This is a repository copy of *Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/128778/

Version: Accepted Version

Article:

Silva de Miranda, PL, Oliveira-Filho, AT, Pennington, RT et al. (3 more authors) (2018) Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. Global Ecology and Biogeography, 27 (8). pp. 899-912. ISSN 1466-822X

https://doi.org/10.1111/geb.12749

© 2018 John Wiley & Sons Ltd. This is the peer reviewed version of the following 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, which has been published in final form at https://doi.org/10.1111/geb.12749. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

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

Pedro Luiz Silva de Miranda[1]*, Ary Oliveira-Filho[2], R. Toby Pennington[3,4], Danilo M. Neves[5], Timothy R. Baker[6], Kyle G. Dexter[1,3]

Short-running tittle: South American Biomes

[1]School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3FF, UK

[2]Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte, Minas Gerais, 31270–901, Brazil

[3]Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, EH3 5LR, UK

[4]Department of Geography, University of Exeter, Rennes Drive, Exeter, EX4 4RJ, UK

[5]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson 85721, USA

[6]School of Geography, University of Leeds, Leeds, LS2 9JT, UK

*Correspondence: P.L.Silva-De-Miranda@sms.ed.ac.uk

Phone: +44 7871462345

Address: Room 401, Crew Building, King's Buildings, Edinburgh EH9 3FF

Key-words: Amazon Forest, Atlantic Forest, Cerrado, Chaco, Cluster Analysis, NeoTropTree, Savanna, Seasonally Dry Tropical Forest,

Number of words in abstract: 287

Number of words in main text: 5620

Number of references: 84

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 biome. 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 biome.

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 may represent different biomes as they are distinct in both climate and species composition. Our results show an important environmental overlap amongst biomes, with error rates to classify sites into biomes of 19-21% and 16-18% when 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, likely due 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.

Key-words: Cluster Analysis, Atlantic Forest, Amazon Forest, Chaco, Savanna, Cerrado, Seasonally Dry Tropical Forest, NeoTropTree.

Introduction

The biome concept has existed for over a century with the overarching purpose of delimiting recognisable, ecologically meaningful vegetation units. Humboldt (1816) used the term phytophysigonomy 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 et al., 1992; Prentice, 1990). More recently, biomes have been used to categorise 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 employ here, is that of a widespread vegetation formation with distinct ecosystem function.

The term ‘biome’ itself was first employed 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 latter 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 biome by simply knowing the MAT and MAP (e.g., as in Qian, Jin, & Ricklefs, 2017; Siepielski et al., 2017). More recently, large-scale remotely sensed data have become available, which has led researchers to map biomes using simple characterisations of vegetation physiognomy or ecosystem function, including average vegetation height, percent 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 indinstinguiable 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 Whitakker, 1975 vs. Friedl et al., 2002 vs. Woodward et al., 2004 vs. 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 actually 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 (Baker et al., 2017;

Sandel et al., 2015; Violle, Borgy, & Choler, 2015). The premise of this paper 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 (a.k.a. ecological niche modelling) of indicator species can be used to map biomes (as in 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). Bioregionalisation schemes that partition space into geographic units based on species composition and environmental data, such as the global ecoregions proposed by Olson & 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 Olson & Dinerstein’s (1998) scheme that Myers et al.(2000) and Mittermeier et al. (1998, 2004) proposed the global biodiversity hotspots, which are biomes or geographic subsets of biomes (i.e. ecoregions), that present 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 bioregionalisation 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 forests, woodlands and savanna vegetation formations (Neves, Dexter, Pennington, Bueno, & Oliveira Filho, 2015; Vanzolini, 1963). The Cerrado Domain is comprised 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). 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 its 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, e.g. hypersaline soils (DRYFLOR, 2017; Prado & Gibbs, 1993). The Pantanal Domain has heterogeneous vegetation including SDTFs, savanna and swamps, while the Pampa Domain is a largely subtropical grassland that has forest patches along river courses and on certain edaphic conditions.

Lowland Tropical South America, due to its size, diversity and non-continuous geographic distribution of biomes and vegetation types, is an ideal system 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 point to the necessity of developing a new approach for biome delimitation that is better linked to biodiversity. Biome schemes centred on species composition may be more useful for comparative biology, conservation, and 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 be able to distinguish wet forests from the dry biomes, but that it will fail to distinguish SDTF from savanna as they 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 Olson et al. (2001) system), and then for Brazil only, against the Domain classification of IBGE (2012). Our use of floristics data may allow for the delimitation and mapping of biomes in a manner directly relevant to managing ecosystems and developing conservation strategies, for example by enabling the modelling of future climate change effects on tropical vegetation (Prieto-Torres & Rojas-Soto, 2016; Prieto-Torres et al., 2016).

Methods

The NeoTropTree dataset

Floristic inventories of tree communities were obtained from the NeoTropTree (NTT) dataset (Oliveira-Filho, 2017), which contains tree species inventories for more than 6,000 geo-referenced sites across South America. Trees are here defined as free-standing woody plants greater than three metres 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, 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, Zuloaga, & Belgrano, 2015) – http://www.darwin.edu.ar/), with additional direct consultation of taxonomists. Our data excludes checklists with < 10 species, because in lowland tropical regions, this is invariably due 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 in Oliveira-Filho (2017; see also 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.neotroptree.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 above 1,000 m elevation or below 36^o S latitude. Montane areas were excluded because biogeographic barriers may 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.

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 β_{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 randomising 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 at least 50% of the clusters are distinguished (Omland, Cook, & Crisp, 2008). This analysis was performed in R (R Team, 2016) 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 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 if there is floristic integrity to these previously proposed biomes, and we found clear evidence that there was, i.e. higher-level groups were comprised largely of one broad biome type (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 was categorised to biome.

We assessed which sites may be intermediate or transitional between our biomes using a silhouette analysis, via the R package cluster (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2016). 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, 2002) of sites in two dimensions based on the Simpson distance amongst sites.

Using climate and edaphic data to distinguish biomes

To assess if 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, & Jarvis, 2005), which quantify various aspects of temperature and precipitation regimes, as well as 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 (cmol/kg) and 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 cm, 5 cm, 15 cm and 30 cm, which were then averaged. Two different classifications were performed, one considering climatic data alone and another considering both climatic and soil data.

In order to assess the success rate of the classification tree approach in assigning sites to biome 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).

Comparison to existing biome maps

We compared how two commonly used vegetation maps for South America classify sites to biome compared to 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 was 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).

Results

Biomes of Lowland Tropical South America

Hierarchical cluster analysis produced five higher-level groups (Fig. 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 comprised of sites along the Atlantic coast, which we refer to as the Atlantic Forest biome (Fig. 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 geographic areas of the forest Domains (Fig. 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 (Fig. 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 Seasonally Dry Tropical Forest (SDTF), based on the original vegetation classifications of sites (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 Caatinga Domain of northeastern Brazil (Fig. 2). It is spatially interdigitated with the Savanna biome. The last group, which we distinguish as a separate biome is the Chaco, comprising woodlands in Bolivia, Argentina and Paraguay and extending to the borders of southern Brazil. While 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 degrees latitude that are unlikely to experience freezing and can be considered tropical (Fig. 2). See Supplementary Materials (Appendix 1) for further description of the biomes. Our continuous biome map, developed using the Thiessen's polygons method, shows the LTSA biomes' overall spatial distribution 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 (Fig 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 (Fig. 4a, Table S2). An ordination of sites (NMDS with two axes, stress value= 0.1816) also suggests that these sites are compositionally transitional (Fig. 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), while they were infrequent between other biomes, including between any pair of dry biomes. Floristically transitional sites are common in the Dry Diagonal (Fig. 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.

Using climate and edaphic data to distinguish biomes

We find that biomes overlap substantially in climatic space, both in terms of water availability (Fig. 5) and temperature (Fig. 6). For example, all five biomes defined here occupy at least two of the climatic biomes proposed by Whittaker (1975) (Fig. 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% (Table S3). The most common misclassifications involved Amazon or Atlantic Forest sites being classified as belonging to the Savanna biome or vice versa, while 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 geographic space, were removed (Table S4, error rate: 20.3%).

The inclusion of edaphic variables slightly increased overall classification success by 3.2% (Table 2), and 3% when transitional sites were included (Table S5). There were a total of 124 sites that switched from being classified incorrectly (with just climatic data) 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 main most important variables for classification were Mean Annual Precipitation, 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).

Comparison to existing biome maps

The classification systems developed by Olson and Dinerstein et al. (2001, 2017) and IBGE (2012) assigned 74-75% of the NTT sites to the same biomes as they were placed according to our analyses (74.7% Dinerstein et al., 2017, Table S6; 74.5% IBGE, 2012, 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 (Figure S2). In IBGE's system, the error rate stems from SDTF sites being classified as Cerrado and vice-versa (Figure S3).

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 due to pronounced climatic overlap of biomes (Figs 5, 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. Recently, researchers have begun assigning study sites to biomes, generally those of Whittaker (1975) based solely on climatic values, e.g. mean annual precipitation and temperature (e.g. Díaz et al., 2016; Qian & Ricklefs, 2017; Siepielsky et al*.,* 2017). Our results suggest this is potentially problematic (Fig. 6). For example, the Amazon and Atlantic Forests can both occur in areas that are more seasonal than 'tropical rain forest' (sensu Whittaker), while 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 are restricted to a single biome in Whittaker's climatic biome classification (Fig. 6).

We were able to employ 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). While floristic approaches to mapping biomes are unlikely to succeed inter-continentally 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, www.forestplots.net, www.neotroptree.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 due to high edaphic and floristic heterogeneity coupled with incomplete sampling. We have generated a continuous map (Fig. 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 Pennington et al. (2006), Werneck (2011), Collevatti et al. (2013).

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 since 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 (Fig. 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) that have argued 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 supplementary material (Appendix 1).

Using climate and edaphic data to distinguish biomes

Mean annual precipitation (MAP), 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). 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, & Chave, 2014). Nevertheless, it is notable that measures of temperature variability, particularly across seasons, were more important than other temperature measures, including mean annual temperature (MAT) and minimum 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).

While a classification success rate of 80% seems high, this would result in 1 in 5 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 may 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 as they comprise 93.4% of our sites. In order to improve classification of these sites to biome based on environmental data, environmental data in better resolution are needed. Publicly available environmental data are derived from interpolation. For climate, which varies at a relatively broad spatial grain, this may not be problematic. However, edaphic data vary at a small spatial grain, and interpolation-based methods may be inadequate to capture edaphic conditions at many sites. Also, the edaphic data from SoilGrids does 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, may 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).

Comparing to 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 ~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 IBGE's system) as a distinct biome or domain, while the IBGE Domain system also delimits the Pampas (a.k.a. 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 at South Brazil are floristically similar, in relation to tree species composition, to the rest of the subtropical portion of the Atlantic Forest biome (Oliveira-Filho, Budke, Jarenkow, Eisenlohr, & Neves, 2015).

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 prioritisation. The biomes we have delimited differ in tree species composition and therefore likely differ in ecosystem function. Ecosystem management plans should therefore be developed separately for each. Similarly, these biomes have almost no species in common, yet have many species unique to them. Our schematic map (Fig. 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 have been recognised 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 (though see DRYFLOR 2016 for first steps). Current conservation planning for SDTF in Brazil focuses solely on the Caatinga Domain, but many Brazilian SDTFs are found in disjunct patches outside of this area, especially in the Cerrado, placing them under laws designed to protect savanna diversity. As another example, the Chaco is under great threat due to an increase of habitat destruction and fragmentation during the last 30 years (Hansen et al., 2013, Nori et al. 2016), but if recognised as a separate biome, as our analyses suggest, the urgency of its conservation may be better recognised (Kuemmerle et al., 2017).

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, etc.) must be measured and accounted for in models. The interdigitiation of biomes, especially in the dry diagonal across Brazil, is not recognised in the current IBGE (2012) system on which Brazilian conservation legislation is based, leading to the neglect of highly threaten SDTF vegetation outside of the Caatinga Domain. Our analyses also show 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.

Acknowledgements

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.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). 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/.

Biosketch

Pedro L. Silva de Miranda is a PhD student at the University of Edinburgh under the supervision of Dr. Kyle G. Dexter, Dr. Toby Pennington, Dr. Caroline Lehmann and Professor 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 Dr. Kyle Dexter and focuses primarily on tropical vegetation.

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*, *110*, 6442–6447.

Augusto, L., Davies, T. J., Delzon, S., De Schrijver, A., & Chave, J. (2014). The enigma of the rise of angiosperms: can we untie the knot?. *Ecology Letters, 17,* 1326-1338.

Baker, T. R., Pennington, R. T., Dexter, K. G., Fine, P. V. A., Fortune-Hopkins, H., Honorio, E. N., … Vasquez, R. (2017). Maximising Synergy among Tropical Plant Systematists, Ecologists, and Evolutionary Biologists. *Trends in Ecology & Evolution*, *32*, 258–267.

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. In *Ecological Society of America* (Vol. 5, pp. 119–128). New York.

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*, *04*, 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. (2015). Package 'recluster' - Ordination methods for the analysis of beta-diversity indices. Retrieved from: https://cran.r-project.org/web/packages/recluster/recluster.pdf

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.

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*(9): e0161323.

Dexter, K. G., Smart, B., Baldauf, C., Baker, T. R., ..., 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). Response - Forest conservation: Remember Gran Chaco. *Science*, *355*, 465–466.

Engemann, K., Sandel, B., Enquist, B. J., JØrgensen, P. Mø., 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.

ESRI. (2017). ArcGIS 10.4.1. Redlands: CA: Environmental Systems Research Institute.

Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X. Y., Muchoney, D., & Strahler. (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., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2): 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*, 267–268.

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 (Ed.). (2012). *Manual técnico da vegetação brasileira* (2a edição revista e ampliada). Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística-IBGE.

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 conservation: Humans' handprints. *Science*, *355*, 465.

Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., Šímová, I., … Enquist, B. J. (2014). Functional trait space and the latitudinal diversity gradient. *Procedings of the National Academy of Sciences*, *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 randomForest. *R News*, *2*, 18–22.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2016). cluster: Cluster Analysis Basics and Extensions.

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*, *106*, 20610–20615.

McCune, Bruce, James B. Grace, and Dean L. Urban. *Analysis of ecological communities*. Vol. 28. Gleneden Beach, OR: MjM software design, 2002.

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. & Di Minin, E. (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.neotroptree.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 *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 Jounal 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 EarthA 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., Heskel, 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., 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., & Rojas-Soto, O. R. (2016). Reconstructing the Mexican Tropical Dry Forests via an Autoecological Niche Approach: Reconsidering the Ecosystem Boundaries. *PLOS ONE*, *11*.

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.

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.

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., Schrodt, 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 Adam. *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.

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

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 of the United States of America*, *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: 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*, *111*, 13690–13696.

Walter, H. (1973). *Vegetation of the earth in relation to climate and the eco-physiological conditions*. New York: Springer.

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 of London. Series B, Biological Sciences*, *359*, 1465–76.

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.

Table 1: Confusion matrix between sites categorised based on floristic composition via hierarchical clustering (rows) and sites categorised using climate and a classification tree approach (columns). The diagonal gives the number of sites that are correctly classified by climate, while the off-diagonal elements give mis-classifications (18.6%). Only non-floristically transitional sites were considered. Accuracy: 81%; Average precision: 81%; Average recall: 80%.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF
Amazon Forest	989	6	45	0	0
Atlantic Forest	3	1290	199	5	50
Cerrado	58	167	357	0	50
Chaco	0	7	0	76	1
SDTF	0	51	65	1	408

Table 2: Confusion matrix between sites categorised based on floristic composition via hierarchical clustering (rows) and sites categorised using climate + soil and a classification tree approach (columns). The diagonal gives the number of sites that are correctly classified by climate, while the off-diagonal elements give mis-classifications (15.2%). Accuracy: 84%; Average precision: 84%; Average recall: 83%.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF
Amazon Forest	1001	4	37	0	0
Atlantic Forest	4	1331	161	4	49
Cerrado	48	121	423	0	40
Chaco	0	7	0	76	1
SDTF	0	55	52	1	417

Table 3: The mean variable importance value (± one standard error) for all climatic variables included in the Random Forest analysis across 100 runs of the Bremnans' algorithm utilizing rarefactions of the main dataset (116 sites per biome**)**.

Environmental Variables	**Climate**	**Climate + Soil**
	Mean ± SE	**Mean ± SE**
Mean Annual Precipitation (mm)	356.81 ± 1.09	318.8 ± 1.18
Temperature Seasonality (C^o)	319.73 ± 1.23	287.14 ± 1.13
Maximum Climatological Water Deficit (mm/yr)	273.2 ± 0.69	232.07 ± 0.71
Isothermality (%)	233.29 ± 0.98	211.53 ± 0.87
pH (KCl)	*	188.98 ± 0.84
Mean Temperature of Coldest Quarter (C^o)	187.06 ± 0.95	163.07 ± 0.97
Precipitation of Wettest Quarter (mm)	155.06 ± 0.56	120.57 ± 0.48
Cation Exchange Capacity (cmol/Kg)	*	119.89 ± 0.23
Precipitation of Driest Quarter (mm)	148.46 ± 0.53	119.37 ± 0.51
Precipitation of Driest Month (mm)	133.16 ± 0.49	109.94 ± 0.44
Mean Annual Temperature (C^o)	122.75 ± 0.71	96.15 ± 0.66
Precipitation of Wettest Month (mm)	119.83 ± 0.42	91 ± 0.35
Mean Temperature of Driest Quarter (C^o)	106.46 ± 0.57	81.93 ± 0.49
Amount of Sand (%)	*	81.73 ± 0.17
Maximum Temperature of Warmest Month (C^o)	103.8 ± 0.33	81.69 ± 0.31
Amount of Silt (%)	*	76.89 ± 0.13
Temperature Annual Range (C^o)	101.51 ± 0.32	75.32 ± 0.23
Precipitation Seasonality (%)	99.22 ± 0.24	74.3 ± 0.31
Minimum Temperature of Coldest Month (C^o)	99.21 ± 0.23	73.38 ± 0.37
Precipitation of Warmest Quarter (mm)	98.61 ± 0.3	70.77 ± 0.18
Precipitation of Coldest Quarter (mm)	97.11 ± 0.47	69.21 ± 0.25
Temperature's Diurnal Range (C^o)	91.45 ± 0.19	68.67 ± 0.16
Amount of Clay (%)	*	65.97 ± 0.13
Mean Temperature of Wettest Quarter (C^o)	79.01 ± 0.22	61.57 ± 0.24
Mean Temperature of Warmest Quarter (C^o)	60.71 ± 0.12	46.52 ± 0.16

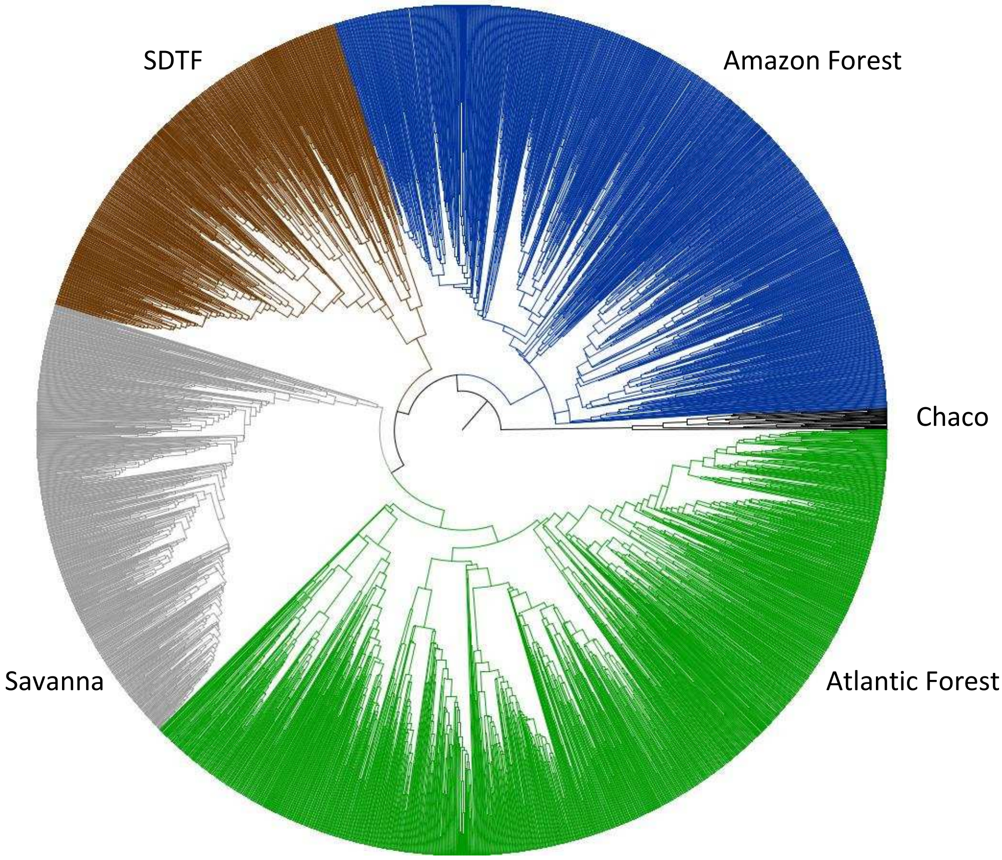

Figure 1: Hierarchical cluster of 4,103 sites in lowland (<1,000 m.a.s.l.) tropical South America and neighbouring subtropical areas based on tree species composition. Five principal higher-level groups can be observed, which were refer to as the Amazon Forest (blue), Atlantic Forest (green), Savanna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

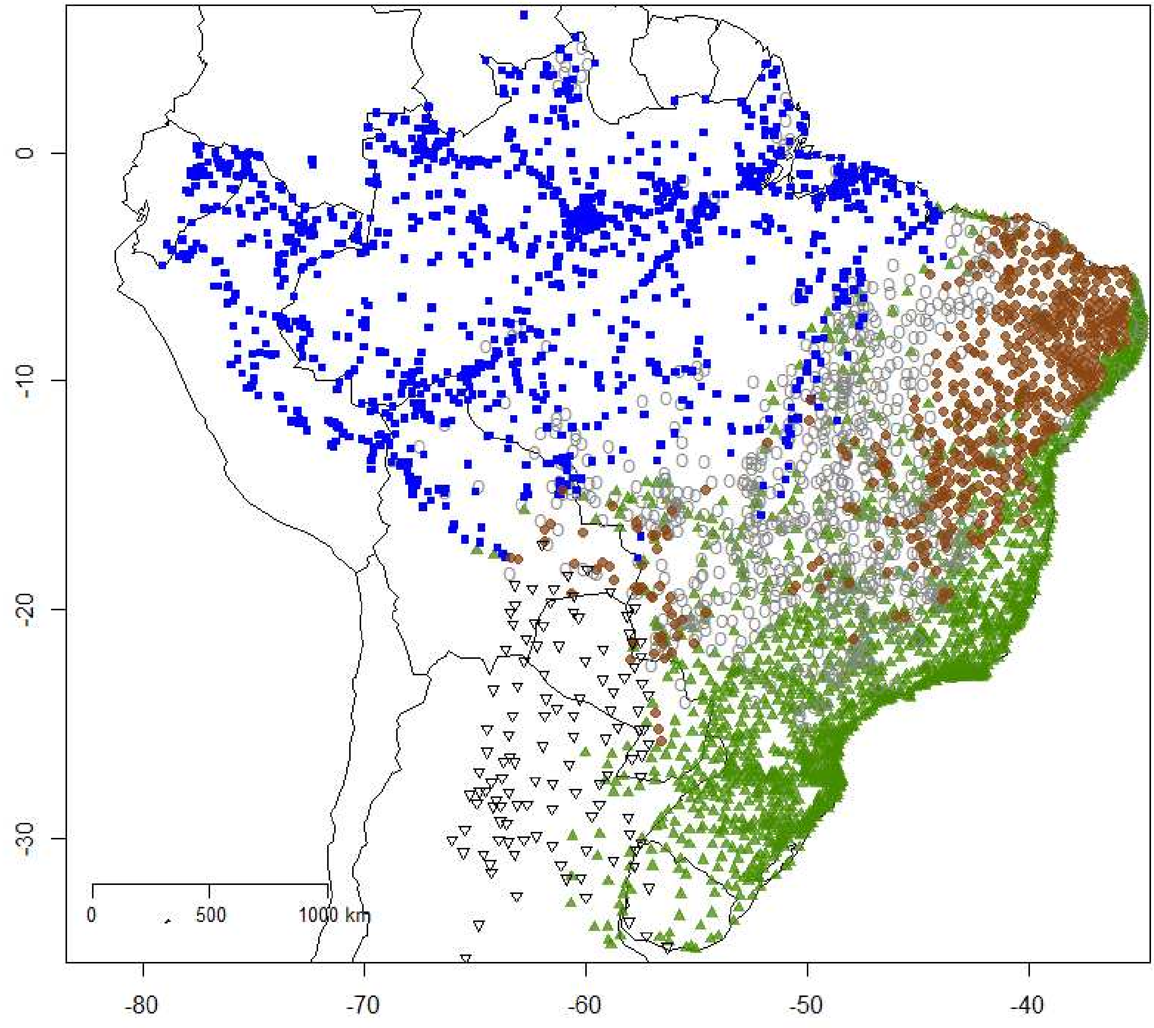

Figure 2 - Map of Lowland Tropical South America with sites classified into biomes based on hierarchical cluster analysis of tree species composition: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow grey circles), Amazon Forest (blue squares), Chaco (inverted hollow 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.

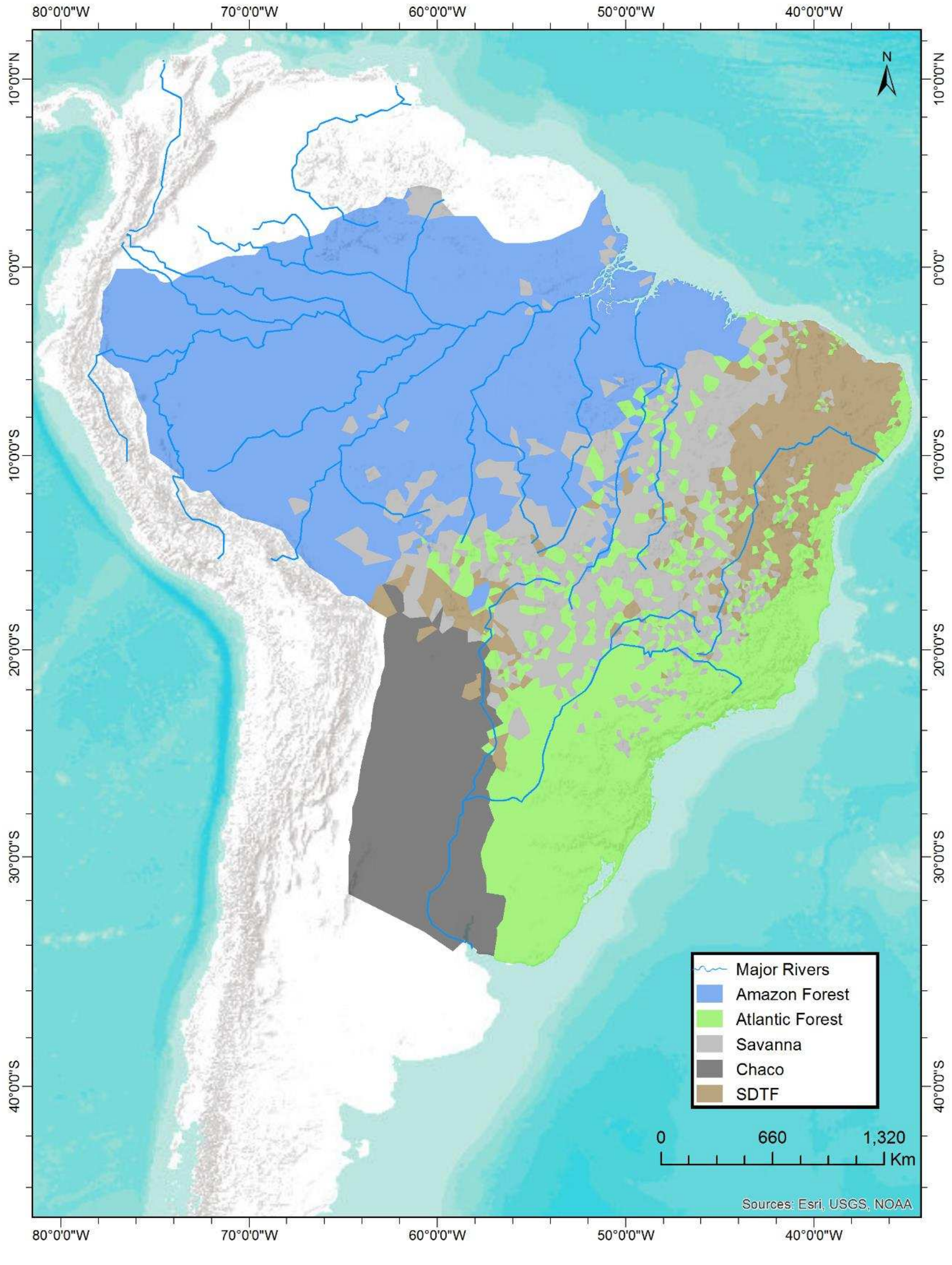

Figure 3 – Map of South America with a schematic representation of the biomes delimited via hierarchical cluster analysis in the present contribution (Amazon Forest, Atlantic Forest, Savanna, Chaco and Seasonally Dry Tropical Forests – SDTF). The map was created by applying the Thiessen polygons method on the categorised points presented in figure 2. See text for further details.

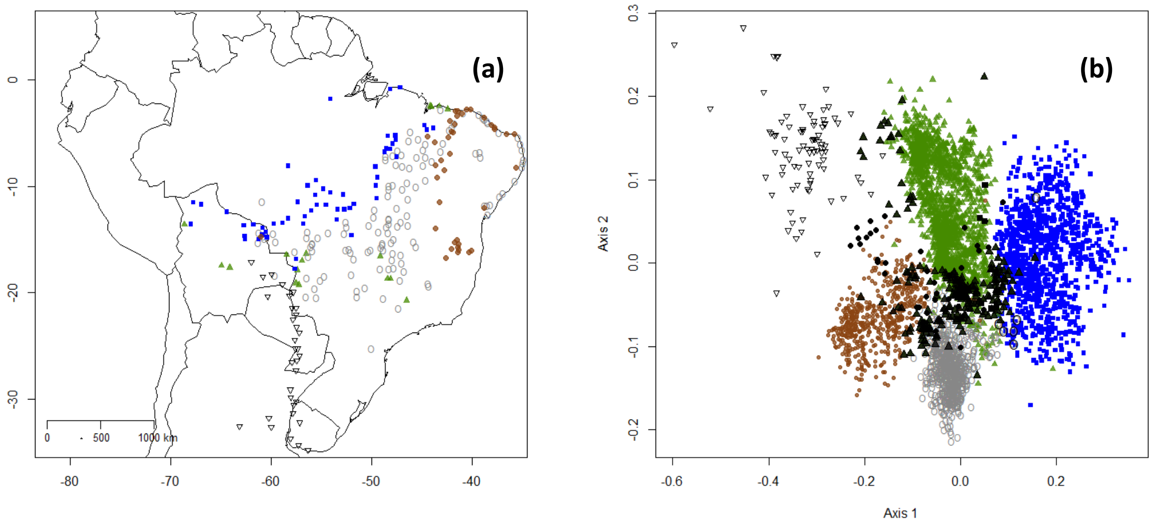

Figure 4: NeoTropTree sites which have a transitional/ambiguous floristic identity, as revealed by the silhouette analysis, and how they are distributed in geographic (a) and species compositional (b) spaces. In (a), sites are categorised according the biome to which they are floristically more similar. 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 correspond to: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow grey circles), Amazon Forest (blue squares), Chaco (inverted hollow black triangles).

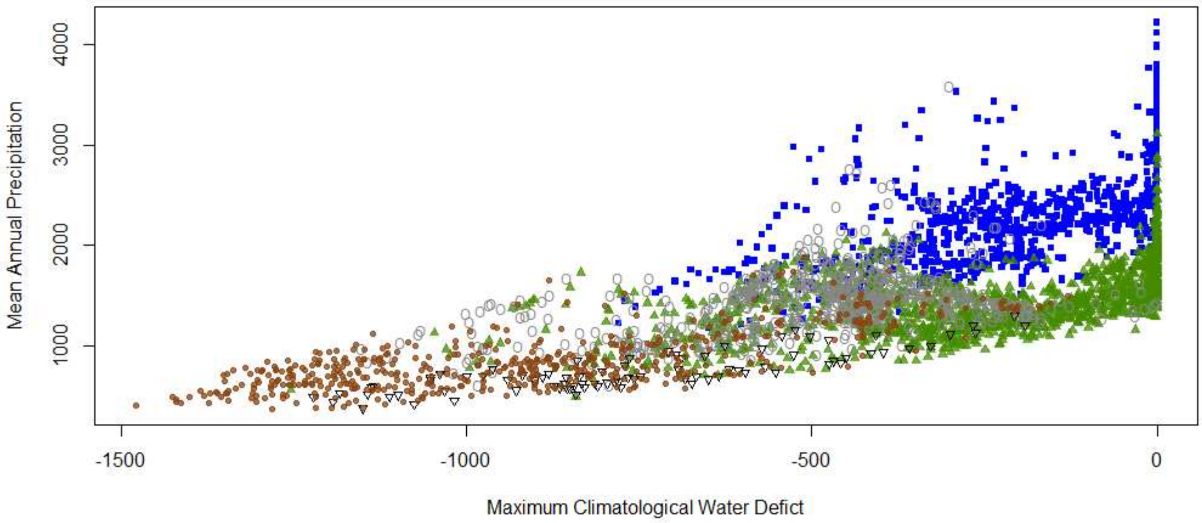

Figure 5: Distribution of sites with respect to precipitation regime. Mean annual precipitation values come from worldclim (Hijmans et al. 2005) and maximum climatological water deficit comes from Chave et al. (2014). Symbols correspond to: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow grey circles), Amazon Forest (blue squares), Chaco (inverted hollow black triangles). Modelled after Fig. 1 in Malhi *et al.* (2009), which suggested that savannas were drier than seasonal forests, contrary to the pattern here.

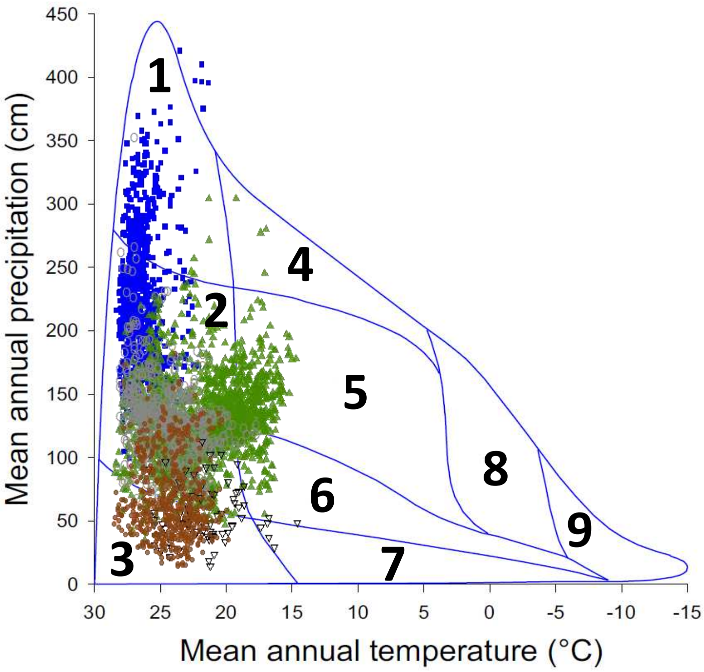

Figure 6: Distribution of sites in climatic space across the nine biomes proposed by Wittaker (1975) considering mean annual precipitation (cm) and mean annual temperature (C^o). Numbers correspond to: Tropical rain forest (1), Tropical seasonal forest/savanna (2), Tropical and subtropical desert (3), Temperate rainforest (4), Temperate deciduous forest (5), Woodland/scrubland (6), Temperate grassland/dessert (7), Boreal forest (8), and Tundra (9). While symbols and colors correspond to: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow gray circles), Amazon Forest (blue squares), and Chaco (inverted hollow black triangles).

Appendix to:

Silva de Miranda, P., L., Oliveira-Filho, A., T., Pennington, R., T., Neves, D., M., Baker, T., R., Dexter, K., G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America

Appendix 1. Main Biomes of Lowland Tropical South America – Brief descriptions

Wet Forest Biomes (Amazon and Atlantic Forests)

All rain forests, moist forests, evergreen forests and most semideciduous forests fell within two overarching groups in the cluster analysis, which we termed the Atlantic and Amazon Forest biomes. While we have argued that a floristics approach can be used to delimit biomes at continental scales where biogeographic factors are not the main driver of turnover in species composition, it may be that the floristic differentiation between the Atlantic and Amazon Forests is due in part to their biogeographic isolation by the Dry Diagonal. However, to definitely determine whether these forests represent distinct biomes, further comparative research is needed to determine how they compare in terms of ecosystem function.

The Atlantic Forest biome can be further divided into three different floristic groups, a northern group, completely tropical, encompassing all the coastal Atlantic forests ranging from northeast Brazil south to the state of Rio de Janeiro; a second group, largely sub-tropical, beginning at Sao Paulo’s coast and harbouring all of the forests covering the South of Brazil, Uruguay, Southeast Paraguay and portions of Northeast Argentina, especially the Missiones region; and a last group, also tropical, formed by semideciduous forests further inland, scattered mostly across Brazil, but also present as far west as Bolivia. This distribution matches the *sensu-latissimo* definition proposed by (Oliveira-Filho *et al.*, 2006) with the additional inclusion of forest patches amongst the subtropical grasslands in the south of Brazil, Southern Paraguay, most of Uruguay and Northeast Argentina. This region has been distinguished from the Atlantic Forest in the past based on its overall physiognomy of forest patches in a grassy landscape, which contrasts with contiguous forest. However, these forest patches clearly show strong floristic continuity with the Atlantic Forest, as was also observed by Oliveira-Filho *et al.* (2013), and likely have similar ecosystem function to the now heavily fragmented Atlantic Forest.

The Amazon Forest biome does not show as clear subdivisions as the Atlantic Forest biome. However, there is evident floristic differentiation between “terra firme” and seasonally flooded forests, and these two subgroups can be further divided between sites in the western Amazon (from Peru, Bolivia, Ecuador and the Brazilian state of Acre) and the eastern Amazon (encompassing most of the Brazilian portion of the Amazon Forest, including the states of Amazonas, Pará, Mato Grosso, Maranhão and Roraima). These divisions between eastern and western Amazon and between “terra firme” and seasonally flooded forests have been reported before in the literature (e.g. Prance 1982; ter Steege et al., 2006).

The gallery forests within the Cerrado Domain do not cluster with the prevailing Savanna biome in that Domain, nor do they form their own unique cluster. Instead, they are floristically most similar to the most geographically proximal wet forest biome, either the Atlantic or the Amazon Forest. Similarly, sites found in sandy coastal areas of Brazil, often termed “restingas” or “matas de maré”, do not comprise a single group in our hierarchical cluster, but cluster with the closest wet forest biome (Atlantic or Amazon Forest).

Dry Biomes (Savanna, Seasonally Dry Tropical Forest and Chaco)

Our analyses confirm that the savannas distributed across LTSA form a single floristic unit. There are no clear subdivisions within this Biome. Savanna is a disturbance driven system, which may allow for the ready establishment of dispersing propagules of dominant tree species and a homogenisation of the tree flora over large spatial scales. Indeed, savannas in SA have been shown to possess a consistent set of dominant oligarchic tree species (Bridgewater *et al.*, 2004), which may be why clear subgroups are not evident. In addition, the high disturbance in the system may prevent tree communities from reaching an equilibrium or ‘climax’ in species composition, which may inhibit sites from converging on similar species composition in similar environments, which could in turn inhibit the formation of clear floristic groups.

Our analyses suggest that the SDTF scattered across lowland tropical South America should be regarded as one single biome, as has been suggested by previous studies (DRYFLOR, 2016; Pennington et al., 2000, 2009; Prado & Gibbs, 1993). As found by Neves et al. (2015) and DRYFLOR (2016), our results suggest two main groups across the Dry Diagonal, one comprising the various forests of the Caatinga Domain and the other comprising SDTF patches scattered throughout the Cerrado Domain and into regions of the Pantanal and Chiquitania. The Misiones floristic group here shows greater floristic affinity with the Atlantic Forest than it does with other SDTF. The Misiones forests receive more rainfall than other STDF (Neves et al. 2015) and are semi-deciduous in nature (DRYFLOR 2016). Meanwhile, the Piedmont forests are found to be floristically more similar to sites in the Chaco than to other SDTF. This is perhaps not surprising given their proximity to the Chaco and that both environments receive significant frost in the winter season (Neves et al. 2015).

The Chaco is floristically different, in terms of tree species composition, from other sites across LTSA. While this difference has been noted in the past, particularly in comparison with SDTF (Pennington et al., 2000; Prado & Gibbs, 1993; Spichiger et al., 2004), it has often been attributed to the Chaco experiencing heavy frost. While many of the sites in our Chaco biome do experience frost, a large number of sites in eastern Bolivia, western Paraguay and south central Brazil (Mato Grosso do Sul state) do not experience frost, and could be considered tropical in nature. We refer to these northern Chaco sites as the 'tropical Chaco'. It is floristically distinct from other SDTF and may have different ecosystem function, but further research is needed to compare ecosystem function in SDTF versus tropical and subtropical Chaco sites.

Chiquitania and Pantanal

Two regions that have always been a challenge to place in floristic or biome classification schemes are the Chiquitania and Pantanal regions of eastern Bolivia and southwestern Brazil. The Chiquitania region is the site of contact between savannas (composed mostly of the savanna wetlands from the Pantanal region and the Llanos de Moxos region in Bolivia), Amazon Forest, SDTF and the Chaco (Killeen et al., 2006; Pennington et al., 2009). This region is composed of a mosaic of SDTF mixed with savannas, overlying diverse old geological formations (Navarro, 2011), and its northern portion grades into the Amazon Forest. Chiquitania is notable for its lack of endemic plant species, which is attributed to its recent geological past and to its transitional nature (Killeen et al., 2006). Our analyses show that sites within the Chiquitania's geographic range (Navarro, 2011) alternatively cluster together with the SDTF, Savanna and Amazon Forest biomes, and that perhaps the region should not be considered as a distinct vegetation entity on its own.

The floristic identity of forests and woodlands in the Pantanal also do not stand out as distinct within a continental context, although such was proposed by Veloso et al. (1991) and Navarro (2011). The Brazilian government also classifies it as a unique Domain (IBGE, 2012). However, just like Chiquitania, the Pantanal is composed of sites that belong to the Savanna and SDTF biomes as well as a wet forest biome, but in this case the Atlantic Forest biome. The lack of endemic species in this region is also evidence of its lack of floristic distinctness and recent geological history (Pott *et al.*, 2011)

References:

Bridgewater, S., Ratter, J. A., & Ribeiro, J. F. (2004). Biogeographic patterns , b-diversity and dominance in the cerrado biome of Brazil. *Biodiversity and Conservation*, *13*, 2295–2318.

DRYFLOR. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. *Science*, *353*, 1383–1387.

IBGE. (2012). *Manual Técnico da Vegetação Brasileira* (segunda ed). Rio de Janeiro: Instituto de Brasileiro de Geografia e Estatística - IBGE.

Killeen, T., Chavez, E., Peña-Claros, M., Toledo, M., Arroyo, L., Caballero, J., … Steininger, M. (2006). The Chiquitano Dry Forest, the Transition between Humid and Dry Forest in Eastern Lowland Bolivia. In R. . T. Pennington, G. P. . Lewis, & J. A. Ratter (Eds.), *Neotropical Savannas and Seasonally Dry Forests Plant Diversity, Biogeography, and Conservation* (pp. 213–233). CRC Press.

Navarro, G. (2011). *Classificación de la Vegetación de Bolivia*. Santa Cruz de la Sierra: Centro de Ecología Difusión Simón I. Patiño.

Neves, D. M., Dexter, K. G., Pennington, R. T., Bueno, M. L., & Oliveira-Filho, A. T. (2015). *Environmental and historical controls on floristic composition across the South American Dry Diagonal*.

Oliveira-Filho, A. T., Budke, J. C., Jarenkow, J. A., Eisenlohr, P. V., & Neves, D. R. M. (2013). 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., 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 *Neotropical Savannas and Seasonally Dry Forests: Plant Diversity, Biogeography, and Conservation* (pp. 159–192).

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.

Pott, a, Oliveira, a K. M., Damasceno-Junior, G. a, & Silva, J. S. V. (2011). Plant diversity of the Pantanal wetland. *Brazilian Journal of Biology = Revista Brasleira de Biologia*, *71*(1 Suppl 1), 265–273.

Prado, D. E., & Gibbs, P. E. (1993a). Patterns of Species Distributions in the Dry Seasonal Forests of South America. *Annals of the Missouri Botanical Garden*, *80*, 902–927.

Prado, D. E., & Gibbs, P. E. (1993b). 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). *Biological Diversity in the Tropics*. (N. York, Ed.). Columbia University Press.

Spichiger, R., Calenge, C., & Bise, B. (2004). Geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin. *Journal of Biogeography*, *31*, 1489–1501.

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.

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: Fundação Instituto Brasileiro de Geografia e Estatística, IBGE.

Appendix to:

Silva de Miranda, P., L., Oliveira-Filho, A., T., Pennington, R., T., Neves, D., M., Baker, T., R., Dexter, K., G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America

Appendix 2. Indicator Species Analysis

Methods

In order to determine the biome affiliation of species, we used a modified version of the phi coefficient of Tichy & Chytrý (2006) that leverages presence/absence data and accounts for variation in sampling amongst groups. Specifically, we used the the r^g correlation index of De Cáceres & Legendre (2009). This varies from -1 to 1, with positive values indicating a non-random association of a species with a group, or biome in this case, and negative values indicating a non-random anti-association. To test if the associations between a given species and biomes were significant, we randomized occurrences across sites 1000 times and assessed if a species was found more or less frequently in a biome than expected by chance, using an 0.05 alpha significance threshold, with a multiple significance test (Šidák's test) to avoid Type I error. Species with a significant and positive association with any biome are henceforth referred to as diagnostic species for that biome. Indicator species analyses were conducted using functions in the indicspecies package for the R Statistical Software (De Cáceres & Legendre, 2009).

Results:

In total, 8231 out of 10306 tree species were found to be significantly positively associated with at least one biome ($p < 0.05$ after Šidák's test). The Atlantic Forest has 2492 diagnostic species, the Amazon Forest has 4786 and the other Biomes – Savanna, SDTF and Chaco – have 318, 459 and 177 respectively. Table S8 reports all species and their association values (r^g) for each Biome.

References:

Cáceres, M. De, & Legendre, A. N. D. P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*, 3566–3574.

Tichy, L., & Chytrý, M. (2006). Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science*, *17*, 809–818.

Appendix to:

Silva de Miranda, P., L., Oliveira-Filho, A., T., Pennington, R., T., Neves, D., M., Baker, T., R., Dexter, K., G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America

Appendix 3. Supplementary figures and tables

Table S2: Summary of results for silhouette analysis. The rows correspond to totals under the original classification, derived from the hierarchical clustering analysis, while the columns correspond to totals based on looking at the overall similarity of sites to the multidimensional centroid of each major group in the cluster. The diagonal corresponds to sites where the two approaches agree, while the off-diagonal elements correspond to sites where the two approaches disagree, which we consider to indicate sites that are transitional between the two biomes.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF	Original Classification
Amazon Forest	1042	7	0	0	0	1049
Atlantic Forest	46	1549	115	19	39	1768
Cerrado	7	0	632	0	0	639
Chaco	0	0	0	84	0	84
SDTF	2	10	13	13	525	563
Corrected Classification	1097	1566	760	116	564	4103

Table S3: Confusion matrix between sites categorised based on floristic composition via hierarchical clustering (rows) and sites categorised using climate + soil and a classification tree approach (columns), for all sites including ones identified as transitional via a silhouette analysis. The diagonal gives the number of sites that are correctly classified by climate + soil, while the off-diagonal elements give mis-classifications (20.7%). Accuracy: 79%; average precision: 78%; average **recall** rate: 76%.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF
Amazon Forest	1021	8	66	0	2
Atlantic Forest	7	1281	209	11	58
Cerrado	79	179	439	1	62
Chaco	0	22	0	89	5
SDTF	2	60	75	5	422

Table S4: Confusion matrix between sites categorised based on floristic composition via hierarchical clustering (rows) and sites categorised using climate and a classification tree approach (columns), for all non-geographically overlapping sites (those with centres >10 km apart). The diagonal gives the number of sites that are correctly classified by climate, while the off-diagonal elements give mis-classifications (20.3%). Accuracy: 79%; average precision: 79%; average recall rate: 77%.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF
Amazon Forest	812	6	63	0	1
Atlantic Forest	6	1051	158	7	44
Cerrado	66	137	361	1	56
Chaco	0	14	1	78	4
SDTF	2	42	64	4	353

Table S5: Confusion matrix between sites categorised based on floristic composition via hierarchical clustering (rows) and sites categorised using climate + soil and a classification tree approach (columns), for all sites including ones identified as transitional via a silhouette analysis. The diagonal gives the number of sites that are correctly classified by climate + soil, while the off-diagonal elements give mis-classifications (17.7%). Accuracy: 82%; average precision: 81%; average recall rate: 80%.

	Amazon Forest	Atlantic Forest	Cerrado	Chaco	SDTF
Amazon Forest	1038	6	51	0	2
Atlantic Forest	9	1317	174	10	56
Cerrado	69	136	500	1	54
Chaco	0	22	0	87	7
SDTF	3	57	64	4	436

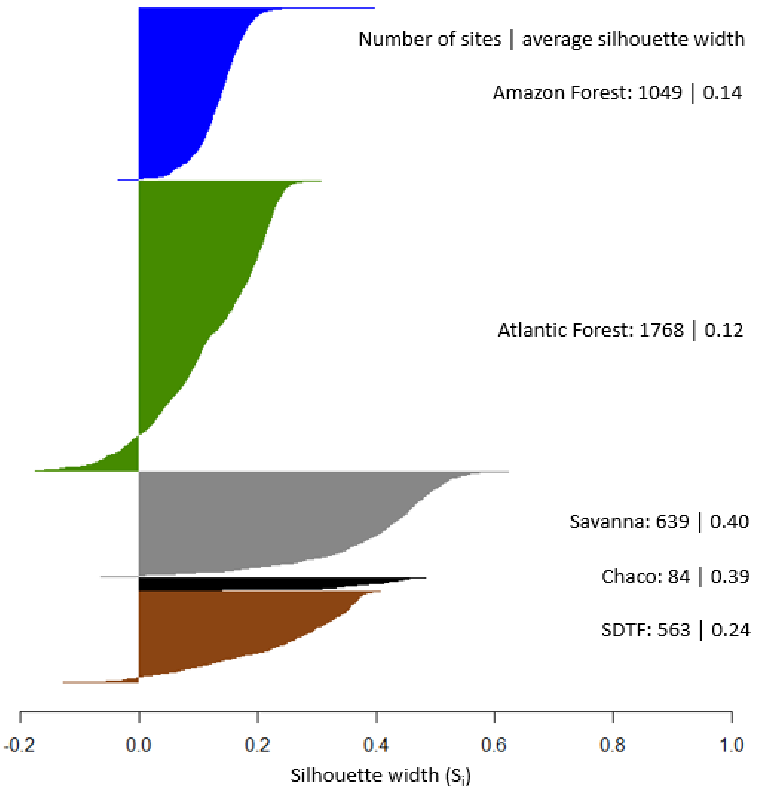

Figure S1 – Silhouette plot with all 4103 NeoTropTree sites included in the cluster analysis. Positive Silhouette width values (S_i) indicate that a site is indeed most similar, in terms of tree species composition, to the other sites in the biome it has been assigned to, whereas negative values indicate that a given site is compositionally more similar to one of the other biomes delimited through the cluster analysis than it is to the biome with which it clustered. The plot also presents the number of sites that compose each biome and their average silhouette width value (S_i).

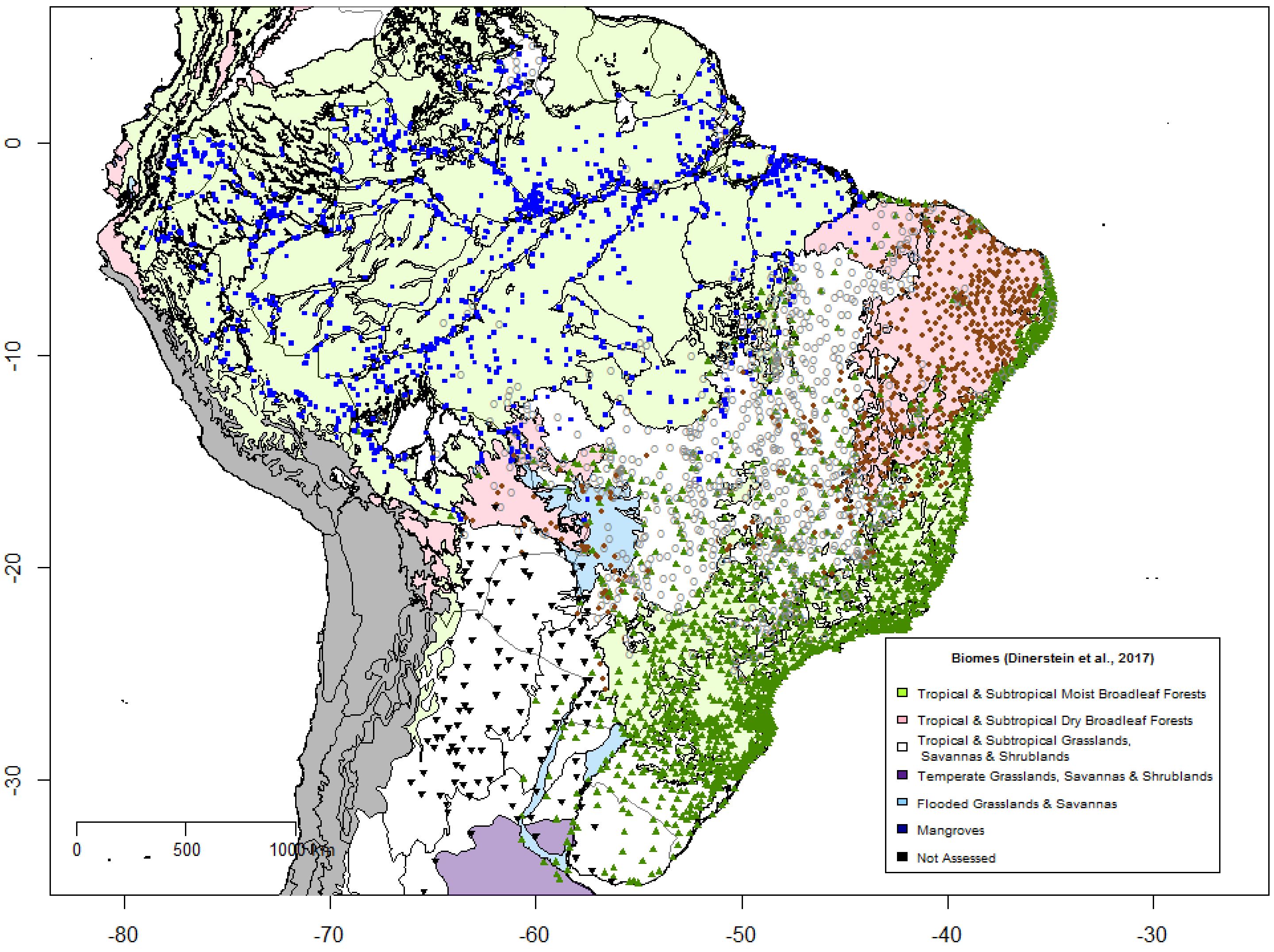

Figure S2 – Map of South America with areas coloured according to Dinerstein *et al.,* (2017), which combines ecoregions into biomes and is a reviewed and updated version of Olson *et. al.* (2001). The points on the map are the NeoTropTree tree species inventory sites classified into biomes by this study: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow gray circles), Amazon Forest (blue squares), and Chaco (inverted hollow black triangles).

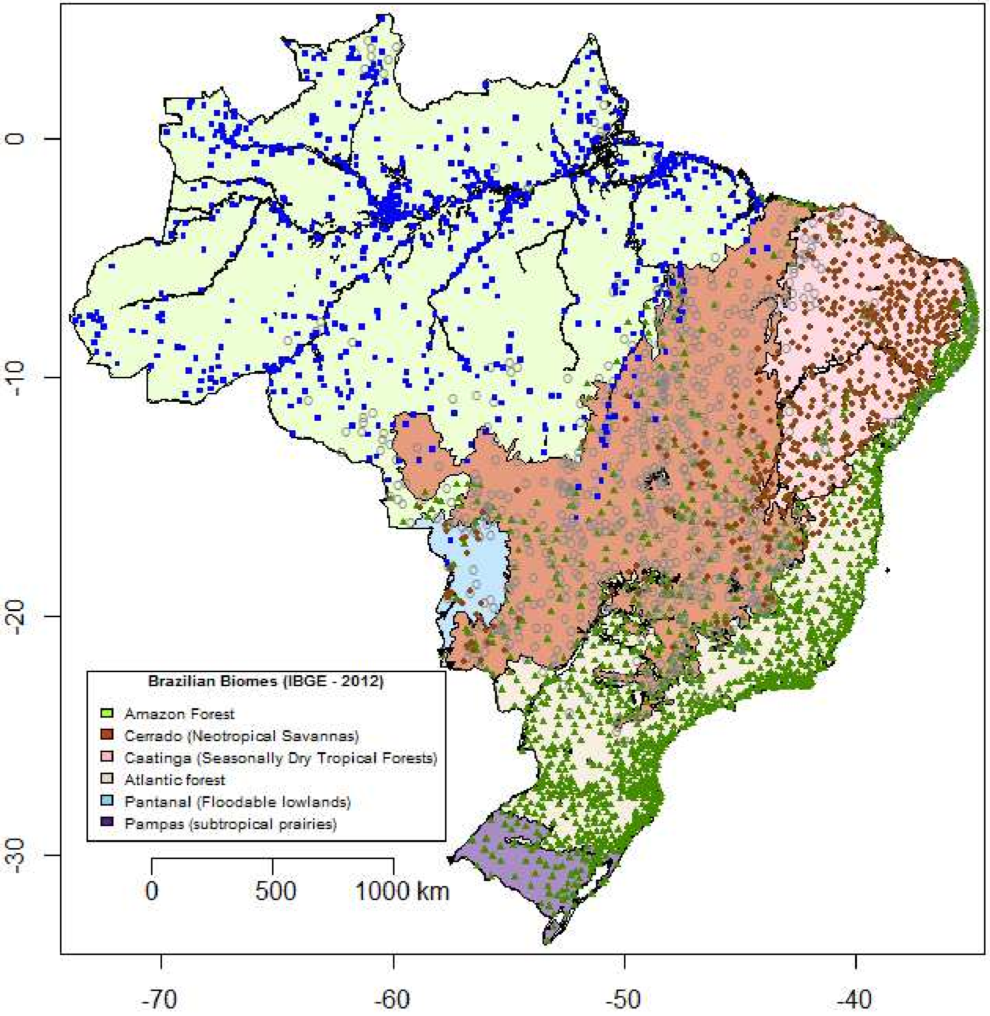

Figure S3 – Map of South America with areas coloured according to the Domain system of IBGE (2012), which are also sometimes referred to as biomes. The points on the map are the Brazilian NeoTropTree tree species inventory sites classified into biomes by this study: Atlantic Forest (green triangles), Seasonally Dry Tropical Forest (brown circles), Savanna (hollow gray circles), Amazon Forest (blue squares), and Chaco (inverted hollow black triangles).

Table S6: Confusion matrix between sites categorised according to Dinerstein *et al.* (2017) biome classification system (rows), which was adapted from Olson *et al.*(2001), and sites as categorised into biomes in this study (columns). The underlined numbers represents the sites that were assigned to matching categories between the two systems. The other elements are treated here as mis-classifications.

	Existing Lowland Tropical South America Biomes according to tree species composition					
Dinerstein *et al.* (2017) - Adapted from Olson *et al.* (2001)	Amazon Forest	Atlantic Forest	Savanna	Chaco	SDTF	Total
Tropical and subtropical moist broadleaf forests (tropical and subtropical humid)	994	994	117	0	41	2146
Tropical and subtropical grasslands, savannas and shrublands (tropical and subtropical semiarid)	57	403	544	99	90	1193
Tropical and subtropical dry broadleaf forests (tropical and subtropical semihumid)	21	88	78	1	420	608
Flooded grasslands and savannas (temperate to tropical fresh or brackish water inundated)	1	18	16	4	15	54
Temperate grasslands, savannas and shrublands (temperate semiarid)	0	10	0	12	0	22
Mangrove (subtropical and tropical salt water inundated)	24	51	5	0	0	80

Table S7: Confusion matrix between sites categorised according to IBGE (2012) biome/phytogeographic domain classification system (rows), which was adapted from Veloso (1992), and sites as categorised into biomes in this study (columns). The underlined numbers represents the sites that were assigned to matching categories between the two systems. The other elements are treated here as mis-classifications.

	Existing Lowland Tropical South America Biomes according to tree species composition					
Brazilian biomes	Amazon Forest	Atlantic Forest	Savanna	SDTF	Chaco	Total
Amazon Forest	635	20	74	1	0	730
Atlantic Forest	0	911	37	31	0	979
Cerrado	35	287	539	114	2	977
Caatinga	0	57	54	369	0	480
Pampa	0	78	0	0	1	79
Pantanal	2	9	21	18	1	51
Continental water mass	85	21	4	10	1	121
Oceanic water mass	6	89	2	2	0	99