



OPEN

DATA DESCRIPTOR

A global dataset of terrestrial biological nitrogen fixation

Carla R. Reis Ely^{1,2}✉, Steven S. Perakis³✉, Cory C. Cleveland⁴, Duncan N. L. Menge⁵, Sasha C. Reed⁶, Sarah A. Batterman^{7,8,9}, Timothy E. Crews¹⁰, Katherine A. Dynarski⁴, Maga Gei¹¹, Michael J. Gundale¹², Sarah E. Jovan¹³, Sian Kou-Giesbrecht¹⁴, Mark B. Peoples¹⁵, Emilio Rodríguez-Caballero^{16,17}, Verity G. Salmon¹⁸, Fiona M. Soper¹⁹, Anika P. Staccone²⁰, Benton N. Taylor²¹, Bettina Weber^{17,22} & Nina Wurzburger²³

Biological nitrogen fixation (BNF) is the main natural source of new nitrogen inputs in terrestrial ecosystems, supporting terrestrial productivity, carbon uptake, and other Earth system processes. We assembled a comprehensive global dataset of field measurements of BNF in all major N-fixing niches across natural terrestrial biomes derived from the analysis of 376 BNF studies. The dataset comprises 32 variables, including site location, biome type, N-fixing niche, sampling year, quantification method, BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$), the percentage of nitrogen derived from the atmosphere ($\%N_{\text{dfa}}$), N fixer or N-fixing substrate abundance, BNF rate per unit of N fixer abundance, and species identity. Overall, the dataset combines 1,207 BNF rates for trees, shrubs, herbs, soil, leaf litter, woody litter, dead wood, mosses, lichens, and biocrusts, 152 herb $\%N_{\text{dfa}}$ values, 1,005 measurements of N fixer or N-fixing substrate abundance, and 762 BNF rates per unit of N fixer abundance for a total of 424 species across 66 countries. This dataset facilitates synthesis, meta-analysis, upscaling, and model benchmarking of BNF fluxes at multiple spatial scales.

Background & Summary

Nitrogen (N) availability is one of the main factors regulating terrestrial productivity, carbon (C) uptake, and organic matter decomposition in the biosphere^{1–3}. Biological N fixation (BNF) is the process whereby atmospheric dinitrogen gas (N_2) is converted into biologically available N. BNF is carried out by specific prokaryotes that possess the enzyme nitrogenase. N-fixing prokaryotes occupy myriad N-fixing niches, both as free-living bacteria (e.g., in soil, litter, and dead wood) and in symbiosis or association with plants or other organisms (e.g.,

¹Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR, 97331, USA. ²Oak Ridge Institute for Science and Education, Oak Ridge, TN, 37830, USA. ³United States Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, OR, 97331, USA. ⁴Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT, 59812, USA. ⁵Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY, 10027, USA. ⁶United States Geological Survey, Southwest Biological Science Center, Moab, UT, 84532, USA. ⁷Cary Institute of Ecosystem Studies, Millbrook, NY, 12545, USA. ⁸School of Geography, University of Leeds, Leeds, UK. ⁹Smithsonian Tropical Research Institute, Ancon, Panama. ¹⁰The Land Institute, Salina, KS, 67401, USA. ¹¹Association for Tropical Biology and Conservation, Minneapolis, MN, 55405, USA. ¹²Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, 90183, Umea, Sweden. ¹³USDA Forest Service, PNW Research Station, Portland, OR, USA. ¹⁴Department of Earth and Environmental Sciences, Dalhousie University, Halifax, NS, B3H 4R2, Canada. ¹⁵CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT 2601, Australia. ¹⁶Centro de Investigación de Colecciones Científicas de la Universidad de Almería (CECOUAL) y Departamento de Agronomía, Universidad de Almería, Almería, Spain. ¹⁷Multiphase Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany. ¹⁸Environmental Science Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA. ¹⁹Department of Biology and School of the Environment, McGill University, Montreal, QC, H3A 1B1, Canada. ²⁰Earthshot Labs, Sebastapol, CA, 95472, USA. ²¹Department of Organismic and Evolutionary Biology, Harvard University, 1300 Centre St., Rosindale, MA, 02131, USA. ²²Division of Plant Sciences, Institute for Biology, University of Graz, Graz, Austria. ²³Odum School of Ecology, University of Georgia, Athens, GA, 30602, USA. ✉e-mail: carlargreis@gmail.com; sperakis@usgs.gov

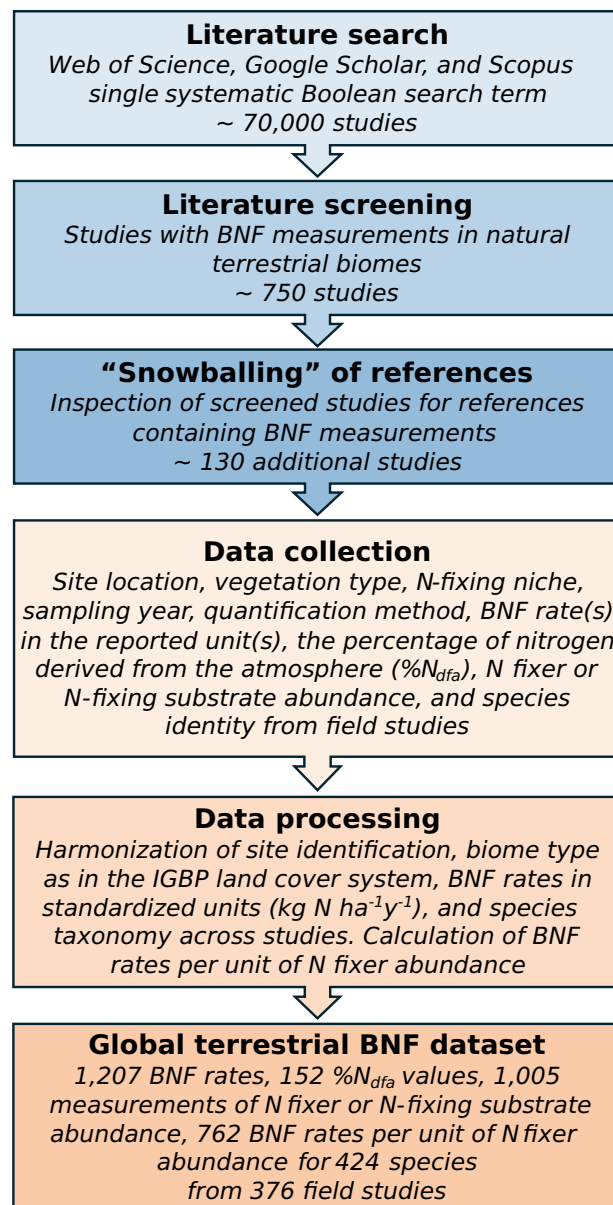


Fig. 1 Workflow and systematic approaches for developing the global dataset of biological nitrogen fixation (BNF) in natural terrestrial biomes¹¹.

trees, shrubs, herbs, mosses, lichens, and biocrusts)^{4,5}. BNF represents the primary natural source of new N inputs in most terrestrial ecosystems worldwide, replenishing N losses and supporting new growth^{6,7}. However, terrestrial BNF estimates at multiple spatial scales remain considerably uncertain^{8–10}. This uncertainty is partly due to the challenge of constraining BNF measurements across the multitude of N-fixing niches within ecosystems^{4,5} and may reflect sampling bias in field BNF studies that favor geographic hotspots where N fixers and N-fixing substrates (*i.e.*, the material in which N fixation occurs, for example, dead wood) are abundant⁸. A comprehensive accounting of BNF fluxes by individual N-fixing niches and their abundances is essential for improving accuracy and reducing uncertainty in terrestrial BNF estimates.

Here, we assembled a global dataset of field measurements of BNF in all major N-fixing niches across natural terrestrial biomes¹¹. Using systematic approaches of literature search and data collection and processing (Fig. 1), we generated a dataset with 32 variables, including site location, the International Geosphere-Biosphere Program (IGBP) land cover class¹², N-fixing niche (root-nodulating N-fixing trees, shrubs, and herbs, free-living BNF in soil, leaf litter, woody litter, dead wood, and BNF associated with mosses, lichens, and biocrusts), sampling year, quantification method, BNF rate (kg N ha⁻¹ y⁻¹), the percentage of N derived from the atmosphere (%N_{dfa}), N fixer or N-fixing substrate abundance, BNF rate per unit of N fixer abundance, and species identity. The dataset includes 376 BNF studies and combines 1,359 BNF measurements (1,207 BNF rates and 152 %N_{dfa} values) (Fig. 2), 1,005 measurements of N fixer or N-fixing substrate abundance, and 762 BNF rates per unit of N fixer abundance for a total of 424 species across 66 countries. This dataset facilitates synthesis, meta-analysis,

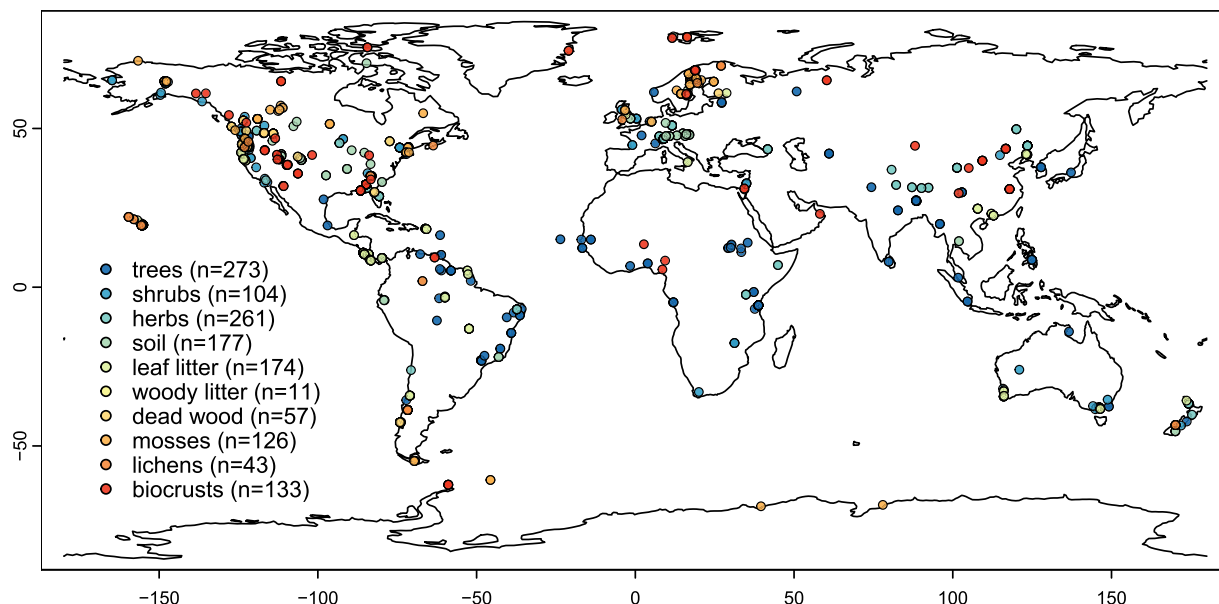


Fig. 2 Location of field measurements of BNF in the global BNF dataset in natural terrestrial biomes¹¹. Each point represents one BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$) and/or percentage of nitrogen derived from the atmosphere ($\%N_{\text{dfa}}$) value ($n = 1,359$) for root-nodulating N-fixing trees, shrubs, or herbs, free-living BNF in soil, leaf litter, woody litter, or dead wood, or BNF associated with mosses, lichens, or biocrusts. Not all points are visible due to overlap. Maps for each niche are in Supplementary Figures S2 to S4.

upsampling, and model benchmarking of BNF rates, as well as evaluation of the effects of environmental change on BNF fluxes at multiple spatial scales^{4,5,8}. To our knowledge, no other dataset with site-level BNF rates in natural terrestrial ecosystems is currently available in repositories or data journals. Global gridded datasets of natural terrestrial BNF fluxes upscaled using the BNF dataset presented here, and spatially explicit abundances of N-fixing niches globally are available at <https://doi.org/10.5066/P13THKNR>^{8,13}.

Methods

Literature search. We performed a literature search using a single systematic Boolean search term (*i.e.*, using “AND” and “OR” filters) to combine N fixation terms with lists of natural terrestrial biomes, countries, N-fixing taxa, and N-fixing niches (Supplementary Text S1). We used this search term in the software Publish or Perish (<https://harzing.com/resources/publish-or-perish>) to search for BNF publications in the Web of Science (<https://www.webofscience.com>), Google Scholar (<https://scholar.google.com>), and Scopus (<https://www.scopus.com>) databases. This search resulted in a list with ~ 70,000 references (Fig. 1). We screened studies manually and selected those reporting BNF measurements in natural and semi-natural terrestrial ecosystems. We selected studies in unmanaged and managed ecosystems (*e.g.*, silviculture/forestry, rangelands, and agroforestry systems) but excluded studies in grain croplands and forage/livestock-intensive production systems due to more intensive human interference. This screening resulted in ~750 studies that met the criteria for inclusion in the dataset. We also used a snowballing approach wherein we inspected the screened studies to help identify additional publications containing BNF measurements that were not located in the primary search. This snowballing added ~ 130 studies to the dataset. The complete dataset includes ~880 studies conducted in field, growth chamber, greenhouse, shade house, and/or laboratory settings. Here, we describe the subset dataset of 376 studies carried out in the field or other settings simulating field conditions (hereafter “field studies”). The sources included journal articles, dissertations, theses, book chapters, and technical reports published in English, Spanish, French, Chinese, or Russian between 1955 and 2020.

Data collection. We extracted data manually into a spreadsheet template⁵. Data published as figures were extracted using the software WebPlotDigitizer (<https://automeris.io>). We recorded BNF measurements from observations and control treatments of field studies. We also recorded BNF measurements for treatments using prescribed fire or grazing in ecosystems where these disturbances occur naturally. For managed ecosystems, we also recorded BNF measurements for treatments using pruning or thinning but excluded data from recent clear-cuts and young plantations (<1 year old) due to more intensive human disturbance. We also excluded data for pure cultures of N-fixing prokaryotes isolated from field samples, as these do not represent N-fixing activity in the presence of a symbiont and/or community.

For each study we recorded the site name, country name (later converted to ISO code), geographic coordinates in decimal degrees, vegetation type as described by authors (later converted to an IGBP land cover class; see Data processing), and N-fixing niche (root-nodulating N-fixing trees, shrubs, and herbs [legume forbs], free-living BNF in soil [mineral and/or organic soil, including humus], litter [all non-woody litter, usually leaf litter], woody litter [dead branches and stems ≤ 7.5 cm diameter], and dead wood [logs >7.5 cm diameter], and

BNF associated with mosses, lichens, and biocrusts). We also recorded the sampling year(s), quantification method(s) (acetylene reduction assay (ARA) [indirect measurement of nitrogenase enzyme activity based on the reduction of acetylene to ethylene by nitrogenase, and the rate of ethylene accumulation], $^{15}\text{N}_2$ incorporation [measurement of nitrogenase enzyme activity based on the rate of incorporation of ^{15}N -labelled N_2 into tissue], ^{15}N natural abundance [measurement of the proportion of N derived from BNF based on the N stable isotopic composition of tissues], ^{15}N dilution [measurement of the dilution of an applied ^{15}N isotopic label by fixed N], N accumulation [measurement of the difference in N content of a system at two points in time], N accretion [measurement of the difference in N content in a chronosequence to approximate the accumulation in a single site], and mass balance [measurement of the difference between other N fluxes and accrual in a bounded system])^{5,14,15}, reported BNF measurements and their units, N fixer or N-fixing substrate abundance from on-site surveys (N fixer relative basal area and/or relative stem density for trees, N fixer percent ground cover for shrubs, herbs, mosses, and biocrusts, and biomass for dead wood), and N fixer family, genus, and species names (Fig. 1 and Table 1). Additionally, we recorded sampled soil depth and bulk density, which we used in soil BNF rate conversions (see Data Processing). Finally, we recorded whether the studied mosses and lichens were epiphytes or located on the ground (“habit”). For field studies relying on methods that require sample incubation (*i.e.*, ARA and $^{15}\text{N}_2$ incorporation), we recorded only BNF measurements taken at temperature and moisture levels within the range experienced in the field at any given time in the year, and in the absence of supplemental C or nutrient additions. We note that the number of significant figures of numeric variables varies among observations, reflecting the precision of reported values.

Data processing. Multiple entries in the dataset were either from the same administrative research areas or the same study sites within research areas. We harmonized these location names so that the same research areas and study sites have the same names across the dataset. We identified unique study sites as plots/stands that differed in vegetation composition or dominance, age, disturbance history, N fixer abundance, elevation/climate, and/or topographic setting within the same research area. Although we identified individual study sites as plots/stands with different characteristics within research areas, these can be aggregated into larger spatial scales depending on the study goals (*e.g.*, by research area or within a radius or grid cell). We filled gaps in geographic coordinates, vegetation type, N fixer or N-fixing substrate abundance, and soil bulk density using information gathered from sister publications by the same authors, provided directly by authors, or from other publications in the same study site/research area, whenever possible and appropriate.

We harmonized the vegetation type within and across study sites into one of the IGBP land cover classes¹² (Supplementary Figure S1) based on the reported vegetation descriptions and definitions in the IGBP land cover system (Supplementary Table S2). For plantations of N-fixing species, we considered IGBP land cover classes where the N-fixing species naturally occur. For plantations of early-mid seral N-fixing species and plots/stands under ecological succession, we classified the land cover class as the corresponding late-successional vegetation type. We combined “Closed Shrublands” and “Open Shrublands” as “Shrublands” and combined “Woody Savannas” and “Savannas” as “Savannas” because most BNF studies lacked the necessary information for these finer classifications. We classified bogs and fens as “Permanent Wetlands” due to their water table being constantly close to the surface, and marshes as “Grasslands” and swamps as a forest biome, due to their more dynamic water table.

We standardized and updated family, genus, and species names in the dataset using The World Flora Online (WFO) database (<http://www.worldfloraonline.org>) and the Taxonomic Name Resolution Service (TNRS) online tool (<https://tnrs.biendata.org>)¹⁶ for species of trees, shrubs, herbs, mosses, and lichens. We checked names individually and used matched accepted names or accepted names converted from synonyms. For lichen species not listed in the WFO database, we checked for accepted names in the Consortium of Lichen Herbaria (<https://lichenportal.org/portal/>). We also used the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/taxonomy>) and Algaebase (<https://www.algaebase.org>) taxonomy databases to check for accepted names of prokaryotes. Unresolved names (<0.1%) were kept as reported in the original publication.

Published BNF rates were reported in 48 different units (Table 1). We provide BNF rates both in the reported units and standardized to $\text{kg N ha}^{-1} \text{y}^{-1}$. The rates in the reported units are either the exact reported values or the mean or sum of reported group values. We averaged reported rates obtained using different methods or incubation conditions. We also averaged reported rates for replicates and seasonal or annual rates. We either summed or averaged rates to aggregate species-level into community-level rates. We summed species-level rates calculated using the local abundances of each species, while we averaged species-level rates calculated using 100% coverage (*e.g.*, from samples or assumed). We also summed rates for different soil layers or depth intervals, and litter or dead wood classes or decomposition stages. The standardized BNF rates were either originally reported in $\text{kg N ha}^{-1} \text{y}^{-1}$ or similar units (*e.g.*, $\text{g N m}^{-2} \text{y}^{-1}$) ($n = 1,073$) or were converted to these units from cumulative or sub-annual BNF or ARA rates per unit of area or N fixer abundance, or $\%N_{\text{dfa}}$ values ($n = 134$) (Fig. 3), as detailed below for each niche. All standardized BNF rates were adjusted to the local abundance of N fixers and N-fixing substrates, either by the authors in the original publication or in our analysis.

For **trees**, we converted cumulative BNF rates across multiple years to mean annual rates by dividing by the number of years, though any calculation of mean rates might obscure interannual rate variation. We converted BNF rates per unit of N-fixing tree abundance to rates per unit of area using local N-fixing tree abundance data. We converted tree BNF rates per unit of nodule mass to rates per unit of area using local nodule biomass data for studies that measured both BNF activity and nodule biomass throughout the entire growing season. We also converted tree $\%N_{\text{dfa}}$ to BNF rates per area using local N fixer abundance and N content data.

Variable	Type	Unit	Description	# Records	# Values	Range
record_ID	Integer	NA	Record identification code	1,529	1,529	1; 1,529
site_name	Character	NA	Study site name	1,529	987	NA
country	Character	NA	3-digit ISO country code	1,529	66	NA
lat	Numeric	°	Latitude coordinate	1,529	500 [†]	−69; 79
lon	Numeric	°	Longitude coordinate	1,529	500 [†]	−164.8262; 175.37
IGBP	Character	NA	International Geosphere-Biosphere Program land cover class	1,528	10	NA
niche	Character	NA	N-fixing niche	1,529	10	NA
ecological_level	Character	NA	If species or community-level data	1,240	2	NA
year	Integer	NA	Sampling year	1,301	57	1952; 2018
method	Integer	NA	Quantification method(s) code	1,359	13	10; 70
BNF_central	Numeric	varies, as reported	Central BNF rate reported	1,188	739	0; 9,432.6
BNF_min	Numeric	varies, as reported	Minimum BNF rate reported	37	29	0; 282
BNF_max	Numeric	varies, as reported	Maximum BNF rate reported	39	34	0.13; 363
BNF_unit	Character	varies, as reported	The unit of reported BNF rate(s)	1,207	48	NA
BNF_final	Numeric	kg N ha ^{−1} y ^{−1}	Final BNF rate in standardized units, based on reported BNF rate(s) and local abundance of N fixers or N-fixing substrate	1,188	718	0; 361.6
BNF_final_type	Integer	NA	Type of final BNF rate code: if originally reported or converted from similar units or other types of BNF data	1,188	6	1; 6
pct_BA	Numeric	%	N fixer relative basal area	269	85	0.8; 100
pct_stems	Numeric	%	N fixer relative stem density	282	66	0.69; 100
pct_cover	Numeric	%	N fixer percent ground cover	401	227	0; 100
AGB	Numeric	Mg ha ^{−1}	N-fixing substrate aboveground biomass	53	45	0.66; 296.2
BNF_BA	Numeric	kg N ha ^{−1} y ^{−1} %BA ^{−1}	BNF rate per unit of N fixer relative basal area	221	173	0; 4.625
BNF_stems	Numeric	kg N ha ^{−1} y ^{−1} %stems ^{−1}	BNF rate per unit of N fixer relative stem density	230	178	0; 7.602
BNF_cover	Numeric	kg N ha ^{−1} y ^{−1} %cover ^{−1}	BNF rate per unit of N fixer percent ground cover	311	217	0; 11.2
NDFA	Numeric	%	Percentage of nitrogen derived from atmosphere (%N _{dfa})	152	84	0.63; 100
family	Character	NA	Taxonomic family	2,378	85	NA
genus	Character	NA	Taxonomic genus	2,378	233	NA
species	Character	NA	Taxonomic species	2,186	424 [‡]	NA
habit	Integer	NA	Habit of N-fixing mosses and lichens code: if epiphyte or on the ground	213	2	1; 2
soil_depth	Numeric	cm	Soil depth sampled	152	18	0.5; 40
bulk_density	Numeric	g cm ^{−3}	Soil bulk density	13	8	0.0000241; 1
ref_code	Character	NA	Source code	1,529	376	NA
ref_complete	Character	NA	Complete source reference	376	376	NA

Table 1. Variables in the global BNF dataset in natural terrestrial biomes¹¹. Detailed descriptions are in Supplementary Table S1. [†]Number of unique pair of coordinates. [‡]Number of unique species (*i.e.*, Genus + species names). “NA” stands for not applicable.

For *shrubs*, we converted cumulative BNF rates across multiple years to mean annual rates by division. When necessary, we converted shrub ARA rates per area to BNF rates using a conversion ratio of 3 mol of C₂H₄: 1 mol of N₂ (R ratio)¹⁷. We used an R ratio of 3:1 for shrubs and other niches (see below) as it is the most applied theoretical ratio and is the single ratio that represents the majority of BNF rate data, even though the distribution of ¹⁵N₂ calibrated R ratios from individual studies ranges widely above and below 3:1, which could affect individual BNF rates by 2 to >8 fold¹⁸. When available, direct calibrations provide more accurate results, so we used BNF rates estimated using direct calibrations rather than the assumed R ratio of 3:1 when presented in individual studies¹⁸. We also converted shrub sub-annual BNF or ARA rates to annual rates. We converted sub-annual ARA rates for studies that measured BNF activity at least twice during a growing season. We converted mean growing season rates to annual rates using 24 hours of activity per day and assumed growing season lengths (365 days for study sites located between latitudes 30N–30S, 180 days between 30–60 N/S, and 100 days between 60–90 N/S).

For *herbs*, we converted ARA rates per area to BNF rates using an R ratio of 3:1¹⁸. We also converted herb sub-annual BNF to annual rates as described above for shrubs.

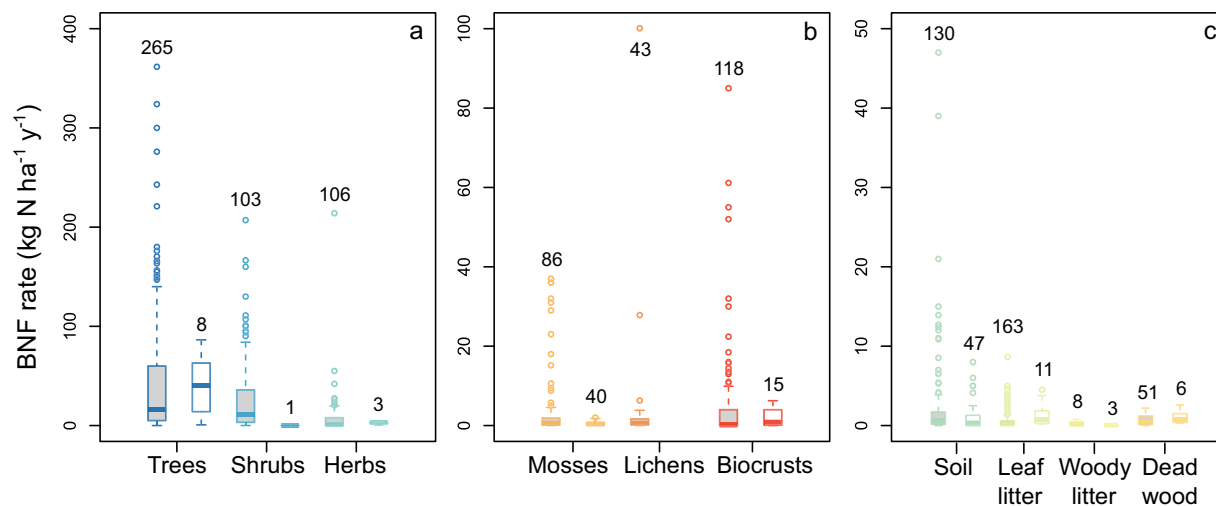


Fig. 3 BNF rates originally reported in $\text{kg N ha}^{-1} \text{y}^{-1}$ or similar units (grey filled boxplots) versus BNF rates converted to these units from cumulative or sub-annual BNF or ARA rates per unit of area or N fixer abundance, or $\%N_{\text{dfa}}$ values (white filled boxplots) for root-nodulating N-fixing trees, shrubs, and herbs (a), BNF associated with mosses, lichens, and biocrusts (b), and free-living BNF in soil, leaf litter, woody litter, and dead wood (c) in the global BNF dataset in natural terrestrial biomes¹¹. Boxplots depict the median and interquartile range, with whiskers representing values within 1.5 times the interquartile range. Values outside this range are shown as individual data points. The sample sizes are indicated above each boxplot.

For **soil**, we converted ARA to BNF rates using an R ratio of 3:1¹⁸. We converted ARA or BNF rates per unit of soil dry mass to rates per unit of area using local soil bulk density and applied these rates across the mean soil depth in our dataset (~ 10 cm, $n = 152$). If local soil bulk density was not available, we used the mean soil bulk density for the same biome type in the dataset. We also converted soil sub-annual ARA or BNF rates to annual rates as described above for other niches.

For **litter**, **woody litter**, and **dead wood**, we converted ARA to BNF rates using an R ratio of 3:1¹⁸. We converted ARA or BNF rates per unit of dry mass to rates per unit of area using local biomass data. We also converted sub-annual ARA or BNF rates to annual rates as described above for other niches.

For **mosses**, we converted ARA rates per unit of area to BNF rates using an R ratio of 3.3:1¹⁹. We did not use ARA rates for *Sphagnum* as these mosses exhibit a high abundance of methanotrophs, which can result in highly variable R ratios²⁰. We also converted moss sub-annual ARA to annual rates as described above for other niches.

For **biocrusts**, we converted ARA rates per unit of area to BNF rates using an R ratio of 3:1¹⁸. We also converted biocrust sub-annual ARA or BNF rates to annual rates using the mean number of days with precipitation > 1 mm per year (average of 1991–2020) from CPC Global Unified Gauge-Based Analysis of Daily Precipitation provided by the NOAA PSL (<https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html>)^{21,22}. We assumed 12 hours of BNF activity following a precipitation event for BNF measurements taken under light or dark conditions and 24 hours for measurements taken under both light and dark conditions.

We also calculated BNF rates per unit of N fixer abundance using the standardized BNF rates and reported local abundances of N fixers⁸ (Table 1). For trees, we calculated BNF rates per unit of N fixer relative basal area and/or N fixer relative stem density, *i.e.*, the BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$) when N-fixing tree relative basal area is 1% ($\text{kg N ha}^{-1} \text{y}^{-1} \%BA^{-1}$) and when N-fixing tree relative stem density is 1% ($\text{kg N ha}^{-1} \text{y}^{-1} \%stems^{-1}$), respectively. For shrubs, mosses, and biocrusts, we calculated BNF rates per 1% N fixer ground cover ($\text{kg N ha}^{-1} \text{y}^{-1} \%cover^{-1}$). BNF rates per unit of N fixer abundance are available for N-fixing herbs as $\%N_{\text{dfa}}$ values. BNF rates per unit of N fixer abundance indicate the additive increment for each 1% increase in N fixer abundance and allow scaling BNF rates per area using N fixer abundance data⁸. For example, for trees, $0.7 \text{ kg N ha}^{-1} \text{y}^{-1} \%BA^{-1}$ yields $0.7 \text{ kg N ha}^{-1} \text{y}^{-1}$ for 1% N-fixing tree relative basal area, $1.4 \text{ kg N ha}^{-1} \text{y}^{-1}$ for 2% N-fixing tree relative basal area, and $7 \text{ kg N ha}^{-1} \text{y}^{-1}$ for 10% N-fixing tree relative basal area.

For most niches, we provide species-level data for studies with one N-fixing species or community-level data for studies with more than one N-fixing species. For N-fixing herbs and mosses, we provide both species and community-level data, the latter weighted by the abundance of N fixers whenever species-level abundance data was available.

Data Records

The dataset is available in the ScienceBase repository (<https://doi.org/10.5066/P1MFVBVHK>)¹¹. It consists of 8 data files in the “csv” format, each combining a subset of the variables (Table 1):

The “SITE.csv” file includes the variables study site name, ISO country code, geographical coordinates in decimal degrees, IGBP land cover class, N-fixing niche, ecological level (if species or community-level data), and the source code that links data to BNF study references in the “REFERENCES.csv” file, listed below.

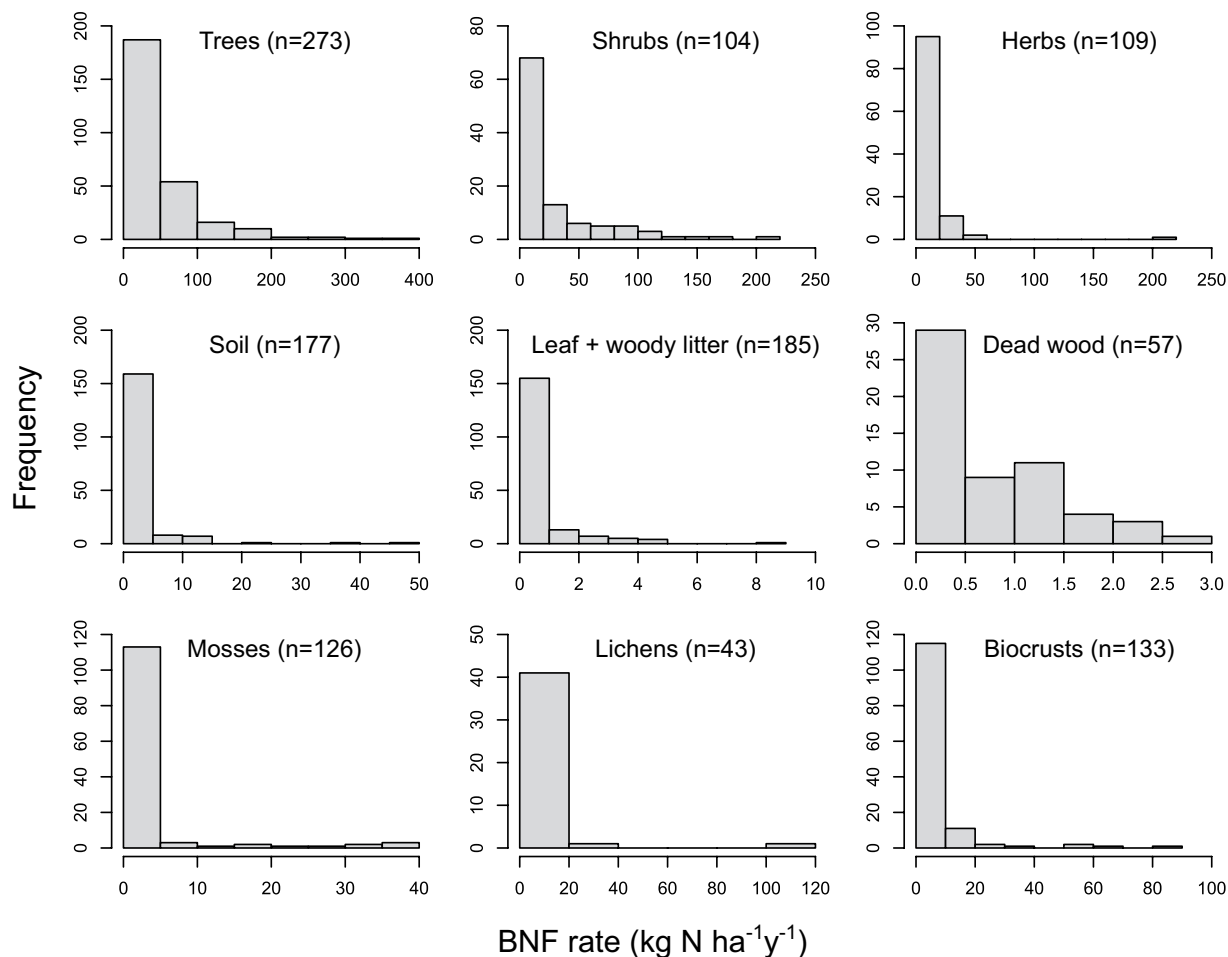


Fig. 4 Frequency distribution of BNF rates (kg N ha⁻¹ y⁻¹) for root-nodulating N-fixing trees, shrubs, and herbs, free-living BNF in soil, leaf litter, woody litter, and dead wood, and BNF associated with mosses, lichens, and biocrusts in the global BNF dataset in natural terrestrial biomes¹¹. The sample sizes are indicated in parentheses.

The “BNF_AREA.csv” file includes the variables central, minimum, and maximum reported BNF rates and their unit, sampling year, and the quantification method(s) of BNF rates. This file also provides the variable final BNF rate in standardized units (kg N ha⁻¹ y⁻¹) and a variable that indicates whether final BNF rates were originally reported in kg N ha⁻¹ y⁻¹ or converted from similar units or other types of BNF data.

The “NDFA.csv” file provides the variables herb %N_{dfa}, sampling year, and the quantification method(s) of %N_{dfa} values.

The “ABUNDANCE.csv” file includes the variables relative basal area and relative stem density for N-fixing trees, the percent ground cover for N-fixing shrubs, herbs, mosses, and biocrusts, and aboveground biomass for dead wood.

The “BNF_ABUNDANCE.csv” file provides the variables BNF rate per 1% N-fixing tree relative basal area, BNF rate per 1% N-fixing tree relative stem density, and BNF rate per 1% ground cover for N-fixing shrubs, herbs, mosses, and biocrusts.

The “SPECIES.csv” file lists the names of taxonomic families, genera, and species of N fixers.

The “AUXILIARY_DATA.csv” file has additional variables, including soil bulk density and sampled soil depth that supported rate conversions (see Data processing) and the habit of N-fixing mosses and lichens (if epiphyte or on the ground).

The “REFERENCES.csv” file lists the BNF study codes and complete references.

Except for the “REFERENCES.csv” file, all other data files include the variable “record identification code” linking data across files. Most files, namely “SITE.csv,” “BNF_AREA.csv,” “NDFA.csv,” “SPECIES.csv,” and “REFERENCES.csv” have individual variables as columns and observations as rows. The files “ABUNDANCE.csv,” “BNF_ABUNDANCE.csv,” and “AUXILIARY_DATA.csv” are “molten” files, combining variables into a single column and observations into another column²³.

Detailed descriptions of all variables and values are in the accompanying “METADATA.csv” file and Supplementary Table S1. For manipulating the data using R programming language, see the accompanying “RCode_Data_manipulation_examples.txt” file. For manipulating the data using spreadsheet software, see Supplementary Text S2.

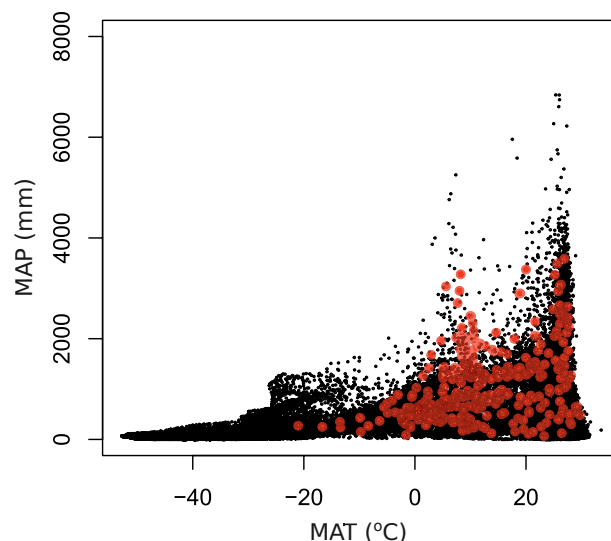


Fig. 5 Climate space of study sites in the global BNF dataset¹¹. The background black points represent the density of mean annual temperature (MAT) and mean annual precipitation (MAP) grid cell values at 0.004-degree resolution across natural terrestrial biomes from TerraClimate²⁶ (average of 2000–2020), including areas permanently covered by snow or ice. Study sites in the global dataset are depicted as red points.

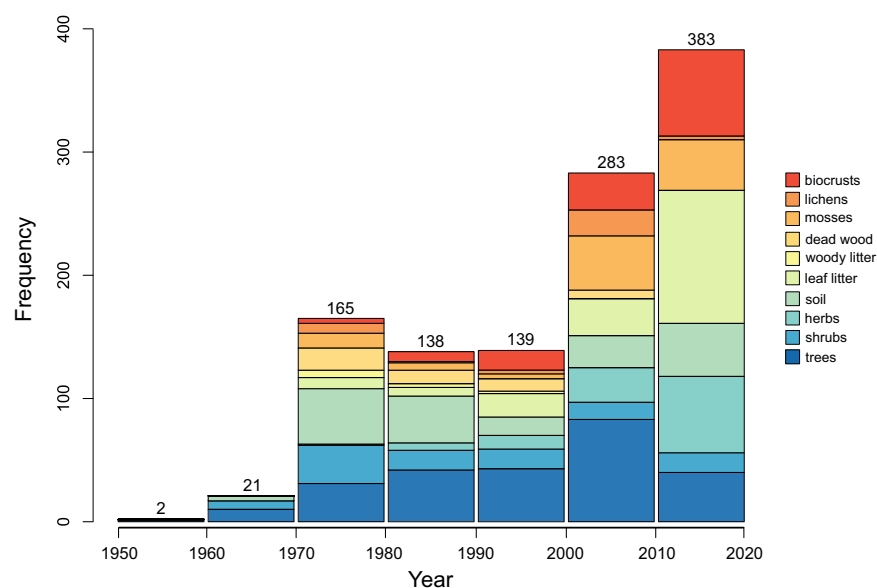


Fig. 6 Frequency distribution of sampling year of BNF rates ($\text{kg N ha}^{-1} \text{y}^{-1}$) for root-nodulating N-fixing trees, shrubs, and herbs, free-living BNF in soil, leaf litter, woody litter, and dead wood, and BNF associated with mosses, lichens, and biocrusts in the global BNF dataset in natural terrestrial biomes. Totals by decade are indicated above each bar.

Technical Validation

We reviewed and removed duplicate studies during the literature screening. If the same data were published in multiple formats, we selected data published as a journal article over other formats. We checked each data entry against its original publication at least twice, with particular attention to whether potential outliers represented data entry errors. We did not exclude potential outliers from the dataset as extreme values can represent hotspots or hot moments of BNF activity, which can account for a large proportion of BNF in natural ecosystems^{5,24}.

We also checked BNF rate conversion calculations at least twice and enforced systematic procedures within and across niches, as described in Data Processing. All BNF rates converted from cumulative or sub-annual BNF or ARA rates per unit of area or N fixer abundance, or $\%N_{\text{dfa}}$ values, are within the range of BNF rates originally reported in $\text{kg N ha}^{-1} \text{y}^{-1}$ or similar units (Fig. 3). The BNF rates in the dataset show a right-skewed frequency distribution across niches (Fig. 4), indicating frequent low and infrequent high values, typical of BNF fluxes^{5,25}.

The location of study sites in the dataset spans all continents but with a larger representation of mid-latitude areas in the Northern Hemisphere. Still, the location of study sites encompasses nearly the entire global range of temperature and precipitation (average of 2000–2020 from TerraClimate²⁶) in natural terrestrial biomes (Fig. 5), except for extremely cold locations ($\text{MAT} < -20^\circ\text{C}$) where BNF is likely very low, and a handful of extremely wet locations ($\text{MAP} > 4000\text{ mm}$). The extensive coverage of Earth's climate space suggests that the dataset is broadly representative of the heterogeneity of environmental conditions that drive BNF rates. Most of the BNF rates were measured between the years 2000 and 2020 (59%) (Fig. 6).

Code availability

Data visualization and manipulation were conducted using standard code packages in R version 4.3.0²⁷. To facilitate user data manipulation, we provide accompanying R code with examples at <https://doi.org/10.5066/P1MFBVHK11>.

Received: 16 February 2025; Accepted: 1 May 2025;

Published online: 05 August 2025

References

1. LeBauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**, 371–379, <https://doi.org/10.1890/06-2057.1> (2008).
2. Du, E. *et al.* Global patterns of terrestrial nitrogen and phosphorus limitation. *Nat Geosci* **13**, 221–226, <https://doi.org/10.1038/s41561-019-0530-4> (2020).
3. Craine, J. M., Morrow, C. & Fierer, N. Microbial nitrogen limitation increases decomposition. *Ecology* **88**, 2105–2113, <https://doi.org/10.1890/06-1847.1> (2007).
4. Cleveland, C. C. *et al.* Exploring the Role of Cryptic Nitrogen Fixers in Terrestrial Ecosystems: A Frontier in Nitrogen Cycling Research. *Ecosystems* **25**, 1653–1669, <https://doi.org/10.1007/s10021-022-00804-2> (2022).
5. Soper, F. M. *et al.* A roadmap for sampling and scaling biological nitrogen fixation in terrestrial ecosystems. *Methods Ecol Evol* **12**, 1122–1137, <https://doi.org/10.1111/2041-210X.13586> (2021).
6. Cleveland, C. C. *et al.* Patterns of new versus recycled primary production in the terrestrial biosphere. *Proc Natl Acad Sci USA* **110**, 12733–12737, <https://doi.org/10.1073/pnas.1302768110> (2013).
7. Wieder, W. R., Cleveland, C. C., Smith, W. K. & Todd-Brown, K. Future productivity and carbon storage limited by terrestrial nutrient availability. *Nat Geosci* **8**, 441–444, <https://doi.org/10.1038/ngeo2413> (2015).
8. Reis Ely, C. R. *et al.* Global terrestrial biological nitrogen fixation and its modification by agriculture. *Nature*, accepted, (2025).
9. Reis, C. R. G. *et al.* Biological nitrogen fixation across major biomes in Latin America: Patterns and global change effects. *Sci Total Environ* **746**, <https://doi.org/10.1016/j.scitotenv.2020.140998> (2020).
10. Staccone, A. *et al.* A Spatially Explicit, Empirical Estimate of Tree-Based Biological Nitrogen Fixation in Forests of the United States. *Global Biogeochem Cy* **34**, <https://doi.org/10.1029/2019GB006241> (2020).
11. Reis Ely, C. R. *et al.* A global dataset of terrestrial biological nitrogen fixation: U.S. Geological Survey data release, <https://doi.org/10.5066/P1MFBVHK> (2025).
12. Friedl, M. A. & Sulla-Menashe, D. MCD12Q1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/MODIS/MCD12Q1.006> (2019).
13. Reis Ely, C. R. *et al.* Global gridded dataset of terrestrial biological nitrogen fixation across natural and agricultural biomes (rasters): U.S. Geological Survey data release, <https://doi.org/10.5066/P13THKNR> (2025).
14. Boddey, R. *et al.* Use of the ^{15}N natural abundance technique to quantify biological nitrogen fixation by woody perennials. *Nutr Cycl Agroecosyst* **57**, 235–270, <https://doi.org/10.1023/A:1009890514844> (2000).
15. Unkovich, M. *et al.* Measuring plant-associated nitrogen fixation in agricultural systems. (Australian Center for International Agricultural Research, 2008).
16. Boyle, B. L. *et al.* The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**, 16, <https://doi.org/10.1186/1471-2105-14-16> (2013).
17. Hardy, R. W. F. *et al.* The acetylene-ethylene assay for N_2 fixation: Laboratory and field evaluation. *Plant Physiol* **43**, 1185–1207, <https://doi.org/10.1104/pp.43.8.1185> (1968).
18. Soper, F. M., Simon, C. & Jauss, V. Measuring nitrogen fixation by the acetylene reduction assay (ARA): is 3 the magic ratio? *Biogeochemistry* **152**, 345–351, <https://doi.org/10.1007/s10533-021-00761-3> (2021).
19. Saiz, E., Sgouridis, F., Drijfhout, F. P. & Ullah, S. Biological nitrogen fixation in peatlands: Comparison between acetylene reduction assay and $^{15}\text{N}_2$ assimilation methods. *Soil Biol Biochem* **131**, 157–165, <https://doi.org/10.1016/j.soilbio.2019.01.011> (2019).
20. Leppänen, S. M., Salemaa, M., Smolander, A., Mäkipää, R. & Tirola, M. Nitrogen fixation and methanotrophy in forest mosses along a N deposition gradient. *Environ Exp Bot* **90**, 62–69, <https://doi.org/10.1016/j.envexpbot.2012.12.006> (2013).
21. Chen, M. *et al.* Assessing objective techniques for gauge-based analyses of global daily precipitation. *J Geophys Res Atmos* **113**, <https://doi.org/10.1029/2007JD009132> (2008).
22. Xie, P. *et al.* A gauge-based analysis of daily precipitation over East Asia. *J Hydrometeorol* **8**, 607–626, <https://doi.org/10.1175/JHM583.1> (2007).
23. Wickham, H. Reshaping Data with the reshape Package. *J Stat Softw* **21**, <https://doi.org/10.18637/jss.v021.i12> (2007).
24. Reed, S. C., Cleveland, C. C. & Townsend, A. R. Functional ecology of free-living nitrogen fixation: a contemporary perspective. *Annu Rev Ecol Evol Syst* **42**, 489–512, <https://doi.org/10.1146/annurev-ecolsys-102710-145034> (2011).
25. Winbourne, J. B. *et al.* A new framework for evaluating estimates of symbiotic nitrogen fixation in forests. *Am Nat* **192**, 618–629, <https://doi.org/10.1086/699828> (2018).
26. Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci Data* **5**, <https://doi.org/10.1038/sdata.2017.191> (2018).
27. R Core Team. R: A Language and Environment for Statistical Computing. <https://www.R-project.org/> (2023).

Acknowledgements

This paper is a contribution from a working group on biological nitrogen fixation supported by the U.S. Geological Survey (USGS) John Wesley Powell Center for Analysis and Synthesis (G19AC00036, S.S.P., D.N.L.M., C.C.C., S.C.R.). This research was supported in part by an appointment to the USGS Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between U.S. Departments of Energy (DOE) and Interior (DOI) under DOE contract number DE-SC0014664. S.A.B. was funded by the Leverhulme Trust and the United Kingdom Natural Environment Research

Council (NE/M019497/1, NE/N012542/1). E.R-C was supported by the Ramon y Cajal fellowship (RYC2020-030762-I) and CRUST R-Forze (PID2021-127631NA-I00) project, both funded by FEDER/Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación. We thank E. Drumm, C. Lauria, J. Shostrand, I. van Duuren, M. Blackburn, and K. Copeland for their help with data collection. This manuscript was reviewed by USGS and USDA Forest Service for technical and policy content and approved for publication. We thank Linda Geiser for her comments. The views and conclusions in this article represent the views solely of the authors from ORAU/ORISE and USDA Forest Service but do represent the views of USGS. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions

C. Reis Ely, S. Perakis, D. Menge, C. Cleveland, and S. Reed led the investigation. C. Reis Ely prepared the dataset files and wrote the first draft. All authors contributed to methods development, literature search, data collection and processing, and editing the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05131-4>.

Correspondence and requests for materials should be addressed to C.R.R. or S.S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

The Authors. Parts of this work were authored by US Federal Government authors and are not under copyright protection in the US; foreign copyright protection may apply 2025

Supplementary Information

Reis Ely *et al.* A global dataset of terrestrial biological nitrogen fixation. *Sci Data*, 10.1038/s41597-025-05131-4 (2025).

Contents

Supplementary Text S1, page 1

Supplementary Text S2, page 2

Supplementary Table S1, page 4

Supplementary Table S2, page 7

Supplementary Figure S1, page 8

Supplementary Figure S2, page 9

Supplementary Figure S3, page 10

Supplementary Figure S4, page 11

Supplementary Text S1. Systematic Boolean search term.

The search term structure:

(N fixation terms) AND (terrestrial biomes list OR country list OR N-fixing taxa list OR N-fixing niches list)

("nitrogen fix*" OR "N2 fix*" OR "dinitrogen fix*" OR "N fix*" OR nitrogenase) AND ("temperate forest" OR "mixed forest" OR "tropical forest" OR "subtropical forest" OR rainforest OR "moist forest" OR "wet forest" OR "dry forest" OR arctic OR boreal OR tundra OR "high latitude" OR taiga OR subarctic OR grassland OR savanna OR desert OR shrubland OR dryland OR steppe OR meadow OR Cerrado OR Mediterranean OR fynbos OR chaparral OR matorral OR maquis OR kwongan OR "Antigua and Barbuda" OR "Bosnia and Herzegovina" OR "British Virgin Islands" OR "Burkina Faso" OR "Cabo Verde" OR "Cayman Islands" OR "Central African Republic" OR "Costa Rica" OR "Cote d'Ivoire" OR "Czech Republic" OR "Democratic Republic of the Congo" OR "Dominican Republic" OR "East Timor" OR "El Salvador" OR "Equatorial Guinea" OR "French Guiana" OR "Latin America" OR "Marshall Islands" OR "New Zealand" OR "North America" OR "North Macedonia" OR "Papua New Guinea" OR "Puerto Rico" OR "Saint Kitts and Nevis" OR "Saint Lucia" OR "Saint Vincent and the Grenadines" OR "San Marino" OR "Sao Tome and Principe" OR "Saudi Arabia" OR "Sierra Leone" OR "Solomon Islands" OR "South Africa" OR "South America" OR "South Sudan" OR "Sri Lanka" OR "State of Palestine" OR "Trinidad and Tobago" OR "Turks and Caicos" OR "U.S. Virgin Islands" OR "United Arab Emirates" OR "United Kingdom" OR "United States" OR Afghanistan OR Africa OR Alaska OR Albania OR Algeria OR Andorra OR Angola OR Anguilla OR Antarctica OR Argentina OR Armenia OR Aruba OR Asia OR Austria OR Australia OR Azerbaijan OR Bangladesh OR Bahamas OR Bahrain OR Barbados OR Belgium OR Belize OR Belarus OR Benin OR Bhutan OR Bolivia OR Botswana OR Brazil OR Brunei OR Burundi OR Bulgaria OR Cambodia OR Cameroon OR Canada OR Chad OR Chile OR China OR Colombia OR Comoros OR Congo OR Croatia OR Cuba OR Curacao OR Cyprus OR Djibouti OR Denmark OR Dominica OR Ecuador OR Egypt OR Eritrea OR Eswatini OR Estonia OR Ethiopia OR Europe OR Fiji OR Finland OR France OR Gabon OR Gambia OR Georgia OR Ghana OR Greece OR Greenland OR Grenada OR Guadeloupe OR Guatemala OR Guinea OR "Guinea-Bissau" OR Guyana OR Haiti OR Honduras OR Hungary OR Iceland OR India OR Indonesia OR Iran OR Iraq OR Ireland OR Israel OR Italy OR Jamaica OR Japan OR Jordan OR Kazakhstan OR Kenya OR Kiribati OR Korea OR Kuwait OR Kyrgyzstan OR Laos OR Latvia OR Lebanon OR Lesotho OR Liberia OR Libya OR Liechtenstein OR Lithuania OR Luxembourg OR Madagascar OR Malawi OR Malaysia OR Maldives OR Mali OR Malta OR Martinique OR Mauritania OR Mauritius OR Mexico OR Micronesia OR Moldova OR Monaco OR Mongolia OR Monserrat OR Montenegro OR Morocco OR Mozambique OR Myanmar OR Namibia OR Nauru OR Nepal OR Netherlands OR Nicaragua OR Niger OR Nigeria OR Norway OR Oman OR Palau OR Pakistan OR Panama OR Paraguay OR Peru OR Philippines OR Portugal OR Poland OR Qatar OR Romania OR Russia OR Rwanda OR Samoa OR Senegal OR Serbia OR Seychelles OR Singapore OR Slovenia OR Slovakia OR Somalia OR Spain OR Sudan OR Suriname OR Sweden OR Switzerland OR Syria OR Tajikistan OR Tanzania OR Thailand OR Togo OR Tunisia OR Turkey OR Turkmenistan OR Tuvalu OR Tonga OR Uganda OR UK OR Ukraine OR Uruguay OR US OR USA OR Uzbekistan OR Vanuatu OR Venezuela OR Vietnam OR Yemen OR Zambia OR Zimbabwe OR legum* OR Fabaceae OR actinorhizal OR Coriaria OR Datisca OR Alnus OR Allocasuarina OR Casuarina OR Ceuthostoma OR Gymnostoma OR Comptonia OR Myrica OR Morella OR Elaeagnus OR Hippophae OR Shepherdia OR Colletia OR Discaria OR Kentrothamnus OR Retanilla OR Talguenea OR Trevoa OR Ochtophila OR Ceanothus OR Cercocarpus OR Chamaebatia OR Cowania OR Dryas OR Purshia OR Alder OR Dryas OR Sphagnum OR Hylocomium OR Pleurozium OR Ptilium OR Hylocomium OR Nephroma OR Peltigera OR Sticta OR Stereocaulon OR Collema OR Nostoc OR tree OR shrub OR scrub OR herb OR *lichen OR cyanobacteria OR moss OR liverwort OR free-living OR heterotroph* OR "soil crust" OR "biological soil crust" OR biocrust OR "cryptogamic crust" OR *symbio* OR cyanolichen OR endophyt*)

Supplementary Text S2. Data manipulation using spreadsheet software.

Microsoft Excel

We suggest using the Power Query tool for data manipulation in Microsoft Excel. To merge two different data files into a single spreadsheet, follow these steps:

1. Open the dataset files as separate sheets in the same Excel workbook.
2. Select the data in the first sheet, then go to the “Data” tab and click on the “From Table/Range” icon in the “Get and Transform Data” section to load the selected data into Power Query. Do the same for the second sheet.
3. In Power Query, select the first table, go to the “Home” tab, and click on “Merge Queries.” Select the second table for merging and the common column, “record_ID”, on each table. To keep all rows from the first sheet and add matching data from the second sheet, choose the “Left Outer” join kind. Other types of joins are also available.
4. Click on the small icon in the top right corner of the new column to expand and select which columns you want to add to your data. Then, click “Close & Load” to load the combined data back into Excel.

To merge more than two data files, merge two data files at a time.

Molten files (combining variables into a single column and observations into another column) must be reshaped before merging with other files. To do that, follow these steps:

1. Open the molten data file in Excel.
2. Select the data, then go to the “Data” tab and click on the “From Table/Range” icon in the “Get and Transform Data” section to load the selected data into Power Query.
3. In Power Query, select the “variable” column, then go to the “Transform” tab and click on “Pivot Column”. Select the “value” column.
4. Click on “Close & Load” to load the reshaped data back into Excel.

Google Sheets

To manipulate data using Google Sheets, use the function “query” to merge two data files:

1. Open the dataset files as separate sheets in the same Google Sheets file.
2. Create a blank sheet and copy and paste the “query” function below (adjust as needed):

```
=QUERY({Sheet1!A:B, Sheet2!B:C}, “SELECT Col1, Col2, Col3, Col4”, 1)
```

Where,

“{Sheet1!A:B, Sheet2!B:C}” combines data from Sheet1 columns A to B and Sheet2 columns B to C, with column A being the variable “record_ID” in both sheets,

“SELECT Col1, Col2, Col3, Col4” selects the columns to display in the result. Col1= “record_ID” in Sheet1, Col2= column B in Sheet1, Col3= column B in Sheet2, and Col4= column C in Sheet2,

and “1” indicates that the first row contains variable names.

The default of the “query” function is an inner join kind, that is, the result only includes rows with matching “record_ID” in both sheets. Other types of joins, as well as merging more than two sheets at once, are not as straightforward but possible using a combination of “QUERY”, “ARRAYFORMULA”, “IFERROR,” and “VLOOKUP” formulas.

To reshape molten data files before merging, follow these steps:

1. Open the molten data file in Google Sheets.
2. Select the data, then go to the “Insert” tab, click on “Pivot table” and create a new sheet.
3. In the Pivot table editor, select “record_ID” as rows, “variable” as columns, and “value” as values. By default, Google Sheets will also sum values in a separate column and row. Copy and paste the data, excluding summary statistics, into a new sheet.

Supplementary Table S1. Metadata of the global BNF dataset in natural terrestrial biomes.

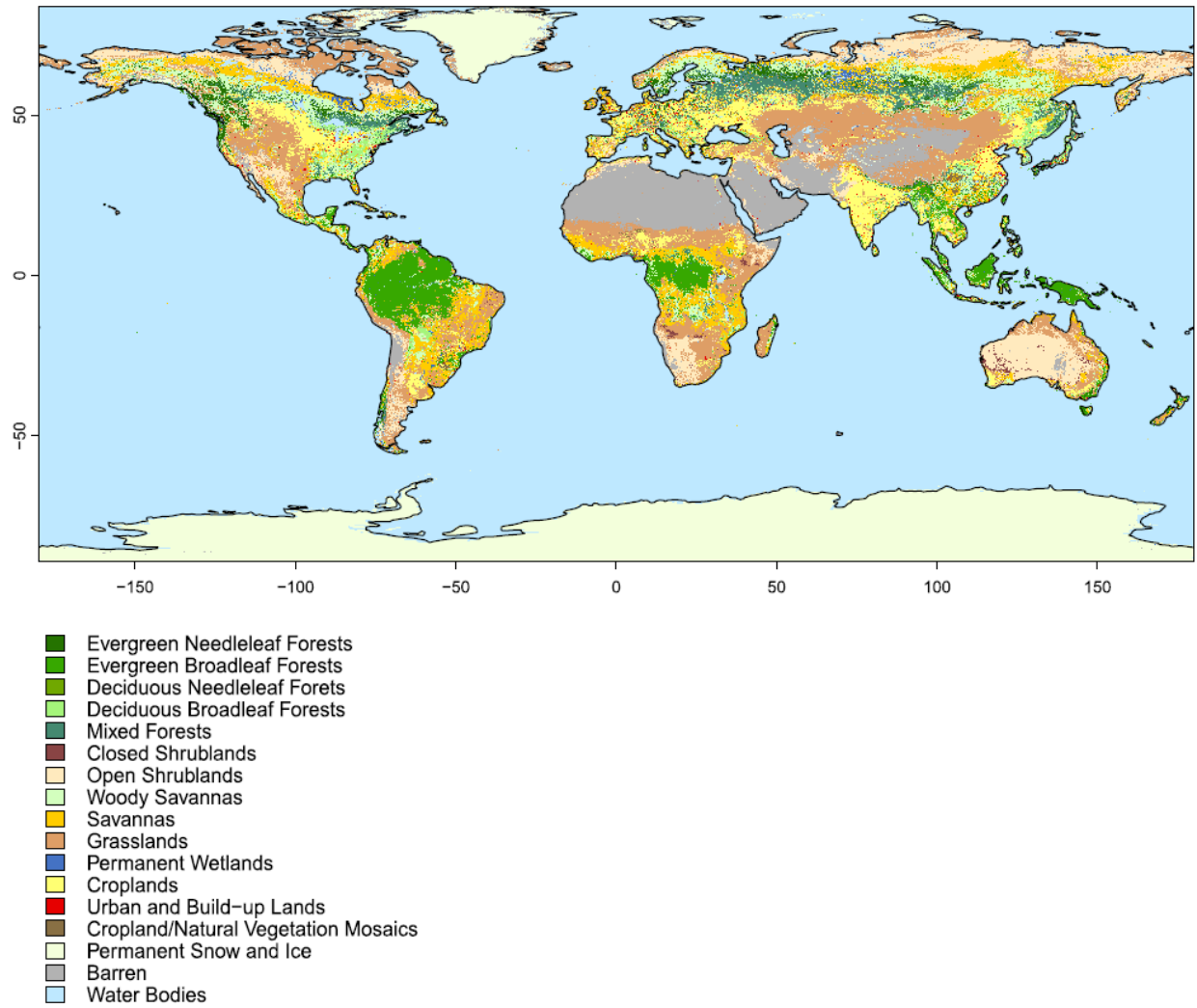
File	Field	Description	Values
SITE	record_ID	Record identification code that links data across files	
SITE	site_name	Study site name. Similar site names followed by different numbers indicate plots/stands that differ in vegetation composition or dominance, age, disturbance history, N-fixer abundance, elevation/climate, and/or topoeadaphic setting within the same study area	
SITE	Country	3-digit ISO country code	"Multiple" if data spans more than one country
SITE	Lat	Latitude coordinate in decimal degrees	
SITE	Lon	Longitude coordinate in decimal degrees	
SITE	IGBP	International Geosphere-Biosphere Program land cover class	"ENF" for Evergreen Needleleaf Forests, "EBF" for Evergreen Broadleaf Forests, "DBF" for Deciduous Broadleaf Forests, "MF" for Mixed Forests, "Shrub" for Closed Shrublands or Open Shrublands, "Sav" for Woody Savannas or Savannas, "Grass" for Grasslands, "Bare" for Barren, "Wet" for Permanent Wetlands, and "Snow" for Permanent Snow and Ice
SITE	Niche	N-fixing niche	"Trees", "Shrubs", "Herbs", "Soil" (mineral and/or organic soil, including humus), "Litter" (all non-woody litter, usually leaf litter), "Woody litter" (dead branches and stems ≤ 7.5 cm diameter), "Dead wood" (logs > 7.5 cm diameter), "Mosses", "Lichens", and "Biocrusts"
SITE	ecological_level	Ecological organization level the data refer to	"Species" or "Community"
SITE	ref_code	Source reference code in the form of "First author's last name_year of publication"	
BNF_AREA	record_ID	Record identification code that links data across files	
BNF_AREA	year	The calendar year that BNF was measured	The mean was entered if BNF was measured over multiple years
BNF_AREA	BNF_central	Central BNF rate reported	Value is exactly as reported, or the mean or sum of reported group values
BNF_AREA	BNF_min	Minimum BNF rate reported	

BNF_AREA	BNF_max	Maximum BNF rate reported	
BNF_AREA	BNF_unit	The unit BNF_central, BNF_min, and BNF_max are expressed	
BNF_AREA	method	The method(s) used to quantify BNF rates	10 = Acetylene Reduction Assay (ARA), 11 = ARA + $^{15}\text{N}_2$ incorporation, 12 = ARA + N accretion, 13 = ARA + N accumulation, 20 = $^{15}\text{N}_2$ incorporation, 21 = $^{15}\text{N}_2$ incorporation + N accretion, 30 = ^{15}N natural abundance, 31 = ^{15}N natural abundance + ^{15}N dilution, 32 = ^{15}N natural abundance + N accretion, 40 = ^{15}N dilution, 50 = N accretion, 60 = N accumulation, or 70 = Mass balance of other N fluxes
BNF_AREA	BNF_final	Final BNF rate in standardized units ($\text{kg N ha}^{-1} \text{y}^{-1}$) based on BNF_central or BNF_min and BNF_max, and local N-fixer or N-fixing substrate abundance	If BNF_central is not available, BNF_final is based on the mean of BNF_min and BNF_max values
BNF_AREA	BNF_final_type	Indication on whether BNF_final was reported in the original publication, converted from similar temporal and/or spatial units, or converted from other types of BNF data. It also indicates whether rates were adjusted for local N-fixer or N-fixing substrate abundance in the original publication or here	1 = reported in the original publication and already adjusted for local N-fixer or N-fixing substrate abundance, 2 = reported in the original publication and adjusted here for local N-fixer or N-fixing substrate abundance, 3 = converted from similar temporal and/or spatial units already adjusted for local N-fixer or N-fixing substrate abundance, 4 = converted from similar temporal and/or spatial units and adjusted here for local N-fixer or N-fixing substrate abundance, 5 = converted from other types of BNF data already adjusted for local N-fixer or N-fixing substrate abundance, or 6 = converted from other types of BNF data and adjusted here for local N-fixer or N-fixing substrate abundance
ABUNDANCE	record_ID	Record identification code that links data across files	
ABUNDANCE	variable	N-fixer or N-fixing substrate abundance variable	"pct_BA" for N-fixer relative basal area (%BA), "pct_stems" for N-fixer relative stem density (%stems), "pct_cover" for N-fixer percent ground cover (%cover), or "AGB" for N-fixing substrate aboveground biomass (Mg ha^{-1})
ABUNDANCE	value	N-fixer or N-fixing substrate abundance value	
BNF_ABUNDANCE	record_ID	Record identification code that links data across files	

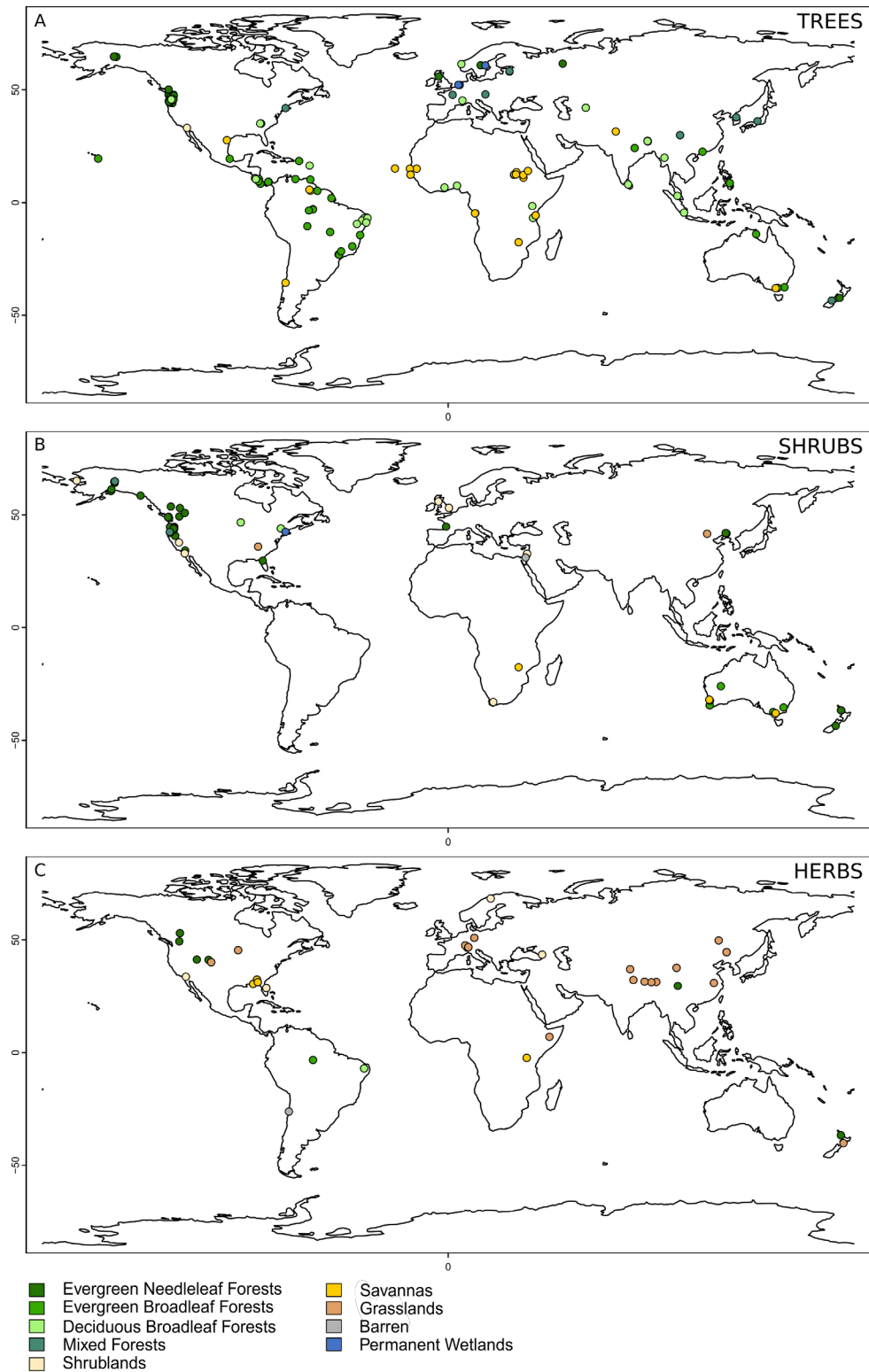
BNF_ABUNDANCE	variable	BNF per unit of N-fixer abundance variable	"BNF_BA" for BNF rate per unit of N-fixer relative basal area ($\text{kg N ha}^{-1} \text{y}^{-1} \% \text{BA}^{-1}$), "BNF_stems" for BNF rate per unit of N-fixer relative stem density ($\text{kg N ha}^{-1} \text{y}^{-1} \% \text{stems}^{-1}$), or "BNF_cover" for BNF rate per unit of N-fixer percent ground cover ($\text{kg N ha}^{-1} \text{y}^{-1} \% \text{cover}^{-1}$)
BNF_ABUNDANCE	value	BNF per unit of N-fixer abundance value	
NDFA	record_ID	Record identification code that links data across files	
NDFA	year	The calendar year that %N _{dfa} was measured	The mean was entered if %N _{dfa} was measured over multiple years
NDFA	method	The method(s) used to quantify %N _{dfa}	30 = ¹⁵ N natural abundance, 31 = ¹⁵ N natural abundance + ¹⁵ N dilution, or 40 = ¹⁵ N dilution
NDFA	Ndfa	Percentage of nitrogen derived from atmosphere (%N _{dfa})	
SPECIES	record_ID	Record identification code that links data across files	
SPECIES	family	Taxonomic family	
SPECIES	genus	Taxonomic genus	
SPECIES	species	Taxonomic species	"sp" indicates one and "spp" indicates more than one unidentified species
AUXILIARY_DATA	record_ID	Record identification code that links data across files	
AUXILIARY_DATA	variable	Auxiliary data variable	"habit" for the habit of sampled N-fixing mosses and lichens, "soil_depth" for the soil depth sampled for BNF (cm), or "bulk_density" for soil bulk density (g cm^{-3})
AUXILIARY_DATA	value	Auxiliary data value	1 = on the ground, or 2 = epiphyte for N-fixer habit
REFERENCES	ref_code	Source reference code in the form of "First author last name_year of publication"	
REFERENCES	ref_complete	Complete source reference	

Supplementary Table S2. The International Geosphere-Biosphere Program (IGBP) land cover classification system¹².

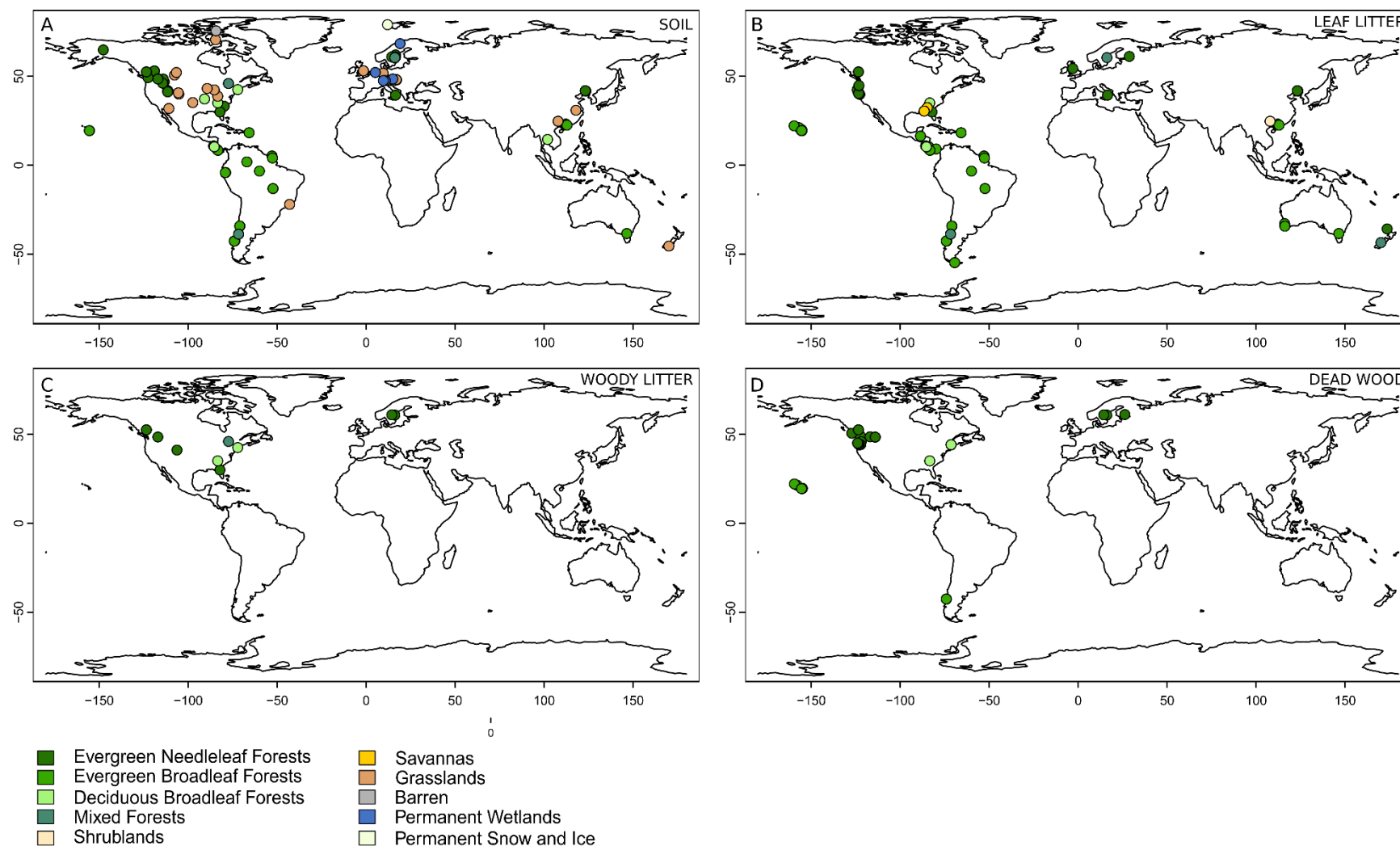
Land cover class	Description
Evergreen Needleleaf Forests	Dominated by evergreen conifer trees (canopy >2m). Tree cover >60%
Evergreen Broadleaf Forests	Dominated by evergreen broadleaf and palmate trees (canopy >2m). Tree cover >60%
Deciduous Needleleaf Forests	Dominated by deciduous needleleaf (larch) trees (canopy >2m). Tree cover >60%
Deciduous Broadleaf Forests	Dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%
Mixed Forests	Dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m). Tree cover >60%
Closed Shrublands	Dominated by woody perennials (1-2m height) >60% cover
Open Shrublands	Dominated by woody perennials (1-2m height) 10-60% cover
Woody Savannas	Tree cover 30-60% (canopy >2m)
Savannas	Tree cover 10-30% (canopy >2m)
Grasslands	Dominated by herbaceous plants (<2m)
Permanent Wetlands	Permanently inundated lands with 30-60% water cover and >10% vegetated cover
Croplands	At least 60% of the area is cultivated cropland
Urban and Built-up lands	At least 30% impervious surface area, including building materials, asphalt, and vehicles
Cropland/Natural vegetation mosaics	Mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation
Permanent Snow and Ice	At least 60% of the area is covered by snow and ice for at least 10 months of the year
Barren	At least 60% of the area is non-vegetated barren (sand, rock, soil) areas with less than 10% vegetation



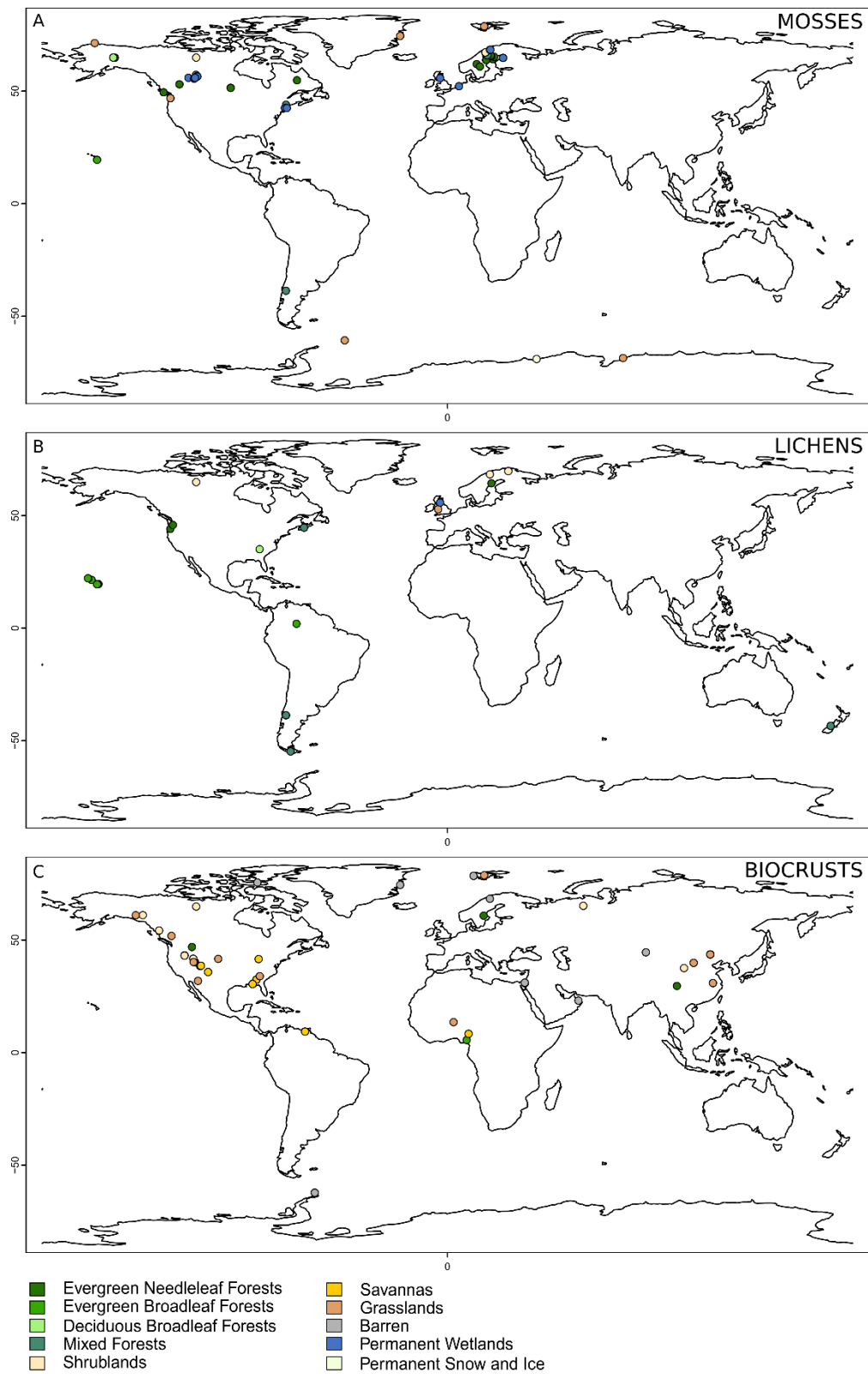
Supplementary Figure S1. The IGBP land cover classification system in 2019¹².



Supplementary Figure S2. Location of field measurements of BNF in natural terrestrial biomes for N-fixing trees (n=273) (A), shrubs (n=104) (B), and herbs (n=261) (C) by IGBP biome. Each point represents one BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$) and/or percentage of nitrogen derived from the atmosphere ($\%N_{\text{dfa}}$) value. Not all points are visible due to overlap.



Supplementary Figure S3. Location of field measurements of BNF in natural terrestrial biomes in soil (n=177) (A), non-woody, mostly leaf litter (n=174) (B), woody litter (n=11) (C), and dead wood (n=57) (D) by IGBP biome. Each point represents one BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$). Not all points are visible due to overlap.



Supplementary Figure S4. Location of field measurements of BNF in natural terrestrial biomes for N-fixing mosses (n=126) (A), lichens (n=43) (B), and biocrusts (n=133) (C) by IGBP biome. Each point represents one BNF rate ($\text{kg N ha}^{-1} \text{y}^{-1}$). Not all points are visible due to overlap.