sPlot open - An environmentally-balanced, open-access, global dataset of vegetation plots

This manuscript is still work in progress

This manuscript ([permalink]) was automatically generated from [fmsabatini/sPlotOpen_Manuscript@47aa8e1] on October 22, 2020.

Authors

Francesco Maria Sabatini[1,2], Jonathan Lenoir[3], Tarek Hattab[4], Elise Arnst[5], Milan Chytrý[6], Jürgen Dengler[7,8,9], Valério De Patta Pillar[10], Patrice De Ruffray[11], Stephan M. Hennekens[12], Ute Jandt[2], Florian Jansen[13], Borja Jiménez-Alfaro[14], Jens Kattge[15], Aurora Levesley[16], Oliver Purschke[17], Brody Sandel[18], Fahmida Sultana[19], Marten Winter[9], Svetlana Aćić[20], Emiliano Agrillo[21], Miguel Alvarez[22], Iva Apostolova[23], Mohammed A.S. Arfin Khan[24], Isabelle Aubin[25], Yves Bergeron[26], Erwin Bergmeier[27], Idoia Biurrun[28], Anne D. Bjorkman[29], Laura Casella[30], Luis Cayuela[31], Tomáš Černý[32], Victor Chepinoga[33], János Csiky[34], Renata Ćušterevska[35], Els De Bie[36], Michele De Sanctis[21], Panayotis Dimopoulos[37], Mohamed Abd El-Rouf Mousa El-Sheikh[38,39], Brian Enquist[40], Manfred Finckh[41], Emmanuel Garbolino[42], Melisa Giorgis[43], Valentin Golub[44], Alvaro G. Gutierrez[45], Mohamed Z. Hatim[46], Norbert Hölzel[47], Jürgen Homeier[48], Adrian Indreica[49], Kim Sarah Jacobsen[50], John Janssen[12], Birgit Jedrzejek[51], Norbert Jürgens[41], Zygmunt Kącki[52], Ali Kavgacı[53], Michael Kessler[54], Andrey Korolyuk[55], Hjalmar Kühl[9,56], Flavia Landucci[6], Ching-Feng Li[57], Hongyan Liu[58], Tatiana Lysenko[59], Corrado Marcenò[28], Jesper Erenskjold Moeslund[60], Jonas V. Müller[61], Jérôme Munzinger[62], Jalil Noroozi[63], Arkadiusz Nowak[64], Viktor Onyshchenko[65], Gerhard E. Overbeck[66], Aníbal Pauchard[67], Robert K. Peet[68], Aaron Pérez-Haase[69,70], Tomáš Peterka[6], Gwendolyn Peyre[71], Oliver L. Phillips[16], Vadim Prokhorov[72], Valerijus Rašomavičius[73], Rasmus Revermann[41], John S. Rodwell[74], Eszter Ruprecht[75], Solvita Rūsiņa[76], Cyrus Samimi[77], Joop H.J. Schaminée[12], Marco Schmidt[78], Urban Šilc[79], Željko Škvorc[80], Anita Smyth[81], Zvjezdana Stančić[82], Zhiyao Tang[58], Ioannis Tsiripidis[83], Milan Valachovič[84], Kim André Vanselow[85], Kiril Vassilev[23], Eduardo Vélez-Martin[86], Roberto Venanzoni[87], Alexander Christian Vibrans[88], Risto Virtanen[9,89,90], Henrik von Wehrden[91], Viktoria Wagner[92], Donald A. Walker[93], Desalegn Wana[94], Karsten Wesche[9,95,96], Timothy Whitfeld[97], Wolfgang Willner[98], Susan Wiser[5], Thomas Wohlgemuth[99], Sergey Yamalov[100], Helge Bruelheide[1,2],

✉ — To whom correspondence should be addressed: francesco.sabatini@botanik.uni-halle.de

1. German Centre for Integrative Biodiversity Research (iDiv) - Halle-Jena-Leipzig. Germany
2. Martin-Luther University Halle-Wittenberg, Institute of Biology, Am Kirchtor 1, 06108, Halle, Germany
3. Unité de Recherche "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN), UMR 7058 CNRS, Université de Picardie Jules Verne, 80037 Amiens Cedex 1, France
4. MARBEC, University of Montpellier, CNRS, IFREMER and IRD, Sète, France
5. Manaaki Whenua – Landcare Research, PO Box 69040, 7640, Lincoln, New Zealand
6. Masaryk University, Department of Botany and Zoology, Kotlářská 2, 611 37, Brno, Czech Republic
7. Zurich University of Applied Sciences (ZHAW), Vegetation Ecology Group, Institute of Natural Resource Sciences (IUNR), Grüentalstr. 14, 8820, Wädenswil, Switzerland
8. University of Bayreuth, Plant Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), Universitätsstr. 30, 95447, Bayreuth, Germany
9. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103, Leipzig, Germany
10. Federal University of Rio Grande do Sul, Ecology, Av. Bento Gonçalves 9500, 91501-970, Porto Alegre, Brazil
11. IBMP, 12, rue du Général-Zimmer, 67084, Strasburg, France
12. Wageningen University and Research, Wageningen Environmental Research (Alterra), P.O.Box 47, 6700 AA, Wageningen, Netherlands
13. University of Rostock, Faculty of Agricultural and Environmental Sciences, Justus-von-Liebig-Weg 6, 18059, Rostock, Germany
14. University of Oviedo, Research Unit of Biodiversity (CSIC/UO/PA), C. Gonzalo Gutiérrez Quirós s/n, 33600, Mieres, Spain
15. Max Planck Institute for Biogeochemistry, Hans Knöll Str. 10, 07745, Jena, Germany
16. University of Leeds, School of Geography, Woodhouse Lane, LS2 9JT, Leeds, United Kingdom

17. NA,
18. Aarhus University, Aarhus, Denmark
19. Shahjalal University of Science & Technology, Forestry & Environmental Science, 3114, Sylhet, Bangladesh
20. Faculty of Agriculture, Department of Agrobotany, Nemanjina 6, 11080, Belgrade-Zemun, Serbia
21. Sapienza University of Rome, Department of Environmental Biology, P.le Aldo Moro 5, 00185, Rome, Italy
22. University of Bonn, Plant Nutrition, INRES, Karlrobert-Kreiten-Str., 53115, Bonn, Germany
23. Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Department of Plant and Fungal Diversity and Resources, Acad. Georgi Bonchev 23, 1113, Sofia, Bulgaria
24. Shahjalal University of Science & Technology, Forestry & Environmental Science, Akhalia, 3114, Sylhet, Bangladesh
25. Canadian Forest Service, Natural Resources Canada, Great Lakes Forestry Centre, 1219 Queen St. East, P6A 2E5, Sault Ste Marie (Ontario), Canada
26. Université du Québec en Abitibi-Témiscamingue, Forest Research Institute, 445 boul. de l'Université, J9X5E4, Rouyn-Noranda, Canada
27. University of Göttingen, Vegetation Ecology and Phytodiversity, Untere Karspüle 2, 37073, Göttingen, Germany
28. University of the Basque Country UPV/EHU, Plant Biology and Ecology, P.O. Box 644, 48080, Bilbao, Spain
29. Aarhus University, Section for Ecoinformatics & Biodiversity, Department of Bioscience, Ny Munkegade 114, 8000, Aarhus C, Denmark
30. ISPRA - Italian National Institute for Environmental Protection and Research, Biodiversity Conservation Department, Via Vitaliano Brancati, 60, 00144, Roma, Italy
31. Universidad Rey Juan Carlos, Department of Biology, Geology, Physics and Inorganic Chemistry, c/ Tulipán s/n, 28933, Móstoles, Spain
32. Czech University of Life Sciences Prague, Department of Forest Ecology, Faculty of Forestry and Wood Sciences, Kamýcká 1176, 165 21, Praha 6 - Suchdol, Czech Republic
33. V.B. Sochava Insitute of Geography SB RAS, Laboratory of Physical Geography and Biogeography, Ulan-Batorskaya, 1, 664033, Irkutsk, Russian Federation
34. University of Pécs, Department of Ecology, Ifjúság u. 6., 7624, Pécs, Hungary
35. Faculty of Natural Sciences and Mathematics, Institute of Biology, Arhimedova 3, 1000, Skopje, Republic of Macedonia
36. Research Institute for Nature and Forest (INBO), Departement of Biodiversity and Natural Environment, Havenlaan 88, bus 73, 1000, Brussels, Belgium
37. University of Patras, Institute of Botany, Division of Plant Biology, Department of Biology, University Campus, 26504, Patras, Greece
38. College of Science, King Saud University, Botany and Microbiology Department, P.O. Box 2455, 11451, Riyadh, Saudi Arabia
39. Damanhour University, Botany Department, Faculty of Science, Damanhour, Egypt
40. University of Arizona, Ecology and Evolutionary Biology, 1041 E. Lowell St., AZ 85721, Tucson, United States
41. University of Hamburg, Biodiversity, Ecology and Evolution of Plants, Institute for Plant Science & Microbiology, Ohnhorststr. 18, 22609, Hamburg, Germany
42. MINES ParisTech, Crisis and Risk research Centre (CRC), 1 rue Claude Daunesse, BP 207, 06904, Sophia Antipolis, France
43. Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET), ECOLOGÍA VEGETAL Y FITOGEOGRAFÍA, Av. Vélez Sársfield 1611, 5000, Córdoba, Argentina
44. Institute of Ecology of the Volga River Basin, Laboratory of Phytocoenology, Komzina, 10, 445003, Toljatty, Russia
45. Universidad de Chile, Departamento de Ciencias Ambientales y Recursos Naturales Renovables, Facultad de Ciencias Agronomicas, Santa Rosa 11315, La Pintana, 8820808, Santiago, Chile
46. Tanta University, Botany, Faculty of Science, El Geish St., 31527, Tanta, Egypt
47. University of Muenster, Institute of Landscape Ecology, Heisenbergstr. 2, 48149, Münster, Germany
48. University of Göttingen, Plant Ecology and Ecosystems Research, Untere Karspüle 2, 37073, Göttingen, Germany
49. Transilvania University of Brasov, Department of Silviculture, Sirul Beethoven 1, 500123, Brasov, Romania
50. Ghent University, Dept Environment, Coupure Links 653, 9000, Ghent, Belgium
51. University of Münster, Institute of Landscape Ecology, Heisenbergstr. 2, 48149, Münster, Germany
52. University of Wrocław, Botanical Garden, Sienkiewicza 23, 50-335, Wrocław, Poland
53. Soutwest Anatolia Forest Research Institute, Silviculture and Forest Botany, POB 264, 07002, Antalya, Turkey
54. University of Zurich, Department of Systematic and Evolutionary Botany, Zollikerstrasse 107, 8008, Zurich, Switzerland
55. Central Siberian Botanical Garden, Siberian Branch, Russian Academy of Sciences, Geosystem Laboratory, Zolotodolinskaya str. 101, 630090, Novosibirsk, Russian Federation
56. Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Primatology, Deutscher Platz 6, 04103, Leipzig, Germany
57. National Taiwan University, School of Forestry and Resource Conservation, No. 101, Section 2, Kuang-Fu Road, 30013, Hsinchu, Taiwan
58. Peking University, College of Urban and Environmental Sciences, Yiheyuan Rd. 5, 100871, Beijing, China
59. Institute of Ecology of the Volga River Basin RAS, Dept. of the Phytodiversity Problems, Komzin str. 10, 445003, Togliatti, Russia
60. Aarhus University, Department of Bioscience, Grenaavej 14, 8410, Roende, Denmark
61. Royal Botanic Gardens, Kew, Conservation Science, Wakehurst Place, RH17 6TN, Ardingly, West Sussex, United Kingdom
62. IRD, CIRAD, CNRS, INRA, Université Montpellier, AMAP - Botany and Modelling of Plant Architecture and Vegetation, Boulevard de la Lironde, 34398, Montpellier, France
63. University of Vienna, Department of Botany and Biodiversity Research, Rennweg 14, 1030, Vienna, Austria
64. Polish Academy of Sciences, Botanical Garden - Center for Biological Diversity Conservation, Prawdziwka 2, 02-976, Warszawa, Poland

65. National Academy of Sciences of Ukraine, M.G. Kholodny Institute of Botany, Tereshchenkivska 2, 01601, Kyiv, Ukraine
66. Universidade Federal do Rio Grande do Sul, Department of Botany, Av. Bento Gonçalves 9500, 91501-970, Porto Alegre, Brazil
67. University of Concepción, Laboratorio de Invasiones Biológicas (LIB), Victoria 631, 4030000, Concepción, Chile
68. University of North Carolina, Department of Biology, CB3280, South Road, 27599-3280, Chapel Hill, NC, United States
69. University of Barcelona, Department of Evolutionary Biology, Ecology and Environmental Sciences, Diagonal 643, 08028, Barcelona, Spain
70. Center for Advanced Studies of Blanes, Spanish Research Council (CEAB-CSIC), Continental Ecology, Carrer d'accés a la Cala St. Francesc, 14, 17300, Blanes, Girona, Spain
71. University of the Andes, Department of Civil and Environmental Engineering, Carrera 1 Este No. 19A-40, Edificio Mario Laserna, Piso 6 , 111711, Bogota, Colombia
72. Kazan Federal University, Institute of Environmental Sciences, Kremlevskaya 18, 420008, Kazan, Russia
73. Nature Research Centre, Institute of Botany, Zaliuju Ezeru 49, 08406, Vilnius, Lithuania
74. NA, 7 Derwent Road, LA1 3ES, Lancaster, United Kingdom
75. Babeș-Bolyai University, Hungarian Department of Biology and Ecology, Faculty of Biology and Geology, Republicii street 42., 400015, Cluj-Napoca, Romania
76. University of Latvia, Department of Geography, 1 Jelgavas Street, 1004, Riga, Latvia
77. University of Bayreuth, Climatology, Bayreuth Center of Ecology and Environmental Research (BayCEER), Universitätsstr. 30, 95447, Bayreuth, Germany
78. Stadt Frankfurt am Main - Der Magistrat, Palmengarten, Siesmayerstraße 61, 60323, Frankfurt am Main, Germany
79. Research Centre of Slovenian Academy of Sciences and Arts (ZRC SAZU), Institute of Biology, Novi trg 2, 1000, Ljubljana, Slovenia
80. University of Zagreb, Faculty of Forestry, Svetošimunska 25, 10000, Zagreb, Croatia
81. University of Adelaide, TERN, North Terrace, 5005, Adelaide, Australia
82. University of Zagreb, Faculty of Geotechnical Engineering, Hallerova aleja 7, 42000, Varaždin, Croatia
83. Aristotle University of Thessaloniki, School of Biology, 54124, Thessaloniki, Greece
84. Plant Science and Biodiversity Centre Slovak Academy of Sciences, Institute of Botany, Dubravska cesta 9, 84523, Bratislava, Slovakia
85. University of Erlangen-Nuremberg, Department of Geography, Wetterkreuz 15, 91058, Erlangen, Germany
86. Universidade Federal do Rio Grande do Sul, Department of Ecology, Av Bento Gonçalves 9500, 91501-970, Porto Alegre, Brazil
87. University of Perugia, Department of Chemistry, Biology and Biotechnology, Borgo XX giugno 74, 06124, Perugia, Italy
88. Universidade Regional de Blumenau, Departamento de Engenharia Florestal, Rua São Paulo, 3250, 89030-000, Blumenau, Brazil
89. University of Oulu, Ecology and Genetics Research Unit, Biodiversity Unit, Kaitoväylä 5, 90014, Oulu, Finland
90. Helmholtz Center for Environmental Research - UFZ, Department of Physiological Diversity, Permoserstr. 15, 04318, Leipzig, Germany
91. Leuphana University of Lüneburg, Institute of Ecology, Universitätsallee 1, 21335, Lüneburg, Germany
92. University of Alberta, Department of Biological Sciences, Biological Sciences Building, T6G2E9, Edmonton, Canada
93. University of Alaska, Institute of Arctic Biology, P. O. Box 7570000, 99775, Fairbanks, United States
94. Addis Ababa University, Department of Geography & Environmental Studies, Sidist Kilo SQ, 150178, Addis Ababa, Ethiopia
95. Senckenberg Museum of Natural History Görlitz, Botany Department, PO Box 300 154, 02806, Görlitz, Germany
96. Technische Universität Dresden, International Institute Zittau, Markt 23, 02763, Zittau, Germany
97. Brown University, Department of Ecology and Evolutionary Biology/Brown University Herbarium, 34 Olive Street, 02912, Providence, United States
98. Vienna Institute for Nature Conservation & Analyses, Giessergasse 6/7, 1090, Vienna, Austria
99. Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Research Unit Forest Dynamics, Zürcherstrasse 111, 8903, Birmensdorf, Switzerland
00. Ufa Scientific Centre, Russian Academy of Sciences, Laboratory of Wild-Growing Flora, Botanical Garden-Institute, Mendeleev str., 195/3, 450080, Ufa, Russia

Abstract

Vegetation provides the foundation of life on Earth. Assessing biodiversity status and trends in plant communities is therefore critical to understand and quantify the effects of global change on ecosystems. Here, we present the largest dataset of vegetation plots (i.e. species co-occurrence or community composition data) ever released in open access. It contains information on 91,031 vegetation plots recording the cover or abundance of each plant species that occurs in a plot of a given surface area at the date of the botanical survey. Plots were derived from 103 local to regional datasets. To improve the representation of Earth's environmental conditions, plots were resampled from a larger pool of vegetation plots using an environmentally stratified sampling design. Each vegetation plot comes with information on community-weighted means and variances of key plant functional traits. Our open-access dataset can be used to explore global patterns of diversity at the plant community level, as ground truthing data in remote sensing applications or as a baseline for biodiversity monitoring.

Background & Summary

Biodiversity is facing a global crisis (1). As many as 1 million species are estimated to be already facing extinction, mostly as a consequence of anthropogenic impacts, land-use and climate change (1). The rates of biodiversity redistribution and homogenization are also accelerating (2; 3). Biological assemblages are becoming progressively more similar to each other globally, as local biodiversity and endemic species go extinct and are replaced by introduced exotic species or by more widespread and competitive native species (1; 3). This has profound potential impacts on human and ecosystem health (4; 5). For instance, many terrestrial and marine species are shifting their geographical distribution as a response to climate change (2), including animals hosting pathogens transmissible to humans (6; 7; 8).

Vegetation, i.e., the assemblage of plant species, is no exception to this biodiversity crisis (9; 10; 3). This is worrisome, since terrestrial vegetation accounts for 80% (450 Gt C) of the living biomass on Earth (11). Given the central role of vegetation in ecosystem productivity, stability and functioning (10), assessing biodiversity status and trends in plant communities is paramount, for other life compartments and human societies alike.

Monitoring plant biodiversity trends requires adequate data across a range of scales ([**???**]). Large independent collections of plant occurrence data do exist at the global or continental extent via the Botanical Information and Ecology Network (BIEN) (12), the Global Inventory of Floras and Traits (GIFT) (13) or the Global Biodiversity Information Facility (GBIF) (https://www.gbif.org/). However, all these occurrence-only databases either neglect how individual plant species co-occur and interact locally to form plant communities, or are collected at spatial resolutions (e.g., one-degree grid cells) which are too coarse to assess biodiversity trends at the most relevant scale of local plant communities (14).

Yet, there is a long-lasting tradition among botanists to record the cover or abundance of each plant species that occurs in a vegetation plot of a given size (i.e. surface area) at a given time (e.g. 15). Compared to species-level data, vegetation-plot data present many advantages. First, they contain information on which plant species co-occur together in the same locality at a given moment in time (16). This built-in feature of vegetation plots is a necessary prerequisite for testing hypotheses related to biotic interactions among plant species (i.e. plant-plant interactions). It can also provide crucial information on where and when a species is absent, therefore improving current species distribution models (17). Being spatially explicit, vegetation plots can be resurveyed through time to assess potential changes in plant species composition relative to a baseline (18; 19, 3). As they normally contain also information on the relative cover or abundance of each species, vegetation plots are more adequate to detect subtle biodiversity changes, compared to data based on the occurrence of individual species only (20).

Vegetation-plot data are very fragmented, though, as they typically stem from a myriad of research projects. As such, these data often suffer from the usual trade-off in biodiversity data: Collections have either fine-grain spatial resolutions but small spatial extents, or vice versa (21). Furthermore, with their disparate sampling protocols, standards and taxonomic resolutions, aggregating and harmonizing vegetation plot data proves extremely challenging (22). It is not surprising, therefore, that these data have only been rarely used in global-scale biodiversity research until recently (23; 24).

The sPlot initiative tries to close this data gap. It leverages on several existing local to regional vegetation-plot datasets, to create a harmonized and comprehensive global geo-database of terrestrial plant species assemblages (25). Established in 2013, sPlot currently contains more than 1.9 million vegetation plots, and is fully integrated with the TRY database (26), from which it derives information on plant functional traits. The sPlot database is increasingly being used to study continental- to global-scale vegetation patterns, such as the relative contribution of regional vs. local factors on the global patterns of fern richness (27), the mechanisms underlying the spread and abundance of native vs. invasive tree species (28), and worldwide trait–environment relationships in plant communities (22).

Here, we provide an open-access data set composed of 91,031 plots, which is representative of the environmental space covered by the sPlot database. Plots stem from 103 databases, and span across 115 countries (Figure 1). This resampled dataset (sPlot Open - hereafter) is composed of: (1) plot-level information, including metadata and basic vegetation structure descriptors; (2) the species composition of each vegetation plot, including species cover or abundance information when available; and (3) community-level functional diversity indices derived from the TRY database (26).

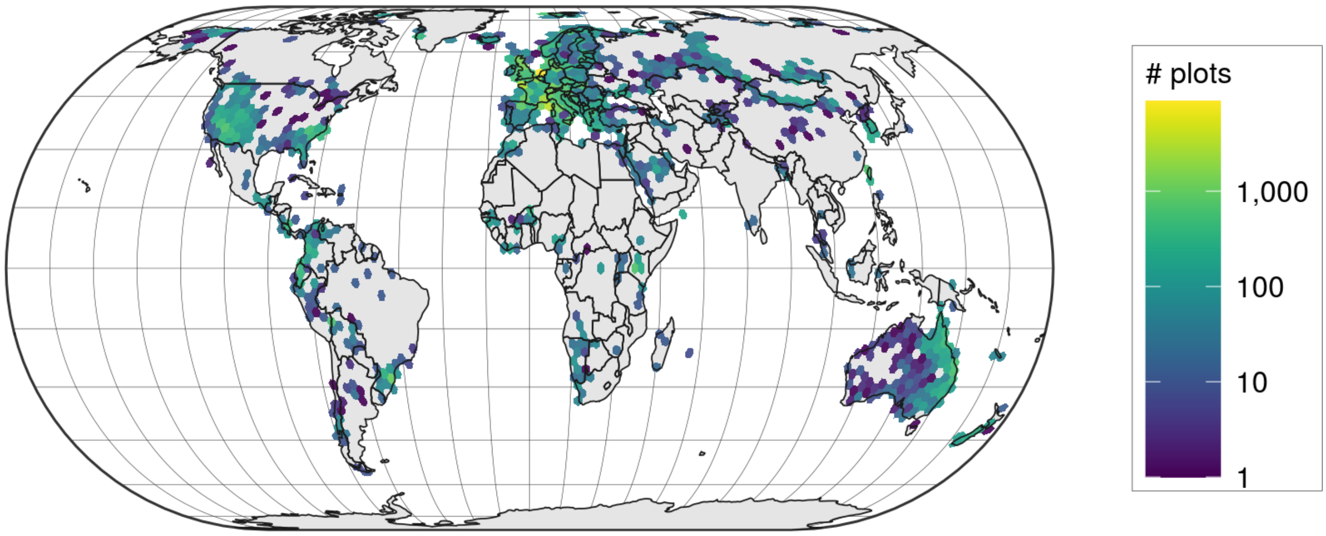

Figure 1: Global map of sPlot Open (n = 91,031) and spatial distribution of vegetation plot density per hexagonal cell with a spatial resolution of approximately 70.000 km². Map projection is Eckert IV.

Methods

Vegetation plot data sources

We started from the sPlot database v2.1 (created October 2016), which contains 1,121,244 vegetation plots and 23,586,216 species records stemming from 110 different vegetation-plot datasets of regional, national or continental extent. Some of the 110 datasets stem from regional or continental initiatives (see 25 for more information). For instance: 48 vegetation-plot datasets derive from the European Vegetation Archive (EVA) (16), three major African datasets from the Tropical African Vegetation Archive (TAVA), multiple vegetation datasets in the USA from the VegBank archive (29; 30). Data from other continents (South America, Asia) or countries were contributed as separate datasets. The metadata of each of the 110 vegetation-plot datasets stored in sPlot are managed through the Global Index of Vegetation-Plot Databases (GIVD; 31), using the GIVD identifier as the unique dataset identifier.

Resampling method

Data in the sPlot database are unevenly distributed across continents and biomes (see 22). Mid-latitude regions in developing countries (mostly Europe, the USA and Australia) are overrepresented, while regions in the tropics and subtropics are underrepresented, which is a typical geographical bias in biodiversity data (e.g., 32; 2). To reduce this imbalance to the extent possible, we performed a stratified resampling approach, using several environmental variables available at the global extent as sampling strata. We considered 30 climatic and soil variables. For climate we complemented the 19 bioclimatic variables from CHELSA (33), as well as two variables reflecting growing-season warmth (growing degree days above 1 °C - GDD1 - and 5 °C - GDD5), which we calculated based on CHELSA bioclimatic variables. In addition we considered an index of aridity (AR) and a model for Potential Evapotranspiration (PET - 34). For soil, we extracted seven variables from the SOILGRIDS database (35), namely: soil organic carbon content in the fine earth fraction, cation exchange capacity, pH, as well as the fractions of coarse fragments, sand, silt and clay.

We stratified our sampling effort based on the following procedure. First we ran a global principal component analysis (PCA) of the 30 above-mentioned environmental variables. We considered the full environmental space of all terrestrial habitats on Earth at a spatial resolution of 2.5 arcmin, totaling 8,384,404 terrestrial grid cells, irrespective of whether a grid cell hosted vegetation plots from the sPlot database v2.1 or not. We then subdivided the environmental space represented by the first two principal components (PC1–PC2), accounting for 47% and 23% of the total variation on PC1 and PC2, respectively, into a 100 × 100 grid. This PC1-PC2 bidimensional space was subsequently used to balance our sampling effort across all PC1-PC2 grid cells for which vegetation plots are available. Before projecting vegetation plots from the sPlot database v2.1 onto this PC1-PC2 environmental space, we removed vegetation plots: from wetlands; from anthropogenic vegetation types; without geographical coordinates; and with a location uncertainty higher than 3 km for those having geographical coordinates. This led to a total of 799,400 out of the initial set of 1,121,244 vegetation plots. When projecting the 799,400 vegetation plots in the PC1-PC2 grid, we calculated how many vegetation plots occurred in each PC1-PC2 grid cell. For those grid cells with more than 50 vegetation plots (n = 858), we randomly selected up to 50 vegetation plots using the heterogeneity-constrained random resampling algorithm from [36]. This approach optimizes the selection of a random subset of vegetation plots that encompasses the highest variability in species composition while avoiding peculiar and rare communities, which may represent outliers. We based the quantification of variability in plant species composition among the 50 randomly selected vegetation plots by computing the mean and the variance of the Jaccard's dissimilarity index (37) between all possible pairs of vegetation plots for a given random selection of 50 vegetation plots (n = 1225). We chose this dissimilarity index because it is not influenced by differences in species richness among vegetation plots. More precisely, for a given PC1-PC2 grid cell containing more than 50 vegetation plots, we generated 1,000 random selections of 50 vegetation plots and ranked the 1,000 random selections according to the mean (ascending order) and variance (descending order) value. Ranks from both sortings were summed for each random selection, and the random selection with the lowest summed rank was considered as the most representative of the focal grid cell. In case a grid cell contained fewer than 50 plots, we retained all of them. In this way, we reduced the imbalance towards over-sampled climate types, while ensuring the resampled dataset to be representative of the entire environmental gradient covered by the sPlot database. We repeated the resampling procedure three times to get three different possibilities of a random selection of 50 vegetation plots per PC1-PC2 grid cell with, initially, more than 50 vegetation plots. Vegetation plots selected during the first iteration were our first choice, while we considered the vegetation plots additionally selected in the second and third iteration as reserves when asking for the permission to release the data as open access to each dataset's contributor(s).

Permission to release the data as open access

The resampling procedure resulted in a preliminary potential selection of 98,383 vegetation plots (first choice) and 51,634 vegetation plots flagged as reserves (second or third choice for the subset of PC1-PC2 grid cells with more than 50 vegetation plots available). Being the sPlot database a consortium of independent datasets, whose copyright belongs to the data contributor, we used this preliminary potential selection to ask each dataset's custodian (i.e., either the owner of a dataset or its authorized representative in case of a collective dataset) for permission to release the data of each selected vegetation plot as open access. For 8,070 vegetation plots, permission could not be granted, for instance because the data are unpublished, confidential or sensitive. For these vegetation plots, we used the reserve pool to randomly select replacements, for which such permission could be granted. We imposed the constraint that each vegetation plot in the reserve should belong to the same environmental strata, i.e., the same PC1-PC2 grid cell, of the confidential vegetation plot. Note that a given PC1-PC2 grid cell may have one or more confidential vegetation plots (max = xx) that could not be replaced from the reserve pool.

Trait information

For each vegetation plot for which open access has been granted, we computed the community weighted means for eighteen plant functional traits derived from the TRY database v3.0 (26). These traits were selected among those traits that describe the leaf, wood and seed economics spectra (38; 39), and are known to either affect different key ecosystem processes or respond to macroclimatic drivers or both (25). The eighteen plant functional traits were: (1) leaf area [mm^2]; (2) stem specific density [$g\ cm^{-3}$]; (3) specific leaf area [$m^2 kg^{-1}$]; (4) leaf carbon concentration [$mg\ g^{-1}$]; (5) leaf nitrogen concentration [$mg\ g^{-1}$]; (6) leaf phosphorus concentration [$mg\ g^{-1}$]; (7) plant height [m]; (8) seed mass [mg]; (9) seed length [mm]; (10) leaf dry matter content [$g\ g^{-1}$]; (11) leaf nitrogen per area [$g\ m^{-2}$]; (12) leaf N:P ratio [$g\ g^{-1}$]; (13) leaf $\delta^{15}N$ [per million]; (14) seed number per reproductive unit; (15) leaf fresh mass [g]; (16) stem conduit density [mm^{-2}]; (17) dispersal unit length [mm]; and (18) conduit element length [μm].

Because missing values were particularly widespread in the species-trait matrix, we employed a gap-filling procedure based on hierarchical Bayesian modeling (R package 'BHPMF', 40; 41). Gap-filling was performed at the level of individual observations. We then loge-transformed all gap-filled trait values and averaged each trait by taxon (i.e., at species, or genus level). Additional information on the gap-filling procedure are available in [25].

Community-weighted means (CWM) and the variances (CWV) were calculated for every plant functional trait j and every vegetation plot k as follows (42):

$$CWM_{j,k} = \sum_{i}^{n_k} p_{i,k} t_{i,j} \tag{1}$$

$$CWV_{j,k} = \sum_{i}^{n_k} p_{i,k} (t_{i,j} - CWM_{j,k})^2 \tag{2}$$

where n_k is the number of species with trait information in vegetation plot k, $p_{i,k}$ is the relative abundance of species i in vegetation plot k calculated as the species' fraction in cover or abundance of total cover or abundance, and $t_{i,j}$ is the mean value of species i for trait j.

Data Records

The final dataset that is provided here as open access contains 91,031 vegetation plots from 115 countries and all continents except Antarctica (Figure 1) and stems from 103 constitutive datasets (11). Information on the size (surface area) of the vegetation survey is available for 61,898 vegetation plots, and ranges between 0.01 m^2 and 4 ha (mean = 270 m^2; median = 78.5 m^2). The average number of vascular plant species per vegetation plot ranges between 1 (i.e. monospecific stands) and 270 species (mean = 17.6; median = 13). Most plots only include information on vascular plants, while a minority also includes information on lichens (n = 3,045) or mosses (n = 4,963). By reducing the overrepresentation of vegetation plots in specific environmental conditions, the resampling procedure described above strongly reduced the bias in the distribution of vegetation plots within the environmental niche space. Yet, due to the lack or scarcity of data from some geographical regions, like the tropics, the spatial distribution of vegetation plots remains unbalanced across geographical regions (Figure 1). This is evident when comparing the number of plots across continents or biomes. Europe is by far the best represented continent, with 53,884 vegetation plots. In contrast, Africa and South America have only 4507 and 5515 vegetation plots, respectively. The representation of biomes is equally unbalanced. The biomes 'Temperate midlatitudes' and 'Subtropics with winter rain' have 37,507 and 16,510 vegetation plots, respectively, while none of the other biomes have more than 10,000 vegetation plots (Figure 2).

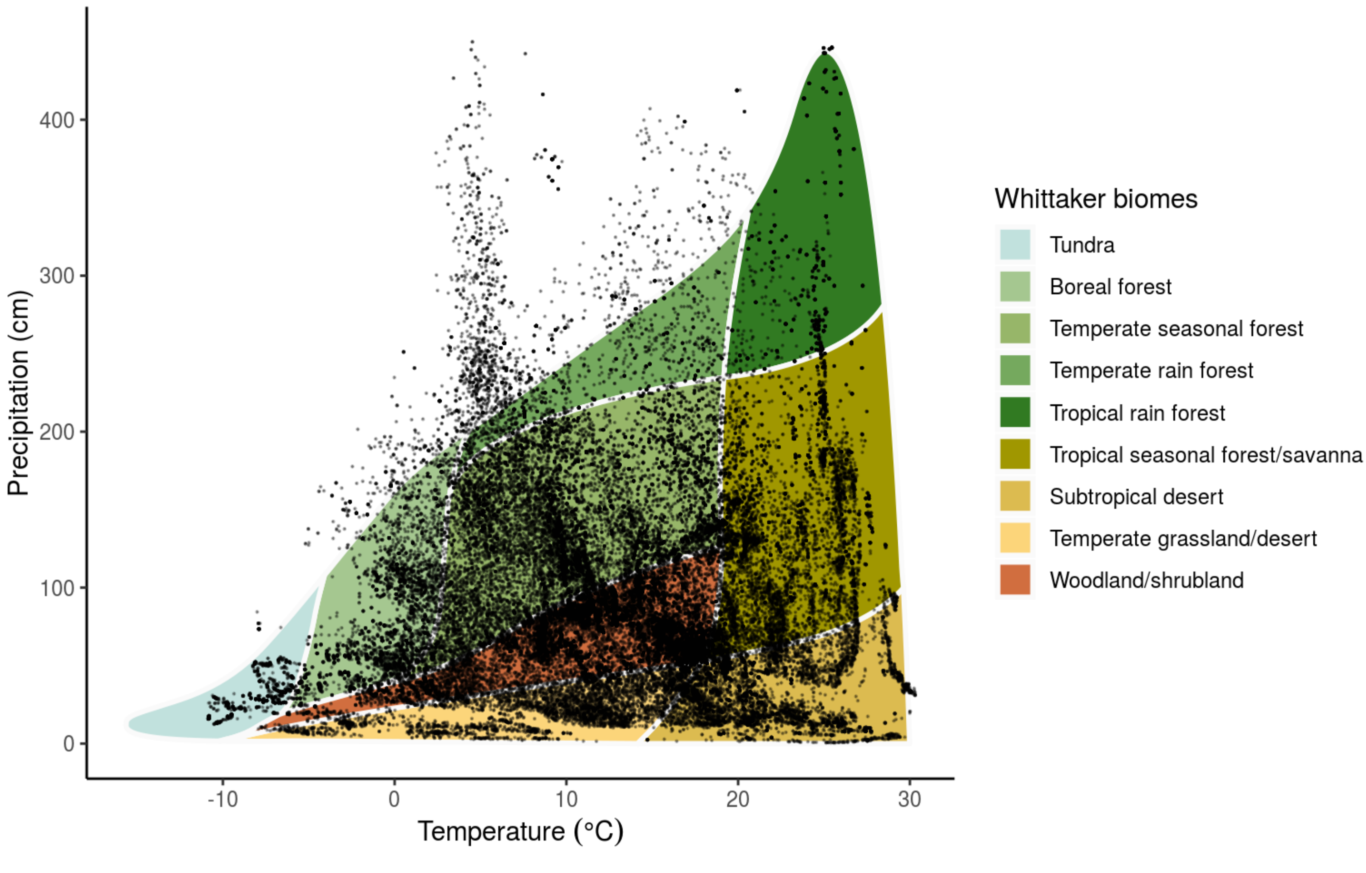

Figure 2: Distribution in environmental space (R package plotbiomes by Valentin Stefan).

Finally, the dataset contains a relatively balanced number of forest (n = 25,832) vs. non forest (n = 38,203) vegetation plots, with a minor proportion of plots remaining unassigned (n = 10,050). The assignment of plots to forests and non-forests is based on multiple lines of evidence, including the plot-level information on the cover of the tree layer, as well as traits of species composing a plot, such as growth form and height. In short, a plot record was considered a forest if the cover of the tree layer, or alternatively, the sum of relative cover of all tree taxa, was greater than 0.25. It was instead considered a non-forest record if the sum of relative cover of low-stature, non-tree and non-shrub taxa was greater than 0.90. For an extensive explanation on this classification scheme, we refer the reader to [25]. Even if the proportion of forest vs. non-forest vegetation plots is relatively well-balanced, the geographical distribution of vegetation plots belonging to different vegetation types is likely not balanced in the geographical space, as it depends on the idiosyncrasies of the individual datasets composing the sPlot database. For instance, the data from New Zealand only include plots collected in non-forest ecosystems, while data from Chile only refer to forests. We invite potential users to carefully read the description of each individual dataset before using this open-access dataset.

Database Organization

The open-access dataset is organized into three matrices.

The *'header'* matrix contains plot level information for the 91,031 vegetation plots provided in this open access dataset, including metadata (e.g., plot ID, ownership, sampling date, geographical location, positional accuracy), sampling design information (e.g., the total surface area used during the vegetation survey), and a plot-level description of vegetation structure (e.g., vegetation type, percentage cover of each vegetation layer). A brief description of all the xx variables contained in the header matrix is provided in 12.

The *'DT'* matrix contains data on the species composition of each plot. It is structured in a long format and contains 1,607,826 records, from 39,922 taxa, mostly resolved at the species level. For each record we report both the taxon name as originally contributed by the data custodian (column *'Matched_concept'*), and the taxon name after taxonomic standardization (column *'Species_name_harmonized'*). For each entry, we report the species cover//abundance values. These follow different standards across the datasets constituting the sPlot database. We therefore provide both the cover//abundance value as reported in the oringal data, which often is given on a cover//abundance scale (column *'Cover'*), and a *'Relative_cover'* field, i.e., the cover//abundance of each taxon in each vegetation layer divided by the total cover//abundance of all taxa in that vegetation layer. Finally, for each entry, we provide a *'Taxon_group'* field, reporting whether the corresponding taxa is a vascular plant, moss, lichen or alga.

Finally the *'CWM'* matrix contains the community-weighted means and variances calculated for each of the 18 functional traits mentioned above. It also contains three additional columns. The column *'Species_richness'* returns the number of species recorded in each plot. The columns *'Trait_coverage_cover'* and *'Trait_coverage_pa'* return respectively the proportion of total cover and species in a plot for which functional trait information was available.

Functional trait information was available for 20,932 species. The average proportion of species in each plot for which we have functional trait information is 0.88 (median = 1). For 47,177 plots the coverage is complete, while only in one plot we have no functional trait information for any of the occurring species. When considering relative cover, the average trait coverage is 0.89. As many as 68,234 and 74,388 plots have functional trait information for more than 80% of the species or 80% of relative cover, respectively.

Technical Validation

The sPlot database has a nested structure, and is composed of several individual datasets, each validated and maintained by its respective dataset custodian. Each individual dataset also has individual vegetation plots, each provided by its owner (the person who performed the actual vegetation survey) or by someone who digitized the original data from the scientific or grey literature. We obviously have no direct control on the individual vegetation plots that we provide here in an open access dataset. Yet, each of these vegetation plots are stemming from trained professional botanists, or published scientific work, and are accompanied by detailed information on the sampling protocols used, thus ensuring data quality and reliability.

Before having been integrated into the sPlot database, each dataset was further checked for consistency and, if having a different format, was converted to a Turboveg 2 database (43). During this conversion into a Turboveg format, we checked that all datasets contained the required metadata information and we converted this information to the sPlot database standards, if necessary. Furthermore we cross-checked that each plot is located within the geographic scopes of its respective dataset. Finally, we harmonized all the taxonomic names from a dataset, based on the sPlot's taxonomic backbone (Purschke 2017). This backbone matched all the taxonomic names (without nomenclatural authors) from all datasets in sPlot 2.1 and TRY v3.0 (26) to their resolved version based on the Taxonomic Name Resolution Service web application (TNRS version 4.0; 44; iPlant Collaborative, 2015). This allowed to (1) harmonize all datasets to a common nomenclature, and (2) to link the sPlot database to the TRY database (26). All taxa originally denoted at taxonomic ranks lower than species, were aggregated at species level. Additional detail on the taxonomic resolution is reported in [25], while a description of the workflow, including R-code, is available in [45]

Usage Notes

The sPlot Open database can be downloaded from https://www.idiv.de (link to PlantHub). The use of data contained in BioTIME should cite original data citations in addition to the present paper. The data included in the present paper represent the subset of sPlot for which we were able to secure licences for making these data open. The additional studies in sPlot are avalable under sPlot's Governance and Data Property Rules (www.idiv.de/sPlot).

Code Availability

The R code used to produce sPlot Open from the sPlot 2.1 database is found here (https://portal.idiv.de/nextcloud/index.php/s/YjMZtwFDwtoefGi).

Acknowledgements

We are grateful to thousands of vegetation scientists who sampled vegetation plots in the field or digitized them into regional, national or international databases. We also appreciate the support of the German Research Foundation for funding sPlot as one of the iDiv (DFG FZT 118, 202548816) research platforms, and the organization of three workshops through the sDiv calls. We acknowledge this support with naming the database "sPlot", where the "s" refers to the sDiv synthesis workshops. The study was supported by the TRY initiative on plant traits (http://www.try-db.org). For all further acknowledgements see Appendix S10 in [25].

Author contributions

FMS wrote the first draft of the manuscript, with considerable input by JL and HB. JL and TH wrote the resampling algorithm. FMS set up the GiHub project and produced the graphs. He also coordinated the sPlot consortium. SMH wrote the Turboveg v3 software, which holds the sPlot database. JK provided the trait data from TRY and FS performed the trait data gap filling. HB secured the funding for sPlot as a strategic project of iDiv. All other authors contributed data. All authors contributed to revisng the manuscript.

Competing interests

[*A competing interests statement is required for all papers accepted by and published in Scientific Data. If there is no conflict of interest, a statement declaring this must still be included in the manuscript*.]

References

1. **Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services**
IPBES
IPBES secretariat (2019)
ISBN: [978-3-947851-13-3](978-3-947851-13-3)

2. **Species better track climate warming in the oceans than on land**
Jonathan Lenoir, Romain Bertrand, Lise Comte, Luana Bourgeaud, Tarek Hattab, Jérôme Murienne, Gaël Grenouillet
Nature Ecology & Evolution (2020-05-25) https://doi.org/ggx3np
DOI: 10.1038/s41559-020-1198-2 · PMID: 32451428

3. **Replacements of small- by large-ranged species scale up to diversity loss in Europe's temperate forest biome**
Ingmar R. Staude, Donald M. Waller, Markus Bernhardt-Römermann, Anne D. Bjorkman, Jörg Brunet, Pieter De Frenne, Radim Hédl, Ute Jandt, Jonathan Lenoir, František Máliš, … Lander Baeten
Nature Ecology & Evolution (2020-04-13) https://doi.org/ggrs73
DOI: 10.1038/s41559-020-1176-8 · PMID: 32284580

4. **Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being**
Gretta T. Pecl, Miguel B. Araújo, Johann D. Bell, Julia Blanchard, Timothy C. Bonebrake, I-Ching Chen, Timothy D. Clark, Robert K. Colwell, Finn Danielsen, Birgitta Evengård, … Stephen E. Williams
Science (2017-03-31) https://doi.org/f9xmpm
DOI: 10.1126/science.aai9214 · PMID: 28360268

5. **Managing consequences of climate-driven species redistribution requires integration of ecology, conservation and social science**
Timothy C. Bonebrake, Christopher J. Brown, Johann D. Bell, Julia L. Blanchard, Alienor Chauvenet, Curtis Champion, I-Ching Chen, Timothy D. Clark, Robert K. Colwell, Finn Danielsen, … Gretta T. Pecl
Biological Reviews (2018-02) https://doi.org/gc2dvc
DOI: 10.1111/brv.12344 · PMID: 28568902

6. **Global trends in emerging infectious diseases**
Kate E. Jones, Nikkita G. Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L. Gittleman, Peter Daszak
Nature (2008-02) https://doi.org/cbxh9h
DOI: 10.1038/nature06536 · PMID: 18288193 · PMCID: PMC5960580

7. **The range of Ixodes ricinus and the risk of contracting Lyme borreliosis will increase northwards when the vegetation period becomes longer**
Thomas G. T. Jaenson, Elisabet Lindgren
Ticks and Tick-borne Diseases (2011-03) https://doi.org/fmn4dp
DOI: 10.1016/j.ttbdis.2010.10.006 · PMID: 21771536

8. **Altitudinal Changes in Malaria Incidence in Highlands of Ethiopia and Colombia**
A. S. Siraj, M. Santos-Vega, M. J. Bouma, D. Yadeta, D. R. Carrascal, M. Pascual
Science (2014-03-06) https://doi.org/f5vb47
DOI: 10.1126/science.1244325 · PMID: 24604201

9. **A Significant Upward Shift in Plant Species Optimum Elevation During the 20th Century**
J. Lenoir, J. C. Gegout, P. A. Marquet, P. de Ruffray, H. Brisse
Science (2008-06-27) https://doi.org/bnhhj8
DOI: 10.1126/science.1156831 · PMID: 18583610

10. **The functional role of producer diversity in ecosystems**
Bradley J. Cardinale, Kristin L. Matulich, David U. Hooper, Jarrett E. Byrnes, Emmett Duffy, Lars Gamfeldt, Patricia Balvanera, Mary I. O'Connor, Andrew Gonzalez
American Journal of Botany (2011-03) https://doi.org/fnh8qs
DOI: 10.3732/ajb.1000364 · PMID: 21613148

11. **The biomass distribution on Earth**
Yinon M. Bar-On, Rob Phillips, Ron Milo
Proceedings of the National Academy of Sciences (2018-06-19) https://doi.org/cp29
DOI: 10.1073/pnas.1711842115 · PMID: 29784790 · PMCID: PMC6016768

12. **Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity**
Brian J Enquist, Rick Condit, Robert K Peet, Mark Schildhauer, Barbara M. Thiers
PeerJ (2018-01-13) https://doi.org/ghfnsx
DOI: 10.7287/peerj.preprints.2615v2

13. **GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography**
Patrick Weigelt, Christian König, Holger Kreft
Journal of Biogeography (2019-06-09) https://doi.org/gf38t6
DOI: 10.1111/jbi.13623

14. **Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data**
Elizabeth H. Boakes, Philip J. K. McGowan, Richard A. Fuller, Ding Chang-qing, Natalie E. Clark, Kim O'Connor, Georgina M. Mace
PLoS Biology (2010-06-01) https://doi.org/brfdq6
DOI: 10.1371/journal.pbio.1000385 · PMID: 20532234 · PMCID: PMC2879389

15. **Versuch einer Übersicht über die Wiesentypen der Schweiz**
F. G. Stebler, C. Schröter
Landwirt. Jahrb. Schweiz (1893)

16. **European Vegetation Archive (EVA): an integrated database of European vegetation plots**
Milan Chytrý, Stephan M. Hennekens, Borja Jiménez-Alfaro, Ilona Knollová, Jürgen Dengler, Florian Jansen, Flavia Landucci, Joop H. J. Schaminée, Svetlana Aćić, Emiliano Agrillo, … Sergey Yamalov
Applied Vegetation Science (2016-01) https://doi.org/bc7k
DOI: 10.1111/avsc.12191

17. **Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data**
Steven J. Phillips, Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, Simon Ferrier
Ecological Applications (2009-01) https://doi.org/dx4s78
DOI: 10.1890/07-2153.1 · PMID: 19323182

18. **Global environmental change effects on plant community composition trajectories depend upon management legacies**
Michael P. Perring, Markus Bernhardt-Römermann, Lander Baeten, Gabriele Midolo, Haben Blondeel, Leen Depauw, Dries Landuyt, Sybryn L. Maes, Emiel De Lombaerde, Maria Mercedes Carón, … Kris Verheyen

Global Change Biology (2018-04) https://doi.org/gc6mjp
DOI: 10.1111/gcb.14030 · PMID: 29271579

19. **Accelerated increase in plant species richness on mountain summits is linked to warming**
Manuel J. Steinbauer, John-Arvid Grytnes, Gerald Jurasinski, Aino Kulonen, Jonathan Lenoir, Harald Pauli, Christian Rixen, Manuela Winkler, Manfred Bardy-Durchhalter, Elena Barni, … Sonja Wipf
Nature (2018-04-04) https://doi.org/gdfwk3
DOI: 10.1038/s41586-018-0005-6 · PMID: 29618821

20. **Exploring large vegetation databases to detect temporal trends in species occurrences**
Ute Jandt, Henrik von Wehrden, Helge Bruelheide
Journal of Vegetation Science (2011-12) https://doi.org/d8b4jv
DOI: 10.1111/j.1654-1103.2011.01318.x

21. **Biodiversity data integration—the significance of data resolution and domain**
Christian König, Patrick Weigelt, Julian Schrader, Amanda Taylor, Jens Kattge, Holger Kreft
PLOS Biology (2019-03-18) https://doi.org/c3xz
DOI: 10.1371/journal.pbio.3000183 · PMID: 30883539 · PMCID: PMC6445469

22. **Global trait–environment relationships of plant communities**
Helge Bruelheide, Jürgen Dengler, Oliver Purschke, Jonathan Lenoir, Borja Jiménez-Alfaro, Stephan M. Hennekens, Zoltán Botta-Dukát, Milan Chytrý, Richard Field, Florian Jansen, … Ute Jandt
Nature Ecology & Evolution (2018-11-19) https://doi.org/gfj595
DOI: 10.1038/s41559-018-0699-8 · PMID: 30455437

23. **Big data for forecasting the impacts of global change on plant communities**
Janet Franklin, Josep M. Serra-Diaz, Alexandra D. Syphard, Helen M. Regan
Global Ecology and Biogeography (2017-01) https://doi.org/f9hdp3
DOI: 10.1111/geb.12501

24. **Achievements and challenges in the integration, reuse and synthesis of vegetation plot data**
Susan K. Wiser
Journal of Vegetation Science (2016-09) https://doi.org/ghfnr5
DOI: 10.1111/jvs.12419

25. **sPlot – A new tool for global vegetation analyses**
Helge Bruelheide, Jürgen Dengler, Borja Jiménez-Alfaro, Oliver Purschke, Stephan M. Hennekens, Milan Chytrý, Valério D. Pillar, Florian Jansen, Jens Kattge, Brody Sandel, … Andrei Zverev
Journal of Vegetation Science (2019-04-08) https://doi.org/gfvhkm
DOI: 10.1111/jvs.12710

26. **TRY plant trait database – enhanced coverage and open access**
Jens Kattge, Gerhard Bönisch, Sandra Díaz, Sandra Lavorel, Iain Colin Prentice, Paul Leadley, Susanne Tautenhahn, Gijsbert D. A. Werner, Tuomas Aakala, Mehdi Abedi, … Christian Wirth
Global Change Biology (2020) https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14904
DOI: 10.1111/gcb.14904

27. **Global fern and lycophyte richness explained: How regional and local factors shape plot richness**
Anna Weigand, Stefan Abrahamczyk, Isabelle Aubin, Claudia Bita-Nicolae, Helge Bruelheide, Cesar I. Carvajal-Hernández, Daniele Cicuzza, Lucas Erickson Nascimento da Costa, János Csiky, Jürgen Dengler, … Michael Kessler
Journal of Biogeography (2019-12-30) https://doi.org/ggf4gr
DOI: 10.1111/jbi.13782

28. **Similar factors underlie tree abundance in forests in native and alien ranges**
Masha T. Sande, Helge Bruelheide, Wayne Dawson, Jürgen Dengler, Franz Essl, Richard Field, Sylvia Haider, Mark Kleunen, Holger Kreft, Joern Pagel, … Tiffany M. Knight
Global Ecology and Biogeography (2019-12) https://doi.org/ggftj7
DOI: 10.1111/geb.13027 · PMID: 32063745 · PMCID: PMC7006795

29. **Vegetation-plot database of the Carolina Vegetation Survey**
Richard K. Peet, Michael T. Lee, M. Forbes Boyle, Thomas R. Wentworth, Michael P. Schafale, Alan S. Weakley
Vegetation databases for the 21st century (2012) https://doi.org/10.7809/b-e.00081

30. **VegBank – a permanent, open-access archive for vegetation-plot data**
Richard K. Peet, M. T. Lee, M. D. Jennings, D. Faber-Langendoen
Vegetation databases for the 21st century (2012) https://doi.org/10.7809/b-e.00080

31. **The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science**
Jürgen Dengler, Florian Jansen, Falko Glöckler, Robert K. Peet, Miquel De Cáceres, Milan Chytrý, Jörg Ewald, Jens Oldeland, Gabriela Lopez-Gonzalez, Manfred Finckh, … Nick Spencer
Journal of Vegetation Science (2011-08) https://doi.org/ctx2s7
DOI: 10.1111/j.1654-1103.2011.01265.x

32. **Climate-related range shifts - a global multidimensional synthesis and new research directions**
J. Lenoir, J.-C. Svenning
Ecography (2015-01) https://doi.org/f6xz9h
DOI: 10.1111/ecog.00967

33. **Climatologies at high resolution for the earth's land surface areas**
Dirk Nikolaus Karger, Olaf Conrad, Jürgen Böhner, Tobias Kawohl, Holger Kreft, Rodrigo Wilber Soria-Auza, Niklaus E. Zimmermann, H. Peter Linder, Michael Kessler
Scientific Data (2017-09-05) https://doi.org/gbvksk
DOI: 10.1038/sdata.2017.122 · PMID: 28872642 · PMCID: PMC5584396

34. **Global High-Resolution Soil-Water Balance**
Antonio Trabucco, Robert J. Zomer
figshare (2019) https://figshare.com/articles/Global_High-Resolution_Soil-Water_Balance/7707605/3
DOI: 10.6084/m9.figshare.7707605.v3

35. **SoilGrids250m: Global gridded soil information based on machine learning**
Tomislav Hengl, Jorge Mendes de Jesus, Gerard B. M. Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, Marvin N. Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, … Bas Kempen
PLOS ONE (2017-02-16) https://doi.org/f9qc5p
DOI: 10.1371/journal.pone.0169748 · PMID: 28207752 · PMCID: PMC5313206

36. **Heterogeneity-constrained random resampling of phytosociological databases**
Attila Lengyel, Milan Chytrý, Lubomír Tichý
Journal of Vegetation Science (2011-02) https://doi.org/dvjzbz
DOI: 10.1111/j.1654-1103.2010.01225.x

37. **The relationship between species replacement, dissimilarity derived from nestedness, and nestedness**
Andrés Baselga
Global Ecology and Biogeography (2012-12) https://doi.org/gddc72
DOI: 10.1111/j.1466-8238.2011.00756.x

38. **A leaf-height-seed (LHS) plant ecology strategy scheme**
Mark Westoby
Plant and Soil (1998-02-01) https://doi.org/10.1023/A:1004327224729
DOI: 10.1023/a:1004327224729

39. **The world-wide "fast-slow" plant economics spectrum: a traits manifesto**
Peter B. Reich
Journal of Ecology (2014-03) https://doi.org/gfc4z9
DOI: 10.1111/1365-2745.12211

40. **Uncertainty Quantified Matrix Completion Using Bayesian Hierarchical Matrix Factorization**
Farideh Fazayeli, Arindam Banerjee, Jens Kattge, Franziska Schrodt, Peter B. Reich
Institute of Electrical and Electronics Engineers (IEEE) (2014-12) https://doi.org/ghfnw3
DOI: 10.1109/icmla.2014.56

41. **BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography**
Franziska Schrodt, Jens Kattge, Hanhuai Shan, Farideh Fazayeli, Julia Joswig, Arindam Banerjee, Markus Reichstein, Gerhard Bönisch, Sandra Díaz, John Dickie, … Peter B. Reich
Global Ecology and Biogeography (2015-12) https://doi.org/f76qw8
DOI: 10.1111/geb.12335

42. **Scaling from Traits to Ecosystems**
Brian J. Enquist, Jon Norberg, Stephen P. Bonser, Cyrille Violle, Colleen T. Webb, Amanda Henderson, Lindsey L. Sloat, Van M. Savage
Advances in Ecological Research (2015) https://doi.org/ghfnsw
DOI: 10.1016/bs.aecr.2015.02.001

43. **TURBOVEG, a comprehensive data base management system for vegetation data**
Stephan M. Hennekens, Joop H. J. Schaminée
Journal of Vegetation Science (2001-02-24) https://doi.org/cgmn6m
DOI: 10.2307/3237010

44. **The taxonomic name resolution service: an online tool for automated standardization of plant names**
Brad Boyle, Nicole Hopkins, Zhenyuan Lu, Juan Antonio Raygoza Garay, Dmitry Mozzherin, Tony Rees, Naim Matasci, Martha L Narro, William H Piel, Sheldon J Mckay, … Brian J Enquist
BMC Bioinformatics (2013-01-16) https://doi.org/gb8vxz
DOI: 10.1186/1471-2105-14-16 · PMID: 23324024 · PMCID: PMC3554605

45. **Oliverpurschke/Taxonomic_Backbone: First Release Of The Workflow To Generate The Taxonomic Backbone For Splot V.2.1 And Try V.3.0**
Oliver Purschke
Zenodo (2017-08-18) https://doi.org/ghf4ph
DOI: 10.5281/zenodo.845445

Supplementary Material

Table 11: List of databases contributing to the open access dataset extracted from the sPlot database. Databases are ordered based on their ID in the Global Index of Vegetation Databases (GVID ID).

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Table 12: Description of the variables contained in the 'header' matrix, together with their range (if numeric) or possible levels (if nominal or boolean). Variable type can be c - character (i.e. text), f - factor (i.e. qualitative or ordinal variable), i - integer (e.g. binomial), n - numeric (i.e., double) or l - logical (i.e., boolean).

Variable	Range/Levels	No.records	Type.of.variable
GIVD_ID		91031	character
Dataset		91031	character

|Continent |

- Africa
- - Asia
  - - Australia
    - - Europe
      - - North America
        - - Oceania
          - - South America
              |90729 |factor | |Country | |91031 |character | |Biome |
            - Alpine
            - - Boreal zone
              - - Dry midlatitudes
                - - Dry tropics and subtropics
                  - - Polar and subpolar zone
                    - - Subtrop. with year-round rain
                      - - Subtropics with winter rain
                        - - Temperate midlatitudes
                          - - Tropics with summer rain
                            - - Tropics with year-round rain
                                |91031 |factor | |Date |-29764 - 16469 |75798 |numeric |
                                |Latitude |-54.73863 - 80.149116 |91031 |numeric |
                                |Longitude |-162.741433 - 179.590053 |91031 |numeric |
                                |Location_uncertainty |1 - 2500 |91002 |integer |
                                |Releve_area |0.01 - 40000 |61898 |numeric |
                                |Herbs_identified |FALSE = 4876; TRUE = 6323 |11199
                                |logical | |Plant_recorded |
                              - All trees & dominant understory
                              - - All vascular plants
                                - - All vascular plants and dominant cryptogams

- - All woody plants
  - - Dominant trees
    - - Only dominant species
      - - Dominant woody plants >= 2.5 cm dbh
        - - Woody plants >= 10 cm dbh
          - - Woody plants >= 1 m height
            - - Woody plants >= 1 cm dbh
              - - Woody plants >= 20 cm dbh
                - - Woody plants >= 2.5 cm dbh
                  - - Woody plants >= 5 cm dbh
                    - - NA |91015 |factor | |Altitude |-25 - 4819 |52121 |numeric | |Aspect |0 - 360 |30796 |numeric | |Slope |0 - 99 |37784 |numeric | |is_forest |FALSE = 20396; TRUE = 25832 |46228 |logical | |is_nonforest |FALSE = 50870; TRUE = 38203 |89073 |logical | |ESY | |55457 |character | |Naturalness |1 - 2 |68011 |integer | |Forest |FALSE = 38295; TRUE = 23735 |62030 |logical | |Shrubland |FALSE = 38233; TRUE = 11081 |49314 |logical | |Grassland |FALSE = 10213; TRUE = 46947 |57160 |logical | |Sparse_vegetation |FALSE = 33381; TRUE = 11315 |44696 |logical | |Wetland |FALSE = 29078; TRUE = 18038 |47116 |logical | |Cover_total |1 - 313 |24712 |integer | |Cover_tree_layer |0.5 - 150 |7245 |numeric | |Cover_shrub_layer |0.5 - 145 |10197 |numeric | |Cover_herb_layer |0.2 - 180 |26679 |numeric | |Cover_moss_layer |1 - 100 |9643 |integer | |Cover_lichen_layer |1 - 95 |734 |integer |

Cover_algae_layer	1 - 100	221	integer
Cover_litter_layer	1 - 100	4500	integer
Cover_bare_rocks	1 - 100	1897	integer
Cover_cryptogams	1 - 95	593	integer
Cover_bare_soil	0.1 - 99	1412	numeric
Height_trees_highest	1 - 99	6115	numeric
Height_trees_lowest	1 - 90	221	numeric
Height_shrubs_highest	0.1 - 9.9	2880	numeric
Height_shrubs_lowest	0.1 - 9	328	numeric
Height_herbs_average	0.1 - 440	10125	numeric
Height_herbs_lowest	1 - 250	2785	integer
Height_herbs_highest	1 - 600	1733	integer