The impacts of landscape composition, marginality and climatic stability on the patterns of endemism of Cerrado woody plants

Running title: Endemism and climate of the Cerrado

João de Deus Vidal Júnior[1], Anete Pereira de Souza*[1,2] & Ingrid Koch[2]

[1]Center of Molecular Biology and Genetic Engineering (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil

[2] Plant Biology Department, Biology Institute, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil

*Email: jd.vidal27@gmail.com

Corresponding author:

Anete Pereira de Souza,

Plant Biology Department, Biology Institute, UNICAMP, 13083-862, Campinas, SP, Brazil

Telephone number: +55 19-3521-1132 31 Fax number: +55 19-3521-1089

E-mail: anete@unicamp.br

Acknowledgements:

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) in the form of a scholarship to J.D. Vidal. This research was also supported by CAPES - Computational Biology Program.

We would like to thank professors Fábio Pinheiro and Paulo De Marco Júnior for reading and suggesting important improvements to the manuscript.

Conflict of interest:

The authors have no conflicts of interest to declare.

Author contributions:

JDV designed the study, performed the analysis, and wrote the manuscript; APS and IK assisted with the discussion of results and wrote the manuscript.

Biosketch

João de Deus Vidal Jr. is a postdoctoral researcher in the Laboratory of Genetics and Molecular Analysis at Universidade Estadual de Campinas. He is interested in spatial patterns of diversity of Neotropical plants, with a special focus on the role of past climatic changes in the generation of modern patterns of diversity, at both community and population levels. Ingrid Koch is a professor at the Institute of Plant Biology of Universidade Estadual de Campinas. She is interested in the systematics, evolution and species distribution of plants from Neotropical environments. Anete Pereira de Souza is a professor at the Institute of Plant Biology of Universidade Estadual de Campinas. She is interested in the population genetics and molecular evolution of plants, with an emphasis on bioinformatics and genome evolution.

ABSTRACT

Aim: Many theories were proposed to explain the outstanding endemism of plants in the Cerrado. Among the diversification mechanics suggested, the most widely discussed are the plateau/valley, stable/unstable, and core/peripheral hypotheses. The first argues that plateaus harbor more ancient lineages than valleys and should therefore present higher levels of endemism. The second theory suggests that climate stability in certain regions promoted more suitable environments for the maintenance of paleoendemic species. The last scenario attributes the distribution of endemism in the Cerrado to gradients of optimal conditions available to locally adapted species, predicting higher endemism levels in core regions of the domain than in marginal areas. We compared the endemism patterns of Cerrado flora with the predictions of each theory to discuss their role in the distribution of endemism nuclei across the domain.

Location: Brazil.

Time period: Quaternary.

Major taxa studied: Angiosperms.

Methods: We generated an endemism map using records of 174 woody plant species, comprising most vegetation types of the Cerrado domain. We applied spatial analysis and distribution models to summarize the importance of each predictor of endemism.

Results: We found correlations among endemism, elevation, and marginality, supporting the plateau/valley and core/peripheral hypotheses. Although plateaus are more stable climatic environments, their elevation and centrality, not their climatic stability, predict higher endemism. Our vegetation models using simulated datasets supported the expansion of Cerrado plateau vegetation and retraction in valleys during the Last Glacial. However, real occurrence data models for gallery and riverine forests indicated that species in valleys presented contrasting responses to past climatic shifts.

Main conclusions: We identified regions of higher plant endemism within the Cerrado domain and verified elevation and centrality as stronger predictive factors of this endemism compared with climatic stability. Our results also indicate that different vegetation types within valleys presented different responses to the Quaternary climate.

Keywords: Biogeography, Community ecology, Species distribution modeling, Last Glacial Maximum, Savanna, SDTF

INTRODUCTION

The Cerrado is the largest savanna and the second largest major bioregion in South America, with a complex and unique biogeographic history. The Cerrado harbors more plant species than any other savanna in the world (> 7000 species; Mendonça et al., 1998), with over 44% of its flora being endemic, making it the richest tropical savanna in the world in this regard (Klink & Machado, 2005). The Cerrado landscape is divided into ancient plateaus, which are large extension areas with relatively high elevations (500–1700 m), and peripheral valleys, which are lower elevation regions (100–500 m) that surround the highlands (Silva, 1997; Silva & Bates, 2002). These landscape unities present different biotic compositions and geological histories, indicating that their biogeographical histories were also distinct. The biota of valleys is heterogeneous, composed of gallery, riverine, and dry forests, marshlands, and enclaves of species from adjacent biomes (Silva & Bates, 2002), while plateaus present the savanna as a dominant vegetation type (Silva, 1997).

The outstanding diversity of plant species in the Cerrado is a topic of present discussion in Neotropical biogeography. Many theories have been proposed to explain the high degree of endemism found for several taxa, but due to the variety of responses of organisms and the limited studies concerning this environment (Turchetto-Zolet, Pinheiro, Salgueiro, & Palma-Silva, 2013), no prevailing process acting on the biogeographic history of this domain has emerged thus far. Regarding the currently discussed theories, three particular hypotheses of diversification are the most supported scenarios: (1) the geographic compartmentalization of the landscape in highlands and depressions of the Late Miocene as a driver of allopatric speciation between disjoint landscape units; (2) the differences in Pleistocene climatic stability along regions within the domain promoting the differential maintenance of diversity in stable and unstable regions; and (3) the gradient of optimal environmental conditions between the core and periphery of the Cerrado that promoted lower

extinction rates in nuclear regions, which favored the maintenance of endemism in the center of the domain. Although they are not mutually exclusive, the relative importance of these theories remains a major focus of debate because most individual patterns of taxa seem to support different scenarios (e.g., Turchetto-Zolet et al., 2013).

The first hypothesis attributes differential endemism across the Cerrado to the older geological origin of plateau formations, which date from the Late Tertiary, while valleys are the result of recent erosive processes (Ab'Sáber, 1983; Silva, 1997; Werneck, 2011). Given this scenario, ancient geomorphological processes acting on these environments might have promoted the isolation of older lineages in the highlands. Consequently, the recently formed valleys were gradually occupied by neoendemic species (Silva, 1997). Therefore, for this hypothesis, higher elevation regions are expected to present more endemism than depressions. This model is consistent with endemism patterns reported for birds (Silva, 1997), plants of the genus *Mimosa* (Simon & Proença, 2000), and squamate lizards (Nogueira, Ribeiro, & Costa, 2011).

In addition to the geomorphological processes of the Late Miocene, the Pleistocene climate was also proposed as a major driver of diversification of species and lineages within the Cerrado (Werneck, Nogueira, Colli, Sites, & Costa, 2012; Santos, Nogueira, & Giugliano, 2014; Bueno et al., 2017; Arruda, Schaefer, Fonseca, Solar, & Fernandes-Filho, 2018; Costa et al., 2018). This hypothesis is closely associated with both plateau/valley and core/periphery hypotheses, as both valleys (Ab'Sáber, 1983) and marginal regions (Costa et al., 2018) of the Cerrado were more sensitive to Quaternary climatic shifts than highlands or the central regions of the domain, as suggested by their pollen record and the frequent presence of species from adjacent biomes (Ledru, 2002; Costa et al., 2018). In this scenario, during cycles of cold and dry climates, the distribution of moist vegetation from valley environments underwent strong retractions. By contrast, the plateau vegetation expanded into

valleys, possibly even penetrating the modern distribution of adjacent domains (van der Hammen, 1974). Indeed, stable climatic regions are considered to predict diversity at both interspecific (Werneck, 2011) and intraspecific levels (Carnaval & Moritz, 2008; Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009), as areas presenting less stable environmental conditions are more prone to face local extinction (Harrison & Noss, 2017).

The third hypothesis argues that core regions of the Cerrado historically presented relatively more optimal conditions for the maintenance of regionally adapted populations of its species, supporting the diversification of endemics through time (Soule, 1973; Eckert, Samis, & Lougheed, 2008). In this scenario, nuclear regions of the Cerrado are expected to show higher levels of endemism, whereas communities in peripheral regions are more prone to undergo stochastic events of local extinction, given the exposure of populations of species to sub-optimal conditions in comparison with nuclear regions (Soule, 1973). An example supporting this scenario is the genetic diversity of populations of the squamate lizards of the genus *Micrablepharus* (Santos et al., 2014).

Multiple species niche models are useful for detecting general patterns, which are important for assembling large-scale scenarios. Large-scale hypothesis testing for the identification of refugia has been conducted for several regions around the world, with important biogeographic outcomes (Thuiller, 2004; Svenning, Normand, & Kageyama, 2008). Mostly, single species patterns or simulated dataset models have been applied in past distributional reconstructions for the Cerrado (Collevatti et al., 2012; Werneck et al., 2012; Lima, Telles, Chaves, & Lima-Ribeiro, 2017; Arruda et al., 2018), while multiple species approaches are less common (de Siqueira & Peterson, 2003; Bonaccorso, Koch, & Peterson 2006; Bueno et al., 2017). These studies addressed future climatic changes based on a conservation biology approach (de Siqueira & Peterson, 2003), simulated the history of Amazonian forest fragmentation and the potential occupation of open areas by Cerrado

species (Bonaccorso et al., 2006) and reconstructed the distribution of woody savanna and seasonally dry tropical forests (SDTF) vegetation under past scenarios (Bueno et al., 2017). These studies also identified an important relationship between landscape composition and shifts in the distribution of the savanna during glaciations (Bueno et al., 2017). However, none of these studies included simulations that explicitly incorporated the differences in the abiotic and biotic compositions of valleys and plateaus. Here, we compared the levels of endemism of the woody flora of the Cerrado domain with the three scenarios proposed to explain its diversification. We tested whether higher elevations, central regions or climatically stable areas show higher levels of endemism. In addition, we modeled real and simulated occurrence data of woody plants from valley and plateaus to test the scenarios of retraction/expansion of different vegetation types occurring in these landscapes during glacial cycles.

METHODS

Species occurrence data and endemism calculation

To generate the maps of endemism for woody species of the Cerrado, we gathered occurrence data for 63 savanna species, 54 SDTF species, 36 gallery forest species, and 21 riverine forest species (Table 1). We elaborated the species list for savanna, gallery and riverine forests based on the NeoTropTree database (Oliveira-Filho 2017), the Cerrado physiognomy description published by Ribeiro and Walter (1998), and the SDTF catalogue proposed by Prado and Gibbs (1993). The savanna list included species solely present in Cerrado arid woodlands following the NeoTropTree database. The SDTF species list included all valid species listed in Prado and Gibbs (1993), not considering infraspecific classification. The riverine forest species list and the gallery forest species list included

typical species of each vegetation type, following the classification system proposed by Ribeiro and Walter (1998). We considered riverine forests and gallery forests to be different vegetation types due to their composition and the restricted distribution of riverine forests in association with major rivers of the domain (Ribeiro & Walter, 1998; Silva & Bates, 2002). We excluded species present in multiple vegetation types to reduce the bias of widespread species in models. We retrieved species occurrence data from GBIF using the R package "rgbif" (Chamberlain, Ram, Barve, & Mcglinn, 2016). To ensure the quality of the data, we adopted search parameters that restrict data without original coordinates and with spatial issues. A full list of material data sources is presented in Appendix 2. We manually checked points by plotting individual maps and discarding outliers. We also discarded species with poor occurrence information or a low number of registries (< 10 registries).

We generated a map of weighted endemism using the algorithm proposed by Guerin, Ruokolainen, and Lowe (2015), applying a geographic grid with 0.25 × 0.25 degrees cells to generate the presence/absence matrix. We calculated the number of unique species that occurred in each cell and weighted this statistic for each species by dividing it by the number of grid cells in which they occurred (Guerin et al., 2015). By doing so, we expected to reduce the bias of widespread generalist species in the estimates of endemism.

Endemism correlation with elevation, marginality and climate

To evaluate the relative climatic stability of plateaus and valleys, we divided the entire Cerrado domain into highlands (≥ 500 m elevation) and depressions (< 500 m elevation). We randomly sampled 1000 points for each landscape unity and extracted values of 19 WorldClim bioclimatic variables (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) to compare the present environmental conditions between areas. To compare climatic stability, we sampled the same random points for past climatic scenarios (mid-Holocene, Last Glacial

Maximum and Last Interglacial; Hijmans et al., 2005). We calculated the variances of each environmental variable of plateaus and valleys and compared them using a bootstrap approach by sampling 1000 repetitions of the variances and comparing the significance of outcomes with Welch's t-test (Welch, 1938) for each variable in each landscape unit.

To test the correlation of endemism with elevation and marginality, we extracted values of elevation and distance from the Cerrado shape centroid for 1000 random points. We extracted the values of endemism for these same points and fitted them into linear models using Pearson's product-moment correlation test to estimate the significance of these associations. To test the relationship between endemism and stability, we fitted a generalized linear model with endemism as a response variable, with the variances of each of the 19 bioclimatic variables between the present, mid-Holocene, LGM, and LIG as predictors, as well as elevation and distance to the core. We then compared the p-values of each predictor to discuss their relative importance.

Historical vegetation distribution

To test the historical climate hypothesis, we assessed whether plateau vegetation expanded or retracted during glaciations and whether the response of valleys was the opposite, as proposed in the refugia scenario. We generated models for simulated points of valleys and plateaus and for the real occurrence of species associated with gallery forests, riverine forests, SDTF and savanna (Table 1). We sampled 1000 random points from all 19 Worldclim bioclimatic variables to calculate Pearson's pairwise correlation and discard variables with a high correlation (≥ 0.75). We obtained a set of 12 less correlated variables: annual precipitation, precipitation seasonality, precipitation of the driest quarter, precipitation of the warmest quarter, precipitation of the coldest quarter, mean diurnal range, isothermality, temperature seasonality, maximum temperature of the warmest month, mean temperature of

228 the wettest quarter, mean temperature of the driest quarter, and elevation. For past scenarios
229 (mid-Holocene, Last Glacial Maximum and Last Interglacial), we used the variables from
230 MIROC-ESM and CCSM4-scaled climatic models (Hijmans et al., 2005). We averaged
231 models between both scenarios (MIROC-SM and CCSM4) and presented a consensus model
232 for each of these time periods. For the Last Interglacial (LIG; ~120,000 – 140,000 years BP),
233 we adopted a projection of bioclimatic values from WorldClim following Otto-Bliesner et al.
234 (2006).

235 We used the maximum entropy machine-learning algorithm Maxent v. 3.3.3 (Phillips,
236 Anderson, & Schapire, 2006) to model the potential distribution for each species and for the
237 simulated datasets. We processed the models and bioclimatic layers using R packages
238 "dismo" (Hijmans, Phillips, Leathwick, & Elith, 2013) and "raster" (Hijmans & van Etten,
239 2014). For species distribution models, we used 1000 background points for each species to
240 estimate background information with pseudo-absences. The model performance was
241 evaluated by AUC-ROC values. We converted present models into binary models by
242 establishing a threshold based on the lower presence training (LPT) value from which a
243 presence point is recorded in the actual data for the present model (Pearson, Raxworthy,
244 Nakamura, & Townsend Peterson, 2007). Then, we applied this same procedure to past
245 scenarios using the present model threshold. We summarized information from models by
246 counting per species presence cells for each scenario reconstruction. To facilitate comparison
247 between species, we converted the values into percentages by dividing areas by the larger
248 range registered for each species. In so doing, we were able to detect under which climatic
249 scenario a species reached the maximum range of its distribution, and we were able to
250 compare glacial and interglacial distributions of species. We also combined all binary
251 models, for each species and for the whole dataset, to identify individual- and community-
252 level stable areas. The same approach was implemented for the simulated datasets.

RESULTS

Species list and endemism calculation

In our survey, we obtained 9452 registries for 54 SDTF species, 1085 occurrence records for 21 riverine forest species, and 6903 records for 36 gallery forest species. For the savanna (Cerrado *sensu stricto*), we recovered 9145 occurrences for 63 species (Appendices S1, S2, S3 and S4). Our estimates of weighted endemism for the Cerrado domain returned higher levels of endemism in the states of Minas Gerais, Goiás, Mato Grosso, Mato Grosso do Sul, São Paulo, and Bahia (Figure 1). We were able to identify endemism nuclei in (1) Chapada dos Guimarães, (2) Serra do Roncador, (3) Araguaia Valley, (4) Chapada das Emas, (5) Serra dos Pirineus, (6) Chapada dos Veadeiros, (7) Chapada do São Francisco, (8) Planalto do Distrito Federal, (9) Serra do Cipó, and (10) Serra de Botucatu.

Endemism correlation with elevation, marginality and climate

All 19 bioclimatic variables analyzed presented significant differences between valleys and plateaus (Table 1). A total of 16 variables (84%) presented lower variances for plateaus than for valleys, suggesting that plateau climatic conditions were slightly more stable during the time periods that we compared. The other three variables displayed lower variances in valleys. Welch's two-sample t-test returned p-values lower than 0.01 for all comparisons. We observed a positive correlation for elevation and endemism, while distance to the core and endemism showed a negative correlation (Figure 2).

Historical vegetation distribution

The AUC for simulated datasets models was 0.736 for plateaus and 0.722 for valleys. For these data, our results recovered a pattern of glacial expansion for plateaus, while valleys presented glacial retraction in terms of suitability (Figure 3). In fact, the lowest proportional distribution for valleys and the highest for plateaus occurred during the LGM (Figure 4). For real occurrence data, average AUC values were 0.892 for SDTF models, 0.956 for riverine forest models, 0.879 for gallery forest models and 0.915 for savanna models, indicating high predictive accuracy of models. We found that 38% of savanna species presented their maximum range during glacial periods, while gallery species, riverine species and SDTF species presented variable patterns (Figure 5; Appendices S1, S2, S3 and S4). For the forest vegetation types, 81% of the modeled riverine species displayed wider distribution ranges during interglacial periods. SDTF and gallery forests presented an opposite pattern to that of the riverine forests, with 72% of gallery forest species and 54% of SDTF species presenting their wider distribution during the LGM. With respect to the savanna, we found that 65.08% of the species presented wider LGM ranges than the present distribution. Combined models for all species binary distributions resulted in community-level reconstructions of vegetation types (Figure 5). For these combined reconstructions, we also observed a similar response of a wider occurrence of SDTF and gallery forests during the LIG and LGM, while the savanna presented a more local expansion within the Cerrado. Riverine forests, in comparison, presented a slightly narrower distribution during the LGM and were greatly restricted during the LIG.

The potential distribution of species from the different vegetation types indicated that suitable areas for the occurrence of species from the savanna, gallery forests, and riverine forests expanded. Expansion for the savanna, however, was much more subtle than that for the gallery and riverine forests. SDTF models showed a clearer retraction in overall species

potential distribution, with a notable retraction in the LGM. From the mid-Holocene to the present scenario, potential areas for all vegetation types decreased.

DISCUSSION

We generated endemism maps for Cerrado woody plants and compared the patterns observed with biogeographical theories of the Cerrado, which assumed plateau/depression, stability/instability and core/peripheral processes involved in the generation of endemism. We demonstrated that the climate presented an overall higher stability in the plateaus than in the valleys. This stability, however, showed a weak correlation with endemism. Additionally, we reconstructed the distribution of these species and environments for the recent Quaternary period. We managed to link the distribution of higher endemism nuclei to elevation and centrality, reinforcing the importance of these elements in the biogeographic processes that generated the Cerrado diversity. For vegetation history reconstructions, we concluded that the savanna, a vegetation type associated with plateaus, expanded during the Last Glacial period, while the different valley vegetation types presented variable responses. With respect to valleys, SDTF retracted during the Last Glacial period, while riverine forests and gallery forests expanded. Moreover, we found evidence that plateaus configure stable scenarios for most climatic variables, but this stability is not necessarily linked with higher levels of endemism. In sequence, we discuss our results within the context of the climate changes, tectonic history and biogeographic processes of the domain.

Elevation and marginality as predictors of endemism

The identification of nuclei of endemism in Chapada dos Guimarães, Chapada dos Veadeiros, Serra do Cipó, and Distrito Federal is consistent with endemism patterns identified for other taxa (Simon & Proença 2000; Nogueira et al., 2011). Most of these areas

are located at high elevation, with a single valley region (Vale do Araguaia) identified as highly endemic, supporting the results of our correlation analysis. Elevation, along with landscape compartmentalization, is considered an important driver of diversification worldwide, as differences in elevation often promote isolation, a key process for speciation (Slaton, 2015). Although the concentration of endemism levels of plants at high elevations has already been suggested on a global scale (Steinbauer et al., 2016), a finer scale pattern for the Cerrado, as we presented here, reinforces the role of this relationship for this specific domain, as previously suggested (e.g., Fiaschi & Pirani, 2009).

In addition to elevation, marginality is also an important predictor of endemism. This pattern might be explained by the fact that species that are adapted to a particular environment tend to present higher survivability, reproductive success and population growth under their optimal conditions (Eckert et al., 2008). These conditions are gradually distributed across space and are expected to be more available in nuclear areas of distributions. This result is also supported by the higher historical suitability of nuclear region habitats for plants of the Cerrado, proposed by Bueno et al. (2017) and Costa et al. (2018). The real importance of interactions between marginality and endemism is controversial, however, as differential performances for nuclear populations have not been demonstrated by the available studies so far (Abeli, Gentili, Mondoni, Orsenigo, & Rossi, 2014; Pironon, Villellas, Morris, Doak, & García, 2015). However, we were able to demonstrate a convergence of endemism and centrality for the Cerrado. Hence, these variables are not independent of each other; due to the uplift of the Central Brazilian plateau in the late Miocene, the central region of Brazil displays considerably higher elevations than the marginal regions of the Cerrado (e.g., Werneck, 2011).

Climate dynamics in the Cerrado, vegetation changes and climatic refugia

The plateaus presented higher climatic stability for most (16) of the environmental variables included in our analysis. Valleys, comparatively, only presented lower variances than plateaus for annual mean temperature, precipitation seasonality, and precipitation of the warmest month. The absence of a correlation between climatic stability and endemism is justified by the fact that although the climatic component is a major driver of species distribution, it does not determine the species distribution by itself. Distributions and range shifts are determined by complex interactions between abiotic and biotic components and species mobility (Soberón & Peterson, 2005). Therefore, constraints to dispersal, such as the heterogeneous landscape composition of the central Brazilian region, most likely configured fundamental barriers, shaping the distribution of the flora of the Cerrado throughout the Quaternary.

For the potential distribution models, we observed very similar patterns for simulated and real occurrence datasets of the savanna and plateau points, supporting the close association of this vegetation with highlands in the Cerrado. Our results disagree, however, with the widespread savanna scenario in the LIG with posterior glacial retraction proposed by Bueno et al. (2017). Their models indicated a retraction for plateaus of the savanna during the LGM, which was interpreted as evidence of the role of the valleys as refugia during glaciations. Our models, in contrast, suggested a progressive retraction from the LGM to the present, with no significant glacial retraction, coinciding with the pollen registry for Central West Brazil (van der Hammen, 1974). From our perspective, interglacial climates enhanced humidity, instead of glacial aridity, promoting retraction of the range of most species from the savanna formations of the Cerrado into interglacial refugia; this scenario was previously proposed for the lineage diversification of highland restricted dry-adapted taxa in the Cerrado, such as *Pilosocereus aurisetus* (Bonatelli et al., 2014) and *Tibouchina papyrus* (Collevatti et al., 2012).

While simulated dataset models fit the previously suggested (Werneck, 2011) scenario of glacial expansion under plateau and retraction under valley environmental conditions, real occurrence data models agreed with the savanna expansion but showed a variety of responses for the different valley vegetation types. For the glacial scenario, most savanna, gallery forest, and riverine forest species presented expansion, while SDTF retracted. The closest relationships of riverine forests and gallery forests with rivers and watersheds (Ribeiro & Walter, 1998) are potential indicators that these species, although also occurring in valleys, show different tolerance and adaptation than species of SDTF. This may also be interpreted as an effect of the heterogeneity of climatic conditions observed for valley landscapes, which comprise more complex environmental gradients than plateaus. Our models support this assumption, indicating that even though the overall environmental space in valleys decreased during the LGM, riverine and gallery forests were able to expand their ranges. With respect to the mid-Holocene (6000 ybp), the driest phase of Central Brazil, most regions in South America maintained their vegetation types or experienced the advance of more arid/dry vegetation types (Marchant et al., 2009; Costa et al., 2018). As the savanna is the dominant vegetation type in plateaus (Silva & Bates, 2002; Werneck, 2011), the overall stability of savanna models during the climatic scenarios recovered for our models supports the assumption of more stable climates for these regions compared to valleys.

Among the vegetation types of valleys, riverine forests and gallery forests showed different responses than SDTF, which was coincident with the simulated dataset model for valleys. STDF models were consistent with a scenario of glacial retraction, followed by mid-Holocene expansion. Previous distribution models for SDTF also refuted the proposal of a wider distribution for this vegetation type during glacial periods (Werneck, Costa, Colli, Prado, & Sites, 2011). Our results differ from those of Werneck et al. (2011) by suggesting the expansion of SDTF during the LGM into the eastern corridor of the Amazon Basin, which

was a scenario previously proposed by Pennington, Prado and Pendry (2000). The discrepancy between the simulated models and the species models is mostly associated with the fact that taxa associated with plateaus or valleys are not restricted to their environments, as species often occur in different habitats than those to which they are mostly adapted, due to ecological plasticity (e.g., Hortal, Triantis, Meiri, Thébault, & Sfenthourakis, 2009). Using real occurrence data, we were able to identify a wide corridor in the eastern Amazon, potentially suitable for most savanna, SDTF and gallery forest species throughout the LGM, as well as gradually reduced suitability over the last 22,000 years. Although we noticed that models of species from the savanna and SDTF responded similarly to the simulated models, gallery and riverine forests did not. The variety of vegetation types occurring in valleys is also an indicator of the higher habitat diversity (Silva & Bates, 2002) and, possibly, of a complex biogeographic history, which may not be fully explained by a single vegetation response model.

CONCLUSION

We conclude that climatic stability alone is not able to predict endemism distribution in the Cerrado, which is mostly strongly correlated with the elevation and centrality of the environments in this domain. The majority of endemics is concentrated in higher elevation nuclear regions, while peripheral depressions present lower endemism levels. The Quaternary climate was indeed more stable in plateaus than in valleys, which demonstrates the overlap of these hypotheses. We conclude that the three hypotheses interacted in the origin and distribution of the diversity of the domain, with different magnitudes. Additionally, we demonstrate how the different vegetation types of the Cerrado responded individually to the climatic shifts of the last 130,000 years. We expect our results to provide new hypotheses and possible scenarios for further investigations about how past climatic changes affected the

endemism distribution of Neotropical biota. By demonstrating that vegetation types across a domain may present differential responses to climatic changes, we recommend more careful approaches in reconstructions aiming to estimate the impacts of future climatic changes on wider scales. Additionally, we recommend further studies on community-level approaches, as we implemented here, with other organisms, to test the extent of our interpretations.

REFERENCES

Ab'Sáber, A. N. (1983). O domínio dos cerrados: introdução ao conhecimento. *Revista do Serviço Público*, 111, 41–55.

Abeli, T., Gentili, R., Mondoni, A., Orsenigo, S. & Rossi, G. (2014). Effects of marginality on plant population performance. *Journal of Biogeography*, 41, 239–249. doi: 10.1111/jbi.12215.

Arruda, D. M., Schaefer, C. E. G. R., Fonseca, R. S., Solar, R. R. C. & Fernandes-Filho, E. I. (2018). Vegetation cover of Brazil in the last 21 ka: new insights into the Amazonian refugia and Pleistocenic arc hypotheses. *Global Ecology and Biogeography*, 27, 47–56. doi: 10.1111/geb.12646.

Bonaccorso, E., Koch, I. & Peterson, A. T. (2006). Pleistocene fragmentation of Amazon species' ranges. *Diversity & Distributions*, 12, 157–164. doi: 10.1111/j.1366-9516.2005.00212.x.

Bonatelli, I. A. S., Perez, M. F., Peterson, A. T., Taylor, N. P., Zappi, D. C., Machado, M. C., … Moraes, E. M. (2014). Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeodistributional reconstructions for Pilosocereus aurisetus and allies. *Molecular Ecology*, 23, 3044–3063. doi: 10.1111/mec.12780.

Bueno, M. L., Pennington, R. T., Dexter, K. G., Kamino, L. H. Y., Pontara, V., Neves, D. M., … de Oliveira-Filho, A. T. (2017). Effects of Quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. *Ecography*, 40, 403–414. doi: 10.1111/ecog.01860.

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. doi: 10.1111/j.1365-2699.2007.01870.x.

Carnaval, A. C., Hickerson, M. J., Haddad, C. F., Rodrigues, M. T. & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785–789. doi: 10.1126/science.1166955.

Chamberlain, S., Ram, K., Barve, V. & Mcglinn, D. (2016) *Rgbif: Interface to the Global Biodiversity Information Facility API. R Package Version 0.9.2*.

Collevatti, R. G., de Castro, T. G., de Souza Lima, J. & de Campos Telles, M. P. (2012). Phylogeography of Tibouchina papyrus (Pohl) Toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and Evolution*, 2, 1024–1035. doi: 10.1002/ece3.236.

Costa, G. C., Hampe, A., Ledru, M. P., Martinez, P. A., Mazzochini, G. G., Shepard, D. B., … Carnaval, A. C. (2018). Biome stability in South America over the last 30 kyr: inferences from long-term vegetation dynamics and habitat modelling. *Global Ecology and Biogeography*, 27, 285–297. doi: 10.1111/geb.12694.

da Silva, J. M. C. & Bates, J. M. (2002). Biogeographic patterns and conservation in the South American Cerrado: a tropical savanna hotspot. *BioScience*, 52, 225.

da Silva, J. M. C. (1997). Endemic bird species and conservation in the Cerrado Region, South America. *Biodiversity and Conservation*, 6, 435–450. doi: 10.1023/A:1018368809116.

de Siqueira, M. F. & Peterson, A. T. (2003). Consequences of global climate change for geographic distributions of Cerrado tree species. *Biota Neotropica*, 3, 1–14.

Eckert, C. G., Samis, K. E. & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170–1188. doi: 10.1111/j.1365-294X.2007.03659.x.

Fiaschi, P. & Pirani, J. R. (2009). Review of plant biogeographic studies in Brazil. *Journal of Systematics and Evolution*, 47, 477–496. doi: 10.1111/j.1759-6831.2009.00046.x.

Guerin, G. R., Ruokolainen, L. & Lowe, A. J. (2015). A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution*, 6, 845–852. doi: 10.1111/2041-210X.12361.

Harrison, S. & Noss, R. (2017). Endemism hotspots are linked to stable climatic refugia. *Annals of Botany*, **119**, 207–214. doi: 10.1093/aob/mcw248.

Hijmans, R. J. & van Etten, J. (2014) *Raster: Geographic Data Analysis and Modeling. R Package Version 2*.

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. doi: 10.1002/joc.1276.

Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. (2013) *Dismo: Species Distribution Modeling. R Package Version 0.9–3*.

Hortal, J., Triantis, K. A., Meiri, S., Thébault, E. & Sfenthourakis, S. (2009). Island species richness increases with habitat diversity. *American Naturalist*, 174, E205–E217. doi: 10.1086/645085.

Klink, C. A. & Machado, R. B. (2005). Conservation of the Brazilian Cerrado. *Conservation Biology*, 19, 707–713. doi: 10.1111/j.1523-1739.2005.00702.x.

Ledru, M. P. (2002). Late Quaternary history and evolution of the Cerrados as revealed by palynological records. In: P.S. Oliveira, R.J. Marquis (Eds.), The Cerrados of Brazil: ecology and natural history of a neotropical savanna (Vol. 1, pp. 33–50). New York, NY: Columbia University Press.

Lima, J. S., Telles, M. P. C., Chaves, L. J., Lima-Ribeiro, M. S., Collevatti, R. G. (2017).Demographic stability and high historical connectivity explain the diversity of

a savanna tree species in the Quaternary. *Annals of Botany*, 119, 645–657. doi: 10.1093/aob/mcw257.

Marchant, R., Cleef, A., Harrison, S. P., Hooghiemstra, H., Markgraf, V., van Boxel, J., … Wille, M. (2009). Pollen-based biome reconstructions for Latin America at 0, 6000 and 18 000 radiocarbon years ago. *Climate of the Past*, 5, 725–767. doi: 10.5194/cp-5-725-2009.

Mendonça, R. C., Felfili, J. M., Walter, B. M. T., Júnior, M. C. S., Rezende, A. V., Filgueiras, T. S. & Nogueira, P. E. (1998). Flora vascular do Cerrado. In S. M. Sano & S. P. Almeida (Eds.), *Cerrado: Ambiente e Flora* (Vol. 1, pp. 289–556). Brasília, DF: EMBRAPA.

Nogueira, C., Ribeiro, S., Costa, G. C., Colli, G. R. (2011).Vicariance and endemism in a Neotropical savanna hotspot: distribution patterns of Cerrado squamate reptiles. *Journal of Biogeography*, 38, 1907–1922. doi: 10.1111/j.1365-2699.2011.02538.x.

Oliveira-Filho, A. T. (2017) *NeoTropTree, Flora arbórea da Região Neotropical: um banco de dados envolvendo biogeografia, diversidade e conservação*. Universidade Federal de Minas Gerais. <http://www.neotroptree.info>.

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H. & Hu, A. (2006). Simulating arctic climate warmth and icefield retreat in the Last Interglaciation. *Science*, 311, 1751–1753. doi: 10.1126/science.1120808.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Townsend Peterson, A. (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. doi: 10.1046/j.1365-2699.2000.00397.x.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026.

Pironon, S., Villellas, J., Morris, W. F., Doak, D. F. & García, M. B. (2015). Do geographic, climatic or historical ranges differentiate the performance of central versus peripheral populations? *Global Ecology and Biogeography*, 24, 611–620. doi: 10.1111/geb.12263.

Prado, D. E. & Gibbs, P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, 80, 902–927. doi: 10.2307/2399937.

Ribeiro, J. F. & Walter, B. M. T. (1998). Fitofisionomias do bioma Cerrado. In S.M. Sano & S.P. Almeida (Ed.), *Cerrado: ambiente e flora* (pp. 87–166). Brasília: Embrapa Cerrados.

Santos, M. G., Nogueira, C., Giugliano, L. G., Colli, G. R. (2014). Landscape evolution and phylogeography of Micrablepharus atticolus (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, 41, 1506–1519. doi: 10.1111/jbi.12291.

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. doi: 10.1016/S0006-3207(00)00085-9.

Slaton, M. R. (2015). The roles of disturbance, topography and climate in determining the leading and rear edges of population range limits. *Journal of Biogeography*, 42, 255–266. doi: 10.1111/jbi.12406.

Soberón, J. & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 9–32. doi: 10.17161/bi.v2i0.4.

Soule, M. (1973). The epistasis cycle: a theory of marginal populations. *Annual Review of Ecology and Systematics*, 4, 165–187. doi: 10.1146/annurev.es.04.110173.001121.

Steinbauer, M. J., Field, R., Grytnes, J. A., Trigas, P., Ah-Peng, C., Attorre, F., … Beierkuhnlein, C. (2016). Topography-driven isolation, speciation and a global increase of endemism with elevation. *Global Ecology and Biogeography*, 25, 1097–1107. doi: 10.1111/geb.12469.

Svenning, J. C., Normand, S. & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, 96, 1117–1127. doi: 10.1111/j.1365-2745.2008.01422.x.

Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020–2027.

Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F. & Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, 22, 1193–1213. doi: 10.1111/mec.12164.

van der Hammen, T. (1974). The Pleistocene changes of vegetation and climate in tropical South America. *Journal of Biogeography*, 1, 3. doi: 10.2307/3038066.

Welch, B. L. (1938). The significance of the difference Between two means when the population variances are unequal. *Biometrika*, 29, 350–362. doi: 10.1093/biomet/29.3-4.350.

Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. *Quaternary Science Reviews*, 30, 1630–1648. doi: 10.1016/j.quascirev.2011.03.009.

Werneck, F. P., Costa, G. C., Colli, G. R., Prado, D. E. & Sites Jr, J. W. (2011). Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidences. *Global Ecology and Biogeography*, 20, 272–288.

Werneck, F. P., Nogueira, C., Colli, G. R., Sites, J. W. & Costa, G. C. (2012). Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, 39, 1695–1706. doi: 10.1111/j.1365-2699.2012.02715.x.

Data accessibility statement: The weighted endemism raster data generated in this study are available as raster grids from the Pangaea database: https://issues.pangaea.de/browse/PDI-17312. The GBIF material list used for the endemism maps and models is available as supplementary material.

Table 1. Mean values and confidence intervals of the variances of 19 bioclimatic variables for valley and plateau environments in the Brazilian Cerrado based on 1000 bootstrap replicates. Variances were calculated by comparing cell values for 1000 random points within each landscape unity for the present, mid-Holocene (6,000 years before present), Last Glacial Maximum (22,000 years before present), and Last Interglacial (130,000 years before present). *t-values*: t-values calculated with a paired t-test. (*) p-value > 0.01.

		Valleys		Plateaus			
Variable	Description	Mean	95% ci	Mean	95% ci	t-value	
bio1	Mean Annual Temperature	386	6	392	7	27	*
bio2	Mean Diurnal Range	132	-19	113	-20	-221	*
bio3	Isothermality	22	-5	16	-5	-440	*
bio4	Temperature Seasonality	285582	-82242	203040	-82843	-538	*
bio5	Max. Temperature of Warmest Month	569	47	615	46	152	*
bio6	Min. Temperature of Coldest Month	909	-232	676	-234	-501	*
bio7	Annual Temperature Range	468	-139	329	-140	-469	*
bio8	Mean Temperature of Wettest Quarter	282	47	329	46	287	*
bio9	Mean Temperature of Driest Quarter	820	-115	705	-116	-252	*
bio10	Mean Temperature of Warmest Quarter	422	-6	416	-7	-31	*
bio11	Mean Temperature of Coldest Quarter	555	-29	525	-30	-81	*
bio12	Annual Precipitation	89629	-18262	71271	-18455	-373	*
bio13	Precipitation of Wettest Month	3440	-1174	2263	-1180	-764	*
bio14	Precipitation of Driest Month	137	-44	92	-45	-313	*
bio15	Precipitation Seasonality	194	-45	149	-45	-304	*
bio16	Precipitation of Wettest Quarter	25724	-9928	15775	-9970	-923	*
bio17	Precipitation of Driest Quarter	1871	-623	1245	-630	-338	*
bio18	Precipitation of Warmest Quarter	46822	-9568	37211	-9655	-431	*
bio19	Precipitation of Coldest Quarter	58886	-28905	29912	-29043	-821	*

Table 2. Generalized linear model for endemism, elevation, marginality and variances of bioclimatic variables among the Last Interglacial, Last Glacial Maximum, mid-Holocene and present.

	Description	Estimate	Std. error	t-value	Pr(>|t|)	
(Intercept)		9.14E-02	4.24E-02	2.156	0.0312	*
Elevation		1.30E-04	1.39E-05	9.406	<2e-16	***
Marginality	Distance to centroid	-1.13E-07	1.01E-08	-11.18	<2e-16	***
bio1	Mean Annual Temperature	1.09E-05	2.48E-04	0.044	0.9649	
bio2	Mean Diurnal Range	-2.09E-04	2.54E-04	-0.825	0.4093	
bio3	Isothermality	-4.65E-04	5.89E-04	-0.789	0.4299	
bio4	Temperature Seasonality	1.28E-08	4.26E-08	0.301	0.7634	
bio5	Max. Temperature of Warmest Month	-7.00E-05	3.97E-05	-1.763	0.0781	.
bio6	Min. Temperature of Coldest Month	-7.21E-05	7.96E-05	-0.906	0.3653	
bio7	Annual Temperature Range	8.14E-05	7.38E-05	1.104	0.2698	
bio8	Mean Temperature of Wettest Quarter	1.14E-04	8.39E-05	1.354	0.1758	
bio9	Mean Temperature of Driest Quarter	-4.92E-05	8.67E-05	-0.567	0.5709	
bio10	Mean Temperature of Warmest Quarter	4.50E-05	8.47E-05	0.532	0.5951	
bio11	Mean Temperature of Coldest Quarter	1.04E-04	1.01E-04	1.028	0.304	
bio12	Annual Precipitation	1.12E-06	1.08E-06	1.04	0.2985	
bio13	Precipitation of Wettest Month	4.85E-06	8.27E-06	0.586	0.5578	
bio14	Precipitation of Driest Month	-2.86E-04	3.44E-04	-0.833	0.4049	
bio15	Precipitation Seasonality	2.01E-05	2.77E-04	0.073	0.9421	
bio16	Precipitation of Wettest Quarter	-2.31E-06	2.11E-06	-1.097	0.273	
bio17	Precipitation of Driest Quarter	8.16E-06	2.94E-05	0.277	0.7817	
bio18	Precipitation of Warmest Quarter	2.17E-07	1.33E-07	1.63	0.1032	

bio19	Precipitation of Coldest Quarter	6.99E-08	9.61E-08	0.727	0.467

604

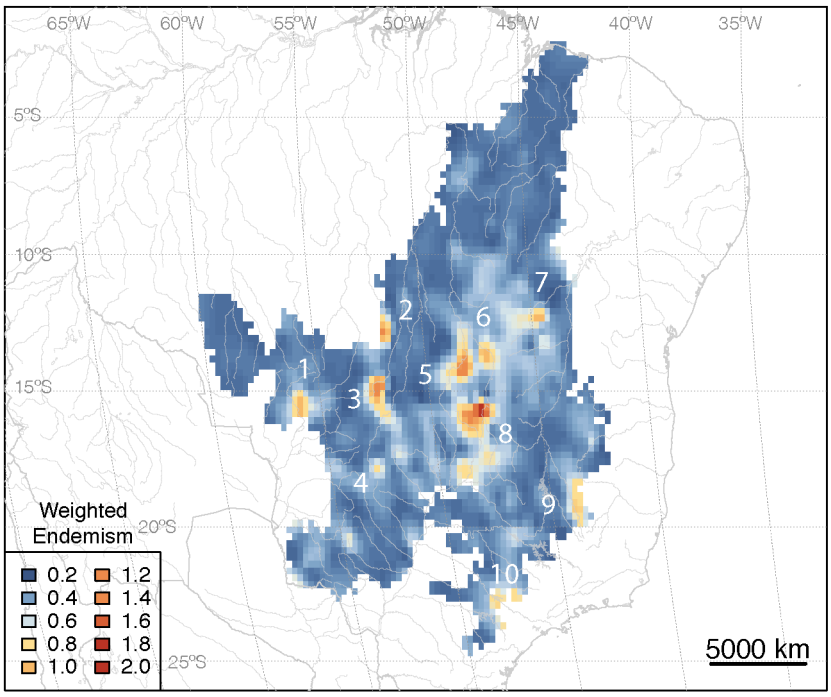

Figure 1. Weighted endemism (Guerín et al., 2015) map for woody species of the Cerrado domain. Grid cells are 0.25 × 0.25 degrees wide. Map in Mollweide projection. Numbers indicate geographic regions with higher endemism, identified as the following: (1) Chapada dos Guimarães; (2) Serra do Roncador; (3) Araguaia Valley; (4) Chapada das Emas; (5) Serra dos Pirineus; (6) Chapada dos Veadeiros; (7) Chapada do São Francisco; (8) Planalto do Distrito Federal; (9) Serra do Cipó; (10) Serra de Botucatu.

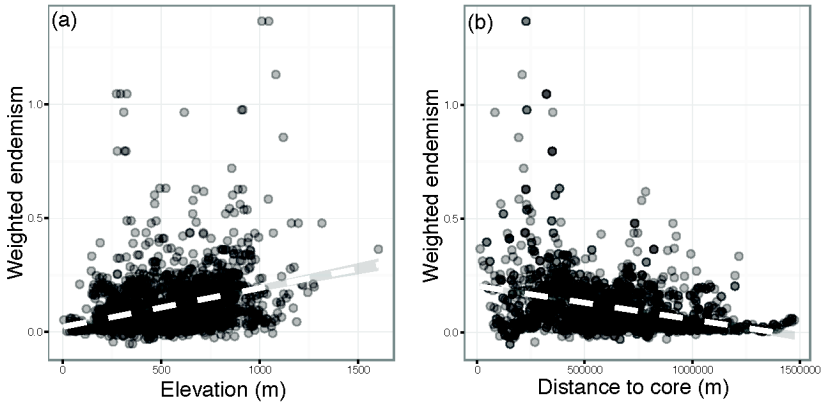

Figure 2. Linear models showing the correlation between weighted endemism, elevation, and distance to the core. (a) Weighted endemism vs. elevation; (b) Weighted endemism vs. distance to the core.

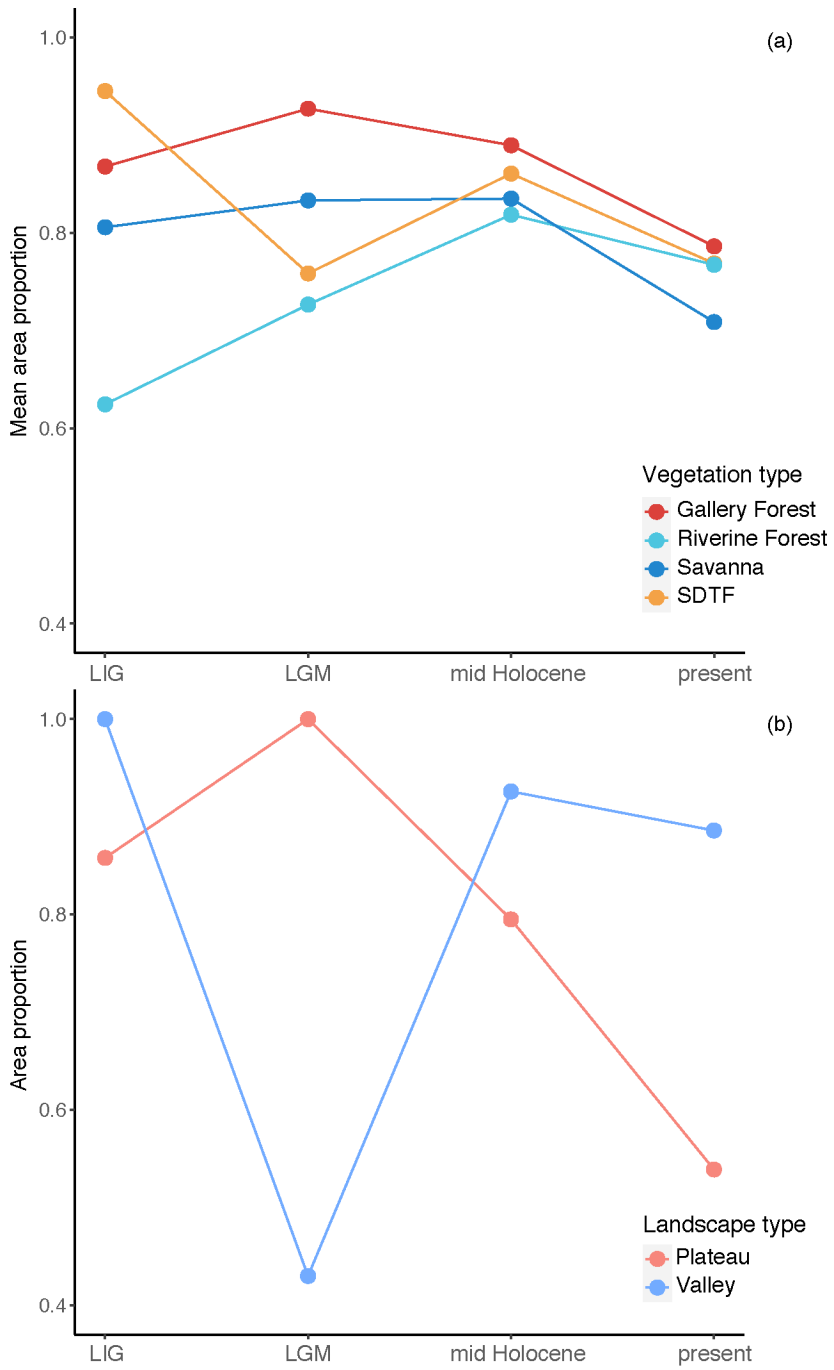

Figure 3. Potential suitable area for vegetation types of the Cerrado and environmental spaces for valley and plateau biotas based on simulated and real occurrence species dataset models of presence/absence generated with Maxent (Phillips et al., 2006). (a) Mean area proportion of species for each vegetation type calculated by the relative size of the predicted presence area compared to the observed maximum range for each landscape unit among the

four modeled scenarios. (b) Estimated environmental space proportion for plateaus and valleys among the four modeled scenarios.

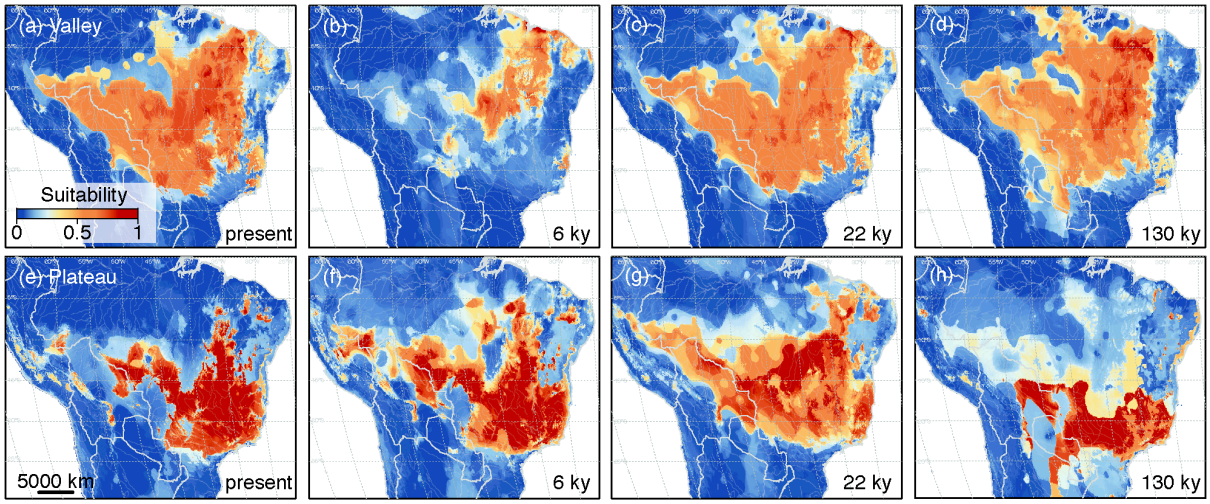

Figure 4. Potential suitable area for valley and plateau biotas of the Cerrado domain based on simulated dataset models of presence/absence generated with Maxent (Phillips et al., 2006). The area proportion is the relative size of the predicted presence compared to the maximum observed range for each landscape unit among the four modeled scenarios.

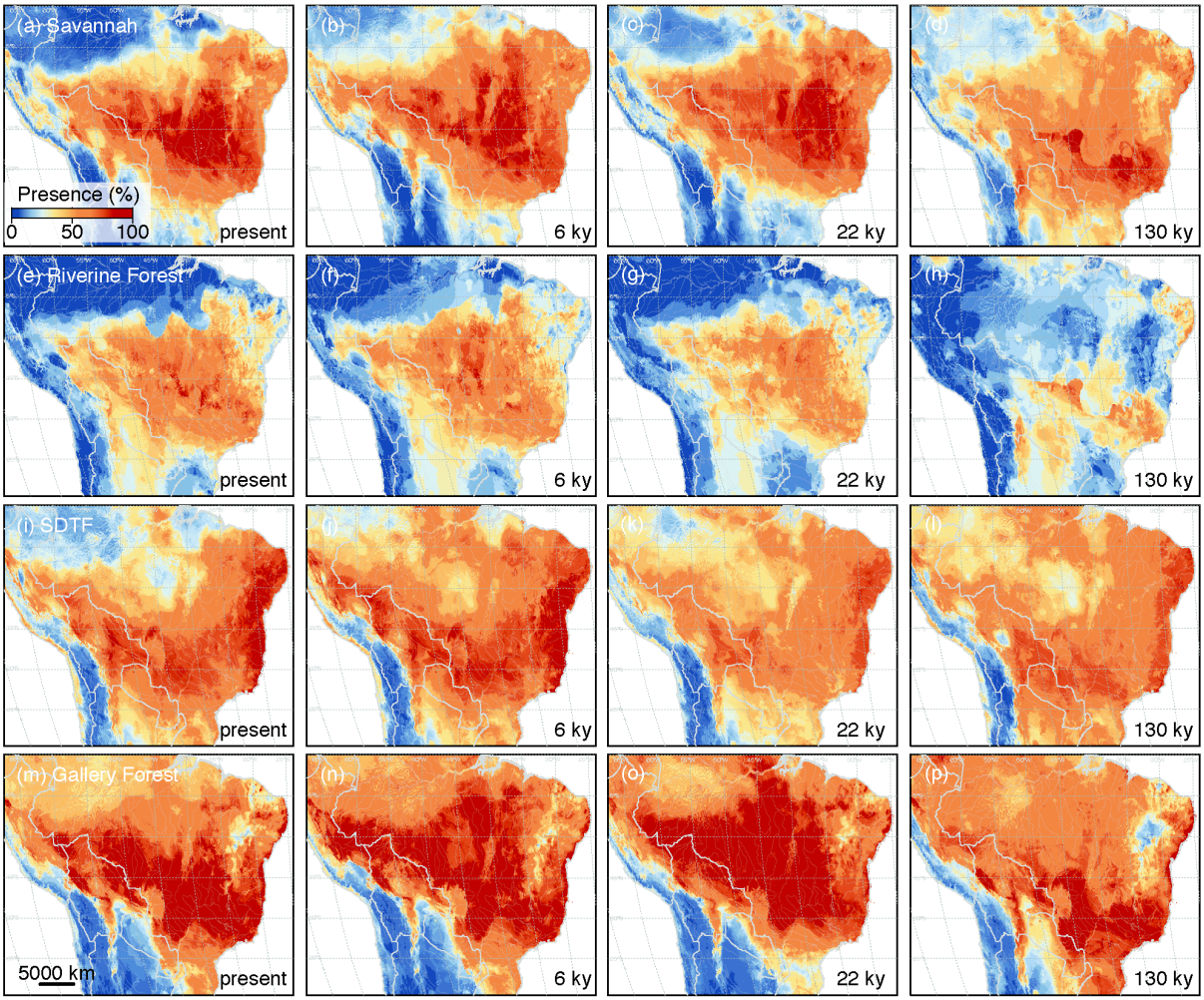

Figure 5. Potential suitable area for valley and plateau biotas of the Cerrado domain based on simulated dataset models of presence/absence generated with Maxent (Phillips et al., 2006). The area proportion is the relative size of the predicted presence compared to the maximum observed range for each landscape unit among the four modeled scenarios.