR.

RETRIEVING PHYLOGENETIC, SPATIAL AND ENVIRONMENTAL DATA

We used the dated seed plant phylogeny of Smith & Brown (2018) as an estimate of phylogenetic relationships. This phylogeny contains over 350 000 seed plant species and was assembled based on the

Open Tree of Life (Hinchliff *et al.*, 2015); backbone relationships were inferred based on DNA sequence data from GenBank, and taxa for which no such data were available were added based on taxonomic placements (version 'ALLOTB'; see Smith & Brown, 2018).

Occurrence data were retrieved from herbarium specimen records in iDigBio (https://www.idigbio. org/portal, 2020) and GBIF (GBIF.org, 2020; https:// www.gbif.org/), the two largest online databases for georeferenced occurrence data (see Supporting Information 1 for detailed references of occurrence points). To define the study area, we used bounding boxes to delimit the biogeographical regions to be included in the analyses, and occurrence records were downloaded from within these bounding boxes. To ensure adequate sampling of occurrence records from both OCBIL and non-OCBIL regions, we selected all of Australia and all of South America for analyses of OCBILs on those continents. Delimitation of the GCFR was less straightforward given the greater continuity of African geography and biotas with adjacent continents, so we considered two alternatives: sub-equatorial Africa and sub-Saharan Africa. Subequatorial Africa is large enough to include the OCBIL of interest as well as adjacent non-OCBIL regions; however, this designation cuts through biological communities and does not reflect a biogeographically relevant area. We therefore selected the larger, but more biologically meaningful region of sub-Saharan Africa. Because the dimensions of the bounding box determine the species that will be sampled, and therefore affect the results, we performed a secondary more focused analysis using sub-equatorial Africa for comparison (see Supporting Information 2, Figures S1, S2). We defined WGS84 bounding boxes [min x, min y, max x, max y – values reflecting latitude (x) and longitude (y)] corresponding to the entire continent of South America (-82, -56, -32, 15), Australia (112, -44, 154, -10) and sub-Saharan Africa (-18, -36, 52, 15). We performed data cleaning to ensure quality of records, a crucial step when dealing with large datasets (Sosef et al., 2017). Cleaning scripts (available at https:// github.com/lifemapper/lmpy/) ensured occurrence data were retained only for accepted species (324 009) in the phylogeny of Smith & Brown (2018) occurring within the limits of the bounding boxes described above with further filtering steps as follows. First, occurrence points were required to include at least four decimal places. For a species to be included, it had to be represented at 12 or more distinct localities. These requirements decreased the number of species to 171 377 (~53%). Then, we removed points with GBIF or iDigBio quality flags (invalid, mismatched and suspect taxonomy, datum missing and errors, geocode errors, points at the origin), reducing the number to 171 168 species. Duplicate collections of a

plant specimen are commonly made and distributed to multiple herbaria, resulting in the same locality being counted redundantly in data aggregators. Duplication can also derive from overlap between our two main occurrence sources, iDigBio and GBIF, as they contain many of the same records. Thus, we removed duplicate localities and constrained occurrence data to the limits of each bounding box described above.

We used species range data from the Plants of the World Online (POWO, 2019; http://powo.science. kew.org/), accessed via Kew Python APIs, to retain only localities thought to represent the native range of the study species. To do this, we intersected the occurrence data retrieved from iDigBio and GBIF with regions specified as level 3 geographical descriptors from the World Geographical Schema for Recording Plant Distributions (WGSRPD; https://www.tdwg.org/ standards/wgsrpd/). These regions define localities where the taxa are found based on expert opinion; POWO also distinguishes between native and anthropogenic areas of species ranges. Therefore, if a locality point for a given taxon was situated outside of its expert-assessed geographical range, we removed it from our records. We included in the analysis all occurrence points from taxa for which there was no information available on POWO on expert-assessed geographical range. After this step, 44 482 species (represented by 4 166 988 occurrence points) remained for South America, 7033 species (represented by 1 410 804 occurrence points) for Australia and 3990 species (represented by 135 873 occurrence points) for sub-Saharan Africa. Before filtering steps there were 6 955 497 occurrence points for 69 981 species for South America, 6 588 920 occurrence points for 24 135 species for Australia and 2 906 103 occurrence points for 51 963 species for sub-Saharan Africa.

Contemporary environmental data were retrieved from three different databases: WorldClim (Fick & Hijmans, 2017; https://worldclim.org/data/bioclim. html), Soil Grids (https://www.isric.org/explore/ soilgrids) and Global 30 Arc-Second Elevation (GTOPO30) (https://www.usgs.gov/centers/eros/ science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science_center_ objects=0#qt-science_center_objects).From WorldClim, we used Bioclim 1 (annual mean temperature), Bioclim 7 (temperature annual range), Bioclim 12 (annual precipitation) and Bioclim 17 (precipitation of driest quarter). From Soil Grids, we used sand content, soil organic carbon content, soil pH in water and coarse fragment volumetric percent. From GTOPO30 (https:// doi.org/10.5066/F7DF6PQS), we used *elevation*.

We calculated a novel metric representing relative climatic stability from the mid-Pliocene to the present. From Paleoclim (Dolan *et al.*, 2015; Hill, 2015; Karger

et al., 2017; Brown et al., 2018; http://www.paleoclim.org/), we obtained six time points [specifically the models Current (0 kya), Last Glacial Maximum (0.021 Mya), Last Interglacial (0.13 Mya), MIS19 (0.787 Mya), mid-Pliocene Warm Period (3.205 Mya) and M2 (3.3 Mya)] and then calculated pairwise distances between all pairs of the six different time frames, performing these calculations separately for all available temperature and precipitation data. The resulting environmental datasets therefore are a measure of climatic stability via temporal variation in temperature and precipitation.

INTEGRATING SPATIAL, PHYLOGENETIC AND ENVIRONMENTAL DATA

Only species with locality records that satisfied the filters described above were used to produce species distribution models (SDMs) in Maxent (Phillips et al., 2017) using the Lifemapper platform (Beach et al., 2020). SDMs expand the use of locality data by providing broader geographical distributions that help alleviate incomplete sampling biases (Barthlott et al., 2007). Each of the three study areas was organized into a grid of 0.5° square cells, with each cell representing a local plant community (although we recognize that a cell may actually comprise multiple biological communities). SDMs for species occurring at the community level were overlaid to produce a presence-absence matrix (PAM). We required at least 10% coverage of a grid cell with non-zero model probability as the criterion of presence of a given species in a grid cell.

For each grid cell, a community subtree was trimmed from the phylogeny of Smith & Brown (2018) to calculate phylogenetic metrics. The 'age' of a plant community can be interpreted in multiple ways, such as the age of extant species or of higher level clades. To attempt to capture both of these concepts, we defined two novel community metrics, tip age, corresponding to the age of the immediate parent node of each tip, and node height, corresponding to the age of every internal node in the phylogeny. As defined, each of these returns a vector of ages for each grid cell, that is a list of tip or node ages, which we summarized by calculating a series of summary statistics, including four different percentiles (2.5, 25, 75 and 97.5%), as well as means and medians. Unless otherwise noted, we focus discussion hereafter on mean values. Overall, higher summary statistics capture older community ages, while the lower percentiles capture younger age estimates. These metrics are similar to measures used in previous investigations. Mean node height is similar to the age metric in a recent investigation of grid cell ages of angiosperms in China (Lu et al.,

2018), which was specifically defined for a phylogeny collapsed to higher level clades with species richness data, while mean tip age is similar to the age metric in a recent investigation of diversification in rosids (Sun *et al.*, 2020).

Finally, we combined the grid cell age metrics with the environmental variables described above and exported metrics as a single matrix for each bounding box. The analyses described above were conducted using the workflow available through Lifemapper lmpy, at https://github.com/lifemapper/lmpy/, https://github.com/biotaphy/BiotaPhyPy/ and https://github.com/biotaphy/projects/.

REGRESSION ANALYSES

We performed univariate and multiple linear regressions using community age as a response variable (specifically, mean node height and mean tip length) and the set of environmental variables as predictors. For the multivariate analyses, we implemented model comparison using four different variable combinations: all variables, climatic stability variables ('temperature distance' and 'precipitation distance'), bioclimatic variables ('annual mean temperature', 'temperature annual range', 'annual precipitation' and 'precipitation of driest quarter'), and just soil variables ('coarse fragments', 'sand percent', 'soil pH' and 'organic carbon').

Before performing the linear regression analysis, we executed further cleaning steps to remove missing data values. After this final filtering step, we had at most 5778 grid cell values for sub-Saharan Africa, 2313 grid cells for Australia and 5189 grid cells for South America. Calculations and plots were conducted in R (R Core Team, 2013; https://www.r-project.org/). Maps were plotted in QGIS (QGIS.org, 2019) using publicly available shapefiles for South America (South America shapefile, 2015; http://tapiquen-sig.jimdo. com), Australia (Australia shapefile, 2020; https:// www.igismap.com/australia-shapefile-download/) and Africa (Africa shapefile, 2020; https://open. africa/dataset/africa-shapefiles). The shapefiles for the SWAFR, CFR (part of the GCFR) and campos rupestres were obtained from previously published studies (Mucina & Rutherford, 2006; Silveira et al., 2016; Gioia & Hopper, 2017).

RESULTS

UNDERSTANDING THE COMBINED EFFECT OF MULTIPLE ENVIRONMENTAL PREDICTORS

For South America, sub-Saharan Africa and Australia, the full multiple regression models explained substantial amounts of variation for both tip and node age metrics. The adjusted R^2 values for mean tip length were 0.510 for South America, 0.257 for sub-Saharan Africa and 0.644 for Australia. Similarly, adjusted R^2 values for mean node height were 0.591 for South America, 0.270 for sub-Saharan Africa and 0.603 for Australia. For South America and sub-Saharan Africa, the multivariate analysis including only soil variables as predictors yielded the second highest adjusted R^2 values: 0.429/0.522 and 0.203/0.210, respectively (values displayed as mean tip length/mean node height). For Australia, the second highest adjusted R^2 was observed for the analyses including only bioclimatic variables. 0.585/0.547. For all areas, the lowest adjusted R^2 values observed included the variables representing climatic stability (Table 1). Interestingly, mean tip length models generally captured more of the variance in the data. For example, the analysis for Australia including only soil variables produced an adjusted R^2 value of 0.416 when age was measured in terms of mean node height, but 0.542 when measured in terms of mean tip length.

UNDERSTANDING THE INDIVIDUAL EFFECT OF ENVIRONMENTAL PREDICTORS

We also executed a series of univariate linear regression analyses to understand age relationships with individual environmental predictors. In general, these models had low adjusted R^2 values, suggesting limited explanatory power for individual variables, but many relationships were significant. Moderately strong adjusted R^2 values were arbitrarily defined as values > 0.1. While some predictors were important across the three regions, the nature of the relationship was sometimes idiosyncratic. For example, soil pH was a positive and relatively strong predictor of age in South America (0.145/0.217), but a negative and also relatively strong predictor for Australia (0.247/0.135). Overall, univariate linear regression analysis exhibited no clear and consistent pattern across the three regions (Table 2).

Although individual predictors were inconsistently correlated with age in the three regions, results using the two age metrics were generally consistent within a single region. For example, in South America, elevation had a positive relationship with an adjusted R^2 value of 0.238 for mean tip length and 0.291 for mean node height (Table 2). Only two predictor variables produced conflicting results for the same area when comparing the two age metrics: annual mean temperature and coarse fragments. However, the differences were not significant. For Australia, for example, annual mean temperature was not a predictor of mean tip length

 $\textbf{Table 1.} \ \, \textbf{Adjusted} \, R^2 \, \textbf{values} \, \, \textbf{and} \, \, \textbf{Akaike information criterion} \, \, (\textbf{AIC}) \, \textbf{scores for each of the multiple linear regression} \, \\ \textbf{analyses conducted} \, \, \textbf{analyses conducted} \, \, \textbf{analyses} \, \, \textbf{conducted} \, \, \textbf{analyse} \, \, \textbf{conducted} \, \, \textbf{conduc$

Area	Mean node l	neight	Mean tip ler	Type of analysis		
	$\overline{R^2}$	AIC score	$\overline{R^2}$	AIC score		
South America	0.591	32 160.12	0.510	34 757.68	All variables	
	0.031	36 096.53	0.013	$38\ 362.52$	Climate stability	
	0.390	$33\ 986.2$	0.318	36 456.39	Bioclimatic	
	0.522	$32\ 861.94$	0.429	35 533.1	Soil	
Sub-Saharan Af-	0.270	$46\ 566.24$	0.257	478 98.78	All variables	
rica	0.002	48 334.37	0.008	49 536.6	Climate stability	
	0.116	47 646.55	0.092	49 036.14	Bioclimatic	
	0.210	46 997.87	0.203	48 284.31	Soil	
Australia	0.603	10 628.91	0.644	11507.67	All variables	
	0.054	12 293.26	0.154	13 479.63	Climate stability	
	0.547	10 868.31	0.585	11 842.75	Bioclimatic	
	0.416	11 364.24	0.542	$12\ 071.7$	Soil	

The all variables analyses included all the environmental predictors, climate stability analyses included temperature distance and precipitation distance, bioclimatic analyses included annual mean temperature, temperature annual range, annual precipitation and precipitation of driest quarter, and soil analyses included coarse fragments, sand percent, soil pH and organic carbon.

(adjusted $R^2 = 0$) but showed a significant, weak and slightly positive relationship (adjusted $R^2 = 0.027$) as a predictor of mean node height (Table 2).

Overall, South America and Australia had higher, more significant adjusted R^2 values compared with sub-Saharan Africa, suggesting that the predictors we assembled explain less of the community age variation in the latter area. For most analyses, the relationships observed for sub-Saharan Africa were close to zero; the analyses with the highest adjusted R^2 values and most significant relationships included the predictor variables: soil pH (0.037/0.031, positive association), elevation (0.014/0.024, positive association), annual precipitation (0.018, negative association for mean tip length) and annual mean temperature (0.020, negative association for mean node height) (Fig. 1 and Table 2). For South America, the analyses producing the highest adjusted R^2 values and most significant relationships included the predictor variables: coarse fragments (0.262/0.332, positive association), elevation (0.238/0.291, positive association) and annual mean temperature (0.166/0.249, negative association) (Fig. 2 and Table 2). For Australia, the highest adjusted R^2 values and most significant relationships were obtained for the predictor variables: annual precipitation (0.356/0.195, positive association), sand percent (0.317/0.154, negative association) and soil pH (0.247/0.135, negative association) (Fig. 3 and Table 2). Supporting Information 3 provides graphs corresponding to the remaining regressions (Figures S3-S50).

For Australia, the univariate analysis using mean tip length as the response variable showed consistently higher adjusted R^2 values, regardless of the predictor variable. For example, with precipitation of driest quarter as a predictor variable and mean node height as the response variable, an adjusted R^2 value of 0.026 was observed, while this value increased to 0.154 when the response variable was mean tip length (Table 2). For sub-Saharan Africa, the difference in adjusted R^2 values between analyses including mean tip length or mean node height was very small and followed no consistent trend. For South America, the analysis using mean tip length as a response variable, regardless of predictors, presented lower, but still significant, adjusted R^2 values.

VISUALIZING COMMUNITY AGE ON A MAP

To visualize areas corresponding to old communities in South America, sub-Saharan Africa and Australia, we plotted six different maps, three displaying statistics for mean node height and three for mean tip length. For South America and sub-Saharan Africa, both age metrics yielded similar spatial patterns. Old communities overlapped mostly with the Andes and a few areas in the Amazon in South America (Fig. 4). For sub-Saharan Africa, most of the oldest communities are concentrated in (1) western and southern coastal regions in the south, (2) western Zambia and central eastern/north-eastern Angola, (3) western Guinea, (4) south-eastern South Sudan, (5) northern Kenya and

Table 2. Results obtained with the univariate linear regression analyses are organized by area investigated (South America, sub-Saharan Africa, Australia) and community age metric (mean node height and mean tip length).

Area	South America					Sub-Saharan Africa					Australia							
Variable	Mean node height		Mean tip length		Mean node height		Mean tip length		Mean node height			Mean tip length						
Temperature distance	N	0.003	***	N	0.001		P	0.000		P	0.000		N	0.054	***	N	0.154	***
Precipitation distance	N	0.031	***	N	0.013	***	N	0.003	***	N	0.008	***	N	0.006	***	N	0.025	***
Annual mean temperature	N	0.249	***	N	0.166	***	N	0.020	***	N	0.006	***	P	0.027	***	N	0.000	
Temperature annual range	P	0.082	***	P	0.024	***	P	0.004	***	P	0.001	**	N	0.123	***	N	0.195	***
Annual precipitation	N	0.161	***	N	0.110	***	N	0.014	***	N	0.018	***	P	0.195	***	P	0.356	***
Precipitation of driest quarter	N	0.092	***	N	0.063	***	N	0.007	***	N	0.005	***	P	0.026	***	P	0.154	***
Elevation	P	0.291	***	P	0.238	***	P	0.024	***	P	0.014	***	P	0.000		P	0.000	
Coarse fragments	P	0.332	***	P	0.262	***	P	0.014	***	P	0.009	***	P	0.001		N	0.018	***
Sand percent	P	0.012	***	P	0.005	***	N	0.006	***	N	0.005	***	N	0.154	***	N	0.317	***
Soil pH	P	0.217	***	P	0.145	***	P	0.031	***	P	0.037	***	N	0.135	***	N	0.247	***
Organic carbon	P	0.000		P	0.000		P	0.001	**	P	0.001	**	P	0.052	***	P	0.157	***

N = negative and P = positive (direction) and numbers correspond to adjusted R^2 values. "", " and " indicate significance of P-values: *** (0), ** (0.001), * (0.1), " (0.1) and " (1).

(6) the Horn of Africa (Fig. 5). For Australia, tip- and node-based metrics gave contrasting results. The map based on mean node height showed old communities concentrated near the south-eastern coast and in some areas of the western desert, such as the Gibson Desert. For the map based on mean tip length statistics, old communities are mostly concentrated near the south-eastern coast, with fewer old communities scattered across the north of Australia, including the western desert (Fig. 6). We did not observe a significant overlap between old communities and either the *campos rupestres* or the SWAFR. A higher incidence of old communities was observed in the CFR compared to the other OCBIL areas (Fig. 7).

DISCUSSION

Our analyses suggest that plant community ages are influenced by small effects from many environmental predictors with no one predictor predominantly determining community composition (Table 1). A similar conclusion was reached for plant species in the Cerrado of Brazil, where a high incidence of endemism is correlated with central areas, high elevation and climate stability (Vidal *et al.*, 2019).

Different variables were predictors of community age in different areas, sometimes with opposing effects (Table 2). For example, annual precipitation and soil pH were significant predictors of community age for South America and Australia, but these relationships had opposite effects in each area. Whereas old communities seem to occur in dry areas with alkaline soils in South America, in Australia they tend to occur in humid areas and where soils are acidic. Interestingly, extremely acidic, and alkaline soils lead to soil infertility because pH conditions impact availability, absorption, and retention of nutrients and other materials in the soil (Merry, 2009).

In general, climate stability (as captured by temperature distance and precipitation distance) did not strongly predict community age, even when relationships were significant. The multiple linear regressions including temperature distance and precipitation distance had the lowest adjusted R^2 values among all multiple regression analyses. Even when these predictors were analysed individually, results were generally not significant. One exception was observed; temperature distance showed a relatively strong, negative relationship with community age for Australia (e.g. more climatic disturbance predicts lower mean tip length) (Table 2). Climatic stability

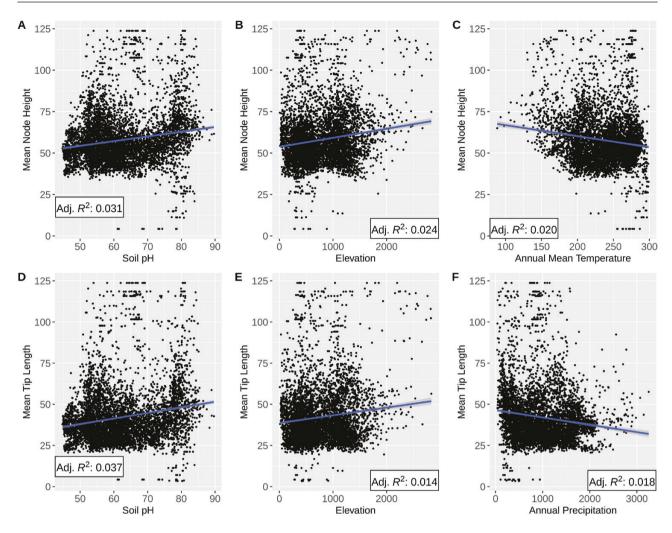


Figure 1. Univariate linear regression graphs for the four most significant variables for sub-Saharan Africa: soil pH (A, D), elevation (B, E), annual mean temperature (C) and annual precipitation (F). The upper row (A–C) and the lower row (D–F) show relationships between these environmental variables with mean node height and mean tip length, respectively.

has been considered an important driver of other diversification patterns across the globe, including endemism (Jansson, 2003), but our results suggest it is not universally associated with the occurrence of old plant communities. Nevertheless, generating high-resolution estimates of palaeoclimate is challenging, and alternative ways to calculate the influence of climatic stability could have yielded different results (Eiserhardt *et al.*, 2015).

For South America, five environmental variables significantly and relatively strongly predicted community age: annual precipitation, soil pH, annual mean temperature, elevation and coarse fragments, with the last three showing the strongest relationships (Fig. 2 and Table 2). This means that old communities in South America occur in dry, cool and elevated areas, where the soil is alkaline and contains a high volume of coarse fragments. Surprisingly, in our analyses most

of the old communities in South America overlap with the Andes (Fig. 4). The orogenic history of the Andean Plateau has either been understood as a recent and rapid event, with some areas undergoing uplift of approximately half of their height over the last 10 Myr (Gregory-Wodzicki, 2000), or as a more gradual uplift, starting around 25 Mya (Ehlers & Poulsen, 2009). Under either of these models, the Andes are geologically young compared to the *campos rupestres*; most of the *campos rupestres* mountain chains date from the Precambrian and are over 540 Myr old (Schaefer *et al.*, 2016; Walker *et al.*, 2018).

Although surprising, our finding that the oldest communities overlapped with the Andes may relate to the heterogeneous geological formations across this mountain range. The Southern and Central Andes probably formed earlier than the Northern Andes, and mid-elevation areas probably harbour older lineages than

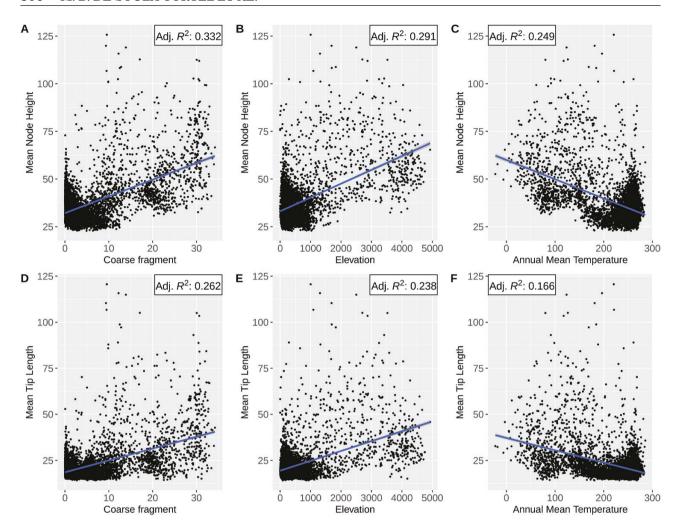


Figure 2. Univariate linear regression graphs for the three most significant variables for South America: coarse fragment (A, D), elevation (B, E) and annual mean temperature (C, F). The upper row (A-C) and the lower row (D-F) show relationships between these environmental variables with mean node height and mean tip length, respectively.

high-altitude areas (Luebert & Weigend, 2014). A finer scale grid-cell resolution would be required to test if old communities are prevalent at mid-elevations, but most of the old communities did overlap with the Southern and Central Andes rather than with the Northern region, consistent with this explanation. The occurrence of old seed plant communities does not appear to be strongly associated with geologically ancient landscapes in South America, but they do tend to occur in rocky soils as measured by coarse fragment volumetric percent. Such soils are resistant to erosion and may favour low moisture accumulation, which can lead to soil infertility (Darmody et al., 2004; Rodrigue & Burger, 2004). Hence, soil composition may be a major feature driving the locations of old communities in South America.

For Australia, four environmental predictors had significant and relatively strong relationships with community age, in terms of both mean node height and mean tip length: annual precipitation, sand percent, soil pH and temperature annual range (Fig. 3 and Table 2). Three other variables were significant predictors of community age when measured in terms of mean tip length: organic carbon, temperature distance and precipitation of driest quarter. When considering only the four strong predictors of both age metrics, old communities in Australia tend to occur in humid, nonsandy, acidic and cooler areas. When incorporating the three predictors of community age measured in terms of mean tip length, old communities in Australia also tend to occur in areas where organic carbon is high, and temperature has been stable through time. Australia's map based on mean node height indicates that old communities overlap both with mesic areas on the eastern coast, as well as with desert areas in western Australia (Fig. 6). However, when the map is plotted based on mean tip length, old communities overlap predominantly with mesic areas on the eastern coast, with only a few in the arid western region (Fig. 6).

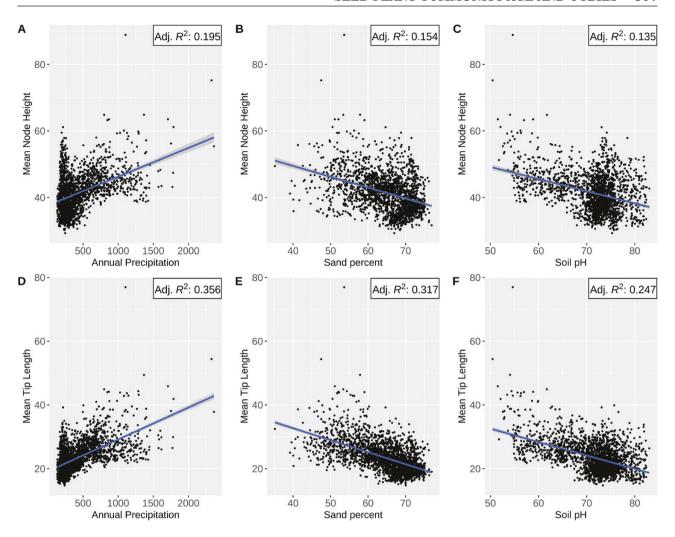


Figure 3. Univariate linear regression graphs for the three most significant variables for Australia: annual precipitation (A, D), sand percent (B, E) and soil pH (C, F). The upper row (A–C) and the lower row (D–F) show relationships between these environmental variables with mean node height and mean tip length, respectively.

This difference between the east and west of Australia indicates crown diversification in mesic areas is probably older than in desert areas in Australia. This idea is supported by suggestions that the mesic biome is ancestral to the semi-arid regions and deserts in Australia (Byrne et al., 2011). In fact, evidence from drilled cores suggests that aridification began only around 2 Mya in stony desert areas and intensified around 1 Mya in the Simpson Desert, when dune formation initiated (Fujioka et al., 2009). Many older lineages probably went extinct during aridification, which would have resulted in younger extant lineages in desert areas, as recovered in our results. Moreover, relative phylogenetic diversity (RPD) calculated for 90% of angiosperm genera across Australia revealed high values for the Wet Tropics, New England, Tasmania and the Australian Alps, indicating the presence of long branches for communities in these

areas (Thornhill *et al.*, 2016). When plotting our map over the 89 biogeographical regions of Australia (IBRA version 7, 2012), we recovered a similar pattern, with all of the areas cited above also showing a high incidence of old communities. All of these areas are more mesic than desert areas occupying central and western Australia.

For Africa, most of the environmental variables were significant, yet weak, predictors of community age. Soil pH, elevation, annual mean temperature (for mean node height) and annual precipitation (for mean tip length) showed the highest adjusted R^2 values among the predictors tested (Fig. 1 and Table 2). These results indicate that old communities in Africa occur where soils are acidic, elevation is high, and annual mean temperature and precipitation are low. However, these results should be interpreted cautiously given the relatively weak relationships observed.

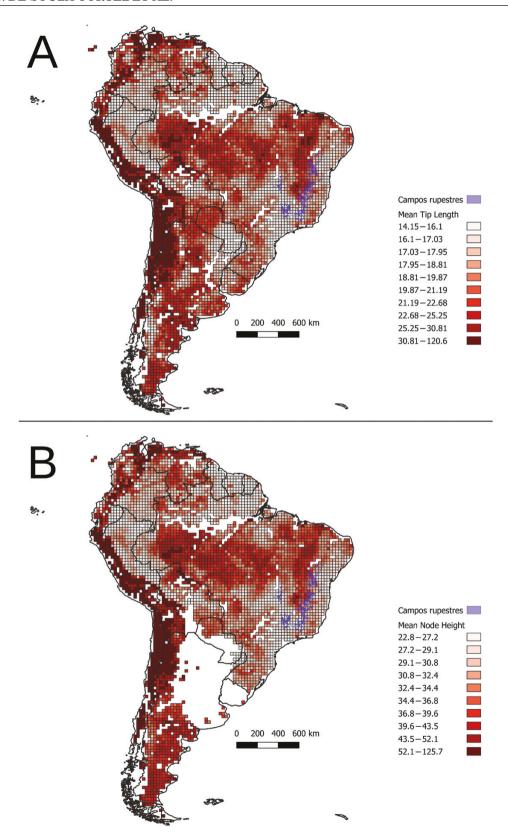


Figure 4. Maps of South America colour coded by community age in terms of mean tip length (A) and mean node height (B). Communities are represented by grid cells. Older communities are coloured in darker tones of red. The *campos rupestres* are coloured in blue. Age is measured in millions of years.

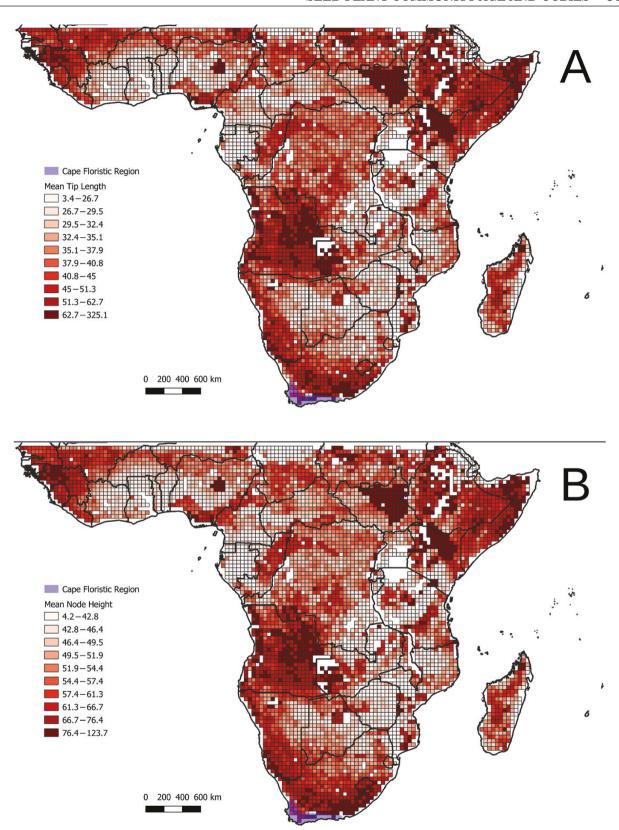


Figure 5. Maps of sub-Saharan Africa colour coded by community age in terms of mean tip length (A) and mean node height (B). Communities are represented by grid cells. Older communities are coloured in darker tones of red. The Cape Floristic Region (CFR) is coloured in blue. Age is measured in millions of years.

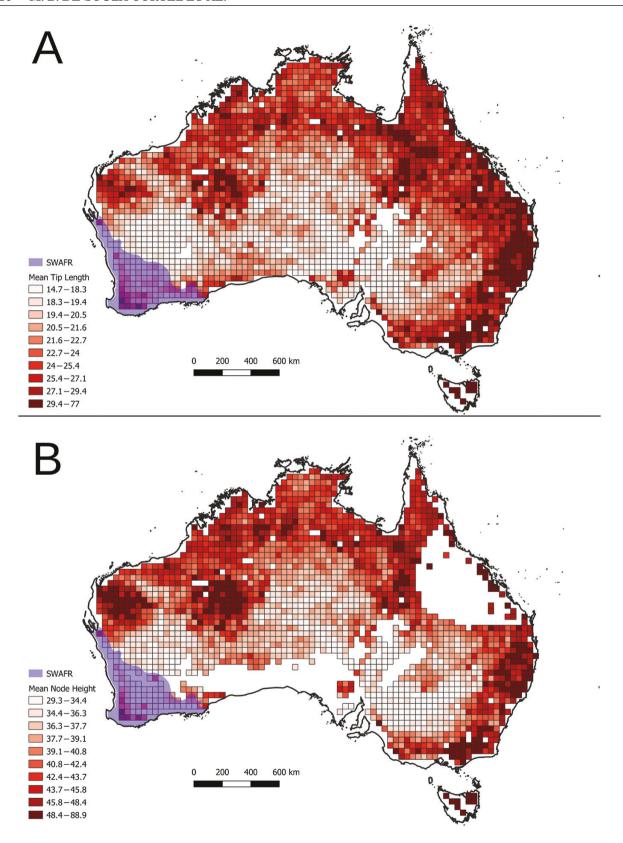


Figure 6. Maps of Australia colour coded by community age in terms of mean tip length (A) and mean node height (B). Communities are represented by grid cells. Older communities are coloured in darker tones of red. The Southwestern Australian Floristic Region (SWAFR) is coloured in blue. Age is measured in millions of years.

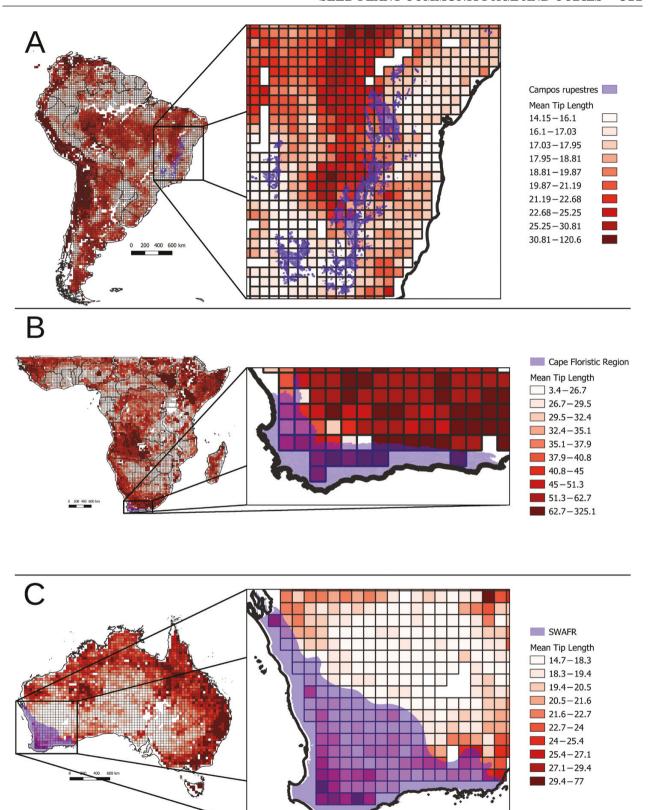


Figure 7. The OCBIL regions investigated in detail colour coded in terms of mean tip length. From top to bottom: part of South America and the campos rupestres (in blue), part of sub-Saharan Africa and the Cape Floristic Region (CFR - in blue), and part of Australia and the Southwestern Australian Floristic Region (SWAFR - in blue). Age is measured in millions of years. © 2020 The Linnean Society of London, $Biological\ Journal\ of\ the\ Linnean\ Society,\ 2021,\ 133,\ 297-316$

Our analyses based on mean tip length and mean node height metrics indicate that old communities in sub-Saharan Africa mainly overlap with six different areas (Fig. 5). Southern sub-Saharan Africa (1) is framed by the Great Escarpment, an ~5000-km-long discontinuous assembly of mountains rich in biodiversity, extending from Angola to Mozambique (Clark et al., 2011). The area between northern Zambia and central Angola (2) is influenced by the Zambezi and Congo watershed and is characterized by miombos rich in biodiversity and endemism (Linder, 2001). Western Guinea (3) is affected by the West African monsoon and is dominated by rainforests (Couvreur et al., 2020). South-eastern South Sudan (4) is dominated by a tropical savanna climate, advancing into northern Kenya (5), where it mostly intermingles with a warm arid climate (Beck et al., 2018). The Horn of Africa (6) is dominated by a warm desert climate (Beck et al., 2018).

The areas described above exhibit variable environmental characteristics, reflected in the weak relationships observed for sub-Saharan Africa. Elevation, however, seems to play an important role; apart from occupying the Great Escarpment, some of the old communities identified through our analyses are also situated near or at the Ethiopian Dome and the Guinea Rise (Couvreur et al., 2020). In fact, mountains in tropical Africa have been found to harbour both recent and anciently diversified lineages, serving as both cradle and museum (Dagallier et al., 2020).

Another important 'museum' in tropical Africa is the Guineo-Congolian region; significantly high RPD values were identified across this region (Dagallier et al., 2020), reinforcing the idea that lowland forests in Africa harbour the oldest lineages in the continent (Linder, 2014). This finding agrees with the geological history of Africa; physiognomic evidence suggests that forests were formed before woodlands and savannas (Jacobs, 2004). Fossil records date to as early as the Palaeocene (65.5 Mya) for forests, Middle Eocene (46 Mya) for woodlands and Early Miocene for savannas (17–23 Mya) (Jacobs, 2004). Although some of the old communities in our analyses overlapped with western Guinea (part of the Guineo-Congolian region), the pattern observed is not as clear as that presented by Dagallier et al. (2020). This difference may be related to the different metrics employed in each study, but it is more likely to be a consequence of different sources for occurrence data.

Given the large-scale focus of our study, we used only GBIF and iDigBio as sources of occurrence records and not more specialized databases, such as RAINBIO (Dauby et al., 2016, for tropical Africa, used by Dagallier et al., 2020), SpeciesLink (for Brazil, SpeciesLink, 2020; http://www.splink.org.br/) and the Atlas of Living Australia (ALA) (for Australia, Atlas

of Living Australia, 2020; https://www.ala.org.au/). Although most records in ALA and SpeciesLink are also included in GBIF, the same is not true for RAINBIO (Dauby et al., 2016). Including these resources in future studies could improve data completeness; data deficiency is a problem often associated with studies investigating regional and global biodiversity patterns. Biased collections have also been suggested as another limiting factor; non-random sampling leads to poorly sampled areas that are usually filtered out due to insufficient data, leading to under-representation of these areas (Barthlott et al., 2007; Sosef et al., 2017). However, digital collections of natural history specimens are currently the best providers of large datasets (Gioia & Hopper, 2017), particularly for studies capturing regional and/or global patterns.

A more accurate geographical delimitation of OCBIL and non-OCBIL areas would enable more fine-scale geographical analyses at a global extent. Studies at local (Ball et al., 2020) and regional (Zappi et al., 2017) scales reveal patterns associated with fine-scale geographical differences. For example, old lineages in the *campos rupestres* are more prevalent in quartzitic soils than forested ironstones, consistent with the second OCBIL prediction (Zappi et al., 2017). Conversely, adopting a finer resolution can also hinder precision; artificially empty cells are more easily generated, leading to distortion in regression analyses (Linder, 2001), which have largely been used in this study. In this context, improvement both in sampling efforts and in mapping technologies (i.e. through remote sensing; Schut et al., 2014) will provide further opportunities to test OCBIL theory, allowing for a comparison between patterns revealed by both largeand small-scale studies, ideally more appropriate for comparing OCBIL and non-OCBIL areas.

Our study revealed that community age, regardless of how it is measured, is influenced by a combination of different environmental predictors in different areas of the world. No single predictor showed consistently meaningful results for the three areas investigated (South America, Australia and sub-Saharan Africa), and results for these regions were very different. Climatic stability predictors showed the least meaningful results overall. According to our analyses, old communities in South America are associated with dry, cool, elevated areas, where soil is alkaline and the volume of coarse fragments is high, with our maps indicating an overlap of these areas with the Andes. Old communities in Australia are associated with humid, cool areas, and with non-sandy, acidic soil, and analyses based on mean tip length also indicated a relationship with high organic carbon content and stable temperatures through time. Australia showed different patterns; mesic areas such as the Wet Tropics, New England, Tasmania and the Australian Alps

overlapped with old communities, and desert areas in western Australia overlapped with old communities only when the map was plotted in terms of mean node height. Our results for sub-Saharan Africa were not comparatively as significant as those obtained for South America and Australia, but elevation seems to influence the occurrence of old communities; our analyses indicated overlap with the Great Escarpment, the Guinea Rise and the Ethiopian Dome. Finally, although old communities appear not to be associated with OCBILs, some limitations associated with the global scope of this study have probably influenced our results. Thus, further continental-scale tests of OCBIL theory should focus on improved sampling (by including specialized databases) and use of finer grid cells to differentiate between OCBIL and surrounding non-OCBIL areas.

ACKNOWLEDGEMENTS

We thank the reviewers and editor for valuable comments and suggestions. We are also grateful for the institutions that contributed data to digital online repositories and for iDigBio and GBIF for compiling and distributing these data, without which this work would not have been possible. We thank Aimee Stewart for assistance with data compilation and coding, Jim Beach and Jose Fortes for helpful discussion, and everyone involved with the BiotaPhy project. This work was supported in part by United States National Science Foundation grants DBI-1458640 and DBI-1930007 to D.E.S., P.S.S. and J. A. B. Fortes; DBI-1547229 to P.S.S.; DBI-1930030 to S.A.S.; DBI-1930005 to J. H. Beach. The paper is a contribution to the Special Issue of the Biological Journal of the Linnean Society entitled OCBIL theory: a new science for old ecosystems.

REFERENCES

- Africa shapefile. 2020. Available at: https://open.africa/dataset/africa-shapefiles Accessed 14 September 2020.
- Alcantara S, Ree RH, Mello-Silva R. 2018. Accelerated diversification and functional trait evolution in Velloziaceae reveal new insights into the origin of the campos rupestres' exceptional floristic richness. *Annals of Botany* 122: 165–180.
- Allen JM, Germain-Aubrey CC, Barve N, Neubig KM, Majure LC, Laffan SW, Mishler BD, Owens HL, Smith SA, Whitten WM, Abbott JR, Soltis DE, Guralnick R, Soltis PS. 2019. Spatial phylogenetics of Florida vascular plants: the effects of calibration and uncertainty on diversity estimates. iScience 11: 57-70.
- Alves RJV. 1994. Morphological age determination and longevity in some Vellozia populations in Brazil. Folia Geobotanica & Phytotaxonomica 29: 55–59.

- Alves RJV, Silva NG, Fernandes Júnior AJ, Guimarães AR. 2013. Longevity of the Brazilian underground tree Jacaranda decurrens Cham. Annals of the Brazilian Academy of Sciences 85: 671-677.
- Atlas of Living Australia. 2020. Available at: http://www.ala.org.au. Accessed 20 September 2020.
- Australia shapefile. 2020. Available at: https://www.igismap.com/australia-shapefile-download/. Accessed 20 August 2020
- Ball JW, Robinson TP, Wardell-Johnson GW, Bovill J, Byrne M, Nevill PG. 2020. Fine-scale species distribution modelling and genotyping by sequencing to examine hybridization between two narrow endemic plant species. Scientific Reports 10: 1562.
- Barthlott W, Hostert A, Kier G, Kuper W, Kreft H, Mutke J, Rafiqpoor D, Sommer JH. 2007. Geographic patterns of vascular plant diversity at continental to global scales. *Erkunde* 61: 305–315.
- Beach J, Stewart A, Grady C. 2020. Lifemapper web services. Available at: http://lifemapper.org. Accessed 17 April 2020.
- Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. 2018. Data descriptor: present and future Koppen-Geiger climate classification maps at 1-km resolution. Scientific Data 5: 180214.
- Bell C, Kutschker A, Kalin M. 2012. Phylogeny and diversification of Valerianaceae (Dipsacales) in the southern Andes. Molecular Phylogenetics and Evolution 63: 724–737.
- Born J, Linder HP, Desmet P. 2007. The Greater Cape Floristic region. *Journal of Biogeography* 34: 147–162.
- Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Scientific Data 5: 18025.
- Byrne M, Steane DA, Joseph L, Yeates DK, Jordan GJ, Crayn D, Aplin K, Cantrill DJ, Cook LG, Crisp MD, Keogh JS, Melville J, Moritz C, Porch N, Sniderman JMK, Sunnucks P, Weston PH. 2011. Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography* 38: 1635–1656.
- Cardillo M, Pratt R. 2013. Evolution of a hotspot genus: geographic variation in speciation and extinction rates in Banksia (Proteaceae). *Evolutionary Biology* 13: 155.
- Carpenter RJ, McLoughlin S, Hill RS, McNamara K, Jordan GJ. 2014. Early evidence of xeromorphy in angiosperms: stomatal encryption in a new Eocene species of *Banksia* (Proteaceae) from Western Australia. *American Journal of Botany* 101: 1486–1497.
- Clark VR, Barker NP, Mucina L. 2011. The great escarpment of southern Africa: a new frontier for biodiversity exploration. Biodiversity Conservation 20: 2543–2561.
- Couvreur TLP, Dauby G, Blach-Overgaard A, Deblauwe V, Dessein S, Droissart V, Hardy OJ, Harris DJ, Janssens SB, Ley AC, Mackinder BA, Sonké B, Sosef MSM, Stévart T, Svenning JC, Wieringa JJ, Faye A, Missoup AD, Tolley KA, Nicolas V, Ntie S, Fluteau F, Robin C, Guillocheau F, Barboni D, Sepulchre P. 2020. Tectonics, climate and the diversification

- of the tropical African terrestrial flora and fauna. Biological Reviews 000-000.
- Cowling RM, Lombard AT. 2002. Heterogeneity, speciation/ extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic region. *Diversity and Distribution* 8: 163–179.
- Dagallier LPML, Janssens SB, Dauby G, Blach-Overgaard A, Mackinder BA, Droissart V, Svenning JC, Sosef MSM, Stévart T, Harris DJ, Sonké B, Wieringa JJ, Hardy OJ, Couvreur TLP. 2020. Cradles and museums of generic plant diversity across tropical Africa. New Phytologist 225: 2196–2213.
- Darmody RG, Thorn CE, Schlyter P, Dixon JC. 2004. Relationship of vegetation distribution to soil properties in Kärkevagge, Swedish Lapland. *Arctic, Antarctic, and Alpine Research* 36: 21–32.
- Dauby G, Zaiss R, Overgaard A-B, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC, Engledow H, Fadeur G, Figueira R, Gereau RE, Hardy OJ, Harris DJ, Heij J, Janssens S, Klomberg Y, Ley AC, Mackinder BA, Meerts P, van de Poel JL, Sonké B, Sosef MSM, Stévart T, Stoffelen P, Svenning J-C, Sepulchre P, van der Burgt X, Wieringa JJ, Couvreur TLP. 2016. RAINBIO: a megadatabase of tropical African vascular plants distributions. *PhytoKeys* 74: 1–18.
- Dolan AM, Haywood AM, Hunter SJ, Tindall JC, Dowsett HJ, Hill DJ, Pickering SJ. 2015. Modelling the enigmatic Late Pliocene Glacial Event—Marine Isotope stage M2. Global and Planetary Change 128: 47–60.
- **Dynesius M, Jansson R. 2000.** Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences USA* **97:** 9115–9120.
- Ehlers TA, Paulsen CJ. 2009. Influence of Andean uplift on climate and paleoaltimetry estimates. *Earth and Planetary Science Letters* 281: 238–248.
- Eiserhardt WL, Borchenius F, Sandel B, Kissling WD, Svenning JC. 2015. Late Cenozoic climate and the phylogenetic structure of regional conifer floras world-wide. *Global Ecology and Biogeography* 24: 1136–1148.
- **Fick SE**, **Hijmans RJ**. **2017**. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302–4315.
- Fujioka T, Chappell J, Fifield LK, Rhodes EJ. 2009. Australian desert dune fields initiated with Pliocene-Pleistocene global climatic shift. *Geology* 37: 51–54.
- **Furley PA**, **Metcalfe SE. 2007.** Dynamic changes in savanna and seasonally dry vegetation through time. *Progress in Physical Geography* **31:** 633–642.
- GBIF.org. 2020. GBIF Occurrence Download. Available at https://doi.org/10.15468/dl.gtgtt5. Accessed 13 April 2020.
- Gioia P, Hopper SD. 2017. A new phytogeographic map for the Southwest Australian Floristic Region after an exceptional decade of collection and discovery. *Botanical Journal of the Linnean Society* 184: 1–15.
- **Global 30 Arc-Second Elevation (GTOPO30)** Digital Object Identifier (DOI) number: /10.5066/F7DF6PQS.

- Goldblatt P, Manning JC, Snijman D. 2005. Cape plants: corrections and additions to the flora. *Bothalia* 35: 35-46
- **Gregory-Wodzicki KM. 2000.** Uplift history of the Central and Northern Andes: a review. *Geological Society of America Bulletin* **112:** 1091–1105.
- Hill DJ. 2015. The non-analogue nature of Pliocene temperature gradients. Earth and Planetary Science Letters 425: 232–241.
- Hilton-Taylor C. 1996. Patterns and characteristics of the flora of the Succulent Karoo Biome, southern Africa. In: van der Maesen LJE, van der Bergt XM, van Medenbach de Rooy JM, eds. *The biodiversity of African plants*. Dordrecht: Kluwer, 58–72.
- Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, Crandall KA, Deng J, Drew BT, Grazis R, Gude K, Hibbett DS, Katz LA, Laughinghouse HD, McTavish EJ, Midford PE, Owen CL, Ree RH, Rees JA, Soltis DE, Williams T, Cranston KA. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences USA 112: 12764–12769.
- Hopper SD. 2009. OCBIL theory: towards an integrated understanding of the evolution, ecology and conservation of biodiversity on old climatically buffered, infertile landscapes. *Plant Soil* 322: 49–86.
- Hopper SD, Silveira FAO, Fiedler PL. 2016. Biodiversity hotspots and Ocbil theory. Plant Soil 403: 167–216.
- http://www.idigbio.org/portal. 2020. Query: {"filtered": {"filter": {"and": [{"term": {"kingdom": "plantae"}}, {"exists": {"field": "geopoint"}}, {"exists": {"field": "scientificname"}}]}}}, 22049228 records, accessed on 2020-04-14T01:51:06.491029, contributed by 738 Recordsets.
- **IBRA version 7. 2012.** Map produced by ERIN. Canberra: Australian Government Department of the Environment and Energy.
- Jacobs BF. 2004. Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philosophical Transactions of the Royal Society B* 359: 1573–1583.
- Jansson R. 2003. Global patterns in endemism explained by past climatic change. Proceedings of the Royal Society B. 270: 583–590.
- Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2008. Plant systematics: a phylogenetic approach, 3rd edn. Sunderland: Sinauer Associates.
- Karger DN, Conrad O, Bohner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M.
 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 170122.
- Klak C, Reeves G, Hedderson T. 2014. Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature* 427: 63–65.
- Lagomarsino LP, Condamine FL, Antonelli A, Mulch A, Davis CC. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). New Phytologist 210: 1430–1442.

- **Linder HP. 2001.** Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography* **28:** 169–182.
- **Linder HP. 2014.** The evolution of African plant diversity. *Frontiers in Ecology and Evolution* **2:** 38.
- Lu L, Mao L, Yang T, Ye J, Liu B, Li H, Sun M, Miller JT, Mathews S, Hu H, Niu Y, Peng D, Chen Y, Smith SA, Chen M, Xiang K, Le C, Dang V, Lu A, Soltis PS, Soltis DS, Li J, Chen Z. 2018. Evolutionary history of the angiosperm flora of China. Nature 554: 234–238.
- Luebert F, Weingend M. 2014. Phylogenetic insights into Andean plant diversification. Frontiers in Ecology and Evolution 2: 27.
- Marques AR, Atman APF, Silveira FAO, Lemos-Filho JP. 2014. Are seed germination and ecological breadth associated? Testing the regeneration niche hypothesis with bromeliads in a heterogeneous neotropical montane vegetation. *Plant Ecology* 215: 517–529.
- Merry RH. 2009. Acidity and alkalinity of soils. In: Sabljic A, ed. *Environmental and ecological chemistry Vol. II*. Oxford: Eolss Publishers Co. Ltd, 115–132.
- Mishler BD, Guralnick R, Soltis PS, Smith SA, Soltis DE, Barve N, Allen JM, Laffan SW. 2020. Spatial phylogenetics of the North American Flora. *Journal of Systematics and Evolution* 58: 393–405.
- Morellato LPC, Silveira FAO. 2018. Plant life in campo rupestre: new lessons from an ancient biodiversity hotspot. *Flora* 238: 1–10.
- Mucina L, Rutherford MC (eds). 2006. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. Pretoria: South African National Biodiversity Institute.
- Mucina L, Wardell-Johnson GW. 2011. Landscape age and soil fertility, climatic stability, and fire regime predictability: beyond the OCBIL framework. *Plant Soil* 341: 1–23.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hot spots for conservation priorities. *Nature* 403: 853–858.
- Neves DM, Dexter KG, Pennington RT, Valente ASM, Bueno ML, Eisenlohr PV, Fontes MAL, Miranda PLS, Moreira SN, Rezende VL, Saiter FZ, Oliveira-Filho AT. 2017. Dissecting a biodiversity hotspot: the importance of environmentally marginal habitats in the Atlantic forest domain of South America. Diversity and Distributions 23: 898–909.
- Paton AJ, Brummitt N, Govaerts R, Harman K, Hinchcliffe S, Allkin B, Lughadha EN. 2008. Towards target 1 of the global strategy for plant conservation: a working list of all known plant species-progress and prospects. *Taxon* 57: 602–611.
- Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40: 887–893.
- **Pidwirny M. 2006.** "Introduction to the Oceans". In: Fundamentals of physical geography, 2nd edn. http://www.physicalgeography.net/fundamentals/80.html. Accessed 2 July 2020.

- POWO. 2019. Plants of the world online. Kew: Facilitated by the Royal Botanic Gardens. Available at: http://www.plantsoftheworldonline.org/. Accessed 17 April 2020.
- QGIS.org. 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: http://ggis.org. Accessed 16 June 2020.
- R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.
- Rodrigue JA, Burger JA. 2004. Forest soil productivity of mined land in the midwestern and eastern coalfield regions. Soil Science Society of America Journal 68: 833–844.
- Schaefer CEGR, Corrêa GR, Candido HG, Arruda DM, Nunes JA, Araujo RW, Rodrigues PMS, Filho EIF, Pereira AFS, Brandão PC, Neri AV. 2016. The physical environment of Rupestrian grasslands (Campos Rupestres) in Brazil: geological, geomorphological and pedological characteristics, and interplays. In: Fernandes GW, ed. Ecology and conservation of mountaintop grasslands in Brazil. Cham: Springer International Publishing, 15–53.
- Schut AGT, Wardell-Johnson GW, Yates CJ, Keppel G, Baran I, Franklin SE, Hopper SD, Van Niel KP, Mucina L, Byrne M. 2014. Rapid characterisation of vegetation structure to predict refugia and climate change impacts across a global biodiversity hotspot. PLoS One 9: e82778.
- Silveira FAO, Negreiros D, Barbosa NPU, Buisson E, Carmo FF, Carstensen DW, Conceição AA, Cornelissen TG, Echternacht L, Fernandes GW, Garcia QS, Guerra TJ, Jacobi CM, Lemos-Filho JP, Stradic SL, Morellato LPC, Neves FS, Oliveira RS, Schaefer CE, Viana PL, Lambers H. 2016. Ecology and evolution of plant diversity in the endangered campo rupestre: a neglected conservation priority. Plant and Soil 403: 129–152.
- Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany. 105: 302–314.
- Sosef MSM, Dauby G, Blach-Overgaard A, van der Burgt X, Catarino L, Damen T, Deblauwe V, Dessein S, Dransfield J, Droissart V, Duarte MC, Engledow H, Fadeur G, Figueira R, Gereau RE, Hardy OJ, Harris DJ, Heij J, Janssens S, Klomberg Y, Ley AC, Mackinder BA, Svenning JC, Sepulchre P, Zaiss R, Wieringa JJ, Couvrer TLP. 2017. Exploring the floristic diversity of tropical Africa. BMC Biology 15: 15.
- South America shapefile. 2015. Carlos Efraín Porto Tapiquén. Orogénesis Soluciones Geográficas. Porlamar, Venezuela 2015. Based on shapes from Environmental Systems Research Institute (ESRI). Free Distribution. Available at http://tapiquen-sig.jimdo.com. Accessed 1 July 2020.
- **SpeciesLink**. **2020**. Available at: http://www.splink.org.br/. Accessed 20 September 2020.
- Sun M, Folk RA, Gitzendanner MA, Soltis PS, Chen Z, Soltis DE, Guralnick RP. 2020. Recent accelerated diversification in rosids occurred outside the tropics. *Nature Communications* 11: 3333.

- Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen TP, Ackerly DD, Mishler BD. 2017. Spatial phylogenetics of the native California flora. BMC Biology 15: 96.
- Thornhill AH, Mishler BD, Knerr NJ, González-Orozco CE, Costion CM, Crayn DM, Laffan SW, Miller JT. 2016. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography* 43: 2085–2098.
- Valente LM, Vargas P. 2013. Contrasting evolutionary hypotheses between two Mediterranean-climate floristic hotspots: the Cape of southern Africa and the Mediterranean Basin. *Journal of Biogeography* 40: 2032–2046.
- Verboom GA, Archibald JK, Bakker FT, Bellstedt DU, Conrad F, Dreyer LL, Forest F, Galley C, Goldblatt P, Henning JF, Mummenhoff K, Linder HP, Muasya AM, Oberlander KC, Savolainen V, Snijman DA, van der Niet T, Nowell TL. 2008. Origin and diversification

- of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution* **51**: 44–53.
- Vidal JD Jr, Souza AP, Koch I. 2019. Impacts of landscape composition, marginality of distribution, soil fertility and climatic stability on the patterns of woody plant endemism in the Cerrado. Global Ecology and Biogeography 28: 904–916.
- Walker JD, Geissman JW, Bowring SA, Babcock LE, compilers. 2018. Geologic Time Scale v. 5.0. Denver: Geological Society of America.
- Werneck F. 2011. The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews* 30: 1630-1648.
- Zappi DC, Moro MF, Meagher TR, Nic Lughadha E. 2017. Plant biodiversity drivers in Brazilian Campos Rupestres: insights from phylogenetic structure. Frontiers in Plant Science 8: 2141.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Part 1

Detailed references for the data retrieved from GBIF and iDigBio

Part 2

Figure S1. Map of subequatorial Africa colour coded by community age in terms of mean tip length.

Figure S2. Map of subequatorial Africa colour coded by community age in terms of mean node height.

Part 3

Figures S3–S50. Graphs showing the relationship between community age (measured in terms of both mean tip length and mean node height) and the following environmental predictors: annual mean temperature, temperature annual range, annual precipitation, precipitation of driest quarter, coarse fragments, elevation, organic carbon, precipitation distance, sand percent, soil pH and temperature distance. Figures S3–S18 depict graphs for Africa, Figures S19–S34 for Australia and Figures S35–S50 for South America.