

**Ecography**

**ECOG-05026**

Stropp, J., Umbelino, B., Correia, R. A., Campos-Silva, J. V., Ladle, R. J. and Malhado, A. C. M. 2020. The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon. – Ecography doi: 10.1111/ecog.05026

**Supplementary material**

## **Appendix 1**

### **1. Description of data filtering**

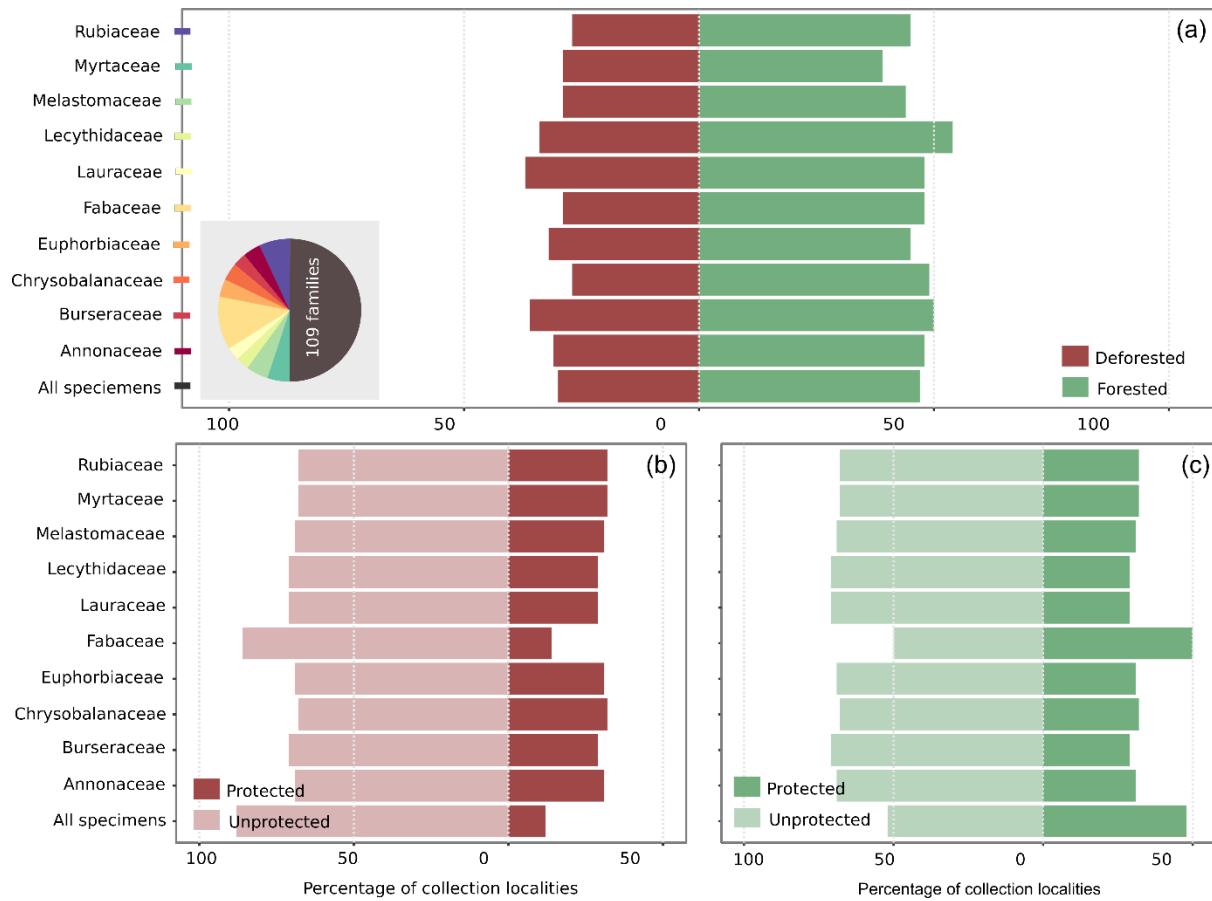
Our combined dataset initially contained 399,147 herbarium specimens of 7,383 tree species. We screened this dataset and flagged specimens holding 1) uncertain geographical coordinates, and 2) a missing and/or uncertain date of collection and 3) duplicate specimens. This filtering led to a dataset containing 129,252 specimens of 5,750 tree species (see electronic supplementary material for details on data filtering).

We considered coordinates as uncertain if 1) the decimals of latitude and longitude contained only zeros, 2) the information provided in the field ‘country’ did not refer to Brazil and 3) the specified latitude and longitude coincided with a city or village (i.e. places classified as village, city, or capital by the IBGE). We then verified the date of collection by assessing the plausibility of values given in the field “eventDate”. All specimens with a date of collection referring to after May 2018 - the date of data download - were classified as errors. We also considered as errors a year of collection previous to 1500. Furthermore, we considered specimens bearing uncertain date of collection if they were collected between 1600 and 1899, as early collected specimens more frequently bear an incorrect year of collection (see Supplementary Information in da Costa et al. 2019). Although this approach may flag specimens with a correct year of collection as uncertain, we think the effect on our results is negligible because the number of botanical collections in the Amazon was very low prior to 1900 (ter Steege et al. 2016). We also flagged duplicate specimens, i.e., specimens holding identical species name, geographical coordinates, and date of collection. We identified duplicates by comparing the species names of specimens with the species names as standardized by TNRS (Boyle et al. 2013). Identical geographical coordinates were assessed after rounding the original information on latitude and longitude stored in our dataset to three decimals, which is equivalent to an accuracy of 111 meters. Our filtering resulted in a final dataset of 129,252 specimens of 5,750 tree species. Data filtering was performed by using customized *R* scripts and functions of the package *scrubr* (Chamberlain 2017).

**Table A1.** List of tree species and their respective number of specimens, species conservation status, species sub-region, total number and percentage of specimens collected at currently protected or deforested localities, and loads of the first two axes of FAMD analysis. Available at <https://figshare.com/s/076f8e31341382010fba>

**Table A2.** Significance values of pairwise comparisons between groups using Wilcoxon rank sum test. Groups were identified by the Hierarchical Clustering on Principal Components (HCPC) of the Factorial Analysis for Mixed Data (FAMD).

	EA, SA, WAS & Vulnerable or endangered	EA, SA, WAS & Not assigned	GS, CA, WAS & Not assigned	GS, CA, WAS & Not threatened
EA, SA, WAS & Vulnerable or endangered	-			
EA, SA, WAS & Not assigned	0.180	-		
GS, CA, WAS & Not assigned	<0.005	<0.005	-	
GS, CA, WAS & Not threatened	<0.005	<0.005	<0.005	-



**Fig. A1.** The percentage of collections localities of Amazonian tree specimens placed in areas that are still forested (green) and that were deforested by 2017 (red) (a). Panel (b) depicts the percentage of collection localities that were deforested and that are placed in protected (dark red) or unprotected (light red) areas. Panel (c) depicts the percentage of collection localities that are still covered by forest and are placed in protected (dark green) or unprotected (light green) areas. The pie chart in panel (a) shows the percentage of specimens belonging to the ten most abundant families (color gradient from purple to green (50% of the total number of specimens) and to the other 109 families.

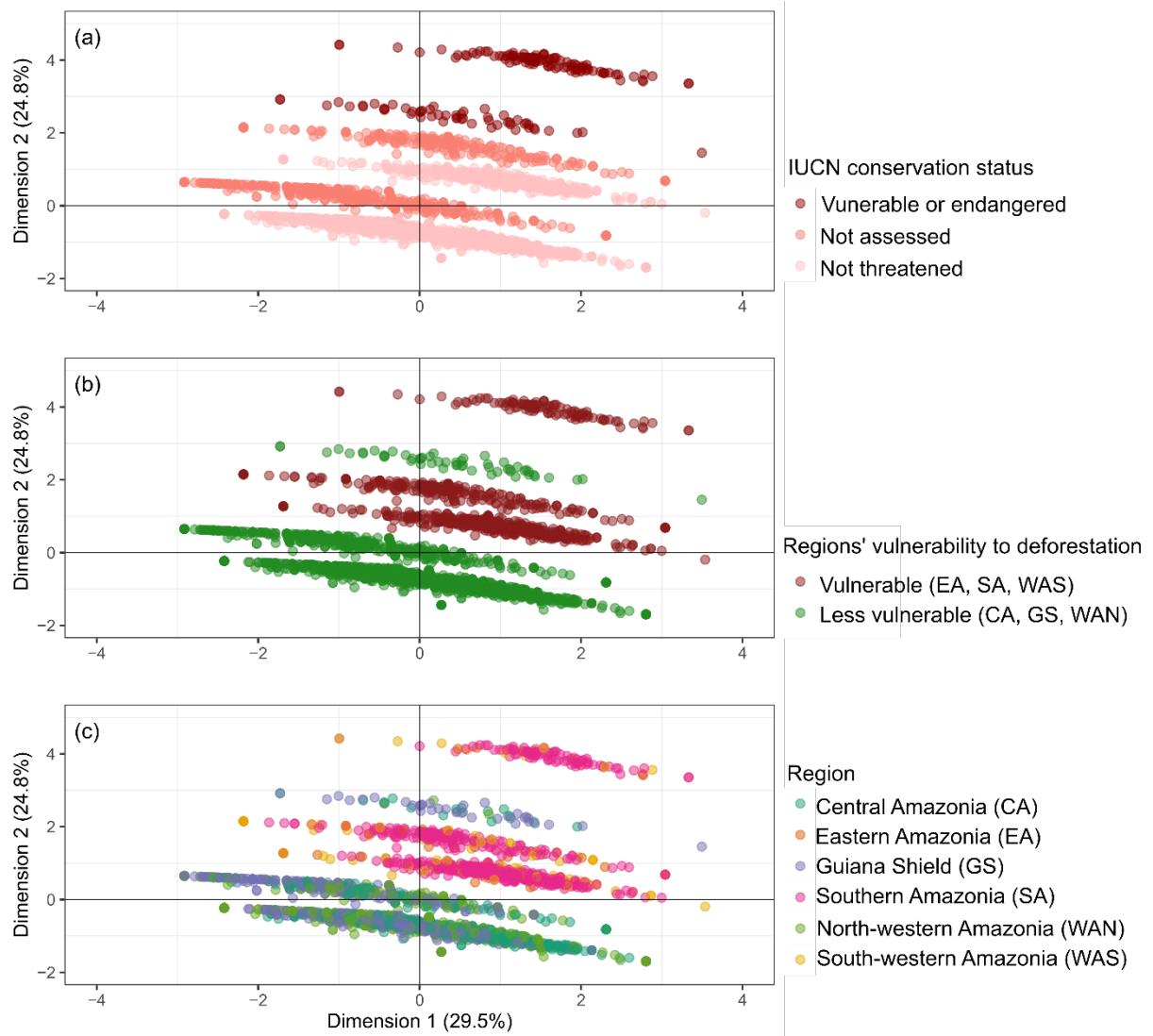
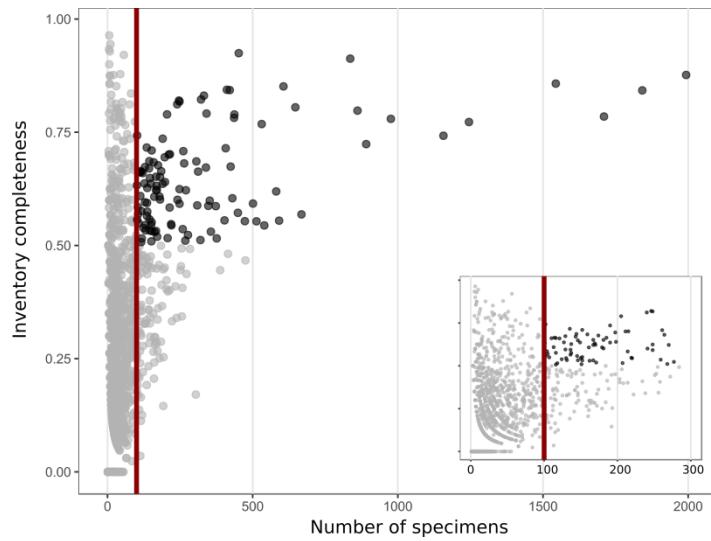


Fig. A2. Scatterplot of the Factor Analysis for Mixed Data (FAMD) Hierarchical Clustering on Principal Components (HCPC). Each dot represents a tree species. In panels (a) and (b) colours represent categorical variables included in the FAMD analysis. In panel (b) colours represent the region in which among six Amazonian regions a species occurs according to Gomes et al. (2019). The list of 3469 tree species, their respective attributes included and scores for the first and second dimension, and is given in Table A1.



**Fig. A3.** Relationship between inventory completeness obtained as the complementary value of the slope at the last point smoothed of species accumulation curves and number of specimens (i.e. 1-slope). Well-sampled cells (dark grey) were define as those possessing at least 100 specimens (red line) and inventory completeness  $\geq 0.5$  ( $N = 120$ ), whereas under-sampled cells were considered as those possessing at least 100 inventory completeness  $< 0.5$  ( $N = 3,446$ ). Cells with less than 100 specimens ( $N = 3,342$ ) may contain spurious values of inventory completeness and therefore were not included in our analysis (see inlet graph). Each dot in the graph represents a grid cell of 25 km  $\times$  25 km; one grid cell containing 10,677 records and inventory completeness equals 0.97 is not shown.

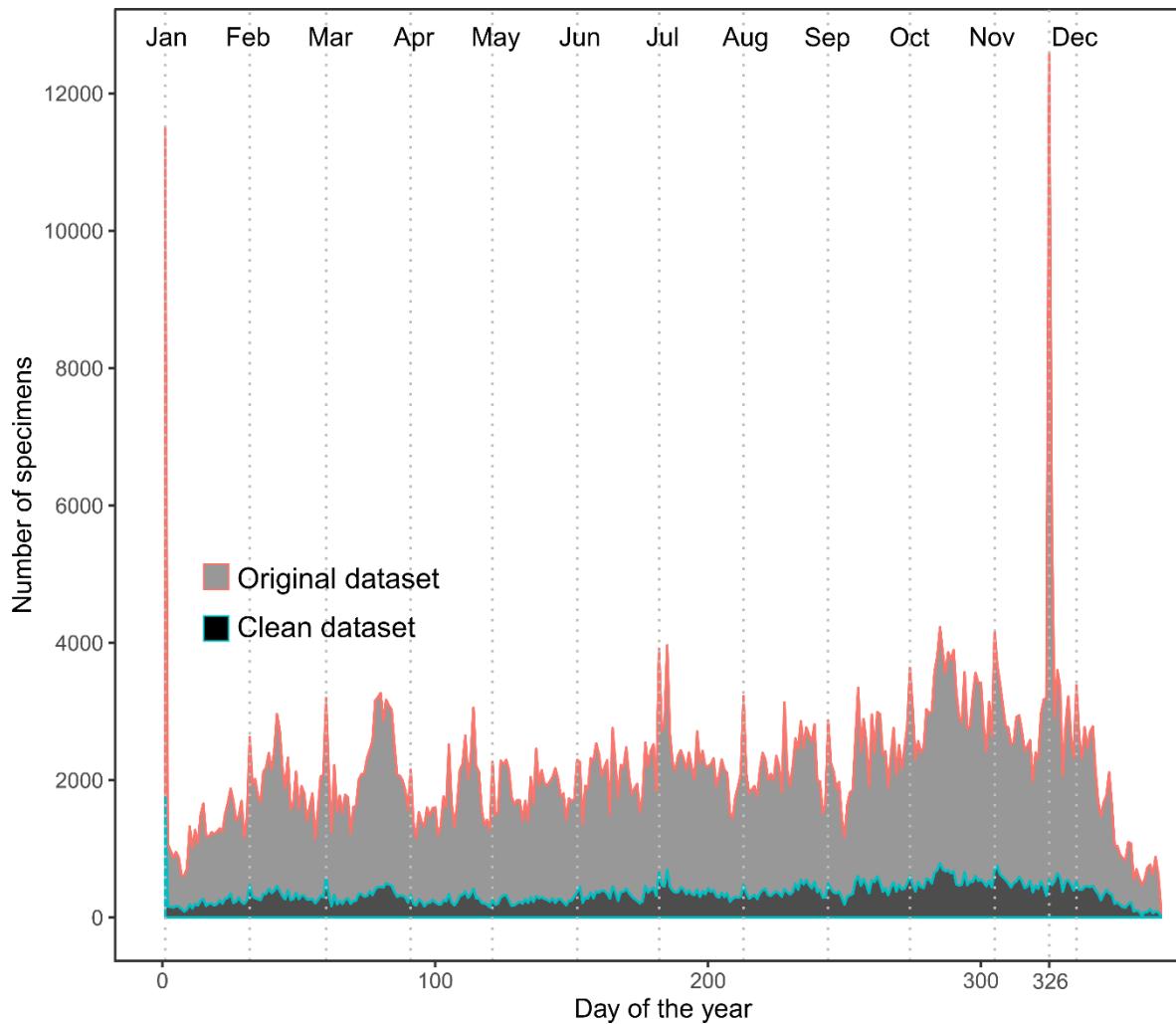


Fig. A4. Change in date of collection of digital specimens stored in our original and clean datasets. Spikes often correspond to the first day of each month, with a disproportionately large spike for the first of January (in both datasets). Spikes on the first day of each month is partially due to incomplete date of collection that are often recorded as the first day of the month (see Groom et al. 2019). The spike on the day 326 (21<sup>st</sup> of November) in the original dataset is likely due to typographic errors on the date of collection. Out of the 12,041 records that had the 21<sup>st</sup> of November as their day and month of collection, only 555 were kept on our clean dataset. Ten thousand records recorded as collected on this day were flagged by our data filtering as having uncertain date of collection (year of collection after download date), 595 were flagged as bearing invalid species name, 1247 as uncertain geographic coordinates, and 3518 as duplicated specimens.

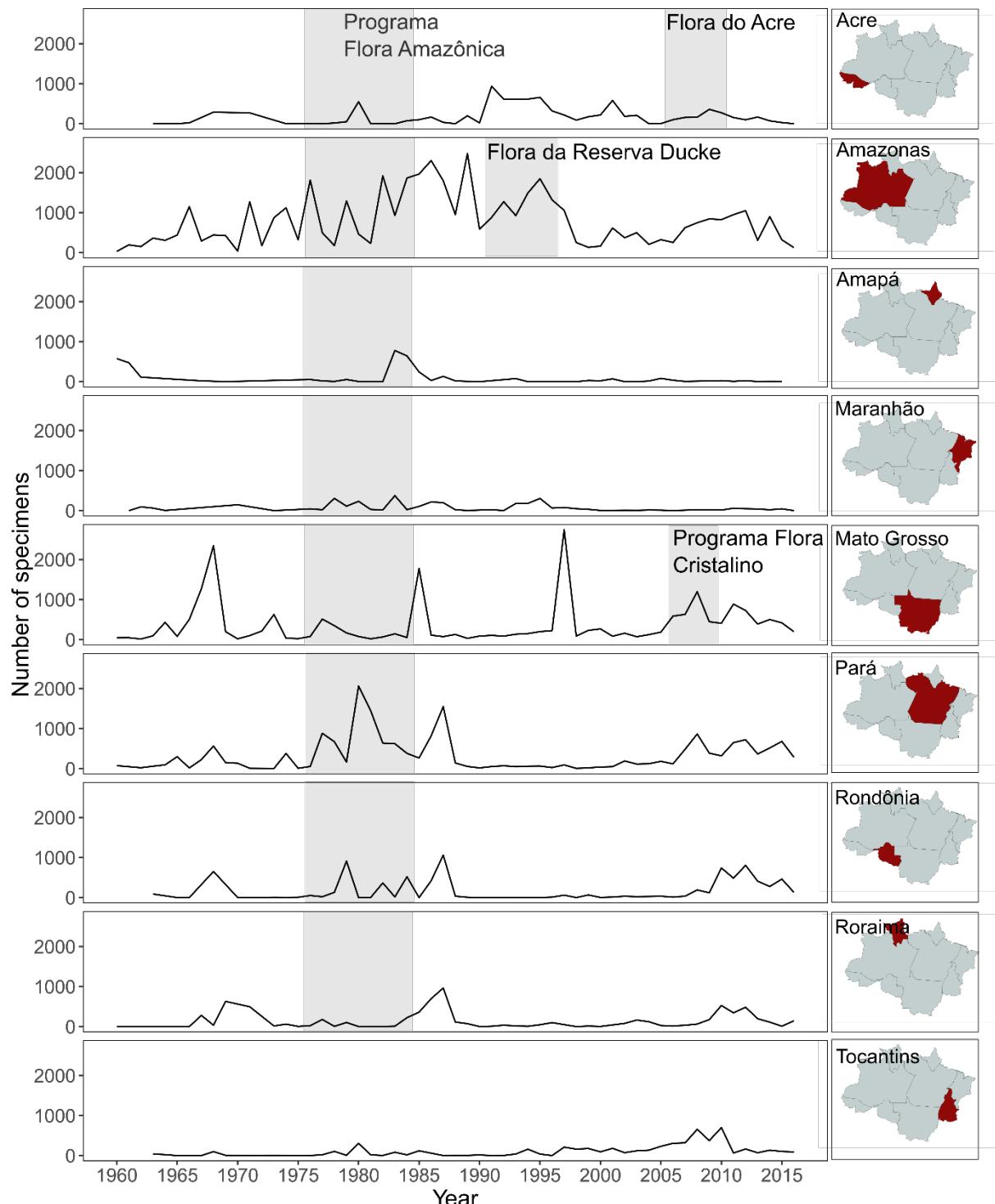


Fig. A5. Number of tree specimens collected between Number of tree specimens and species collected each year between 1960 and 2017 and 2017 in each of the nine states of the Brazilian Amazon.

**Complete citation of occurrence data retrieved from SpeciesLink**

Herbário Alexandre Leal Costa (ALCB), Herbário da Universidade Federal de Sergipe (ASE), Arizona State University Vascular Plant Herbarium (ASU-Plants), Herbarium Berolinense (B), Herbário Antônio Nonato Marques (BAH), Xiloteca Calvino Mainieri(BCTw), Herbário da Universidade Federal de Minas Gerais (BHCB), Herbário UFMG - Samambaias e Licófitas (BHCB-SL), Herbário do Jardim Botânico da Fundação de Parques Municipais e Zoobotânica (BHZB), Brazilian Laboratory of Agrostology(BLA), Herbário Irina Delanova Gemtchújnicov (BOTU), Xiloteca "Profa. Dra. Maria Aparecida Mourão Brasil" (BOTUw), CAS Botany (BOT) (CAS-BOT), Herbário da Embrapa Recursos Genéticos e Biotecnologia (CEN), Herbário Leopoldo Krieger(CESJ), Herbário da Fundação Universidade Federal de Mato Grosso do Sul (CGMS), Herbário Centro Norte Mato Grossense(CNMT), Convolvulaceae of Brazil (Convolvulaceae\_BR), Herbário da Universidade Federal de Mato Grosso do Sul, Campus Pantanal (COR), Herbário CPAP da Embrapa Pantanal (CPAP), Coleção de plantas medicinais e aromáticas (CPMA), Herbário Pe. Dr. Raulino Reitz (CRI), Coleção de Plantas Vivas do Jardim Botânico da Fundação de Parques Municipais e Zoobotânica(CVJBFZB), Herbário da Reserva Natural Vale (CVRD), Duke University Herbarium (DUKE), Herbário da Universidade Tecnológica Federal do Paraná - Dois Vizinhos (DVPR), Royal Botanic Garden Edinburgh Herbarium (E), Herbário Prisco Bezerra (EAC), Herbário do Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (EAFM), Herbário Jaime Coelho de Moraes (EAN), Herbário Escola de Florestas Curitiba (EFC), Herbário da Escola Superior de Agricultura Luiz de Queiroz(ESA), Herbário ESAL (ESAL), Evaldo Buttura (EVB), Field Museum of Natural History - Brazilian records (F), Herbário Friburguense (FCAB), Coleção Botânica de Plantas Medicinais (Fiocruz-CBPM), University of Florida Herbarium (FLAS)(FLAS), Herbário do Departamento de Botânica da Universidade Federal de Santa Catarina (FLOR), Fototeca Mauricio Mercadante (FMM), Fundación Puerto Rastrojo (FPR-Colombia), Fototeca Paulo Schwirkowski (FPS), Robert K. Godfrey Herbarium (FSU), Herbário da Universidade Estadual de Londrina (FUEL), Coleção Ficológica do Herbário da Universidade Estadual de Londrina (FUEL-Algae), Coleção de lâminas de grãos de pólen (Funed-Pol), Herbário Dr. Roberto Miguel Klein(FURB), Fototeca Vinícius Dittrich (FVD), Geneva Herbaria Catalogue

with species Brazil (G), Geneva Herbaria Catalogue with species Brazil - De Candolle herbarium (G-DC), Herbário Amapaense (HAMAB), Herbário Alarich Rudolf Holger Schultz (HAS), Coleção Ficológica do HAS (HAS-Algae), Herbário do Instituto de Estudos Costeiros da Universidade Federal do Pará (HBRA), Herbário Virtual Flora Brasiliensis (HbVirtFlBras), Herbário Caririense Dárdano de Andrade-Lima (HCDAL), Herbário da Universidade Tecnológica Federal do Paraná Campus Campo Mourão (HCF), Herbário do Departamento de Ciências Florestais(HDCF), Herbário Delta do Parnaíba (HDELTA), Herbário Dendrológico Jeanine Felfili (HDJF), Herbário Ezechias Paulo Heringer (HEPH), Herbário da Amazônia Meridional (HERBAM), Herbário Dr. Ary Tupinambá Penna Pinheiro (HFSL), Herbário de Ilha Solteira (HISA), Herbario Jataiense Prof. Germano Guarim Neto (HJ), Herbário de Montes Claros (HMC), Herbário do Pantanal "Vali Joana Pott" (HPAN), Herbário Padre Balduino Rambo (HPBR), Herbário do Jardim Botânico Plantarum (HPL), Herbário do Museu de Ciências Naturais da PUC-Minas (HPUC-MG), Herbário Rioclarense (HRCB), Herbário Sérgio Tavares(HST), Herbário da Universidade Federal do Oeste do Pará (HSTM), Herbário do Trópico Semiárido (HTSA), Carpoteca do Trópico Semiárido (HTSA-Carpoteca), Herbário da Universidade Estadual do Centro-Oeste (HUCO), Herbário da Pontifícia Universidade Católica do Paraná (HUCP), Herbário da Universidade Católica de Pernambuco (HUCPE), Herbário da Universidade de Caxias do Sul (HUCS), Liqueenoteca do Herbário da Universidade de Caxias do Sul (HUCS-Liqueenoteca), Micoteca do Herbário da Universidade de Caxias do Sul (HUCS-Micoteca), Herbario da Universidade Estadual de Feira de Santana (HUEFS), Herbário da Universidade Estadual de Goiás (HUEG), Herbário UEM (HUEM), Herbário da Universidade Estadual do Sudoeste da Bahia (HUESB), Herbário da Universidade Federal de São João Del Rei (HUFSJ), Herbarium Uberlandense (HUFU), Herbário Mogiense (HUMC), Herbário da Universidade do Estado da Bahia (HUNEB), Herbário da Universidade Estadual de Ponta Grossa (HUPG), Herbário do Recôncavo da Bahia (HURB), Herbário Unisanta (HUSC), Herbário da Universidade de Tocantins (HUTO), Herbário Vale do São Francisco (HVASF), Xiloteca do Herbário Vale do São Francisco (HVASFw), Herbário do Vale do Taquari (HVAT), Herbário do Instituto SENAI de Tecnologia em Meio Ambiente(HXBH), Herbário do Instituto Agronômico de Campinas (IAC), Herbário do Instituto de

Ciências Naturais (ICN), Herbário de Rio Verde (IFRV), Herbário INPA (INPA), Carpoteca INPA (INPA-Carpoteca), Coleção de Madeiras - Xiloteca INPA (INPAw), Herbário - IPA Dárdano de Andrade Lima (IPA), Herbário do Parque da Ciência Newton Freire Maia (IRAI), Herbário Joinvillea(JOI), Xiloteca Joinvillea (JOIw), Herbário Lauro Pires Xavier (JPB), Laboratório de Botânica e Ecologia Vegetal (LABEV), Laboratório de Fitoplâncton (LabFito), Orquidário do Laboratório de Biologia Molecular e Biossistemática de Plantas (LBMBP), Herbário de Lages da Universidade do Estado de Santa Catarina (LUSC), Herbarium of the Botanische Staatssammlung München (M), Herbário do Instituto do Meio Ambiente do Estado de Alagoas (MAC), Herbário MACK (MACK), Herbário do Maranhão (MAR), Herbário do Museu Botânico Municipal (MBM), Herbário Mello Leitão (MBML-Herbario), Herbário Norte Mineiro (MCCA), Herbário Profª. Drª. Marlene Freitas da Silva (MFS), Herbário Profª. Drª. Marlene Freitas da Silva - Coleção de Flores (MFS-Flores), Herbário Profª. Drª. Marlene Freitas da Silva - Coleção de Frutos (MFS-Frutos), Herbário Profª. Drª. Marlene Freitas da Silva - Coleção de Plântulas (MFS-Plantulas), University of Michigan Herbarium (MICH), Herbário do Museu Integrado de Roraima (MIRR), Missouri Botanical Garden - Brazilian records (MO), Herbário Dárdano de Andrade Lima(MOSS), Herbário do Museu da Pontifícia Universidade Católica do Rio Grande do Sul (MPUC), Botanical Collections (NHM-London-BOT), Herbário Nova Xavantina (NX), The New York Botanical Garden - Brazilian records (NY), OBIS Brasil(OBIS\_BR), Herbário "Professor José Badini" (OUPR), MNHN - Herbário Virtual A. de Saint-Hilaire (P), Herbarium Anchieta(PACA-AGP), Herbarium Anchieta - Aloysio Sehnem (PACA-Bryophytes), Herbário do Parque Estadual do Rio Doce (PERD), Herbário Professor Vasconcelos Sobrinho (PEUFR), Herbário do Museu Nacional (R), Herbário do Departamento de Botânica da Universidade Federal Rural do Rio de Janeiro (RBR), Herbário do Museu Nacional - Criptogamos (R-Criptogamos), Coleção Biológica Realeza (REAL), RECOLNAT - Herbário Virtual A. Glaziou (RECOLNAT\_Glaziou), Herbário Rondoniense (RON), Herbário do Museu Nacional - Tipos (R-Tipos), Sistema de Informação do Programa Biota/Fapesp (SinBiota), Herbário de São José do Rio Preto (SJRP), Herbário de algas de São José do Rio Preto (SJRP-Algae), Herbário de Bryophyta de São José do Rio Preto (SJRP-Bryophyta), Herbário de Pteridophyta de São José do Rio Preto (SJRP-Pteridophyta),

Herbário Rosa Mochel(SLUI), Herbário da Universidade Federal de Santa Maria (SMDB), Solanaceae  
Source - a taxonomic resource for the nightshade family (Solanaceae\_Source\_BR), Herbário do Centro  
de Ciências e Tecnologias para a Sustentabilidade (SORO), Herbário do Estado "Maria Eneyda P.  
Kaufmann Fidalgo" - Coleção de Fanerógamas (SP), Herbário do Estado "Maria Eneyda P. Kaufmann  
Fidalgo - Coleção de Algas (SP-Algae), Maria Eneyda P. Kauffman Fidalgo (SP-Bryophyta), Herbário da  
Universidade de São Paulo (SPF), Herbário da Universidade de São Paulo – Coleção de Algas (SPF-  
Algae), Herbário do Departamento de Biologia da FFCLRP-USP (SPFR), Maria Eneyda Pacheco  
Kauffman Fidalgo (SP-Fungi), Xiloteca do Instituto de Biociências da Universidade de São Paulo  
(SPFW), Herbário Dom Bento José Pickel (SPSF), Herbário Tangará(TANG), Herbário Graziela Barroso  
(TEPB), Herbário da Universidade de Brasília (UB), Herbário da Universidade Estadual de Campinas  
(UEC), Herbário Universidade Estadual de Santa Cruz (UESC), Herbário da Universidade Federal do  
Acre(UFACPZ), Herbário da Universidade Federal de Goiás (UFG), Herbário UFMT (UFMT), Herbário  
UFP - Geraldo Mariz (UFP), Carpoteca UFP (UFP-Carpoteca), Herbário da Universidade Federal de  
Roraima (UFRR), Herbário da Universidade Estadual do Oeste do Paraná (UNOP), Herbário da  
Universidade Estadual do Oeste do Paraná (UNOP-Algae), Herbário da Universidade Federal do Paraná  
(UPCB), Smithsonian Department of Botany - Brazilian records (US), Herbário da Universidade Federal  
de Viçosa (VIC), Herbário Central da Universidade Federal do Espírito Santo VIES (VIES) disponível na  
rede speciesLink(<http://www.splink.org.br>) em 15 de Maio de 2018 às 07:24.