HAL
open science

The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa

Hans ter Steege, Rens W Vaessen, Dairon Cárdenas-López, Daniel Sabatier, Alexandre Antonelli, Sylvia Mota de Oliveira, Nigel Pitman, Peter Møller Jørgensen, Rafael P Salomão

To cite this version:

Hans ter Steege, Rens W Vaessen, Dairon Cárdenas-López, Daniel Sabatier, Alexandre Antonelli, et al.. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. Scientific Reports, 2016, 6, pp.29549. 10.1038/srep29549 . hal-02127709

HAL Id: hal-02127709

https://hal.umontpellier.fr/hal-02127709v1

Submitted on 13 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SCIENTIFIC REPORTS

OPEN

The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa

Received: 01 March 2016
Accepted: 17 June 2016
Published: 13 July 2016

Hans ter Steege[1,2], Rens W. Vaessen[1,†], Dairon Cárdenas-López[3], Daniel Sabatier[4], Alexandre Antonelli[5,6], Sylvia Mota de Oliveira[1], Nigel C. A. Pitman[7,8], Peter Møller Jørgensen[9] & Rafael P. Salomão[2]

Amazonia is the most biodiverse rainforest on Earth, and the debate over how many tree species grow there remains contentious. Here we provide a checklist of all tree species collected to date, and describe spatial and temporal trends in data accumulation. We report 530,025 unique collections of trees in Amazonia, dating between 1707 and 2015, for a total of 11,676 species in 1225 genera and 140 families. These figures support recent estimates of 16,000 total Amazonian tree species based on ecological plot data from the Amazonian Tree Diversity Network. Botanical collection in Amazonia is characterized by three major peaks, centred around 1840, 1920, and 1980, which are associated with flora projects and the establishment of inventory plots. Most collections were made in the 20th century. The number of collections has increased exponentially, but shows a slowdown in the last two decades. We find that a species' range size is a better predictor of the number of times it has been collected than the species' estimated basin-wide population size. Finding, describing, and documenting the distribution of the remaining species will require coordinated efforts at under-collected sites.

The forests and savannahs of the Amazon basin and Guiana Shield (here Amazonia) arguably hold the greatest biodiversity on Earth: an estimated 1300 species of birds, 427 species of mammals and 50,000 species of seed plants[1,2]. Whereas the collection record for birds and mammals may be relatively complete and the estimates for those taxa relatively precise, plants in Amazonia remain hugely under-collected[3,4]. Indeed, 300 years of plant collecting in Amazonia have resulted in a modern-day density of only 10 collections/100 km^2[5,6]. Feeley and Silman[7] suggested that the lack of data is so strong that there exists a '*gaping data void' such that many species and many habitats remain… functionally invisible for most studies*". Data is not only scarce; it is also often unavailable[3,5], and biased spatially and taxonomically[8–10].

As a result, the number of tree species in Amazonia remains unknown. No checklist of all Amazonian trees has ever been compiled, and estimates of the size of the Amazonian tree flora are hotly contested because data is scant and extrapolation techniques are abundant. Recently, ter Steege and colleagues estimated the number of Amazonian tree species that exceed 10 cm diameter at breast height (DBH) at ~16,000, by fitting a logseries curve to population estimates of ~5000 species identified in 1170 inventory plots across Amazonia[11]. That hypothesis has been embraced as plausible by some authors[12] but criticized by others[13,14]. Non-parametric estimates, that were suggested by an anonymous reviewer of the above paper, predict considerably lower plant diversity: 6000–7000 species in the case of Amazonian trees[11].

[1]Naturalis Biodiversity Center, Vondellaan 55, Postbus 9517, 2300 RA Leiden, The Netherlands. [2]Coordenação de Botânica, Museu Paraense Emílio Goeldi, Av. Magalhães Barata 376, C.P. 399, Belém, PA 66040–170, Brazil. [3]Herbario Amazónico Colombiano, Instituto SINCHI, Calle 20 No 5-44, Bogotá, DF, Colombia. [4]Institut de Recherche pour le Développement (IRD, UMR AMAP), TA A-51/PS2, Bd. de la Lironde, 34398 Montpellier cedex 5, France. [5]Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, SE-405 30 Göteborg, Sweden. [6]Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, SE-413 19, Göteborg, Sweden. [7]Science and Education, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL 60605–2496, USA. [8]Nicholas School of the Environment, Duke University, Durham, North Carolina 27705, USA. [9]Missouri Botanical Garden, P.O. Box 299, St. Louis, Missouri 63166-0299, USA. [†]Present address: Ecology and Biodiversity Group, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. Correspondence and requests for materials should be addressed to H.T.S. (email: hans.tersteege@naturalis.nl)

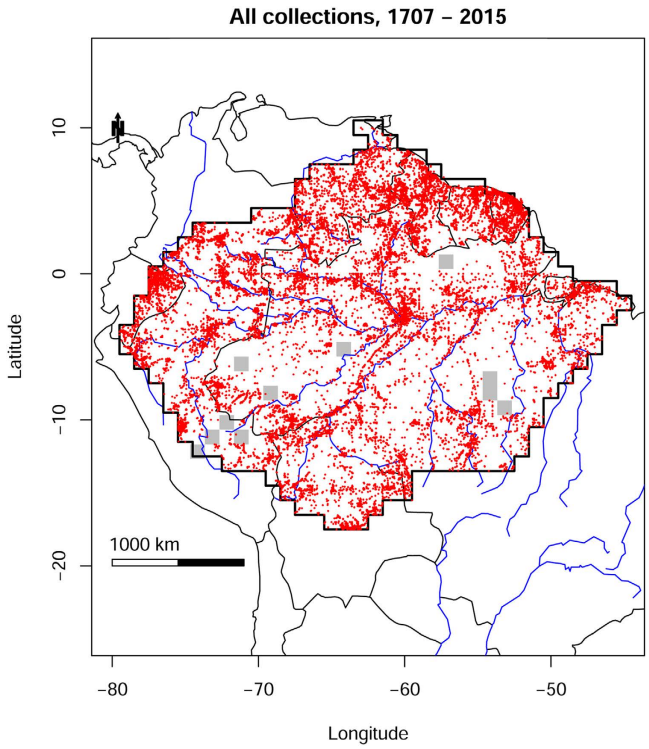

Figure 1. Collection localities of herbarium specimens of Amazonian trees, collected between 1707 and 2015, for which geographical coordinates were available and considered reliable. Grey squares are 1-degree grid cells that contain no collections. Map created with custom R script. Base map source (country.shp, rivers.shp): ESRI (http://www.esri.com/data/basemaps, © Esri, DeLorme Publishing Company).

To provide a stronger empirical foundation for this debate, here we provide a preliminary checklist of all valid tree species collected to date in Amazonia. We then analyse the list to explore why some Amazonian tree species are more frequent in herbaria and other floristic datasets than others. Does a species' frequency in these datasets reflect its abundance and range size, the date it was first discovered in Amazonia, the spatial pattern of Amazonian exploration, or some combination of those factors? Answering these questions will allow us to discuss the dynamic behind the discovery of new species and consequently the best approach to complete the inventory of the Amazonian tree flora.

Results

Diversity and collection density. The checklist contains 11,187 valid species names (Appendix S1) from 530,025 unique collections of trees made in Amazonia over the last three centuries, from 1707 to 2015, as well as 489 valid species names known to occur in Amazonia but lacking occurrence record data. The total number of species in the checklist is 11,676.

Average collection density (530,025 specimens from an area of 5.6 million km^2) is just under 10 per 100 km^2. Collections are spread across Amazonia, but concentrations of collecting effort are clearly visible in the Guianas, around large Amazonian cities (e.g. Belém, Manaus), and along major rivers and roads (Fig. 1). Thus, while a 1-degree grid cell ($111 \times 111\,km^2$) in Amazonia contains ~950 collections on average, 31 grid cells contain more than 2500. The grid cell in which Manaus is located has 24,598 collections (Fig. S1). At the poorly collected end of the spectrum, half of all grid cells contain fewer than 200 collections. Eleven grid cells contain no tree collections at all, and together these cover an area of approximately 320,000 km^2.

The vast majority of the Amazonian tree species in our preliminary checklist (96%) come from our initial collections database (GBIF, SpeciesLink, INPA, MG, and COAH). Roughly 41% of all Amazonian tree species (4657) have been recorded in ATDN tree plots to date (Appendix S1) and 104 of these were not found in the herbarium databases we examined. Additional species (i.e., species not in the initial collections dataset or ATDN) were added from checklists, catalogues, and IPNI (385). Correcting all names of Annonaceae with Maas *et al.*[15] resulted in 22 changes in the 498 names of Amazonian Annonaceae (down to 481). Five of these were simply a change in name, while 17 were the merging of two species into one new synonymy. The latest *Guatteria* treatment made more of an impact: 62 species were put in synonymy of 15 species (10 into *G. hirsuta* and 26! into *G. punctata*) and 10 new Amazonian species were described[16].

Taxonomic patterns. The 11,676 tree species in the checklist belong to 140 families and 1225 genera. Fabaceae was, as expected, the family with the largest number of accepted tree species (1611), followed by Rubiaceae (1058), Melastomataceae (624), Myrtaceae (606), Lauraceae (566), Annonaceae (480), Euphorbiaceae

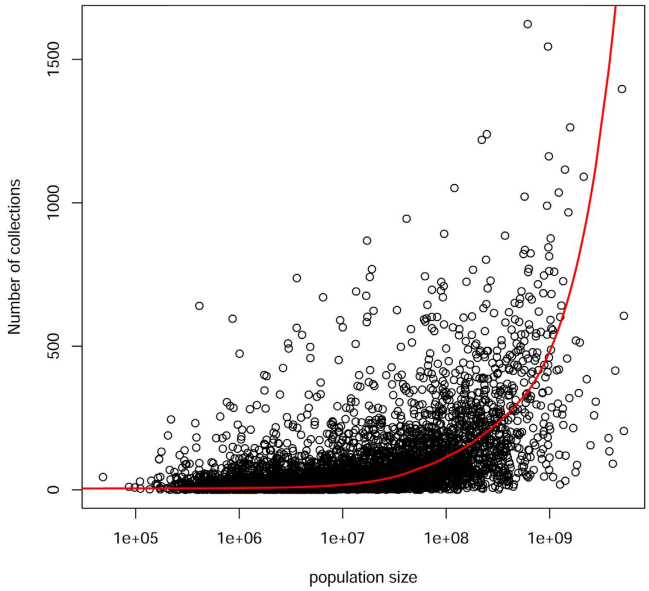

Figure 2. The effect of estimated basin-wide population size on the number of specimens in herbarium databases, for 11,187 Amazonian tree species. Note that x-axis is log-scale. Red line: local weighted regression (loess) through all data.

(351), Chrysobalanaceae (312), Malvaceae (304), and Sapotaceae (278). These 10 families accounted for 53% of all tree species (Appendix S3). Fifteen families were represented by a single species (Appendix S3). The ten most speciose genera were *Miconia* (307 species), *Eugenia* (209), *Ocotea* (190), *Inga* (184), *Psychotria* (182), *Myrcia* (173), *Licania* (159), *Pouteria* (154), *Solanum* (147), and *Swartzia* (144), which together accounted for 16% of all species (Appendix S3). More than a third of all genera (451 out of 1225) were monospecific in the Amazonian tree collections.

Collection frequency, estimated population size, and range size. The tree species with the highest number of collections was *Hirtella racemosa* (Chrysobalanaceae), with 1623 collections, followed by *Tapirira guianensis, Eschweilera coriacea, Licania heteromorpha, Siparuna guianensis, Cordia nodosa, Virola elongata,* and *Micropholis guyanensis*. Fifty percent of all collections were accounted for by the 745 most collected species; these represent 6.7% of all tree species in the collections dataset.

Estimated population size had a moderate effect on the number of collections per species (Fig. 2, linear regression $R^2 = 22\%$, $p \ll 0.001$). Only 139 of the 227 hyperdominants, sensu ter Steege *et al.*[11], were among the most frequently collected species that accounted for 50% of all collections. Among species with population sizes ranging from 10^5 to 10^7 individuals, the mean number of collections was nearly constant at 14 collections per species. For species with estimated populations between 1 and 10^6 individuals[11], the mean number of collections was also around 14 (data not shown, see Appendix S1). Among species with population sizes over 10^8, the mean number of collections rose sharply to 200 collections per species, for a final average of close to 700 for species with populations over 10^9 individuals. There was a similarly small effect of rank population size on rank number of collections (Fig. S2), with the former explaining 35% of the variation in the latter. Population size had no noticeable correlation with the date a species was first collected (Fig. S3). Species discovery has increased through time starting in 1800 (Fig. 3), with peaks in discovery centred around 1820–1850, 1900–1920, and a large peak around 1975–1985.

Range size, as defined by the number of plots in which a species has been found, had a moderately strong effect on the number of collections (Fig. 4A, linear regression: $R^2 = 44\%$, $p \ll 0.001$). Species that were found in 100 plots or more had on average more than ten times the number of collections than species observed in fewer than ten plots. However, some species only observed in 1–5 plots had close to 500 collections. The number of grid cells in which a species has been collected, another measure for its range size (Fig. 4B), is strongly correlated with the number of collections ($R^2 = 88\%$, $p \ll 0.001$). With a slope of 3, on average species are collected 3 times per grid cell in which they have been collected. After accounting for the number of grid cells in which a species occurs, neither the number of plots nor the population of the species adds more than 1% explained variation to the number of collections.

We found only a few thousand collection records (5696) from the period before 1900, after which the average collection rate increased to 839 specimens/year. After 1960, the collection rate rose again sharply to 10,380 collections/year. There have been fewer new collections in the last 10 years (4727/year from 2000 to 2010 and 2636/year for the last five years (Fig. 5). The rate of tree species discovery in Amazonia has risen faster than the rate of collections (Fig. 5), with a rate of 35 newly collected tree species per year between 1820 and 1850. This rate slowed considerably (13 new species per year) between 1850 and 1875, after which it rose to an average of 84 newly collected tree species/year from 1900 to 2000, peaking from 1975 to 1985 with an average of 164 newly collected

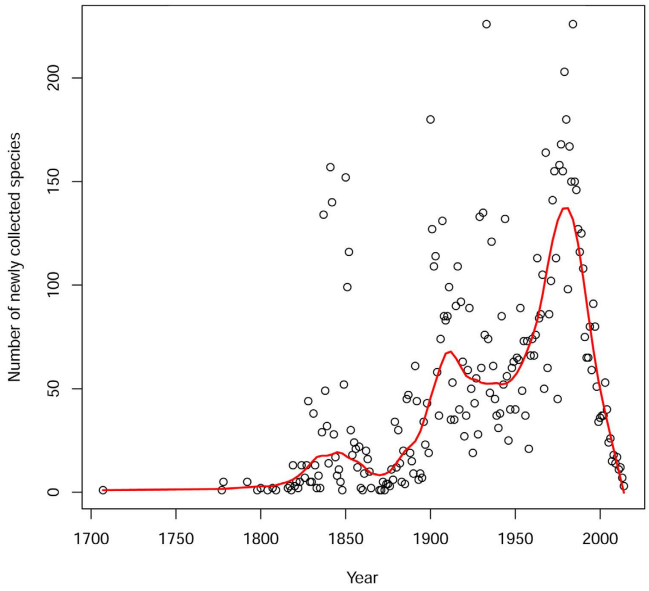

Figure 3. The number of first-time collections of Amazonian tree species each year between 1707 and 2015. The red line (loess regression) suggests 3 peaks in discovery. The decline after 1980 is partly the effect of decreasing efficiency in the rate of new species per number of collections (see Fig. 6) and partly the effect of the lag time between the collection of new species, the description of new species, and the appearance of new species in herbarium databases.

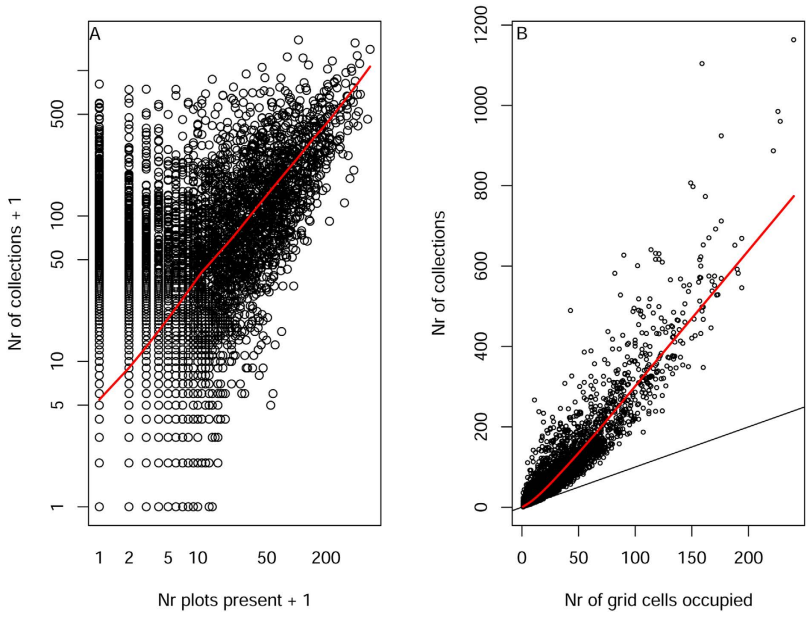

Figure 4. (**A**) The number of ATDN plots in which a species is present[11], a proxy for range size across Amazonia, as a predictor for the number of collections. (**B**) The number of grid cells in which a species has been collected, a non-independent proxy for range size, as a predictor for the number of collections.

species/year. The randomized species-collection curve (Fig. 6) suggests that more tree species will be found in Amazonia if substantial collecting effort continues (see discussion). But the number of collections required to find a new species has increased substantially over the last 300 years, from a few collections per new species in the 1800 s to ~300 collections for a new species now (Fig. S4).

A historical perspective on the discovery of the Amazonian tree flora. *1700.* Our data suggests that the earliest collections of Amazonian trees were mainly made in the three Guianas (Figs S5 and S6). More than half of the collections (3562) before 1900 came from this area. By area the Guianas are still the best collected countries in Amazonia by far (Table 1). Jean Baptiste Christian Fusée Aublet may have been the first serious

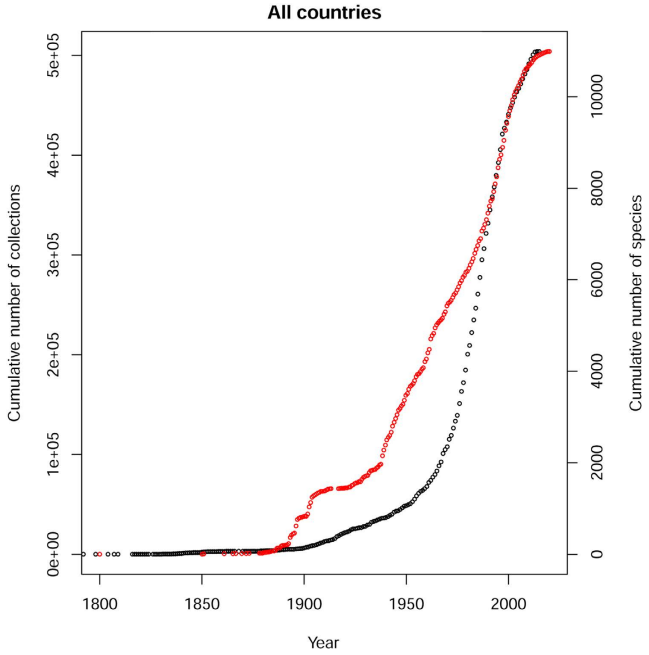

Figure 5. Cumulative number of tree species specimens collected in Amazonia from 1800 to 2015 (Black line, left y-axis). Cumulative number of tree species collected in Amazonia from 1800 to 2015 (red line, right y-axis).

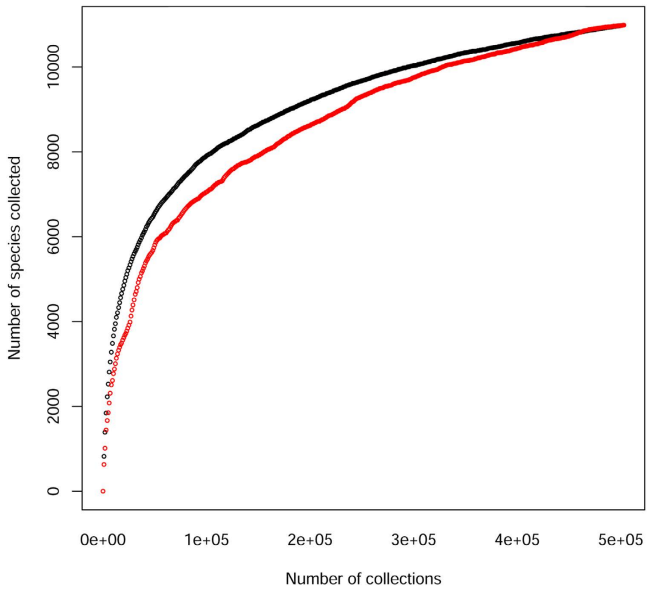

Figure 6. The number of species as a function of the number of collections. Black line: data randomized with regard to year of collection. Red line: data in order of date collected.

collector in the region, responsible for the first collections of 211 tree species (Appendix S1), including many type specimens from French Guiana (1762–1764). Based on his collections he published the '*Histoire des plantes de la Guiane françoise*'[17].

In the same period, practically all of the other Amazonian countries were botanically explored during the expeditions of Joseph de Jussieu (1704–1779). Although he made many other scientific collections in addition to plants, Jussieu was the first naturalist to collect plant specimens in Ecuador, Peru, and Bolivia, as a participant of the French Geodesic expedition from 1735 to 1743[18], together with the geographer and naturalist Charles Marie de La Condamine. After the French Geodesic expedition, de La Condamine, together with Louis Godin, Pierre Bouguer and the Spaniards Jorge Juan and Antonio de Ulloa, carried out an expedition from Quito to the Atlantic Ocean, through Ecuador, Colombia, and Brazil (and from there to France). This is considered to be the beginning of the great travels of European scientists in Amazonia, and led to the description of the quinine tree (*Cinchona officinalis*)

country	#collections	grid cells	Area (km^2)	collections/100 km^2	#species collected	#species estimated	%species collected
Bolivia	11721	40	492840	2.38	2404	6060	40
Brazil	278165	333	4102893	6.78	7694	12655	61
Colombia	35277	41	505161	6.98	3511	10073	35
Ecuador	16202	9	110889	14.61	2366	6827	35
French Guiana	64762	7	86247	75.09	2303	*	*
Guyana	36445	15	184815	19.72	2820	*	*
Peru	39851	62	763902	5.22	4422	9336	47
Suriname	31246	11	135531	23.05	2030	*	*
Venezuela	16356	38	468198	3.49	3189	6563	49
total	**530025**	**556**	**6850476**	**7.74**	**11194**	**16000**	**70**
*Guianas	132453	33	406593	32.58	3744	4581	82

Table 1. Collection data by country. Number of collections; Number of grid cells; Area of country based on grid cells (km^2); Collection density (collections/100 km^2); Number of species collected; Number of species collected based on ref. 11; Collection completeness as the percentage species collected of those expected. The latter two are combined for the three Guianas separate species estimate was made[11].

and the rubber tree (*Hevea* spp.). Also Alexandre Rodrigues Ferreira (1758–1815), merchant from Bahia, traveled through Amazonia to Mato Grosso (1783–1792) and collected many plants, now located at Paris and LISU.

Despite being visited by Jussieu, Bolivia was – and still is – one of the least explored western Amazonian countries[19] (Table 1). Some other early botanical exploration in Bolivia, of minor impact, include the French botanist Louis Feuillée (1660–1732) and the Austro-Hungarian naturalist and explorer Thaddaus Peregrinus Xaverius Haenke (1761–1816). Another pioneering expedition worth mention, although not centred in Amazonia, was that of Hipólito Ruiz, who travelled to Peru in 1777–1784 and later to Chile, writing the '*Florae Peruvianae et Chilensis prodromus*'.

The only country that was not visited by the two main collecting efforts of the 18th century (Aublet and Jussieu) was Venezuela. The first documented expedition in the country was the '*Comisión de Límites*' in 1755. The main botanist on the trip, Pehr Loefling, a student of Linnaeus, fell seriously ill and died and his plants were lost[20]. Later, Alexander von Humboldt and Aimé Bonpland collected in Venezuelan Guayana in 1800 and their excellent collections were the basis for further studies in the area. They travelled to the Orinoco and the Rio Negro, making important botanical collections in Venezuela, Brazil, and Colombia.

1800s. After 1800 many more naturalists were active in Amazonia. In the Guianas, there were two increases in collecting between 1804 and 1853 (Fig. S6B,C). The first peak occurred in 1837 due to the collecting effort of Frederick Louis Splitberger in Suriname, the second to collections by Robert and Richard Schomburgk, who travelled extensively between Guyana and Venezuela from 1835 to 1844 and published considerably[21,22].

Richard Spruce (1817–1893) spent almost seven years in Ecuador, from 1857 to 1863, but collected mainly in the highlands and left for Peru in 1863[18]. The most representative work of Spruce for the Amazonian Flora is the result of his collections in Venezuela and Brazil. He made over 800 botanical collections, many of which were new to science[20]. Apart from the general botanical collections, his study of Amazonian bryophytes is a benchmark in the Neotropics, with the description of hundreds of new species. Also in the first half of the 19th century, Juan Isern y Batlló (1825–1866) and companions on the Royal Spanish Pacific expedition left Ecuador through the Amazonian lowlands and travelled down the Amazon river to Manaus, Belém, and finally Pernambuco[18].

But the expedition that had the greatest impact on knowledge of the Amazonian tree flora in the 19th century was probably that of Carl Friedrich Philipp von Martius (1794–1868). With zoologist Johann Baptist von Spix he travelled four years (1817–1821) from Rio de Janeiro to São Paulo (in south-eastern Brazil) to Belém, and from there up the Amazon River by canoe. Overall, he and his team travelled 10,000 km and made 3,541 collections of birds, insects, and other animals, and 25,000–30,000 herbarium collections[23,24], which contain ca. 7,300 plant species, and are housed at the Botanische Staatssammlung München. Perhaps even more importantly, von Martius initiated the '*Flora Brasiliensis*' project, containing taxonomic treatments of 22,767 species, mostly Brazilian angiosperms. He completed 46 of the 130 volumes before his death in 1868, with the monograph being completed in 1906 (http://florabrasiliensis.cria.org.br/).

1900s. In the 20th century, the number of tree specimen collections and the number of tree species recorded for Amazonia increased by two orders of magnitude (Fig. 5). Collection effort also varied among countries (Fig. S6), because efforts became much more divided. While earlier efforts were concentrated in large expeditions that crossed country borders, the increased sovereignty of countries after independence from colonial rule led to localized and national programs of botanical exploration. Furthermore, the start of several flora projects intensified the study of the region.

Below a brief account of the collecting history of the 20th century for each of the Amazonian countries is given in order of collecting intensity. Table S1 lists the most important collectors for this period by country.

French Guiana. French Guiana is the best collected country by area in Amazonia, with a collecting density of 75 collections/100 km^2 (Table 1). The cumulative number of collections increases sharply in the second half of the century (Fig. S6A); this is related to the creation of the herbarium in Cayenne by Roelof Oldeman, and the consecutive collecting efforts of Oldeman's team and Jean-Jaques de Granville, widely spread over French Guiana, and of Scott Mori in Saul[21]. Most botanists were involved in both plot inventories and traditional collecting, making it difficult to disentangle their relative contributions. Nearly 265 ha of forest botanical inventories have been carried out in French Guiana since 1980, and these have added 350 new tree species for the country. During the same time, only ~60 tree species new to French Guiana were collected outside the plots.

Suriname. Up to 1953 fieldwork of botanists collecting under the 'Boschwezen' number series, mainly in the Zanderij, Sectie O, and Brownsberg areas and along the Nickerie River, added greatly to the number of collections in Suriname (Fig. S6B). In the same period collections intensified with the start of the '*Flora of Suriname*'[21]. Jan Lindeman added greatly to the collections of trees in Suriname and produced one of the first tree guides of the region[25].

Guyana. In the first half of the century, tree collections increased with large contributions from Noel Yvri Sandwith around the Mazaruni Station and from Albert Charles Smith around the Cuyuwini landing and in the Rupununi Savannah (Fig. S6C). The continued field trip and collections of the Forest Department led to a great increase in the knowledge of forest types and their dominant tree species[26]. A second increase, after the 1980s, reflected a change in collecting strategy, with large expeditions targeting specific geographical areas within the scope of projects such as the '*Flora of the Guianas*', started in 1984[21]. At this moment 3744 species of trees have been collected in the Guianas (Appendix S2, Table 1) and using the estimate of expected species richness for the Guianas of 4581 species[11], 82% of the tree species have been collected locally.

Ecuador. The number of species increased sharply in the 1950s and collecting intensified in the 1970s (Fig. S6D) with the start of the '*Flora of Ecuador*' in 1973[27]. Jørgensen[18] estimated that 50,000 plant collections were made in Ecuador prior to 1900, while ten times that number were made in the 20th century. Collecting in Ecuador initially focused on the Pacific lowlands; the Amazonian lowlands were explored much later, most notably by a group of botanists associated with the National Herbarium (David Neill, Walter Palacios, Carlos Cerón, Alberto Dik). By 1993 the total number of plant collections in Ecuadorian Amazonia was estimated at 61,000[27]. In our dataset the total number of tree collections for Amazonian Ecuador is 16,202 (Appendix S2, Table S1). While this makes the region the fourth-best collected in Amazonia (Table 1), our data suggest that much more collecting will be needed to record the full tree flora. With 2366 tree species collected to date, just 35% of the estimated total number of tree species in the Ecuadorian Amazon have been recorded there.

Colombia. Between 1940 and 1950 botanical collecting increased (Fig. S6E). Richard Evans Schultes made important collections in the department of Vaupés, adding a great number of new species and new records to the Colombian flora. José Cuatrecasas visited the forests of the Orteguaza River (Caquetá), the Vaupés River, the Serranía de la Macarena (Meta) and the Serranía de la Lindosa (Guaviare), where he collected valuable botanical material. Between 1972 and 1979 the Project Radargramétrico del Amazonas (PRORADAM) generated important collections of timber species in the Colombian Amazon. In the 1990s many collections were made for ecological studies of Amazonian forests. The Herbario Amazónico Colombiano (COAH) del Instituto SINCHI, created in 1983 to focus on the country's Amazonian flora, now holds ~95,000 specimens and 7,700 species of vascular plants. A new catalogue of all plants of Colombia was published in 2015[28]. In our dataset the total number of tree collections for Amazonian Colombia is 35,277 (Appendix S2, Table 1). With 3511 tree species collected, 35% of the estimated number of tree species have been collected in the Colombian Amazon.

Brazil. Brazil is the country with the highest number of collections (278,165) for an average density 6.8 collections/100 km^2. The first peak of collections occurred very early in the century (Fig. S6F), when the herbarium of the Museu Paraense Emílio Goeldi em Belém (MG) was created. Later, many Amazonian expeditions organized by the Rio de Janeiro Botanical Garden contributed to the increase of collections thanks to Jacques Huber (1867–1914), João Geraldo Kuhlmann (1882–1958), and especially Adolpho Ducke (1876–1959)[29]. Nelson A. Rosa recorded more than 6,000 collections in Amazonia and João Murça Pires (1946–1962) collected in Amazonas and eastern Amazonia (close to 5,000 collections in the data used here) and also established the first permanent plots in upland forests[30]. George Alexander Black (1910–1957) collected all over Amazonia[31]. Black also established plots and made "*attempts to estimate species diversity and population density of trees in Amazonian forests*"[32]. Ghillean Prance is arguably the collector who made the most collections of trees in Amazonia (>20,000 in the data presented here). Much new material was collected during the '*Projeto Flora Amazonica*', which increased the number of collections in Amazonian herbaria by about 50%[33]. Also, the start of '*Flora Neotropica*' in 1967 helped spur an increase in collections and the discovery of new species in Brazilian Amazonia, as was the case with the monographs of Chrysobalanaceae and Lecythidaceae by Prance[6]. In our dataset the total number of tree collections for Amazonian Brazil is 278,165 (Appendix S2, Table 1). With 7696 tree species collected, 61% of the estimated number of tree species have been collected in the Brazilian Amazon.

Peru. Twentieth-century botany in Peru was initiated by the work of Augusto Weberbauer and Ignatz Urban, which led to an immediate increase in the number of species[19] (Fig. S6F). The '*Flora of Peru*' project was started in 1922 by Francis Macbride, who added greatly to Peruvian collections himself and attracted many other collectors. Specimens generated from this project provided the Field Museum herbarium (F) with one of the world's best overall collections of Peruvian plants; however, most of those collections have not yet been digitized. Collectors

from the Missouri Botanical Garden, especially Alwyn Gentry and Rodolfo Vásquez, also contributed large numbers of Amazonian tree collections in the period 1970–2000. In 1993 another major step was completed with the '*Catalogue of the flowering plants and gymnosperms of Perú*'[34]. In our dataset the total number of tree collections for Amazonian Peru is 39,851 (Appendix S2, 1). With 4422 tree species collected, 47% of the estimated number of tree species have been collected in the Peruvian Amazon.

Venezuela. Botanical collecting picked up in the 1940s and then increased steeply in the 1970s (Fig. S6G), yet Huber[20] considers lowland Venezuela to still be severely under-collected. The '*Flora of the Venezuelan Guayana*', initiated by Julian Steyermark in the early 1980s and completed under the guidance of Paul Berry, Kay Yatskievych, and Bruce Holst, has helped the latest increase in the cumulative number of species. In our dataset the total number of tree collections for Amazonian Venezuela is 16,356 (Appendix S2, 1). With 3189 tree species collected, 49% of the estimated number of tree species has been collected in the Venezuelan Amazon.

Bolivia. With just 2.4 tree collections per 100 km^2, Bolivia is by far the most under-collected Amazonian country for trees (Table 1). Remarkably, for every 1 tree specimen collected per unit area in Bolivia, 31 tree specimens have been collected in French Guiana (the best-collected country). This partly reflects a much later start in botanical exploration. At the beginning of the 20th century, four independent projects were undertaken but the flora of Bolivia received virtually no study between the 1920s and 1946[35]. In our dataset the total number of tree collections for Amazonian Bolivia is 11,721 (Appendix S2, 1). With 2404 tree species collected, 40% of the estimated number of tree species has been collected in the Bolivian Amazon.

Discussion

Nearly 12,000 tree species have been collected or observed in Amazonia to date. This makes the estimate of 16,000 total tree species for Amazonia by ter Steege *et al.*[11] seem entirely plausible, and supports their conclusion that non-parametric extrapolation methods are not very suitable for large areas with low sampling intensities (see also[36–38]). We acknowledge that the estimate of 16,000 remains speculative and is based on a debated sampling theory[39], but the logseries has not only been demonstrated empirically in a multiple of studies, but also derived by theoretical frameworks such as Neutral Theory[40], the application of Quantum Field Theory for Biogeography[41], and the Theory of Maximum Entropy in Ecology[42].

Trees have not been collected randomly across Amazonia with regard to time, space, or species identity (Fig. 1, Fig. S5). This is a well-known bias, which is understood to reduce the value of collection databases for describing patterns of diversity and planning conservation[6,43–45]. What we were surprised to find is that the estimated population size of a species in Amazonia is a poor predictor of the number of times it has been collected, and of the date it was first collected in Amazonia (Figs 2 and S3). This appears to be the result of how trees have been collected. Collecting started from a limited number of points, where many or perhaps most species, both common and rare, were collected. The fact that many collectors tend to maximize the number of species for a given number of collections[44] causes a strong disconnect between local abundance and local number of collections of tree species (Fig. 2). When new areas are visited the same pattern repeats itself. The data of one of the first Amazonian collectors, Aublet, is a good example of the efficiency of collectors. Aublet collected 211 tree species in French Guiana (Appendix S1). These species differ in estimated population size by 5 orders of magnitude and 12 of them have not been encountered in tree inventory plots in French Guiana. Because in most areas common as well as rare species are collected, range size affects the number of collections much more than abundance, since wide-ranging species are likely to be collected in many local inventories. This effect is so strong that after accounting for range size, population size did not add a significant amount of explained variation.

That plant collectors have recorded nearly 12,000 species with a collection effort of fewer than 10 collections per 100 km^2 likewise reflects their high efficiency at capturing rare taxa[44]. Ter Steege *et al.*[11] estimated that 11,000 species of Amazonian trees have populations of $\leq$1 million individuals, and that the rarest 5000 species are represented by $\leq$1000 individuals. The chance of randomly collecting a tree belonging to a species with $\leq$1 million individuals is $10^6/{\sim}4 * 10^{11} = 2.5 * 10^{-6}$. For species with populations of $\leq$100 individuals, the probability drops to $2.5 * 10^{-9}$. And these very low probabilities significantly overestimate the probability that a collector who finds the tree will collect it, since most collectors will only collect specimens if they have flowers or fruits (see below). It would seem that finding a needle in a haystack might be easier. Yet, nearly 7000 of these 11,000 rare species have been recorded to date, and are found in the herbarium record.

The species accumulation curve by collections (Fig. 6) makes it clear that more species will be found with additional collection effort. New species are also expected to arise from careful inspection of herbarium material[16,46,47], especially as more material is made available online through platforms such as GBIF and SpeciesLink.

Our list of tree species contains many species that rarely exceed 10 cm DBH. These species are necessarily rare in tree inventory plots (where 10 cm is the diameter cut-off) but they are not necessarily rare in Amazonia. A good example is *Hirtella racemosa*, a small tree which ranks #1 in number of collections (Appendix S1) but only #133 in estimated population size based on the ATDN plots[11], appendix S1. *Tococa guianensis* is another example, ranking #25 in number of collections and occasionally reaching 8 m in height in Amazonia (e.g., Douglas C. Daly 7148) but mostly smaller and never identified in the ATDN plots. The opposite is also true. Palms are extremely abundant in the ATDN plots, accounting for 6 of the ten tree species with the largest estimated populations[11], but are much less collected due to the enormous specimens they generate and the amount of work that this requires. Other factors that may influence the frequency of collecting are the phenology of species with regard to the timing of collecting trips, which are most often in dry periods[21,22]; irregular flowering and fruiting, which means that some species in tree plots are not found in fertile state despite frequent visits over 10 consecutive years

(D. Sabatier, personal observation); the fact that it is easier to collect understory trees than canopy trees; and the varying (in)conspicuousness of reproductive structures in different taxa[48]. Additional factors may include temporal fluctuations in taxonomic research funding, inventory campaigns in connection with new protected areas, environmental assessments, and floristic projects.

Finally, very rare species may actually be species common elsewhere; species from the Atlantic forest, Caatinga and Andes have all been collected rarely in Amazonia (Fig. S7). This may suggest that singleton species are mainly found on the edges of Amazonia. That is, however, not the case – they are more often found in rich areas with high collecting intensity (e.g. the area surrounding Manaus and French Guiana [Fig. S8]).

Data from public repositories contain various errors[8]. In the data used here at least five types of errors can be found: taxonomy, life-form, identification, location, and incomplete digitization. The third and fourth errors occur at the first stage of data collecting (i.e., at the individual stem level) and are very difficult to find and correct after data have been aggregated.

(1) Taxonomy: This error is independent of data collection. New taxonomic treatments of families and genera will move species currently on our checklist into synonymy. The impact on our list was relatively small in the case of a nomenclatural treatment of all Annonaceae (~3%, see above) but much greater in the full systematic treatment of the genus *Guatteria* (of the same family)[16]. Further study of specimens will also add to the checklist via the formal description of new species, as shown by the several new species added from IPNI but not yet present in the herbarium databases and the 10 new species of *Guatteria* described in[16]. The taxonomy of Amazonian tree families and genera is in a continuous state of flux and improving molecular techniques may cause great changes in the current classification and number of species, genera, and families. Recent phylogenetic analyses of Chrysobalanaceae[49], Protieae (Burseraceae)[50], and Lecythidaceae[51,52] show that our view of large Amazonian tree genera and families will change dramatically in the near future. Work on other large families is also ongoing (e.g. Annonaceae – Paul Maas, pers. comm.; Sapotaceae, Lauraceae - Jerome Chave, pers. comm.).
(2) Life form: When we were unsure whether a given species qualified as a tree by our definition, our main references were a recent checklist of South American trees[53] and the ATDN species list[11]. Other sources included the 'Tree flora of the Neotropical region'[54], tree checklists from TROPICOS (http://mobot.mobot.org/W3T/Search/vast.html) and we checked all these new species on their protologue and on herbarium labels for their habit. Because life form can be fluid within a species, it is possible that our checklist includes a small number of species that are primarily shrubs or lianas. We believe the amount of non-tree species in our list is quite small.
(3) Species identifications: An unknown but significant number of specimens in our collections dataset are incorrectly identified[55,56]. Incorrect identifications in herbaria have a tendency to snowball, as botanists match unidentified specimens to wrongly identified material. In some cases, duplicates of the same collection may receive different names in different herbaria. In public repository data such errors are difficult to check and may have slightly increased our species list.
(4) Location: Many digitized collections contain location errors, many of which we were able to correct. Errors in location data lead to errors in species richness prediction and tend to overestimate local richness as species are wrongly positioned outside their actual range. With many species these errors add up[8]. As we constructed one large list and checked quite substantially for errors in geography, removing data with dubious location, we believe this is not a major problem in our list.
(5) Incomplete digitization: Many large herbaria have not digitized their collections and deposited the data in GBIF, and their collections are only partially represented in our checklist. For example, most collections from the major expeditions outside the Guianas before 1900 were deposited in P and M (Paris and München), and records from these two herbaria were not present in GBIF when we downloaded our data. This bias might potentially explain why our dataset shows that early collecting was focused on the three Guianas (Figs S5 and S6). However, the patterns and species counts reported here for the period before 1900 are likely robust to this potential shortcoming, since most historical collectors made numerous duplicates that were widely distributed among or sold to herbaria and private collectors. Collections from the 20th century are probably less vulnerable to this type of bias, since they have more complete label data, label data in English, and were collected under more established herbarium practices that make digitization easier and more likely (including the sharing of duplicates with multiple herbaria).

We have shown that collecting of trees is not random in Amazonia, consistent with[44].The non-random nature of collections has strong implications for species distribution modelling, where the collections are assumed to be a random sample of an (unknown) probability distribution. The very effective collecting strategy of botanical collectors, stratified by geographical access, violates this assumption. This suggests that species distribution models for poorly sampled regions, such as Amazonia, may not only be compromised by the low number of observations[6] but also by the poor representativeness of their total ranges[5,57].

The accumulation of numbers of collections and species has slowed in the last 1–2 decades. This may reflect a number of factors. First, it may be largely artefactual: a result of the lag time between collecting a specimen and making its data available[47]. Specimens need to be identified, inserted in herbaria, databased, and in some cases described as new species, and then the data need to be uploaded to an online aggregator (such as SpeciesLink or GBIF). The latter is done on a regular basis by most large herbaria, but not on short time intervals and not by many smaller herbaria. Many undescribed species may also be found in recent[47] or older[46] collections. As an example, the 10 newly described *Guatteria* species in ref. 16 were collected for the first time in Amazonia between 1943 and 1991 (Appendix S1). Second, the decline could reflect a shift from general collection and plot work to

more phylogenetic research, which targets specific groups and does not lead to broad collections. Third, it may reflect a decrease in collecting expeditions due to increased difficulties in getting permits and funding. We suggest that the first option is most likely, as the lag-time is well established and the second and third options are more speculative - establishment of plots in Amazonia does not seem to show such decrease (Hans ter Steege, pers. obs.) and plant collection is done mainly by national researchers, who, in most of the countries, have permits linked to their institutes or research projects.

What are the most efficient strategies to complete the documentation of Amazonia's tree flora? Where, how, and how fast can we collect the species of Amazonian trees that remain undescribed? A simple extrapolation of the collection-species curve (Fig. S9) suggests that if the current logarithmic relationship between the number of species and collections remains valid, at least 4.5 million unique new collections will be required to discover all 16,000 species predicted to occur in Amazonia. At the current rate of collections per year (13,500), this will take over 300 years. And although deforestation is currently lower[58] than previously predicted[59], further loss of forest will inevitably lead to the loss of rare, small-ranged species before they can be discovered. If deforestation were to increase to levels of the early 2000s, most of the rare - and possibly unknown - species in eastern and southern Amazonia would face threat of extinction[58].

There is no doubt that intensifying collecting efforts in general would result in new species for science or new species records for Amazonia. But the probability that a botanist will collect a tree species not yet in our list is small and will continue to decline (see above). Intensifying collection effort therefore should not be restricted to collecting more specimens. We recommend focusing on data exchange and curation, digitization of specimens, support of taxonomic descriptions and monographs of families, in addition to field collecting that targets specific areas or taxa.

1) Digitize all Amazonian herbarium specimens. It has been estimated that perhaps 50% of all undescribed species are already present in natural history collections[46]. If this is true, then several hundred Amazonian tree species have been collected and deposited in herbaria but not yet 'discovered' or described, as was shown in the recent example of *Guatteria*[16]. Discovering those new species in herbaria is a slow process - because huge numbers of specimens remain overlooked by taxonomists. Digitization efforts combined with search engines will quickly inform taxonomists about the specimens of the focal taxon available in the different herbaria, helping the choice of material to study and facilitating the spotting of possible new species that should be studied more carefully. To resolve these problems, we recommend aggressive digitization of herbarium collections combined with data curation by specialists, such as proposed in Reflora (http://floradobrasil.jbrj.gov.br/). Naturalis's recent advances at the Naturalis Biodiversity Center in the Netherlands may serve as a model[60–62]. With a grant of €13 million, and over a period of two and a half years, 4.4 million herbarium specimens (and 4.1 million other specimens) were digitized at an average price of €1.52 per specimen[61]. In three automated 'herbarium digistreets' 22,000–24,000 herbarium sheets were imaged per day, with a record of 35,000 on a single day. Label data was then added into a database using the digital images. The digital material is available through Naturalis' Bioportal (http://bioportal.naturalis.nl/). A similar effort involving MNHN-Paris and IRD-Cayenne is being carried out in France as part of the e-ReColNat project (https://recolnat.org/) and will soon give higher visibility to historical Brazilian collections. This massive digitization should be coupled with equally intensive quality checking of the digital label data, ensuring quality for the geographical locations, typological correctness of the names of the species and collectors, etc.

2) Support and develop taxonomic and floristic expertise. Taxonomy seems a science of the past. Universities have largely abandoned natural history collections, together with the field of taxonomy[63,64]. In some cases this has led to a concentration of all natural history collections in one institute, as in the case of Naturalis[60]. However, unfortunately even natural history museums are focusing more and more on non-systematic research[65]. '*Taxonomy and systematics are NOT stamp collecting*'[66], however, and '*taxonomy can justly be called the pioneering exploration of life on a little known planet*'[67]. Taxonomy is non-linear hypothesis-driven research[64,68] and the hypotheses of species delimitations can only be made by taxonomists who have spent ample time studying a particular group. The inclusion of new techniques and the increase of known species have slowed this process instead of speeding it up, as more material needs to be consulted and more tests carried out[68]. As new species are found among the collections, primarily during family and genus revisions or monographs, and as these can only reliably be assessed by experts, we need to maintain and extend taxonomic expertise and train and support young taxonomists so that they can work for years on large and difficult groups. If new identifications can easily be passed on to the large data repositories[8], this would also improve the quality of digital information. The importance of flora projects in this process cannot be overstressed.

3) Accelerate and facilitate information exchange on Amazonian trees. All taxonomic literature on Amazonian trees (and in general) should be available online. Ideally all floras, revisions and monographs should be produced as e-documents with a standard format that can be updated easily. Existing alternatives are for example the Lecythidaceae pages of S.A. Mori (http://sweetgum.nybg.org/lp/) or the Scratchpad-based Lauraceae pages (http://lauraceae.myspecies.info/). All new descriptions should be online as fast as possible, either incorporated in the online e-monographs and e-floras or by being published in e-journals that also provide data exchange with e.g. IPNI[69]. Wen *et al.*[62] provide an outline for the workflow of e-monographs. It would be an asset for the Amazonian tree flora if 'Flora of Ecuador', 'Flora of the Guianas', 'Flora Neotropica,' and similar floras were online with electronic keys, links to collections, and image databases. Such floras democratize floristic and taxonomic information and are updatable 'on the fly'. It should be noted that, while e-floras and e-monographs will speed access to floristic and systematic information, it will not make taxonomic work much faster (see point 2).

We have started to amass all information known on Amazonian tree species. Based on the checklist in this paper, we created a platform (http://amazoniantreeflora.myspecies.info) where descriptions, images, DNA barcodes (GenBank, BOLD, new sequences) can be added by a growing and open community of Amazonian tree specialists. Ideally, we will assemble all collected species, all unique herbarium specimens, all species of the plots and we will have them sequenced. Identification tools and high-quality phylogenies of Amazonian tree species are also high priorities.

To aid in the identification of unidentified stems in the tree plots some of us (Daniel Sabatier and Julien Engel) are developing a platform (http://atdnmorphospecies.myspecies.info/) to share and manage images and information about "morphospecies" i.e. numbered species. These are supported by specimens expected to belong to valid but locally unknown and sometimes undescribed species. The site was developed first for French Guiana, to share information among the teams working in the same area, but will be modified make the platform useful for any tree plot contributor and for any part of Amazonia.

4) More focus on Amazonian research. Amazonia is still largely under-collected[3,4,6], except perhaps for the Guianas (Table 1, S1). In Brazil most research is concentrated in the southeast and south, which have 59 of the 92 Brazilian herbaria and 67% of all Brazilian collections. Brazilian Amazonia, while covering half of Brazil, has just five registered herbaria containing just 11% of the botanical collections[70]. Also, current systematic research focuses on non-Amazonian areas. Of the 2875 new Brazilian angiosperm species described from 1990 to 2006, only 20% were Amazonian. While close to 50% of the new species occurring outside Amazonia were described by Brazilian researchers, the figure for Amazonian species was only 20%. Just 0.9% of research funding from the Brazilian government between 1997 and 2002 was dedicated to Amazonia[70].

5) Target geographic areas where collection effort is low and expected diversity is high. Although many new species may be found in herbaria (point 1), collecting is still needed and may be a limiting factor as in other groups[66]. Where should botanists focus their efforts, in order to avoid having to collect 4,500,000 specimens? Probably areas and with both high diversity and low collecting density. Hopkins[6] suggested four main regions where botanical knowledge is especially weak in Amazonia, but where biodiversity is expected to be high (1. lowland Colombia, centred in the area comprising parts of the departments of Vichada, Meta, Guainia, Guaviare and Vaupes; 2. western Amazonian Brazil, within the state of Amazonas, approximately between the cities of Tefé and Envira, comprising the interfluvial region between the Rio Purus and the Rio Juruá , and extending north of the Amazon River as far as the Jaú National Park; 3. northern Amazonian Brazil, extending from north-east Amazonas State across southern Roraima and the portion of Pará State about 300 km north of the Amazon River, and including the southern extremity of Guyana; 4. south-eastern Amazonian Brazil, extending from the southeast corner of Amazonas State (headwaters of the Rio Sucunduri and middle course of the Rio Aripuanã) and southern Pará State, especially the upper reaches of the Rio Irirí and Rio Curuá.). Based on diversity (Fig. S10B), collection density (Fig. S10A, Table 1), level of botanical knowledge (Fig. S10C) and completeness of the botanical survey (Table S1), we suggest that collecting should focus on three areas: 1) the border area of Brazil and the Guianas, 2) the large interfluves of southwestern Brazilian Amazonia, and 3) the eastern slopes and foothills of the Andes (Ecuador, Peru, Bolivia), where many new species are collected on nearly every botanical expedition (Robin Foster pers. comm.). In addition to targeting specific areas, efficiency in finding the remaining species might greatly improve if collection efforts were taxon-targeted. Focusing collecting efforts specifically on species that are not identifiable in the field and ignoring common species, would increase the likelihood of finding the missing species. We also need to keep collecting more common species, however, to widen the geographic range of collections as herbarium specimens rather than published observations are the only primary data that can be checked for correctness of identification[71]. Rare species will remain difficult to find. Hopkins[6] suggests that well executed florula projects[29], accompanied by consistent long-term collecting will result in the discovery of many rare species. However, even for Reserva Ducke we estimate (ter Steege *et al.* in prep), that over 2600 tree species will be found, while less than 2000 have been collected to date.

6) Embrace new technologies. Aublet would have been surprised to learn that most botanists still search for new plant species the same way he did in 1762. If we are to locate the rarest Amazonian tree species before they go extinct, we need to incorporate new technology, experimentation, and other emerging methods that allow us to search larger areas with greater precision. Botanists may need to spend more time in the air, or in the canopy, than on the ground in the field. One example of an emerging method for large-scale plant identification is spectranomics[72], which in one study was able to identify 85% of 1449 species in Peru based on a 20-trait spectranomic signature for leaf reflectance. If the authors can take the next step and collect those signatures from near-remote-sensing platforms, it may become possible to map large numbers of individual trees and identify undescribed trees with unknown signatures. DNA-Barcoding, phylogenetic research, near infrared spectroscopy[73,74], and other techniques will also help to identify species or discover new species through an alternative way other than monographing, yet combined with taxonomic expertise. Many of these techniques require new collection methods, which is why every Amazonian tree specimen collected from this point on should be accompanied by silica gel-dried leaf samples.

Conclusions and prospects. We have shown that recent estimates of Amazonian tree diversity based on statistical extrapolation from a logseries[11] are fully plausible: close to 12,000 of the ~16,000 species estimated to occur in the region have been recorded to date. The vast majority of this discovery has taken place in the last century, thanks to the effort of thousands of botanists from all over the world, flora projects, collecting expeditions, and the establishment of tree inventory plots. Finding the remaining species and better documenting all species'

distributions constitute a major scientific challenge of Amazonian botany. We argue that the most time- and cost-efficient strategy to tackle this problem is to expand the collecting effort to understudied sites, followed by taxon-focused collection campaigns and the application of new technologies. The remaining work in discovering the tree diversity of Amazonia will have far-reaching effects in ecology, systematics, conservation biology and evolution. This research program will be costly but needs to be put into perspective, as proposed by Hubbell[75]: '*we need far better data on the geographic ranges and abundances of tropical tree species to finally put the "how many species?" question to rest. It seems to me that our priorities are misplaced. We spend many billions of dollars to look for extra-terrestrial life but far less to understand life and its distribution on our own planet*'.

Methods

Data sets. All plant collection records for the nine Amazonian countries available at GBIF (http://www.gbif.org; 1.5 million records) and SpeciesLink (http://splink.cria.org.br/; 300,000 records) were retrieved in May 2014. We complemented these two large datasets with all collection records from the 2015 institutional databases of three primarily Amazonian herbaria (INPA, MG, and COAH; 170,000; 180,000, and 28,000 records respectively). We also included a previously compiled dataset of plant collections in all herbaria in Guyana, Suriname, and French Guiana, containing data from these three countries; at the time of the study these records had not been added to GBIF[22] (190,000 records). All families and genera that do not contain any tree species were removed. A tree was defined as a species with a woody trunk of at least 10 cm diameter at breast height (DBH, where breast height = 1.30 m). Tree-like species that may have a trunk but that are not always taken into account when establishing tree plots (e.g., tree ferns, *Montrichardia*, *Phenakospermum*, bamboos, *Zamia*) were removed. Hereafter we refer to this first database as the collections dataset.

In order to determine which names in the collections dataset are considered taxonomically valid tree species, we independently produced a list of all known tree species (including synonyms) for South America, based on three literature sources[11,53,54]. We checked the spelling and synonymy of all ~56,000 names gathered in this preliminary tree species list with the Taxonomic Name Resolution Service (http://tnrs.iplantcollaborative.org/TNRSapp.html), the Plant List (http://www.theplantlist.org/), and the Brazilian flora checklist (http://reflora.jbrj.gov.br), to achieve a standard taxonomy (e.g., APG3 plant families) and yield a list of all South American trees synchronized with the ATDN (http://atdn.myspecies.info/) plot database[11]. This required the merging of *Crepidospermum* and *Tetragastris* into *Protium*[50]. Hereafter we refer to this second database as the South American trees reference list, or reference list.

All names in the collections dataset were compared with the corrected names in the South American trees reference list. Species in the collections dataset that matched a species in the reference list were considered a tree species with a valid name or a proper synonym and added to a third database, which is the preliminary checklist of Amazonian trees we present in Appendix S1 (hereafter the checklist). Species in the collections dataset that did not match any species in the reference list were checked against TNRS and then re-matched with the reference list. After this automated procedure, remaining problems were assessed one by one and solved by literature search and with the Plantminer (http://www.plantminer.com/), which mines various sources through the Plant List. If after inspection some of the unmatched tree species names in the collections dataset were indeed correct, they were added to the reference list and to the checklist. All doubtful names (unresolved, illegitimate, invalid) were checked against the original sources where possible and added to the checklist if they had been published correctly and removed if not. Annonaceae names were standardized to the most recent index to this family[15] and *Guatteria* was standardized to the most recent monograph of the genus[16].

In order to keep only the Amazonian collections in our collections dataset, we checked all coordinates, corrected them where possible, and removed doubtful coordinates. We accepted as Amazonian all collections located inside Amazonia[11], as well as all collections without coordinates, but arguably made inside Amazonia based on country (Guyana, Suriname, French Guiana, all collections) or departments that are fully or mainly Amazonian (e.g. Pará, Rondônia, Roraima and Amapá, in Brazil; Amazonas, in all countries).

At this point, the new collections dataset consisted of all Amazonian herbarium records identified to a valid species name. However, because a given herbarium record may occur in GBIF, SpeciesLink, and our herbarium databases, our collections dataset contained a large number of duplicate records. To select unique collections we pulled out all unique combinations of species, collection number, year, country, and province. While this may have omitted some collections that lack a collection number, we believe this number is relatively small. Importantly, the method retained all species names in the data.

During the analyses we noticed that several tree species listed in recent checklists[22,54] and catalogues of Ecuador, Peru and Bolivia (www.tropicos.org) were not in our South American trees reference list. We added these species after checking the names and locations as described above. Finally, we consulted the International Plant Name Index (http://www.ipni.org/) for all species that were newly added to IPNI from 1990 to 2015 in the nine Amazonian countries (9998 species) and added those tree species that we could reliably attribute to Amazonia. The checklists and new species from IPNI added species with no occurrence records to our checklist. For these and species with very few records (together ~1000 species) we searched GBIF once more on a species-by-species basis and added unique record locations within Amazonia. Species for which we found no collection or plot data were added to the checklist but were not included in the numerical analyses, except in the analysis to compare range sizes with the number of collections.

Data analysis. For species confirmed to occur in the Amazon Tree Diversity Network (ATDN) of 1170 tree plots established across the region, we included in the checklist how many plots it occurred in, how many stems occur in those plots, and estimated Amazonian population size[11] (see Appendix S1). We used simple regression to test the effect of a species' estimated population size and range size (which we here defined as the number of ATDN tree plots in which a species was found) on the number of times it had been collected and the earliest date

it was collected in Amazonia, to test the hypothesis that more abundant and widespread species are collected more often and earlier than rarer species. Finally we produced a list of all unique collections by country to serve as country checklists. For these lists we only used the collection data. All tests and graphing the data were carried out with custom scripts in R[76].

Maps. All maps were created with custom scripts in R[76]. The geographic information (country.shp and river.shp) are shape files from ESRI (http://www.esri.com/data/basemaps, © Esri, DeLorme Publishing Company).

References

1. Da Silva, J. M. C., Rylands, A. B. & da Fonseca, G. A. B. The Fate of the Amazonian Areas of Endemism. *Conservation Biology* **19,** 689–694, doi: 10.1111/j.1523-1739.2005.00705.x (2005).
2. Lewinsohn, T. M. & Prado, P. I. How Many Species Are There in Brazil? *Conservation Biology* **19,** 619–624, doi: 10.1111/j.1523-1739.2005.00680.x (2005).
3. Sousa-Baena, M. S., Garcia, L. C. & Peterson, A. T. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* **20,** 369–381, doi: 10.1111/ddi.12136 (2014).
4. Schulman, L., Toivonen, T. & Ruokolainen, K. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography* **34,** 1388–1399, doi: 10.1111/j.1365-2699.2007.01716.x (2007).
5. Feeley, K. 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, doi: 10.1371/journal.pone.0125629 (2015).
6. Hopkins, M. J. G. Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography* **34,** 1400–1411, doi: 10.1111/j.1365-2699.2007.01737.x (2007).
7. Feeley, K. J. & Silman, M. R. The data void in modeling current and future distributions of tropical species. *Global Change Biology* **17,** 626–630, doi: 10.1111/j.1365-2486.2010.02239.x (2011).
8. Maldonado, C. *et al.* Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography* **24,** 973–984, doi: 10.1111/geb.12326 (2015).
9. Beck, J., Boller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* **19,** 10–15, doi: 10.1016/j.ecoinf.2013.11.002 (2014).
10. Vale, M. M. & Jenkins, C. N. Across-taxa incongruence in patterns of collecting bias. *Journal of Biogeography* **39,** 1744–1748, doi: 10.1111/j.1365-2699.2012.02750.x (2012).
11. ter Steege, H. *et al.* Hyperdominance in the Amazonian tree flora. *Science* **342,** 1243092, doi: 10.1126/science.1243092 (2013).
12. Harte, J. & Kitzes, J. Inferring regional-scale species diversity from small-plot censuses. *PLoS One* **10,** e0117527, doi: 10.1371/journal.pone.0117527 (2015).
13. Ricklefs, R. E. How tree species fill geographic and ecological space in eastern North America. *Ann Bot* **115,** 949–959, doi: 10.1093/aob/mcv029 (2015).
14. Fung, T., Villain, L. & Chisholm, R. A. Analytical formulae for computing dominance from species-abundance distributions. *J Theor Biol* **386,** 147–158, doi: 10.1016/j.jtbi.2015.09.011 (2015).
15. Maas, P. J. M., Westra, L. Y. T., Rainer, H., Lobao, A. Q. & Erkens, R. H. J. An updated index to genera, species, and infraspecific taxa of Neotropical Annonaceae. *Nordic Journal of Botany* **29,** 257–356, doi: 10.1111/j.1756-1051.2011.01092.x (2011).
16. Maas, P. J. M. *et al.* Confronting a morphological nightmare: revision of the Neotropical genus Guatteria (Annonaceae). *Blumea* **60,** 1–219, doi: 10.3767/000651915X690341 (2015).
17. Aublet, J. B. C. F. *Histoire des plantes de la Guiane françoise*. (P. F. Didot jeune, 1775).
18. Jørgensen, P. M. In *Catalogue of the vascular plants of Ecuador* (eds P.M. Jørgensen & S. León-Yánez) (Missouri Botanical Garden, 1999).
19. Jørgensen, P. M., Ulloa, C. & Maldonado, C. In *Botánica Económica de los Andes Centrales* (eds M. R. Moraes *et al.*) (Universidad Mayor de San Andrés, 2006).
20. Huber, O. In *Flora of the Venezuelan Guayana*. Vol. Volume 1: Introduction (eds P. E. Berry, B. K. Holst & K. Yatskievych) 63–95 (Smithsonian Institution Press, 1995).
21. Haripersaud, P., ter Steege, H., de Granville, J. J., Chevillotte, H. & Hoff, M. Species abundance, distribution and diversity in time and space after centuries of botanical collecting in the Guianas. *Taxon* **59,** 592–597 (2010).
22. Haripersaud, P. H. *Botanical collecting in the Guianas*, PhD Thesis, Utrecht University (2009).
23. Anonymous. *Brasilianische Reise 1817–1820: Carl Friedrich Philipp von Martius zum 200er Geburtstag herausgegeben von Jörg Helbig*. (Hirmer Verlag, 1994).
24. Förther, H. Die Geschichte des Martius-Herbariums: seine Brasilienkollektion und Empfehlungen zur Typuswahl. *Sendtnera* **2,** 5–24 (1994).
25. Lindeman, J. C. & Mennega, A. M. W. *Bomenboek voor Suriname: herkenning van Surinaamse houtsoorten aan hout en vegetatieve kenmerken*. (Utrecht University/Landsbosbeheer, 1963).
26. Fanshawe, D. B. *The Vegetation of British Guyana. A Preliminary Review*. (Imperial Forestry Institute, 1952).
27. Renner, S. S. *A history of botanical exploration in Amazonian Ecuador, 1739–1988*. Vol. 82 (Smithsonian Institution Press, 1993).
28. Bernal, R., Gradstein, S. R. & Celis, M. *Catálogo de plantas y líquenes de Colombia*, 2015).
29. Hopkins, M. J. G. Flora da Reserva Ducke, Amazonas, Brasil. *Rodriguésia* **56,** 9–25 (2005).
30. Pires, J. M. & Salomão, R. P. Dinâmica da Diversidade Arbórea de um Fragmento de Floresta Tropical Primária na Amazônia Oriental - 1. Período: 1956 a 1992. *Boletim do Museu Paraense Emílio Goeldi. Botânica* **16,** 63–110 (2000).
31. Prance, G. T. An Index of Plant Collectors in Brazilian Amazonia. *Acta Amazonica* **1,** 25–65 (1971).
32. Black, G. A., Dobzhansky, T. & Pavan, C. Some Attempts to Estimate Species Diversity and Population Density of Trees in Amazonian Forests. *Botanical Gazette* **111,** 413–425, doi: 10.1086/335612 (1950).
33. Prance, G. T., Nelson, B. W., Silva, M. F. d. & Daly, D. C. Projeto Flora Amazônica: eight years of binational botanical expeditions. *Acta Amazonica* **14,** 5–30 (1984).
34. Brako, L. & Zarruchi, J. L. Catalogue of the flowering plants and gymnosperms of Perú. *Monographs in Systematric Botany* **45,** 1–1286 (1993).
35. Funk, V. A. & Mori, S. A. *A bibliography of plant collectors in Bolivia*. Vol. 70 (Smithsonian Institution Press, 1989).
36. Brose, U., Martinez, N. D. & Williams, R. J. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* **84,** 2364–2377, doi: 10.1890/02-0558 (2003).
37. Wang, J.-P. & Lindsay, B. G. Wang J-PZ, Lindsay BG. A penalized nonparametric maximum likelihood approach to species richness estimation. *J Am Statist Assoc*. **100,** 942–959 (2005).
38. Xu, H., Liu, S. R., Li, Y. D., Zang, R. G. & He, F. L. Assessing non-parametric and area-based methods for estimating regional species richness. *Journal of Vegetation Science* **23,** 1006–1012, doi: 10.1111/j.1654-1103.2012.01423.x (2012).
39. Fisher, R. A., Corbet, A. S. & Williams, C. B. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *The Journal of Animal Ecology* **12,** 42, doi: 10.2307/1411 (1943).
40. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. (Princeton University Press, 2001).

41. O'Dwyer, J. P. & Green, J. L. Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecol Lett* **13,** 87–95, doi: 10.1111/j.1461-0248.2009.01404.x (2010).
42. Harte, J. Maximum Entropy and Ecology. *A Theory of Abundance,Distribution, and Energetics*. (Oxford University Press, 2011).
43. Nelson, B. W., Ferreira, C. A. C., Dasilva, M. F. & Kawasaki, M. L. Endemism Centers, Refugia and Botanical Collection Density in Brazilian Amazonia. *Nature* **345,** 714–716, doi: 10.1038/345714a0 (1990).
44. ter Steege, H., Haripersaud, P. P., Banki, O. S. & Schieving, F. A model of botanical collectors' behavior in the field: never the same species twice. *Am J Bot* **98,** 31–37, doi: 10.3732/ajb.1000215 (2011).
45. Tobler, M., Honorio, E., Janovec, J. & Reynel, C. 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. *Biodiversity and Conservation* **16,** 659–677, doi: 10.1007/s10531-005-3373-9 (2007).
46. Bebber, D. P. *et al.* Herbaria are a major frontier for species discovery. *Proc Natl Acad Sci USA* **107,** 22169–22171, doi: 10.1073/pnas.1011841108 (2010).
47. Fontaine, B., Perrard, A. & Bouchet, P. 21 years of shelf life between discovery and description of new species. *Current biology: CB* **22,** R943–944, doi: 10.1016/j.cub.2012.10.029 (2012).
48. ter Steege, H. & Persaud, C. A. The phenology of Guyanese timber species: a compilation of a century of observations. *Vegetatio* **95,** 177–198 (1991).
49. Bardon, L. *et al.* Origin and evolution of Chrysobalanaceae: insights into the evolution of plants in the Neotropics. *Botanical Journal of the Linnean Society* **171,** 19–37, doi: 10.1111/j.1095-8339.2012.01289.x (2013).
50. Fine, P. V., Zapata, F. & Daly, D. C. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* **68,** 1988–2004, doi: 10.1111/evo.12414 (2014).
51. Huang, Y. Y., Mori, S. A. & Kelly, L. M. Toward a phylogenetic-based Generic Classification of Neotropical Lecythidaceae-I. Status of Bertholletia, Corythophora, Eschweilera and Lecythis. *Phytotaxa* **203,** 85–121, doi: 10.11646/phytotaxa.203.2.1 (2015).
52. Mori, S. A. *et al.* Toward a Phylogenetic-based Generic Classification of Neotropical Lecythidaceae-II. Status of Allantoma, Cariniana, Couratari, Couroupita, Grias and Gustavia. *Phytotaxa* **203,** 122–137, doi: 10.11646/phytotaxa.203.2.2 (2015).
53. Grandtner, M. M. & Chevrette, J. *Dictionary of Trees, Volume 2: South America. Nomenclature, Taxonomy and Ecology*. (Academic Press, 2013).
54. Oliveira Filho, A. T. NeoTropTree: Tree flora of the Neotropical Region: A database involving biogeography, diversity and conservation, 2014).
55. Dexter, K. G., Pennington, T. D. & Cunningham, C. W. Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter? *Ecological Monographs* **80,** 267–286, doi: 10.1890/09-0267.1 (2010).
56. Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. & Scotland, R. W. Widespread mistaken identity in tropical plant collections. *Current biology: CB* **25,** R1066–1067, doi: 10.1016/j.cub.2015.10.002 (2015).
57. Loiselle, B. A. *et al.* Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* **0,** 070908043732003-???, doi: 10.1111/j.1365-2699.2007.01779.x (2007).
58. Ter Steege, H. *et al.* Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci Adv* **1,** e1500936, doi: 10.1126/sciadv.1500936 (2015).
59. Soares-Filho, B. S. *et al.* Modelling conservation in the Amazon basin. *Nature* **440,** 520–523, doi: 10.1038/nature04389 (2006).
60. Anonymous. Uit het depot, op het web. *Twee eeuwen nationaal natuurhistorisch erfgoed in het digitale domein*. (Naturalis Biodiversity Center, Leiden, 2015).
61. Heerlien, M. *et al.* The Natural History Production Line: An Industrial Approach to the Digitization of Scientific Collections. *Acm J Comput Cult He* **8,** 1–11, doi: 10.1145/2644822 (2015).
62. Wen, J., Ickert-Bond, S. M., Appelhans, M. S., Dorr, L. J. & Funk, V. A. Collections-based systematics: Opportunities and outlook for 2050. *Journal of Systematics and Evolution* **53,** 477–488, doi: 10.1111/jse.12181 (2015).
63. Raven, P. H. Taxonomy: where are we now? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359,** 729–730, doi: 10.1098/rstb.2004.1462 (2004).
64. Wheeler, Q. D. Taxonomic triage and the poverty of phylogeny. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359,** 571–583, doi: 10.1098/rstb.2003.1452 (2004).
65. Funk, V. A. The Erosion of Collections-Based Science: Alarming Trend or Coincidence? *Plant Press* **17** (2014).
66. May, R. M. Tomorrow's taxonomy: collecting new species in the field will remain the rate-limiting step. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359,** 733–734, doi: 10.1098/rstb.2003.1455 (2004).
67. Wilson, E. O. Taxonomy as a fundamental discipline. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359,** 739, doi: 10.1098/rstb.2003.1440 (2004).
68. Sluys, R. The unappreciated, fundamentally analytical nature of taxonomy and the implications for the inventory of biodiversity. *Biodiversity and Conservation* **22,** 1095–1105, doi: 10.1007/s10531-013-0472-x (2013).
69. Penev, L., Kress, W. J., Knapp, S., Li, D. Z. & Renner, S. Fast, linked, and open - the future of taxonomic publishing for plants: launching the journal PhytoKeys. *PhytoKeys*, 1–14, doi: 10.3897/phytokeys.1.642 (2010).
70. Sobral, M. & Stehmann, J. R. An analysis of new angiosperm species discoveries in Brazil (1990-2006). *Taxon* **58,** 227–232 (2009).
71. Schilthuizen, M., Vairappan, C. S., Slade, E. M., Mann, D. J. & Miller, J. A. Specimens as primary data: museums and 'open science'. *Trends Ecol Evol* **30,** 237–238, doi: 10.1016/j.tree.2015.03.002 (2015).
72. Asner, G. P. *et al.* Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *The New phytologist* **204,** 127–139, doi: 10.1111/nph.12895 (2014).
73. Lang, C., Costa, F. R., Camargo, J. L., Durgante, F. M. & Vicentini, A. Near Infrared Spectroscopy Facilitates Rapid Identification of Both Young and Mature Amazonian Tree Species. *PLoS One* **10,** e0134521, doi: 10.1371/journal.pone.0134521 (2015).
74. Durgante, F. M., Higuchi, N., Almeida, A. & Vicentini, A. Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. *Forest Ecology and Management* **291,** 240–248, doi: 10.1016/j.foreco.2012.10.045 (2013).
75. Hubbell, S. P. Estimating the global number of tropical tree species, and Fisher's paradox. *Proc Natl Acad Sci USA* **112,** 7343–7344, doi: 10.1073/pnas.1507730112 (2015).
76. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2011).

Acknowledgements

We thank thousands of naturalists, botanists, ecologists and all others who collected thousands of collections of trees and the hundreds of herbaria that made their data available through GBIF and SpeciesLink. Special thanks to Mike Hopkins who made available the latest version of the collection database of INPA. HtS and RS are supported by grant 407232/2013-3 - PVE - MEC/MCTI/CAPES/CNPq/FAPs; AA is supported by grants from the Swedish Research Council, the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement n. 331024), and a Wallenberg Academy Fellowship. We thank Miles Silman and one anonymous reviewer for very constructive comments that helped to improve the manuscript.

Author Contributions

H.t.S. conceived and designed the study. R.V. carried out GBIF and SpeciesLink data collection. R.V. and H.t.S. carried out error checking and correction of primary input data. P.J.M. and N.P. contributed Tropicos checklists. All co-authors contributed to checking and re-checking of the species list and all co-authors contributed to the writing.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: ter Steege, H. *et al.* The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.* **6**, 29549; doi: 10.1038/srep29549 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/