

# ECOGRAPHY

## Research

### The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon

J. Stropp, B. Umbelino, R. A. Correia, J. V. Campos-Silva, R. J. Ladle and A. C. M. Malhado

J. Stropp (<https://orcid.org/0000-0002-2831-4066>) ✉ ([juliana.stropp@gmail.com](mailto:juliana.stropp@gmail.com)), B. Umbelino (<https://orcid.org/0000-0002-7602-8400>), R. A. Correia (<https://orcid.org/0000-0001-7359-9091>), J. V. Campos-Silva (<https://orcid.org/0000-0003-4998-7216>), R. J. Ladle (<https://orcid.org/0000-0003-3200-3946>) and A. C. M. Malhado (<https://orcid.org/0000-0003-3621-779X>), *Inst. of Biological Sciences and Health, Univ. Federal de Alagoas, s/n Tabuleiro dos Martins, Maceió, AL, Brazil. Present address of JS: Dept of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain, and Dept of Computer Science, Univ. of Luxembourg, Esch-sur-Alzette, Luxembourg. Present address of RAC: Helsinki Lab of Interdisciplinary Conservation Science, Dept of Geosciences and Geography, Univ. of Helsinki, Finland, and Helsinki Inst. of Sustainability Science (HELSUS), Univ. of Helsinki, Finland. RAC also at: DBIO and CESAM-Centre for Environmental and Marine Studies, Univ. of Aveiro, Aveiro, Portugal. JVC-S also at: Faculty of Environmental Sciences and Natural Resource Management, Norwegian Univ. of Life Sciences, Ås, Norway.*

#### Ecography

43: 979–989, 2020

doi: 10.1111/ecog.05026

Subject Editor: Nathalie Butt  
Editor-in-Chief: Miguel Araújo  
Accepted 18 February 2020



The remarkable biodiversity of the Brazilian Amazon is poorly documented and threatened by deforestation. When undocumented areas become deforested, in addition to losing the fauna and flora, we lose the opportunity to know which unique species had occupied a habitat. Here we quantify such knowledge loss by calculating how much of the Brazilian Amazon has been deforested and will likely be deforested until 2050 without having its tree flora sufficiently documented. To this end, we analysed 399 147 digital specimens of nearly 6000 tree species in relation to official deforestation statistics and future deforestation scenarios. We find that by 2017, 30% of all the localities where tree specimens had been collected were mostly deforested. Some 300 000 km<sup>2</sup> (12%; 485 25 × 25 km grid cells) of the Brazilian Amazon had been deforested by 2017, without having a single tree specimen recorded. An additional 250 000–900 000 km<sup>2</sup> of severely under-collected rainforest will likely become deforested by 2050. If future tree sampling is to cover this area, sampling effort has to increase two- to six-fold. Nearly 255 000 km<sup>2</sup> or 7% of rainforest in the Brazilian Amazon is easily accessible but does yet but remain under-collected. Our study highlights how progressing deforestation increases the risk of losing undocumented species of a hyper-diverse tree flora.

Keywords: Amazonian trees, Brazilian Amazon, collection effort, deforestation, GBIF, open-access biodiversity databases

#### Introduction

The Amazon basin harbours one of the most diverse terrestrial floras and faunas on Earth (Gentry 1992). Its species richness and relevance for the global climate (Nobre et al. 2016) have elevated the region to a cause célèbre of the global conservation movement (Dinerstein et al. 2019). Nevertheless, habitat loss and deforestation continue to threaten forests throughout the basin (Fearnside 2015). After several years



[www.ecography.org](http://www.ecography.org)

© 2020 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

of stability, deforestation – the clear cut of mature forest – is increasing again, with an estimated 9.762 km<sup>2</sup> being clear-cut only in the Brazilian Amazon between August 2018 and July 2019 (Barlow et al. 2019, INPE 2019).

Widespread deforestation will likely lead to massive species extinctions. It remains unknown, however, how many and which species will be most affected (Grelle 2005, Hubbell et al. 2008, Crooks et al. 2017). Knowledge shortfalls arise in part because of incomplete species inventories and taxonomic descriptions of species (dos Santos et al. 2015), which are most severe for small, poorly studied taxa (Hortal et al. 2015), but also exist for relatively well studied groups, such as trees (Hopkins 2007, Feeley 2015). The inventory of Amazonian trees (currently ranging from ~7000 to ~10 000 species (Cardoso et al. 2017, ter Steege et al. 2019)) is far from complete and as many as 5000 new tree species may still be undiscovered (ter Steege et al. 2016).

As inventories of Amazonian trees remain limited to a few well-studied taxa and well-surveyed areas (Nelson et al. 1990, Schulman et al. 2007, Feeley 2015), deforestation may lead to the extinction of unknown tree species. The risk of losing poorly documented species is aggravated by decreasing research budgets for biological surveys (Fernandes et al. 2017, Magnusson et al. 2018). In times of increasing deforestation and decreasing exploration, it is therefore important to know: 1) how much of the Amazonian rainforest has been deforested and will likely be deforested in the future without having its tree flora documented? 2) Whether, and by how much, botanical sampling needs to advance to ensure the Amazonian tree flora is well-documented before it eventually becomes deforested? To address these questions, we compiled data on the occurrence of nearly 6000 tree species. Species occur in various regions across the Amazon basin, differ in the degree to which they are threatened with extinction and in the intensity with which they have been sampled.

## Material and methods

### Study area

We focus on the Brazilian Amazon, a region that covers approximately 60% of the Amazon rainforest and for which detailed deforestation statistics have been published since 1988 (Kintisch 2007, INPE 2019). The Brazilian Amazon also has the highest number of tree collections among all South American countries (ter Steege et al. 2016).

### Selection of tree species

We compiled a list of known tree species in the entire Amazon using three checklists: ter Steege et al. (2016), Beech et al. (2017) and Cardoso et al. (2017). ter Steege et al. (2016) collated the first checklist, which includes the names of 11 676 tree species collected between 1707 and 2015. Cardoso et al. (2017) revised the checklist of ter Steege et al (2016) and

proposed a new list of 6727 species. Differences between the two checklists arise mainly from different baseline data or source used to verify species names, the applied geographical boundary of Amazonia and the definition of what constitutes a ‘tree’. Our third checklist is derived from Beech et al. (2017), who provided a list of 60 065 names of tree species recorded worldwide.

From these three lists, we established a broadly accepted checklist of all tree species so far recorded in the Brazilian Amazon. We considered the names of all tree species as valid if they were present in at least two of the three checklists described above. We established our checklist by querying all tree names presented by ter Steege et al. (2016) or Beech et al. (2017) that were flagged as ‘no.error’ in the revised checklist presented by Cardoso et al. (2017). The query resulted in 9667 species, which we then checked for repetitions (due to spelling and synonyms) by submitting them to the Taxonomic Name Resolution Service – TNRS (Boyle et al. 2013) in May 2018. We only retained species names for which TNRS rendered: 1) the taxonomic status as ‘accepted name’ and ‘no opinion’; and 2) an overall match score of > 0.9. Match scores range from zero to one, where one indicates a complete match between the name to be checked and a taxon name in the TNRS database. A score of zero indicates no match. After this taxonomic standardisation, our final checklist contained 9506 valid names of tree species that occur in the Brazilian Amazon.

### Dataset: collection of herbarium specimens of Amazonian trees

We retrieved occurrence data for the 9506 tree species from three major open-access biodiversity databases: the Botanical Information and Ecological Network (BIEN; Enquist et al. 2009), the Global Biodiversity Information Facility (GBIF; <www.gbif.org>), and SpeciesLink (<www.splink.cria.org.br>). The BIEN database (ver. 4.0.1; release date 14 May 2018) was queried using the valid names of the 9506 tree species. The GBIF and SpeciesLink databases were queried using the phyla Magnoliopsida and Liliopsida, as these encompass most of the Amazonian tree species. We then selected from the occurrence data retrieved from GBIF and SpeciesLink only those entries that contained a valid tree name according to our checklist.

After retrieving occurrence data from the three databases, we retained only those that: 1) referred to preserved specimens; 2) listed the geographical coordinates of the location where the respective specimen was collected; and 3) contained the date of collection. We then filtered the selected entries and kept only those with geographical coordinates located within the boundaries of the Brazilian Legal Amazon according to the Brazilian Institute of Geography and Statistics (IBGE; IBGE.gov.br). Finally, we combined the entries retrieved from the three databases into a single dataset. Data were retrieved with R (R Core Team) from BIEN with the ‘BIEN’ package (Maitner et al. 2018) and from the GBIF database with the ‘rgbif’ package (Chamberlain 2017, GBIF 2018a, b), whereas data from SpeciesLink were downloaded

online from SpeciesLink (2018; complete citation is given in the Supplementary material Appendix 1). Our combined dataset contained 399 147 herbarium specimens of 7383 tree species. These entries comprise species identity, the location and the date of collection of herbarium specimens that are available through open-access biodiversity databases. For simplicity, we refer to these entries as herbarium specimens.

We screened this dataset and flagged duplicated specimens and those holding uncertain geographical coordinates and/or a missing or uncertain date of collection. This filtering led to a dataset containing 129 252 specimens of 5750 tree species (see Supplementary material Appendix 1 for details on data filtering).

We added to this dataset information on the area of occupancy and the conservation status of individual species. Information on the area of occupancy was extracted from Gomes et al. (2019) and indicates in which among six Amazonian regions a species occurs: eastern (EA), southern (SA), southwestern (WAS), central (CA), north-western Amazonia (WAN) and the Guiana Shield (GS). Gomes et al. (2019) modelled the area of occupancy for individual species based on the suitability of environmental conditions by applying species distribution models. We assigned to each region in which a given species occurs a binary value representing high or low vulnerability to deforestation. The eastern, southern and south-western regions of the Brazilian Amazon were considered vulnerable to deforestation, whereas central, north-western Amazonia and the Guiana Shield were considered less vulnerable to deforestation. Information on conservation status was obtained from ter Steege et al. (2015), who adhered to the International Union for Conservation of Nature (IUCN) Red List criteria and classify species as vulnerable (VU) or endangered (EN) based on their estimated population density and historical deforestation rates as of 2013. Information on both area of occupancy and conservation status was retrieved for 3469 tree species.

### Data on historical and future deforestation

Spatial layers of cumulative historical deforestation up to 2017 were retrieved from the Amazon Deforestation Estimation Project, PRODES ('Projeto de Estimativa do Desflorestamento da Amazônia'). PRODES provides official statistics on deforestation in the Brazilian Amazon dating back to 1988 at a resolution of 120 × 120 m (Camara et al. 2013). The spatial layers of future deforestation, modelled for the year 2050 at a resolution of 1 × 1 km, were retrieved from the 'Distributed Active Archive Center for Biogeochemical Dynamics (DAAC)' (Soares-Filho et al. 2006). We used two scenarios: 1) business-as-usual deforestation (BAU), and 2) controlled deforestation under improved governance (IG). The business-as-usual scenario estimates that by 2050 a total of 1 500 000 km<sup>2</sup> of the Brazilian Amazon will be deforested; whereas the improved governance scenario estimates that roughly 500 000 km<sup>2</sup> will be deforested (Fig. S6 in Soares-Filho et al. 2006). Spatial layers of historical and future deforestation contain four land-cover classes: deforested, forest,

non-forest and no data. Deforestation was not computed for areas classified as non-forest, such as the seasonally dry and white-sand forests (Camara et al. 2013).

Land-cover data were analysed at two spatial resolutions: 1) one-kilometre around each collection locality and 2) a regular 25 × 25-km grid, the latter covering the entire Brazilian Amazon. We determined the land-cover around each collection locality by assigning it with the land-cover class covering more than 50% of a 1-km buffer surrounding each locality. Similarly, we determined the land-cover at 25 km resolution, by assigning the land-cover class that occupied more than 50% of a grid cell. To assign future deforestation to each collection locality, we extracted the land-cover class corresponding to the single 1 km pixel that overlaps with the geographic coordinates of a collection locality. To assign information on future deforestation to each grid cell at 25 × 25 km resolution, we attributed to each cell the future land-cover class that occupied more than 50% of a grid cell. All spatial layers were manipulated using 'South America Albers Equal Area Conic' projection, using the R package 'rgdal' (Bivand et al. 2018).

### Data on land protection and accessibility

Spatial layers of protected areas and indigenous land were retrieved from IBGE and converted to a raster format at a spatial resolution of 1 km. Accessibility, measured as travel time (in hours) to the nearest city with more than 50 000 inhabitants as of the year 2000, was obtained at a 1 km resolution from Nelson (2008). Information on land protection was then analysed at the same resolution of 1 km around each collection locality and 25 × 25 km grid covering the Brazilian Amazon. For obtaining land protection around each collection locality, we extracted a binary class of land protection (protected or unprotected) corresponding to the single 1 km pixel that overlapped with a collection locality. We attributed to each 25 × 25 km grid cell the binary class of land protection depending on whether more than 50% of a grid cell was classified as protected area (including indigenous territories). A mean accessibility value expressed in travel time to the nearest (h) major city was attributed to each grid cell.

### Analysis

#### Historic deforestation and tree sampling

Descriptive statistics highlighted that the deforestation affects collection localities of species to a varying degree. We used this information to identify groups of species whose collection localities were subject to a similar degree of deforestation, occur in the same region and share the same conservation status. Next, we assessed whether collection effort varied among the groups of species. To this end, we used Factor Analysis for Mixed Data (FAMD) and a hierarchical clustering of its principal components (HCPC). FAMD and HCPC are ordination techniques to visualize data points with similar values of continuous and categorical variables (Kassambra 2017). For the FAMD, we include for each species the proportion of collection localities located in protected areas and deforested

areas, as continuous variables. As categorical variables, we used the binary value representing the region's vulnerability to deforestation and the conservation status of individual species. We then tested with a Wilcoxon rank-sum test whether the collection effort varies among groups of species.

We also explored how deforestation and collection effort varied across the Brazilian Amazon by computing descriptive statistics for land-cover classes and the number of specimens and species in each 25×25-km grid cell. Moreover, we calculated inventory completeness of tree species for grid cells with at least 100 specimens. For this, we considered the cumulative number of specimens and species collected from 1900 until 2017. We quantified inventory completeness by estimating the final slope of smoothed species accumulation curves (Lobo et al. 2018). This metric indicates the rate at which sampling of specimens yields new species added to the dataset (Hortal et al. 2008).

After estimating the slope, we calculated the complementary slope value (i.e. 1-slope) to represent inventory completeness; values approaching one indicate a high completeness. We defined better-sampled cells as those with at least 100 specimens and an inventory completeness of  $\geq 0.5$ . We tested whether these relatively better-sampled cells were spatially clustered by applying a Monte Carlo test on homogeneous point pattern. Finally, we produced a grid of distance to better-sampled cells by attributing to each 25×25-km cell the geographical distance (km) between a given cell and the closest better-sampled cell.

#### **Future deforestation and tree sampling**

We calculated the total number of grid cells that are predicted to be deforested by 2050 and assessed the extent to which future deforestation will affect both better- and poorly-sampled cells.

Next, we estimated the necessary increase in the number of cells with at least 100 specimens in order to ensure that future tree sampling covers an area equivalent to the size of area that is predicted to be deforested by 2050, but until 2017 did not contain a single specimen with complete labelling in open-access biodiversity databases. We fitted linear models with explanatory variables representing the hypothesized increase in the number of grid cells with at least 100 specimens necessary to add up to 1040 or 400 cells by 2050. These numbers represent the total number of cells predicted to have more than 50% of their area deforested under the business-as-usual or improved governance scenario, respectively, but remained poorly collected by 2017.

#### **Identifying opportunities for botanical sampling**

We identified areas of the Brazilian Amazon that are vulnerable to deforestation but still offer opportunities to document poorly collected species. First, we overlaid the grid of distances to better-sampled cells with the grid of travel time to the nearest major city. The overlap of these two grids resulted in a third 25×25-km grid, in which each cell was assigned a value corresponding to one of nine distance categories, representing the distance to better-sampled cells and

the travel time to the next major city. Second, we computed the area (km<sup>2</sup>) of all cells that are assigned as forest and that were above the 50-percentile of distances to better-sampled cells and below the 25-percentile of travel times to the nearest major city. Such cells are easily accessible and therefore vulnerable to deforestation but at the same time provide opportunities to document new species for the following reasons: First, locations that are far from better-sampled cells are more likely to hold unknown flora (Ladle and Hortal 2013) and thus are more likely to yield new records of species occurrences; second, locations that are easily accessed, i.e. those near roads, cities or navigable rivers may also be preferred for botanical surveying.

## **Results**

### **Herbarium specimens of Amazonian trees: an overview**

The data retrieved in May 2018 from BIEN, GBIF and SpeciesLink contained 399 147 specimens of 7383 tree species collected within the Brazilian Amazon. Data filtering led to an exclusion of duplicated specimens (43% of the initial records), specimens with a missing or uncertain date of collection (32% of the initial records) and specimens with erroneous geographic coordinates (13% of the initial records; Table 1). As individual specimens can be associated with multiple errors, data filtering led to the exclusion of 68% of the initial records. The filtered dataset contained 129 252 specimens, comprising 5750 tree species from 119 families.

The ten most frequently collected families in descending order of number of specimens were Fabaceae, Rubiaceae, Melastomataceae, Myrtaceae, Annonaceae, Euphorbiaceae, Chrysobalanaceae, Lecythidaceae, Lauraceae and Burseraceae. These families contain half of all collected specimens. The ten most frequently collected species were *Myrcia splendens* (804 specimens), *Tapirira guianensis* (428), *Siparuna guianensis* (427), *Licania heteromorpha* (346), *Byrsonima crassifolia* (340), *Myrcia guianensis* (339), *Protium heptaphyllum* (311), *Casearia javitensis* (307), *Virola sebifera* (303) and *Xylopia aromatica* (302).

### **Historic deforestation and tree sampling**

Thirty percent of all collection localities were deforested by 2017 (n=38 944), i.e. more than half of the internal area of their respective 1-km buffer had been deforested. Most of these were in unprotected areas (Supplementary material Appendix 1 Fig. A1). Whereas all collection localities of 264 individual species had been completely deforested, all collection localities of 1764 species were still covered by forest in 2017 (Table 2).

Factor analysis for mixed data (FAMD) and hierarchical clustering of its principal components (HCPC) revealed that species sharing the same region and conservation status tend to be subject to a similar amount of deforestation in their collection localities. The first two FAMD axes explained 54%

Table 1. Percentage and absolute number of specimens excluded through data filtering. As individual specimens can suffer from multiple errors, the sum of percentages does not equal the total 68% of excluded specimens.

Criteria of data filtering	Percentage of specimens flagged in the data filtering (absolute number)
Imprecise coordinates: decimals of latitude and longitude contained only zeros	6% (28 845)
Coordinates near administrative locations of cities or villages	6% (25 052)
Uncertain country: the information provided in the field 'country' did not refer to Brazil	0.5% (2338)
Uncertain date: records collected between 1600 and 1899 and after the date of data download	16% (63 098)
Missing date	16% (62 379)
Duplicates: identical species name, geographical coordinates, and date of collection	43% (171 320)
Total number of specimens retained after data filtering	33% (129 252)

of the variation and revealed two gradients. The first FAMD axis represents the current forest-cover of collection localities. The left side of this axis shows species with a large percentage of specimens collected in protected areas whereas the right side shows species with a high percentage of specimens collected in deforested localities (Fig. 1a, Supplementary material Appendix 1 Table A1). The second FAMD axis shows species grouped into four clusters according to their region and conservation status (Fig. 1a, Supplementary material Appendix 1 Fig. A2). Collection effort varied significantly across the four clusters (Kruskal–Wallis chi-squared = 408.7;  $p < 0.005$ ,  $df = 3$ ). Species that occur in the Guiana shield, central and north-western Brazilian Amazon are represented by a significantly lower number of specimens (median = 12; median absolute deviation = 13.3) than species that occur in the eastern, southern and south-western Brazilian Amazon (median = 25.5; median absolute deviation = 27.4) (Fig. 1b, Supplementary material Appendix 1 Table A2).

Overall, we find that vast areas of the Brazilian Amazon remain under-collected. About half of the region (~2.6 million km<sup>2</sup>; 4125 of 7691 grid cells of 25 × 25 km) does not have a single specimen labelled with complete information in open-access biodiversity databases. Only 224 grid cells, 3% of all cells, contained 100 or more specimens. Only such cells allow us to estimate inventory completeness. Values of inventory completeness for these 224 grid cells ranged from 0.03 to 0.97 (median = 0.52; median absolute deviation = 0.2; Supplementary material Appendix 1 Fig. A3). For 120 grid cells, corresponding to 1.5% of the Brazilian Amazon, tree flora is relatively well documented (N specimens ≥ 100 and inventory completeness ≥ 0.5). Thirty seven percent of these better-sampled cells have been deforested by 2017. Past deforestation across poorly sampled areas caused loss of approximately 300 000 km<sup>2</sup> or 12% of rainforest (485 grid

cells), which not a single specimen with complete labelling had been recorded.

Better-sampled cells are spatially clustered (pseudo- $p = 0.001$ ; Monte Carlo simulation on homogeneous point pattern; Fig. 2a). The longest distance of any cell to such better-sampled cells is observed in northern Amapá and southwestern Amazonas, where cells can be located up to 430 km from the nearest better-sampled grid cell (AP and AM in Fig. 2b) (median distance to a better-sampled cell in the Brazilian Amazon = 106 km; median absolute deviation = 78 km).

### Future deforestation and tree sampling

If deforestation were to follow the business-as-usual scenario of Soares-Filho et al. (2006), up to 47% of the ~2.6 million km<sup>2</sup> (1939 grid cells of 25 × 25 km) of the Brazilian Amazon that remained severely under-collected until 2017 will have been deforested by 2050 (i.e. > 50% of internal area classified as deforested; Fig. 3). Approximately 900 000 km<sup>2</sup> (1407 grid cells) of rainforest that are predicted to become deforested by 2050 under a business-as-usual scenario, remain severely under-collected to date. Under the scenario of improved governance, instead of 900 000 km<sup>2</sup> only 250 000 km<sup>2</sup> (400 grid cells) of severely under-collected grid cells will be deforested by 2050.

Our data indicate that documenting tree flora in severely under-collected areas before they become deforested will require a tremendous increase in sampling effort. For example, the sampling of tree specimens between 1960 and 2017 yielded on average four new grid cells per year for which 100 or more specimens had been collected (Fig. 4a–b). This rate is insufficient to prevent that rainforest in the Brazilian Amazon is deforested before its tree flora has been documented. As our linear model indicates, the number of cells with ≥ 100

Table 2. Percentage and absolute number of collection localities that were deforested by 2017 according to the land-cover classification of PRODES (Camara et al. 2013) and that may be deforested in 2050 assuming business-as-usual and improved governance following Soares-Filho et al. (2013). Deforestation is not quantified for areas classified as 'non-forest' or for which no data are available in PRODES. The total number of collection localities is 129 252.

Land-cover class	Deforestation recorded by 2017	Deforestation under business-as-usual scenario until 2050	Deforestation under improved governance scenario until 2050
Deforested	30% (38 944)	64% (83 357)	51% (65 835)
Forested	47% (60 463)	15% (19 283)	28% (36 723)
Non-forest	18% (23 359)	20% (24 729)	20% (24 811)
No data	5% (6486)	1% (1883)	1% (1883)

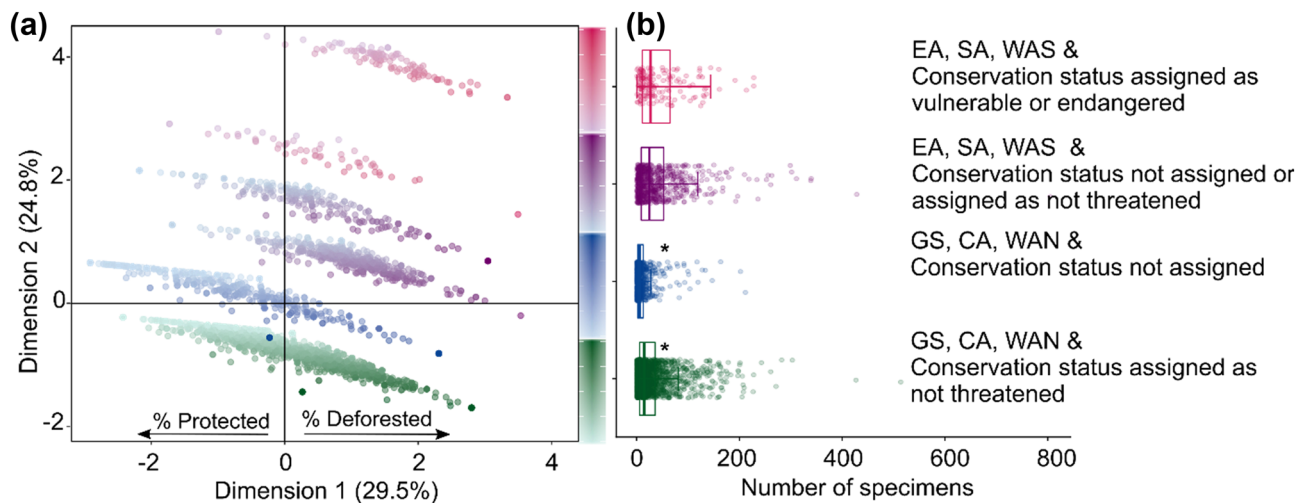


Figure 1. (a) Scatterplot of the factor analysis for mixed data (FAMD) and the hierarchical clustering on principal components (HCPC). Each dot represents a tree species, with lighter shades indicating a higher proportion of specimens collected in protected areas and darker shades indicating a higher proportion of specimens collected in deforested areas. Colours represent the four clusters identified by the FAMD analysis and indicate species region and conservation status. Pink dots indicate species that occur in the eastern, southern and southwestern portions of the Brazilian Amazon (i.e. the arc of deforestation) and are classified as either vulnerable to, or endangered by, extinction. Purple dots indicate species occurring in the eastern, southern and southwestern portions of the Brazilian Amazon; but are not threatened with extinction or have not been assigned a conservation status. Dark-blue dots indicate species that occur in the Guiana shield, central and north-western Brazilian Amazon and did not have their conservation status assigned. Green dots indicate species that occur in the Guiana shield, central and the north-western Brazilian Amazon and have their conservation status assigned as not threatened. Note that collection localities of species that occur in the eastern (EA), southern (SA) and southwestern (WAS) Brazilian Amazon tend to be subject to higher deforestation. (b) Boxplot of the median number of specimens per species grouped into four clusters.

specimens recorded each year must rise from four to 26 ( $y = -51689.9 + 25.7 \times x$ ) to ensure that 1407 cells become relatively better-sampled before 2050 (Fig. 4c). These 1407 cells represent the total number of cells that is predicted to be deforested in the BAU scenario.

If protected areas and indigenous territories remain protected from deforestation under improved governance, deforestation may not affect 1407 but only 400 under sampled cells. To ensure that 400 new cells will be sampled before 2050, the number of cells with  $\geq 100$  specimens recorded each year has to increase only from four to six ( $y = -10988.1 + 5.5 \times x$ ) (Fig. 4c).

### Opportunities for botanical sampling

When considering accessibility to the forest (measured as travel time to the nearest major city), it appears that easily accessible areas can be in close proximity or far away, from a better-sampled cell (Fig. 2c–d). We found that 219 375 km<sup>2</sup> of the Brazilian Amazon, corresponding to 351 cells, 59% of the 8% of cells marked in green in Fig. 2c are: 1) located in close proximity to a major city but far away from a better-sampled cell; 2) are not represented by a single specimen and 3) are still covered by forest. Such areas are attractive for future botanical explorations because they can be easily reached and are likely to yield new records of occurrence or even the discovery of new species. Yet, such cells are also vulnerable to deforestation as land conversion rates are highest around major cities, roads and rivers.

### Discussion

In an ideal world we would have accurate, up-to-date and detailed information about the fauna and flora of the Amazon basin, including data on historical species occurrence for areas that were subsequently deforested. Such knowledge has immense value for biogeographical analysis (Cox et al. 2016) and can provide a baseline for conservation and restoration initiatives (Gillson et al. 2011). In reality, missing or incomplete data about historic species occurrence in deforested areas presents a problem because this type of information is irretrievable and can only be approximated.

We first quantified how much of the Brazilian Amazon has been deforested and is likely to be deforested in the future without having its tree flora well documented. Our results show that by 2017, 485 grid cells of 25 × 25 km, or roughly 300 000 km<sup>2</sup> (12%) of the Brazilian Amazon, had at least half of their internal area cleared by deforestation without having recorded a single tree specimen with complete label. The loss of poorly-documented forest was concentrated in the south, eastern and southwestern Brazilian Amazon, where deforestation rates have been historically highest.

Our results show that by 2017, 30% of all historic collection localities were deforested, i.e. more than half of the internal areas of their respective 1-km buffer were deforested. This means that herbarium specimens poorly represent the current species occurrences in these localities (Ladle and Hortal 2013, Tessarolo et al. 2017). Such losses highlight the

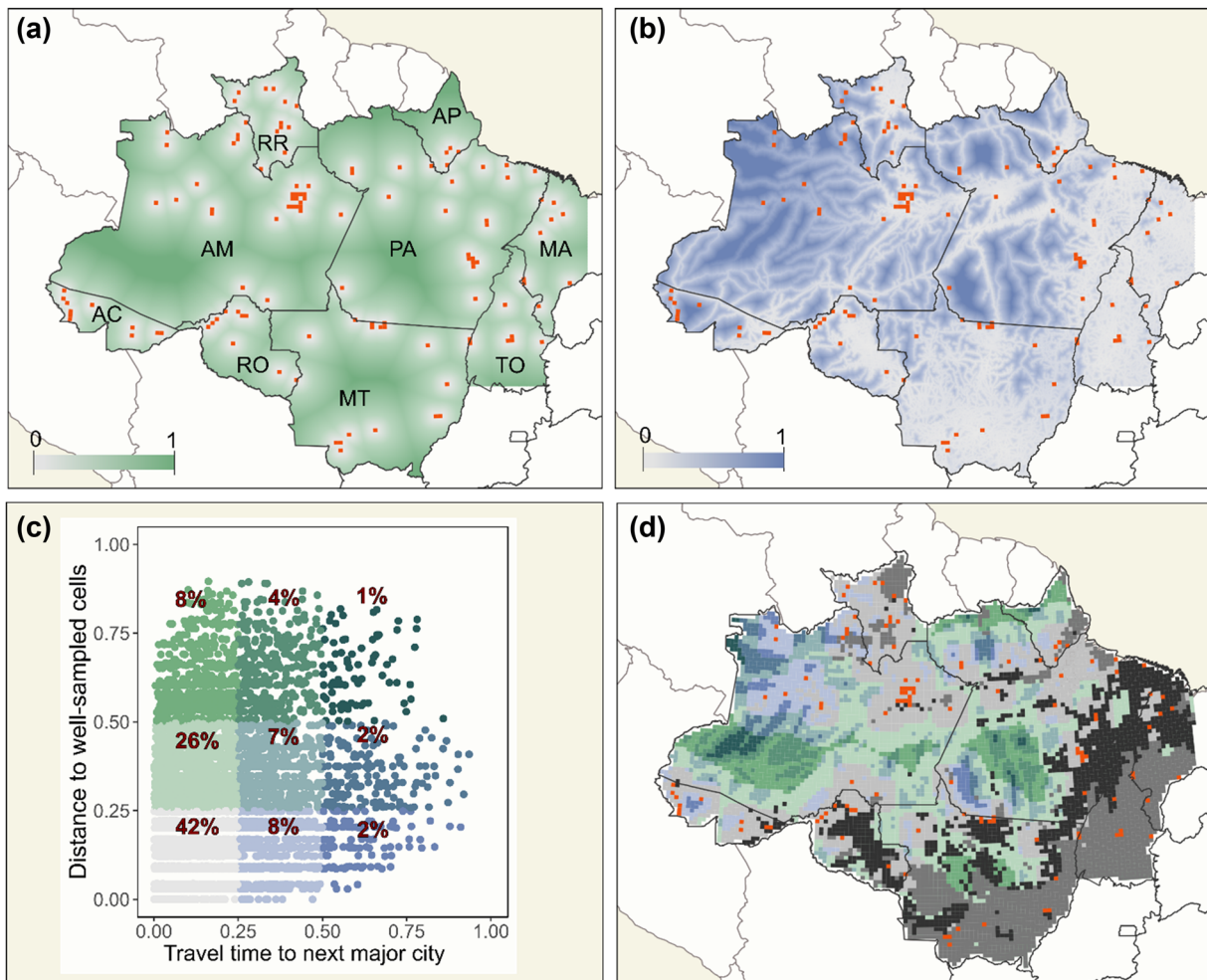


Figure 2. (a) Map of distance from any grid cell in the Brazilian Amazon to a better-sampled cell (orange squares). (b) Map of travel time to a city with more than 50 000 inhabitants in 2000; in both maps values were scaled to zero and one with zero indicating the shortest distance to a better-sampled cell (a) and the longest travel time to the next major city (b). (c) Scatterplot of distance to better-sampled cells and travel time to next major city; numbers indicate the percentage of cells in each of nine categories (indicated by unique colour). (d) Spatial overlap between distance to better-sampled cells and travel time to next major city; deeper shades of green indicate regions far from better-sampled cells, whereas deeper shades of blue indicate remote areas; black and dark-grey cells represent deforested and non-forested areas according to PRODES, respectively. Better-sampled grid cells are considered as those containing at least 100 specimens and showing an inventory completeness  $\geq 0.5$ .

Grid cells (N; %)	Land-cover 2017 Prodes		Land-cover 2050 Business-as-usual		Land-cover 2050 Governance	
	Deforested (%)	Forest (%)	Deforested (%)	Forest (%)	Deforested (%)	Forest (%)
Without specimen (4,125; 54)	12	75	47	41	21	66
1 to 100 specimens (3,342; 43)	18	55	52	23	33	42
$\geq 100$ specimens (224; 3)	20	59	66	19	46	37
Well-sampled (120; 2)	37	59	63	15	47	30

Figure 3. Summary of land-cover and sampling of tree specimens in grid cells of  $25 \times 25$  km. Sums of percentages of cells classified as deforested and forested do not add up to 100% because cells may be classified as 'non-forest' or 'no-data'.

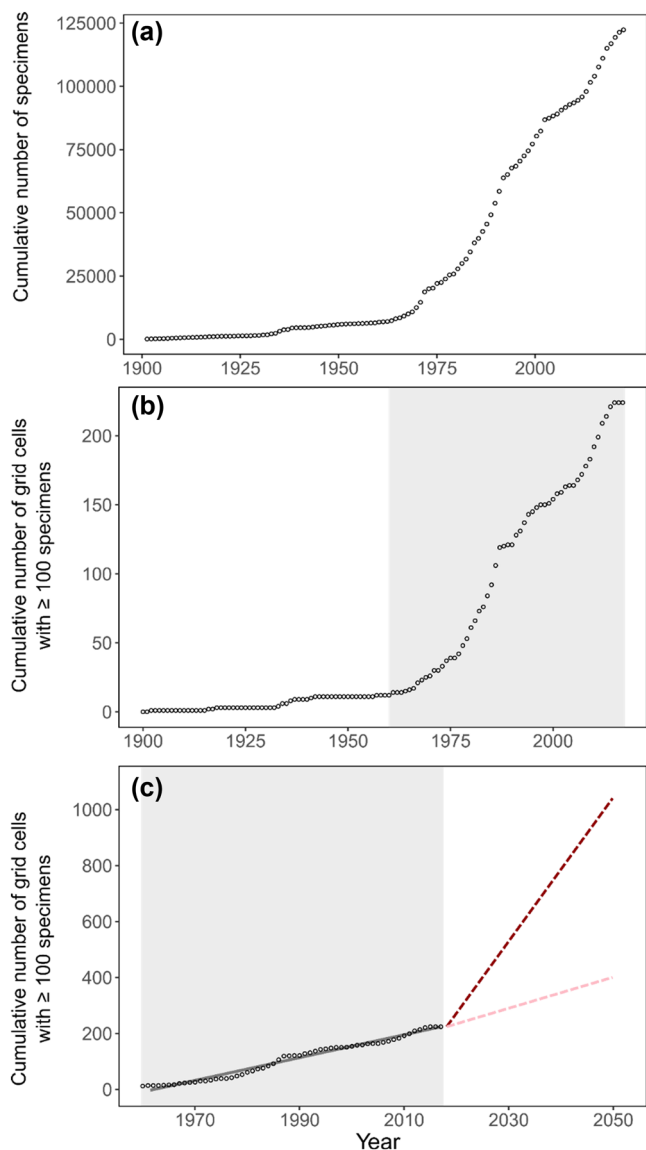


Figure 4. (a) Cumulative number of tree specimens and (b) cumulative number of grid cells in which at least 100 specimens were collected between 1900 and 2017. (c) Cumulative number of grid cells with at least 100 specimens as observed from 1960 and 2017 (grey line) and projected until 2050. The number of grid cells with  $\geq 100$  specimens was projected to add up to 1040 (dark red line) or 400 (pink line); which is equivalent to the total number of currently unsampled cells that may become deforested under the business-as-usual and improved governance scenario. The grey areas highlight the period between 1960 and 2017, when the number of grid cells with  $\geq 100$  specimens increased linearly ( $y = -8007.0 + 4.08 \times x$ ;  $r^2 = 0.97$ ,  $p < 0.05$ ).

value of past surveys, since they may be the only evidence for the occurrence of species in a given area (Vellend et al. 2013), thereby providing important baseline data for conservation and restoration. For example, Devey et al. (2013) reconstructed genetic connectivity between populations of the genus *Eligmodon* (Fabaceae) by sequencing the DNA

of recently collected specimens and specimens that were collected in areas that later became deforested. In the Brazilian Amazon, however, herbarium specimens are rarely used as a source of baseline data for biodiversity conservation and restoration (Durigan et al. 2013, Bustamante et al. 2019), partially due to low collection effort (Nelson et al. 1990, Hopkins 2007, 2019, Schulman et al. 2007, Feeley 2015).

Moreover, analysing herbarium specimens requires first to identify errors in specimens' labels and digital records (Rapacciuolo and Blois 2019). The data filtering applied in this study found 68% of the initial number of specimens contained errors in species names, place or date of collection. Errors still persist in our clean dataset, especially if related to the original herbarium labels. Such errors include different species names or dates of collection assigned to duplicate specimens housed in different herbaria (Hopkins 2007) or incorrect species identification (Goodwin et al. 2015). These errors are often copied onto records that become available online. Moreover, errors can occur when transcribing information contained in the specimens' labels (Groom et al. 2019; Supplementary material Appendix 1 Fig. A4). Errors like these can prevent the identification of duplicate specimens. As a consequence, the number of unique plant collections made in the Brazilian Amazon is probably even lower than the number of digital specimens analysed here.

If past trends persist, some 250 000–900 000 km<sup>2</sup> of poorly documented rainforest may be lost by 2050. Our second research question sought to understand by how much botanical sampling needs to increase so that the tree flora is documented before it is deforested. Our estimates show that, if future tree sampling is to cover the 250 000–900 000 km<sup>2</sup> that is predicted to be deforested, but remain under-collected, sampling has to increase two to six-fold. Note that deforestation scenarios used here are based on high historical deforestation rates between 1997 and 2002. However, deforestation after 2005 has been lower than predicted by the business-as-usual scenario (Nepstad et al. 2009). The decline in deforestation is attributed to the expansion of the protected areas network, more strict access to rural credits and imprisonment of illegal loggers. Some of these measures are no longer in place and recent changes in public policy in Brazil risk bringing back the high deforestation rates observed in the past (Levis et al. 2020).

Considerable uncertainty also remains about how many and which tree species in the Brazilian Amazon are affected by deforestation. Given the low density of tree collections, it is only possible to model the distribution of roughly half of the of tree species that are hypothesized to occur in the Amazon (Gomes et al. 2019). Reducing uncertainties in species distribution models requires improving the quality and coverage of species occurrence data. Looking at historic tree sampling, we find that 80% of all tree species collected in the Brazilian Amazon are represented by less than five correctly labelled herbarium specimens. This finding is in line with previous studies conducted in the Amazon, reporting that, on average, species are represented by three (Milliken et al.



2010) or four herbarium specimens (ter Steege et al. 2011). As shown here, species that occur in the Guiana shield, central and western–northern have a lower number of specimens than those that occur in eastern, southern and south-western areas of the Brazilian Amazon (Fig. 1), where accessibility and deforestation are highest. Past sampling efforts in these accessible regions did not ensure that range-restricted and threatened species, such as *Pouteria psammophila* (O’Brien 1998), are sufficiently represented in herbaria (Supplementary material Appendix 1 Table A1).

Both botanical sampling and deforestation are not random but subject to strong biases towards easily accessible areas (Barber et al. 2014), where both collectors and loggers can more conveniently engage in their respective activities. As a result, we find that better-sampled cells, which cover only 1.5% of the Brazilian Amazon, are spatially clustered and occur in close proximity to major roads or large rivers (Fig. 2b). Yet, 219 375 km<sup>2</sup> (approximately the size of the United Kingdom) or 7% of forested areas in the Brazilian Amazon (351 cells) do not provide a single specimen to herbarium collections but are relatively easy to reach. Our proxy for accessibility (i.e. travel time to major cities) does not consider, however, additional barriers to botanical sampling, such as bureaucratic hurdles to conduct biological surveys in protected areas (dos Santos et al. 2015) and land-use conflicts. For instance, Southern Pará, a region we identify as more easily accessible and poorly sampled, overlaps with areas of historic land-use conflict (Escada et al. 2005) and the protected area ‘Estação Ecológica Terra do Meio’.

Biases, gaps and an overall low botanical sampling have been recurrently reported for the Amazon since the 1990s (Nelson et al. 1990, Hopkins 2007, 2019, Schulman et al. 2007, Feeley 2015). The low rate of botanical sampling in the Brazilian Amazon place the region 65 yr behind other Brazilian regions, such as Southeast and South Brazil, in terms of botanical knowledge (Hopkins 2019). Why do low sampling rates persist in the Brazilian Amazon? Part of the problem is the low accessibility and an unequal geographical distribution of resources, both human and financial. Even though the Amazon accounts for more than half of the Brazilian territory, recent botanical programs have allocated only 5–10% of available resources to the region (Hopkins 2019).

Historical peaks in botanical sampling coincide with high profile research programs, specifically the ‘Flora projects’ that focused on intensive field surveys and accurate species descriptions (Supplementary material Appendix 1 Fig. A5). The number of collections peaked in all Amazonian states, except Tocantins, between 1977 and 1984 (a period that coincides with the ‘Projeto Flora Amazônica’), increasing the number of collections in Amazonian herbaria by about 50% (Prance et al. 1984, ter Steege et al. 2016). Between 1992 and 1997 the peak in botanical sampling in the Amazonas state may be explained by the ‘Projeto Flora da Reserva Ducke’. Intensive botanical sampling in the 10 000 ha of this nature reserve added 1000 plant species to the known species in the

area and resulted in the discovery of at least 48 new plant species (Hopkins 2005). The projects ‘Flora do Cristalino’ (Zappi et al. 2011) and ‘Flora do Acre’ (Medeiros et al. 2014) contributed to peaks in botanical sampling in the states of Mato Grosso and Acre. These flora projects not only boosted botanical collections across the Brazilian Amazon but they also increased the taxonomic expertise in the region (Hopkins 2005).

If future tree sampling is to cover an area of 250 000–900 000 km<sup>2</sup> (predicted to be deforested), current sampling rates will need to increase two to six-fold. Collectors could adopt three strategies, each with distinct implications for herbarium collections.

1) Focus sampling of difficult-to-reach and under-sampled areas. This strategy would expand knowledge about species distribution and lead to new species discoveries. Difficult-to-reach areas are located in southwestern Amazonas and northern Amapá and include large protected areas and indigenous territories (e.g. Vale do Javari indigenous territory or Tumucumaque National Park) (Fig. 2a). Increasing the frequency of surveys in remote or protected areas may require an improved research infrastructure, simplified bureaucratic procedures for biodiversity sampling and increased funding for more costly botanical expeditions (Martins et al. 2017).

2) Sampling of accessible but under-sampled areas. This strategy can help establishing complete species inventories for a given location (Ribeiro et al. 1999) and could provide important historical baseline data. Efforts in this direction should focus on the large and botanically unexplored areas in southern Pará and central-northern Mato-Grosso (Fig. 3c). Sampling these regions may require overcoming bureaucratic hurdles and land-use conflicts.

3) Sampling of areas vulnerable to deforestation and targeting under-collected species. This strategy can help documenting species before their habitats are lost. It can also generate knowledge about species that are poorly represented in herbaria but face the greatest threat from deforestation (Supplementary material Appendix 1 Table A1). Such species, including those that classify as vulnerable or endangered according to the IUCN Red List, occur predominantly in the eastern and southern portions of the Brazilian Amazon. While it is crucial to sample threatened species, a one-sided focus could overlook opportunities to better document species that are missing from the IUCN Red List or similar inventories (Possingham et al. 2002).

Implementing a strategy of targeted sampling, ideally through ‘flora projects’, would considerably expand our knowledge of the tree flora in the Brazilian Amazon. Flora projects deepen botanical knowledge because they cover several aspects of botanical exploration, from establishing field inventories to accurately identifying species. The latter may lead to taxonomic re-assignments and the discovery of new species (Bebber et al. 2010). The success of flora projects in the Brazilian Amazon can be attributed to the work of taxonomists and para-taxonomists in the region (Prance et al. 1984, Ribeiro et al. 1999). Yet, taxonomic research is among

the most underfunded of disciplines in the biological sciences (Wilson 2003, 2004). Changing the situation in times of limited research budgets (Magnusson et al. 2018) is a major challenge. Finally, it is important to recognize that trees are one of the best sampled taxa in the Brazilian Amazon. Knowledge of other taxa, such as arthropods, is characterized by even larger shortfalls (Oliveira et al. 2016).

## Data availability statement

Data available from Figshare Digital Repository: <<https://dx.doi.org/10.6084/m9.figshare.11839020>> (Stropp et al. 2020).

*Acknowledgements* – We thank Martin Weiss and three anonymous reviewers for their suggestions in earlier versions of this manuscript. The ideas presented in this study owe much to discussions with Joaquín Hortal.

*Funding* – This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES; Finance Code 001; PPG-DIBICT fellowship to BU and PNPD fellowship to JVS), the Brazilian National Council for Scientific and Technological Development (CNPq research project (#448688/2014-0) and post-doctoral fellowship to JS (#434391/2016-6)). RAC is currently funded by the Helsinki Inst. for Sustainability Science (HELSUS) and the Univ. of Helsinki. RJL and ACMM are supported by CNPq (#309879/2019-1 and #309980/2018-6). This study also benefited from financial support from ‘GBIF’s 2016 Young Researchers’ Award’ and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Action (grant agreement #843234; project: TAXON-TIME).

*Author contributions* – The first and second author contributed equally to this paper.

## References

- Barber, C. P. et al. 2014. Roads, deforestation and the mitigating effect of protected areas in the Amazon. – *Biol. Conserv.* 177: 203–209.
- Barlow, J. et al. 2019. Clarifying Amazonia’s burning crisis. – *Global Change Biol.* 26: 319–321.
- Bebber, D. P. et al. 2010. Herbaria are a major frontier for species discovery. – *Proc. Natl Acad. Sci. USA* 107: 22169–22171.
- Beech, E. et al. 2017. GlobalTreeSearch: the first complete global database of tree species and country distributions. – *J. Sustain. For.* 36: 454–489.
- Bivand, R. et al. 2018. rgdal: bindings for the geospatial data abstraction library 2017. – R package ver. 0.8-13, pp. 1–55.
- Boyle, B. et al. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. – *BMC Bioinform.* 14: 16.
- Bustamante, M. M. C. et al. 2019. Ecological restoration as a strategy for mitigating and adapting to climate change: lessons and challenges from Brazil. – *Mitig. Adapt. Strateg. Global Change* doi: 10.1007/s11027-018-9837-5
- Camara, G. et al. 2013. Metodologia para o cálculo da taxa anual de desmatamento na Amazônia legal. – *Inst. Nac. Pesqui. Espac. Rep.* <[www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/metodologia\\_taxaprodes.pdf](http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/metodologia_taxaprodes.pdf)>.
- Cardoso, D. et al. 2017. Amazon plant diversity revealed by a taxonomically verified species list. – *Proc. Natl Acad. Sci. USA* 114: 10695–10700.
- Chamberlain, S. 2017. rgbif: interface to the global ‘biodiversity’ information facility ‘API’. – R package ver. 0.9.8, <<https://CRAN.R-project.org/package=rgbif>>.
- Cox, C. B. et al. 2016. Biogeography: an ecological and evolutionary approach. – Wiley.
- Crooks, K. R. et al. 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. – *Proc. Natl Acad. Sci. USA* 114: 7635–7640.
- Devey, D. S. et al. 2013. A snapshot of extinction in action: the decline and imminent demise of the endemic *Eligmocarpus Capuron* (Caesalpinioideae, Leguminosae) serves as an example of the fragility of Madagascan ecosystems. – *S. Afr. J. Bot.* 89: 273–280.
- Dinerstein, E. et al. 2019. A global deal for nature: guiding principles, milestones and targets. – *Sci. Adv.* 5: eaaw2869.
- dos Santos, J. G. et al. 2015. Geographic trends and information deficits in Amazonian conservation research. – *Biodivers. Conserv.* 24: 2853–2863.
- Durigan, G. et al. 2013. Ecological restoration of Xingu Basin headwaters: motivations, engagement, challenges and perspectives. – *Phil. Trans. R. Soc. B* 368: 20120165.
- Enquist, B. J. et al. 2009. The Botanical Information and Ecology Network (BIEN): cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. – *iPlant Collab.*
- Escada, M. I. S. et al. 2005. Processos de ocupacao das novas fronteiras da Amazonia: o interfluvio do Xingu/Iriri. – *Estud. Avancados* 19: 9–23.
- Fearnside, P. M. 2015. Deforestation soars in the Amazon. – *Nature* 521: 423.
- Feeley, K. 2015. Are we filling the data void? An assessment of the amount and extent of plant collection records and census data available for tropical South America. – *PLoS One* 10: e0125629.
- Fernandes, G. W. et al. 2017. Dismantling Brazil’s science threatens global biodiversity heritage. – *Perspect. Ecol. Conserv.* 15: 239–243.
- GBIF 2018a. GBIF occurrence download <<https://doi.org/10.15468/dl.bj5o3m>>. – Accessed from R via rgbif (<<https://github.com/ropensci/rgbif>>) on 10 May 2018.
- GBIF 2018b. GBIF occurrence download <<https://doi.org/10.15468/dl.yftoxc>>. – Accessed from R via rgbif (<<https://github.com/ropensci/rgbif>>) on 14 May 2018.
- Gentry, A. H. 1992. Tropical forest biodiversity: distributional patterns and their conservational significance. – *Oikos* 63: 19–28.
- Gillson, L. et al. 2011. Baselines, patterns and process. – In: Whitaker, R. J. and Ladle, R. J. (eds), *Conservation biogeography*. Blackwell–Wiley, pp. 31–44.
- Gomes, V. H. F. et al. 2019. Amazonian tree species threatened by deforestation and climate change. – *Nat. Clim. Change* 9: 547–553.
- Goodwin, Z. A. et al. 2015. Widespread mistaken identity in tropical plant collections. – *Curr. Biol.* 25: R1066–R1067.
- Grelle, C. E. V. 2005. Predicting extinction of mammals in the Brazilian Amazon. – *Oryx* 39: 347–350.
- Groom, Q. et al. 2019. Improved standardization of transcribed digital specimen data. – *Database* 2019: 1–4.

- Hopkins, M. J. G. 2005. Flora da reserva Ducke, Amazonas, Brazil. – *Rodriguesia* 56: 9–25.
- Hopkins, M. J. G. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. – *J. Biogeogr.* 34: 1400–1411.
- Hopkins, M. J. G. 2019. Are we close to knowing the plant diversity of the Amazon? – *An. Acad. Bras. Cienc.* 91: e20190396.
- Hortal, J. et al. 2008. Historical bias in biodiversity inventories affects the observed realized niche of the species. – *Oikos* 117: 847–858.
- Hortal, J. et al. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. – *Annu. Rev. Ecol. Evol. Syst.* 46: 523–549.
- Hubbell, S. P. et al. 2008. How many tree species are there in the Amazon and how many of them will go extinct? – *Proc. Natl Acad. Sci. USA* 105: 11498–11504.
- INPE 2019. Monitoramento da Floresta Amazônica Brasileira por Satélite. – [www.obt.inpe.br/OBT/noticias/a-estimativa-da-taxa-de-desmatamento-por-corte-raso-para-a-amazonia-legal-em-2019-e-de-9-762-km2](http://www.obt.inpe.br/OBT/noticias/a-estimativa-da-taxa-de-desmatamento-por-corte-raso-para-a-amazonia-legal-em-2019-e-de-9-762-km2).
- Kassambra, A. 2017. Practical guide to principal component methods in R. – [sthda.com](http://sthda.com).
- Kintisch, E. 2007. Improved monitoring of rainforests helps pierce haze of deforestation. – *Science* 316: 536–537.
- Ladle, R. J. and Hortal, J. 2013. Mapping species distributions: living with uncertainty. – *Front. Biogeogr.* 5: 8–9.
- Levis, C. et al. 2020. Help restore Brazil's governance of globally important ecosystem services. – *Nat. Ecol. Evol.* 4: 172–173.
- Lobo, J. M. 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.
- Magnusson, W. E. et al. 2018. Effects of Brazil's political crisis on the science needed for biodiversity conservation. – *Front. Ecol. Evol.* 6: 163.
- Maitner, B. S. et al. 2018. The bien r package: a tool to access the Botanical Information and Ecology Network (BIEN) database. – *Methods Ecol. Evol.* 9: 373–379.
- Martins, E. et al. 2017. Challenges and perspectives for achieving the global strategy for plant conservation targets in Brazil. – *Ann. Missouri Bot. Gard.* 102: 347–357.
- Medeiros, H. et al. 2014. Botanical advances in southwestern Amazonia: the flora of Acre (Brazil) five years after the first Catalogue. – *Phytotaxa* 177: 101–117.
- Milliken, W. et al. 2010. Amazon vegetation: how much don't we know and how much does it matter? – *Kew Bull.* 65: 691–709.
- Nelson, A. 2008. Travel time to major cities: a global map of accessibility. – *Off. Off. Publ. Eur. Communities, Luxemburg*, doi: 10.2788/95835.
- Nelson, B. W. et al. 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. – *Nature* 345: 714.
- Nepstad, D. et al. 2009. The end of deforestation in the Brazilian Amazon. – *Science* 326: 1350–1351.
- Nobre, C. A. et al. 2016. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. – *Proc. Natl Acad. Sci. USA* 113: 10759–10768.
- O'Brien, P. 1998. *Pouteria psammophila*. – The IUCN Red List of threatened species 1998: e.T35883A9958778.
- Oliveira, U. et al. 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. – *Divers. Distrib.* 22: 1232–1244.
- Possingham, H. P. et al. 2002. Limits to the use of threatened species lists. – *Trends Ecol. Evol.* 17: 503–507.
- Prance, G. T. et al. 1984. Projeto flora Amazonica: eight years of binational botanical expeditions. – *Acta Amaz.* 14: 5–30.
- Rapacciuolo, G. and Blois, J. L. 2019. Understanding ecological change across large spatial, temporal and taxonomic scales: integrating data and methods in light of theory. – *Ecography* 42: 1247–1266.
- Ribeiro, J. E. L. da S. et al. 1999. Flora da Reserva Ducke. – INPA/DFID.
- Schulman, L. et al. 2007. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. – *J. Biogeogr.* 34: 1388–1399.
- Soares-Filho, B. S. et al. 2006. Modelling conservation in the Amazon basin. – *Nature* 440: 520–523.
- Soares-Filho, B. S. et al. 2013. LBA-ECO LC-14 modeled deforestation scenarios, Amazon Basin: 2002–2050. – Data set, available on-line from Oak Ridge Natl Lab. Distrib. Act. Arch. Center, Oak Ridge, TN, USA, <<http://dx.doi.org/10.3334/ORNDAAC/>>.
- SpeciesLink 2018. SpeciesLink occurrence download. – <[www.splink.org.br](http://www.splink.org.br)> accessed 15 May 2018.
- Stropp, J. et al. 2020. Data from: The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon. – Figshare Digital Repository, <<https://dx.doi.org/10.6084/m9.figshare.11839020>>.
- ter Steege, H. et al. 2011. A model of botanical collectors' behavior in the field: never the same species twice. – *Am. J. Bot.* 98: 31–37.
- ter Steege, H. et al. 2015. Estimating the global conservation status of more than 15,000 Amazonian tree species. – *Sci. Adv.* 1: e1500936.
- ter Steege, H. et al. 2016. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. – *Nat. Sci. Rep.* 29549: 1–15.
- ter Steege, H. et al. 2019. Towards a dynamic list of Amazonian tree species. – *Sci. Rep.* 9: 3501.
- Tessarolo, G. et al. 2017. Temporal degradation of data limits biodiversity research. – *Ecol. Evol.* 7: 6863–6870.
- Vellend, M. et al. 2013. Historical ecology: using unconventional data sources to test for effects of global environmental change. – *Am. J. Bot.* 100: 1294–1305.
- Wilson, E. O. 2003. The encyclopedia of life. – *Trends Ecol. Evol.* 18: 77–80.
- Wilson, E. O. 2004. Taxonomy as a fundamental discipline. – *Phil. Trans. R. Soc. B* 359: 739.
- Zappi, D. C. et al. 2011. Plantas vasculares da região do Parque Estadual Cristalino, norte de Mato Grosso, Brasil. – *Acta Amaz.* 41: 29–38.

Supplementary material (available online as Appendix ecog-05026 at <[www.ecography.org/appendix/ecog-05026](http://www.ecography.org/appendix/ecog-05026)>). Appendix 1.