Biodiversity and Conservation
https://doi.org/10.1007/s10531-021-02195-7

ORIGINAL PAPER

Recovering the drivers of sampling bias in Bignonieae (Bignoniaceae) and identifying priority areas for new survey efforts

Juan Pablo Narváez-Gómez[1] · Thaís B. Guedes[2,3] · Lúcia G. Lohmann[1]

Received: 27 September 2020 / Revised: 20 April 2021 / Accepted: 29 April 2021

Abstract
Identifying knowledge gaps and the potential biases and limitations of biological databases is essential for biogeographical research, to efficiently plan biodiversity surveys, and to accurately design conservation efforts. Here we describe the taxonomic, temporal, and spatial coverage of a comprehensive database mainly composed of data collected between 1963 and 2003 of the largest clade of Neotropical lianas, the tribe Bignonieae (Bignoniaceae). We assess the level of database completeness and propose new survey areas to fill knowledge gaps and optimize sampling coverage. The Bignonieae database includes 28,763 records representing 98% of the known species. The database covers 72% of the Neotropical region and includes data collected mainly during the last 40 years of the 20th century. Members of the tribe are conspicuous components of lowland forests, with most species showing narrow range sizes. The Amazon rainforest is the most under-sampled region and the area with the lowest sampling rate. On the other hand, the best sampled areas are scattered across Central America, the Peruvian and Bolivian Amazon, and selected Brazilian cities. Sampling rate across the geographical extent of Bignonieae was best predicted by the distance from cities. Collection effort is needed across the Neotropics so that a higher number of localities can be sampled, especially in the Amazon, where Bignonieae is centered. New surveys are urgently needed to maximize new species discoveries and to effectively design conservation plans that maximize biodiversity-rich regions facing increased threat.

Keywords Botanical inventories · Herbarium data · Lianas · Neotropics · Occurrence data · Wallacean shortfall

Communicated by Daniel Sanchez Mata

✉ Juan Pablo Narváez-Gómez
narvaez-gomez.jp@alumni.usp.br

✉ Lúcia G. Lohmann
llohmann@usp.br

Extended author information available on the last page of the article

Published online: 18 May 2021

Introduction

Biological databases are key resources for the understanding of biodiversity distribution patterns (Meyer et al. 2015). Current biodiversity threats contributed to expediting the digitization of specimens deposited in natural history museums around the world, allowing for an increase in the inclusion of distribution data in several macroecological and biogeographical studies (Soberón and Peterson 2004). Correlations among occurrence data and biotic and/or abiotic factors have allowed us to test hypotheses about the drivers of species distribution patterns (Pyke and Ehrlich 2010; Peterson et al. 2011; Wiens et al. 2011), and to implement efficient strategies for biodiversity management and conservation (Soberón and Peterson 2009). Despite the recent efforts and the easy access to biodiversity information, data quality remains a concern (Guralnick et al. 2007). Issues such as spatial and temporal biases in sampling effort, poor georeferencing quality, taxonomic errors, and lack of standards to effectively store specimen information can drastically affect the results of spatial analyses and their respective biological inferences (Newbold 2010; Daru et al. 2018; Yost et al. 2018). Therefore, understanding inherent biases of biodiversity databases allow us to evaluate the limitations of our data and the significance of our conclusions.

Biases in collection effort result from logistic limitations and opportunistic collection efforts, causing species occurrences to be clustered in space and time (Hortal et al. 2015). This aggregation and patchiness of species records is determined by the accessibility to remote areas, the relative biological importance of some regions (i.e., protected or high endemism richness areas), and the detectability of species given their phenology and biology, among others (Meyer et al. 2015). The richness peak around rivers and roads (Zizka et al. 2020) and the inflation of richness values around well-collected areas has been documented across taxa (e.g., Meyer et al. 2015; Guedes et al. 2018). Moreover, socioeconomic factors influencing research institutions and the costs and challenges involved in sampling remote places also increase the patchiness of records across and within species ranges (Meyer et al. 2015, 2016a). Indeed, most biodiversity databases include an imbalanced sampling of species ranges to a certain degree.

Biases can also emerge from the very process of assembling databases, including the lack of metadata documenting the assembly process, and the quality of the information contained in the database. The low quality of biodiversity databases might arise from different types of errors, including mechanical errors (e.g., typos, empty fields, mixed specimen information), georeferencing mistakes (e.g., wrong specimen localities, high geographical uncertainty, lack of georeferencing metadata), and/or taxonomic errors (e.g., lack of taxonomic information, specimen misidentification, outdated classifications) (Maldonado et al. 2015). These mistakes lead to false estimates of the habitats and environments occupied by each species, leading to erroneous inferences about species traits and their associated landscapes (Hortal et al. 2015). Accounting for these uncertainties and associated errors is critical for precise and accurate understanding of species distribution patterns and their underlying processes (Meyer et al. 2016b).

Several analyses can address biases in biodiversity databases. For example, by determining the spatial distribution of collection density from a database and associating this data with factors such as the distance from populated places and roads, we can estimate the biases produced by collection activity (Kadmon et al. 2004; Pautasso and McKinney 2007). By determining the completeness of individual biodiversity databases, we are able to identify localities with: (i) unknown survey efforts, caused by a lack of any documentation of the sampling effort; (ii) unknown absences, caused by a lack of data on species

absences; and (iii) unknown recurrences, caused by the exclusion of duplicate records of a single species from an individual location (Lobo et al. 2018). This information allows us to drawn ignorance maps where well-collected and poorly sampled areas can be identified (Lobo et al. 2018).

Sampling biases are especially problematic in tropical regions of the world, where species diversity is the highest (Prance et al. 2000; Kier et al. 2005; Raven et al. 2020). For example, the Neotropical region houses three plant diversity hotspots (i.e., Costa Rica-Choco, Tropical Eastern Andes, the Atlantic Rainforest, and the Cerrado; Myers et al. 2000). Even though the Neotropics are among the less known regions of the world floristically (Gentry 1982; Barthlott et al. 2005), with many possible unknown species to science (Pimm and Joppa 2015), efforts conducted in the recent decades have narrowed the knowledge gaps in this region (Ulloa Ulloa et al. 2017). Amazonia continues to represent the most under-collected biome in the world, with collecting efforts focused on the most populous regions, along roads, and rivers (Nelson et al. 1990; Hopkins 2007). Likewise, higher levels of richness and endemism in the Atlantic Forest are correlated with the location of major museums and herbaria, illustrating past collecting preferences and decisions by collectors. On the other hand, very severe Wallacean shortfall were found in plants of neglected biomes such as the Cerrado and Caatinga (Simon and Proença 2000; Proença et al. 2010; Santos et al. 2011). Similar patterns have been recovered for other neotropical countries such as Mexico (Bojórquez-Tapia et al. 1995), Colombia (Arbeláez-Cortés 2013), Peru (Rodríguez and Young 2009), Ecuador (Engemann et al. 2015), and Guyana (Funk et al. 2005). Given that spatial bias in sampling effort is so pervasive across the Neotropics, detailed analyses of the biases and uncertainties of Neotropical databases are greatly needed.

Bignonieae (Bignoniaceae) is a monophyletic tribe, broadly distributed throughout the Neotropics, including the Antilles (Lohmann and Taylor 2014). Taxonomic studies of this group conducted by one of us (LGL) during the past 30 years have led to the compilation of a comprehensive database with ca. 30,000 occurrence points of verified specimens. Previous versions of the Bignonieae database were used to evaluate potential correlations between range size and detectability (Sheth et al. 2008), as well as to estimate biases in range size estimation (Sheth et al. 2012). While detectability and range size are not correlated in this group (Sheth et al. 2008), the greater the area of occupancy of individual species, the lower the spatial bias (Sheth et al. 2012). Other studies have used the Bignonieae database to investigate species richness and the relationship with species traits and the environment (Meyer et al. 2018, 2019, 2020). Despite that, the geographical biases caused by accessibility and differential collection effort of the Bignonieae database have not been addressed. Here, we conduct a thorough analysis of the Bignonieae database, including an assessment of its level of completeness. We further explore its spatial, temporal, and taxonomic coverage, as well as propose new survey areas and taxa to fill knowledge gaps and optimize coverage.

Methods

Database assembly

The compilation of the Bignonieae database followed three stages that were conducted during the course of 20 years. During the first stage, data for 30,227 specimens of

Bignonieae that were available in the Tropicos database until the year 2003 were downloaded (http://legacy.tropicos.org/SpecimenSearch.aspx). LGL inspected and verified the taxonomic identity of all of these herbarium specimens deposited at the Missouri Botanical Garden (MO). Geographic coordinates were extracted from herbarium specimens whenever available and plotted in ArcMap 9.1 (ESRI 2005) to confirm that the coordinates really belonged to the described location. Whenever coordinates were not included in the specimen label, the most specific locality was identified and its coordinate determined using regional maps and online gazetteers, especially the "Getty Thesaurus of Geographic Names Online" (http://www.getty.edu/research/tools/vocabularies/tgn/). Duplicated records or records with limited information for georeferencing were excluded producing a database of 26,660 records. This database became the basic source of geographic information for a synopsis of tribe Bignonieae (Lohmann and Taylor 2014).

During the second stage, species names were updated following subsequent taxonomic treatments of selected genera, namely: *Adenocalymma* (Fonseca and Lohmann 2019), *Bignonia* (Zuntini et al. 2015a, b), *Dolichandra* (Fonseca et al. 2017), *Pachyptera* (Francisco and Lohmann 2018), *Martinella* (Kataoka and Lohmann 2021), *Tanaecium* (Frazão and Lohmann 2019), *Tynanthus* (Medeiros and Lohmann 2015), and *Xylophragma* (Kaehler et al. 2019). For species with less than five specimens, new records were included from the herbarium SPF (University of São Paulo) adding 503 new records to the database. Besides, 1600 newly curated records of *Adenocalymma* were included following an updated synopsis of this genus (Fonseca and Lohmann 2019). Therefore, the database reached 28,763 records for which the taxonomic data is fully curated. To increase the geographic accuracy, we searched for geographical outliers in the individual species distributions by plotting specimen coordinates in QGIS 2.18.16 (QGIS Development Team 2018) and visual inspecting distant occurrence points for the 386 species of Bignonieae included here (for further details see Meyer et al. 2019). The taxonomic identity and the correspondence between locality descriptions and geographic coordinates of each outlier were verified by LGL, Leila Meyer (University of Goiás, Brazil), and JPNG. Mismatches between the geographical coordinates and locality descriptions were corrected. The georeferencing process involved the interpretation of locality descriptions in specimens' labels and the extraction of geographical coordinates in Google Earth (https://earth.google.com/web/) following the BioGeomancer Consortium manual of best practices for georeferencing (Chapman and Wieckzoreck 2006). Details of this version of the Bignonieae database are provided in Fig. S1, see Online Resource in Electronic Supplementary Material.

The third and final stage of the database compilation consisted in further curation of geographical information for the current 28,763 records included in the database. Two rounds of revisions were implemented. The first round occurred after identifying outliers in the altitudinal profile of species distributions (see below altitudinal range profile). Extreme altitudinal data points were checked for possible georeference errors and corrected whenever necessary following the same procedure described earlier. The second round consisted in the identification of potential errors in geographic coordinates such as geographic coordinates located over centroids, laying in the sea, occurring around research institutions, and the presence of zero coordinates. To flag these issues, we used the function clean_coordinates() from the package CoordinateCleaner (Zizka et al. 2020) implemented in R programming language (R Core Team 2020). All records flagged as problematic were mapped and verified using the locality descriptions in Google Earth.

Temporal, taxonomic, and spatial coverage

We described the Bignonieae database in detail. Taxonomic coverage was represented as the number of records per genus, with unique and duplicated localities (i.e., records for the same species occurring in the same locality, but collected by different collectors at different times) being accounted for. The temporal coverage of Bignonieae database was measured as the number of collected specimens per collection date per year and month across the complete temporal span covered. The most representative collectors were identified by counting the total number of collections per collector in the Bignonieae database. As indicated above, most records included in the current database were downloaded from Tropicos (MO) in 2003, meaning that collections made after this date were not included in this dataset, with a few exceptions.

We accessed the spatial coverage of the Bignonieae Database by superimposing the records to different geographic operational units (one degree lat/long cells, administrative areas, biogeographical regions sensu Morrone 2014, and ecoregions sensu Olson et al. 2001). Range size and the altitudinal profile of each species was estimated as described below.

Range size profile

We used the function CalcRange available in the speciesgeocodeR R package (Töpel et al. 2017) to calculate the range of each species through convex hull. We defined species range size categories using the kmean clustering technique available in the *stats* package of the R programming language (R Core Team 2020). This cluster analysis technique partitions a set of observations of range sizes into k groups, where membership to the k group is determined by the shortest distance to the group mean range size value. It applies an iterative and heuristic algorithm that sets an arbitrary number of cluster centroids around which observations are grouped together based on the minimum mean distance to centroids. In a second step, this algorithm recalculates the centroid value from the observations in the group; this procedure is repeated 100 times. We assigned the number of centroids and clusters to classify the species range size in four categories, as follows: narrow, medium-narrow, medium-wide, and wide. We divided the medium category in two (i.e., medium-narrow and medium-wide) because the amplitude of the wide cluster and the variance of range sizes included was high when only three categories were used (i.e., narrow, medium, and wide). Species with less than three unique locality records were manually assigned to an additional category named “micro”.

Altitudinal range profile

An elevational database was created by cross checking the Bignonieae database with elevation data obtained from the U.S. Geological Survey’s EROS Data Center (https://www.usgs.gov/centers/eros) and Google Earth Pro 7.3 (https://earth.google.com/web/). This procedure included four steps: (i) the GTOPO30 global digital elevation model was downloaded from the USGS service in raster format from different global zones encompassing the Neotropics (courtesy of the U.S. Geological Survey, 2004) and using the geographical extension of Bignonieae as proxy; (ii) all layers were merged into a unique raster file using QGIS 2.18.16 (QGIS Development Team 2018); (iii) elevation values for all point records

in the occurrence database were obtained using the extract function from the raster package in R; and, (iv) outliers were identified using boxplots for each species and compared with the elevation data already available from locality descriptions in the occurrence database (when these values were different we opted to keep the value from the collector). New georeferences were provided only when points were erroneously georeferenced (see last paragraph of the database assembly section). The same procedure was applied to maximum and minimum elevation values for each one of the 386 species. Some altitudinal outliers remained even after cleaning the elevation data because these outliers were interpreted to reflect true species occurrences.

Spatial biases and database completeness

We used the sampbias R package (Zizka et al. 2020) to assess the effect of accessibility over the geographic biases of the Bignonieae database. This method calculates the expected change in the sampling rates of the individual records as a function of the distance from rivers, airports, and populated places. This approach describes observed sampling rate as a Poisson process and models the expected species records as exponentially decaying from these geographic features using a Bayesian statistical framework. Because different biasing factors might be correlated (e.g., cities, airports), the sampbias R package estimates the joint effect of all factors. This method operates under the assumption that species are distributed across the entire study region and calculates how strong the biases are, while identifying unexplored places. The bias effect is then interpreted as the proportion of records missed in each cell as a function of distance to geographical features. If the biases are strong, a fast-decaying sampling intensity is expected from the specific type of geographic bias factor being examined. Unexplored and under-collected places are identified as those with the lowest sampling rates and no observed records, reflecting the difficulty to access the region. This analysis was run using a spatial scale of one degree, with a buffer of two, using the sampbias default gazetteer.

We used the R package KnownBR (Lobo et al. 2018) to analyze the geographical distribution of survey completeness and identify places with the highest and lowest knowledge of Bignonieae diversity. This analysis estimates species accumulation curves for each geographic unit under examination and estimates the survey coverage intrinsic to the database. To achieve this, the analysis assumes that the distributional database is the most comprehensive possible and uses the number of records and species to calculate species accumulation curves for each geographical unit included in the analysis. Under the assumption of infinite survey, these curves are fitted to theoretical functions with asymptotic behavior to predict how many species would be expected in each geographic unit (Lobo et al. 2018). The percentage of observed against expected records defines the completeness of the database, representing a surrogate of the survey effort and knowledge contained in the database. The final slope of the accumulation curve tells the amount of effort necessary to complete the survey within a particular geographic unit. The values of slope, completeness, and ratio of records per species indicate the quality of the survey conducted in each geographic unit. Lower values of slope, greater values of completeness, and higher observed–expected species ratios define the best surveyed areas.

We applied this method to the Bignonieae database to assess the quality of the geographic information and how well the species diversity is known for each 1° cell in which species occurrences are recorded across the whole geographic extent of the tribe. This analysis was implemented using a format A matrix of species occurrences using grid cells

of 1°, a ratio between records and species of one (R/S), and by applying the exact estimator of Ugland et al. (2003) to obtain the species accumulation curves and to estimate sampling completeness. Although higher ratios of species records are preferable, we used an R/S = 1 because the point occurrence density is low across the geographic extent of the tribe. This R/S ratio allowed us to discriminate between cells with higher completeness and lower slopes from cells with lower completeness and higher slopes. We also calculated the quality of the survey effort in each cell using the function SurveyQ of the "KnownBR" package (Lobo et al. 2018). This function uses the completeness, final slope of the species accumulation curves, and the R/S ratio to identify localities with high and low sampling effort in order to help decide future survey efforts (Lobo et al. 2018). We used the default definitions for well (slope < 0.02, completeness > 90%, and R/S ratio > 15) and poorly (slope > 0.3, completeness < 50%, and R/S ratio < 3) sampled localities.

Identifying priority areas for new survey efforts

To visualize the geographical biases and the knowledge gaps that must be alleviated in further survey efforts, we combined the information produced by the survey effort analysis, the sampling bias analysis, and the cells without occurrence records using basic raster operations. As a surrogate of our ignorance of the geographical distribution, we extracted the following information in different raster layers: (1) the cells with no records in the geographical extend of Bignonieae, (2) the poorly surveyed cells, and (3) the cells with the lowest sampling rates conditional to geographical accessibility. A cell might be characterized by only one, two, three or none of these sources of information. There are seven possible combinations among these three kinds of data (i.e., 1, 2, 3, 1-2, 1-3, 2-3, and 1-2-3). To numerically codify these categories, we reclassified the layers so the cells without records had a value of one, the poorly surveyed cells had a value of three, and the cells with the lowest sampling rates had a value of five. After combining these three layers, all possible combinations of information in the new raster layer are numerically distinguishable because the sum of the sequence 135 produces the numbers 4, 6, 8, and 9. This way every cell on the layer is classified according to a cell type that describes different combinations of the results of all analyses. Other arithmetic sequences of three numbers can be used provided they fulfill the condition that when these numbers are added they will produce different numbers to identify each one of the different combinations among the analyses. Notice that cells without records are incompatible with cells with poor surveys because the package "KnowBR" only calculates the completeness in cells that have occurrence records. This means that there are no combinations that imply adding cells with these conditions. The final raster layer is plotted in a map to show the regions that are least sampled and their relationship to accessibility. Through this approach we were able to identify priority areas for new survey efforts aiming to alleviate knowledge gaps in the distribution of Bignonieae species that resulted from accessibility limitations.

Results

Temporal, taxonomic, and spatial coverage

The Bignonieae database comprises 386 species representing all 20 Bignonieae genera currently recognized (Lohmann and Taylor 2014; Fonseca and Lohmann 2019). Overall,

the database includes 28,763 records of which 21,170 are unique localities (same unique combination of XCOOR and YCOOR), while 7593 correspond to collections made for the same species at the same locality at different times (Online Resource). Within the Bignonieae database, five genera are more representative accounting for 18,512 (64.36%) records: *Fridericia* (5201 records representing 59 spp.), *Bignonia* (4117 records representing 30 spp.), *Adenocalymma* (3803 records representing 72 spp.), *Amphilophium* (3109 records representing 46 spp.), and *Tanaecium* (2282 records representing 21 spp.) (Fig. 1a, b). Together, these five genera encompass 228 out the 386 sampled species, representing 59% of the overall species diversity in the database. In contrast, *Manaosella* and *Perianthomega*, two of the monospecific genera of Bignonieae, showed the lowest number of records, i.e., 56 and 26 records, respectively. In total, 304 species have more than 10 records each, 60 species have less than 10 records, and 22 species have less than three records.

Most specimens included in this database were collected between 1963 and 2003, with only 657 specimens collected after 2003 incorporated to fill in some species distribution gaps. Overall, 4587 specimens (15.95%) were collected before 1963, while 18,835 specimens (65.48%) were collected between 1963 and 2003, and 2.28% (657 specimens) were collected from 2004 onwards (Fig. 1c). The remaining 4684 specimens (16.29%) do not include information about collection date. The distribution of records by month shows that specimens were collected throughout the year, with a minor decrease in collections from August to December, which correspond to the cooler and drier months in the Tropics (Fig. 1d). The highest peak of collection activity throughout the complete temporal span of the database coincides with the active years of A.H. Gentry, E. Zardini, G. Hatschbach, and J.A. Steyermark. However, a monthly tendency was not identified among the 12 most productive collectors (Online Resource). Overall, the temporal coverage of the database seems to follow a peak of collection activity during the 80's to 90's, with a mild tendency to higher collection activity during the summer.

The geographic extent of Bignonieae currently covered by the occurrence database encompasses the continental platform of America between 39°N and 35°S of latitude and 35° W and 110° W of longitude, and the Antillean Islands in the Caribbean Sea (Fig. 1e). This tribe encompasses the whole Neotropical region, extending some degrees further into North America, where *Bignonia capreolata* occurs. When administrative areas were examined, Brazil presented the highest number of records, species, and endemic taxa (Fig. 1f–h), doubling the numbers of Venezuela, Peru, Bolivia, and Colombia altogether. Likewise, the smaller biogeographical units of the regionalization scheme of the Neotropical region showed the lowest counts of records, species, and endemic taxa (Fig. S4–5, see Online Resource in Electronic Supplementary Material). Namely, Bignonieae was shown to occur predominantly in the Brazilian and Chacoan subregions, with the latter showing the highest number of endemic species (Online Resource). The tribe was also found in both the South American and Mexican transition zones. However, while one endemic species was found in the Mexican transition zone, no endemic species were recovered in the South American transition zone. Similarly, the three dominions with highest numbers of records were the Pacific, the Boreal Brazilian, and the Parana dominions, respectively; the three dominions with the highest number of endemics were the Parana, the Brazilian, and the Pacific dominions (Online Resource). Few provinces included endemic species, with the highest number of endemics located in the Atlantic Forest, followed by the Caatinga and Parana provinces (Online Resource). Ecoregions showed a similar pattern. However, given that ecoregions show a smaller number of subdivisions, the number of endemics recovered in this

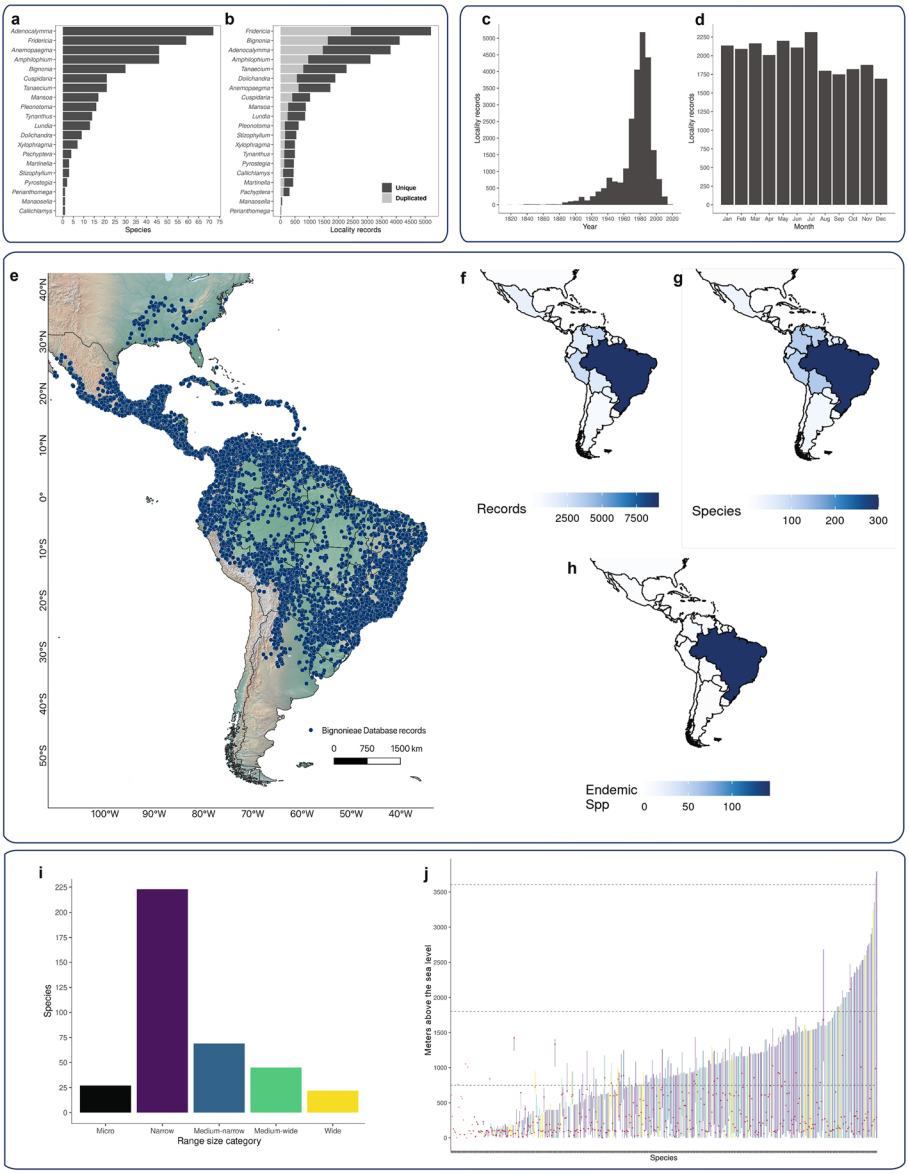

Fig. 1 Taxonomic, temporal, and spatial coverage of the Bignonieae database in the Neotropics: **a**, **b** Number of unique and duplicated records per genus. **c**, **d** Number of specimens sampled throughout months and years (1816–2015). **e** Geographical extension of the Bignonieae database. **f** number of records, **g** number of species, and **h** number of endemic species by administrative regions in tropical America. **i** Number of species in each range size class. **j** Altitudinal range profile, each line represents the linear altitudinal range between the minimum and maximum elevation values per species; red dots indicate that the median value is located within the range, while dashed horizontal lines show the limits of the boundaries of lowlands (750 m.a.s.l.), montane (1800 m.a.s.l.), and paramo (4500 m.a.s.l.) altitudinal zones

region was smaller (Online Resource). In sum, when considering the spatial coverage of different geographical units, the number of endemic species decreased at smaller and less inclusive spatial subdivisions.

Range size

Most species showed narrow range sizes (Fig. 1i; Online Resource). When range size variation was broken into categories using a K-means clustering with four centers, the following species numbers per categories were recovered: (i) 223 species with narrow range sizes (i.e., below 1,361,662 km^2) (ii) 69 species with medium-narrow range sizes (i.e., between 1,410,303 km^2 and 4,357,140 km^2); (iii) 45 species with medium-wide range sizes (i.e., between 4,653,741 km^2 and 7,731,285 km^2); and (iv) 22 species with widespread range sizes (i.e., between 11,425,411 km^2 and 18,642,351 km^2) (Table 1). Area calculations were not possible for the 22 species with less than three locality records classified under the "micro" range size category.

Altitudinal range

The altitudinal profile of the Bignonieae database shows that the tribe is conspicuous in the lowlands, although a relatively high number of species show wide altitudinal ranges due to a few outliers (Fig. 1j). No genera occupy a preferred altitudinal belt (Online Resource). While 137 species are restricted to lowlands (less than 750 m.a.s.l.) and seven species are restricted to mountains, 239 species are found in both of these altitudinal belts. In general, species with smaller altitudinal ranges also show fewer occurrence points and restricted ranges. On the other hand, species with wider altitudinal ranges are also widespread and show a higher number of occurrence points. Although outliers were checked and wrong georeferences were addressed, some outliers remained, displacing upward the altitudinal range of these species. Given that the taxonomic identity of all specimens was carefully verified by a Bignonieae expert (LGL), these outliers are assumed to represent correct occurrences of the species geographic distributions.

Quality of georeferenced occurrence points

We gathered georeferenced occurrence points from two sources: coordinates recovered from specimen labels and coordinates assigned from the interpretation of locality descriptions. In the Bignonieae database, 1114 records (3%) were flagged as possible geographic errors, which is an excellent indication of the uninterrupted database curation. The main

Table 1 Definition of range size categories in square kilometers from Kmeans clustering showing the number of species per class, centroids, quantiles, minimum and maximum values

Size class	Species	Centroid	Min	1st Q	2nd Q	3rd Q	Max
Narrow	223	358,428,5	0	58,218	192,842	592,229,5	1,361,662
Medium-narrow	69	2,435,854,7	1,410,303	2,226,514	2,226,514	3,031,349	4,357,140
Medium-wide	45	6,694,466,1	4,653,741	6,186,717	6,186,717	7,731,285	10,400,739
Wide	22	15,763,456,5	11,425,411	12,511,736	13,543,922	18,642,351,8	26,349,177

issue in this assessment was the presence of records close to capitals and country centroids, where 562 records were flagged as problematic. These records were maintained as such, because the species associated with those records were shown to truly occur nearby capitals and country centroids. The second potential source of geographic errors were 509 records that felt in the sea. A detailed evaluation of those records indicated that most of these records were located near the coastal shores and were not encompassed by the reference polygon used in the CoordinateCleaner package (Zizka et al. 2020). In addition, several records from the Antillean islands were also located within the sea, which is due to the fact that geometry of the islands displace the centroid out of their territories, or due to points laying in the middle of rivers. All of these records were georeferenced using the locality description in order to correct for these kinds of mistakes. Part of these records were also identified by verifying the altitudinal range profile of each species with altitude values of 0 m.a.s.l. The third potential source of geographic errors was associated with 65 records classified as species range outliers. These outliers were also assumed to represent correct occurrences because they were checked in previous stages of data curation and were shown to fall within the geographic distribution of the associated species. The fourth potential source of error corresponded to records located over administrative area centroids that were either georeferenced with imprecise locality descriptions or using natural reserve centroids. In all of these cases the original coordinates were conserved. The last two issues, the presence of zero coordinates and records located near biodiversity institutions, were not problematic because the species ranges were shown to encompass these localities. An inspection for duplicated records identified 7576 records. However, these records corresponded to different collections made at different times, by different collectors, and those records were maintained. In sum, the assessment of georeference quality allowed us to verify the high quality of the Bignonieae database.

Sampling biases and accessibility

A map of the database at 1° grid cells of spatial resolution showed that although the spatial coverage of the Neotropics is high, with 72% of the total number of cells showing at least one record, most cells showed less than 50 records (Fig. 2a). The Amazon was under-collected with huge gaps inside the biome and adjacent areas next to the Cerrado, the Savannas of Colombia, and Peru. The great Chaco was also shown to be under-collected despite a good spatial coverage in surrounding areas. Some centers with a high number of occurrence records (i.e., more than 200 records per grid cell) were identified within and around the following locations: (i) San Jose (Costa Rica); (ii) Barro Colorado Island (Panama); (iii) Iquitos, Manu, and Madre de Dios National Parks (Peru); (iv) Natural Reserves Madidi and Noel Kemp, and Santa Cruz (Bolivia), (v) Asunción (Paraguay), and (vi) Manaus, Belém, São Paulo, Rio de Janeiro, Brasilia, and Belo Horizonte (Brazil), among others. Species richness was high in just few cells scattered inside and around Amazonia, southeastern Brazil, and Central America (Fig. 2b). The highest richness per cell was 65 species; these same cells also included the highest occurrence records count per cell. Most grid cells included less than 20 species.

The distance from cities was the main predictor of sampling rate across the geographical extent of Bignonieae, followed by a moderate effect of rivers and airports, and a negligible effect of roads (Fig. 2c, d). This means that the number of expected records rapidly decreased with distance from cities, while records decreased in a steady fashion with distance from roads. The geographical projection of sampling rates conditional to these bias

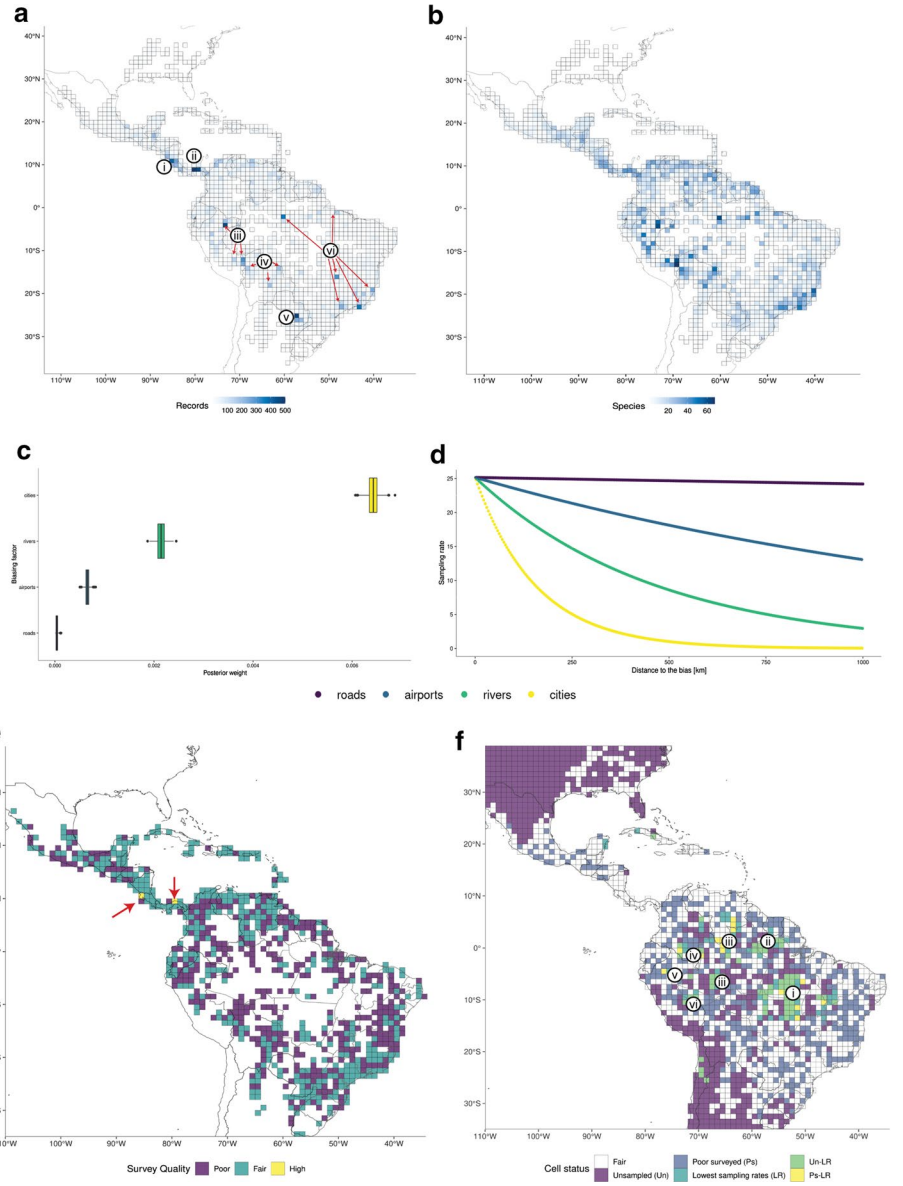

Fig. 2 Spatial bias and ignorance maps of the Bignonieae database throughout the Neotropics. **a**, **b** Number of occurrence records and species (richness) by one degree grid cell. The centers with more than 200 records per grid cell are indicated by red arrows per country (i) Costa Rica, (ii) Panama, (iii) Peru, (iv) Bolivia, (v) Paraguay, and (vi) Brazil. **c** Posterior weight of each category of biasing factors. **d** Change in sampling rate with distance for each biasing factor, showing the distance to cities as the strongest biasing factor. **e** Geographical projection of sampling effort showing High, Fair, and Poor surveys per geographical unit, highlighting the highest quality surveys in Costa Rica and Panama. **f** Map of ignorance on the knowledge of species richness and distribution of Bignonieae. Areas with the highest collection priority located within the Amazon basin. (i) Brazilian states of Mato Grosso and Pará; (ii) northern Pará; (iii) northern and southern portions of the Amazonas state; (iv) Colombian Amazon; (v) northern Perú; and, (vi) region between Acre and Perú

factors showed that the Amazon region has scattered areas with the lowest sampling rates throughout the geographical extent of Bignonieae (Online Resource). These areas are the most affected by accessibility biases. Other regions such as south-eastern Brazil and northern Andes are also biased towards cities despite the higher number of collections and species in these regions.

Completeness of survey efforts

The species accumulation curves and the completeness of each geographical unit showed a moderate number of cells with lower values of slope, and a lower number of cells with higher completeness, suggesting that the sampling effort has been heterogeneous. Lower values of slope (Online Resource) indicate that a higher number of records is necessary to discover new species either because the diversity is well-sampled or because there is a high prevalence of rare species within individual geographic units. These units with lower values of slope are relatively sparse across the Neotropics, but are common in Central America, the Antillean Islands, and south-eastern Brazil. Cells with completeness values higher than 80% are scarce, although a few are located in Central America and the Antilles, while others are dispersed across South America, with a slightly higher concentration in South-eastern Brazil, Paraguay, and northern Argentina (Online Resource). Intermediate values of completeness (i.e., around 50%) and completeness values below 30% are more numerous and well distributed across the geographic extent of the tribe, indicating that additional survey efforts are needed across the whole geographical extent of the tribe (Online Resource).

Survey effort quality showed that higher collection efforts are still needed in order to accurately represent the species diversity of the tribe Bignonieae in the Neotropics (Fig. 2e, f; Online Resource). The relationship between completeness, the ratio between records and species (R/S), and the final slope of the species accumulation curves varied across localities. While the best surveyed localities were characterized by high completeness values, low final slopes, and high number of records relative to the number of species (R/S), the worse surveyed localities were characterized by low completeness, low slopes, and low R/S ratio values (Online Resource). Only two cells showed high quality surveys in contrast to 390 cells with fair quality sampling, and 397 with poor sampling (Online Resource). The two high quality surveyed cells were located in Costa Rica and Panama. In sum, despite the high quality of the Bignonieae database, additional survey efforts are still needed throughout the geographical extent of tribe Bignonieae (Fig. 2e, f).

Priority areas for future surveys

The Amazon is the most under-sampled region throughout the Neotropics, as well as the area with the lowest sampling rate, which is due to accessibility biases (Fig. 2e). The seven priority areas for future sampling within the Amazon are: (i) an area that encompasses the Brazilian states of Mato Grosso and Pará, next to the transition zone between the Cerrado and the Amazon; (ii) an area located in northern Pará, next to Suriname and Guiana; (iii) one area in the northern portion of the state of Amazonas, close to Venezuela; (iv) one area in the southern portion of the state of Amazonas; (v) the whole Colombian Amazon; (vi) an areas located in northern Peru; and, (vii) an area between the Brazilian state of Acre and Perú (Fig. 2f). Similarly, an area encompassing the Brazilian states of Maranhão, Tocantis, Piaui, and Bahia in the Brazilian Cerrado is also under-sampled (Fig. 2e, f). Poor surveyed areas are also scattered across the whole geographical extent of the tribe and should

be considered as secondary priorities when planning new expeditions (Fig. 2e, f; Online Resource). On the other hand, Central America and south-eastern Brazil are the regions where sampling has been most intensive.

Discussion

Identifying knowledge gaps and biases in biodiversity databases is fundamental to appropriately frame research questions and understand the scope of interpretations and conclusions. The Wallacean shortfall (i.e., the lack in the knowledge about species distributions) derives mainly from the strong relationship between geographic information and the collecting effort done by researchers to gather species distribution data, which has the undesirable consequence of aggregating records in space and time (Hortal et al. 2015). The Bignonieae database we describe here include data from Tropicos (MO) downloaded in 2003 and samples collected by many different botanists over the course of ca. 40 years, mainly between 1963 and 2003, although a few collections (2.28% of the whole database) made between 2004 and 2015 were added subsequently in order to complement the distributions of some species with less than ten herbarium records. The high quality of the geographic data contained in the Bignonieae database is due to several rounds of database curation conducted over the course of approximately 20 years. The taxonomic, spatial, and temporal coverage of this database are wide, encompassing most species known, across the whole geographical extent. Knowledge gaps were identified, especially in Amazonia, which is still incompletely surveyed for these lianas although the tribe is centered in this region (Meyer et al. 2018). The completeness assessment and the ignorance maps produced here will help increase the efficiency of future Bignonieae sampling.

Coverage of the Bignonieae database

The Bignonieae database is among the best datasets currently available for the study of plant diversity and distribution in the Neotropics (Hopkins 2007). This dataset was mainly assembled from specimens deposited at the Missouri Botanical Garden, where Alwyn Gentry, the most prolific collector of the Bignoniaceae, worked and deposited his samples. Several works have been published using earlier versions and different subsets of this database, aiming to address different questions about Bignonieae distribution patterns (Elith et al. 2006, 2020; Hopkins 2007; Sheth et al. 2008, 2012; Meyer et al. 2018, 2019, 2020; Paz et al. 2021). The classification adopted in the Bignonieae database follows Lohmann and Taylor (2014), but also includes subsequent published taxonomic updates (i.e., Medeiros and Lohmann 2015; Zuntini et al. 2015a, b; Fonseca et al. 2017; Francisco and Lohmann 2018; Fonseca and Lohmann 2019; Frazão and Lohmann 2019; Kaehler et al. 2019; Kataoka and Lohmann 2021). Despite that, specimens collected since 2004 are still being curated and have not yet been incorporated into the database. The Species Link (http://www.splink.org.br) alone lists more than 2600 Bignoniaceae records collected primarily by L.G. Lohmann (744 specimens), A.R. Zuntini (482), L.H. Fonseca (320), A.F. Nunes (249), A. Nogueira (236), F. Firetti (229), M. Kaehler (202), M.R. Pace (176), R.S. Ribeiro (98), and J.N. Francisco (69). These records might help improve some of the sampling gaps and should be incorporated into the Bignonieae database once this data has been fully curated. Moreover, there is a clear need to gather specimen information that is available in local herbaria which would be crucial to increase the geographical representation of

Bignonieae (see Colombo et al. 2016). In particular, several Brazilian herbaria (e.g., ASE, CEN, EAC, EAN, HCDAL, HDELTA, HUTO, HUEFS, INPA, IPA, JPB, MAC, MAR, MOSS, PEUFR, NY, RON, TEPB, UB, UFMT, and UFP) house important Bignonieae specimens that need to be studied in detail. Additional efforts to collect and curate this information are ongoing and will be integrated into the database in the years to come.

The database temporal sampling seems to be relatively homogeneous throughout the year, with a mild increase in sampling effort during the summer months (Fig. 1d). This suggests that biases on detectability by phenology might not be severe in this database, corroborating earlier findings (Sheth et al. 2012). However, further phenological studies of Bignonieae are still needed. Although the taxonomic coverage of the database includes all twenty genera of Bignonieae recognized to date (Lohmann and Taylor 2014; Fonseca and Lohmann 2019), the number of unique records is high for each genus. Furthermore, the proportion of duplicated localities is also significant (Fig. 1a, b). Further efforts to explore new regions are necessary so that different localities are added to known species ranges. Additional sampling would also increase the probability of finding new species across the geographical extent of the tribe.

The geographic coverage of the Bignonieae dataset is high, with records reported from all Neotropical countries and geographical units identified by various biogeographical classification schemes (e.g., Olson et al. 2001; Morrone 2014) (Fig. 1e). The number of occurrences, species, and endemic species decreases with the size of the geographical unit used to aggregate and count records. Most species appear to have narrow range sizes, with the mean range size within this category being around 358,439 km^2 (Fig. 1i). Range size categories were defined based on the Bignonieae database, with narrow range sizes being seven times that of narrow endemics for other groups of organisms (i.e., 50,000 km^2). Placing range sizes into categories helps us understand the variability of predefined operational geographic units. Further studies about patterns of species co-occurrence are needed to better describe and understand several distribution patterns in Bignonieae, especially regionalization proposals, patterns of endemism, phylogenetic and endemism diversity measures (Guedes et al. 2018). Information about patterns of species co-occurrence helps us to understand how geological and climatic predictors are associated to the origin and diversification of the tribe.

The Bignonieae database supports the observation that this tribe constitutes a conspicuous lowland plant clade (Gentry 1979; Lohmann and Taylor 2014). It also shows that a high proportion of species reach mountain regions (Fig. 2f), although high altitude records correspond to geographical outliers. The amplitude of the geographical range is positively correlated to the number of records, suggesting that a higher collection effort is needed to increase the knowledge of the altitudinal range of the species of this tribe. Reviewing altitude data allowed us to identify several erroneous georeferenced localities that were subsequently corrected.

Geographic biases and survey completeness

The spatial coverage of Bignonieae database throughout the Neotropics was substantial at the spatial resolution of one degree of grid cell size. The cells with the highest occurrence record numbers and species richness are not necessarily coincident (Fig. 2a, b). Records were mainly biased toward cities and secondarily towards rivers (Fig. 2c, d). This pattern of aggregation of occurrence records around populated places and routes that guarantee accessibility to the surveyed regions has been documented in several taxa in the Neotropics,

especially plants (Nelson et al. 1990; Kadmon et al. 2004; Pautasso and McKinney 2007; Vale and Jenkins 2012; Oliveira et al. 2016; Guedes et al. 2018). Cities are the centers of botanical institutions from which expeditions are generally undertaken and the entrance to remote and unexplored places. The Bignonieae database clearly reflects this general pattern. South-eastern Brazil is not only one of its centers of diversity, but also one of the most intensively sampled regions, likely reflecting the high number of research centers and universities located in this region (Sousa-Baena et al. 2014).

The Amazon sensu lato (including the Guiana region) corresponds to the center of diversity of the tribe (Meyer et al. 2018). Despite that, it is by far the most under-sampled region for Bignonieae, with vast areas showing the lowest sampling rates, and often not a single collection record (Fig. 2f). In this region the effect of rivers has clearly biased collection efforts, with collections concentrated along rivers. The Amazon is one of the most remote and under-sampled Neotropical areas for many taxa (Milliken et al. 2010; Guedes et al. 2018). Knowledge from occurrence databases obtained from herbaria are insufficient to account for species diversity and distribution in this region, calling for additional botanical expeditions (Hopkins 2019). Increasing sampling efforts in the Amazon has become even more urgent in recent years given the high deforestation rates, which are eliminating many species-rich yet under-sampled regions (Stropp et al. 2020).

In order to identify priority locations for new survey efforts, we analyzed the survey completeness of one-degree cells across the Neotropics, in locations where Bignonieae species have been recorded (Fig. 2f). For this analysis, we considered the Bignonieae database as the most exhaustive compilation of information available for this tribe to date. Our analyses indicate that new species discoveries are likely to emerge from a high number of locations, while fewer places seem to represent well the diversity of Bignonieae species. Databases where a large group of species are only known from a few geographical units, while widespread species dominate the records in cells in vast areas across the whole geographical extent are common among plants (Tobler et al. 2007).

Priority areas for new survey efforts

To properly identify priority collection areas, we classified cells as poor, fair, and high-quality surveys (Fig. 2e). Half the sites where Bignonieae is known to occur were classified as poor-quality indicating that revisiting those locations can increase the number of species reported. Lots of these cells are within Amazonia, highlighting how important it is to intensify sampling efforts in this region. Although only two cells located in the Neotropics (i.e., around Barro Colorado Island in Panama and Guanacaste in Costa Rica) were classified as high-quality surveys, a lot of cells with fair-quality surveys were also recovered across the Neotropics, a pattern that has been recovered for other groups (Sousa-Baena et al. 2014; Pelayo-Villamil et al. 2018; La Sorte and Somveille 2020).

Our ignorance map compiled from cells without records, poorly surveyed cells, and cells with lowest sampling rates suggested priority areas for new survey efforts (Fig. 2f). According to this analysis, the Amazon and the Cerrado of central-eastern Brazil appeared as the first sampling priorities. Given that Amazonia is the center of diversity of the tribe, it offers the best chance not only to expand the geographical knowledge of species ranges, but also to discover new species. Sites with poor quality surveys within the Amazon represent a second priority given the low completeness of sampling in these regions. Those findings corroborate the recommendations of other studies that have indicated the need to

focus sampling efforts in remote and under-sampled areas, while also revisiting accessible under-sampled areas and sampling highly threatened regions (Hopkins 2019; Stropp et al. 2020).

Conclusions

Well-curated distribution databases are crucial to address conservation issues and provide reliable answers to biogeographical questions. Obtaining raw point locality data is a demanding and costly task, although it is one with high returns. When compared to other kinds of distribution representation techniques such as range polygons, raw point locality data provide a better perspective of what is known and unknown about species geographical ranges (Rocchini et al. 2011; Maldonado et al. 2015; Guedes et al. 2018). The Bignonieae database is well curated, covering its taxonomic diversity, and presenting accurate geographical data, thus providing the most reliable source of species distribution data for these lianas to date. However, our analyses have shown that there is still room for improvement. Additional collection efforts are greatly needed across the Neotropics in order to encompass new localities and temporal scales. In particular, the curation and inclusion of Bignonieae specimens collected since 2004 will provide additional insights into the current status of the representation of Bignonieae specimens deposited in herbaria around the World. This information might help bridge the knowledge gap in the Amazon region, which remains substantial even though this region represents the main center of species diversity of Bignonieae. Further survey efforts will not only tackle the Wallacean shortfall in this group of lianas, but would certainly increase the rate of new species discoveries. The current deforestation pressures in the Amazon (Stropp et al. 2020) is threatening the diversity of Bignonieae, increasing the relevance of this region for conservation efforts. Accelerating the assembly of higher-quality distribution databases for multiple taxa in the Neotropics is urgently needed if we are to effectively design conservation plans for its most diverse regions.

Coding availability

Scripts used here are available on https://github.com/jupanago/RCode_BignonieaDatabase.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10531-021-02195-7.

Acknowledgements We thank all botanists who collected Bignonieae samples analyzed in this study, especially the late Al Gentry, who was the most prolific Bignoniaceae collector to date. We also thank the curators and technical staff of the 196 herbaria hosting the specimens surveyed here, especially the Missouri Botanical Garden (MO) where most samples are deposited; Dr. David Coomes for hosting JPNG in the Forest Ecology and Conservation Group at the Department of Plant Sciences of the University of Cambridge; Luiz Henrique Fonseca for providing us with 1600 curated herbarium records of the genus *Adenocalymma*; Bette Loiselle, Iván Jiménez, Seema Sheth, Trish Consiglio, Tibisay Escalona, and Leila Meyer for assistance with the compilation of earlier versions of the Bignonieae dataset; David Hawksworth and two anonymous reviewers for helpful comments that improved the manuscript. We also thank the following funding sources: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a Ph.D. fellowship to JPNG. (Finance Code 001); Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for a Thematic Project (Grant No. 2018/23899-2), and a collaborative FAPESP/NSF/NASA grant on the "Assembly

and evolution of the Amazonian biota and its environment" to LGL (201/50260-6); the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a Pq-1B grant to LGL (310871/2017-4); and, the Universidade Estadual do Maranhão for a Senior Researcher fellowship to TBG.

Author contributions JPNG, TBG, LGL planned the study. LGL led the compilation of the distribution dataset with input from JPNG. JPNG performed all analyses, prepared the figures and tables assisted by TBG and LGL. JPNG led the writing with contribution from all authors. All authors contributed to the interpretation and discussion of the results and approved the final version of this manuscript.

Funding This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior through a doctoral fellowship to JPNG (CAPES, Finance Code 001); the State University of Maranhão through a Senior Research fellowship to TBG; the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through a Thematic Project (2018/23899-2), and a collaborative FAPESP/NSF/NASA grant to LGL (Grant No. 2012/50260-6); and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) through a Pq-1B grant to LGL (Grant No. 310871/2017-4).

Availability of data and material Additional information supporting all results presented in this paper are available in the Supporting Information section.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Arbeláez-Cortés E (2013) Knowledge of Colombian biodiversity: published and indexed. Biodivers Conserv 22:2875–2906. https://doi.org/10.1007/s10531-013-0560-y

Barthlott W, Mutke J, Rafiqpoor D et al (2005) Global centers of vascular plant diversity. Nova Acta Leopoldina 92:61–83

Bojórquez-Tapia LA, Azurra I, Ezcurra E, Flores-Villela OA (1995) Identifying conservation priorities in Mexico through GIS and modeling. Ecol Appl 5:215–231. https://doi.org/10.2307/1942065

Chapman AD, Wieczorek J (2006) Guide to best practices for georeferencing. Global Biodiversity Information Facility, Copenhagen

Colombo B, Kaehler M, Calvente A (2016) An inventory of the Bignoniaceae from the Brazilian state of Rio Grande do Norte highlights the importance of small herbaria to biodiversity studies. Phytotaxa 278:19–28. https://doi.org/10.11646/phytotaxa.278.1.2

Daru BH, Park DS, Primack RB et al (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytol 217:939–955. https://doi.org/10.1111/nph.14855

Elith J, Graham CH, Anderson RP et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151. https://doi.org/10.1016/S0003-3472(80)80103-5

Elith J, Graham C, Valavi R et al (2020) Presence-only and presence-absence data for comparing species distribution modeling methods. Biodivers Inform 15:69–80. https://doi.org/10.17161/bi.v15i2.13384

Engemann K, Enquist BJ, Sandel B et al (2015) Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. Ecol Evol 5:807–820. https://doi.org/10.1002/ece3.1405

ESRI (Environmental Systems Research Institute) (2005) ArcGIS 9.1. ESRI, Redlands, California

Fonseca LHM, Lohmann LG (2019) An updated synopsis of *Adenocalymma* (Bignonieae, Bignoniaceae): new combinations, synonyms, and lectotypifications. Syst Bot 44:893–912. https://doi.org/10.1600/036364419x15710776741341

Fonseca LHM, Cabral SM, de Fatima Agra M, Lohmann LG (2017) Taxonomic revision of *Dolichandra* (Bignonieae, Bignoniaceae). Phytotaxa 301:1–70. https://doi.org/10.11646/phytotaxa.301.1.1

Francisco JNC, Lohmann LG (2018) Taxonomic revision of Pachyptera (Bignonieae, Bignoniaceae). PhytoKeys 131:89–131. https://doi.org/10.3897/phytokeys.92.20987

Frazão A, Lohmann LG (2019) An updated synopsis of *Tanaecium* (Bignonieae, Bignoniaceae). PhytoKeys 132:31–52. https://doi.org/10.3897/phytokeys.132.37538

Funk VA, Richardson KS, Ferrier S (2005) Survey-gap analysis in expeditionary research: where do we go from here? Biol J Linn Soc 85:549–567. https://doi.org/10.1111/j.1095-8312.2005.00520.x

Gentry AH (1979) Distribution patterns of neotropical Bignoniaceae: some phytogeographic implications. In: Larsen K, Holm-Nielsen L (eds) Tropical botany. Academic Press, London, pp 339–354
Gentry AH (1982) Neotropical floristic diversity: phytogeographical connections between Central and South America, pleistocene climatic fluctuations, or an accident of the andean orogeny? Ann Missouri Bot Gard 69:557–593. https://doi.org/10.2307/2399084
Guedes TB, Sawaya RJ, Zizka A et al (2018) Patterns, biases and prospects in the distribution and diversity of neotropical snakes. Glob Ecol Biogeogr 27:14–21. https://doi.org/10.1111/geb.12679
Guralnick RP, Hill AW, Lane M (2007) Towards a collaborative, global infrastructure for biodiversity assessment. Ecol Lett 10:663–672. https://doi.org/10.1111/j.1461-0248.2007.01063.x
Hopkins MJG (2007) Modelling the known and unknown plant biodiversity of the amazon basin. J Biogeogr 34:1400–1411. https://doi.org/10.1111/j.1365-2699.2007.01737.x
Hopkins MJG (2019) Are we close to knowing the plant diversity of the Amazon? An Acad Bras Cienc 91:1–7. https://doi.org/10.1590/0001-3765201920190396
Hortal J, de Bello F, Diniz-Filho JAF et al (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annu Rev Ecol Evol Syst 46:523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400
Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl 14:401–413. https://doi.org/10.1890/02-5364
Kaehler M, Michelangeli FA, Lohmann LG (2019) Fine tuning the circumscription of *Fridericia* (Bignonieae, Bignoniaceae). Taxon 68:751–770. https://doi.org/10.1002/tax.12121
Kataoka EY, Lohmann LG (2021) Taxonomic revision of *Martinella* Baill. (Bignonieae, Bignoniaceae). PhytoKeys 177:77–116. https://doi.org/10.3897/phytokeys.177.64465
Kier G, Mutke J, Dinerstein E et al (2005) Global patterns of plant diversity and floristic knowledge. J Biogeogr 32:1107–1116. https://doi.org/10.1111/j.1365-2699.2005.01272.x
La Sorte FA, Somveille M (2020) Survey completeness of a global citizen-science database of bird occurrence. Ecography 43:34–43. https://doi.org/10.1111/ecog.04632
Lobo JM, Hortal J, Yela JL et al (2018) KnowBR : an application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. Ecol Indic 91:241–248. https://doi.org/10.1016/j.ecolind.2018.03.077
Lohmann LG, Taylor CM (2014) A new generic classification of tribe Bignonieae (Bignoniaceae). Ann Missouri Bot Gard 99:348–489. https://doi.org/10.3417/2003187
Maldonado C, Molina CI, Zizka A et al (2015) Estimating species diversity and distribution in the era of big data: to what extent can we trust public databases? Glob Ecol Biogeogr 24:973–984. https://doi.org/10.1111/geb.12326
Medeiros MCMP, Lohmann LG (2015) Taxonomic revision of *Tynanthus* (Bignonieae, Bignoniaceae). Phytotaxa 216:1–60
Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6:8221. https://doi.org/10.1038/ncomms9221
Meyer C, Jetz W, Guralnick RP et al (2016a) Range geometry and socio-economics dominate species-level biases in occurrence information. Glob Ecol Biogeogr 25:1181–1193. https://doi.org/10.1111/geb.12483
Meyer C, Weigelt P, Kreft H, Lambers JHR (2016b) Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecol Lett 19:992–1006. https://doi.org/10.1111/ele.12624
Meyer L, Diniz-Filho JAF, Lohmann LG (2018) A comparison of hull methods for estimating species ranges and richness maps. Plant Ecol Divers 10:389–401. https://doi.org/10.1080/17550874.2018.1425505
Meyer L, Diniz-Filho JAF, Lohmann LG et al (2019) Canopy height explains species richness in the largest clade of neotropical lianas. Glob Ecol Biogeogr 29:26–37. https://doi.org/10.1111/geb.13004
Meyer L, Kissling WD, Diniz-filho JAF et al (2020) Deconstructing species richness–environment relationships in neotropical lianas. J Biogeogr. https://doi.org/10.1111/jbi.13924
Milliken W, Zappi D, Sasaki D et al (2010) Amazon vegetation: how much don't we know and how much does it matter? Kew Bull 65:691–709. https://doi.org/10.1007/s12225-010-9236-x
Morrone JJ (2014) Biogeographic regionalisation of the neotropical region. Zootaxa 3782:1–110. https://doi.org/10.11646/zootaxa.3782.1.1
Myers N, Mittermeier R, Mittermeier C et al (2000) Biodiversity hotspots for conservation priorities. Conserv Biol 403:853. https://doi.org/10.1038/35002501
Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. Nature 345:714–716. https://doi.org/10.1038/345714a0
Newbold T (2010) Applications and limitations of museum data for conservation and ecology with particular attention to species distribution models. Prog Phys Geogr 34:3–22. https://doi.org/10.1177/0309133309355630

Oliveira U, Paglia AP, Brescovit AD et al (2016) The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers Distrib 22:1232–1244. https://doi.org/10.1111/ddi.12489

Olson DM, Dinerstein E, Wikramanayake ED et al (2001) Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51:933–938. https://doi.org/10.1641/0006-3568(2001)051

Pautasso M, McKinney ML (2007) The botanist effect revisited: Plant species richness, county area, and human population size in the United States. Conserv Biol 21:1333–1340. https://doi.org/10.1111/j.1523-1739.2007.00760.x

Paz A, Brown JL, Cordeiro CLO et al (2021) Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic Forest. J Biogeogr. https://doi.org/10.1111/jbi.14083

Pelayo-Villamil P, Guisande C, Manjarrés-Hernández A et al (2018) Completeness of national freshwater fish species inventories around the world. Biodivers Conserv 27:3807–3817. https://doi.org/10.1007/s10531-018-1630-y

Peterson AT, Soberón J, Pearson RG et al (2011) Ecological niches and geographic distributions. Princeton University Press, New Jersey

Pimm SL, Joppa LN (2015) How many plant species are there, where are they, and at what rate are they going extinct? Ann Missouri Bot Gard 100:170–176. https://doi.org/10.3417/2012018

Prance GT, Beentje H, Dransfield J, Johns R (2000) The tropical flora remains undercollected. Ann Missouri Bot Gard 87:67. https://doi.org/10.2307/2666209

Proença CEB, Soares-Silva LH, Rivera VL, et al (2010) Regionalização, centros de endemismo e conservação com base em espécies de angiospermas indicadoras da biodiversidade do Cerrado brasileiro. In: Rezende Diniz I, Marinho Filho J, Bomfim Machado R, Brandão Cavalcanti R (eds) CERRADO: conhecimento científico quantitativo como subsídio para ações de conservação. pp 91–148

Pyke GH, Ehrlich PR (2010) Biological collections and ecological/environmental research: a review, some observations and a look to the future. Biol Rev 85:247–266. https://doi.org/10.1111/j.1469-185X.2009.00098.x

QGIS Development Team (2018) QGIS Geographic Information System. QGIS Association. http://qgis.org

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/

Raven PH, Gereau RE, Phillipson PB et al (2020) The distribution of biodiversity richness in the tropics. Sci Adv 6:eabc6228. https://doi.org/10.1126/sciadv.abc6228

Rocchini D, Lobo JM, Jime A et al (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Prog Phys Geogr 32:211–226. https://doi.org/10.1177/0309133311399491

Rodríguez LO, Young KR (2009) Biological diversity of Peru: determining priority areas for conservation. AMBIO A J Hum Environ 29:329–337. https://doi.org/10.1579/0044-7447-29.6.329

Santos JC, Leal IR, Almeida-Cortez JS et al (2011) Caatinga: The scientific negligence experienced by a dry tropical forest. Trop Conserv Sci 4:276–286. https://doi.org/10.1177/194008291100400306

Sheth SN, Lohmann LG, Consiglio T, Jiménez I (2008) Effects of detectability on estimates of geographic range size in Bignonieae. Conserv Biol 22:200–211. https://doi.org/10.1111/j.1523-1739.2007.00858.x

Sheth SN, Lohmann LG, Distler T, Jiménez I (2012) Understanding bias in geographic range size estimates. Glob Ecol Biogeogr 21:732–742. https://doi.org/10.1111/j.1466-8238.2011.00716.x

Simon MF, Proença C (2000) Phytogeographic patterns of mimosa (Mimosoideae, Leguminosae) in the cerrado biome of Brazil: An indicator genus of high-altitude centers of endemism? Biol Conserv 96:279–296. https://doi.org/10.1016/S0006-3207(00)00085-9

Soberón J, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. Philos Trans R Soc London B 359:689–698. https://doi.org/10.1098/rstb.2003.1439

Soberón J, Peterson AT (2009) Monitoring biodiversity loss with primary species-occurrence data: toward national-level indicators for the 2010 target of the convention on biological diversity. Ambio 38:29–34. https://doi.org/10.1579/0044-7447-38.1.29

Sousa-Baena MS, Garcia LC, Peterson AT (2014) Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Divers Distrib 20:369–381. https://doi.org/10.1111/ddi.12136

Stropp J, Umbelino B, Correia RA et al (2020) The ghosts of forests past and future: Deforestation and botanical sampling in the Brazilian Amazon. Ecography 43:1–11. https://doi.org/10.1111/ecog.05026

Tobler M, Honorio E, Janovec J, Reynel C (2007) Implications of collection patterns of botanical specimens on their usefulness for conservation planning: An example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. Biodivers Conserv 16:659–677. https://doi.org/10.1007/s10531-005-3373-9

Töpel M, Zizka A, Calió MF et al (2017) SpeciesGeoCoder: fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. Syst Biol 66:145–151. https://doi.org/10.1093/sysbio/syw064

U.S. Geological Survey, EROS Data Center Distributed Active Archive Center (EDC DAAC) (2004) Global Digital ElevationModel (GTOPO30). https://doi.org/10.5066/F7DF6PQS

Ugland KI, Gray JS, Ellingsen KE (2003) The species-accumulation curve and estimation of species richness. J Anim Ecol 72:888–897. https://doi.org/10.1046/j.1365-2656.2003.00748.x

Ulloa Ulloa C, Acevedo-Rodríguez P, Beck S et al (2017) An integrated assessment of the vascular plant species of the Americas. Science 358:1614–1617. https://doi.org/10.1126/science.aao0398

Vale MM, Jenkins CN (2012) Across-taxa incongruence in patterns of collecting bias. J Biogeogr 39:1743–1744. https://doi.org/10.1111/j.1365-2699.2012.02759.x

Wiens JA, Seavy NE, Jongsomjit D (2011) Protected areas in climate space: what will the future bring? Biol Conserv 144:2119–2125. https://doi.org/10.1016/j.biocon.2011.05.002

Yost JM, Sweeney PW, Gilbert E et al (2018) Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants . Appl Plant Sci 6:1–11. https://doi.org/10.1002/aps3.1022

Zizka A, Antonelli A, Silvestro D (2020) Sampbias, a method for quantifying geographic sampling biases in species distribution data. Ecography. https://doi.org/10.1101/2020.01.13.903757

Zuntini AR, Taylor CM, Lohmann LG (2015a) Deciphering the neotropical *Bignonia binata* species complex (Bignoniaceae). Phytotaxa. 219:69–77. https://doi.org/10.11646/phytotaxa.219.1.5

Zuntini AR, Taylor CM, Lohmann LG (2015b) Problematic specimens turn out to be two undescribed species of *Bignonia* (Bignoniaceae). PhytoKeys 56:7–18. https://doi.org/10.3897/phytokeys.56.5423

Authors and Affiliations

Juan Pablo Narváez-Gómez[1] · Thaís B. Guedes[2,3] · Lúcia G. Lohmann[1]

[1] Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua Do Matão, 277, São Paulo, SP 05508-090, Brazil

[2] Centro de Estudos Superiores de Caxias, Universidade Estadual Do Maranhão, Praça Duque de Caxias, s/n, Caxias, MA 65604-380, Brazil

[3] Gothenburg Global Biodiversity Center, Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden