

# Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America

Pedro Luiz Silva de Miranda<sup>1</sup>  | Ary T. Oliveira-Filho<sup>2</sup> | R. Toby Pennington<sup>3,4</sup> | Danilo M. Neves<sup>5</sup> | Timothy R. Baker<sup>6</sup> | Kyle G. Dexter<sup>1,3</sup>

<sup>1</sup>School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom

<sup>2</sup>Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

<sup>3</sup>Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

<sup>4</sup>Department of Geography, University of Exeter, Exeter, United Kingdom

<sup>5</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona

<sup>6</sup>School of Geography, University of Leeds, Leeds, United Kingdom

## Correspondence

Pedro Luiz Silva de Miranda, Room 401, Crew Building, King's Buildings, Edinburgh EH9 3FF, United Kingdom.

Email: p.l.silva-de-miranda@sms.ed.ac.uk

## Funding information

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award number: 99999.013197/2013-04; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award number: 301644/88-8; National Environment Research Council, Grant/Award number: NE/1028122/1; National Science Foundation - USA (grant NSF/DEB-1556651); Leverhulme Trust International Academic Fellowship (obs: there is no grant number for this fellowship)

Editor: Thomas Gillespie

## Abstract

**Aim:** To define and map the main biomes of lowland tropical South America (LTSA) using data from tree species inventories and to test the ability of climatic and edaphic variables to distinguish amongst them.

**Location:** Lowland Tropical South America (LTSA), including Argentina, Bolivia, Brazil, Ecuador, Paraguay, Peru and Uruguay.

**Time period:** Present.

**Major taxa studied:** Trees.

**Methods:** We compiled a database of 4,103 geo-referenced tree species inventories distributed across LTSA. We used a priori vegetation classifications and cluster analyses of floristic composition to assign sites to biomes. We mapped these biomes geographically and assessed climatic overlaps amongst them. We implemented classification tree approaches to quantify how well climatic and edaphic data can assign inventories to biomes.

**Results:** Our analyses distinguish savanna and seasonally dry tropical forest (SDTF) as distinct biomes, with the Chaco woodlands potentially representing a third dry biome in LTSA. Amongst the wet forests, we find that the Amazon and Atlantic Forests might represent different biomes, because they are distinct in both climate and species composition. Our results show substantial environmental overlap amongst biomes, with error rates for classifying sites into biomes of 19–21 and 16–18% using only climatic data and with the inclusion of edaphic data, respectively.

**Main conclusions:** Tree species composition can be used to determine biome identity at continental scales. We find high biome heterogeneity at small spatial scales, probably attributable to variation in edaphic conditions and disturbance history. This points to the challenges of using climatic and/or interpolation-based edaphic data or coarse-resolution, remotely sensed imagery to map tropical biomes. From this perspective, we suggest that using floristic information in biome delimitation will allow for greater synergy between conservation efforts centred on species diversity and management efforts centred on ecosystem function.

## KEYWORDS

Amazon Forest, Atlantic Forest, Cerrado, Chaco, cluster analysis, NeoTropTree, Savanna, Dry Tropical Forest

## 1 | INTRODUCTION

The biome concept has existed for more than a century, with the overarching purpose of delimiting recognizable, ecologically meaningful vegetation units. Humboldt (1816) used the term phytophysionomy when referring to areas that may be geographically disjunct but share similar vegetation physiognomy or structure. The link between vegetation structure and climatic conditions was detailed by Schimper (1903), who attributed these similarities to physiological and anatomical adaptations to precipitation and temperature. The relationship between vegetation form and climate permeates the majority of vegetation classification schemes proposed during the 20th century (Clements, 1916; Holdridge, 1947; Walter, 1973; Whittaker, 1975), and climate is still regarded as the main driver of plant and biome distributions (Box, 1995; Prentice, 1990; Prentice et al., 1992). More recently, biomes have been used to categorize the function of ecosystems at large spatial scales, including across continents (Higgins, Buitenwerf, & Moncrieff, 2016; Woodward, Lomas, & Kelly, 2004), and the most prevalent biome concept at present, which we use here, is that of a widespread vegetation formation with distinct ecosystem function.

The term 'biome' itself was first used by Clements (1916) when referring to the biotic community, or set of species, occupying a certain habitat. However, subsequently, Holdridge (1947), Walter (1973), Whittaker (1975) and Odum (1975) gave more emphasis to the relationship between climate and vegetation structure when proposing classification systems for vegetation formations or biomes, and distanced themselves from the community composition perspective suggested by Clements (1916). These later authors delimited biomes using standard climatic variables, such as mean annual temperature (MAT) and mean annual precipitation (MAP; e.g., Whittaker, 1975). A motivating factor for these studies was to create practical classification systems that allow researchers to assign sites to biomes by simply knowing the MAT and MAP (e.g., as by Qian, Jin, & Ricklefs, 2017; Siepielski et al., 2017). More recently, large-scale remotely sensed data have become available, which have led researchers to map biomes using simple characterizations of vegetation physiognomy or ecosystem function, including average vegetation height, percentage tree cover, primary productivity and phenology (Higgins et al., 2016; Hirota, Holmgren, Van Nes, & Scheffer, 2011; Staver, Archibald, & Levin, 2011; Woodward et al., 2004). However, remote-sensing approaches can fail when biomes are indistinguishable from satellite images (Beuchle et al., 2015) or when there is high structural heterogeneity within biomes (Särkinen, Iganci, Linares-Palomino, Simon, & Prado, 2011).

Meanwhile, the different global biome schemes, be they derived from climate or remote sensing, often fail to agree on which are the main biomes (e.g., Whittaker, 1975 versus Friedl et al., 2002 versus Woodward et al., 2004 versus Higgins et al., 2016) and can differ dramatically on the mapping of any given biome (Särkinen et al., 2011). Furthermore, the degree to which biome maps genuinely delimit the spatial distribution of ecosystem function is debated (Moncrieff, Hickler, & Higgins, 2015). The need for more ecologically meaningful definitions of biomes has led some to suggest that functional traits, such as wood density or leaf mass per area of the

dominant plant species, should be used to define and delimit biomes (Van Bodegom, Douma, & Verheijen, 2014; Violle, Reich, Pacala, Enquist, & Kattge, 2014). In order to map functional trait distributions at large spatial scales, researchers have used geo-referenced collection localities for species with available trait data (e.g., Engemann et al., 2016; Lamanna et al., 2014). There are challenges with this approach, most importantly, the absence of trait data for many species, especially in tropical vegetation (Sandel et al., 2015; Violle, Borgy, & Choler, 2015). The premise of the present study is that species occupying distinct biomes have different functional traits and therefore that floristic information can be used to map biomes, avoiding the uncertainties associated with linking species composition to trait databases. Species distribution modelling (also known as ecological niche modelling) of indicator species can be used to map biomes (as by Prieto-Torres & Rojas-Soto, 2016; Särkinen et al., 2011), but such distribution modelling usually uses only climatic variables as predictors and therefore is subject to similar concerns as mapping biomes directly based on climatic data. We argue that, at least for some regions, there are now sufficient species distribution data to map biomes directly using the distribution data themselves.

The mapping of biomes based on floristic information also offers the possibility of synergies with conservation (Whittaker et al., 2005). Bioregionalization schemes that partition space into geographical units based on species composition and environmental data, such as the global ecoregions proposed by Olson and Dinerstein (1998) and Olson et al. (2001) (recently reviewed and updated by Dinerstein et al., 2017) have been used by researchers and decision-makers in conservation at local and global scales. For example, it was by relying on the scheme of Olson and Dinerstein (1998) that Myers, Mittermeier, Mittermeier, Fonseca, & Kent (2000), Mittermeier, Myers, Thomsen, da Fonseca, & Olivieri (1998) and Mittermeier et al. (2004) proposed the global biodiversity hotspots, which are biomes or geographical subsets of biomes (i.e., ecoregions) that have high numbers of endemic species and are particularly threatened.

Brazil, which comprises the majority of the land surface of lowland tropical South America (LTSA), has proposed its own bioregionalization scheme, the Domain system, established by Veloso, Rangel Filho, & Lima (1991) and IBGE (2012). The six Domains, which are used to guide conservation and management policy, are the Amazon Forest, Atlantic Forest, Cerrado, Caatinga, Pantanal and Pampa. The first two are wet forests, with the Amazon Forest occupying much of northern LTSA and the Atlantic Forest occurring along the Atlantic coast of South America, principally in Brazil. They are separated by a 'dry diagonal' of seasonally dry forest, woodland and savanna vegetation formations (Neves, Dexter, Pennington, Bueno, & Oliveira Filho, 2015; Vanzolini, 1963). The Cerrado Domain is composed primarily of savanna and sits in the centre of the dry diagonal, occupying much of central Brazil, but there are disjunct patches of savanna found elsewhere in LTSA, particularly within the Atlantic and Amazon Forests (Ratter, Ribeiro, & Bridgewater, 1997). Wet forests intrude into the Cerrado as gallery forests along river courses (Oliveira-Filho & Ratter, 1995). The Caatinga Domain at the northeast corner of the dry diagonal represents the largest extent of seasonally dry tropical forest

(SDTF) in LTSA (Prado & Gibbs, 1993). However, SDTF also occurs in disjunct patches throughout the Cerrado on more fertile soils (DRYFLOR, 2016; Pennington, Prado, & Pendry, 2000; Prado & Gibbs, 1993). The SDTFs and the Cerrado can be distinguished by physiognomy, function and dissimilarities in phylogenetic composition (Oliveira-Filho, Pennington, Rotella, & Lavin, 2014; Oliveira-Filho et al., 2013). The Chaco woodlands at the southwest of the dry diagonal are climatically seasonal, and these woodlands do not experience fire. The Chaco woodlands have been considered distinct from SDTF on the basis that they experience regular frost, greater temperature seasonality and often distinct edaphic conditions, such as hypersaline soils (DRYFLOR 2016; Prado & Gibbs, 1993). The Pantanal Domain has heterogeneous vegetation, including SDTFs, savanna and swamps, whereas the Pampa Domain is a largely subtropical grassland that has forest patches along river courses and on certain edaphic conditions.

Lowland tropical South America, owing to its size, diversity and non-continuous geographical distribution of biomes and vegetation types, is an ideal system in which to study how biomes can be delimited, at a continental scale, through means other than climate and remote sensing. Its complex environmental controls of both climate and soil indicate the necessity of developing a new approach for biome delimitation that is better linked to biodiversity. Biome schemes centred on species composition might be more useful for comparative biology and conservation and might enable a better understanding of the possible mechanistic relationships between vegetation and environment.

Here we test the utility and performance of a floristic approach for mapping biomes at a continental scale, with a particular focus on Brazil and neighbouring countries. We use a dataset of 4,103 geo-referenced floristic inventories of tree species that span the major climatic and edaphic gradients of the region. We first test how well climatic data perform in distinguishing among biomes. We hypothesize that climatic data will enable us to distinguish wet forests from the dry biomes, but that they will fail in allowing us to distinguish SDTF from savanna, because these are often edaphically differentiated (Ratter et al., 1997). We also test the ability of edaphic data, when considered in conjunction with climate, to increase the accuracy of biome delimitation. Lastly, we assess how our floristic approach to mapping biomes compares with the ecoregion-based classification system of Dinerstein et al. (2017) (a revised version of the system of Olson et al., 2001), and then for Brazil only, against the Domain classification of IBGE (2012). Our use of floristics data might allow for the delimitation and mapping of biomes in a manner directly relevant to the management of ecosystems and development of conservation strategies, for example by enabling the modelling of future effects of climate change on tropical vegetation (Prieto-Torres & Rojas-Soto, 2016; Prieto-Torres, Navarro-Siguenza, Santiago-Alarcon, & Rojas-Soto, 2016).

## 2 | METHODS

### 2.1 | The NeoTropTree dataset

Floristic inventories of tree communities were obtained from the NeoTropTree (NTT) dataset (Oliveira-Filho, 2017), which contains tree

species inventories for > 6,000 geo-referenced sites across South America. Trees are defined here as free-standing woody plants > 3 m in height. Every site in the NTT database is based on a tree species list generated via an inventory, phytosociological survey or floristic survey. These data sources are derived from published and unpublished literature (e.g., PhD theses, environmental consultancy reports). Other species are added to the site species list based on surveys of specimens in herbaria in South America, the USA and Europe or online (e.g., CRIA, 2012). All entries are carefully checked for doubtful determinations and synonyms by consulting the taxonomic literature, the 'Flora do Brasil' (<http://floradobrasil.jbrj.gov.br/>) and the 'Flora del Conosur' (Zuloaga & Belgrano, 2015; <http://www.darwin.edu.ar/>), with additional direct consultation of taxonomists. Our data exclude checklists with < 10 species, because in lowland tropical regions, this is invariably attributable to low sampling or collecting efforts, rather than truly low species richness.

The vegetation type for each site, as documented in the original data source, is recorded and standardized to the vegetation types of Oliveira-Filho (2017; see also Supporting Information Table S1). When a herbarium voucher of an additional species is noted to come from within a 5 km radius of the original site, the collection label is checked to ensure that the species is found in the same vegetation type. Where two or more sites of different vegetation types co-occur within 10 km (768 sites; 19.13% of our total), this results in geographically overlapping sites in the NTT database, each for a distinct vegetation type. Further details of NTT history, protocols and data can be found at [www.neotropree.info](http://www.neotropree.info). We restricted analyses to the tropical and neighbouring subtropical lowlands of South America east of the Andes and did not include any NTT site > 1,000 m elevation or < 36° S latitude. Montane areas were excluded because biogeographical barriers might be playing significant roles in floristic differentiation. Including subtropical sites allowed us to contextualize our results from the tropics. In total, we included 4,103 individual sites, containing 10,306 tree species from 1,062 genera and 148 families.

### 2.2 | Statistical analyses

We performed hierarchical clustering based on tree species composition to assign sites to groups in an unsupervised manner (i.e., without reference to any environmental data). For clustering, we used the Simpson floristic distance amongst sites, which is the complement of the number of species shared between two sites divided by the maximum number of species that could be shared between the two sites:  $1 - \text{species}_{\text{shared}} / \text{total\_species}_{\text{minimum}}$  (Baselga, 2010). This is identical to the  $\beta_{\text{sim}}$  metric (Kreft & Jetz, 2010), but we use the term Simpson distance because of its historical precedence (Baselga, 2010). This metric isolates the effects of species turnover and is not confounded by large differences in species richness amongst sites (Baselga, 2010). We built 1,000 clusters, each after randomizing the row order in the matrix (species per site), following the procedure of Dapporto et al. (2013). We removed 24 sites that were unstable in their placement across the 1,000 clusters, which were identified by co-opting an approach used in phylogenetics to identify 'rogue taxa' that reduce resolution in

phylogenetic analyses (Aberer, Krompass, & Stamatakis, 2012). In the final consensus cluster, only those groups that were present in  $\geq 50\%$  of the clusters are distinguished (Omland, Cook, & Crisp, 2008). This analysis was performed in R (R Core Team, 2017) using the 'recluster' package (Dapporto et al., 2015).

To determine the biome identity of clusters, we used a reciprocal illumination procedure of assessing the overall structure of the cluster while considering site vegetation types (see Supporting Information Table S1). This process is inherently fractal, and one could identify increasingly smaller groups of sites. We focused on defining biomes in the broadest sense in order to increase their generality and utility, and our delimitations were performed in the context of the main biomes that have previously been proposed for LTSA, namely wet or moist tropical forests (hereafter wet forests), SDTF, subtropical forests, savanna and chaco woodlands. In essence, our approach tested whether there is floristic integrity to these previously proposed biomes, and we found clear evidence that there is (i.e., higher-level groups were composed largely of one broad biome type; Supporting Information Table S1). For heuristic purposes, we constructed a continuous biome map by applying Thiessen's polygons method in ArcGIS 10.4.1 (ESRI, 2017). This approach expands a polygon of a given biome classification for each NTT site until the polygons from neighbouring NTT sites are encountered. If they represent the same biome, then the polygons are fused, and this procedure is continued until the entirety of the study area is categorized to biome.

We assessed which sites might be intermediate or transitional between our biomes using a silhouette analysis, via the R package cluster (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2017). We also visually assessed where these ambiguously classified sites are located in species compositional space by means of a non-metric multidimensional scaling analysis (NMDS; McCune, Grace, & Urban, 2002) of sites in two dimensions based on the Simpson distance amongst sites.

### 2.3 | Using climate and edaphic data to distinguish biomes

To assess whether the biomes identified could be distinguished using climatic data, with or without edaphic data, we used a random forest classification tree approach (Breiman, 2001), implemented in the randomForest package in the R statistical software (Liaw & Wiener, 2002). We used 19 bioclimatic variables developed by Hijmans, Cameron, Parra, Jones, and Jarvis (2005), which quantify various aspects of temperature and precipitation regimes, and an estimate of average maximum climatological water deficit (CWD) per year (Chave et al., 2014). As edaphic variables, we included pH (extracted with KCl), cation exchange capacity (in centimoles per kilogram) and the percentage of sand, silt and clay extracted from SoilGrids v0.5.5 (<https://soilgrids.org/>; Hengl et al., 2017) at four different soil depths (0, 5, 15 and 30 cm), which were then averaged. Two different classifications were performed, one considering climatic data alone and the other considering both climatic and soil data.

In order to assess the success rate of the classification tree approach in assigning sites to biomes and to determine which biomes

were incorrectly classified, we generated confusion matrices, which show assignment based on climate alone or climate and soil versus assignment done above based on vegetation type and tree species composition. We also estimated the importance of each variable for distinguishing biomes using Breiman's measure of importance (Breiman, 2001). As we had substantial variation in sample size amongst our biomes that could bias importance measures, we equalized the number of sites across all biomes by rarefying to the number of sites present in the most poorly sampled biome. Rarefactions were performed randomly 100 times, and variable importance values were averaged across the 100 replicates. In order to understand climatic overlaps amongst biomes, we additionally plotted sites in a pairwise manner for key climatic variables (MAP, MAT and CWD).

### 2.4 | Comparison with existing biome maps

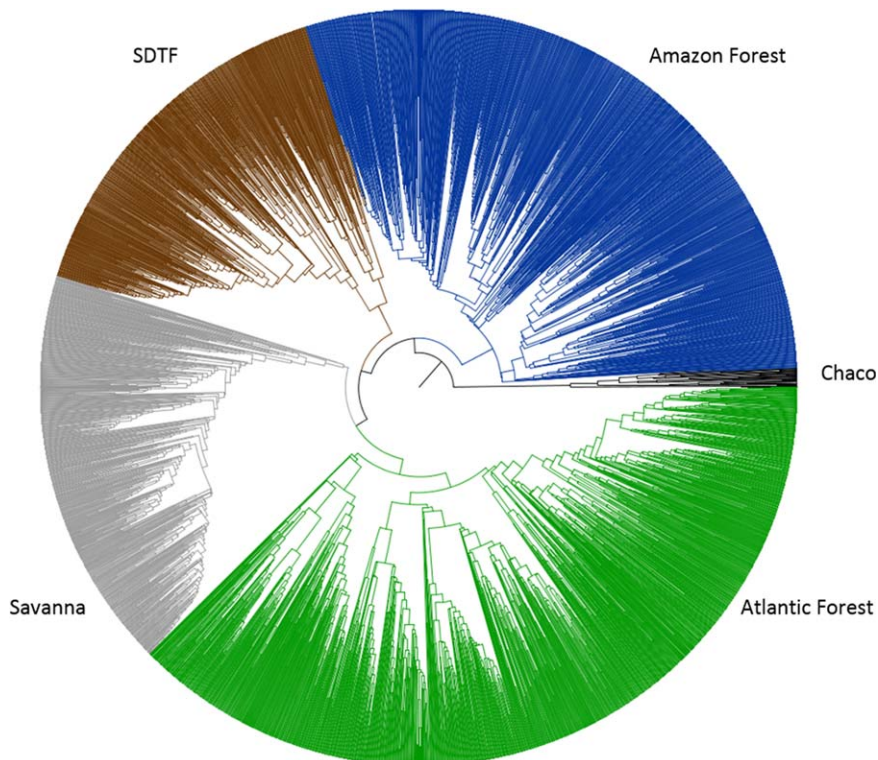
We looked at how two commonly used vegetation maps for South America classify sites to biome compared with our analyses. We focused on the map of Dinerstein et al. (2017), in which ecoregions are grouped into biomes and which is a revised version of Olson et al. (2001), and the Brazilian Domain system (IBGE, 2012). We determined which biomes and domains in these systems conceptually correspond to the biomes we established here and assessed how often these mapping systems gave the same identity to our NTT sites. The ecoregion data layer were obtained from <https://ecoregions2017.appspot.com/> and the IBGE Domain data layer from <http://www.geoservicos.ibge.gov.br/geoserver/web/> (layer CREN:biomas\_5000).

## 3 | RESULTS

### 3.1 | Biomes of lowland tropical South America

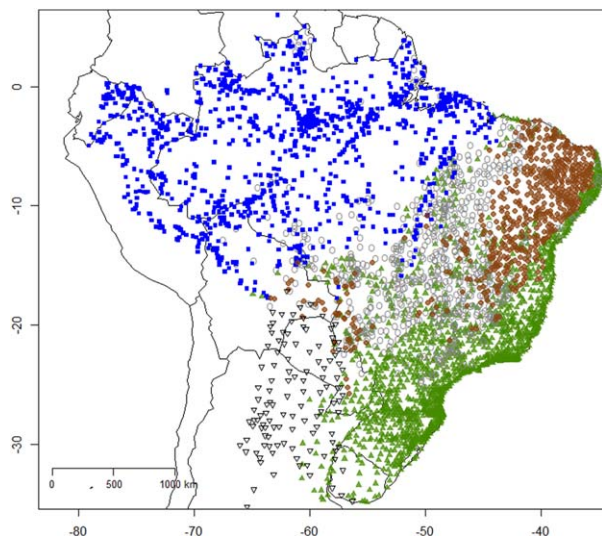
Hierarchical cluster analysis produced five higher-level groups (Figure 1), which we designated as biomes based on a priori vegetation type classifications. Wet forests fell into two different groups, which we tentatively treat as separate biomes. One comprises sites in the Amazon and the Guiana Shield, which we refer to as the Amazon Forest biome, and the other is composed of sites along the Atlantic coast, which we refer to as the Atlantic Forest biome (Figure 2). These two biomes are largely concordant with the Amazon and Atlantic Forest Domains, except that they also include semideciduous and gallery forests, found well outside of the geographical areas of the forest Domains (Figure 2).

The other three major groups in the cluster are found primarily in the dry diagonal, which extends from northeast Brazil to Bolivia, Paraguay and northern Argentina (Figure 2). One, which we refer to as Savanna, comprises sites with a grassy understorey found throughout central Brazil and eastern Bolivia, overlapping with the Cerrado Domain, but with disjunct occurrences in the Amazon Forest and Atlantic Forest biomes. The Savanna biome is clearly distinguished floristically from a biome that we term SDTF, based on the original vegetation classifications of sites (Supporting Information Table S1). The SDTF biome has a discontinuous distribution from the Pantanal and Chiquitania in Bolivia and southern Brazil to its largest extension in the



**FIGURE 1** Hierarchical cluster of 4,103 sites in lowland (< 1,000 m a.s.l.) tropical South America and neighbouring subtropical areas based on tree species composition. Five principal higher-level groups can be observed, which we refer to as the Amazon Forest (blue), Atlantic Forest (green), Savanna (grey), seasonally dry tropical forest or SDTF (brown) and Chaco (black) biomes. See main text for details

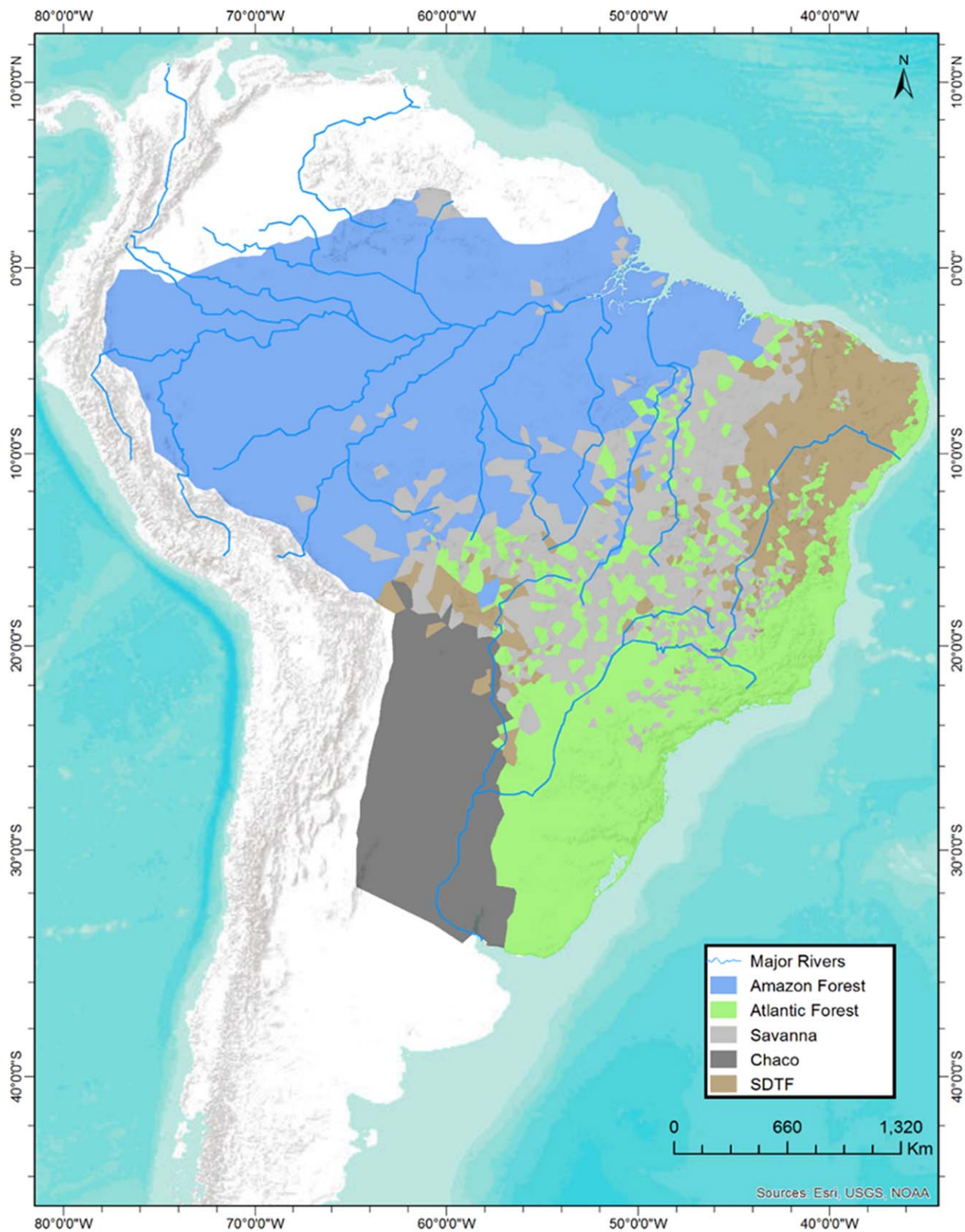
Caatinga Domain of northeastern Brazil (Figure 2). It is spatially interdigitated with the Savanna biome. The last group that we distinguish as a separate biome is the Chaco, comprising woodlands in Bolivia,



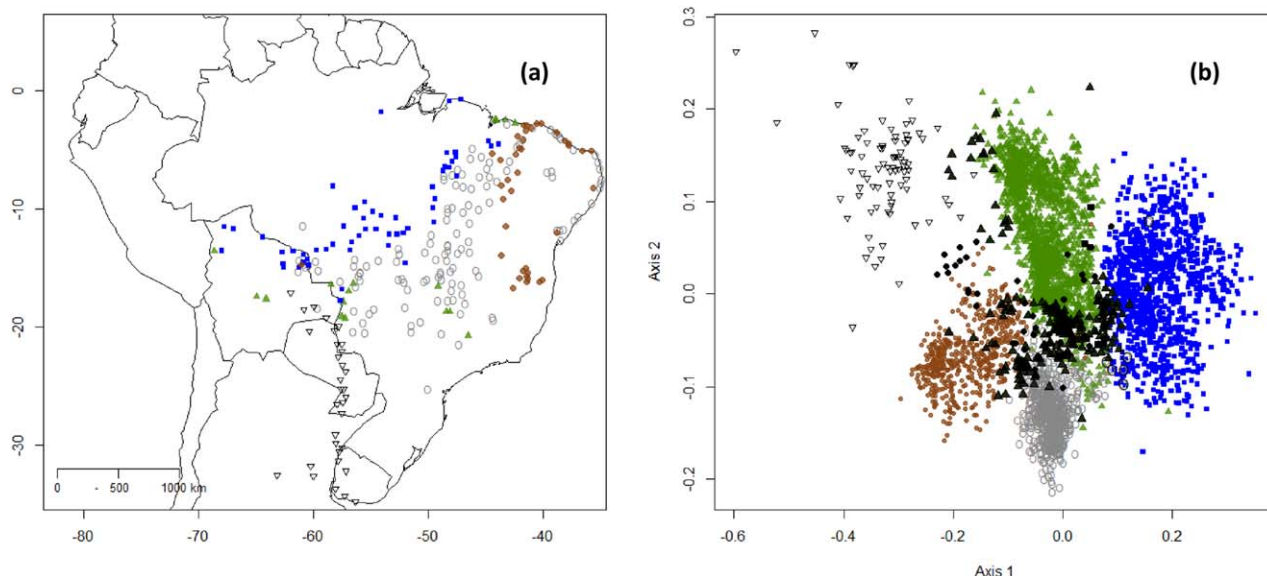
**FIGURE 2** Map of lowland tropical South America, with sites classified into biomes based on hierarchical cluster analysis of tree species composition: Atlantic Forest (green triangles), seasonally dry tropical forest (brown circles), Savanna (open grey circles), Amazon Forest (blue squares), Chaco (inverted open black triangles). Sites that were revealed to be more similar floristically to a different biome from the one with which they originally clustered are here given the symbol of the floristically more similar biome

Argentina and Paraguay and extending to the borders of southern Brazil. Although most of the sites in the Chaco biome cluster are subtropical and experience frost, there are a significant number of sites found north of 23° latitude that are unlikely to experience frost and can be considered tropical (Figure 2). See Supporting Information (Appendix S1) for further description of the biomes. Our continuous biome map, developed using the Thiessen's polygons method, shows the overall spatial distribution of the LTSA biomes and highlights the regions in which they interdigitate (Figure 3).

Of 4,103 sites, 1,097 were classified as Amazon Forest, 1,566 as Atlantic Forest, 760 as Savanna, 564 as SDTF and 116 as Chaco. Silhouette analysis (Supporting Information Figure S1) showed that 271 sites are floristically more similar to a different biome than that with which they were original clustered, which we interpret to indicate that these sites are transitional between two biomes (Figure 4a; Supporting Information Table S2). An ordination of sites (NMDS with two axes, stress value = 0.1816) also suggests that these sites are compositionally transitional (Figure 4b). Floristically transitional sites were common between the Amazon and Atlantic Forest biomes (53 sites), between the Savanna and Atlantic Forest biomes (115 sites) and between the SDTF and Atlantic Forest biomes (49 sites), whereas they were infrequent between other biomes, including between any pair of dry biomes. Floristically transitional sites are common in the dry diagonal (Figure 4a), particularly between the Cerrado and the Amazon Forest and between the Chaco and the Atlantic Forest. Many of the gallery forests within the Cerrado Domain also have an ambiguous tree species compositional identity and are therefore difficult to classify.



**FIGURE 3** Map of South America, with a schematic representation of the biomes delimited via hierarchical cluster analysis in the present contribution [Amazon Forest, Atlantic Forest, Savanna, Chaco and seasonally dry tropical forest (SDF)]. The map was created by applying the Thiessen polygons method on the categorized points presented in Figure 2. See text for further details

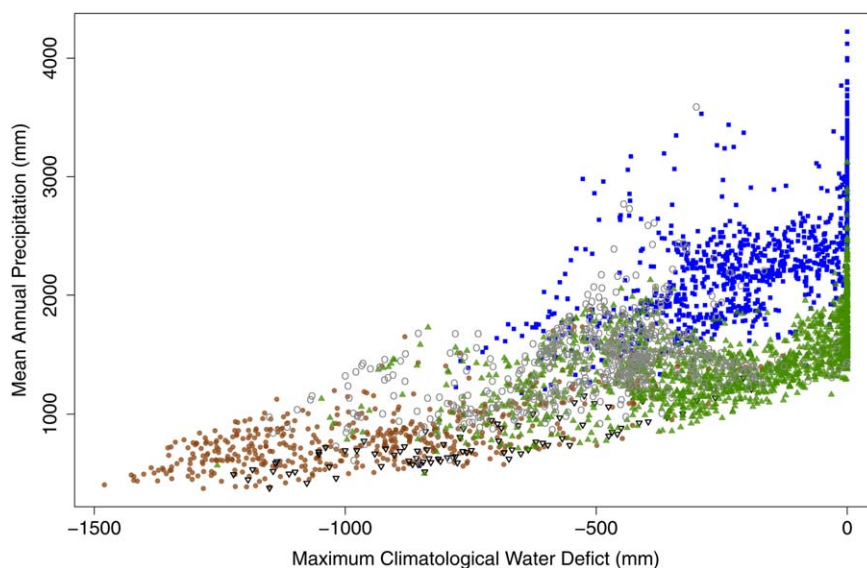


**FIGURE 4** NeoTropTree sites that have a transitional/ambiguous floristic identity, as revealed by the silhouette analysis, and how they are distributed in (a) geographical and (b) species compositional spaces. In (a), sites are categorized according to the biome to which they are floristically more similar. In (b), correctly classified sites are shown in the same colour scheme as Figure 2, whereas misclassified sites are represented in black and in the same shape as the sites of their biome based on the original clustering analysis. Symbols are as follows: green triangles = Atlantic Forest, brown circles = seasonally dry tropical forest, open grey circles = Savanna, blue squares = Amazon Forest and inverted open black triangles = Chaco

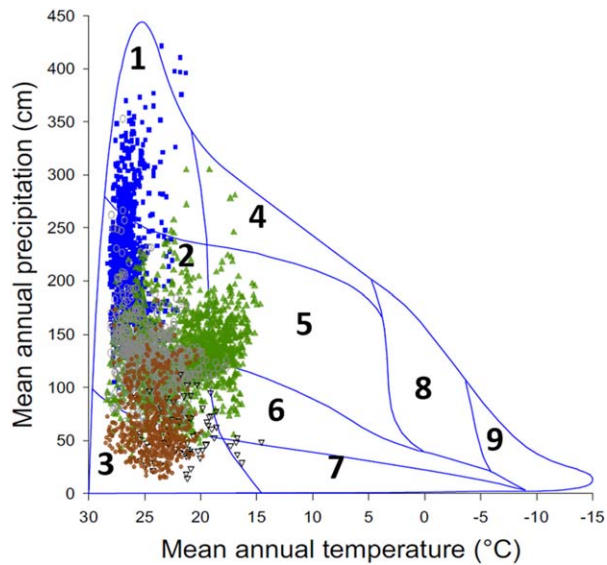
### 3.2 | Using climate and edaphic data to distinguish biomes

We find that biomes overlap substantially in climatic space, in terms of both water availability (Figure 5) and temperature (Figure 6). For example, all five biomes defined here occupy at least two of the climatic

biomes proposed by Whittaker (1975) (Figure 6). Of the 3,832 sites that are not considered transitional in nature, 712 were misclassified based on climate (18.6% of sites; Table 1). Considering all sites together, including transitional ones, we found a slightly higher error rate of 20.7% (Supporting Information Table S3). The most common misclassifications involved Amazon or Atlantic Forest sites being



**FIGURE 5** Distribution of sites with respect to precipitation regime. Mean annual precipitation values come from WorldClim (Hijmans et al., 2005), and maximum climatological water deficit comes from Chave et al. (2014). Symbols are as follows: green triangles = Atlantic Forest, brown circles = seasonally dry tropical forest, open grey circles = Savanna, blue squares = Amazon Forest and inverted open black triangles = Chaco. Modelled after fig. 1 of Malhi et al. (2009), which suggested that savannas were drier than seasonal forests, contrary to the pattern here



**FIGURE 6** Distribution of sites in climatic space across the nine biomes proposed by Whittaker, R. H. (1975), considering mean annual precipitation (in centimetres) and mean annual temperature (in degrees Celsius). Numbers are as follows: 1 = tropical rain forest, 2 = tropical seasonal forest/savanna, 3 = tropical and subtropical desert, 4 = temperate rain forest, 5 = temperate deciduous forest, 6 = woodland/scrubland, 7 = temperate grassland/dessert, 8 = boreal forest and 9 = tundra. Symbols are as follows: green triangles = Atlantic Forest, brown circles = seasonally dry tropical forest, open grey circles = Savanna, blue squares = Amazon Forest and inverted open black triangles = Chaco

classified as belonging to the Savanna biome or vice versa, and climatic misclassifications of SDTF and Savanna were also common (Table 1). Sites in the Amazon and Atlantic Forest wet biomes were distinct climatically. Meanwhile, the Chaco biome was rarely confused climatically with any of the other biomes. These patterns did not change when sites that have centres within 10 km of each other (i.e., overlapping in geographical space) were removed (Supporting Information Table S4; error rate 20.3%).

**TABLE 1** Confusion matrix between sites categorized based on floristic composition via hierarchical clustering (rows) and sites categorized using climate and a classification tree approach (columns)

	Amazon Forest	Atlantic Forest	Savanna	Chaco	SDTF
Amazon Forest	989	6	45	0	0
Atlantic Forest	3	1,290	199	5	50
Savanna	58	167	357	0	50
Chaco	0	7	0	76	1
SDTF	0	51	65	1	408

*Note.* The diagonal gives the number of sites that are correctly classified by climate, and the off-diagonal elements give misclassifications (18.6%). Only non-floristically transitional sites were considered. Accuracy: 81%; average precision: 81%; average recall: 80%. SDTF = seasonally dry tropical forest.

**TABLE 2** Confusion matrix between sites categorized based on floristic composition via hierarchical clustering (rows) and sites categorized using climate + soil and a classification tree approach (columns)

	Amazon Forest	Atlantic Forest	Savanna	Chaco	SDTF
Amazon Forest	1,001	4	37	0	0
Atlantic Forest	4	1,331	161	4	49
Savanna	48	121	423	0	40
Chaco	0	7	0	76	1
SDTF	0	55	52	1	417

*Note.* The diagonal gives the number of sites that are correctly classified by climate+soil data, and the off-diagonal elements give misclassifications (15.2%). Accuracy: 84%; average precision: 84%; average recall: 83%. SDTF = seasonally dry tropical forest.

The inclusion of edaphic variables slightly increased overall classification success by 3.2% (Table 2), and by 3% when transitional sites were included (Supporting Information Table S5). There were a total of 124 sites that switched from being classified incorrectly (with climatic data only) to being classified correctly (once edaphic data were included; Table 2). Most of these were Savanna sites classified as Atlantic Forest and vice versa.

Whether or not edaphic variables are included, the three most important variables for classification were MAP, temperature seasonality and maximum climatological water deficit (Table 3). Overall, climatic variables seem to be more important than edaphic variables for distinguishing biomes, with variables related to precipitation, water availability and temperature seasonality ranking higher than variables related to mean temperature. However, overall we do have fewer edaphic variables, and pH and cation exchange capacity (CEC) are among the top 10 variables (Table 3).

### 3.3 | Comparison with existing biome maps

The classification systems developed by Olson et al. (2001) and Dinerstein et al. (2017) and IBGE (2012) assigned 74–75% of the NTT sites to the same biomes in which they were placed according to our analyses (74.7%, Dinerstein et al., 2017; Supporting Information Table S6; 74.5%, IBGE, 2012; Supporting Information Table S7). In Dinerstein's system, the majority of the misclassification results from Atlantic Forest sites being incorrectly classified as tropical or Subtropical Savannas and Savanna being classified as tropical moist forest (Supporting Information Figure S2). In the IBGE system, the error rate largely stems from SDTF sites being classified as Cerrado and vice versa (Supporting Information Figure S3).

## 4 | DISCUSSION

Our study demonstrates that using climatic data alone, with or without supplementary edaphic data, to map biomes would result in substantial error, causing misclassification of 15.2–20.7% of sites. Such misclassifications are attributable to pronounced climatic overlap of biomes



**TABLE 3** The mean variable importance value ( $\pm$  SE) for all climatic variables included in the random forest analysis across 100 runs of the Breiman's algorithm using rarefactions of the main dataset (116 sites per biome)

Environmental variables	Climate (mean $\pm$ SE)	Climate + soil (mean $\pm$ SE)
Mean annual precipitation (mm)	356.81 $\pm$ 1.09	318.8 $\pm$ 1.18
Temperature seasonality ( $^{\circ}$ C)	319.73 $\pm$ 1.23	287.14 $\pm$ 1.13
Maximum climatological water deficit (mm/yr)	273.2 $\pm$ 0.69	232.07 $\pm$ 0.71
Isothermality (%)	233.29 $\pm$ 0.98	211.53 $\pm$ 0.87
pH (KCl)	-	188.98 $\pm$ 0.84
Mean temperature of coldest quarter ( $^{\circ}$ C)	187.06 $\pm$ 0.95	163.07 $\pm$ 0.97
Precipitation of wettest quarter (mm)	155.06 $\pm$ 0.56	120.57 $\pm$ 0.48
Cation exchange capacity (cmol/kg)	-	119.89 $\pm$ 0.23
Precipitation of driest quarter (mm)	148.46 $\pm$ 0.53	119.37 $\pm$ 0.51
Precipitation of driest month (mm)	133.16 $\pm$ 0.49	109.94 $\pm$ 0.44
Mean annual temperature ( $^{\circ}$ C)	122.75 $\pm$ 0.71	96.15 $\pm$ 0.66
Precipitation of wettest month (mm)	119.83 $\pm$ 0.42	91 $\pm$ 0.35
Mean temperature of driest quarter ( $^{\circ}$ C)	106.46 $\pm$ 0.57	81.93 $\pm$ 0.49
Amount of sand (%)	-	81.73 $\pm$ 0.17
Maximum temperature of warmest month ( $^{\circ}$ C)	103.8 $\pm$ 0.33	81.69 $\pm$ 0.31
Amount of silt (%)	-	76.89 $\pm$ 0.13
Temperature annual range ( $^{\circ}$ C)	101.51 $\pm$ 0.32	75.32 $\pm$ 0.23
Precipitation seasonality (%)	99.22 $\pm$ 0.24	74.3 $\pm$ 0.31
Minimum temperature of coldest month ( $^{\circ}$ C)	99.21 $\pm$ 0.23	73.38 $\pm$ 0.37
Precipitation of warmest quarter (mm)	98.61 $\pm$ 0.3	70.77 $\pm$ 0.18
Precipitation of coldest quarter (mm)	97.11 $\pm$ 0.47	69.21 $\pm$ 0.25
Diurnal range of temperature ( $^{\circ}$ C)	91.45 $\pm$ 0.19	68.67 $\pm$ 0.16
Amount of clay (%)	-	65.97 $\pm$ 0.13
Mean temperature of wettest quarter ( $^{\circ}$ C)	79.01 $\pm$ 0.22	61.57 $\pm$ 0.24
Mean temperature of warmest quarter ( $^{\circ}$ C)	60.71 $\pm$ 0.12	46.52 $\pm$ 0.16

(Figures 5 and 6) and to edaphic heterogeneity at small spatial scales that is not captured by available data, which are derived via interpolation among relatively sparse soil sampling sites. Recently, researchers have begun assigning study sites to biomes, generally those of Whittaker (1975), based solely on climatic values, such as MAP and MAT (e.g., Díaz et al., 2016; Qian, Jin, & Ricklefs, 2017; Siepielski et al., 2017). Our results suggest that this is potentially problematic (Figure 6). For example, the Amazon and Atlantic Forests can both occur in areas that are more seasonal than 'tropical rain forest' (*sensu* Whittaker), whereas the Savanna biome can occur in much wetter areas than indicated by Whittaker (1975; see also Lehmann et al., 2014). It is notable that none of our five major biomes is restricted to a single biome in Whittaker's climatic biome classification (Figure 6).

We were able to take a floristic approach to mapping biomes at a continental scale. Recent biome maps of LTSA, generally based on remote sensing, either fail to include major biomes (e.g., seasonally dry tropical forest is absent from Hirota et al., 2011; Staver et al., 2011) or

are unable to distinguish amongst the dry tropical biomes of Savanna and SDTF (Beuchle et al., 2015). Although floristic approaches to mapping biomes are unlikely to succeed inter-continently because of the lack of shared species or even genera at this scale (Dexter et al., 2015), the increasing availability of floristic composition and species distribution information (e.g., [www.gbif.org](http://www.gbif.org), [www.forestplots.net](http://www.forestplots.net), [www.neotrotree.info](http://www.neotrotree.info)) should allow this approach to be implemented within continents. It is important to note that any complete and continuous (or 'wall-to-wall') map of biome distribution will be inaccurate at small spatial scales owing to high edaphic and floristic heterogeneity coupled with incomplete sampling. We have generated a continuous map (Figure 3), but its purpose is as a heuristic scheme to understand patterns in the distribution of biomes in LTSA. We do not contend that every point on the map is accurately classified, as that would belie one of the principal outcomes of this study, that of high biome heterogeneity at small spatial scales, as previously noted by Collevatti et al. (2013), Pennington et al. (2006) and Werneck (2011).

## 4.1 | Biomes of lowland tropical South America

Our analyses suggest three to five major biomes in LTSA. The Amazon and Atlantic Forests might represent separate biomes, whereas previously they have often been considered as a single tropical wet/moist forest biome. They are floristically distinct, and their climatic niches are almost completely non-overlapping. Our floristic circumscription of the Atlantic Forest matches the *sensu latissimo* definition of Oliveira-Filho, Jarenkow, & Rodal (2006). Our delimitation of the Amazon Forest is similar to previous studies that include the majority of the Amazon Basin drainage and the Guianan Shield (e.g., Prance, 1982; ter Steege et al., 2006), although we note that our sampling of the Guianan Shield is limited.

The Savanna biome is floristically distinct from the other dry biomes, which is expected because it is a uniquely disturbance-driven system, strongly influenced by fire (Archibald, Lehmann, Gómez-Dans, & Bradstock, 2013; Ratter et al., 1997). Many sites in the SDTF biome are often drier, in terms of MAP and CWD, than the majority of sites in the Savanna biome (Figure 5), which runs counter to thinking that tropical wet forest transitions to tropical seasonal forest and then to savanna as water availability declines (e.g., Malhi et al., 2009). Meanwhile, our results from floristic analyses give support to previous studies (DRYFLOR, 2016; Pennington, Lavin, & Oliveira-Filho, 2009; Pennington et al., 2000; Prado & Gibbs, 1993) arguing that the SDTFs scattered across lowland tropical South America should be regarded as a single biome, with the exclusion of the Chaco. We find that the climatic niches of Chaco and SDTF do not overlap, with the Chaco occurring in a colder climate with much higher temperature seasonality. However, further studies are needed that compare ecosystem function in the Amazon versus Atlantic Forests and in the SDTF versus Chaco to verify their status as distinct biomes. For further discussion of floristic patterns within and across biomes, please refer to the Supporting Information (Appendix S1).

## 4.2 | Using climate and edaphic data to distinguish biomes

Mean annual precipitation, several measures of dry season precipitation and water deficit, temperature variability and soil pH were the most important environmental variables in distinguishing major biomes (Table 2). The result that precipitation-related variables are on average more important than temperature-related variables is to be expected, given that the majority of our sampling and most of the biomes under study are within the tropics, and thus represent a limited range of non-freezing temperature regimes (Augusto, Davies, Delzon, De Schrijver, 2014). Nevertheless, it is notable that measures of temperature variability, particularly across seasons, were more important than other temperature measures, including MAT and the minimal temperature of the coldest month. This may be because plant species ranges are often constrained by how much temperature can vary in a given location and by temperature extremes (O'Sullivan et al., 2017).

Although a classification success rate of 80% seems high, this would result in one in five sites being misclassified, which is potentially

problematic for conservation and management decisions. Some sites are floristically transitional in nature and inherently difficult to classify. Such transitional sites might be particularly resilient, and thus important, under future climate change, and they may require their own management regimes (Prieto-Torres et al., 2016). Regardless, the high error rate (18.6%) among non-transitional sites (sites not detected by the silhouette analysis as belonging to a different biome) is still of concern, because they compose 93.4% of our sites. In order to improve classification of these sites to biome based on environmental data, environmental data at better resolution are needed. Publicly available environmental data are derived from interpolation. For climate, which varies at a relatively broad spatial grain, this might not be problematic. However, edaphic data vary at a small spatial grain, and interpolation-based methods might be inadequate to capture edaphic conditions at many sites. Also, the edaphic data from SoilGrids do not include variables, such as soil fertility (sum of bases), phosphorous and aluminium content, which are highly relevant to tree species growth. Meanwhile, other non-climatic and non-edaphic variables, such as fire and disturbance, might play a significant role in determining tree species composition at local sites, and biome identity more widely. For example, SDTF and wet forest can convert to savanna if there is sufficient disturbance via fire or anthropogenic woody biomass removal (Devisscher, Anderson, Aragão, Galván, & Malhi, 2016).

## 4.3 | Comparison with existing biome maps

The comparisons between the classification system presented here and those of Dinerstein et al. (2017) and the Domain system (IBGE, 2012) revealed a c. 25% misclassification rate for the latter two. These high error rates stem from two sources: the intrusion of SDTF and the Atlantic and Amazon Forests (as gallery forest) into the Savanna biome in the dry diagonal, and the existence of non-equivalent categories among these systems. Dinerstein et al. (2017) and IBGE (2012) recognize tropical and subtropical wetlands (named Pantanal in the IBGE system) as a distinct biome or domain, and the IBGE Domain system also delimits the Pampas (also known as Campos Sulinos, southern Brazilian steppes). These two categories have not been detected and classified by our approach. Rather, the region classified as Pantanal by IBGE (2012) is covered by a mix of different vegetation formations that are floristically similar to SDTFs, Savannas and also the semideciduous portion of the Atlantic Forest. The forests within the area known as the Pampas in South Brazil are floristically similar, in relationship to tree species composition, to the rest of the subtropical portion of the Atlantic Forest biome (Oliveira-Filho, Budke, Jarenkow, Eisenlohr, & Neves, 2015).

## 4.4 | Synergies between biodiversity conservation and ecosystem management

Delimiting biomes based on tree species composition offers the possibility of synergy between ecosystem management planning and conservation prioritization. The biomes we have delimited differ in tree species composition and are therefore likely to differ in ecosystem

function. Ecosystem management plans should therefore be developed separately for each. Likewise, these biomes have almost no species in common, yet they have many species unique to them. Our schematic map (Figure 3) also indicates how these biomes are distributed at a continental scale, highlighting how discontinuous biome distribution can be in LTSA. These are important observations that must be considered in conservation and management. As an example, it is only recently that the SDTF has been recognized as a biome (Gentry, 1995; Murphy & Lugo, 1986; Prado & Gibbs, 1993), a definition consistent with our analyses, and there is no synthetic conservation plan that addresses the biome as a whole across the Neotropics (although for first steps, see DRYFLOR, 2016). Current conservation planning for SDTF in Brazil focuses solely on the Caatinga Domain, but many Brazilian SDTFs are found in disjunct patches outside this area, especially in the Cerrado, placing them under laws designed to protect savanna diversity. As another example, the Chaco is under great threat owing to an increase of habitat destruction and fragmentation during the last 30 years (Hansen et al., 2013; Nori et al., 2016), but if recognized as a separate biome, as our analyses suggest, the urgency of its conservation might be better appreciated (Kuemmerle et al., 2017).

## 5 | CONCLUSIONS

We have mapped the principal biomes in LTSA by using information on tree species composition of > 4,000 sites. The Savanna, Amazon and Atlantic Forest and SDTF biomes have an interdigitated distribution in central South America and overlap substantially in climatic space. Biome distribution cannot therefore be fully accounted for by climate, suggesting that climate projections alone will be insufficient to predict future biome shifts. Additional, meaningful environmental variables (e.g., available nitrogen, phosphorous, aluminium) must be measured and accounted for in models. The interdigitation of biomes, especially in the dry diagonal across Brazil, is not recognized in the current IBGE (2012) system on which Brazilian conservation legislation is based, leading to the neglect of highly threatened SDTF vegetation outside the Caatinga Domain. Our analyses also show that Chaco and SDTF are distinct, which must be considered in land management and conservation. We suggest that species composition can be central to delimiting meaningful biomes for comparative research and conservation.

## ACKNOWLEDGMENTS

P.L.S.M. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES) for supporting a full PhD at the University of Edinburgh under the Science without Borders Programme (grant 99999.013197/2013-04). A.T.O.F. was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (CNPq; grant 301644/88-8). D.M.N., R.T.P. and K.G.D. were supported by the National Environment Research Council (grant NE/I028122/1). D.M.N. was also supported by the National Science Foundation – USA (grant NSF/DEB-1556651). K.G.D. was also supported by a Leverhulme Trust International Academic Fellowship (obs: there is no grant number for this fellowship). P.L.S.M. also thanks Chrystiann Lavarini for help in making Figure 3.

## DATA ACCESSIBILITY

The data used to produce this paper can be freely accessed at <http://www.neotroptree.info/>.

## ORCID

Pedro Luiz Silva de Miranda  <http://orcid.org/0000-0002-3008-1556>

## REFERENCES

- Aberer, A. J., Krompass, D., & Stamatakis, A. (2012). Pruning rogue taxa improves phylogenetic accuracy. *Systematic Biology*, 62, 162–166.
- Archibald, S., Lehmann, C. E. R., Gómez-Dans, J. L., & Bradstock, R. A. (2013). Defining pyromes and global syndromes of fire regimes. *Proceedings of the National Academy of Sciences USA*, 110, 6442–6447.
- Augusto, L., Davies, T. J., Delzon, S., & De Schrijver, A. (2014). The enigma of the rise of angiosperms: Can we untie the knot? *Ecology Letters*, 17, 1326–1338.
- Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19, 134–143.
- Beuchle, R., Grecchi, R. C., Shimabukuro, Y. E., Seliger, R., Eva, H. D., Sano, E., & Achard, F. (2015). Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography*, 58, 116–127.
- Box, E. O. (1995). Factors determining distributions of tree species and plant functional types. *Vegetatio*, 121, 101–116.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., ... Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20, 3177–3190.
- Clements, F. E. (1916). The development and structure of biotic communities. *Journal of Ecology*, 5, 119–128.
- Collevatti, R. G., Lima-Ribeiro, M. S., Diniz-Filho, J. A. F., Oliveira, G., Dobrovolski, R., & Terribile, L. C. (2013). Stability of Brazilian seasonally dry forests under climate change: Inferences for long-term conservation. *American Journal of Plant Sciences*, 4, 792–805.
- CRIA (2012). *speciesLink*. Retrieved from <http://www.splink.org.br/index>
- Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R., & Dennis, R. L. H. (2013). Recluster: An unbiased clustering procedure for beta-diversity turnover. *Ecography*, 36, 1070–1075.
- Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R., & Dennis, R. L. H. (2015). Package 'recluster' - Ordination methods for the analysis of beta-diversity indices. Retrieved from <https://cran.r-project.org/web/packages/recluster/recluster.pdf>
- Devisscher, T., Anderson, L. O., Aragão, L. E. O. C., Galván, L., & Malhi, Y. (2016). Increased wildfire risk driven by climate and development interactions in the Bolivian Chiquitania, Southern Amazonia. *PLoS One*, 11, e0161323.
- Dexter, K. G., Smart, B., Baldauf, C., Baker, T. R., Balinga, M. P. B., Brienen, R. J. W., ... Pennington, R. T. (2015). Vegetation in seasonally dry regions of the tropics: Floristics and biogeography. *International Forestry Review*, 17, 10–32.
- Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., ... Gorné, L. D. (2016). The global spectrum of plant form and function. *Nature*, 529, 167–171.

- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., ... Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67, 534–545.
- DRYFLOR. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science*, 353, 1383–1387.
- DRYFLOR, Pennington, R. T., Banda-R, K., Delgado-Salinas, A., Dexter, K. G., Galetti, L., ... Särkinen, T. (2017). Forest conservation: Remember Gran Chaco—Response. *Science*, 355, 465–466.
- Engemann, K., Sandel, B., Enquist, B. J., Jørgensen, P. M., Kraft, N., Marcuse-Kubitza, A., ... Svenning, J.-C. (2016). Patterns and drivers of plant functional group dominance across the Western Hemisphere: A macroecological re-assessment based on a massive botanical dataset. *Botanical Journal of the Linnean Society*, 180, 141–160.
- Environmental Systems Research Institute (ESRI). (2017). *ArcGIS 10.4.1*. Redlands, CA: Author.
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., ... Schaaf, C. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, 83, 287–302.
- Gentry, A. H. (1995). Patterns of diversity and floristic composition in Neotropical montane forest. In S. P. Churchill, H. Balslev, E. Forero, & J. L. Luteyn (Eds.), *Biodiversity and conservation of neotropical montane forest* (pp. 103–126). New York: The New York Botanical Garden.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342, 850–853.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12, e0169748.
- Higgins, S. I., Buitenwerf, R., & Moncrieff, G. R. (2016). Defining functional biomes and monitoring their change globally. *Global Change Biology*, 22, 3583–3593.
- 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.
- Hirota, M., Holmgren, M., Van Nes, E. H., & Scheffer, M. (2011). Global resilience of tropical forest and savanna to critical transitions. *Science*, 334, 232–235.
- Holdridge, L. R. (1947). Determination of world plant formations from simple climatic data. *Science*, 105, 367–268.
- Humboldt, C. A. (1816). XCIII. On the laws observed in the distribution of vegetable forms. *The Philosophical Magazine*, 47, 446–453.
- Instituto Brasileiro de Geografia e Estatística (IBGE). (Ed.). (2012). *Manual técnico da vegetação brasileira* (2ª edição revista e ampliada). Rio de Janeiro, Brazil: Author.
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37, 2029–2053.
- Kuemmerle, T., Baldi, G., Camino, M., Cuellar, E., Cuellar, R. L., Decarre, J., ... Semper-Pascual, A. (2017). Forest conservation: Remember Gran Chaco Forest. *Science*, 355, 465.
- Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., Imova, I., ... Enquist, B. J. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences USA*, 111, 13745–13750.
- Lehmann, C. E. R., Anderson, T. M., Sankaran, M., Higgins, S. I., Archibald, S., Hoffmann, W. A., ... Bond, W. J. (2014). Savanna vegetation-fire-climate relationships differ among continents. *Science*, 343, 548–553.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2, 18–22.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2017). *Cluster: Cluster analysis basics and extensions*. R package version 2.0.6.
- Malhi, Y., Aragao, L. E. O. C., Galbraith, D., Huntingford, C., Fisher, R., Zelazowski, P., ... Meir, P. (2009). Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proceedings of the National Academy of Sciences USA*, 106, 20610–20615.
- McCune, B., Grace, J. B., & Urban, D. L. (2002). *Analysis of ecological communities* (Vol. 28). Gleneden Beach, OR: MjM Software Design.
- Mittermeier, R. A., Gil, P. G., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., ... Fonseca, G. A. B. (2004). *Hotspots revisited* (1st ed.). Mexico City: CEMEX.
- Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, G. A. B., & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biology*, 12, 516–520.
- Moncrieff, G. R., Hickler, T., & Higgins, S. I. (2015). Intercontinental divergence in the climate envelope of major plant biomes. *Global Ecology and Biogeography*, 24, 324–334.
- Murphy, P. G., & Lugo, A. E. (1986). Ecology of tropical dry forest. *Annual Review of Ecology and Systematics*, 17, 67–88.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.
- Neves, D. M., Dexter, K. G., Pennington, R. T., Bueno, M. L., & Oliveira Filho, A. T. (2015). Environmental and historical controls of floristic composition across the South American Dry Diagonal. *Journal of Biogeography*, 42, 1566–1576.
- Nori, J., Torres, R., Lescano, J. N., Cordier, J. M., Periago, M. E., & Baldo, D. (2016). Protected areas and spatial conservation priorities for endemic vertebrates of the Gran Chaco, one of the most threatened ecoregions of the world. *Diversity and Distributions*, 22, 1212–1219.
- Odum, E. P. (1975). *Ecology* (2nd ed.). London: Holt Rinehart and Winston.
- Oliveira-Filho, A. T. (2017). *NeoTropTree, Flora arbórea da Região Neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação*. Retrieved from <http://www.neotropree.info/>
- Oliveira-Filho, A. T., Budke, J. C., Jarenkow, J. A., Eisenlohr, P. V., & Neves, D. R. M. (2015). Delving into the variations in tree species composition and richness across South American subtropical Atlantic and Pampean forests. *Journal of Plant Ecology*, 8, 242–260.
- Oliveira-Filho, A. T., Cardoso, D., Schrire, B. D., Lewis, G. P., Pennington, R. T., Brummer, T. J., ... Lavin, M. (2013). Stability structures tropical woody plant diversity more than seasonality: Insights into the ecology of high legume-succulent-plant biodiversity. *South African Journal of Botany*, 89, 42–57.
- Oliveira-Filho, A. T., Jarenkow, L. A., & Rodal, M. J. N. (2006). Floristic relationships of seasonally dry forests of Eastern South America based on tree species distribution patterns. In R. T. Pennington, J. A. Ratter, & G. P. Lewis (Eds.), *Neotropical savannas and seasonally dry forests: Plant diversity, biogeography, and conservation* (pp. 159–192). London: Taylor & Francis.
- Oliveira-Filho, A. T., Pennington, R. T., Rotella, J., & Lavin, M. (2014). Exploring evolutionarily meaningful vegetation definitions in the

- tropics: A community phylogenetic approach. In D. A. Coomes, D. F. R. P. Burslem, & W. D. Simonson (Eds.), *Forests and global change* (pp. 239–260). Cambridge: Cambridge University Press.
- Oliveira-Filho, A. T., & Ratter, J. A. (1995). A study of the origin of central Brazilian forests by the analysis of plant species distribution patterns. *Edinburgh Journal of Botany*, *52*, 141–194.
- Olson, D. M., & Dinerstein, E. (1998). The global 200: A representation approach to conserving the earth's most biologically valuable ecoregions. *Conservation Biology*, *12*, 502–515.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*, 933–938.
- Omland, K. E., Cook, L. G., & Crisp, M. D. (2008). Tree thinking for all biology: The problem with reading phylogenies as ladders of progress. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *30*, 854–867.
- O'Sullivan, O. S., Heskell, M. A., Reich, P. B., Tjoelker, M. G., Weerasinghe, L. K., Penillard, A., ... Atkin, O. K. (2017). Thermal limits of leaf metabolism across biomes. *Global Change Biology*, *23*, 209–223.
- Pennington, R. T., Lewis, G. P., & Ratter, J. A. (2006). An overview of the plant diversity, biogeography and conservation of neotropical savannas and seasonally dry forests. In: Pennington RT, Lewis GP, Ratter JA, (Eds.), *Neotropical savannas and seasonally dry forests: plant diversity, biogeography and conservation*. (pp. 1–29). Boca Raton: CRC Press.
- Pennington, R. T., Lavin, M., & Oliveira-Filho, A. (2009). Woody plant diversity, evolution, and ecology in the tropics: Perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 437–457.
- Pennington, R. T., Prado, D. E., & Pendry, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, *27*, 261–273.
- 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.
- Prance, G. T. (1982). *Forest refuges: Evidence from woody angiosperms*. In G. T. Prance (Ed.), *Biological diversification in the tropics* (pp. 137–156). New York, NY: Columbia University Press.
- Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Robert, A., Solomon, A. M., ... Solomon, A. M. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, *19*, 117–134.
- Prentice, K. C. (1990). Bioclimatic distribution of vegetation for general circulation models. *Journal of Geophysical Research*, *95*, 11811–11830.
- Prieto-Torres, D. A., Navarro-Siguenza, A. G., Santiago-Alarcon, D., & Rojas-Soto, O. R. (2016). Response of the endangered tropical dry forests to climate change and the role of Mexican Protected Areas for their conservation. *Global Change Biology*, *22*, 364–379.
- Prieto-Torres, D. A., & Rojas-Soto, O. R. (2016). Reconstructing the Mexican tropical dry forests via an autoecological niche approach: Reconsidering the ecosystem boundaries. *PLoS One*, *11*, e0150932.
- Qian, H., Jin, Y., & Ricklefs, R. E. (2017). Patterns of phylogenetic relatedness of angiosperm woody plants across biomes and life-history stages. *Journal of Biogeography*, *44*, 1383–1392.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ratter, J. A., Ribeiro, J. F., & Bridgewater, S. (1997). The Brazilian cerrado vegetation and threats to its biodiversity. *Annals of Botany*, *80*, 223–230.
- Sandel, B., Gutiérrez, A. G., Reich, P. B., Schrod, F., Dickie, J., & Kattge, J. (2015). Estimating the missing species bias in plant trait measurements. *Journal of Vegetation Science*, *26*, 828–838.
- Särkinen, T., Iganci, J. R. V., Linares-Palomino, R., Simon, M. F., & Prado, D. E. (2011). Forgotten forests - Issues and prospects in biome mapping using Seasonally Dry Tropical Forests as a case study. *BMC Ecology*, *11*, 1–15.
- Schimper, A. F. W. (1903). *Plant-geography upon a physiological basis*. Oxford: Clarendon Press.
- Siepielski, A. M., Morrissey, M. B., Buoro, M., Carlson, S. M., Caruso, C. M., Clegg, S. M., ... MacColl, A. D. C. (2017). Precipitation drives global variation in natural selection. *Science*, *355*, 959–962.
- Staver, A. C., Archibald, S., & Levin, S. A. (2011). The global extent and determinants of savanna and forest as alternative biome states. *Science*, *334*, 230–232.
- ter Steege, H., Pitman, N. C. A., Phillips, O. L., Chave, J., Sabatier, D., Duque, A., ... Vásquez, R. (2006). Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, *443*, 444–447.
- Van Bodegom, P. M., Douma, J. C., & Verheijen, L. M. (2014). A fully traits-based approach to modeling global vegetation distribution. *Proceedings of the National Academy of Sciences USA*, *111*, 13733–13738.
- Vanzolini, P. E. (1963). Problemas faunísticos do cerrado. *Simpósio Sobre o Cerrado*, *307*, 1.
- Veloso, H. P., Rangel Filho, A. L. R., & Lima, J. C. A. (1991). *Classificação da vegetação Brasileira, adaptada a um sistema universal*. Rio de Janeiro, Brazil: Ministério da Economia, Fazenda e Planejamento, Fundação Instituto Brasileiro de Geografia e Estatística, Diretoria de Geociências, Departamento de Recursos Naturais e Estudos Ambientais.
- Violle, C., Borgy, B., & Choler, P. (2015). Trait databases: Misuses and precautions. *Journal of Vegetation Science*, *26*, 826–827.
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences USA*, *111*, 13690–13696.
- Walter, H. (1973). *Vegetation of the earth in relation to climate and the eco-physiological conditions*. New York: Springer.
- 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 paleodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, *20*, 272–288.
- Whittaker, R. H. (1975). *Community and ecosystems*. New York: McMillan.
- Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., James, E., Watson, M., ... Willis, K. J. (2005). Conservation biogeography: Assessment and prospect. *Diversity and Distributions*, *11*, 3–23.
- Woodward, F. I., Lomas, M. R., & Kelly, C. K. (2004). Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*, 1465–1476.
- Zuloaga, F. O., & Belgrano, M. J. (2015). The catalogue of vascular plants of the southern cone and the flora of Argentina: Their contribution to the world flora. *Rodriguésia*, *66*, 989–1024.

**BIOSKETCH**

**PEDRO L. SILVA DE MIRANDA** is a PhD student at the University of Edinburgh under the supervision of **KYLE G. DEXTER**, **R. TOBY PENNINGTON**, Caroline Lehmann and **ARY OLIVEIRA-FILHO**. His PhD focuses on identifying the main biomes in lowland tropical South America, understanding the main environmental drivers behind them (climatic and edaphic) and assessing the effects climate change will have on their distribution and diversity. The research group, PLant Evolutionary Ecologists and BiogeographerS (PLEEBS; <http://phylodiversity.net/kdexter/HOME.html>), is led by **KYLE G. DEXTER** and focuses primarily on tropical vegetation.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Silva de Miranda PL, Oliveira-Filho A, Pennington RT, Neves DM, Baker TR, Dexter KG. Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. *Global Ecol Biogeogr.* 2018;27:899–912. <https://doi.org/10.1111/geb.12749>