An Acad Bras Cienc (2022) 94(Suppl. 3): e20211019 DOI 10.1590/0001-3765202220211019
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

ECOSYSTEMS

Spatial structure of the Caatinga woody flora: abundance patterns have environmental, Pleistocene, and indigenous drivers

AUGUSTO C. SILVA & ALEXANDRE F. SOUZA

Abstract: Despite growing knowledge on the distribution and functioning of dryland vegetation types, their internal biodiversity structure (i.e., subregions) is much less studied. In the delineation of subregions, the use of species occurrence or abundance data may reveal different aspects of metacommunity structure. We revisit the issue of the bioregionalization of the woody flora of the Caatinga, the largest block of the dry forest and woodland biome in Latin America, using abundance data. We also evaluated the drivers of the spatial distribution of plant subregions: historical, current environmental and human effects. Using a K-means partition on interpolated NMDS axes, we identified 10 abundance subregions. Aridity, topography and soil, biome stability since the Pleistocene, and historical indigenous effects were retained by a Multinomial Logistic regression model, and their combined fractions explained most of the abundance variability in subregions. The subregions we present may support spatialized conservation and management decisions in the lack of detailed local data. The present results confirm the Caatinga woody flora broad composition patterns uncovered using presence-absence data in previous studies. Additionally, we found larger subregions than those identified with presence and absence data, suggesting the existence of oligarchies of dominant species in distinct parts of the Caatinga biome.

Key words: beta-diversity, biogeographical patterns, environmental gradients, human effects, macroecology, regionalization.

INTRODUCTION

Seasonally dry tropical forests and woodlands comprise non-fire adapted, tree-dominated, succulent-rich, grass-poor, dry tropical forests, woodlands, thickets, and bushlands, and harbours species capable of living with unpredictable rainfall patterns (Queiroz et al. 2017). The global extension of scrub and dry forests is similar to that of tropical moist forests (Bastin et al. 2017). They harbour species-rich ecosystems (Queiroz et al. 2017) and biodiversity hotspots (Myers et al. 2000), and contribute to nearly a third of the global carbon stock (Bastin et al. 2017, Pan et al. 2011). The diverse array of vegetation types that occur under dry climates are going through major changes, since by one hand they are fragile ecosystems susceptible to desertification due to overgrazing and deforestation, and by the other hand they are expanding geographically due to climate change, although accompanied by increased degradation (Pennington et al. 2000, Huang et al. 2015). Despite growing knowledge on the distribution and functioning of seasonally dry tropical forests and woodland vegetation types, their internal biodiversity structure (i.e., compositional subregions) is much less studied. Clear and data-driven identification of subregions detects natural discontinuities in the geographical distribution of the biota (Smith et al. 2018). Natural subregions may be regarded as mapped metacommunities and thus provide a framework to test hypotheses about evolution, ecology, and biogeography (Carnaval et al. 2014, Fine 2015). Plant subregions also function as habitat templates for the distribution and life-history evolution of animal species (Heikinheimo et al. 2012, Southwood 1977). Furthermore, subregion delimitation is a requisite for conservation planning and management due to the conservation of unique biotas (Dinerstein et al. 2017, Whittaker et al. 2005).

In the delineation of subregions, the use of species occurrence or abundance data may reveal different aspects of metacommunity structure. Most macroecological studies rely on occurrence data because they are more abundant and cover wider areas than abundance data (Gomes et al. 2018). They emphasize patterns related to rare and endemic species, which are given equal weight as dominants (Legendre & Legendre 2012). Abundance data, however, reflect dominance by a restricted suit of species and largely reflect the relative success of different ecological strategies that influence population size, resource use patterns, and adaptation to distinct habitats (Fine 2015, Grime & Pierce 2012, Southwood 1977). As stressful ecosystems with severe limitations to resource availability, arid and semi-arid regions frequently present plant communities strongly dominated by a few species (Kessler et al. 2011, Nascimento et al. 2003). In the Americas, recent studies have recognized the continental-scale identity of the dry forest biome (Slik et al. 2018), and the distinctiveness of its 12 nuclei (Banda et al. 2016). This biome is characterized by highly endemic floras, marked niche conservatism, and elevated beta-diversity among its nuclei (Fernandes et al. 2020, Hughes et al. 2013, Moro et al. 2016). Here we approach the issue of the bioregionalization of the Caatinga, the largest block of the dry forest and woodland biome (IBGE 2004, Moro et al. 2016, Queiroz et al. 2017) (Fig. 1a), from the point of view of abundance data. Silva & Souza (2018) have identified nine plant subregions in the Caatinga based on presence and abasence data only (Fig. 1b). We revisit the broader issue of the bioregionalization of the Caatinga by exploring an additional aspect, that of the spatial structure of species abundances in local communities, establishing abundance-based subregions and their determinants. Work in other vegetational domains has indicated that the use of abundance data can reveal compositional patterns different from those obtained with presence and absence data, increasing our understanding of the spatial organization of entire biogeographic regions (Cantidio & Souza 2019).

The Caatinga is a > 840,000 km^2 (IBGE 2004) vegetation complex in north-eastern South America covered by vegetation types ranging from succulent plant communities on rocky substrates to tall deciduous and semideciduous forests (Veloso et al. 1991, Banda et al. 2016, Moro et al. 2016, Queiroz et al. 2017) (Fig. 1c). Together, these diverse vegetation types harbour more than 4500 seed plant species and more than 900 endemic species (BFG 2015, Fernandes et al. 2020) that are spatially structured (Manhães et al. 2016, Cardoso et al. 2021). The basic ecological patterns of the Caatinga are only beginning to be studied (Santos et al. 2011), and protected areas within the domain cover only ca. 8.9% of its surface (Teixeira et al. 2021). Improved knowledge of the spatial structure of this domain is key for both macroecological understanding of semiarid regions and conservation planning. Earlier attempts to discriminate Caatinga plant regions were qualitative and relied heavily on physiognomic subdivisions (Fernandes & Bezerra 1990, Moro et al. 2015a, Santos et al.

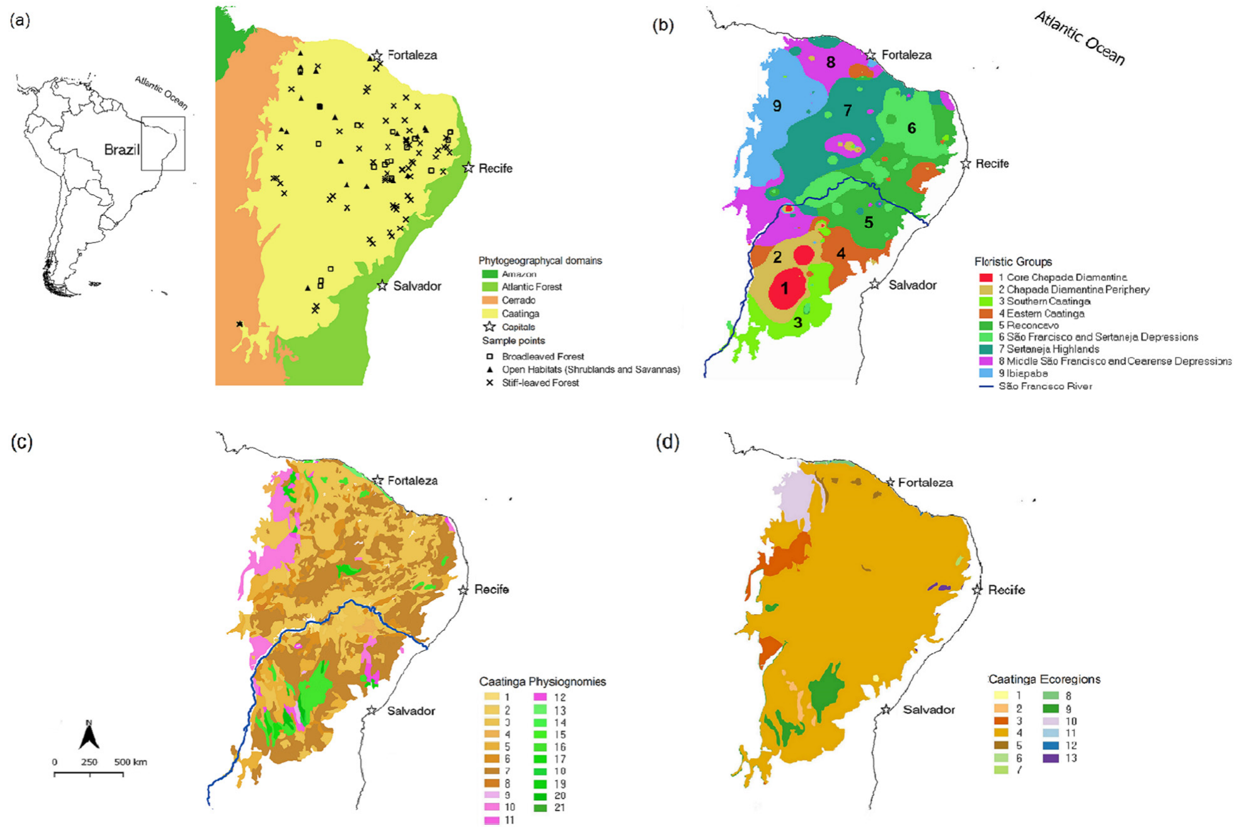

Figure 1. Distribution pattern of vegetation in Caatinga domain. a) sample points with corresponding vegetation types (open square = Broadleaved forest, x symbol = Stiff-leaved forest, open triangle = Open habitats – Shrublands and Savannas). Note that the dark square symbol near the southwest of the state of Ceará corresponds to overlapping white squares. ; b) Caatinga Floristic Groups (Silva & Souza 2018); c) Vegetation types (1 – Semiarid bushy grassland lower plains, 2 – Deciduous stiff-leaved scrub in semiarid lower plains, 3 – Deciduous stiff-leaved scrub in semiarid upper plains , 4 – Deciduous stiff-leaved scrub in semiarid lower highlands, 5 – Semideciduous stiff-leaved scrub in seasonal upper plains, 6 – Deciduous stiff-leaved dwarf-forest in semiarid upper plains, 7 – Deciduous stiff-leaved forest insemiarid lower plains, 8 – Deciduous stiff-leaved forest in upper plains, 9 – Seasonal bushy grassland upper plains, 10 – Semideciduous parkland savanna in seasonal upper plains, 11 – Semideciduous forest savanna in seasonal lower plains, 12 – Semideciduous shrubland savanna in seasonal lower plains, 13 – Semideciduous broadleaved scrub in maritime coastal, 14 – Semideciduous broadleaved scrub in seasonal lower plains, 15 – Semideciduous broadleaved forest in maritime coastal, 16 – Deciduous broadleaved forest in seasoanal upper plains, 17 – Deciduous broadleaved forest in seasoanal lower highlands, 18 – Semideciduous broadleaved forest in seasonal upper plains, 19 – Semideciduous broadleaved forest in seasonal lower highlands, 20 – Semideciduous broadleaved forest in seasonal upper highlands, 21 – Evergreens broadleaved forest in seasonal lower highlands) adapted from (Oliveira-Filho 2009); d) Caatinga Ecoregions (1 – Bahia interior forest, 2 – Campos rupestres montane savana, 3 – Cerrado, 4 – Caatinga, 5 – Caatinga enclaves moist forest, 6 – Atlantic coast restingas, 7 – Pernambuco coastal forest, 8 – Northeast Brazil restingas, 9 – Brazilian atlantic dry forest, 10 – Maranhão Babaçu forest, 11 – Amazon-Orinoco-Southern Caribbean mangroves, 12 – Southern Atlatic Brazilian mangroves, 13 – Pernambuco interior forest) (Dinerstein et al. 2017).

2012), on inferred environmental proxies like topography (lowlands vs. highlands) and soil type (sedimentary vs. crystalline basement) (Moro et al. 2016, Queiroz et al. 2017), or on expert-based landscape subdivisions encompassing geomorphological, vegetation type, animal, and plant knowledge, like the WWF ecoregion approach (Dinerstein et al. 2017, Velloso et al. 2002) (Fig. 1d). Physiognomy, soils, topography and floristic composition, however, are only indirectly related (Saiter et al. 2016, Silva & Souza 2018), and reliance on these environmental features as direct criteria for floristic subdivision may hide important biodiversity structures (Saiter et al. 2016, Silva & Souza 2018).

Our objective was to provide a data-driven identification of the spatial abundance structure of the Caatinga based on statistically reproducible methods that explicitly consider the degree of compositional separation between regions. Additionally, we asked: what are the drivers of the spatial distribution of abundances across the Caatinga plant subregions? Distinct sets of drivers may determine such patterns. First, the stability of the deciduous vegetation since the Last Glacial Maximum (21000 years before present [BP]) could be related to the species composition of each locality (Moura et al. 2017, Silva & Souza 2018). This is because during the Pleistocene the alternation of glacial and inter-glacial periods caused adaptive radiation and allopatric/parapatric speciation related to altitudinal shifts and fragmentation of species ranges (Collevatti et al. 2012, Turchetto-Zolet et al. 2013). Second, niche-based species sorting by current environmental differences found across regions may also produce spatially recognizable metacommunity structure (Vellend 2010). The basic mechanism is that the amount of energy and nutrients available within a region mediates productivity (Mazzochini et al. 2019) and population size, increasing speciation and reducing extinction probabilities (Fine 2015). Ecological strategies mediate species success along resource and abiotic conditions gradients in arid and semiarid regions (Frenette-Dussault et al. 2012, Grime & Pierce 2012). In tropical semiarid regions, factors known to mediate productivity and niche affinities include ridges and mountain ranges that attenuate aridity due to reduced temperatures and precipitation patterns (Dick & Wright 2005, Kreft & Jetz 2007, Rodal et al. 2008, Turchetto-Zolet et al. 2013, Vieira et al. 2015), soil texture and nutrition, and the balance between heat and rainfall (i.e., aridity), all associated with significant changes in species composition (González-Orozco et al. 2014, Moro et al. 2015b, 2016, Neves et al. 2015, Queiroz et al. 2017, Rito et al. 2017a, Silva & Souza 2018). Third, current and pre-historical indigenous human activities like selective logging, deforestation, cattle grazing, and fire may strongly distort biogeographical patterns and alter species composition and dominance patterns in seasonally dry tropical forests (Neves et al. 2015, Ribeiro et al. 2016, Rito et al. 2017b, Medeiros et al. 2018).

MATERIALS AND METHODS

Data

Study area

We adopted a broad, ecologically meaningfull concept of Caatinga, which encompasses not only the semiarid Seasonally Dry Forest and Woodland biome but also savannah, semideciduous, and enclaves of rupestrian fields and 'brejos de altitude' evergreen forest that occur in the region (IBGE 2004, Moro et al. 2016) . This is because diferent vegetation types share a large number of species (Fernandes et al. 2020), frequently present phenopitic plasticity of the same species pool (Cantidio & Souza 2019, Silva et al. 2019a), and are integrated through evolutionary and ecossistem functioning processes (Neves et al. 2020). The Caatinga is located in north-eastern South America (Fig. 1a) on top of Pre-Cambrian granitic and gneissic basement exposed by erosion and in the quartz sandy terrains or limestone outcrops of some large sedimentary basins that occur in the region. During the Tertiary nutrient-rich, stony, and shallow soils developed and are referred to as crystalline terrains (reviews in Moro et al. 2016 and Queiroz et al. 2017). Very high mountain rainges are lacking and the altitudes of lowlands vary around 100-1000 m (Ross 2011). The Borborema, Ibiapaba and Araripe mountain ranges have altitudes that reach over 800 m, while the Chapada Diamantina, which is capped by horizontal strata of sandstone, has altitudes that reach over 2000 m (Ross 2011) (see Additional Information in Figshare at doi:10.6084/m9.figshare.17086685, Fig. S1). Average temperatures are between 24°C and 26°C. The domain limits match areas below the 1000 mm rainfall isohyet (Nimer 1972), and annual rainfall is reduced to up to 500 mm per year in the driest areas. Most of the domain presents semi-arid (Bs) or tropical with dry summer (As) Köppen climate type, (Alvares et al. 2013).

Species and environmental data

We compiled data on species abundances for 109 localities from from papers published between 1969 and 2018 (Additional Information - Table SI). We did not set any spatial proximity limits to localities obtained from these sources. The habit of each species was checked in the database of the Flora do Brasil 2020 project, and only species of shrub and trees, including palms, were included (referred throughout the text as woody flora or woody species). We adopted these criteria because the ecological strategies and structural and functional role of herbaceous and subshrub species are too distinct from those of larger woody species. Abundance data were the number of sampled individuals and came from published plot inventories only. Nomenclature was reviewed in May 2019 using the Flora do Brasil project (http://floradobrasil.jbrj.gov.br/). All non-native species were excluded from analyses. Species data were organized in a species-by-locality matrix containing abundance data converted into relative abundances (more details in the Additional Information).

For each locality, we obtained 23 variables related to climatic, soil, hydrological, elevation, vegetation historical stability, and human footprint. Following Dormann et al. (2013), we excluded from the analysis selected variables from all strongly correlated pairs of predictors ($r > 0.75$, Table SII) (Kumar & Stohlgren 2009), resulting in a final subset of 15 explanatory variables (Table SIII). In order to evaluate the effects of current environmental factors on the abundance structure of the Caatinga, we used climatic, soil, and elevation variables. Climatic variables included annual mean temperature (AMT), temperature annual range (TAR), aridity index (AI) (Trabucco & Zomer 2019), and precipitation seasonality (PS). Soil variables included soil depth, density, sand fraction, silte fraction, organic carbon content (CO), acidity (pH-kcl), and nutrient content (as cation exchange capacity, CEC). Elevation and elevation roughness were included as proxies for environmental heterogeneity not accounted for by the other variables. In order to evaluate the effects of biome stability since the Pleistocene on the abundance structure of the Caatinga, we used vegetation stability since the last glacial maximum (21000 years before present [21 kyr BP]) as provided by Costa et al. (2018). In order to evaluate the effects of past and current human-related factors on the abundance structure of the Caatinga, we used the Human Footprint Index (Venter et al. 2016), which summarizes current human activities. We also measured the distance of each locality to the nearest archaeological site, which was used as a proxy of historical human influence (further details and references in the Additional Information).

Data analysis

Caatinga Abundance Subregions

The bioregionalization we performed wa based on modelling of the compositional dissimilarity at the community level with no constraining by environmental variables following recent work on the subject (Cantidio & Souza 2019, Moura et al. 2017, Silva & Souza 2018, Silva-Souza & Souza 2020). The sub-regions we identified were produced by a partition of contiguous surfaces of estimated community dissimilarities. The procedure began by the calculation of a dissimilarity matrix between each pair of sites (point data), for which we used the $d_{BC\text{-}bal}$ index, that measures the turnover component only of β-diversity present in the Bray-Curtis dissimilarity (Baselga 2013). We employed a step-across procedure on the dissimilarity matrix aiming at correcting for loss of sensitivity of the dissimilarity indices caused when the amount of community change approaches or equals 1.0 (Smith et al. 2001, Tuomisto et al. 2012).

We depicted the dissimilarity matrices in ordination space through non-metric multidimensional scaling ordinations (NMDS) with four axes (Fig. S2). The NMDS has few limiting premises, detects nonlinear patters, and maximizes floristic variation information capture (McCune & Grace 2002). We then interpolated the NMDS site scores over the entire extent of the Caatinga, creating a surface of community dissimilarities. The interpolated NMDS axes (Fig. S3 and S4) represented the main Caatinga woody plant abundance compositional gradients. We used a fine spatial resolution (2.5 arc-min, ca. 5 km) to conserve spatial data structure as much as possible. See the Additional Information for further details on these procedures.

Subregions were identified and mapped at the same time by stacking up and then partitioning the interpolated NMDS axes using K-means, which is a non-hierarchical classification method (Fig. S5). Cells were thus grouped based on overall compositional dissimilarity, not on actual species distributions. The application of K-means requires a previously established number of clusters (k), which was determined through the application of the L-method to a matrix containing all axes scores of a PCoA. The L-method is a kind of "elbow" method which employs statistical regressions to determine the best number of groups (Salvador & Chan 2004).

It would be possible that the spatial patterns obtained from our abundance data were produced simply by the spatial distribution of species occurrences. To evaluate this possibility, we also run the bioregionalization procedure on the abundance dataset after converting the abundances into presence-absence values. Indicator species analysis was used to identify the species associated with different subregions (Cáceres & Legendre 2009). All analyses were run with the software R 3.4.3 (R-Core-Team 2017). See the Additional Information for further details for more details on the methods we used and their implementation. The input data files, the R script that runs the bioregionalization, and the shapefile of tine final bioregionalization are openly available in Figshare at doi:10.6084/m9.figshare.17086685.

The bioregionalization just described identified and maped subregions but did not establish compositional relationships between the subregions. To accomplish this, we pooled the species data by subregion and then performed hierarchical clustering analyses on the pooled matrices using the $d_{BC\text{-}bal}$ dissimilarities and the UPGMA linkage method (Kreft & Jetz 2010). We regarded as significant the clusters corresponding to at least 90% of the variation in the ecological distances (Holt et al. 2013).

Drivers of subregion distribution

To determine the drivers of the subregions, which are categorical variables with multiple levels, we employed Multinomial Logistic Regression (Additional information). Following Zuur et al. (2009), we first built a full model including all variables from the soil, elevation, current climate, historical stability, human effects datasets. We then used a stepwise backward model selection based on Akaike's information criterion (AIC) to identify the smallest set of variables explaining the deviance in the abundance subregions (Burnham & Anderson 2002). In order to control for spatial autocorrelation in the data, 14 Moran Eigenvector Maps were included as explanatory variables alongside the environmental variables in the final model (Borcard et al. 2011). The regression model with the lowest AIC had its deviance partitioned to obtain the unique and shared contribution of the different predictor sets (Moura et al. 2017).

RESULTS

Caatinga Biogeographic Subregions Based on Species Abundance

We found a total of 758 native woody species belonging to 315 genera and 79 botanical families in the 109 localities. Species restricted to a single locality totalled 426 (56%). The average number of species per locality was 27.07 (SD = 14.09; median = 28; range: 10 - 81). The NMDS stress value was 0.147. The K-means-based L-method identified 10 abundance subregions as the best solution for our dataset (Fig. 2, Fig. S5). The subregions' names we suggested try to associate regional characteristics to each region, but extrapolation procedures made by computer will result in limits that are different from actual ground limits in more than one case. Abundance subregions 2 and 8 were the most speciose, but subregion 2, in the Core Chapada Diamantina, had by far the highest proportion of restricted species (52%). See Table S4 for information on the extension, number of species, and environmental averages of the subregions. The map produced by analyzing abundance data transformed into presence and absence was revealed a spatial arrangement of subregions distinct from that produced by the same data but including species abundance information (Fig. S6). This indicates that the spatial patterns identified in Figure 2 are in fact the result of species abundance patterns and not just their occurrences. The R script that runs the bioregionalization procedure and the shapefile of the Caatinga abundance subregions are openly available at Figshare . The resulting map is presented as a macroecological hypothesis of the subdivision of the Caatinga woody species communities that can be improved with additional data in the future. The Indicator species analyses obtained a total of 111 species for the 10 subregions (Table SV).

Hierarquical relations

The UPGMA classification using pooled data by subregion recognized five clusters of subregions (Fig. 3). A broad northwest-southeast division was found, with compositional differentiation of the southern Caatinga around the Chapada Diamantina region. A large block of subregions dominates most of central Caatinga.

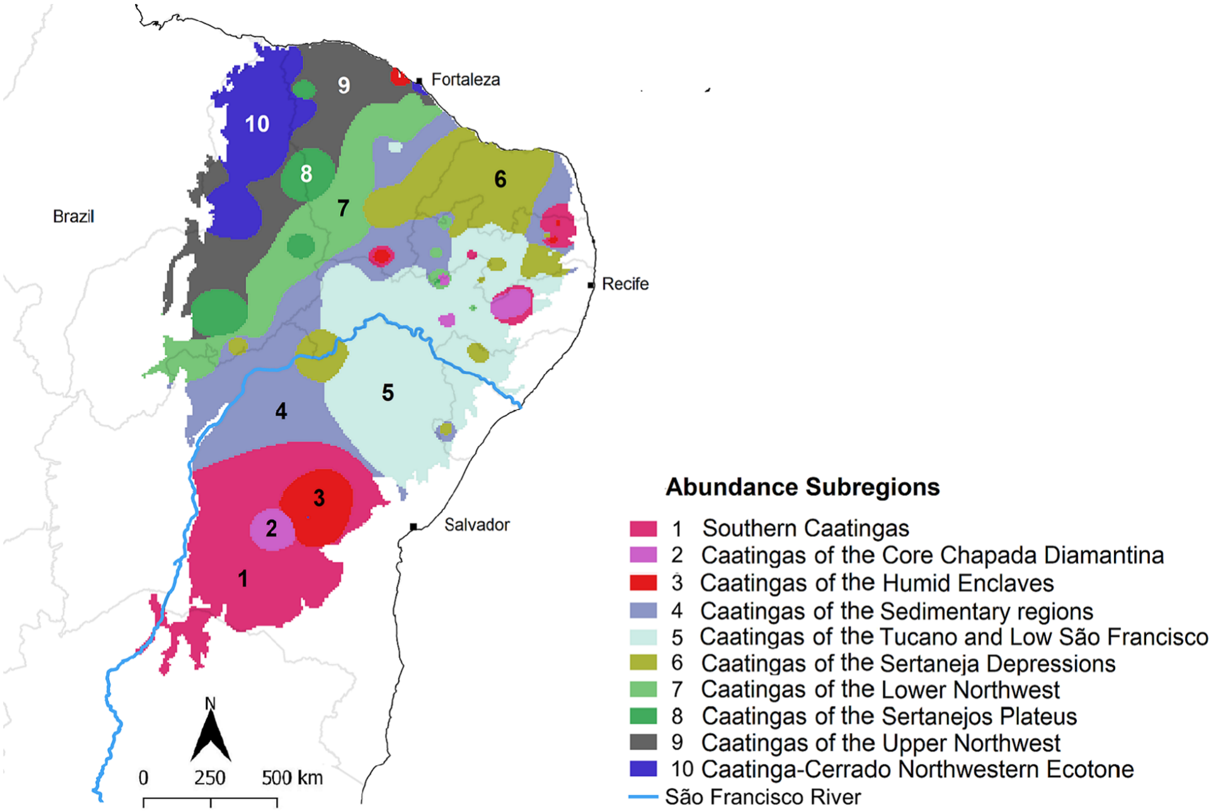

Figure 2. Tree compositional variation in the Seasonally Tropical Dry Forest of South America (STDF). The colours of the map have no absolute meaning, only the colour differences between locations are meaningful. Woody plant regionalization into 10 abundance subregions based on K-means partitioning of the interpolated NMDS axes.

Drivers of subregion distribution

The best multinomial logistic model included aridity, elevation, elevation roughness, historical biome stability, soil silt content, sand content, soil depth, soil nutrition, soil acidity, as well as the distance to the nearest archaeological site. This means that these variables were related to the occurrence probability of the subregions we found. Total explained deviance, including the Moran Eigenvector Maps that described the spatial structure of the compositional data was 0.95%. The variance partition analysis revealed that by far most of the model deviance was explained by the joint effect of the topography and soil, climatic, and human-related variables combined fractions (82.0%), with pure and pairwise fractions explaining negligible portions of the total deviance (Fig. 4).

The distribution of the explanatory variables selected in the final model revealed different environmental profiles for the subregions (Fig. 5). Southern Caatingas are located on plateaus with very variable topography, silty and clayey soils relatively poor in exchangeable cations, abundant archaeological sites, and a history of variable degrees of vegetation stability since the Lasta Glacial Maximum. The Core Chapada Diamantina subregion, which included an important disjunction in the Borborema highlands in Pernambuco, had the highest elevations, soils with relatively high cation exchange capacity, and high vegetation stability since the Last Glacial Maximum. Its northern neighbor Caatingas of the Humid Enclaves stood out for its rugged terrain, higher humidity, and relatively fertile soils. The Caatingas of the

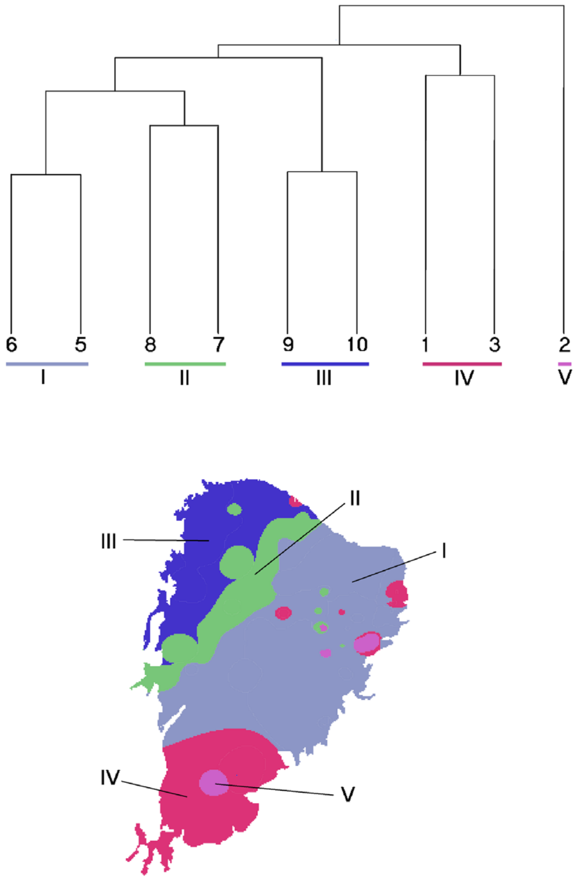

Figure 3. Clustering of woody plant groups of the Seasonally Tropical Dry Forest South America (STDF). Hierarchical classification of the 10 abundance subregions using $d_{BC\text{-}bal}$ dissimilarities and UPGMA as the linkage method. These analyses highlight the relationships between the subregions identified by the K-means analyses and identified significant higher-level clusters of groups. Horizontal bars indicate the final higher-level clusters of groups. Maps of the main clusters of subregions are shown for abundance datasets. Subregion 4 was not included because it was identified based on compositional distance only, with no associated local communities. Subregion 4 was tentatively included into the UPGMA zone I.

Tucano and Low São Francisco occurred over low-lying terrain of shallow, acidic, fertile soils with high cation exchange capacity, under the region's most arid climate, and with a history of high vegetation cover instability since the Last Glacial Maximum. The Caatingas of the Sertaneja Depressions to the nort was similar to the Caatingas of the Tucano and Low São Francisco in aridity, topography, and soil properties, except for that it displayed much wider variation in soild depth and biome stability, and had an increased proximity of indigenous archaeological sites. The Caatingas of the Lower Northwest include flat areas as well as Sertanejo Residual Pateaus such as the Serra de Baturité, relatively acidic silty soils poor in exchangeable cations. The Caatingas of the Sertanejos Plateus covers parts of the the Ibiapaba and associated mountain ranges. This subregion has high elevations, sandy or silty acidic soils, which are poor with low cation exchange capacity. This subregion has had fairly stable vegetation cover since the Last Glacial Maximum. The Caatingas of the Upper Northwest occur mostly on Ceará and Piauí lowlands on sandy soils distant from any indigenous archaeological sites. Finally, the Caatingas of the Caatinga-Cerrado Ecotone covers mostly the humid lowlands of the Parnaiba River basin in Piauí with deep acidic sandy soils.

DISCUSSION

Here we present a woody flora regionalization of the entire Caatinga seasonally dry forest and woodland using a data-driven approach based on abundance, and therefore dominance patterns of community composition. The all-encompassing definition of the Caatinga domain we adopted, which encompasses not only the Seasonally Dry Forest and Woodland biome but also savannah, semideciduous, and evergreen forest enclaves like 'brejos de altitude' (Queiroz et al. 2017) agrees with previous work that indicated interrelationships of the flora of the entire region (Moro et al. 2016, Cardoso et al. 2021). It also provides an ecological and biogeographical framework to our results, since diferent vegetation types share many species

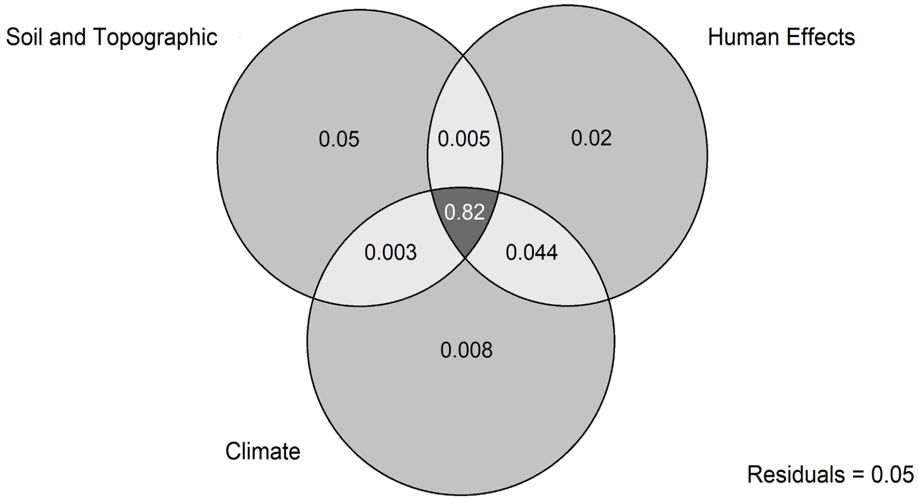

Figure 4. **Variance partition explained by each set of variables selected in the best Stepwise Multiple Logistic Regression model for data of abundance.**

(Fernandes et al. 2020), frequently represent phenopitic plasticity of the same species pool (Cantidio & Souza 2019, Silva et al. 2019a), and are integrated through evolutionary and ecossistem functioning processes (Neves et al. 2020, Cardoso et al. 2021). It is worth noting that we kept all native species occurring in the sampling localities regardless of their occurrence in other South American domains or ecosystems, and with this we ensured an inclusive and ecologically meaningful representation of the Caatinga flora.

Comparison with other regionalization schemes

Our compositional abundance-based results differed from previous attempts to map the internal spatial structure of the Caatinga domain based on vegetation types or the combined animal, plant, and habitat patterns (Dinerstein et al. 2017, Olson et al. 2001, Velloso et al. 2002), including the official Brazilian vegetation map (Veloso et al. 1991). This confirms the importance of seeking the identification of plant macroecological units separately from animal patterns and from physiognomic variation, as pointed out by our recent work (Silva-Souza & Souza 2020, Silva & Souza 2018). The subregions we found based on species abundances bear notable differences and resemblances to the ones obtained by Silva & Souza (2018) for the Caatinga with presence-absence data only, generated using the same methods but a considerably larger number of localities (260 localities in their case vs. 109 in the present analysis) (See 'Methodological issues' in Additional Information for considerations on the methodological limitations of the present contribution). Resemblances include the distinctiveness of southern Caatinga with a differentiated Chapada Diamantina nucleus, a broad north-south subregion at the northwestern Cerrado border, two large eastern subregions (numbered 5 and 6 in both works) and an intermediary north-south subregion numbered 7 in both works. These similarities are striking given we have dropped hundreds of species during the revision of species names for the present work, and excluded cultivated, naturalized, and invasive species. This confirms that Silva & Souza's (2018) analyses identified natural patterns in the spatial organization of the Caatinga flora, despite the retention of cultivated, naturalized, and invasive species and some issues regarding species names (Fernandes et al. 2020), and highlights that macroecological and

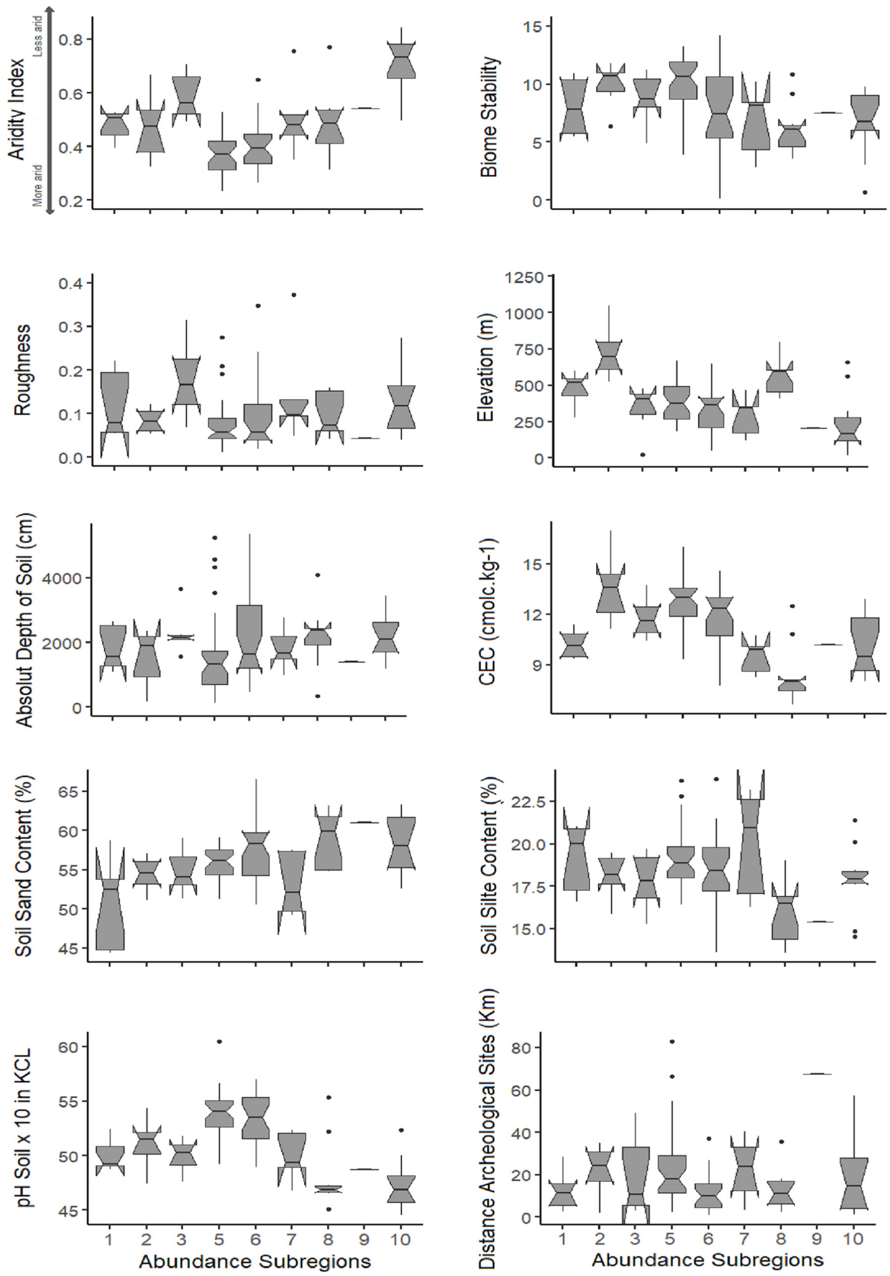

Figure 5. Comparison of the distribution of environmental variables retained in the final logistic multinomial model among compositional subregions of the Caatinga based on local abundances of shrub and tree species. Lower Aridity Index values indicate more arid conditions and lower Biome Stability (arrow direction) values indicate more stable vegetation types since the Last Glacial Maximum. Subregion 4 was not included because it was identified based on compositional distance only, with no associated local communities. Central horizontal bars represent median values; notches surrounding the median represent median 95% confidence intervals. Non-overlapping of notches indicates significant difference at 95% confidence level.

biogeographical analyses may identify robust patterns that are not apparent in checklists. Despite these similarities, in the occurrence-based map there was a greater fragmentation of subregions, with more numerous disjunct areas likely produced by the much more scattered distribution of rare species. Southern Caatinga, for instance, possibly related to the Chapada Diamantina flora but also to the taller stiff-leaved (arboreal caatinga) flora of norther Minas Gerais and southern Bahia, which is partly associated with extensive limestone deposits (Santos et al. 2012). This area is mostly covered by subregion 1 in the present analysis but split more evenly between three occurrence subregions in Silva & Souza results, indicating that distinct suits of rare species occur in different parts of that area, while a smaller suit of species like *Leucochloron* *limae* (Fabaceae), *Pseudobombax simplicifolium* (Malvaceae), and *Coccoloba alnifolia* (Polygonaceae) dominate regional communities. Strictly speaking, it is not possible to know to what extent the differences we found with abundance data compared to Silva & Souza's (2018) results with presence and absence data are due to differences in the number of localities studied. However, our evaluation of the results obtained with abundance data converted to presence and absence (Fig. S6) suggests that the dominance and rarity structure in abundance data produce markedly different patterns than those produced with presence and absence data alone. Our results also highlight the fact that analysis of the spatial organization of species abundances can produce insights distinct from those obtained when considering phylogenetic and functional aspects of herbaceous and woody species together (Cardoso et al. 2021), suggesting that different approaches are needed for a more complete understanding of the biodiversity structure of a rich region like the Caatinga.The correspondence between physiognomies and subregions was limited, but consistent in some areas. Some inland rainforest enclaves on isolated highlands and inselbergs regionally known as *brejos de altitude* appeared as disjunctions of the southern subregions 1, 2, and 3, and a predominantly savannah vegetation matched the disjunct occurrence subregions 8, 9, and 10, with subregions 9 and 10 corresponding to the vegetational complex of *Campo Maior* (Fernandes et al. 2020, Velloso et al. 2002). However, our results confirm that vegetation maps should not be used as proxies to plant diversity subregions, due to a prevalent mismatch between vegetation types and biodiversity subregions, as has been found repeatedly in the last years (Cantidio & Souza 2019, Silva & Souza 2018). For example, stiff-leaved deciduous forests and scrub cover most of the Caatinga domain, but different compositional subregions were identified within this single physiognomy. On the other hand, entire vegetation units like riverine forests and coastal *restinga* heath vegetation did not correspond to well-defined compositional subregions. The mismatch between composition and physiognomy supports recent claims that species occurring across different physiognomies proably display considerable phenotypic plasticity in traits like deciduousness and height (Silva et al. 2019a). Therefore, pysiognomic variation should probably be viewed as an ecosystem-level result of variation in productivity and stress, and not always a good indicator of floristic composition. Despite the points mentioned above, we do not advocate the replacement of multi-taxa or physiognomy-based regionalization methods by composition-based subregions. We believe they represent complementary approaches that are applicable to different situations.

Regional patterns

Our abundance subregions largely portray dominance patterns mediated by variation in ecological strategies (Arellano et al. 2017, Grime & Pierce 2012) that are related to ecosystem structure and function. Fifty-six percent of the species in our sample were restricted to a single locality, what implies widespread small population sizes and promotes high levels of β diversity.

The origins of species inhabiting diferent Caatinga subregions are hardly uniform. Several Caatinga lineages originated from Mesoamerican seasonally dry forests, with subsequent diversification during pre-Pleistocenic times (Queiroz et al. 2017). Evolution in the Seasonally Dry Forest and Woodland nuclei have been strongly shaped by niche conservatism and dispersal limitation, imprinting current low phylogenetic diversity to the region, largely because of stressful conditions that pose strict limits to the establishment of immigrant lineages which are not pre-adapted to long and erratic dry seasons (Neves et al. 2020, Pennington et al. 2006, Queiroz et al. 2017). Likely examples of subregions rich in evolutionarily unique lineages are our Core Chapada Diamantina subregion (subregion 2) as well as those prone to the severe semiarid climates like subregions 4, 5, and 6. Our results confirmed the key role of the Chapada Diamantina as a center of diversity and endemism (Moro et al. 2016, Queiroz et al. 2017), with hundreds of unique genera and species (Velloso et al. 2002, Zappi et al. 2003), in agreement with estimated species richness and restricted species distribution presented by Manhães et al. (2016). The Core Chapada Diamantina subregion included a locality whose vegetation type corresponds to a rupestrian field, and other points corresponding to savanna and dry forest vegetation formations. This means that the high proportion of species restricted to this subregion occurred in a mosaic of vegetational types including stiff-leaved dry forest, savannas, and rupestrian fields. This has been attributed to evolutionary divergence promoted by gene flow reduction caused by barriers as valleys, rifts, and steep walls (Dick & Wright 2005, Turchetto-Zolet et al. 2013, Vieira et al. 2015) and is reinforced by the separation of the Chapada Diamantina subregion to the south of the Caatinga from all other subregions in our UPGMA analysis.

The Caatinga borders highly biodiverse biogeographical provinces as the Atlantic rain forest and the Cerrado Savannah and is close to the eastern Amazon. During the Pleistocene, vast extensions of the Caatinga have been estimated to have been covered by moist tropical forest of mixed Atlantic and Amazon origins (Costa et al. 2018, Silveira et al. 2019), which are likely to have contributed to the Caatinga Domain species pool (Banda et al. 2016) and biogeographical affinities of species in high-altitude rain forest refugia (Santos et al. 2007). Some of our subregions, therefore, are likely ecotones to different Cerrado, Amazon, and Atlantic Forest subregions, and vegetation types other than the stiff-leaved forests and scrubs known as Caatinga vegetation are likely to be more common in these transitional areas (e.g., the Cerrado savanna areas in the Caatinga-Cerrao Ecotone subregion). Probable examples were the subregions clustered in our UPGMA zone III near the northern Cerrado, as well as subregion 1 surrounding the Chapada Diamantina, that have been described as Arboreal Caatinga (Moro et al. 2016) and that have strong floristic links with the Cerrado savannah (Santos et al. 2012). Abundance subregion 1 and 5 are closer to the northeastern Atlantic forest subregions (Cantidio & Souza 2019), and was repeatedly covered by Atlantic forest expansions during the Pleistocene (Costa et al. 2018, Werneck 2011, Werneck et al. 2011).

Drivers of subregion distribution

Our Multiple Logistic models pointed towards strong selection effects (sensu Vellend 2010) by the current environment, biome stability since the Pleistocene, and human drivers on the structure of Caatinga woody species. Results thus confirmed the importance of the expansion and contraction of dry and wet climates in the Pleistocene to the species distribution in the Neotropics (Antonelli & Sanmartín 2011, Hughes et al. 2013), and suggest that the stability of dry forest nuclei may represent more favourable locations to drought-prone species, and less-favourable sites for immigrants from wetter regions. The Caatinga suffered extensive vegetation shifts during the last two million years (Auler et al. 2004, Costa et al. 2018, Oliveira et al. 1999) that produced genetic signtures of range expansion for dry forest and woodland species from xeric refugia (Melo et al. 2016, Vieira et al. 2015), and of recent divergence in xeric plant lineages (Menezes et al. 2016). The retention of biome stability in our results indicate that dry forest refugia probably offered extended time for selection to occur on drought-prone traits than in more unstable localities (Carnaval et al. 2014).

The confirmation of the importance of environmental climate, soil, and elevation effects indicate that the subregions we detected may represent a proxy to the delimitation of metacommunities in the Caatinga, since they are influenced by niche and dispersal mechanisms as suggested by metacommunity theory (Leibold et al. 2004). The indicator species found for our subregions are likely to reflect the hyperdominance of an oligarchy of species in different habitat types (ter Steege et al. 2013), with the dominant species exhibiting substantial habitat specialization, and therefore, strong association with environmental factors. This is an interesting untested phenomenon in the Caatinga and desserves future investigation. The balance of energy appeared as a selective forece through aridity and elevation/elevation roughness structuring abundances in the Caatinga. These variables are proxies to local productivity because they reflect the balance between heat load and hydrological dynamics (González-Orozco et al. 2014), and their selection confirms the findings of Silva & Souza (2018). Productivity is strongly restricted in north-eastern South America due to a combination of aridity, seasonal drought, high temperatures, and interannual rainfall impredictability, that frequently causes several dry years in a row (Garreaud et al. 2009). Semiarid conditions select for stress-tolerant plant ecological strategies that favour resource conservation over resource acquisition (Frenette-Dussault et al. 2012, Silva et al. 2018). Several functional traits have been found in drylands that integrate strategies and adaptations to drought including slow growth, deep roots, cavitation-resistant xylem, deciduousness, regulation of gas exchange, metabolic and structural leaf traits related to water-use efficiency, and low cuticular conductance during extended drought (Santiago et al. 2016). Many of these traits have been found in Caatinga species (Fernandes 2007, Moro et al. 2016, Rizzini 1997).

The spatial distribution of subregions was affected by variation in edaphic factors, including soil nutrients, texture, and depth. This confirms recent claims that assign broad physiognomic (Oliveira et al. 2019) and floristic subdivisions of the Caatinga to soil texture (i.e. cyrstalline versus sedimentary Caatinga) (Moro et al. 2016, Oliveira et al. 2019, Queiroz et al. 2017). However, our results suggest that the relationship between soil variables and species distribution is rather complex and cautions against simplifications of the floristic complexity of the domain to binary subdivision schemes like crystalline × sedimentary Caatinga (Oliveira et al. 2019). This is because most of the deviation in the distribution pattern of abundance subregions was explainded by the joint fraction of all variables. Different suits of species have been found in subregions whose probability of occurrence likely peaked at distinct portions of different gradients in soil nutrition and texture, in agreement with earlier findings (Arruda et al. 2015a, b, Neves et al. 2015, Santos et al. 2012). Adaptations to variation in soil properties in the Caatinga and deserve further investigation.

The inclusion of elevation among the selected drivers indicates that the main mountain ranges of the Caatinga (Chapada Diamantina, Araripe, Borborema highlands, and Ibiapaba) act on floristic and abundance patterns mediating productivity. This is because they attenuate aridity through orographic rainfall and reduced temperatures (Kreft & Jetz 2010), and promote niche diversity through heterogenous mountain environment (Dick & Wright 2005). Despite their relatively low altitude, Caatinga mountain ranges have functioned as refugia for plant lineages adapted to wetter conditions (Pessenda et al. 2010, Santos et al. 2007, Silveira et al. 2019). Additionally, orographic rain and rain shadows associated with mountains and inselbergs create spatial structure in aridity and physiognomy in the Caatinga, largely due to strong humid easterly trade winds. The result of these humidity and rainshadows is the formation of the so-called *brejos de altitude*, which are forest refugia on mountain tops and eastern refugia that harbour plant lineages of distinct origins (Rodal et al. 2008, Santos et al. 2007, Silveira et al. 2019).

The Caatinga has been occupied by humans in the last 9000 years (Medeiros et al. 2018), and the footprint of this pre-historic occupation has become visible in the inclusion of the distance to the nearest archaeological site among our explanatory variables as drivers of the distributions of abundance subregions. Indigenous populations are known to impact the distribution of consumed native species across entire landscapes and geographical regions through passive or active seeding of semidomesticated plants (Levis et al. 2017). Studied localities in subregions 1, 6, and 10 were closer to archaeological sites and could have suffered more intense influence of pre-Columbian human activity on the abundance and distribution of some of its species. This inference, however, is highly uncertain, since there is much archaeological field effort yet to be performed in the Caatinga, and also because the density of archaeological sites is not necessarily a good indicator of the cumulative interaction between indigenous peoples and the flora over long time periods. Detailed accounts of pre-Columbian and even contemporary indigenous interactions with the Caatinga flora are still almost absent, and more detailed research on pre-Columbian human interactions with the geographic distribution of Caatinga species is certainly a line of research worth exploring in the future.

Conservation implications

Because the Caatinga flora is subdivided in different subregions with distinct composition and abundance structure, including hundreds of rare and restricted species, it needs multiple and large protected areas, a principle implemented in the Amazon (Silva-Souza & Souza 2020). The Caatinga presents a reduced and poorly-connected protected area network: 1.79% of the domain original area is protected by Conservation Units of Integral Protection, and 7.13% by Conservation Units of sustainable or multiple use (Teixeira et al. 2021), and 91% of its fragments are smaller than 500 ha (Antongiovanni et al. 2018). A wider protected area cover is clearly necessary given the existence of several subregions in the Caatinga. The biogeographical regionalization we proposed here can be employed for conservation planning and management, help measuring and modelling environmental change, and test hypotheses about the ecology and evolution of dry forest plant and animal biotas (Fine 2015, Heikinheimo et al. 2012, Metzger et al. 2005). The subregions we found may represent also functional units as suggested by previous work (Higgins et al. 2011, Manhães et al. 2016). This is because abundance patterns corresponds to changes in the distribution of community-weighted traits like wood density and leaf size as well as to changes in ecosystem-level properties like forest dynamics and biomass (Malhi et al. 2008, ter Steege et al. 2006). As suggested by Silva-Souza

& Souza (2020) for the Amazon, conservation and carbon sequestration strategies for the Caatinga should be implemented in a subregion-by-subregion manner. Lastly, the predicted increase in the frequency and intensity of droughts produced by global warming will reduce the habitat of endemic plants (Silva et al. 2019b), and induce changes in the distribution of flora and subregions (Esquivel-muelbert et al. 2017), reflecting the effects of aridity on the distribution of plant diversity in the biome.

Climate change has been also predicted to increase the functional homogenization of floristic subregions, leading to the loss of biodiversity (Inague et al. 2021), a process that should be monitored carefully and, hopefully, mitigated, and which may be supported by the subregions we identified here.

Acknowledgments

The assistance of Luiza S. Cantídio, Lara M. Alves, Thalita W. P. Nunes, Greyce Kelly F. Silva, Maíra G. Pivato, Vinicius C. Silva in data entry is gratefully acknowledged. Stimulating suggestions by Marcelo F. Moro, Gislene M. S. Ganade, Alice C. Versieux, and an anonymous reviewer greatly helped to improve the manuscript. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil - (CAPES) – Finance Code 001. This research is registered in the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGEN) under code AA38B00.

REFERENCES

ALVARES CA, STAPE JL, SENTELHAS PC, GONÇALVES JLM & SPAROVEK G. 2013. Köppen's climate classification map for Brazil. Meteorol Zeit 22: 711-728. https://doi.org/10.1127/0941-2948/2013/0507.

ANTONELLI A & SANMARTÍN I. 2011. Why are there so many plant species in the Neotropics? Taxon 60: 403-414. https://doi.org/10.2307/41317138.

ANTONGIOVANNI M, VENTICINQUE EM & FONSECA CR. 2018. Fragmentation patterns of the Caatinga drylands. Landsc Ecol 33: 1353-1367. https://doi.org/10.1007/s10980-018-0672-6.

ARELLANO G, UMAÑA MN, MACÍA MJ, LOZA MI, FUENTES A, CALA V & JØRGENSEN PM. 2017. The role of niche overlap, environmental heterogeneity, landscape roughness and productivity in shaping species abundance distributions along the Amazon-Andes gradient. Glob Ecol Biogeogr 26: 191-202. https://doi.org/10.1111/geb.12531.

ARRUDA DM, SCHAEFER CEGR, CORRÊA GR, RODRIGUES PMS, DUQUE-BRASIL R, FERREIRA-JR WG & OLIVEIRA-FILHO AT. 2015a. Landforms and soil attributes determine the vegetation structure in the Brazilian semiarid. Folia Geobot 50: 175-184. https://doi.org/10.1007/s12224-015-9221-0.

ARRUDA DM, SCHAEFER CEGR & MORAES MLB. 2015b. Relações entre Atributos do Solo e Vegetações da Região Ecotonal do Médio Rio São Francisco, Brasil. Rev Bras Cienc do Solo 39: 1524-1532. https://doi.org/10.1590/01000683rbcs20150241.

AULER AS, WANG X, EDWARDS RL, CHENG H, CRISTALLI PS, SMART PL & RICHARDS DA. 2004. Quarternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. J Quat Sci 19: 693-701. https://doi.org/10.1002/jqs.876.

BANDA KR ET AL. 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. Science 353: 1383-1387. https://doi.org/10.1126/science.aaf5080.

BASELGA A. 2013. Separating the two components of abundance-based dissimilarity: Balanced changes in abundance vs. abundance gradients. Methods Ecol Evol 4: 552-557. https://doi.org/10.1111/2041-210X.12029

BASTIN JF ET AL. 2017. The extent of forest in dryland biomes. Science 356: 635-638. https://doi.org/10.1126/science.aam6527.

BFG. 2015. Growing knowledge: an overview of seed plant diversity in Brazil. Rodriguesia 66: 1085-1113. https://doi.org/10.1590/2175-7860201566411.

BORCARD D, GILLET F & LEGENDRE P. 2011. Numerical ecology with R. Springer, New York.

BURNHAM KP & ANDERSON DR. 2002. Model selection and multi model inference: a practial information-theoretic approach, 2nd ed. Springer, New York.

CÁCERES M & LEGENDRE P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90: 3566-3574. https://doi.org/10.1890/08-1823.1.

CANTIDIO LS & SOUZA AF. 2019. Aridity, soil and biome stability influence plant ecoregions in the Atlantic Forest, a biodiversity hotspot in South America. Ecography 42: 1887-1898. https://doi.org/10.1111/ecog.04564.

CARDOSO D, MOONLIGHT PW, RAMOS G, OATLEY G, DUDLEY C, GAGNON E, QUEIROZ LP, PENNINGTON RT & SÄRKINEN TE. 2021. Defining Biologically Meaningful Biomes Through Floristic, Functional, and Phylogenetic Data. Front Ecol Evol 9: 723558. doi: 10.3389/fevo.2021.723558.

CARNAVAL AC ET AL. 2014. Prediction of phylogeographic endemism in an environmentally complex biome. Proc R Soc B Biol Sci 281: 20141461. https://doi.org/10.1098/rspb.2014.1461.

COLLEVATTI RG, TERRIBILE LC, LIMA-RIBEIRO MS, NABOUT JC, OLIVEIRA G, RANGEL TF, RABELO SG & DINIZ-FILHO JAF. 2012. A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. Mol Ecol 21: 5845-5863.

COSTA GC, HAMPE A, LEDRU MP, MARTINEZ PA, MAZZOCHINI GG, SHEPARD DB, WERNECK FP, MORITZ C & CARNAVAL AC. 2018. Biome stability in South America over the last 30 kyr: Inferences from long-term vegetation dynamics and habitat modelling. Glob Ecol Biogeogr 27: 285-297. https://doi.org/10.1111/geb.12694.

DICK CW & WRIGHT SJ. 2005. Tropical mountain cradles of dry forest diversity. Proc Natl Acad Sci 102: 10757-10758. https://doi.org/10.1073/pnas.0505013102.

DINERSTEIN E, OLSON D, JOSHI A, VYNNE C, BURGESS ND, WIKRAMANAYAKE E, HAHN N, PALMINTERI S, HEDAO P & NOSS R. 2017. An ecoregion-based approach to protecting half the terrestrial realm. Biosci 67: 534-545. https://doi.org/10.1093/biosci/bix014.

ESQUIVEL-MUELBERT A, GALBRAITH D, DEXTER KG, BAKER TR, LEWIS L, MEIR P, ROWLAND L, CARLOS A, NEPSTAD D & PHILLIPS OL. 2017. Biogeographic distributions of neotropical trees reflect their directly measured drought tolerances. Sci Rep 7: 8334. https://doi.org/10.1038/s41598-017-08105-8.

FERNANDES A. 2007. Fitogeografia brasileira: fundamentos biogeográficos. Edições UFC, Fortaleza. p. 183.

FERNANDES AG & BEZERRA P. 1990. Estudo fitogeográfico do Brasil. Stylus Comunicações, Fortaleza. Stylus Comunic, p. 205.

FERNANDES MF, CARDOSO D, QUEIROZ LP. 2020. An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. J Arid Environ 104079. https://doi.org/10.1016/j.jaridenv.2019.104079.

FINE PVA. 2015. Ecological and evolutionary drivers of geographic variation in species diversity. Annu Rev Ecol Evol Syst 46: 369-392. https://doi.org/10.1146/annurev-ecolsys-112414-054102.

FRENETTE-DUSSAULT C, SHIPLEY B, LÉGER J-F, MEZIANE D & HINGRAT Y. 2012. Functional structure of an arid steppe plant community reveals similarities with Grime's C-S-R theory. J Veg Sci 23: 208-222. https://doi.org/10.1111/j.1654-1103.2011.01350.x.

GARREAUD RD, VUILLE M, COMPAGNUCCI R & MARENGO J. 2009. Present-day South American climate. Palaeogeogr Palaeoclimatol Palaeoecol 281: 180-195. https://doi.org/10.1016/j.palaeo.2007.10.032.

GOMES VHF ET AL. 2018. Species distribution modelling : contrasting presence-only models with plot abundance data. Sci Rep 8: 1-12. https://doi.org/10.1038/s41598-017-18927-1.

GONZÁLEZ-OROZCO CE ET AL. 2014. Quantifying phytogeographical regions of Australia using geospatial turnover in species composition. PLoS One 9. https://doi.org/10.1371/journal.pone.0092558.

GRIME JP & PIERCE S. 2012. The evolutionary strategies that shape ecosystems. Wiley-Blackwell, Chichester, West Sussex, UK; Hoboken, NJ. DOI:10.1002/9781118223246.

HEIKINHEIMO H, ERONEN JT, SENNIKOV A, PRESTON CD, OIKARINEN E, UOTILA P, MANNILA H & FORTELIUS M. 2012. Convergence in the distribution patterns of Europe's plants and mammals is due to environmental forcing. J Biogeogr 39: 1633-1644. https://doi.org/10.1111/j.1365-2699.2012.02723.x.

HIGGINS MA, RUOKOLAINEN K, TUOMISTO H, LLERENA N, CARDENAS G, PHILLIPS OL, VÁSQUEZ R & RÄSÄNEN M. 2011. Geological control of floristic composition in Amazonian forests. J Biogeogr 38: 2136-2149. https://doi.org/10.1111/j.1365-2699.2011.02585.x.

HOLT BG ET AL. 2013. An Update of Wallace's Zoogeographic Regions of the World. Science 339: 74-78. https://doi.org/10.1126/science.1228282.

HUANG J, YU H, GUAN X, WANG G & GUO R. 2015. Accelerated dryland expansion under climate change. Nat Clim Chang 6. https://doi.org/10.1038/NCLIMATE2837.

HUGHES CE, PENNINGTON RT & ANTONELLI A. 2013. Neotropical plant evolution: assembling the big picture. Bot J Linn Soc 171: 1-18. https://doi.org/10.1111/boj.12006.

IBGE 2004. MAPA de biomas do Brasil: primeira aproximação. Rio de Janeiro. 1 mapa, color. Escala 1:5000 000. Projeção policônica.

INAGUE GM, ZWIENER VP & MARQUES MCM. 2021. Climate change threatens the woody plant taxonomic and functional diversities of the Restinga vegetation in Brazil. Perspect Ecol Conserv 19: 53-60. https://doi.org/10.1016/j.pecon.2020.12.006.

KESSLER M, GRYTNES J-A, HALLOY SR, KLUGE J, KRÖMER T, LEÓN B, MACÍA MJ & YOUNG KR. 2011. Gradients of plant diversity: local patterns and processes, in: Herzog SK, Martínez R, Jørgensen PM & Tiessen H (Eds), Climate Change and Biodiversity in the Tropical Andes. Inter-American Institute for Global Change Research (IAI) and Scientific Committee on Problems of the Environment (SCOPE), p. 204-2019.

KREFT H & JETZ W. 2007. Global patterns and determinants of vascular plant diversity. Proc Natl Acad Sci 104: 5925-5930. https://doi.org/10.1073/pnas.0608361104.

KREFT H & JETZ W. 2010. A framework for delineating biogeographical regions based on species distributions. J Biogeogr 37: 2029-2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x.

KUMAR S & STOHLGREN TJ. 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree Canacomyrica monticola in New Caledonia. J Eco Nat Sci 1(4): 094-098.

LEGENDRE P & LEGENDRE LF. 2012. Numerical ecology, 3rd ed. Elsevier, Amsterdam.

LEIBOLD MA ET AL. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7: 601-613.

LEVIS C, COSTA FRC ET AL. 2017. Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. Science 931: 925-931. https://doi.org/10.1126/science.aal0157.

MALHI Y, ROBERTS JT, BETTS RA, KILLEEN TJ, LI W & NOBRE CA. 2008. Climate Change, Deforestation, and the Fate of the Amazon. Science 319: 169-173.

MANHÃES AP, MAZZOCHINI GG, OLIVEIRA-FILHO AT, GANADE G & CARVALHO AR. 2016. Spatial associations of ecosystem services and biodiversity as a baseline for systematic conservation planning. Divers Distrib 1-12. https://doi.org/10.1111/ddi.12459.

MAZZOCHINI GG, FONSECA CR, COSTA GC, SANTOS RM, OLIVEIRA-FILHO AT & GANADE G. 2019. Plant phylogenetic diversity stabilizes large-scale ecosystem productivity. Glob Ecol Biogeogr 12963-12963. https://doi.org/10.1111/geb.12963.

MCCUNE B & GRACE JB. 2002. Analysis of ecological communities. MJM Software, Gleneden Beach.

MEDEIROS VB, OLIVEIRA PE, SANTOS RA, BARRETO AMF, OLIVEIRA MAT & PINAYA JLD. 2018. New holocene pollen records from the Brazilian Caatinga. An Acad Bras Cienc 90: 2011-2023. https://dx.doi.org/10.1590/0001-3765201820170161.

MELO WA, LIMA-RIBEIRO MS, TERRIBILE LC & COLLEVATTI RG. 2016. Coalescent simulation and paleodistribution modeling for tabebuia rosealba do not support south American dry forest refugia hypothesis. PLoS One 11. https://doi.org/10.1371/journal.pone.0159314.

MENEZES MOT, ZAPPI DC, MORAES EM, FRANCO FF, TAYLOR NP, COSTA IR & LOIOLA MIB. 2016. Pleistocene radiation of coastal species of Pilosocereus (Cactaceae) in eastern Brazil. J Arid Environ 135: 22-32. https://doi.org/10.1016/j.jaridenv.2016.08.006.

METZGER MJ, BUNCE RGH, JONGMAN RHG, MÜCHER CA & WATKINS JW. 2005. A climatic stratification of the environment of Europe. Glob Ecol Biogeogr 14: 549-563. https://doi.org/10.1111/j.1466-822X.2005.00190.x.

MORO MF, MACEDO MB, MOURA-FÈ MM, FARIAS CASTRO AS & COSTA RC. 2015a. Vegetação, unidades fitoecológicas e diversidade paisagística do estado do Ceará. Rodriguesia 66: 717-743. https://doi.org/10.1590/2175-7860201566305.

MORO MF, SILVA IA, ARAÚJO FS, LUGHADHA EN, MEAGHER TR & MARTINS FR. 2015b. The role of edaphic environment and climate in structuring phylogenetic pattern in seasonally dry tropical plant communities. PLoS One 10: 1-18. https://doi.org/10.1371/journal.pone.0119166.

MORO MF, LUGHADHA EN, ARAÚJO FS & MARTINS FR. 2016. A phytogeographical metaanalysis of the semiarid Caatinga domain in Brazil. Bot Rev 82: 91-148. https://doi.org/10.1007/s12229-016-9164-z.

MOURA MR, ARGÔLO AJ & COSTA HC. 2017. Historical and contemporary correlates of snake biogeographical subregions in the Atlantic Forest hotspot. J Biogeogr 44: 640-650. https://doi.org/10.1111/jbi.12900.

MYERS N, MITTERMEIER RA, MITTERMEIER CG, FONSECA GAB & KENT J. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853-858.

NASCIMENTO CES, RODAL MJN & CAVALCANTI AC. 2003. Phytosociology of the remaining xerophytic woodland associated to an environmental gradient at the banks of the São Francisco river - Petrolina, Pernambuco, Brazil. Rev Bras Botânica 26: 271-287. https://doi.org/10.1590/S0100-84042003000300001.

NEVES DM, DEXTER KG, PENNINGTON RT, BUENO ML & OLIVEIRA FILHO AT. 2015. Environmental and historical controls of floristic composition across the South American Dry Diagonal. J Biogeogr 42: 1566-1576. https://doi.org/10.1111/jbi.12529.

NEVES DM ET AL. 2020. Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. Sci Rep 10: 1188. https://doi.org/10.1038/s41598-019-55621-w.

NIMER E. 1972. Climatologia da região nordeste do Brasil - introdução à climatologia dinâmica. Rev Bras Geogr 34: 3-51.

OLIVEIRA-FILHO AT. 2009. Classificação das fitofisionomias da américa do sul cisandina tropical e subtropical: proposta de um novo sistema - prático e flexível - ou uma injeção a mais de caos? Rodriguésia 6: 237-258.

OLIVEIRA GC, FRANCELINO MR, ARRUDA DM, FERNANDES-FILHO EI & SCHAEFER CEGR. 2019. Climate and soils at the Brazilian semiarid and the forest-Caatinga problem: new insights and implications for conservation. Environ Res Lett 14: 104007. https://doi.org/10.1088/1748-9326/ab3d7b.

OLIVEIRA PE, MAGNO A & SUGUIO K. 1999. Late Pleistocene/Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. Palaeogeogr. Palaeoclimatol Palaeoecol 152: 319-337.

OLSON DM ET AL. 2001. Terrestrial ecoregions of the world: a new map of life on earth. Bioscience 51: 933-938.

PAN Y ET AL. 2011. A large and persistent carbon sink in the world’s forests. Science 333: 988-993. https://doi.org/10.1126/science.1201609.

PENNINGTON RT, PRADO DE & PENDRY CA. 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. J Biogeogr 27: 261.

PENNINGTON RT, RICHARDSON JE & LAVIN M. 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. New Phytol 172: 605-616. https://doi.org/10.1111/j.1469-8137.2006.01902.x.

PESSENDA LCR, GOUVEIA SEM, RIBEIRO AS, OLIVEIRA PE & ARAVENA R. 2010. Late Pleistocene and Holocene vegetation changes in northeastern Brazil determined from carbon isotopes and charcoal records in soils. Palaeogeogr Palaeoclimatol Palaeoecol 297: 597-608. https://doi.org/10.1016/j.palaeo.2010.09.008.

QUEIROZ PG, CARDOSO D, FERNANDES MF & MORO MF. 2017. Diversity and evolution of florwering plants of the Caatinga domain, In: Silva JMC, Leal I & Tabarelli M (Eds), Caatinga: The Largest Tropical Dry Forest Region in South America. Springer, p. 23-63.

R-CORE-TEAM. 2017. R: A language and environment for statistical computing. Version 3.4.3.

RIBEIRO ES, SANTOS B, ARROYO-RODRÍGUEZ V, TABARELLI M, SOUZA G & LEAL IR. 2016. Phylogenetic impoverishment of plant communities following chronic human disturbances in the Brazilian Caatinga. Ecology 97: 1583-1592. https://doi.org/10.1890/15-1122.1.

RITO KF, ARROYO-RODRÍGUEZ V, QUEIROZ RT, LEAL IR & TABARELLI M. 2017a. Precipitation mediates the effect of human disturbance on the Brazilian Caatinga vegetation. J Ecol 105: 828-838. https://doi.org/10.1111/1365-2745.12712.

RITO KF, TABARELLI M & LEAL IR. 2017b. Euphorbiaceae responses to chronic anthropogenic disturbances in Caatinga vegetation: from species proliferation to biotic homogenization. Plant Ecol 218: 749-759. https://doi.org/10.1007/s11258-017-0726-x.

RIZZINI CT. 1997. Tratado de fitogeografia do Brasil: aspectos ecológicos, sociológicos e florísticos. Âmbito Cultural Edições, Rio de Janeiro.

RODAL MJN, BARBOSA MRV & THOMAS WW. 2008. Do the seasonal forests in northeastern Brazil represent a single floristic unit? Brazilian J Biol 68: 467-475. https://doi.org/10.1590/S1519-69842008000300003.

ROSS JLS. 2011. O relevo brasileiro, as superfícies de aplanamento e os níveis morfológicos. RDG Rev. do Dep. Geogr. 5: 7-24.

SAITER FZ, BROWN JL, THOMAS WW, OLIVEIRA-FILHO AT & CARNAVAL AC. 2016. Environmental correlates of floristic regions and plant turnover in the Atlantic Forest hotspot. J Biogeogr 43: 2322-2331. https://doi.org/10.1111/jbi.12774.

SALVADOR S & CHAN P. 2004. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms, in: Proceedings of the International Conference on Tools with Artificial Intelligence. p. 576-584. https://doi.org/10.1109/ICTAI.2004.50.

SANTIAGO LS, BONAL D, GUZMAN ME & ÁVILA-LOVERA E. 2016. Drought survival strategies of tropical trees, Tropical tree physiology, Tree Physiology. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-27422-5.

SANTOS AMM, CAVALCANTI DR, SILVA JMC & TABARELLI M. 2007. Biogeographical relationships among tropical forests in north-eastern Brazil. J Biogeogr 34: 437-446. https://doi.org/10.1111/j.1365-2699.2006.01604.x.

SANTOS JC, LEAL IR, ALMEIDA-CORTEZ JS, FERNANDES GW & TABARELLI M. 2011. Caatinga: the scientific negligence experienced by a dry tropical forest. Trop Conserv Sci 4: 276-286. https://doi.org/10.1177/194008291100400306.

SANTOS RM, OLIVEIRA-FILHO AT, EISENLOHR PV, QUEIROZ LP, CARDOSO DBOS & RODAL MJN. 2012. Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. Ecol Evol 2: 409-428. https://doi.org/10.1002/ece3.91.

SILVA JMC, BARBOSA LCF, LEAL IR & TABARELLI M. 2017. The Caatinga: understanding the challenges, In: Silva JMC, Leal IR & Tabarelli M (Eds), Caatinga: The Largest Tropical Dry Forest Region in South America. Springer, p. 3-19.

SILVA AC & SOUZA AF. 2018. Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America. PLoS One 13: e0196130-e0196130. https://doi.org/10.1371/journal.pone.0196130.

SILVA JLA, SOUZA AF, CALIMAN A, VOIGT EL & LICHSTON JE. 2018. Weak whole-plant trait coordination in a seasonally dry South American stressful environment. Ecol Evol 8: 4-12. https://doi.org/10.1002/ece3.3547.

SILVA JLA, SOUZA AF & SANTIAGO LS. 2019a. Traits uncover quasi-neutral community assembly in a coastal heath vegetation. J Plant Ecol 12: 703-712. https://doi.org/10.1093/jpe/rtz007.

SILVA-SOUZA KJP & SOUZA AF. 2020. Woody plant subregions of the Amazon forest. J Ecol 108: 2321-2335. https://doi.org/10.1111/1365-2745.13406.

SILVA JLSE, CRUZ-NETO O, PERES CA, TABARELLI M & LOPES AV. 2019b. Climate change will reduce suitable Caatinga dry forest habitat for endemic plants with disproportionate impacts on specialized reproductive strategies. PLoS ONE 14(5): e0217028. https://doi.org/10.1371/journal.pone.0217028.

SILVEIRA MHB, MASCARENHAS R, CARDOSO D & BATALHA-FILHO H. 2019. Pleistocene climatic instability drove the historical distribution of forest islands in the northeastern Brazilian Atlantic Forest. Palaeogeogr Palaeoclimatol Palaeoecol 527: 67-76. https://doi.org/10.1016/j.palaeo.2019.04.028.

SLIK JWF ET AL. 2018. Phylogenetic classification of the world's tropical forests. Proc Natl Acad Sci 115: 1837-1842. https://doi.org/10.1073/pnas.1714977115.

SMITH JR ET AL. 2018. A global test of ecoregions. Nat Ecol Evol https://doi.org/10.1038/s41559-018-0709-x.

SMITH RW ET AL. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. Ecol Appl 11: 1073-1087.

SOUTHWOOD TRE. 1977. Habitat, the templet for ecological Strategies? J Anim Ecol 46: 336. https://doi.org/10.2307/3817.

TEIXEIRA MG, VENTICINQUE EM, LION MB & PINTO MP. 2021. The Brazilian Caatinga protected areas: an extremely unbalanced conservation system. Environ Conserv 48: 287-294.

TER STEEGE H ET AL. 2013. Hyperdominance in the Amazonian tree flora. Science 342: 1243092. https://doi.org/10.1126/science.1243092

TER STEEGE H ET AL. 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. Nature 443: 444-447. https://doi.org/10.1038/nature05134.

TRABUCCO A & ZOMER RJ. 2019. Global Aridity Index and Potential Evapo-Transpiration (ET0) Climate Database v2. CGIAR Consortium for Spatial Information (CGIAR-CSI). Published online, available from the CGIAR-CSI GeoPortal at https://cgiarcsi.community.

TUOMISTO H, RUOKOLAINEN L & RUOKOLAINEN K. 2012. Modelling niche and neutral dynamics: on the ecological interpretation of variation partitioning results. Ecography (Cop) 35: 961-971. https://doi.org/10.1111/j.1600-0587.2012.07339.x.

TURCHETTO-ZOLET AC, PINHEIRO F, SALGUEIRO F & PALMA-SILVA C. 2013. Phylogeographical patterns shed light on evolutionary process in South America. Mol Ecol 22: 1193-1213. https://doi.org/10.1111/mec.12164.

VELLEND M. 2010. Conceptual synthesis in community ecology. Q Rev Biol 85: 183-206. https://doi.org/10.1086/652373.

VELLOSO AL ET AL. 2002. Ecorregiões propostas para o bioma Caatinga. Associação de Plantas do Nordeste/Instituto para Conservação Ambiental/The Nature Conservancy do Brasil, Recife.

VELOSO HP, RANGEL-FILHO ALR & LIMA JCA. 1991. Classificação da vegetação brasileira, adaptada a um sistema universal. https://doi.org/ISBN 85-240-0384-7.

VENTER O ET AL. 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nat Commun 7: 1-11. doi:10.1038/ncomms12558.

VIEIRA FA, NOVAES RML, FAJARDO CG, SANTOS RM, ALMEIDA HS, CARVALHO D & LOVATO MB. 2015. Holocene southward expansion in seasonally dry tropical forests in South America: phylogeography of Ficus bonijesulapensis (Moraceae). Bot J Linn Soc 177: 189-201. https://doi.org/10.1111/boj.12241.

WERNECK FP. 2011. The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. Quat Sci Rev 30: 1630e1648-1630e1648.

WERNECK FP, COSTA GC, COLLI GR, PRADO DE & SITES JR JW. 2011. Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. Glob Ecol Biogeogr 20: 272-288. https://doi.org/10.1111/j.1466-8238.2010.00596.x.

WHITTAKER RJ, ARAUJO MB, LADLE RJ, WATSON JE & WILLIS KJ. 2005. Conservation biogeography: asessment and prospect. Divers Distrib 11: 3-23. https://doi.org/10.1111/j.1366-9516.2005.00143.x.

ZAPPI DC ET AL. 2003. Lista das plantas vasculares de Catolés, Chapada Diamantina, Bahia, Brasil. Bol Botânica da Univ São Paulo 21: 345-398. https://doi.org/10.11606/issn.2316-9052.v21i2p345-398.

ZUUR AF, IENO EN, WALKER N, SAVELIEV AA & SMITH GM. 2009. Mixed effects models and extensions in ecology with R. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-87458-6.

How to cite

SILVA AC & SOUZA AF. 2022. Spatial structure of the Caatinga woody flora: abundance patterns have environmental, Pleistocene, and indigenous drivers. An Acad Bras Cienc 94: e20211019. DOI 10.1590/0001-3765202220211019.

Manuscript received on July 17, 2021,
accepted for publication on February 25, 2022

AUGUSTO C. SILVA[1]
http://orcid.org/0000-0003-1486-3897

ALEXANDRE F. SOUZA[2]
http://orcid.org/0000-0001-7468-3631

[1]Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Ecologia, Departamento de Ecologia, CB, Campus Universitário, Lagoa Nova, Caixa Postal 1524, 59078-900 Natal, RN, Brazil

[2]Universidade Federal do Rio Grande do Norte, Departamento de Ecologia, CB, Campus Universitário, Lagoa Nova, Caixa Postal 1524, 59078-900 Natal, RN, Brazil

Correspondence to: **Augusto C. Silva**
E-mail: agst.25@gmail.com

Authors Contributions

A.F.S. conceived the database and coordinated the study; A.C.S. and A.F.S. designed the paper and contributed equally, implemented the database, analysed the data and wrote the manuscript; A.C.S. prepared the figures.