



# Richness pattern and phytogeography of the Cerrado herb–shrub flora and implications for conservation

Aryanne G. Amaral , Cássia B.R. Munhoz, Bruno M.T. Walter, Jesús Aguirre-Gutiérrez & Niels Raes 

## Keywords

Neotropical savannas; PAM clustering; Protected areas; Species distribution model; Species richness

## Nomenclature

Brazilian Flora 2020 in construction (<http://floradobrasil.jbrj.gov.br/>)

Received 6 December 2016

Accepted 3 April 2017

Co-ordinating Editor: Miquel De Cáceres

**Amaral, A.G.** (corresponding author, [amaral.aryanne@gmail.com](mailto:amaral.aryanne@gmail.com))<sup>1,2,3</sup>,

**Munhoz, C.B.R.** ([cbrmunhoz@gmail.com](mailto:cbrmunhoz@gmail.com))<sup>2</sup>,

**Walter, B.M.T.** ([bruno.walter@embrapa.br](mailto:bruno.walter@embrapa.br))<sup>4</sup>,

**Aguirre-Gutiérrez, J.**

([jesus.aguirregutierrez@naturalis.nl](mailto:jesus.aguirregutierrez@naturalis.nl))<sup>3,5</sup>,

**Raes, N.** ([niels.raes@naturalis.nl](mailto:niels.raes@naturalis.nl))<sup>3</sup>

<sup>1</sup>Fundação Brasileira para o Desenvolvimento Sustentável, Engenheiro Álvaro Niemeyer 76, CEP 22610-180, Rio de Janeiro, RJ, Brazil;

<sup>2</sup>Department of Botany, University of Brasília, CP 4457, CEP 70910-900, Brasília, DF, Brazil;

<sup>3</sup>Naturalis Biodiversity Center, Biodiversity Dynamics, Vondellaan 55, Postbus 9517, 2300 RA, Leiden, The Netherlands;

<sup>4</sup>Embrapa Recursos Genéticos e Biotecnologia, W5 Norte, Parque Estação Biológica, CEP 70770-917, Brasília, DF, Brazil;

<sup>5</sup>Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1090 GE, Amsterdam, The Netherlands

## Introduction

Neotropical savannas cover approximately 270 million hectares of the South American continent (Olson et al. 2001). The Cerrado is the largest Neotropical savanna, covering 23% of the Brazilian territory (Silva et al. 2006). It is a hotspot of diversity harbouring around 12 400 plant species with high levels of endemism (Mendonça et al. 2008). The Cerrado is characterized by a mosaic of savannas, grasslands and forests (Myers et al. 2000; Mendonça et al. 2008; Ribeiro & Walter 2008).

## Abstract

**Aims:** The herb–shrub flora has been widely neglected in science and conservation policy throughout the world, so that this biodiversity component remains largely unknown. The objective of this study was to elucidate the spatial patterns of species richness and phytogeographic regions of the Cerrado herb–shrub flora, and to estimate the percentage of areas with high species richness that is still covered by natural vegetation and is located in protected areas or priority areas for conservation.

**Location:** Central Brazil and surrounding territories.

**Methods:** To estimate the pattern of species richness we produced species distribution models (SDMs) for 5362 species using botanical records combined with 17 least correlated environmental variables. All SDMs were tested against a bias-corrected null model and the 5039 significant SDMs were stacked to generate the pattern of botanical richness. The resulting presence–absence matrix was subjected to a partitioning around medoids (PAM) cluster analysis to delineate phytogeographic regions.

**Results:** The pattern of species richness indicated that highest levels of richness are in the central–south and west parts of the Cerrado. The study recognized two floristically distinct clusters at the core of the Cerrado region, and seven additional regions that share floristic elements with the two core clusters. Many areas with high levels of modelled species richness are located outside protected and priority areas for conservation.

**Conclusions:** We recognized patterns of regional variation, reflected in the phytogeographic division. Furthermore, we show that the current protected areas in Brazil do not effectively protect high richness areas of the herb–shrub flora.

Compared to the tree flora, the Cerrado herb–shrub component, with approximately 10 600 species of herb and shrub is much more diverse (Mendonça et al. 2008). Despite its richness, the spatial distribution of richness in the herb–shrub layer is often neglected in studies that provide information for conservation strategies in the Cerrado.

The Cerrado phytogeographic mosaic is related to climatic gradients and spatial patterns of edaphic conditions, such as differences in soil texture, depth and drainage capacity (Furley & Ratter 1988). Besides climatic factors

and soil properties, fire frequency is thought to be an important factor that contributes to the composition and structure of Cerrado vegetation (Durigan & Ratter 2016), and even in the diversification of its flora (Simon et al. 2009). Ratter et al. (2003) proposed six distinct phytogeographic provinces within the Cerrado biome based on surveys of the woody flora at 376 sites (see Fig. 3). Distinctions in these floristic regions were explained by altitude and climate as major gradients, complemented by the influence of soil fertility.

Studies on species richness and phytogeographic patterns of the herb-shrub layer in the Cerrado and the environmental conditions that control the differences in community composition are rare. The few published studies have focused on floristic composition or structural responses of herb-shrub communities to environmental conditions at local scales (e.g. Munhoz et al. 2008). In terms of phytogeographic studies, there is a lack of knowledge about the distribution patterns of the herb-shrub component in relation to climatic and edaphic conditions.

The Cerrado flora has been threatened by deforestation and land-use change for cattle ranching and agriculture since 1970 (MMA 2010). Incentives for economic growth and agricultural projects in the region have resulted in biological invasions, wildfires, deforestation and the conversion to open pastures, with consequent loss of native flora (Durigan et al. 2007). Given the exceptionally high levels of diversity and on-going clearance for cattle ranching and agriculture, there is an urgent requirement to improve conservation policies for the Cerrado. Conservation policies should be based on scientific information on patterns of species distribution, community composition and species richness, to indicate which remaining areas with natural vegetation cover have high conservation value.

For large areas with no systematically gathered data on species occurrence, species distribution models (SDMs) provide a powerful tool to predict their occurrence (Phillips et al. 2009). SDMs identify correlations between known occurrences of species and aspects of abiotic conditions across landscapes of interest to define sets of conditions under which species are likely to occur (Araújo & Peterson 2012). The resulting distribution maps can be used to develop species richness maps, delineate phytogeographic regions (Zhang et al. 2012; Vollering et al. 2016) and guide identification of conservation priority areas (Zhang et al. 2012).

Using a large data set of herb-shrub species collection records and inventory data combined with spatial abiotic data through SDMs, we reached conclusions on patterns of species richness, phytogeographic regions and conservation priority areas for the herb-shrub flora of the Cerrado. Our main objectives were to: (1) identify areas with high herb-shrub species richness derived from SDMs; (2)

delineate phytogeographic regions for the Cerrado based on floristic similarities in the herb-shrub flora; and (3) combine the priority areas and protected area maps with our species distribution models to estimate the percentage of high richness areas that is protected and whether these areas are concentrated in priority regions.

## Methods

### Botanical collection data

To model the Cerrado herb-shrub species richness pattern, we extracted all botanical records for South America north of latitude 40° S from both the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org/>; downloaded 31 Mar 2014) and SpeciesLink (<http://slink.cria.org.br/>; downloaded 10 Apr 2014) databases. To this data set we added 3967 specimens records from our own inventories (Appendix S1) and records extracted from published literature on the Cerrado herb-shrub flora (Appendix S2). After removing duplicate records, records identified to genus level, and those without georeferenced locality data, the data set comprised 986 030 records.

To synonymize and standardize all botanical names we used the package 'flora' (v 0.2.2) in R (v 3.0.3; R Foundation for Statistical Computing, Vienna, AT). This package provides all accepted botanical names for synonyms using the Brazilian Flora (<http://floradobrasil.jbrj.gov.br/>, accessed Aug 2016). First, we obtained the suggested names for all genus and species epithet name combinations; we then discarded infraspecific and variety taxonomic levels. Second, using the suggested names we obtained a list of accepted botanical names for the synonyms. Subsequently, all accepted names were matched with our specimen records data set, resulting in 38 129 accepted botanical names.

We overlaid this data set with a polygon defining the Cerrado biome as identified by the Brazilian Institute of Geography and Statistics – IBGE; (<http://mapas.ibge.gov.br/interativos/arquivos/downloads>, accessed 10 Apr 2014). We obtained 15 062 accepted botanical names of species known to occur in the Cerrado based on botanical presence records. We extracted all records from South America north of latitude 40° S for all 15 062 species known to occur in the Cerrado to prevent modelling truncated niches (*sensu* Raes 2012). Finally, we filtered all records for species that represent herb, sub-shrub, shrub and vine growth forms in the Cerrado Flora Checklist (Mendonça et al. 2008), complemented with species with herb-shrub growth form derived from the R package 'flora'. The data set comprised 9862 accepted herb-shrub species names known to occur in the Cerrado.

We consider the inference of any relationship between presence records and abiotic conditions with less than five

records as unrealistic (Pearson et al. 2007; van Proosdij et al. 2016). Therefore, all the records of species represented by less than five unique collection records at 5 arc-min spatial resolution ( $\sim 9.3 \text{ km} \times 9.3 \text{ km}$  at the equator) were removed. We deem this spatial resolution realistic to match the accuracy at which collection records can be georeferenced. The final data set comprised records of 5362 species, represented by 278 156 unique presence records at 5 arc-min spatial resolution.

### Environmental predictors

To model the herb-shrub species distributions we initially selected 41 environmental predictors. These included 19 bioclimatic predictors plus altitude from the WorldClim database at 5 arc-min spatial resolution (Hijmans et al. 2005). We added potential evapotranspiration estimated as the mean annual temperature ( $^{\circ}\text{C}$ ) divided by total annual precipitation (mm), and multiplied by an empirically derived constant of 60 (Holdridge et al. 1971; Loiselle et al. 2008). Additionally, we selected 20 quantitative soil variables from the ISRIC – World Soil Information database at 5 arc-min spatial resolution (v 1.2, <http://www.isric.org>). All environmental predictors were cropped to the extent of the study area. Finally, we added a mask layer, with a value of one for raster cells where a botanical collection was made based on our total collections database. The mask layer represents the target group background sample used in Maxent (see below; Phillips et al. 2009) and effectively omits all raster cells with no botanical collections. To reduce dimensionality and to avoid issues with multicollinearity between environmental predictors (Dormann et al. 2013), we used Spearman's rank correlation to select the least correlated predictors ( $|\text{Spearman's } \rho| < 0.7$ ; Appendix S3). The bioclimatic and soil predictors were analysed separately. In total, 17 of the 41 predictors were retained as environmental predictors to model the species distributions (Appendix S3).

### Species distribution modelling

To model the species distributions we selected Maxent, v 3.3.3k (Phillips et al. 2006). Maxent was designed to model species distributions with presence-only data, performs well with few presence records (Wisz et al. 2008) and outperforms most other distribution modelling algorithms (Elith et al. 2006; Aguirre-Gutiérrez et al. 2013). We used the Maxent default settings, except the modelling rules were set to use linear and quadratic features only, following Merow et al. (2013). The Maxent models were trained with the masked environmental predictor data set and then projected onto a data set covering the entire study area.

As a measure of model accuracy we used the area under the curve (AUC) value of the receiver operating characteristic (ROC) plot produced by Maxent. All measures of SDM accuracy require absences. When these are lacking, they are replaced by pseudo-absences, or a background sample in the case of Maxent; a large background sample is required to represent the environmental variation of the variables (Phillips et al. 2006; Elith et al. 2011). Background samples, however, are drawn from the species presence area proportional to the species true distribution, which is not known and the reason why SDMs are used. This results in a maximum AUC value of  $1 - a/2$ , where  $a$  represents the species true, but unknown, distribution area. Hence, standard AUC values of  $>0.8$  indicating a good model do not apply (Phillips et al. 2006). Therefore, we used bias-corrected null models (Raes & ter Steege 2007; van Proosdij et al. 2016) to test the AUC value of an SDM developed with all presence records against the distribution of AUC values expected by chance alone. The AUC value of an SDM developed with  $n$  records was tested against the upper AUC value of the lower 95% quantile of AUC values derived from  $100 \times n$  randomly drawn and modelled points. An SDM with an AUC value  $>95\%$  lower quantile AUC value of its null distribution performs significantly better than random expectation ( $P < 0.05$ ). Only significant SDMs were retained for further analyses (Appendix S4).

### Botanical richness pattern and delineation of phytogeographic regions

To generate the pattern of botanical richness of the herb-shrub layer of the Cerrado we converted the continuous Maxent predictions of all significant SDMs to discrete presence/absence values using the '10 percentile training presence threshold'. This threshold is preferred when applied to presence-only data because it does not take absences into account (Merow et al. 2013). Species richness was calculated by summing the number of species with predicted presence in each grid cell at 5 arc-min spatial resolution. High species richness was defined as raster cells where more than 1200 species were predicted to co-occur.

Based on the predicted presence/absence herb-shrub community composition within each cell, we delineated phytogeographic regions for the area covered by the Cerrado region as defined by IBGE plus a buffer of  $5^{\circ}$  ( $2.7\text{--}29.7\text{S}$ ,  $31.5\text{--}65.1^{\circ}\text{W}$ ). Second, we converted the community presence/absence matrix into a community distance matrix using the Jaccard distance (Legendre & De Cáceres 2013). Third, we used the partitioning around medoids (PAM) clustering algorithm to delineate different regions (Kaufman & Rousseeuw 1987). The PAM algorithm uses a technique that is based on the search

for  $k$  representative objects, called medoids, of the data set. After finding a set of  $k$  medoids,  $k$  clusters are constructed by assigning each observation to the nearest medoid. The goal is to find  $k$  representative objects that minimize the sum of dissimilarities of the observations to their closest representative object (Borcard et al. 2011). One limitation of using PAM is that it has no mechanism to detect the optimal number of cluster groups in a data set (Kreft & Jetz 2010). To define the optimal numbers of clusters (values of  $k$  ranging from 2 to 25) we used the Simple Structure Index (SSI). The optimal number of cluster groups has the maximum SSI value (Borcard et al. 2011). These cluster groups were plotted in geographic space using ArcGIS (v 10.1), resulting in a map delineating the phytogeographic regions based on the herb-shrub floristic composition in each grid cell. We overlaid this map with the floristic regions proposed by Ratter et al. (2003) for the woody vegetation of the Cerrado, in order to compare the patterns presented by both studies. To characterize the floristic differences between the cluster groups we used Indicator Species Analysis (Dufrene & Legendre 1997). To calculate the indicator index proposed by Dufrene & Legendre (1997) we used the R package 'indicspecies' (v 1.7.2). Finally, to infer the floristic similarities and relationships between cluster groups we applied an unweighted pair group method with arithmetic mean (UPGMA) clustering of the medoids using the R library 'Cluster' (v 2.0.5) resulting in a dendrogram of floristic similarity between the different phytogeographic regions (Vollering et al. 2016).

### Protected areas and priority areas for conservation

We obtained the current protected areas data set published for Brazil from the World Database on Protected Areas (<http://www.protectedplanet.net>, accessed 20 Nov 2014). Overlaying the species richness map with the protected areas map allowed to assess the percentage of high richness areas that is protected. A recent map showing the priority areas for conservation within the Cerrado provided by the Brazilian Environment Ministry (<http://www.mma.gov.br/biodiversidade/biodiversidadebrasileira/%C3%A1reaspriorit%C3%A1rias/item/489>, accessed 11 Aug 2015) was used to assess whether areas with high-predicted species richness were concentrated in these priority regions.

## Results

### Species distribution models

From the 5362 modelled species, 5039 (94%) showed significant non-random distributions and were retained for further analyses. This data set represents 174 families that

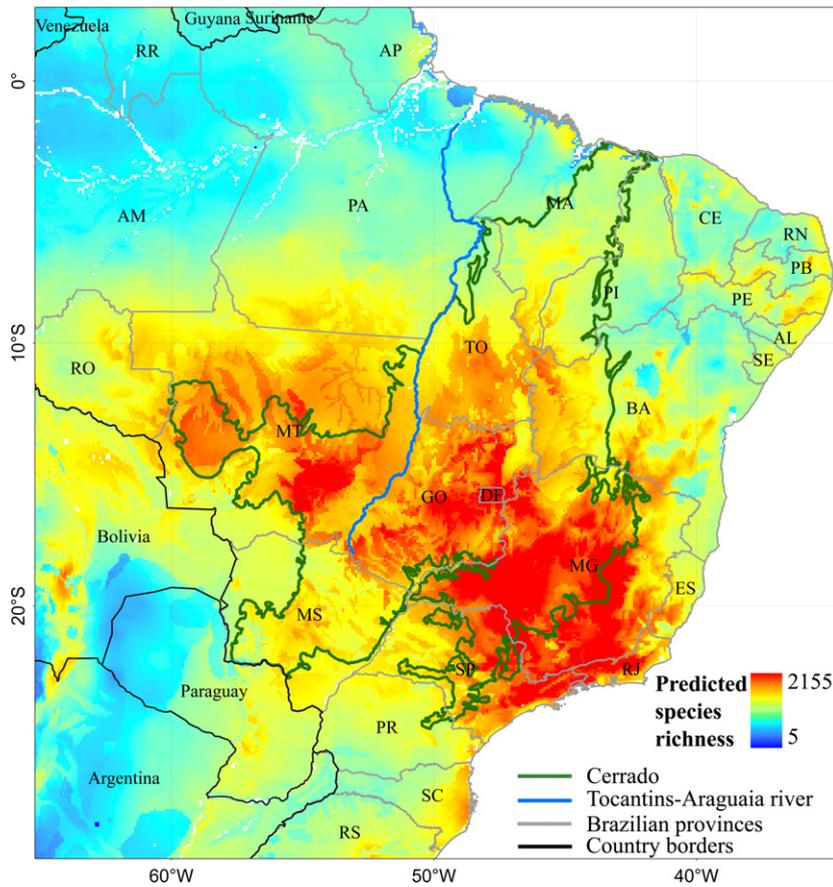
are known to occur in the Cerrado herb-shrub flora. The three most species-rich families were Fabaceae (11%), Asteraceae (10%) and Poaceae (7%). The predictor variables that had the highest relative contribution to significant SDMs were: "precipitation of driest quarter" (bio17), "minimum temperature of coldest month" (bio06) and "temperature seasonality" (bio04). The most important edaphic predictors were "pH in water" (PHAQ), followed by "effective cation exchange capacity" (eCEC), "total available water capacity" (TAWC) and "coarse fragments percentage" (CF). Spatial maps of all used variables are provided in Appendices S5 and S6.

### Botanical richness pattern

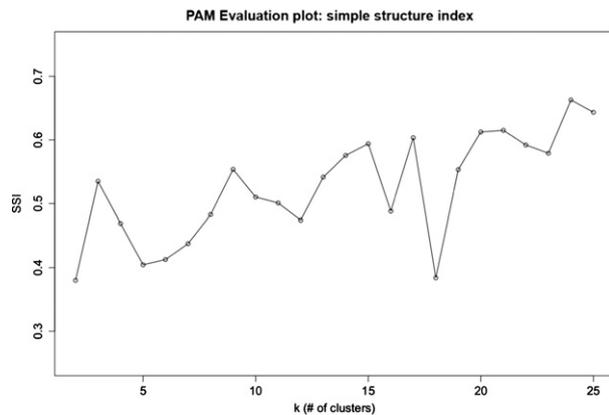
To develop the species richness pattern of the Cerrado herb-shrub flora, we stacked the 5039 significant and thresholded SDMs (Fig. 1). Richness values ranged from 5 towards the Andes (not shown) to 2155 species per 5 arc-min raster cell. Within the study regions two areas with high richness were found, separated by the Tocantins-Araguaia River (Fig. 1). The areas with high-predicted herb-shrub richness extend beyond the Cerrado boundary as defined by IBGE in the east. As expected, areas outside the Cerrado boundary have lower predicted richness of herb-shrub species because only species with known occurrence inside the Cerrado were included. Nonetheless, many species also occur beyond the Cerrado boundary.

### Phytogeographic regions

The SSI analysis indicated that an optimum number of PAM cluster groups were found at nine groups ( $k = 9$ ; Fig. 2), resulting in nine phytogeographic regions (Fig. 3). The UPGMA clustering of the medoids indicated that two major floristic clusters could be identified with two phytogeographic regions that represent the areas with the highest herb-shrub richness: regions 3 and 6 (Fig. 3). Core regions 3 and 6 differ in their species composition and both have the highest number of indicator species of their respective floristic cluster groups. Core region 3 has 1711 indicator species and is part of floristic cluster group 1–5 (Fig. 3, Appendix S7), and region 6 has 647 indicator species and is part of floristic cluster group 6–9 (Fig. 3, Appendix S7). Cluster group 1–5 represents phytogeographic regions that are mainly characterized by dry vegetation types and that have floristic affinities with Cerrado core region 3. Region 3 represents the central-southeast Cerrado and has the closest floristic affinity with disjunct region 4 that represents the Atlantic Forest in the east and the Bolivian Chiquitano dry forests in the west. Region 4 has few species indicative for this region and is characterized by a lack of species that are indicators for region 3.



**Fig. 1.** Predicted species richness pattern for Cerrado herb-shrub flora. (AL, Alagoas; AM, Amazonas; AP, Amapá; BA, Bahia; CE, Ceará; DF, Distrito Federal; ES, Espírito Santo; GO, Goiás; MA, Maranhão; MG, Minas Gerais; MS, Mato Grosso do Sul; MT, Mato Grosso; PA, Pará; PB, Paraíba; PE, Pernambuco; PI, Piauí; PR, Paraná; RJ, Rio de Janeiro; RN, Rio Grande do Norte; RO, Rondônia; RR, Roraima; SC, Santa Catarina; SE, Sergipe; SP, São Paulo; TO, Tocantins).



**Fig. 2.** SSI values for 2–25 cluster groups. The optimal number of PAM cluster groups is  $k = 9$ , resulting in nine phytogeographic regions for the larger Cerrado region.

Cluster group 6–9 represents phytogeographic regions that have floristic affinities with Cerrado core region 6, which are characterized by Amazonian floristic elements. Region 6 has few small disjunct pockets in region 7,

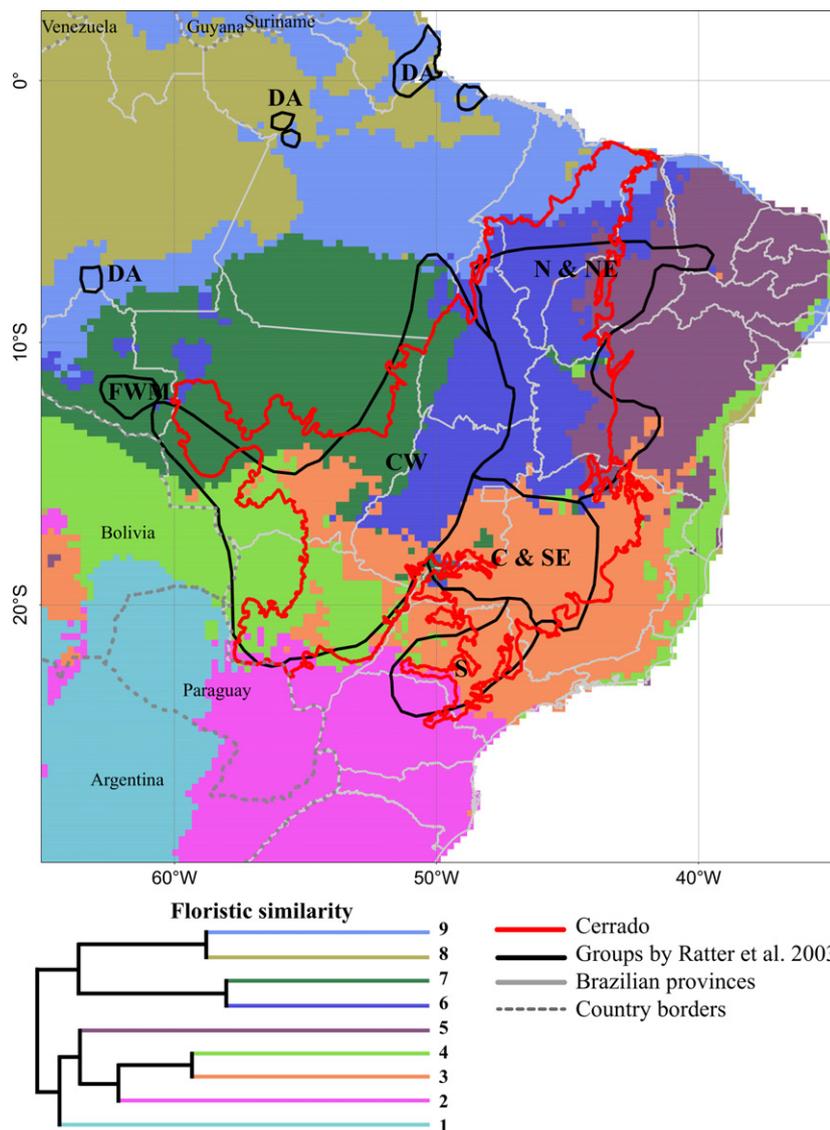
illustrative of the close floristic affinity between the two regions (Fig. 3). Regions 7, 8 and 9 represent Amazonian elements and disjunct savannas that share distribution of herb-shrub species with Cerrado core region 6.

**Protected areas and priority areas for conservation**

Plotting the protected areas over our high richness map indicated that only 18% of areas with high-predicted species richness are located inside protected areas (Fig. 4), and that 22% of the high-richness areas are included in the Cerrado priority areas for conservation (Fig. 4).

**Discussion**

For the first time the Cerrado herb-shrub flora has been quantitatively analysed based on 5039 significant SDMs of species with known occurrence in the Cerrado as defined by IBGE. The highest diversity was in the west and south-east parts of the Cerrado, divided by the Tocantins-Araguaia River (Fig. 1). The areas with high species richness



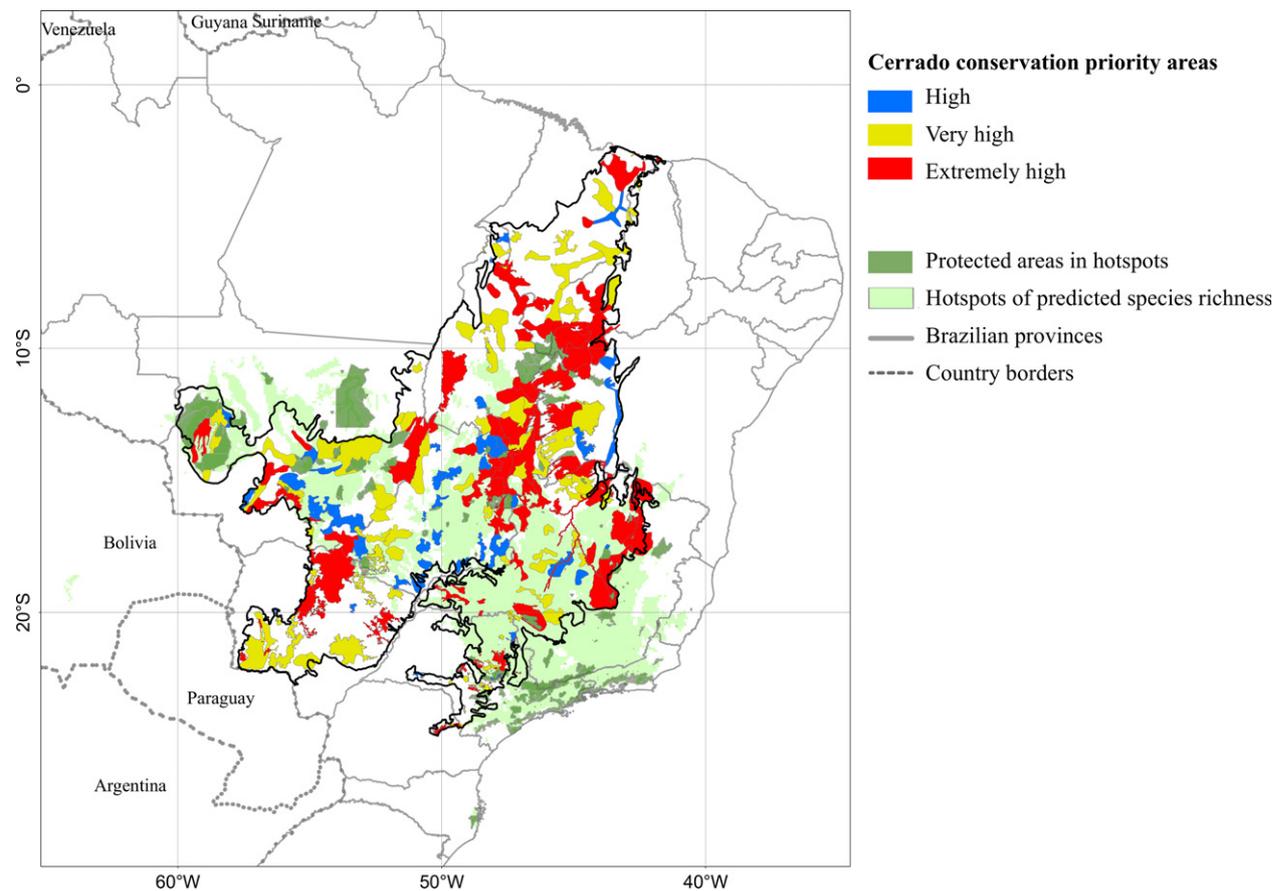
**Fig. 3.** The nine phytogeographic regions of the Cerrado based on predicted distributions of the herb–shrub flora, their compositional relationship (dendrogram in left corner) and floristic regions proposed by Ratter et al. (2003) for the Cerrado woody flora (C & SE, central & southeast; CW, central-west; DA, disjunct Amazonian savannas; FWM, far western mesotrophic sites; N & NE, north & northeast; S, southern).

are largely in the central-southeast phytogeographic region 3 (Fig. 3). Although the north part of the Cerrado is less rich in number of species (Fig. 1), it has a unique and different floristic composition characteristic of phytogeographic region 6 (Fig. 3). Based on distributions of the herb–shrub flora, the Cerrado is delineated by phytogeographic regions 3 and 6, which largely but not completely overlap with the Cerrado definition of IBGE.

**Botanical richness pattern**

The richness pattern of the herb–shrub flora of the Cerrado is defined by species-specific responses to abiotic

environmental gradients, biotic interactions and geographic barriers (Soberón & Peterson 2005; Soberón 2007; Godsoe 2010). SDMs predict species distributions based on identified correlations between species presence records and abiotic gradients (Phillips et al. 2006). Thus, responses of individual species to spatial differences in precipitation, temperature and soil conditions (Appendices S5 and S6) result in the richness pattern presented in Fig. 1. We did not account for biotic interactions or geographic barriers. We deem the potential impacts of geographic barriers negligible given that no major geographic barriers, such as mountain chains, divide the Cerrado region.



**Fig. 4.** The priority areas for conservation within the Cerrado domain according to the Brazilian Environment Ministry and current protected areas in Brazil in areas with high predicted species richness for Cerrado herb–shrub flora.

High richness in the Cerrado flora is related to the fact that Neotropical savannas comprise a mixture of elements of various provenance and floristic affinity (Simon et al. 2009). The Cerrado is bordered by the Amazon in the northwest, the Atlantic Forest in the southeast, Caatinga in the northeast, Pantanal and Chaco in the southwest, which all share floristic elements with the Cerrado flora (Simon et al. 2009). The proximity of these biomes and the Cerrado's central geographic position in South America has contributed to recruitment of many of its plant lineages (Simon et al. 2009).

The areas with high-predicted species richness are located in the west and central/southeast parts of the Cerrado (Fig. 1). An approximate richness pattern was reported for trees (Françoso et al. 2016; Bueno et al. 2017) and for amphibians (Bini et al. 2006). Some areas of high richness, located in central Brazil and in the southeast, represent mountains and high-altitude plateaus with elevations between 1000–2000 m a.s.l. These high-elevation areas (e.g. Serra do Espinhaço complex, Central highlands and southeast Brazilian mountain

tops) are covered by rupestrian grasslands, which are characterized by high herbaceous species richness (Fernandes 2016) and are recognized as centres of endemism for vascular plants (Simon & Proença 2000; Martinelli 2007), especially the herb–shrub flora, which is supported by the high number of indicator species for phytogeographic region 3 (Appendix S7). Most of the global centres of vascular plant richness are located in high-altitude regions of the tropics and subtropics (Barthlott et al. 2005). In these habitats the high number of species is associated with high diversity of habitats and different substrate characteristics (Conceição & Pirani 2005).

Ratter et al. (2006) noted particularly high species richness in the Araguaia and Tocantins drainage regions of Goiás, Tocantins and Mato Grosso. Past surveys have shown that parts of Mato Grosso State, such as Guimarães Plateau and Xavantina-Roncador Range, are among the areas with the highest floristic richness in the Cerrado (Eiten 1975; Ratter et al. 2006). Both findings are supported by our results.

### Phytogeographic regions

Nine phytogeographic regions were retained as the optimal PAM clustering solution (Fig. 2). The UPGMA cluster analysis of the nine medoids indicated that the two core Cerrado regions (3 and 6) are located in distinct floristic cluster groups (Fig. 3). The core Cerrado is represented by region 3 (central-southeast) and region 6 (central-north; Fig. 3). Both regions were recognized in Proença et al. (2010) as different vegetation complexes, together with a central and northwest Cerrado complex.

The cluster group including phytogeographic regions 1–5 and Cerrado core region 3 (Fig. 3) indicates that the central-southeast herb-shrub communities of the Cerrado share floristic elements with flora of the Atlantic Forest (Fig. 3; region 4) and seasonally dry vegetation types of South America, including the Bolivian Chiquitano region (Fig. 3; region 4), the Misiones Nucleus (Fig. 3; region 2), Caatingas (Fig. 3; region 5) and Chaco (Fig. 3; region 1), as described in Pennington et al. (2000) and DRYFLOR (2016). According to our analysis of the herb-shrub flora the boundary of the Cerrado should be extended to the southeast, covering all higher elevation areas. This extension of the Cerrado towards the Atlantic Forest region complies with the narrow definition of the Atlantic Forest of Carnaval & Moritz (2008). Furthermore, this explains the high overlap between the Cerrado and Atlantic Forest tree flora in Françoso et al. (2016), as their definition of the Atlantic Forest includes substantial highland areas that should be classified as Cerrado according to our analysis. Floristic affinities between Cerrado and Atlantic Forest can be explained by past climatic events of the Quaternary (Behling 1998; Carnaval & Moritz 2008) that resulted in expansion of the Atlantic Forest into the present Cerrado region, stretching into Bolivia (Bueno et al. 2017). These findings support our disjunct distribution of region 4 (Fig. 3). Our results imply that the herb-shrub flora followed the same pattern of expansion and contraction through the glacial cycles as the tree flora (Bueno et al. 2017).

Cerrado core region 3, together with phytogeographic regions 1, 2, 4 and 5, largely cover the area known as the South American Dry Diagonal (Neves et al. 2015), stretching from northeast Brazil into Argentina. The UPGMA cluster analysis of the medoids indicated that the herb-shrub communities of Cerrado core region 3 is most similar to the Atlantic Forest/Chiquitanos (Fig. 3; region 4), then to the Misiones (Fig. 3; region 2), followed by Piedmont (Fig. 3; region 1) and Caatinga (region 5; Fig. 3). Similarities between these different vegetation types were also described for woody species of the South American Dry Forests (Neves et al. 2015; DRYFLOR 2016). Neves et al. (2015), however, indicate that the Misiones flora is more

distantly related to the Cerrado flora than the Caatinga, which is not the case in our analysis. It should be noted that Neves et al. (2015) analysed the woody flora. This is likely caused by inclusion of floristic inventories to the Cerrado vegetation types that in our analysis are classified as Cerrado core region 6.

Cerrado core region 6 (Fig. 3) indicates that the floristic composition of the herb-shrub flora in the northern Cerrado is very different from the southern part, with floristic affinities to the Amazonian flora. Both the diversity (Fig. 1) and the number of indicator species in region 6 (Appendix S7) are lower than in region 3. Region 6 has closest affinity with region 7, which represents the southeast Amazon. Within the Amazon biome this area has the longest dry season (ter Steege et al. 2006). Compared to region 3, the amount of precipitation in region 6 is higher (especially in the warmest quarter; Appendix S5), which might explain differences in floristic composition between Cerrado core regions 3 and 6. Castro (1994). Simon & Proença (2000) and Proença et al. (2010) also recognized a northern Cerrado group in their studies, and Castro (1994) recognized the northeast Cerrado as a major subdivision.

We identified both similar and distinct patterns in the division of floristic regions between our results and those of Ratter et al. (2003). Our region 6 corresponds reasonably well to their north and northeast sites (Fig. 3; N and NE). Region 3 is similar to some extent to their central and southeast sites (Fig. 3; C and SE). We were not able to distinguish Ratter's southern (S) region; this area lies within our region 3, and should be lumped into one region (Fig. 3). Neither does our study support Ratter et al. (2003) central-western region (Fig. 3; CW). According to our findings this region is part of four of our phytogeographic regions (3, 4, 6 and 7). However, in the analysis of Ratter et al. (2003), the central-west region is a consensus of a grouping of different regions, and thus not well supported. Furthermore, we were unable to recognize Ratter's disjunct Amazonian sites (DA) and the Far Western Mesotrophic sites (FWM), which group in our Amazonian cluster. Ratter et al. (2003) reported that the FWM region has floristic affinities with the central-west and north and northeast sites, which supports our findings of Amazonian floristic affinities in the areas covered by these regions (Fig. 3).

Given the very large data set we used in our study and the building of relationships between species occurrence and abiotic conditions, we consider our delineation of floristic regions highly objective. It is important to note that the study of Ratter et al. (2003) was based only on the woody flora of Neotropical savannas and, consequently, excluded other vegetation types, such as grasslands and rocky grasslands, which contribute a significant part to Cerrado plant diversity.

### Protected areas and priority areas for conservation

In the Cerrado, 431 priority areas for biodiversity conservation were identified by the Brazilian Environment Ministry. These regions were selected based on biological importance, need for conservation actions and vulnerability. Our study indicated that only a limited part of these priority areas is located in areas with high-predicted herb-shrub species richness (Fig. 4). Furthermore, the percentage overlap between areas with high species richness and protected areas is small (Fig. 4). Considering these results, the current conservation policies for the Cerrado flora do not reflect the biological importance of the herb-shrub flora, and this problem is not exclusive to the herb-shrub flora of the Cerrado. For example, according to Veldman et al. (2015), misconceptions about herb-shrub environments contribute to alarming rates of loss due to conversion for agriculture and tree plantation, as well as to forest encroachment. Conservation and sustainable land-use policies in Brazil need to be extended to non-forest ecosystems in order to avoid losing more of Brazil's biodiversity and ecosystem functions (Overbeck et al. 2015; Veldman et al. 2015).

It is important to discuss possibilities to redesign and reconsider conservation plans for Cerrado vegetation. Ideally a conservation network should protect areas with high species richness in each of the identified core Cerrado phytogeographic regions. Hence, we suggest the need for further studies to expand conservation networks in the central part of Tocantins, west and south of Bahia, the central-southern part of Mato Grosso, south and north of Goiás, and the central, western and northern parts of Minas Gerais (Fig. 1). All these areas have high-predicted species richness. These suggestions should be combined with an assessment of the woody flora of the Cerrado.

Most worrying is that all the above-identified regions are under high pressure due to the expansion of agriculture, pastures and charcoal production (MMA 2010). The Cerrado, in the last two decades, was certainly the most threatened biome through land conversion (Beuchle et al. 2015). Since 1970 producers have been encouraged to open large areas in the core Cerrado, and a new agricultural frontier in the northern Cerrado has been established, comprising Maranhão, Tocantins, Piauí and Bahia provinces, which is called MaToPiBa. The economic incentives and government support for these activities has brought major changes to the Cerrado's natural landscapes, resulting in habitat fragmentation, biodiversity loss, deforestation, soil erosion and water pollution (MMA 2010). This rapid and extensive land-use change to the Cerrado vegetation reinforces the need to implement effective conservation strategies, and our study can help to guide future conservation planning to this end.

### Acknowledgements

We thank Rede ComCerrado and the National Council for Scientific and Technological Development (CNPq) programmes: PELD and SISBIOTA. The first author was supported by two PhD grants (REUNI/PDSE - BEX 12710/13-0) from Coordination for the Improvement of Higher Education Personnel (CAPES). Data collection was conducted by AGA, CBRM and BMTW. AGA and NR were responsible for development of the R scripts and the collection record database, for which they also used of databases provided by BMTW and CBRM. Data analyses were conducted by AGA, NR and JAG. The manuscript was drafted by AGA and edited by CBRM, BMTW, JAG and NR.

### References

- Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. & Biesmeijer, J.C. 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria – Dutch hoverflies as a case study. *PLoS ONE* 8: e63708.
- Araújo, M.B. & Peterson, A.T. 2012. Uses and misuses of bioclimatic envelope modelling. *Ecology* 93: 1527–1539.
- Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G. & Kreft, H. 2005. Global centers of vascular plant diversity. *Nova Acta Leopoldina* 92: 61–83.
- Behling, H. 1998. Late quaternary vegetational and climatic changes in Brazil. *Review of Palaeobotany and Palynology* 99: 143–156.
- Beuchle, R., Grecchi, R.C., Shimabukuro, Y.E., Seliger, R., Eva, H.D., Sano, E. & Achard, F. 2015. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography* 58: 116–127.
- Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F., Bastos, R.P. & Pinto, M.P. 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions* 12: 475–482.
- Borcard, D., Gillet, F. & Legendre, P. 2011. *Numerical ecology with R*. Springer, New York, NY, US.
- Bueno, M.L., Pennington, R.T., Dexter, K.G., Kamino, L.H.Y., Pontara, V., Neves, D.M., Ratter, J.A. & Oliveira-Filho, A.T. 2017. Effects of Quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. *Ecography* 40: 403–414.
- Carnaval, A.C. & Moritz, C. 2008. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* 35: 1187–1201.
- Castro, A.A.J.F. 1994. Comparison floristic-geographical and phytosociology of Cerrado samples. Ph.D. thesis, University of Campinas, São Paulo, BR.
- Conceição, A.A. & Pirani, J.R. 2005. Delimitação de habitats em campos rupestres na Chapada Diamantina, Bahia: substratos,

- composição florística e aspectos estruturais. *Boletim de Botânica da Universidade de São Paulo* 23: 85–111.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Gudrun, C., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., (...) & Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 027–046.
- DRYFLOR 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* 353: 1383–1387.
- Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.
- Durigan, G. & Ratter, J.A. 2016. The need for a consistent fire policy for Cerrado conservation. *Journal of Applied Ecology* 53: 11–15.
- Durigan, G., Siqueira, M.F.D. & Franco, G.A.D.C. 2007. Threats to the Cerrado remnants of the state of São Paulo, Brazil. *Scientia Agricola* 64: 355–363.
- Eiten, G. 1975. The vegetation of the Serra do Roncador. *Biotropica* 7: 112–135.
- Elith, J., Graham, C.H., Anderson, P.R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., (...) & Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57.
- Fernandes, G.W. 2016. The megadiverse rupestrian grassland. In: Fernandes, G.W. (ed.) *Ecology and conservation of mountain-top grasslands in Brazil*, pp. 3–14. Springer, Basel, CH.
- Françoso, R.D., Haidar, R.F. & Machado, R.B. 2016. Tree species of South America central savanna: endemism, marginal areas and the relationship with other biomes. *Acta Botanica Brasilica* 30: 78–86.
- Furley, P.A. & Ratter, J.A. 1988. Soil resources and plant communities of the Central Brazilian Cerrado and their development. *Journal of Biogeography* 15: 97–108.
- Godsoe, W. 2010. I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos* 119: 53–60.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Holdridge, L.R., Grenke, W.C., Hatheway, W.H., Liang, T. & Tosi, J.A. 1971. *Forest environments in tropical life zones: a pilot study*. Pergamon Press, New York, NY, US.
- Kaufman, L. & Rousseeuw, P. 1987. Clustering by means of medoids. In: Dodge, Y. (ed.) *Statistical data analysis based on the L1-norm and related methods*, pp. 405–416. Springer, Dordrecht, NL.
- Kreft, H. & Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* 37: 2029–2053.
- Legendre, P. & De Cáceres, M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecological Letters* 16: 951–963.
- Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35: 105–116.
- Martinelli, G. 2007. Mountain biodiversity in Brazil. *Brazilian Journal of Botany* 30: 587–597.
- Mendonça, R.C., Felfili, J.M., Walter, B.M.T., Silva-Júnior, M.C., Rezende, A.V., Filgueiras, T.S. & Nogueira, P.E. 2008. Flora vascular do Cerrado. In: Sano, S.M., Almeida, S.P. & Ribeiro, J.F. (eds.) *Cerrado: ecologia e flora*, pp. 289–556. Embrapa Cerrados, Brasília, BR.
- Merow, C., Smith, M.J. & Silander, J.A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.
- MMA – Ministério do Meio Ambiente. 2010. *Plano de Ação para prevenção e controle do desmatamento e das queimadas no Cerrado: conservação e desenvolvimento*. Ministério do Meio Ambiente, Brasília, BR.
- Munhoz, C.B.R., Felfili, J.M. & Rodrigues, C. 2008. Species–environment relationship in the herb–subshrub layer of a moist savanna site, Federal District, Brazil. *Brazilian Journal of Biology* 68: 25–35.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A. & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.
- Neves, D.M., Dexter, K.G., Pennington, R.T., Bueno, M.L. & Oliveira-Filho, A.T. 2015. Environmental and historical controls of floristic composition across the South American Dry Diagonal. *Journal of Biogeography* 42: 1566–1576.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., (...) & Kassem, K.R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience* 51: 933–938.
- Overbeck, G.E., Vélez-Martin, E., Scarano, F.R., Lewinsohn, T.M., Fonseca, C.R., Meyer, S.T., Muller, C., Ceotto, P., Dadalt, L., (...) & Pillar, V.D. 2015. Conservation in Brazil needs to include non-forest ecosystems. *Diversity and Distributions* 21: 1455–1460.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Townsend, P. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102–117.
- Pennington, R.T., Prado, D.E. & Pendry, C.A. 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography* 27: 261–273.

- Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19: 181–197.
- Proença, C.E.B., Soares-Silva, L.H., Rivera, V.L., Simon, M.F., Oliveira, R.C., Santos, I.A., Batista, J.A.N., Ramalho, C.L., Miranda, Z.J.G., (...) & Carvalho, S.F. 2010. Regionalização, centros de endemismos e conservação com base em espécies de angiospermas indicadoras da biodiversidade do Cerrado brasileiro. In: Diniz, I.R., Marinho-Filho, J., Machado, R.B. & Cavalcanti, R.B. (eds.) *Cerrado: conhecimento científico quantitativo como subsídio para ações de conservação*, pp. 89–146. Editora UnB, Brasília, BR.
- Raes, N. 2012. Partial versus full species distribution models. *Natureza & Conservação* 10: 127–138.
- Raes, N. & ter Steege, H. 2007. A null-model for significance testing of presence-only species distribution models. *Ecography* 30: 727–736.
- Ratter, J.A., Bridgewater, S. & Ribeiro, J.F. 2003. Analysis of the floristic composition of the Brazilian Cerrado vegetation III: comparison of the woody vegetation of 376 areas. *Edinburgh Journal of Botany* 60: 57–109.
- Ratter, J.A., Ribeiro, J.F. & Bridgewater, S. 2006. The Cerrado of Brazilian Amazonia: a much-endangered vegetation. In: Posey, D.A. & Balick, M.J. (eds.) *Human impacts on Amazonia: the role of traditional ecological knowledge in conservation and development*, pp. 85–97. Columbia University Press, New York, NY, US.
- Ribeiro, J.F. & Walter, B.M.T. 2008. As principais fitofisionomias do Bioma Cerrado. In: Sano, S.M., Almeida, S.P. & Ribeiro, J.F. (eds.) *Cerrado: ecologia e flora*, pp. 151–199. Embrapa Cerrados, Brasília, BR.
- Silva, J.F., Fariñas, M.R., Felfili, J.M. & Klink, C.A. 2006. Spatial heterogeneity, land use and conservation in the Cerrado region of Brazil. *Journal of Biogeography* 33: 536–548.
- Simon, M.F. & Proença, C. 2000. Phytogeographic patterns of Mimosa (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: an indicator genus of high-altitude centers of endemism? *Biological Conservation* 96: 279–296.
- Simon, M.F., Grether, R., Queiroz, L.P., Skema, C., Pennington, R.T. & Hughes, C.E. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences of the United States of America* 106: 20359–20364.
- Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115–1123.
- Soberón, J. & Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2: 1–10.
- ter Steege, H., Pitman, N.C., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., Molino, J.F., Prévost, M.F., Spichiger, R., (...) & Vásquez, R. 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* 443: 444–447.
- van Proosdij, A.S., Sosef, M.S., Wieringa, J.J. & Raes, N. 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39: 542–552.
- Veldman, J.W., Overbeck, G.E., Negreiros, D., Mahy, G., Le Stradic, S., Fernandes, G.W., Durigan, G., Buisson, E., Putz, F.E. & Bond, W.J. 2015. Tyranny of trees in grassy biomes. *Science* 347: 484–485.
- Vollering, J., Schuiteman, A., de Vogel, E., van Vugt, R. & Raes, N. 2016. Phytogeography of New Guinean orchids: patterns of species richness and turnover. *Journal of Biogeography* 43: 204–214.
- Wisn, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS Predicting Species Distributions Working Group 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distribution* 14: 763–773.
- Zhang, M.G., Zhou, Z.K., Chen, W.Y., Slik, J.W.F., Cannon, C.H. & Raes, N. 2012. Using species distribution modeling to improve conservation and land use planning of Yunnan, China. *Biological Conservation* 153: 257–264.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Geographic coordinates of the areas inventoried inside the Cerrado.

**Appendix S2.** Published literature on the Cerrado herb–shrub flora used to generate the models.

**Appendix S3.** Uncorrelated set of 17 environmental predictor variables.

**Appendix S4.** Species' distribution model (SDM) AUC values and 95% CI AUC values of the null models.

**Appendix S5.** The geographic presentation of the eight least correlated bioclimatic predictors and potential evapotranspiration (PET) used to model the species distribution patterns of the herb–shrub flora of the Cerrado.

**Appendix S6.** The geographic presentation of the nine least correlated quantitative continuous soil variables used to model species distribution patterns of the herb–shrub flora of the Cerrado.

**Appendix S7.** Number of species of the Cerrado's herb–shrub flora that best match a given phytogeographic region according to Indicator Species Analysis (ISA).

**Data Accessibility.** Map figures presented in this study are available as raster files from the Pangaea database: <https://doi.org/10.1594/PANGAEA.874914>.