

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

This manuscript ([permalink](#)) was automatically generated from [fmsabatini/sPlotOpen_Manuscript@6b58091](#) on October 14, 2020.

Authors

- **Francesco Maria Sabatini**

 [0000-0002-7202-7697](#) ·  [fmsabatini](#) ·  [sPlot iDiv](#)

German Centre for Integrative Biodiversity Research (iDiv) - Halle-Jena-Leipzig, Germany; Martin-Luther-Universität Halle-Wittenberg, Institut für Biologie, Am Kirchtor 1, 06108, Halle (Saale), Germany · Funded by None

- **Jonathan Lenoir**

 [0000-0003-0638-9582](#) ·  [lenjon](#) ·  [EkoLogIt](#)

Unité de Recherche "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN), UMR 7058 CNRS, Université de Picardie Jules Verne, 80037 Amiens Cedex 1, France · Funded by None

- **Helge Bruelheide**

 [0000-0003-3135-0356](#) ·  [HelgeBruelheide](#)

Martin-Luther-Universität Halle-Wittenberg, Institut für Biologie, Am Kirchtor 1, 06108, Halle (Saale), Germany; German Centre for Integrative Biodiversity Research (iDiv) - Halle-Jena-Leipzig, Germany · Funded by None

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 balanced 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 ([???]). 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 ([???]). The rates of biodiversity redistribution and homogenization are also accelerating (1; 2). 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 ([???]; 2). This has profound potential impacts on human and ecosystem health (3; 4). For instance, many terrestrial and marine species are shifting their geographical distribution as a response to climate change (1), including animals hosting pathogens transmissible to humans (5; 6; 7).

Vegetation, i.e., the assemblage of plant species, is no exception to this biodiversity crisis (8; 9; 2). This is worrisome, since terrestrial vegetation accounts for 80% (450 Gt C) of the living biomass on Earth (10). Given the central role of vegetation in ecosystem productivity, stability and functioning (9), 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 (11). 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. 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 (15). 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 (16). Being spatially explicit, vegetation plots can be resurveyed through time to assess potential changes in plant species composition relative to a baseline (17; 18, 2). 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 (19).

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 (11). Furthermore, with their disparate sampling protocols, standards and taxonomic resolutions, aggregating and harmonizing vegetation plot data proves extremely challenging (20). It is not surprising, therefore, that these data have only been rarely used in global-scale biodiversity research until recently (21; 22).

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 (23). Established in 2013, sPlot currently contains more than 1.9 million vegetation plots, and is fully integrated with the TRY database ([???]), 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 (24), the mechanisms underlying the spread and abundance of native vs. invasive tree species (25), and worldwide trait-environment relationships in plant communities (20).

Here, we provide a 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 ([???]).

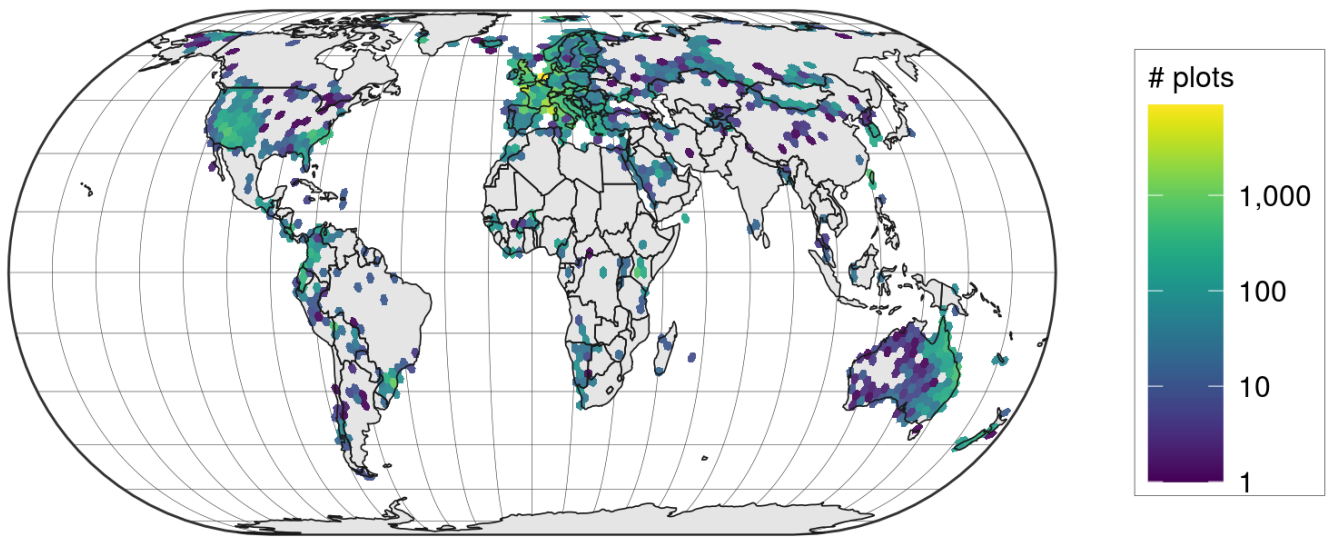


Figure 1: Global map of sPlot Open (n = 91,031) and spatial distribution of vegetation plot density per hexagonal cell with a spatial resolution of approximately 70,000 km². Map projection is Eckert IV.

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 [23](#) for more information). For instance: 48 vegetation-plot datasets derive from the European Vegetation Archive (EVA) ([15](#)), three major African datasets from the Tropical African Vegetation Archive (TAVA), multiple vegetation datasets in the USA from the VegBank archive (Peet, Lee, Boyle, et al., 2012; [26](#)). 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; [27](#)), using the GIVD identifier as the unique dataset identifier.

Resampling method

Data in the sPlot database are unevenly distributed across continents and biomes (see [20](#)). 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., [28](#); [1](#)). 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 ([29](#)), 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 - Trabucco et al. 2010). For soil, we extracted seven variables from the SOILGRIDS database ([30](#)), 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 [\[31\]](#). 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 ([32](#)) 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 ($\text{max} = \text{xx}$) 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 ([\[???\]](#)). These traits were selected among those traits that describe the leaf, wood and seed economics spectra ([33](#); [34](#)), and are known to either affect different key ecosystem processes or respond to macroclimatic drivers or both ([23](#)). The eighteen plant functional traits were: (1) leaf area [mm^2]; (2) stem specific density [g cm^{-3}]; (3) specific leaf area [m^2kg^{-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 $\delta^{15}\text{N}$ [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'; [35](#); [36](#)). 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 [\[23\]](#).

Community-weighted means (CWM) and the variances (CWV) were calculated for every plant functional trait j and every vegetation plot k as follows ([37](#)):

$$CWM_{j,k} = \sum_i^{n_k} p_{i,k} t_{i,j} \quad (1)$$

$$CWW_{j,k} = \sum_i^{n_k} p_{i,k} (t_{i,j} - CWM_{j,k})^2 \quad (2)$$

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 (??). Information on the size (surface area) of the vegetation survey is available for 61,898 vegetation plots, and ranges between 0.01 m² and 4 ha (mean = 270 m²; median = 78.5 m²). 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. Africa and South America, conversely, 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 each, while none of the other biomes have more than 10,000 vegetation plots (Figure 2).

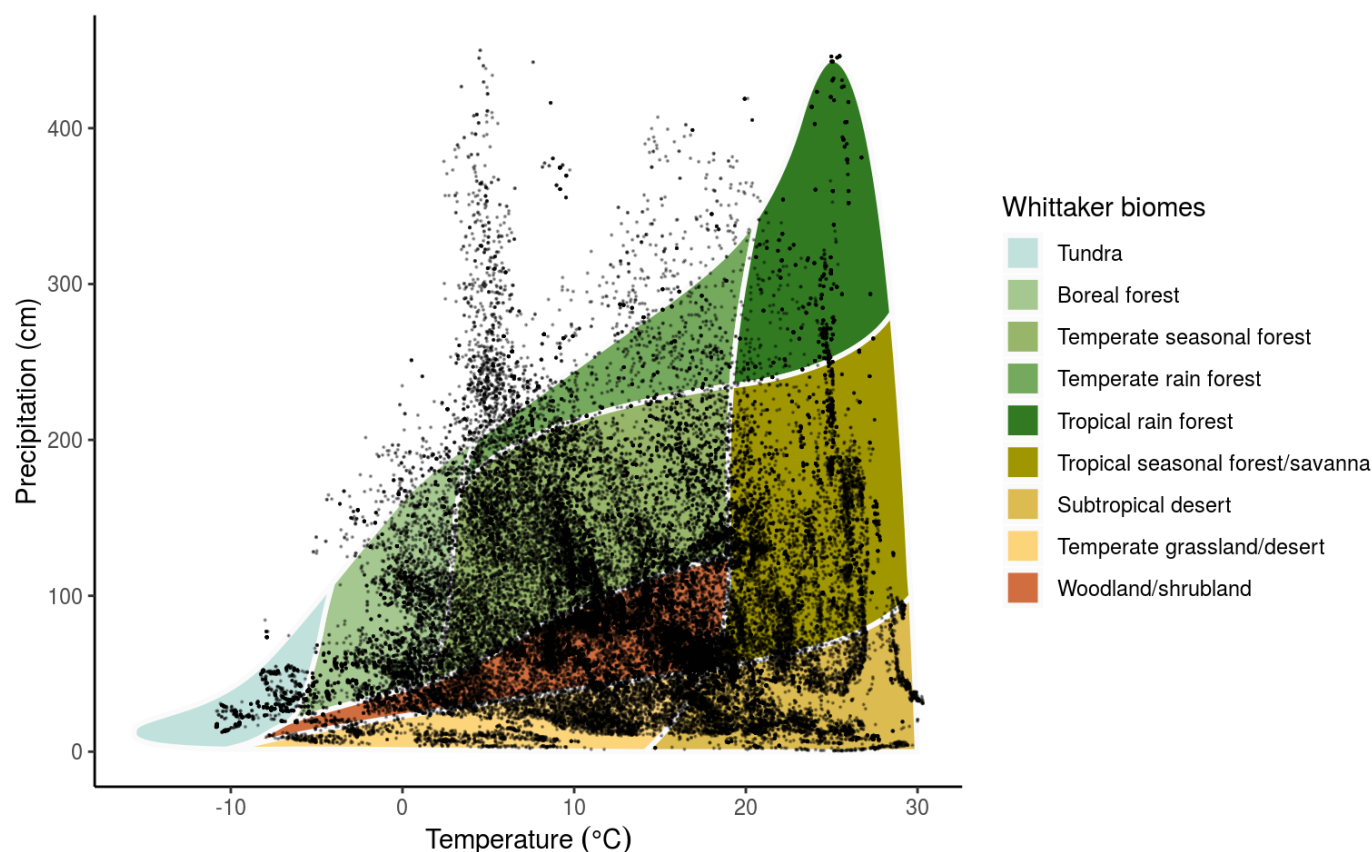


Figure 2: Distribution in environmental space (R package plotbiomes by Valentin Stefan).

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). This classification is based on multiple lines of evidence, including the plot-level information on the cover of the tree layer, and the growth form and height of the species composing a plot. In short, a plot was considered as 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 as non forest 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 [23]. 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.

References

1. **Species better track climate warming in the oceans than on land**
Jonathan Lenoir, Romain Bertrand, Lise Comte, Luana Bourgeaud, Tarek Hattab, Jérôme Muriénne, Gaél Grenouillet
Nature Ecology & Evolution (2020-05-25) <https://doi.org/ggx3np>
DOI: [10.1038/s41559-020-1198-2](https://doi.org/10.1038/s41559-020-1198-2)
2. **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 Hédli, 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](https://doi.org/10.1038/s41559-020-1176-8)
3. **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](https://doi.org/10.1126/science.aai9214) · PMID: [28360268](https://pubmed.ncbi.nlm.nih.gov/28360268/)
4. **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](https://doi.org/10.1111/brv.12344) · PMID: [28568902](https://pubmed.ncbi.nlm.nih.gov/28568902/)
5. **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](https://doi.org/10.1038/nature06536) · PMID: [18288193](https://pubmed.ncbi.nlm.nih.gov/18288193/) · PMCID: [PMC5960580](https://pubmed.ncbi.nlm.nih.gov/PMC5960580/)
6. **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](https://doi.org/10.1016/j.ttbdis.2010.10.006)
7. **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](https://doi.org/10.1126/science.1244325) · PMID: [24604201](https://pubmed.ncbi.nlm.nih.gov/24604201/)
8. **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](https://doi.org/10.1126/science.1156831) · PMID: [18583610](https://pubmed.ncbi.nlm.nih.gov/18583610/)
9. **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](https://doi.org/10.3732/ajb.1000364) · PMID: [21613148](https://pubmed.ncbi.nlm.nih.gov/21613148/)
10. **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](https://doi.org/10.1073/pnas.1711842115) · PMID: [29784790](https://pubmed.ncbi.nlm.nih.gov/29784790/) · PMCID: [PMC6016768](https://pubmed.ncbi.nlm.nih.gov/PMC6016768/)
11. **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](https://doi.org/10.1371/journal.pbio.3000183) · PMID: [30883539](https://pubmed.ncbi.nlm.nih.gov/30883539/) · PMCID: [PMC6445469](https://pubmed.ncbi.nlm.nih.gov/PMC6445469/)
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/ghfnxs>
DOI: [10.7287/peerj.preprints.2615v2](https://doi.org/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](https://doi.org/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](https://doi.org/10.1371/journal.pbio.1000385) · PMID: [20532234](https://pubmed.ncbi.nlm.nih.gov/20532234/) · PMCID: [PMC2879389](https://pubmed.ncbi.nlm.nih.gov/PMC2879389/)

15. **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 Ačić, Emiliano Agrillo, ... Sergey Yamalov
Applied Vegetation Science (2016-01) <https://doi.org/bc7k>
DOI: [10.1111/avsc.12191](https://doi.org/10.1111/avsc.12191)
16. **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](https://doi.org/10.1890/07-2153.1) · PMID: [19323182](https://pubmed.ncbi.nlm.nih.gov/19323182/)
17. **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](https://doi.org/10.1111/gcb.14030) · PMID: [29271579](https://pubmed.ncbi.nlm.nih.gov/29271579/)
18. **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 Bady-Durchhalter, Elena Barni, ... Sonja Wipf
Nature (2018-04-04) <https://doi.org/gdfwk3>
DOI: [10.1038/s41586-018-0005-6](https://doi.org/10.1038/s41586-018-0005-6)
19. **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/d8b4jy>
DOI: [10.1111/j.1654-1103.2011.01318.x](https://doi.org/10.1111/j.1654-1103.2011.01318.x)
20. **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](https://doi.org/10.1038/s41559-018-0699-8)
21. **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](https://doi.org/10.1111/geb.12501)
22. **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/ghfmr5>
DOI: [10.1111/jvs.12419](https://doi.org/10.1111/jvs.12419)
23. **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](https://doi.org/10.1111/jvs.12710)
24. **Global fern and lycophyte richness explained: How regional and local factors shape plot richness**
Anna Weigand, Stefan Abrahamczyk, Isabelle Aubin, Claudia Bitá-Nicolae, Helge Bruelheide, Cesar I. Carvajal-Hernández, Daniele Ciccuzza, 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](https://doi.org/10.1111/jbi.13782)
25. **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](https://doi.org/10.1111/geb.13027) · PMID: [32063745](https://pubmed.ncbi.nlm.nih.gov/32063745/) · PMCID: [PMC7006795](https://pubmed.ncbi.nlm.nih.gov/PMC7006795/)
26. **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 <https://doi.org/10.7809/b-e>
27. **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](https://doi.org/10.1111/j.1654-1103.2011.01265.x)
28. **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](https://doi.org/10.1111/ecog.00967)
29. **Climatologies at high resolution for the earth's land surface areas**
Dirk Nikolaus Karger, Olaf Conrad, Jürgen Böhrer, 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](https://doi.org/10.1038/sdata.2017.122) · PMID: [28872642](https://pubmed.ncbi.nlm.nih.gov/28872642/) · PMCID: [PMC5584396](https://pubmed.ncbi.nlm.nih.gov/PMC5584396/)

30. 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](https://doi.org/10.1371/journal.pone.0169748) · PMID: [28207752](https://pubmed.ncbi.nlm.nih.gov/28207752/) · PMCID: [PMC5313206](https://pubmed.ncbi.nlm.nih.gov/PMC5313206/)

31. 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](https://doi.org/10.1111/j.1654-1103.2010.01225.x)

32. 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](https://doi.org/10.1111/j.1466-8238.2011.00756.x)

33. (unav)

Mark Westoby
Plant and Soil (1998) <https://doi.org/bsvqvq>
DOI: [10.1023/a:1004327224729](https://doi.org/10.1023/a:1004327224729)

34. 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](https://doi.org/10.1111/1365-2745.12211)

35. Uncertainty Quantified Matrix Completion Using Bayesian Hierarchical Matrix Factorization

Farideh Fazayeli, Arindam Banerjee, Jens Kattge, Franziska Schrod, Peter B. Reich
Institute of Electrical and Electronics Engineers (IEEE) (2014-12) <https://doi.org/ghfnw3>
DOI: [10.1109/icmla.2014.56](https://doi.org/10.1109/icmla.2014.56)

36. BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography

Franziska Schrod, 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](https://doi.org/10.1111/geb.12335)

37. 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](https://doi.org/10.1016/bs.aecr.2015.02.001)

Table: 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). `{tbl:Table1 tag='t1'} asd | GVID ID | DB_name | GVID | Custodian | Deputy custodian | n Plots |`

[illegible]

State of Santa Catarina, Brazil (IFFSC Project) | Alexander Christian Vibrans | André Luis de Gasper | 1669 | 1345 | Vibrans et al. (2010) | | SA-BR-003 | Grasslands of Rio Grande do Sul, Brazil | Eduardo Vélez-Martin | Valério De Patta Pillar | 320 | 271 | NA | | SA-BR-004 | Grassland Database of Campos Sulinos | Gerhard E. Overbeck | Valério De Patta Pillar | 161 | 111 | NA | | SA-CL-002 | SSAForests_Plots_db | Alvaro G. Gutierrez | NA | 261 | 163 | NA | | SA-CL-003 | Chilean Park Transects - Fondecyt 1040528 | Aníbal Pauchard | Alicia Marticorena | 165 | 33 | NA | | SA-EC-001 | Ecuador Forest Plot Database | Jürgen Homeier | NA | 172 | 156 | NA |

Table: 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). {#tbl:Table2 tag="t2"} | Variable | Range/Levels | No.records | Type.of.variable | | :-----|:

----- :-----	GIVD ID NA 91031 character	Dataset NA 91031 character	Continent Africa, Asia, Australia, Europe, North America, Oceania, South America 90729 factor	Country NA 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 of recording -29764 - 16469 75798 numeric	Latitude -54.73863 - 80.149116 91031 numeric	Longitude -162.741433 - 179.590053 91031 numeric	Location uncertainty (m) 1 - 2500 91002 integer	POINT_X -162.741433 - 179.590053 91031 numeric	POINT_Y -54.73863 - 80.149116 91031 numeric	Relevé area (m²) 0.01 - 40000 61898 numeric	Herbs identified (y/n) FALSE = 4876; TRUE = 6323 11199 logical	Mosses identified (y/n) FALSE = 19707; TRUE = 4963 24670 logical	Lichens identified (y/n) FALSE = 16027; TRUE = 3045 19072 logical	Plants 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 (m) -25 - 4819 52121 numeric	Aspect (°) 0 - 360 30796 numeric	Slope (°) -5 - 99 37878 numeric	is.forest FALSE = 20396; TRUE = 25832 46228 logical	is.non.forest FALSE = 50870; TRUE = 38203 89073 logical	ESY NA 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 rock (%) 1 - 100 1897 integer	Cover cryptogams (%) 1 - 95 593 integer	Cover bare soil (%) 0.1 - 99 1412 numeric	Height (highest) trees (m) 1 - 99 6115 numeric	Height lowest trees (m) 1 - 90 221 numeric	Height (highest) shrubs (m) 0.1 - 9.9 2880 numeric	Height lowest shrubs (m) 0.1 - 9 328 numeric	Aver. height (high) herbs (cm) 0.1 - 440 10125 numeric	Aver. height lowest herbs (cm) 1 - 250 2785 integer	Maximum height herbs (cm) 1 - 600 1733 integer
--------------	----------------------------------	----------------------------------	---	----------------------------------	---	--	--	--	---	--	---	---	--	--	---	--	---	--	---------------------------------------	---	---	------------------------------	---------------------------------------	--	---	---	---	---	---	---	---	--	---	---	---	---	--	---	---	--	--	--	--	--	---	--