Received: 14 November 2025 | Accepted: 20 January 2026

DOI: 10.1111/2041-210x.70274

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution

REVIEW

BIEN: A biodiversity informatics ecosystem advancing open and reproducible workflows for plant observation, plot and trait data

Brian J. Enquist[1,2] | **Brad Boyle**[1] | **Brian S. Maitner**[3] | **Cory Merow**[4] | **Ben S. Carlson**[4] | **Xiao Feng**[5] | **Gabriel M. Moulatlet**[1] | **Erica A. Newman**[6] | **Daniel S. Park**[7,8] | **Michiel Pillet**[1,3] | **Patrick R. Roehrdanz**[9] | **Josep M. Serra-Diaz**[10] | **Rethvick S. Y. Babu**[1] | **George G. C. Barbosa**[1] | **Nathan Casler**[11] | **John C. Donoghue II**[12,13] | **Daniel Guaderrama**[1] | **Rohith K. Sajja**[1] | **Thomas L. P. Couvreur**[14,15] | **Matthew B. Jones**[16] | **Peter M. Jørgensen**[17] | **Nathan J. B. Kraft**[18] | **Pablo A. Marquet**[19,20,21] | **Brian J. McGill**[22,23] | **Nirav Merchant**[24] | **Naia Morueta-Holme**[25] | **Danilo M. Neves**[26] | **Ary Oliveira-Filho**[26] | **Robert K. Peet**[5] | **Oliver Phillips**[27] | **Brody Sandel**[28] | **Mark Schildhauer**[16] | **Irena Simova**[29,30] | **Hans ter Steege**[15,31] | **Jens-Christian Svenning**[32] | **Barbara Thiers**[33,34] | **Cyrille Violle**[35] | **Jan J. Wieringa**[31] | **Susan K. Wiser**[36]

Correspondence
Brian J. Enquist
Email: benquist@arizona.edu

Funding information
National Science Foundation, Grant/Award Number: HDR-1934790, EF-0553768, DBI-0735191, DBI-1913673 and ABI-1565118; Global Environment Facility SPARC, Grant/Award Number: GEF-5810; Aarhus University Research Foundation; VILLUM FONDEN, Grant/Award Number: 16549; Danish National Research Foundation, Grant/Award Number: DNRF173; Independent Research Fund Denmark, Grant/Award Number: ATEMPO 10.46540/2064-00091B; VILLUM FOUNDATION, Grant/Award Number: VIL53086; Ministerio de Ciencia, Innovación y Universdidades. Programa, Grant/Award Number: RYC2022-035668-I:MCIU/AEI/10.13039/501100011033; Plan de Generación de Conocimiento, Grant/Award Number: PID2024-162433NB-I00; Czech Science

Abstract

1. The rapid expansion of biodiversity data presents new opportunities to understand and forecast biosphere dynamics. However, disparate and dispersed data, taxonomic and geographic inconsistencies, pervasive quality issues, and a lack of reproducable workflows hinder synthesis, introduce biases and limit accurate assessment of biodiversity trends.
2. The Botanical Information and Ecology Network (BIEN) addresses these challenges through two core contributions: (1) an ecosystem of modular, open-source BIEN tools for collecting, harmonizing, validating, integrating, visualizing and analysing biodiversity data, and (2) a standardized, augmented and integrated global database (BIEN db) of plant occurrences, traits and geographic ranges. BIEN tools—including the Taxonomic Name Resolution Service (TNRS), Geographic Name Resolution Service (GNRS), Geocoordinate Validation Service (GVS) and Native Species Resolver (NSR)—enable the construction of analysis-ready data sets. Using this framework, we integrated over 284 million botanical observation records. BIEN db (current version 4.2) is so far the most extensive compilation of

For affiliations refer to page 23.

Foundation, Grant/Award Number: 20-29554X; Korea University; Lovejoy Center for Bridging Biodiversity, Conservation Science, and Policy at the University of Arizona; Arizona Institute for Resilience at the University of Arizona

Handling Editor: Tommaso Russo

standardized observation records and plant traits (25,932,454 unique trait observation records) for land plants, drawn from herbaria, ecological plots, citizen science and trait databases and integrated into a unified geospatial platform. Here, we present the BIEN workflow, the database and several use cases.

3. The BIEN workflow and database support reproducible biodiversity science at scale. BIEN provides improved and unprecedented trait coverage, new estimates of total plant species richness, geographic ranges for over 250,000 species and global maps of plant biodiversity. These tools enable researchers to integrate disparate biodiversity observation records, generate tailored data sets, resolve taxonomic and geographic ambiguities, and estimate species richness across spatial and temporal scales for conservation and forecasting applications.
4. BIEN provides a scalable, open and reproducible informatics platform that integrates global plant biodiversity data. By linking data infrastructure with accessible tools, BIEN democratizes biodiversity science. It enhances transparency, reproducibility and analytical capacity across the research community, enabling more scientists to address a wider range of questions in ecology, evolution and biodiversity science and engage in global change biology and biodiversity forecasting. BIEN offers a robust Open Science platform to understand and respond to the biodiversity crisis.

KEYWORDS
biodiversity informatics, data harmonization, data integration, macroecology, open science, plant biodiversity, reproducible workflows

1 | INTRODUCTION

1.1 | Addressing biodiversity knowledge shortfalls through better data

> The delimitation of separate floral districts and their grouping into more comprehensive combinations are nearly completed, and the time is not far distant when all species of plants and their geographical distribution will be well known. [...] a foundation [...] will have been laid on which science can construct a larger edifice. The essential aim of geographical botany will then be an inquiry into the causes of differences existing among the various floras
>
> (Schimper, 1903).

Well over a century ago, A.F.W. Schimper articulated a bold vision for biodiversity science—one of comprehensive knowledge of global plant diversity and its distribution—laying the groundwork for understanding the processes driving these patterns (Schimper, 1903). Despite significant advances, this vision remains only partially realized (Anderson, 2018; Morueta-Holme & Svenning, 2018). Thousands of new plant species are being described each year, yet shortfalls in biodiversity knowledge persist. These shortfalls include poor knowledge of taxonomy (Linnean), species distributions (Wallacean) and traits (Raunkiaeran) (Hortal et al., 2015a). Closing the gaps in these inventories remains essential to the critical scientific challenge of predicting Earth's biodiversity dynamics (Feng et al., 2022), consistent with Schimper's vision.

Plants are the foundation of terrestrial ecosystem productivity, contributing the vast majority of Earth's biomass and playing central roles in global carbon cycling, climate regulation and ecosystem services (Bar-On et al., 2018). By structuring physical habitats and resource networks, plants underpin biodiversity across trophic levels, from microbes to large herbivores (Li et al., 2024; Schuldt et al., 2019). Despite their critical ecological and socio-economic roles, comprehensive global plant data on diversity, rarity and spatial distribution remain fragmented and often lack integration with ecological and trait data sets (Feng et al., 2022, 2025). Integrating plant data addresses many of these knowledge gaps. It maximizes the potential to advance biodiversity and conservation science (Feng et al., 2022, 2025) and enables more accurate predictions of ecosystem dynamics under changing environmental conditions (Heberling et al., 2021). These efforts may become a more critical societal goal over time, given the essential role of plants in human livelihoods, supplying food, medicine, fuel and materials (Pironon et al., 2024; Reyes-García et al., 2023).

Underlying all of these challenges is the need for reliable and unbiased data, which are essential for determining the total number of plant species on Earth and for comparative biology and ecology, conservation planning and accurate biodiversity forecasts (Davis, 2023; Faurby et al., 2016; Ferrier, 2002; Funk & Richardson, 2002; Heberling et al., 2019, 2021; Mace, 2004; Rushton et al., 2004; Soltis, 2017). For example, generating accurate species lists for specific locations or regions is crucial for ecology, conservation biology and threatened species management, and can inform policy and resource management (Lien et al., 2023; Panter et al., 2020). Furthermore, improved data are essential for the development and integration of theories in ecology (e.g. Marquet et al., 2014), and data on the distribution and abundance of species are critical for resource management and conservation (Jetz et al., 2019; Johnston et al., 2015; Lawton, 1993; Meyer et al., 2016).

The Botanical Information and Ecology Network (BIEN) addresses these challenges by integrating disparate biodiversity observations into a standardized database (Figure 1). BIEN provides tools to explore plant diversity and function across scales and resolutions essential for mitigating biodiversity loss in the Anthropocene. Here, we describe a significantly updated BIEN workflow and database, and provide a list of diverse use cases along with the associated reproducible code. These updates enhance its capacity to quantify global plant diversity, model species distributions and uncover ecological patterns across spatial and temporal scales. The BIEN database and tools thereby present critical resources for advancing biodiversity science.

1.2 | Barriers to data integration for open, reproducible and high-quality biodiversity science

Reproducibility of workflows and data analyses is fundamental to the scientific method, enabling the validation, verification and refinement of findings (Allison et al., 2016; Peng, 2011). In biodiversity science, reproducibility is particularly challenging due to the complexities of data integration and the prevalence of errors in biodiversity data sets (Bowler et al., 2024; Enquist et al., 2016; Feng et al., 2022; Soberón & Peterson, 2004). Integrating diverse data sources requires resolving numerous sources of error and bias, including taxonomic alignments, correcting geographic errors and addressing sampling biases. Addressing each of these often involves methodological choices that vary among researchers (Hampton et al., 2013; Soberón & Peterson, 2004).

Differences in data sets and workflows often make the results from cleaning and harmonization challenging to compare or replicate

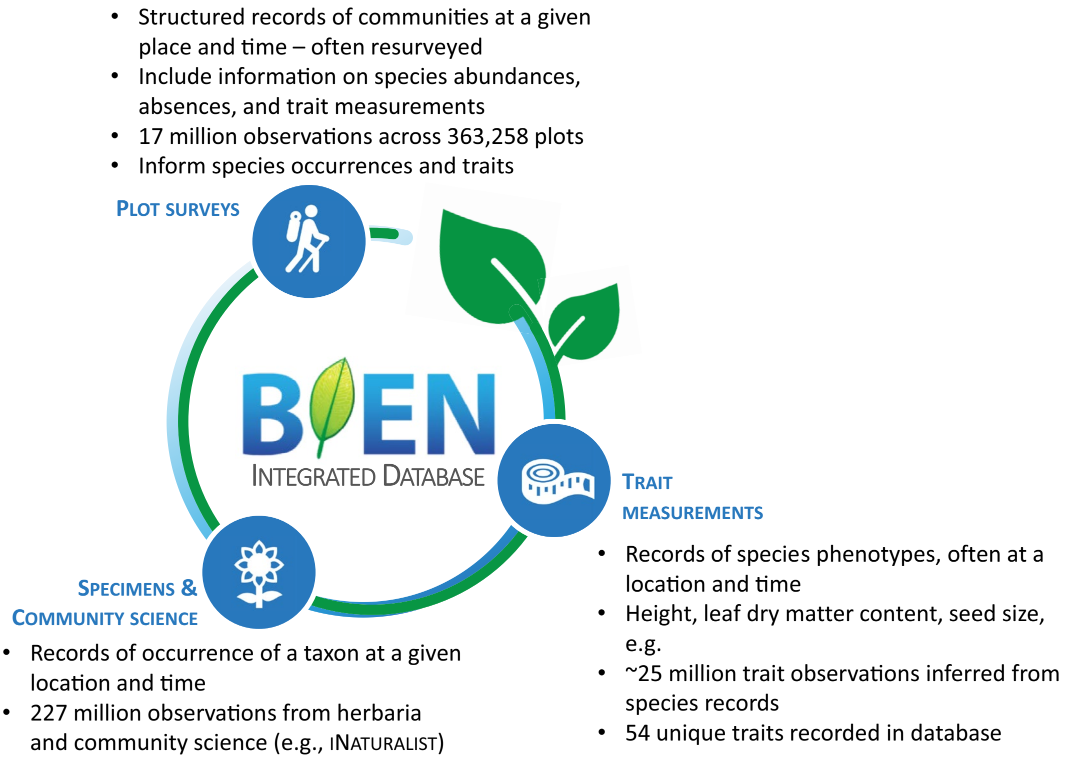


FIGURE 1 The BIEN database (BIEN db) is constructed using a modular workflow that can integrate diverse biodiversity observation data. The BIEN group worked to construct the BIEN db from three primary sources: (i) Specimens—observations from herbaria, museums and citizen science platforms like iNaturalist; (ii) Plot surveys—spatially explicit records documenting community structure and diversity; and (iii) Trait measurements—functional and morphological traits collected in the field. BIEN consolidates hundreds of millions of observations into a unified framework, ensuring data standardization, reducing biases and enabling reproducible workflows. By integrating these data sets, BIEN leverages its unique strengths to address gaps and limitations, fostering a comprehensive approach to biodiversity science through an open and reproducible workflow. This integration enhances cross-scale ecological and evolutionary analyses, supporting applications such as species distribution modelling, functional diversity studies and conservation planning.

(König et al., 2019). For example, efforts to quantify global patterns of species abundances and rarity have been hindered by inconsistent data, taxonomic mismatches and poor data quality (Amano et al., 2016; Boakes et al., 2010; Hortal et al., 2015a). Nevertheless, three primary barriers: (i) disparate and dispersed data, (ii) data incompleteness and biases and (iii) a multitude of errors continue to hinder the integration of biodiversity data for reproducible, high-quality science. In Supporting Information S1.1, we detail the multiple issues underlying each of these barriers.

Focusing effort on integrating data sets and standardizing protocols may offer the best paths forward to addressing these barriers, while additionally improving taxonomic and spatial completeness and coherence, and advancing biodiversity science at a global scale (Feng et al., 2022; Gallagher et al., 2023; Kusumoto et al., 2023; Maitner et al., 2023). Although barriers to data integration have been addressed in some data sets for certain groups (such as vertebrates) (Jetz et al., 2012; Moura & Jetz, 2021), it is important to underscore that the challenges of biodiversity data integration remain complex (Salguero-Gómez et al., 2021). Indeed, a complete data integration workflow for plants lags despite its critical importance for understanding the biosphere's functioning and forecasting future changes (Meyer et al., 2016).

To address these challenges, we created the Botanical Information and Ecology Network (BIEN; biendata.org). The BIEN group has focused on developing two products for the biodiversity community: (i) A comprehensive data workflow defined by a modular open-source informatics services suite to harmonize, validate, augment, share and analyse biodiversity data; and (ii) an integrated, standardized and augmented global plant database (BIEN db) for plant biodiversity research. The updated BIEN informatics ecosystem (Maitner et al., 2017) and the novel applications it enables for biodiversity science, from quantifying global plant diversity to modelling species distributions and uncovering ecological patterns across scales. We illustrate the system's capabilities with case studies that demonstrate its utility for tackling complex ecological questions, fostering global collaboration and promoting reproducible workflows for cleaning, scrubbing and integrating biodiversity data in line with Open Science principles (Hampton et al., 2015; Nielsen, 2020). These tools are freely available to the scientific community, government agencies and conservation organizations.

2 | BIEN SERVICES AND APPLICATIONS

2.1 | The BIEN informatics ecosystem—A computational workflow for creating high-quality, integrated biodiversity data

The BIEN informatics ecosystem comprises reproducible tools for biodiversity data management (Figures 2-3). Each component of the BIEN workflow augments each data observation with transparent measures of quality and accuracy. BIEN is a community-driven initiative that develops standardized protocols to harmonize and integrate biodiversity data across sources, projects and scales.

The BIEN workflow systematically identifies and flags errors in raw data, including misspelled or synonymous taxon names and invalid geographic coordinates. It also resolves inconsistencies in taxonomic and geopolitical names. These steps ensure that users can query and filter data based on data quality and tailor data queries to best match their research needs. By improving data quality and consistency, the BIEN workflow enables the integration of heterogeneous biodiversity data into a single, interoperable data set. When applied to compiled global observation records (see Supporting Information Section S2.3), the BIEN workflow produces a cleaned, standardized and augmented global database (BIEN db) for all known embryophytes ('land plants') with available botanical records. This database integrates herbarium specimens, ecological plots, surveys and trait observations. Users can also apply the BIEN workflow to subsets of botanical data or to their own data sets. In doing so, the workflow addresses key challenges in taxonomy, data quality, scalability and interoperability, making the BIEN informatics ecosystem a flexible foundation for biodiversity research and synthesis.

The BIEN informatics ecosystem is the result of a series of working groups established in 2009 at the National Center for Ecological Analysis and Synthesis (NCEAS). BIEN represents a cross-disciplinary effort involving over 50+ scientists and data specialists to create a global backbone for Open Science in biodiversity research (Enquist et al., 2016, 2019; Feng et al., 2022; Maitner et al., 2017). In Supporting Information S1.1, we provide additional detail on the history of the BIEN working group, additional background regarding BIEN Open Science, and the BIEN Open licence.

2.2 | Database and methods to access data

The BIEN workflow provides a template for how future biodiversity informatics projects can evolve. All validation services within BIEN share a core architecture (Figure 2) that includes components described in the following sections.

2.2.1 | BIEN databases

The BIEN database (BIEN db) was designed to advance Open Science and provide high-quality, reproducible data. The BIEN db integrates flagged and augmented georeferenced observations from specimens, vegetation inventories, regional checklists and individual-level trait data (Figure 1; Supporting Information S2). This schema supports diverse ecological, evolutionary and biodiversity research queries. Built on PostgreSQL, an open-source relational database system (PostgreSQL Global Development Group, 2024), BIEN uses PostGIS to enable advanced geospatial operations, providing a platform for managing and analysing spatial data (Postgis Project Steering Committee & Others, 2018). PostgreSQL and PostGIS combine relational database management with geospatial functionality.

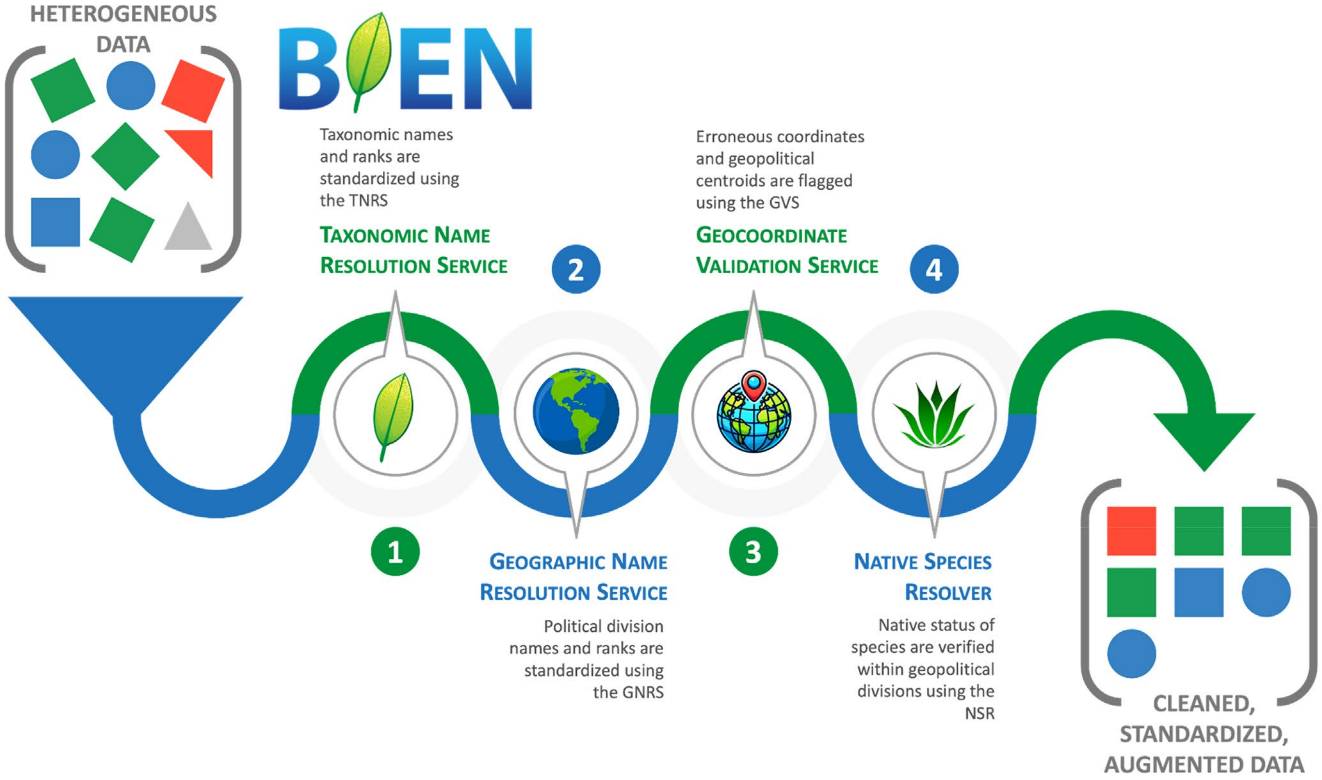


FIGURE 2 The BIEN workflow consists of four key services. This figure illustrates BIEN's modular workflow for integrating and harmonizing diverse observational data sets. All data observations entering into the BIEN workflow are passed through four key services: (1) the Taxonomic Name Resolution Service (TNRS) resolves taxonomic inconsistencies; (2) the Geographic Name Resolution Service (GNRS) standardizes geographic metadata; (3) the Geocoordinate Validation Service (GVS) flags spatial errors and validates coordinates; and (4) the Native Species Resolver (NSR) determines native and cultivated status. These services (detailed in Table 2) augment each observation record with numerous additional information and 'flags' that can be used by the user to filter and subset the data for various higher-level analyses. These services operate in a coordinated manner or independently, and are accessible via APIs and the BIEN R package. The workflow streamlines data cleaning and preparation, enabling researchers to generate standardized, reliable data sets for large-scale biodiversity analyses with minimal additional data harmonization. This flexible, interoperable system supports reproducibility and broad applicability in ecological and evolutionary research.

PostGIS enables BIEN to perform spatial queries, such as calculating species range overlaps with protected areas or analysing biodiversity hotspots. These tools handle the complexities of biodiversity data while integrating with R and BIEN's APIs (see below). The first publicly released version, BIEN db 3.4, (Maitner et al., 2017) contained 81,108,996 observation records. The newest version, reported here, BIEN 4.2 contains 284,466,171 observation records.

The BIEN Geographic Range database (BIEN Open Range db) contains 289,743 plant range maps across 112,953 plant species.

2.2.2 | Application programming Interface (API)

Each step of the BIEN workflow and the resulting BIEN db have associated APIs (Table 1). An API for individual BIEN services helps aid software development and the creation of an 'ecosystem' of linked open-source tools. APIs act as bridges, enabling seamless integration of BIEN's biodiversity resources into analytical workflows, ensuring that data and methods are consistently applied across studies (Manikas, 2016). APIs facilitate direct, automated, standardized interactions with BIEN's core services. This accessibility supports the principles of Open Science (reviewed in Gallagher et al., 2020), ensuring transparency and reproducibility and allowing for broader participation in biodiversity research. These underlying capabilities are essential for individual studies and promote collaborative research by making tools and data sets accessible to the wider scientific community (Supporting Information S2.1, Figure 3).

2.2.3 | R packages

Each BIEN tool in the BIEN workflow has an associated R package (Table 1; Maitner, 2023; Maitner et al., 2017) that links with the associated API to facilitate quick access to each BIEN workflow service and the BIEN database (Table 2). All R package names and links are listed in Table 2. Each R package includes one or more vignettes that contain detailed example R code, and each function within the package includes one or more examples showing how to use it. Each BIEN repository contains documentation of its files, their functions and usage examples; a comprehensive listing of these repositories

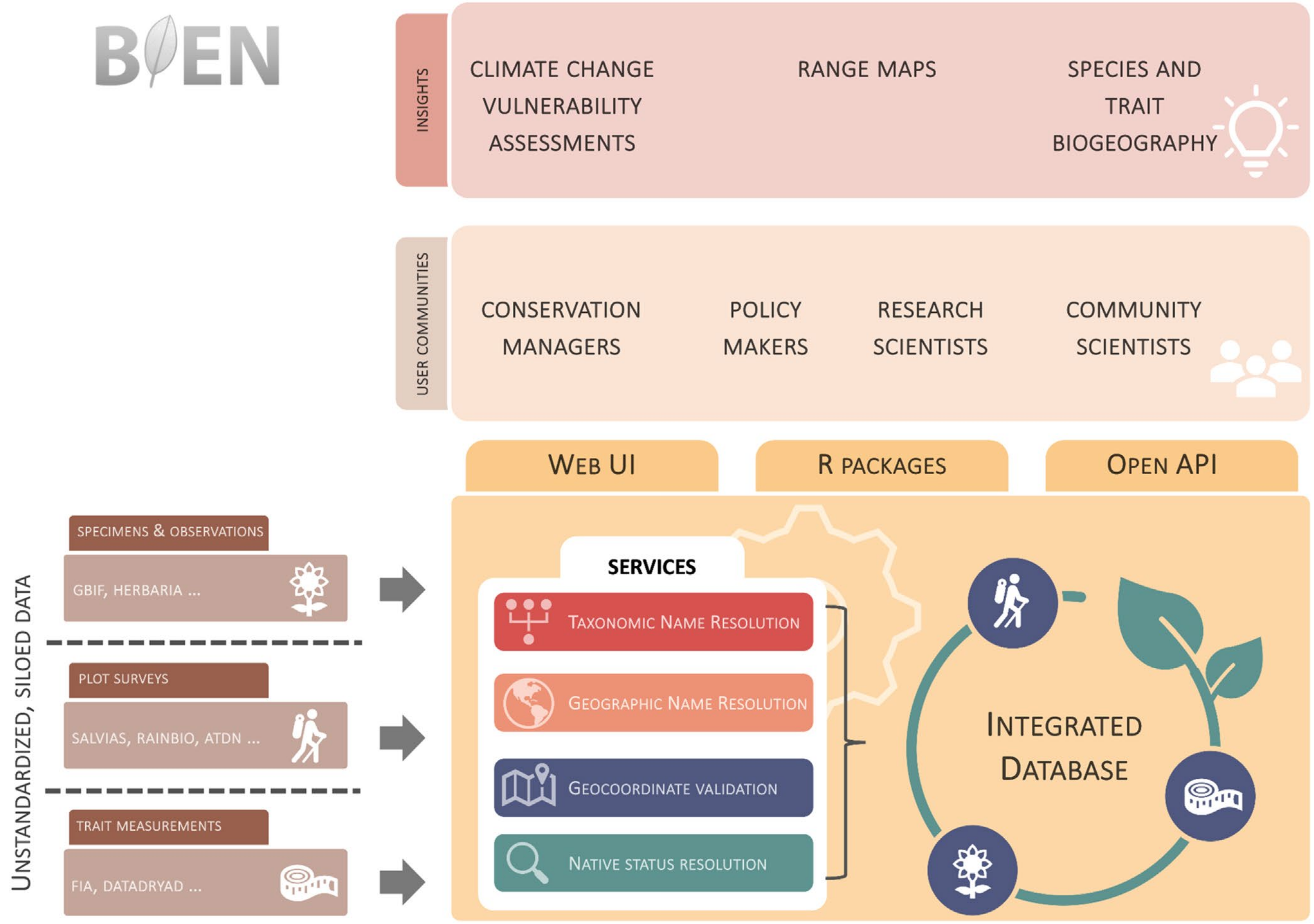


FIGURE 3 An overview of the 'BIEN Ecosystem'. The BIEN ecosystem consists of unstandardized, siloed data that are imported, cleaned, standardized and integrated using openly available data services that are accessed via a web UI, R packages or an API. These cleaned and integrated data are then housed in a single database and used to produce several end products, such as species distribution maps and climate change forecasts. These products can then be used by diverse user communities to yield fundamental insights for biodiversity science.

TABLE 1 The core features of BIEN's infrastructure are designed to enhance the quality, usability and reproducibility of biodiversity data for plant-focused research. These five essential components enable reproducibility and consistency in data analysis.

BIEN informatics infrastructure features	
Feature	**Description**
(1) Standardized database (db) schemas and tools for data integration/validation (BIEN db—global plant observation database; OpenRanges db—global plant range database)	Consistent schemas and validation tools to ensure high-quality geospatial data queries
(2) Reproducible modular workflows for data scrubbing and harmonization with biodiversity observation data (e.g. TNRS, GNRS, GVS, NSR)	Workflows allow reproducibility and consistency in data analysis
(3) Derived products (e.g. range maps, country species lists)	Products like range maps and species lists enable local-to-global ecological, evolutionary, biogeographic, macroecological and conservation work for insights into biodiversity
(4) Improved data quality through feedback (data augmentation and flagging that original data providers can access)	Feedback flags are used to enhance original data for (i) better accuracy, (ii) enhanced usability and (iii) improvement of original data to minimize errors and biases
(5) Standardized data attribution for proper citation (tools for generating automated citations for all data providers and associated authors)	Attribution ensures original collectors and institutions receive proper credit

TABLE 2 An overview of BIEN's open-source infrastructure.

Service or database (db)	BIEN tools	API—application repository	R package	Web service
BIEN Database (BIEN db)	Integrated botanical database of global plant observations. Data are augmented from the BIEN services to identify and flag multiple sources of error and bias.	https://doi.org/10.5281/zenodo.18391665	**BIEN** https://cran.r-project.org/package=BIEN Vignettes—https://cran.r-project.org/web/packages/BIEN/vignettes/BIEN.html	biendata.org
BIEN Open Ranges (Open Ranges db)	OpenRanges db—in-development global plant and animal range database	https://doi.org/10.5281/zenodo.18788209	**OpenRange** https://cran.r-project.org/package=OpenRange	*In development*
Taxonomic Name Resolution Service (TNRS)	A data service to standardize plant scientific names by correcting spelling errors, harmonizing alternative spellings and updating outdated names to currently accepted ones	https://doi.org/10.5281/zenodo.18420398	**TNRS** https://cran.r-project.org/package=TNRS Vignettes—https://cran.r-project.org/web/packages/TNRS/vignettes/TNRS_vignette.html	tnrs.biendata.org
Geographic Name Resolution Service (GNRS)	A data service to resolve and standardize names of political divisions worldwide	https://doi.org/10.5281/zenodo.18420430	**GNRS** https://cran.r-project.org/package=GNRS Vignettes—https://cran.r-project.org/web/packages/GNRS/vignettes/GNRS.html	gnrs.biendata.org
Native Species Resolver (NSR)	A data service to identify whether a species occurrence represents a native or non-native (introduced) individual at a specific location within the political division of observation	https://doi.org/10.5281/zenodo.18420468	**NSR** https://cran.r-project.org/package=NSR Vignettes—https://cran.r-project.org/web/packages/NSR/vignettes/NSR.html	nsr.biendata.org
Geocoordinate Validation Service (GVS)	A data service to identify errors in geographic coordinates associated with biodiversity records, distinguishing between valid geocoordinates and centroids, which are often too imprecise for research purposes	https://doi.org/10.5281/zenodo.18420544	**GVS** https://cran.r-project.org/package=GVS Vignettes—https://cran.r-project.org/web/packages/GVS/vignettes/GVS.html	gvs.biendata.org

Note: BIEN's suite of web applications, services, tools and APIs is designed to enhance the quality and usability of biodiversity data for plant-focused research. These services address critical challenges in biodiversity informatics by offering a general solution for taxonomic standardization, geographic validation and identifying native versus non-native species. The tools include APIs, R packages and web applications that enable researchers to seamlessly integrate, clean and validate data in their workflows. These resources foster reproducibility, scalability and transparency.

and access instructions is provided in the Supporting Information. Designed for researchers with basic knowledge of R, the BIEN R packages provide programmatic access to BIEN's databases and tools, making it easier to pull data into analyses without manually downloading files or formatting data. By streamlining data access and enabling reproducible workflows, the BIEN R packages empower users to integrate high-quality biodiversity data directly into their ecological and statistical models, saving time and reducing errors in large-scale analyses.

Access to the BIEN db can be obtained via the BIEN R package (Maitner et al., 2017). This R package includes several functions for querying the database. We point new users to an associated vignette and tutorial.[1] These resources provide step-by-step instructions, example code, and annotated workflows that demonstrate how to query, clean and analyse BIEN data for a wide range of research applications. The vignettes are available on CRAN and serve as an entry point for both new and advanced users seeking to integrate BIEN into their biodiversity analyses.

2.2.4 | Web-based graphical user Interface (GUI)

Each of the BIEN data services and the BIEN database have a web application (Table 2). Each website links to the associated BIEN API. These user-friendly interfaces allow researchers to explore and access each unique BIEN biodiversity workflow and data component without programming skills. Designed to be intuitive, the GUI enables users to filter, view and download data directly, making BIEN's extensive resources accessible to those unfamiliar with coding. This interface supports open science by widening access to high-quality biodiversity data, empowering a broader range of researchers to engage in data-driven ecological and conservation studies (Supporting Information S2.1).

2.3 | The BIEN web services: Tools for biodiversity data cleaning, scrubbing and harmonization

2.3.1 | BIEN web services description

The BIEN web services are separate from the BIEN db. BIEN's web services harmonize data across multiple dimensions, including taxonomy, geography, species nativity and coordinate precision, producing integrated data sets ready for analysis. These four services—the Taxonomic Name Resolution Service (TNRS), the Geographic Name Resolution Service (GNRS), the Native Species Resolver (NSR) and the Geocoordinate Validation Service (GVS)—provide tools that enhance biodiversity data quality and improve usability for science and decision-making. The services address key challenges in biodiversity data, including taxonomic inconsistencies, geographic errors and identifying non-native occurrences.

All BIEN services are available as R packages through CRAN (e.g. BIEN, TNRS, GNRS, GVS and NSR) and include comprehensive documentation and usage tutorials. As shown below, these services indicate that most observation records that pass through the BIEN pipeline are flagged as potentially erroneous.

Importantly, BIEN's services operate independently of the BIEN db, allowing users to bring their data for cleaning and integration without contributing to or interacting with the BIEN db. This flexibility allows researchers to tailor the tools to their specific data sets and objectives while maintaining control over their data and privacy.

2.3.2 | The taxonomic name resolution service (TNRS)

The TNRS (tnrs.biendata.org) is a critical tool for standardizing plant scientific names, addressing the widespread challenges of taxonomic inconsistencies, synonyms and spelling errors in biodiversity data sets (Boyle et al., 2013; Soberón & Peterson, 2004). Potential issues arise when species names do not align across data sets due to synonymy, unresolved taxa or revisions to historical nomenclature. The TNRS helps resolve taxonomic inconsistencies, corrects spelling errors and standardizes synonyms to accepted names. TNRS version 5.3.1 consults three separate sources for information, including the World Flora Online (WFO; https://www.worldfloraonline.org/), the World Checklist of Vascular Plants (WCVP; https://powo.science.kew.org/) and the Cactaceae checklist (CACT, https://caryophyllales.org/) for sources of nomenclatural and taxonomic information. The TNRS combines capabilities for fuzzy matching, synonym correction, partial matching, returning alternative matches and homonym resolution in both an API and a user-friendly web interface (Boyle et al., 2013). The TNRS also harmonizes species names across data sets, reducing ambiguity and enabling accurate analyses in biodiversity science (Boyle et al., 2013).

In the BIEN db 4.2.8 pipeline, 33% of all ingested records contained non-standard names due to synonymy or misspelling; TNRS resolved about 85% of these errors. Users of TNRS should note that it does not directly address inconsistencies in taxon concepts across authorities, which is a problem that is not addressed in global databases. For example, many large data sets that reference *Quercus prinus* refer to the taxon now known as *Q. montana*, whereas *Quercus prinus* in TNRS points to *Q. michauxii* as the current official synonym, a different taxon with a different range. Divergent global checklists can lead to inconsistencies in species delimitation and downstream analyses, particularly when sources differ in how they treat synonymy or accepted taxa (Ronquillo et al., 2023; Stropp et al., 2025). This is an area needing future attention and standards (Franz et al., 2006; Franz & Peet, 2009).

The TNRS employs a rules-based reconciliation and scoring system to evaluate conflicting records among sources. When conflicts occur, TNRS identifies all potential matches, assigns confidence scores and provides links to the original taxonomic authorities. This user-beware model allows researchers to examine and select the most appropriate name for their analyses while maintaining full

---

[1] https://cran.r-project.org/web/packages/BIEN/vignettes/BIEN_tutorial.html.

transparency of the decision process. Detailed logic for reconciliation, including the order of authority and update cycles, is now documented in the Supporting Information S2.2.1 and Table S1.

2.3.3 | The geographic name resolution service (GNRS)

The GNRS (gnrs.biendata.org) resolves and standardizes names of political divisions worldwide based on data from two standards for geographic names—the GeoNames and Global Administrative Areas (GADM) databases. The GNRS resolves names at three levels—country, state/province and county/parish—through exact and fuzzy matching to accommodate variations in language, spelling, abbreviations and standardized country codes (e.g. ISO and FIPS codes). The service returns the original name, standardized political division names, unique identifiers and an overall match score (0–1) indicating the match accuracy.

Geovalidation of coordinates assesses the validity of the geocoordinates associated with a biological observation. Validity is assessed by determining if a point described by a pair of coordinates falls within that observation's lowest declared political division. Points falling outside the declared political division are flagged as invalid (by setting the column `isGeovalid=0`, where `isGeovalid` takes on the values 0 and 1), and the distance to the nearest point along the boundary of the political division is returned as a measure of error (distErr_km, distance error in km). Observations with `isGeovalid=0` are excluded from most analyses, but users can choose to accept records where distErr_km is, for example, less than 2 km. A more detailed description of the GNRS operations is provided in the Supporting Information S2.2 and Table S2.

In an earlier version of BIEN (BIEN 3.0), we found 439 different countries in the ingested data, despite there being only 193 countries at that time; only 43% of the records had a canonical country name, and 49% of the records had various synonyms for the country (such as abbreviations, alternative spellings or alternative character encodings). In combination, fewer than half the records had valid latitude and longitude values. In the current version of BIEN (BIEN 4.2.8), 22% of records reported invalid latitude/longitude values. Thus, before being passed through the GNRS, only 44,705,332 records (15.72%) matched the geographic name standards precisely. However, after records were passed through the GNRS, the number of records matching the geographic name standards increased to 269,434,901 records (94.72%).

2.3.4 | The native species resolver (NSR)

The NSR (nsr.biendata.org) identifies whether a species occurrence represents a native or non-native (introduced) plant at a specific location within the political division of observation according to expert checklists (Supporting Information S2.2 and Table S3). This service is critical for ecological analyses, as non-native species can both alter and inform biodiversity patterns and ecosystem health, and incorrect assumptions of nativeness can lead to misinterpretation. The NSR makes this determination by comparing it against country-, state- or county-level checklists. The NSR accepts as input one or more observations of a taxon in a political division (country, plus optionally state/province or county/parish). For each observation record, the NSR returns an opinion as to whether that taxon is native or introduced in the lowest political division submitted.

Native and non-native status in BIEN is derived from expert-curated regional checklists and floras, which may distinguish recent introductions from long-established, naturalized taxa. BIEN does not adjudicate invasion status or ecological impact; instead, it relies on community-defined classifications and associated uncertainty to support transparent downstream analyses. Most of the checklists in the BIEN NSR are high-quality published species lists prepared by professional taxonomists as part of Floras or other floristic projects. For details on BIEN NSR sources, see the repository https://zenodo.org/records/18420469. The BIEN NSR can be accessed via all the interfaces listed below. The BIEN NSR provides access to an NSR database containing multiple checklists, offering nearly complete global coverage (see https://nsr.biendata.org/map/ for coverage details). Users can explore checklist sources and native status codes in greater detail via the NSR interface (Maitner & Boyle, 2024).

2.3.5 | The geocoordinate validation service (GVS)

The GVS (gvs.biendata.org) identifies errors in geographic coordinates associated with biodiversity records, distinguishing between valid geocoordinates, erroneous geocoordinates and geopolitical centroids, which are often too imprecise for research purposes (Barker & MacIsaac, 2022; Feng et al., 2024; Park & Davis, 2017; Park et al., 2023). The GVS detects points that may be centroids of political divisions or geographic units (such as islands) using geometric centre, point-on-surface and bounding-box centroids. This distinction is essential, as centroid-based locations can introduce significant errors in species distribution modelling and spatial analyses. In addition to error checks, the GVS returns the corresponding political divisions (with Global Administrative Areas [GADM] identifiers) and calculates inherent uncertainty based on coordinate precision, helping users assess the accuracy of the location data. A more detailed description of the GVS operations is provided in the Supporting Information S2.2 and Table S4.

Information returned by the GVS includes (1) estimates of precision based on the number of decimals placed in the original coordinates, (2) flagging and reporting of errors, such as coordinates out of range, (3) flagging of points in the ocean, (4) names and GADM identifiers of the `admin_0`, `admin_1` and `admin_2` political divisions (e.g. country, state, county) in which a point is located, (5) the absolute and relative distance to the centroid of each political division (see complete list of output fields below). The GVS also assesses the likelihood that the point is a centroid and, if applicable, the type of centroid and political division (country, state or county) of the likeliest centroid.

The GVS web application visually plots each submitted observation point and labels whether each geocoordinate is likely to contain an error. Each observation point is then colour-coded on the map to indicate the type of error associated with that point.

A combined 44% of all occurrence records in BIEN lacked a proper taxonomic name or accurate georeferencing (lat/long coordinates). Also, 36% were identified as *non-native* or had an unknown native status, and 1.3% were identified as cultivated. It is clear that when working with large databases, careful attention must be paid to error detection and scrubbing before performing scientific analyses that assume the data inputs are accurate.

2.4 | Types of biodiversity within the BIEN database

Observations in BIEN include four primary data types:

(i) *Occurrence Records*—consisting of natural history collections (i.e. specimens from herbaria and other botanical efforts to assemble natural history observations from literature and field observations),
(ii) *Ecological Plot and Survey Data*—surveys from plots, transects, and other various survey methods,
(iii) *Citizen science data*, and
(iv) *Trait Data*—trait measurements with associated observations

The observations in the BIEN database are the product of contributions from over 1076 data contributors (see Supporting Information S2 and 2.3.1 and Appendix 2). A large fraction of the observation records from herbaria and citizen science are compiled from the Global Biodiversity Information Facility, or GBIF (Edwards et al., 2000).

2.4.1 | Occurrence records

The BIEN db 4.2 includes over 284 million primary botanical occurrence records, each consisting of a plant observation with an associated taxonomic identifier, geographic coordinates and a date. Herbarium records and citizen science observations (e.g. iNaturalist) capture historical and contemporary plant occurrences, contributing invaluable spatial and temporal insights. Major sources of compiled occurrence record data in BIEN 4.2 stem from several data aggregators, including GBIF https://www.gbif.org/; RAINBIO; http://rainbio.cesab.org/ (Dauby et al., 2016), The Royal Botanical Garden of Sydney, Australia; https://www.rbgsyd.nsw.gov.au/ and NeoTropTree; http://www.neotroptree.info. BIEN includes data from 804 herbaria listed in Index Herbariorum; http://sweetgum.nybg.org/science/ih/. Table S1 in Sup. Doc provides a full description of all data contributors to BIEN db 4.2.

2.4.2 | Ecological plot and survey data

Plot and vegetation surveys provide systematically collected, location-specific data for plant communities, offering a fundamentally different and complementary perspective to species presence-only records derived from herbarium specimens and citizen science observations. Unlike presence-only data, plot and survey data explicitly record species presence, abundance and absence ('zeros'), documenting taxa that were surveyed for but not observed within a defined area. In contrast to pseudo-absences generated for species distribution modelling, plot-based absences reflect explicit sampling effort and therefore provide a stronger basis for ecological inference about species exclusion, environmental filtering and climatic constraints on distributions (e.g. see Elith et al., 2011; Soley-Guardia et al., 2024). Abundance data and true absences are central to community ecology, as they enable analyses of local exclusion, community assembly processes, competitive interactions and the mechanisms by which climate and environment structure ecological communities (e.g. Diamond, 1975; Weiher & Keddy, 1999, 2001).

Plot data in BIEN are derived from established vegetation monitoring programs and forest inventory networks, substantially expanding the geographic, ecological and methodological breadth of the BIEN ecosystem. Plot and survey data systematically sample predefined areas of the landscape, whereas herbarium and citizen science records sample species occurrences opportunistically (Ball-Damerow et al., 2019; Guralnick & Van Cleve, 2005). By sampling predefined areas regardless of which species are present, plot and survey data reduce these landscape-sampling biases and provide a more even representation of environmental space. This difference in sampling design reduces spatial and environmental biases in how landscapes are represented (e.g. see Ball-Damerow et al., 2019; Meyer et al., 2016), making plot and survey data especially valuable for community-level inference, reducing geographic and taxonomic sampling biases from museum, herbarium and citizen science data and model evaluation. The integration of plot-based data with occurrence, trait and range information is a distinctive feature of BIEN, enabling users to seamlessly move between community- and species-level analyses within a single harmonized framework.

Plot sources are available via the *BIEN* R package and are listed via the function `BIEN_plot_list_datasource()`(Maitner, 2023). Plot data includes several plot networks, including CTFS/ForestGEO (Anderson-Teixeira et al., 2015), Carolina vegetation survey, CVS (Peet, Lee, Boyle, et al., 2012), Forest Inventory Analysis, FIA (Gray et al., 2012), SALVIAS (Enquist & Boyle, 2012), TEAM (Rovero & Ahumada, 2017), VegBank (Peet, Lee, Jennings, et al., 2012) and several other sources. Because sampling protocols vary across plots for some data sources (e.g. FIA, VegBank), the function `BIEN_plot_metadata()` provides detailed plot information.

2.4.3 | Trait data

Trait observations add functional and physiological dimensions to BIEN's records. Trait data and observations are at the individual level. BIEN contains over 25 million trait records spanning 54 distinct plant traits, derived primarily from publicly available data sets from more than 800 published trait studies (see Supporting Information; Appendix 2). These traits cover critical plant characteristics such as leaf area, wood density and seed mass, enabling functional analyses across species and ecosystems (Table S5). The most numerous trait data in BIEN are the following: Diameter at breast height (14,929,488); Whole plant height (10,167,913); Whole plant growth form (330,047; see Engemann et al., 2016); Seed mass (93,543); Stem wood density (39,162); Leaf area per leaf dry mass, SLA (51,894).

The different observation data types in BIEN can be leveraged to share relevant information between data types. For trait data, BIEN contains the most information on plant size (e.g. diameter at breast height, height) and growth form. Growth form can be inferred from size measurements in plot and survey data. BIEN's standardized trait data allow for within-species studies of intraspecific variation as well as cross-species comparisons—essential for ecological and evolutionary studies. A list of available trait data can be accessed using the `BIEN_trait_list()` function in the *BIEN* R package (B. Maitner, 2023).

Together, these observation records offer a high-resolution, integrative view of plant biodiversity. These data have provided a key resource for studying how geography, climate, megafaunal extinctions, and human land use shape plant diversity, species distributions and extidnction risks (Doughty et al. 2016; Echeverría-Londoño et al., 2018; Flower et al., 2021; Hannah et al., 2020; Moulatlet et al. 2025; Morueta-Holme et al. 2013; Neves et al. 2021). For more details on each of the above data sources and methods, please see the Supporting Information S1 and S2.

2.5 | Partner applications and linkages

The methodologies and workflows developed by BIEN have catalysed the creation of complementary partner tools that extend its principles of transparency, reproducibility and modularity to other biodiversity data sets. Many of these tools leverage BIEN's application programming interfaces (APIs), ensuring seamless access to its services for taxonomic resolution, geographic validation and data filtering. By providing APIs as a foundation, BIEN facilitates interoperability with external tools. In the Supporting Information S2.4, we detail three add-on partner applications (i.e. occTest Serra-Diaz et al., 2024; Wallace Kass et al., 2018, 2023; occCite Owens et al., 2021) that enable researchers to apply BIEN's core approaches—data integration, quality control and transparency—across various contexts (Owens et al., 2021; Serra-Diaz et al., 2024).

2.6 | Derived biodiversity products—Species geographic ranges and standardized species lists

2.6.1 | Geographic ranges

Using the integrated observation data, the BIEN group has generated species range maps for 98,829 plant species within North and South America (Goldsmith et al., 2016; described in Maitner et al., 2017; Neves et al., 2021). In addition, we have recently generated an unprecedented additional set of 289,743 plant species range maps across the globe, including 112,953 species in the Americas. Updated geographic ranges from BIEN 4.2 are detailed in Merow et al. (ms *in prep.*), and are also highlighted in Moulatlet et al. (2025).

The BIEN OpenRange database (Maitner et al., 2025) contains generated range maps for all terrestrial plant species in the Americas with >15 unique records on a 5 km grid suitable for estimating species distribution models. Previous versions of this workflow and BIEN species range maps are presented in Neves et al. (2021), Feng et al. (2021) and Jung et al. (2021). A summary of the details of Merow et al. (ms *in prep*) is provided in the Supporting Information S2.3.14.

These maps were made available as a proof of concept for the Open Range initiative (part of the Open Nature initiative, https://www.open-nature.org/), which aims to create range maps that anyone can use, subject to attribution. While these new Open Range maps have yet to be peer-reviewed and should be used cautiously, they include metadata that can be used to assess their quality and suitability for different purposes. Geographic range maps are available via an API (https://github.com/EnquistLab/Biendata-Backend-Express) and an R package (Table 2; https://doi.org/10.5281/zenodo.18392177) and can thus be easily integrated with other workflows. Future efforts will expand the availability of the range maps.

2.6.2 | Species lists

BIEN provides a useful repeatable method generating lists of species that occur within a given political division (e.g. country, state/province, county/parish) or within any user-defined polygon (e.g. protected area). Within the BIEN R package, lists can be produced using a function beginning with `"BIEN_list_",` (e.g. BIEN_list_country(); B. Maitner, 2023). These lists are generated from the scrubbed and harmonized taxonomic and geopolitical names in BIEN, along with the geovalidated coordinates. Since these lists are based on occurrence data, they complement other sources of species lists derived from checklists or taxonomic information, such as the Global Inventory of Floras and Traits (GIFT; Weigelt et al., 2020) and the World Checklist of Vascular Plants (WCVP; Govaerts et al., 2021).

2.7 | Where to look to understand different data fields

To clarify the meaning of the different components of the BIEN db, a new function has been added to the `BIEN R` package (Maitner et al., 2017) `BIEN_metadata_data_dictionaries()`. This function provides the user with data dictionaries containing information about the tables, columns and values in the BIEN db.

2.8 | Citation practices

As Robert A. Heinlein aptly stated, 'There ain't no such thing as a free lunch' (TANSTAAFL; Heinlein, 1966). Using BIEN data entails, the responsibility to attribute the original data providers properly. To streamline this process, the BIEN R package (Maitner, 2023) includes the function `BIEN_metadata_citation()`, which automatically generates BibTeX-formatted citations and acknowledgment statements for publications that use BIEN data. Similarly, each BIEN module offers functions for generating citations to the original data providers, specific to the tool used. For example, the `GNRS_citations()` function in the GNRS R package (Boyle & Maitner, 2021) provides citations for geographic name resolution services. See Supporting Information S3.4 for recommendations for when, in some instances, the number of citations required may exceed journal-imposed limits.

3 | BIEN APPLICATIONS

3.1 | The default query

Queries to the BIEN db from the BIEN R package use a default query that excludes cultivated and introduced observations and includes only specific observation types to ensure data reliability and relevance. The default query in SQL (Supporting Information S3) is designed to maximize data quality by filtering out records that are likely to introduce noise or errors into analyses. However, as we discuss below, the user can specify additional queries to obtain unfiltered data.

3.2 | BIEN flags and data augmentation

BIEN supports diverse research needs by augmenting each observation with flags that enable researchers to filter and subset data by quality, geographic relevance and taxonomic accuracy (Supporting Information S3.2). Each data observation record passing through the BIEN's validation services (TNRS, GNRS, GVS and NSR) is augmented with numerous flags (Tables S1–S4). These flags allow researchers to query the BIEN db or tailor their own data sets to specific analytical objectives. Each data record that passes through all BIEN data services is augmented with more than 50 distinct flags (Tables S1–S4). The following is a summary of the most important or 'key' BIEN flags:

1. `is_cultivated`: Identifies whether a record pertains to cultivated specimens. 1: Cultivated specimen. 0: Non-cultivated specimen. null: Cultivation status undetermined.
2. `is_introduced`: Flags observations of non-native species. These flags are used to identify and exclude exotic species from ecological analyses.
3. `observation_type`: Specifies the record's source (e.g. plot, specimen, literature, checklist). This flag enables filtering by data type to suit the study's focus.
4. `is_geovalid`: Indicates whether geographic coordinates are accurate. 1: Verified and accurate. 0 or null: Geographic accuracy is erroneous or unverified.
5. `higher_plant_group`: Distinguishes records by taxonomic group, excluding non-target groups like algae, fungi and bacteria.
6. `is_centroid`: Flags records that are georeferenced to administrative centroids rather than specific locations. 1: Georeferenced to a centroid. 0: Accurately georeferenced.
7. `scrubbed_species_binomial`: Provides a standardized taxonomic name resolved via TNRS, reducing ambiguity and improving consistency across data sets.

3.3 | Filtering and subsetting

BIEN enables users to refine data sets using combinations of flags, metadata columns, geographic information and political divisions (see Tables S1–S4). Combining BIEN's built-in flags with TNRS's detailed outputs allows researchers to filter data sets to achieve the highest possible quality and relevance for their specific scientific goals (see Figure 4). In Tables S1–S4, we describe and list each of the BIEN flags. Key data flags like—'`is_geovalid`,' '`scrubbed_species_name`', '`new.world`' and '`is_cultivated`'—allow researchers to adapt BIEN data to their unique analytical needs. Selecting or excluding data based on the data in these columns will enhance data quality and relevance for analyses of plant biodiversity.

The two primary approaches—flag filtering and geographic/political filtering—offer flexibility for tailoring data sets to specific research needs.

1. Flag filtering: This feature allows you to select or exclude records based on specific flags (e.g. using `is_cultivated` and `is_introduced` to exclude cultivated or introduced species).
2. Geographic and political filtering: Subset data by geographic coordinates or political divisions to focus on regions of interest.

Depending on the question at hand, six additional flags and filters can be used to further assess biodiversity data. In the Supporting Information (see S3.2), we detail these additional flags and filters:

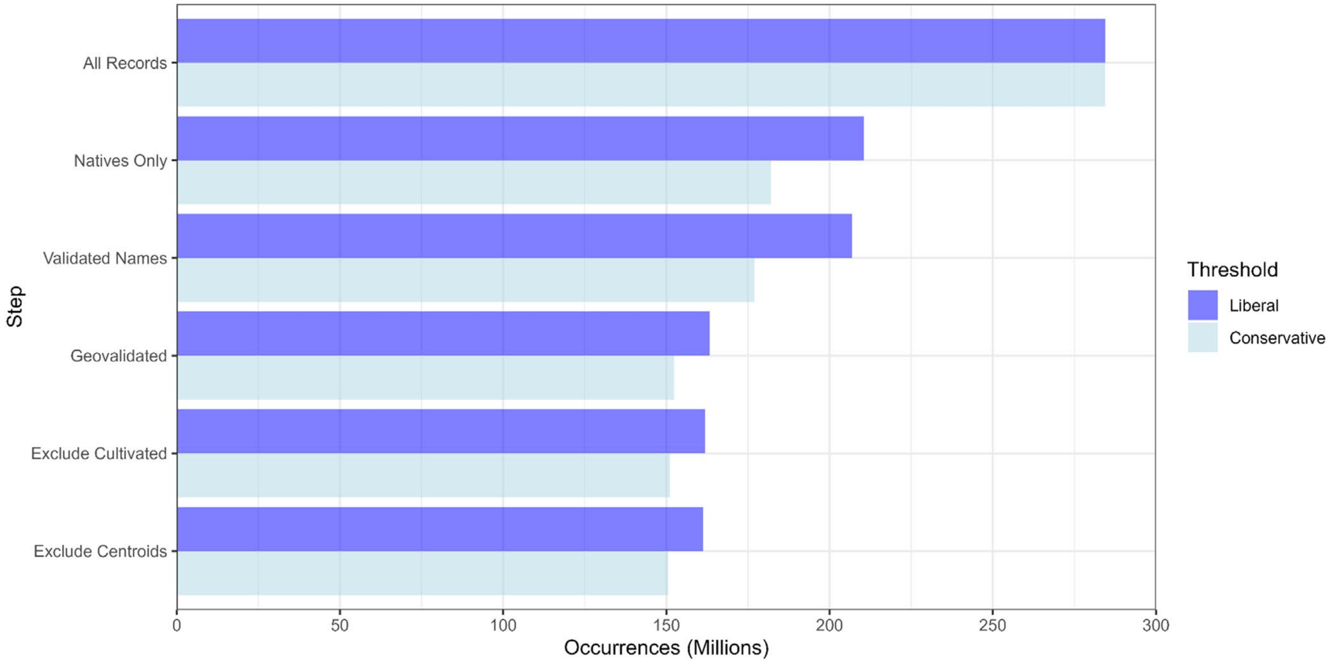


FIGURE 4 Winnowing of observation records through the BIEN workflow for better species distribution modelling. This figure illustrates the sequential filtering of observation records through the BIEN workflow to identify high-quality data for species distribution modelling. Starting with all records, each sequential step applies specific cleaning and validation processes, progressively reducing the data set to remove biased or erroneous observations. Key steps include: (i) All Records to Geovalidated: Incorporates the Geographic Name Resolution Service (GNRS) and Geocoordinate Validation Service (GVS) to ensure accurate geographic metadata and validate coordinates; (ii) Excluding Cultivated and Centroids: Utilizes the Native Species Resolver (NSR) and GVS to exclude records flagged as cultivated or centroid-based; (iii) liberal vs. conservative thresholds: The 'liberal' threshold retains records flagged as 'No opinion' by the Taxonomic Name Resolution Service (TNRS) and records with 'is introduced' = NULL in the NSR. The 'conservative' threshold excludes these records for stricter quality control. At each stage, observation records are filtered to remove potential errors and biases, resulting in a refined data set tailored to the specific use case. This winnowing process highlights the BIEN workflow's effectiveness in producing high-confidence data sets for biodiversity research. It simultaneously underscores that a large fraction of biodiversity data (~50%) does not meet rigorous criteria for species distribution modelling due to various sources of error and bias.

3. Filter by Geographic Validity, 4. Filter by Cultivation Status, 5. Subset by Observation Type, 6. Filter based on Native Status, 7. Filter or subset for Regional Studies, 8. Taxonomic Standardization and give examples of their use.

Users should be aware that some names in large biodiversity data sets may still need to be resolved and require additional research. For further guidance on interpreting TNRS output and best practices for taxonomic name resolution, (see https://tnrs.biendata.org/instructions/). Furthermore, users may use the functions `Name_matched_url` and `Accepted_name_url` to explore further matched or accepted names via linked taxonomic resources for verification or additional metadata.

For users creating checklists or other lists that are highly sensitive to the presence of synonyms, it is recommended to compare all accepted names in the data set with their original names. Any taxonomic name service will be several years behind the current literature and recent changes, including new or outdated synonyms or newly described species, will not be processed correctly by TNRS. For an extensive biodiversity study, this level of inaccuracy may be considered acceptable, but in, for example, a checklist, such names should not be missed or misinterpreted.

3.4 | Flagging of errors, standardization and identification of highest quality data

Most plant biodiversity data contain errors and potential biases. Between 65.4% and 72.5% of the taxonomic names were erroneous or unclear (Tables 3 and 4). Furthermore, of the total number of observation records processed (where an observation name had a taxonomic name that could be parsed by the TNRS), 159,189,390 (55.96%) had an issue with taxonomy or coordinate location. This indicates that most botanical observation records contain at least one potential source of error.

3.5 | BIEN use cases: Examples

The BIEN workflow and standardized database provide a flexible, reproducible framework for addressing core challenges in biodiversity science. Since the release of the first RBIEN package, researchers have applied BIEN tools to a wide range of questions in taxonomy, ecology, biogeography and conservation. Table S6 lists 19 published studies, each accompanied by open-source code and documented

TABLE 3 Key metrics, data types and total number of scrubbed and standardized species within BIEN 4.2, showcasing its role as a standardized, augmented database for global plant ecological and evolutionary research.

Total number of observations	BIEN db 284,466,171	
Observation records from herbarium specimens	BIEN db 73,927,842	
Observation records from citizen science	BIEN db 153,551,098	
Observation records from plots	BIEN db 17,247,823	
Observation records from trait observations	BIEN db 380,417	
Number of plots	BIEN db 363,258	
Trait measurements[a]	BIEN db 25,932,454	
Species geographic range maps	BIEN db 98,829	
	Open Range db 289,743	
Number of species on Earth	Embryophyta (land plants)	Pre-TNRS Total # species names 1,323,320 Post-TNRS Total # species names 363,567–457,949
	Angiosperms (flowering plants)	Post-TNRS Total # species names 323,377–404,043
	Gymnosperms	Post-TNRS Total # species names 1280–1429
	Gymnosperms (conifers)	Post-TNRS Total # species names 856–978
	Gymnosperms (non-conifer)	Post-TNRS Total # species names 424–451
	Ferns and allies	Post-TNRS Total # species names 12,102–20,921
	Bryophytes	Post-TNRS Total # species names 26,808–31,556

Note: The table details the total number of observations, categorized into specimens, plot observations and trait data, along with derived products like geographic range maps (BIEN-curated and open access) and species counts across major plant groups. Species numbers are presented as a range: the lower bound reflects TNRS accepted species, while the upper bound includes names that lack taxonomic opinions. A full description of data sources and contributors is available in Table S1. Code underlying these data is available in Enquist et al. (2026a) in the file 1_BIEN_stats.R.

[a] This includes observation records from trait observations (from the literature) and trait measures obtained from metadata included with observation records in plots (includes measures of size [diameter, height] and growth form [tree, liana etc.]).

workflows that illustrate distinct use cases within the BIEN ecosystem. We encourage users seeking additional examples or reproducible templates to consult the code and workflows linked in these studies.

These applications span a broad range of topics, including taxonomic standardization and species richness estimation, trait–environment modelling, macroecological and biogeographic analyses, species distribution and range modelling, climate change impact and vulnerability assessments, community assembly and diversity–function relationships, phylogenetic and evolutionary studies, biodiversity forecasting and scenario modelling, and data quality assessment and validation.

Below, we present six general use cases that demonstrate how BIEN's modular services and comprehensive plant database enable reproducible workflows for data integration, cleaning and analysis.

3.5.1 | Use case 1—Quantifying the total number of named plant species on earth

BIEN enables researchers to estimate the total number of known land plant species by resolving taxonomic inconsistencies using the TNRS. Taxonomic standardization corrects synonyms, spelling errors and ambiguous names, reducing redundancy and enabling consistent and reproducible global diversity estimates. Incorrect, ambiguous or synonymous taxon names present a persistent obstacle for studies in comparative biology, biodiversity science and conservation (Dayrat, 2005). Large-scale ecological analyses, phylogenetic studies and conservation planning often rely on data sets aggregated from diverse sources. Without taxonomic standardization, such counts can inflate species richness estimates and bias scientific conclusions and conservation policy (Bortolus, 2008).

TABLE 4 The majority of plant observation data have errors and possibly issues flagged by the BIEN workflow.

Validation flag(s)	Description	Number of observation records failing validation
Native status (NSR)	Checks whether or not the organism is known to be native in the region based on published checklists	102,401,972 (**36.00%**) (26.02% introduced, 9.98% unclear)
Taxonomic (TNRS)	Checks whether or not the taxonomic name matches an accepted (or no opinion) name	*Original Name Pre-TNRS* 94,164,849 (**33.10%**; Not 'Accepted') 87,277,419 (**30.68%**; Neither 'Accepted' or 'No Opinion') *Scrubbed Name Post-TNRS* 14,553,538 (**5.12%**; Not 'Accepted') 3,640,826 (**1.28%**; Neither 'Accepted' or 'No Opinion')
Geovalidation (GNRS and GVS)	Checks whether or not coordinates are possible and fall within the specified political divisions	62,510,599 (**21.97%**)
Cultivated status (NSR)	Checks whether or not the organism is cultivated or not cultivated based on location and provided metadata	3,588,539 (**1.26%**)
Centroid (GVS)	Checks whether or not the occurrence likely represents a geopolitical centroid	861,945 (**0.30%**)
TNRS, GVS, GNRS	Occurrence records that have a TNRS perfect match score of 1 and with *no* flags indicating no error in Geovalidation or a Centroid	159,189,390 (**55.96%**)

Note: Summary of validation flags applied to plant observation data. This table summarizes the number and fraction of plant observation records flagged for potential biases and errors during validation using BIEN's key tools: TNRS, GNRS, NSR and GVS. These flags identify issues in taxonomy, geographic coordinates and other metadata recorded with the original observation or specimen. The bold values indicate the proportion of records in that category that are erroneous. Of the total processed records, 159,189,390 (55.96%) passed all taxonomy and geographic accuracy checks, ensuring high confidence for ecological and evolutionary analyses (see Figure 4). Queries for these analyses are available in the supplemental document and Enquist et al. (2026a) in the file 1_BIEN_stats.R.

Prior to taxonomic standardization with the TNRS, the BIEN database contained 1,323,320 unique land plant names (Table 3) across more than 284 million botanical observation records—a number far exceeding any reasoned estimate of global plant diversity. Between 72.5% and 65.4% of the taxonomic names encountered were erroneous, synonymous or unclear. After standardization, the final data set consists of 34,902,348 unique observation records, representing 363,567–457,949 land plant species (Table 3; the lower bound estimate reflects TNRS accepted species, while the upper bound includes names that lack taxonomic opinions). Importantly, without TNRS correction, 959,753–865,371 names (or 72.5%–65.4%) in global biodiversity data sets are redundant, incompatible or erroneous. In Table 3, we list the total number of species for each major land plant clade. The workflow underlying these estimates is available in Enquist et al. (2026a) in the file 1_BIEN_stats.R.

These data, derived from herbarium specimens and ecological plot surveys, provide a comprehensive and standardized global view of plant diversity. The TNRS plays a pivotal role in ensuring that this data set reflects the best available taxonomic knowledge, enabling accurate and reproducible matching of species identities across sources.

3.5.2 | Use case 2—The cumulative number of plant species recorded over time

The gap between the true number of species that actually exist and those that have already been described is known as the Linnean shortfall (Brown & Lomolino, 1998). Accurately estimating the number of known plant species is a foundational challenge in biodiversity science, essential for advancing ecological theory, conservation planning and global biodiversity assessments (Hortal et al., 2015a). Reliable estimates require resolving taxonomic inconsistencies, correcting synonyms and minimizing errors that inflate richness estimates or bias conclusions (Boyle et al., 2013; Lessa et al., 2024; Lu & He, 2017). Open, reproducible workflows for calculating the Linnean shortfall are needed to ensure transparency, validation and adaptability for the broader research community (e.g. Lessa et al., 2024).

Using BIEN's standardized framework, we conducted an updated analysis of the cumulative number of plant species recorded over time, following the methods from Feng et al. (2022). We obtained data from BIEN (version 4.2; accessed December 2024) that compiled plant distribution data from GBIF (https://doi.org/10.15468/dl.87zyez) and non-GBIF sources. These data sets were harmonized through BIEN's data scrubbing and standardization pipeline, including the TNRS (version 5.3.1). Only records with valid collection years and standardized taxonomic names were included in the analysis.

We quantified species accumulation from 1750 to 2020 and fitted cumulative species counts as a function of year (Figure 5). The data were categorized into three groups: GBIF, non-GBIF and the combined full data set. We counted the number of species in all three data sources over time, from 1750 to 2020. We fitted the cumulative numbers (dependent variable) and years (independent variable) with logistic regression using ordinary least squares ['nls' function in stats package version 4.2.0 in R version 4.2.0; R Core Team 2021]. Results indicate a gradual approach towards saturation in the number of recorded species (Figure 5). Based on the logistic model, the predicted asymptote for terrestrial plant species in BIEN is 438,157 ± 1895 (mean ± SD), using taxonomically standardized names resolved by TNRS (version 5.3.1; Boyle et al., 2013). The full workflow is available from Enquist et al. (2026b).

This analysis also enabled an estimate of the Linnean shortfall for plants. The current number of species in the BIEN database is conservatively estimated at 363,567 (Table 3). The difference between the asymptotic estimate (438,157 ± 1895 species) and the current count suggests that approximately 74,590 ± 1895 species remain to be discovered and described (Figure 5). Our analysis is based on currently incorporated records, but this trajectory will shift as new field observations and historical specimen data are integrated. Factors influencing this dynamic include: (i) newly described species often derived from older specimens; (ii) rates of lumping and splitting of taxa due to ongoing taxonomic revision; and (iii) variation in species concepts (Lessa et al., 2024). These factors introduce uncertainty into estimates of the Linnean shortfall, as differing rates of discovery and taxonomic change can underestimate or overestimate the total number of extant species (Lessa et al., 2024). Despite these challenges, our analysis provides an open and reproducible workflow that serves as a basis for improving our understanding of global plant diversity.

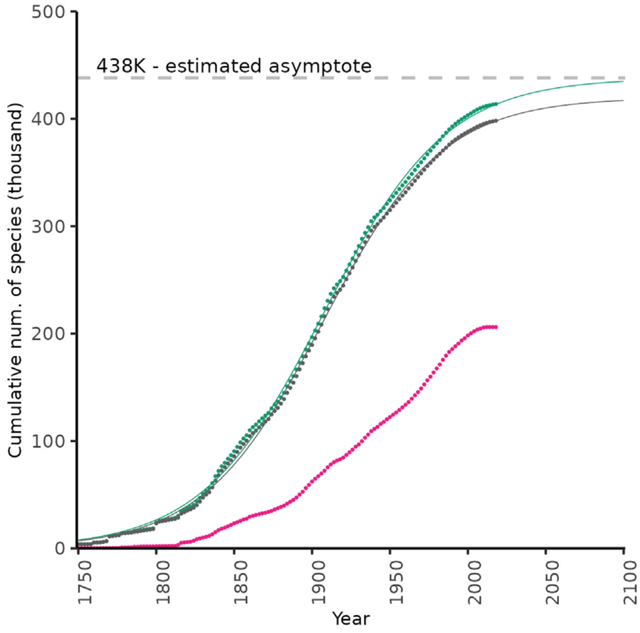


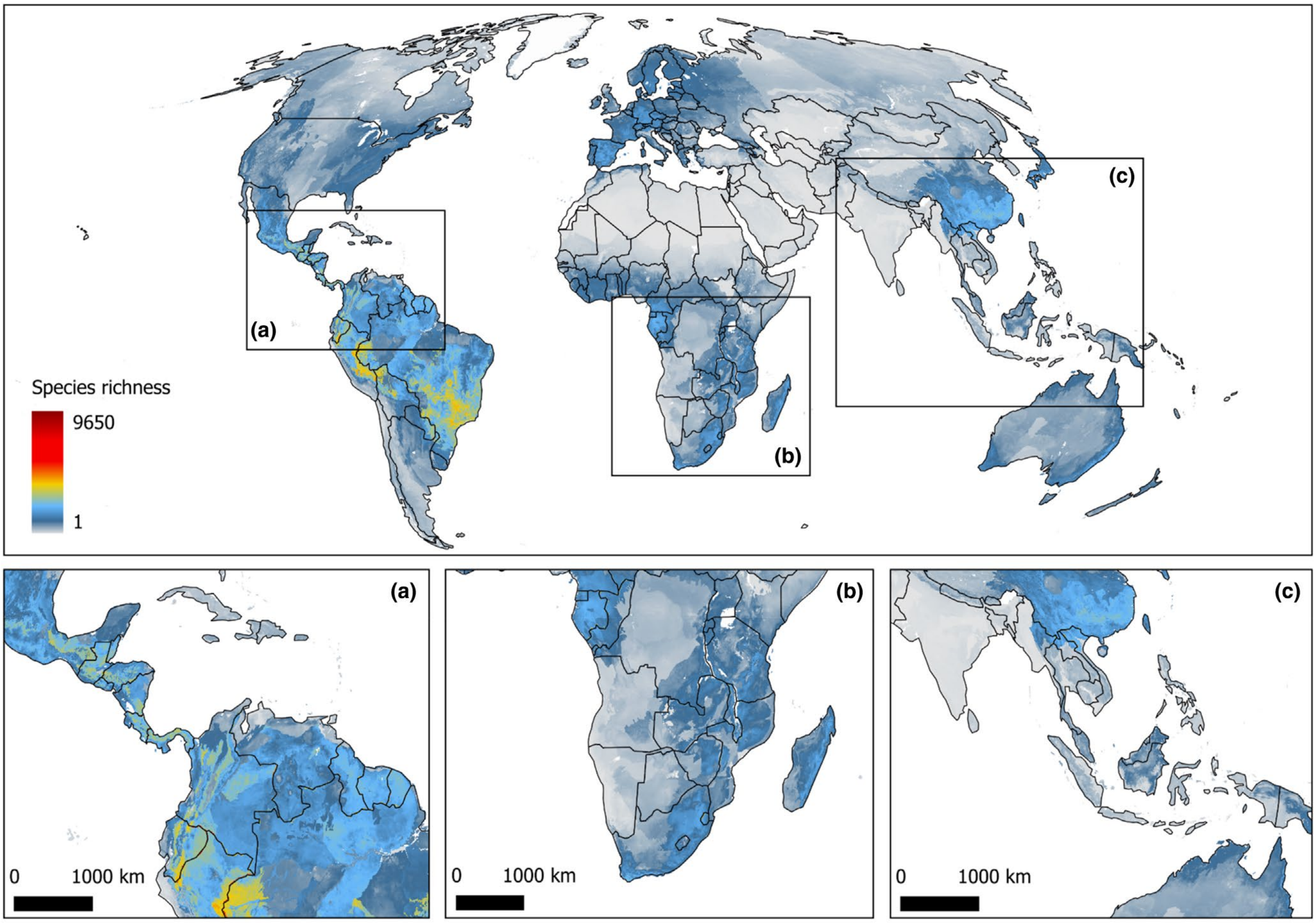
FIGURE 5 The cumulative number of Embryophyte species recorded from 1750 to 2020 across three data sources: BIEN (green), GBIF (grey) and other non-GBIF sources (red). The x-axis represents time in years, while the y-axis shows the cumulative number of unique species identified after taxonomic standardization through the TNRS. The figure builds on the analysis in Feng et al (2022) and highlights each data set's contributions to understanding global plant diversity. BIEN integrates data from GBIF, non-GBIF sources, and additional data sources, providing a comprehensive overview. The curves show the trajectory of species discovery over time, with an initial rapid increase followed by a gradual levelling off. This pattern suggests that species accumulation is approaching an asymptotic limit, indicating that we are nearing a near-complete inventory of land plant species on Earth. This trend reflects the effectiveness of current sampling and data harmonization efforts and underscores the importance of BIEN's harmonized data set as a critical tool for biodiversity science. It provides strong evidence that the global catalogue of land plant species is approaching saturation.

3.5.3 | Use case 3—Mapping global plant species richness

Accurate global species richness maps are essential for identifying biodiversity hotspots, guiding conservation priorities and assessing the ecological impacts of global change (Hannah et al., 2020). BIEN's harmonized and augmented occurrence data enable users to generate high-quality species distribution models that filter out cultivated or uncertain records and standardize taxonomy. BIEN's high-resolution global richness maps incorporate only validated occurrences, ensuring taxonomic and geographic precision. These efforts build on several published studies using BIEN data to generate species geographic ranges and to map plant biodiversity (Echeverría-Londoño et al., 2018; Goldsmith et al., 2016; McFadden et al., 2019; Morueta-Holme et al., 2013; Moulatlet et al., 2025; Neves et al., 2021). New global updates are the focus of several forthcoming papers that showcase the distribution of plant diversity across geographic regions (Merow et al. *in prep*), and assessments of the quality and methodology of the range maps are discussed within.

Using BIEN's standardized Open Range db, we generated global plant biodiversity maps by stacking species distribution models for 289,743 plant species (Figures 6 and 7; Merow et al. in prep; see also Moulatlet et al., 2025). Species range maps were derived from BIEN's integrated occurrence records, which were processed using the TNRS and GNRS to resolve taxonomic inconsistencies and validate spatial information. Native species records were identified and filtered using the NSR to exclude introduced and cultivated species.

These maps reveal global latitudinal diversity gradients and highlight regions of exceptional species richness, including the Amazon Basin, Lower Guinea and Southeast Asia. While sampling intensity remains uneven across regions (Figure 7), particularly outside the Americas, the use of standardized, validated data substantially improves the reliability of diversity estimates (Merow et al., *in prep*). By resolving taxonomic inconsistencies, validating geographic data and filtering for native species, BIEN enhances the precision and reproducibility of large-scale biodiversity mapping. Continued improvement in geographic sampling and data integration will further refine these estimates. Nonetheless, BIEN's range maps may represent a foundation for global analyses of biodiversity patterns, environmental drivers and conservation priorities. They enable researchers to link species richness to environmental gradients, land use change and climate impacts, thereby supporting the development of scalable, reproducible biodiversity forecasts.


Species richness
9650
1
(a)
(b)
(c)
(a)
0
1000 km
(b)
0
1000 km
(c)
0
1000 km

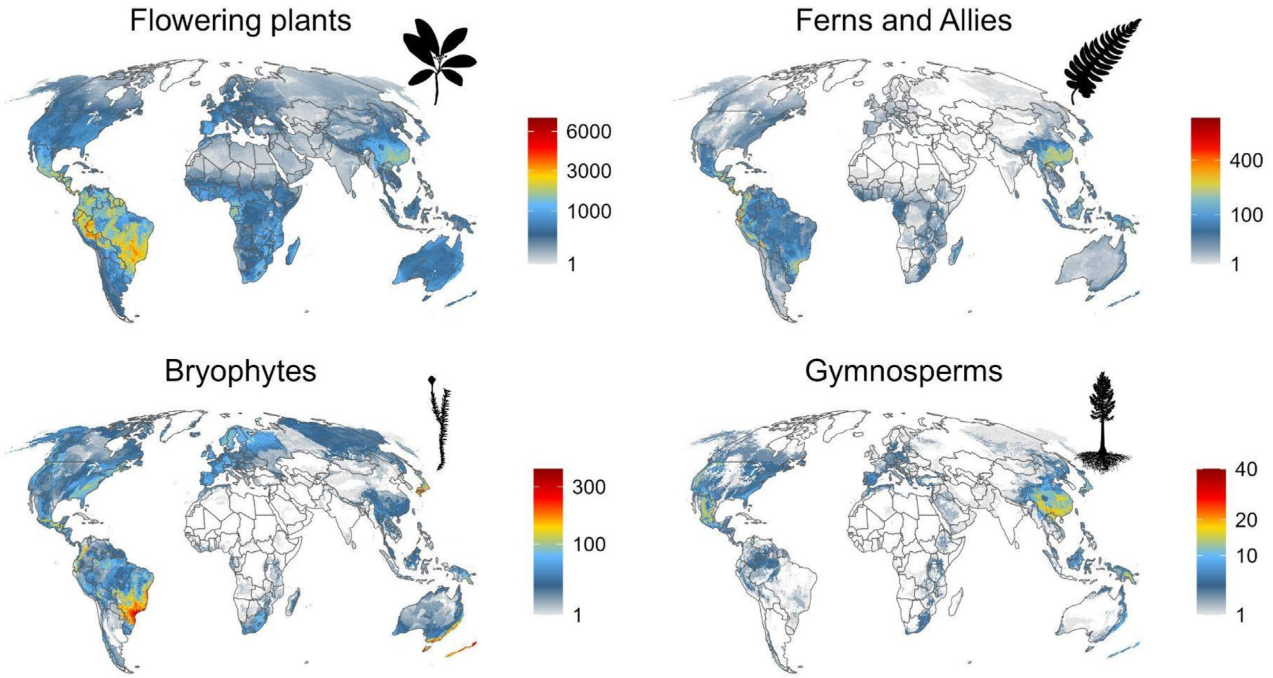

Flowering plants
6000
3000
1000
1
Ferns and Allies
400
100
1
Bryophytes
300
100
1
Gymnosperms
40
20
10
1

FIGURE 6 Global plant biodiversity maps created from stacking 289,743 Embryophyta (Land Plant) species geographic ranges created from the BIEN db. The upper geographic part figures show regional variation in species richness. The lower global part figures show plant species richness broken out by Flowering Plants, Ferns and Allies, Bryophytes, and Gymnosperms. These figures showcase global species richness generated using BIEN's harmonized and augmented data set. The maps were constructed by stacking individual species range rasters, with each raster representing the distribution of a single species. These rasters were projected in a Mollweide equal-area coordinate system at a 5 km spatial resolution (Merow et al. ms in prep). The resulting maps depict the sum of species ranges overlapping within each grid cell, highlighting patterns of biodiversity across geographic regions. Map (i) is the richness map for the Land Plants (Embryophyta) using an absolute linear scale to emphasize species richness hotspots. Map (ii) separates four major land plant clades and groups, Angiosperms, Ferns (and allies), Bryophytes and Gymnosperms.

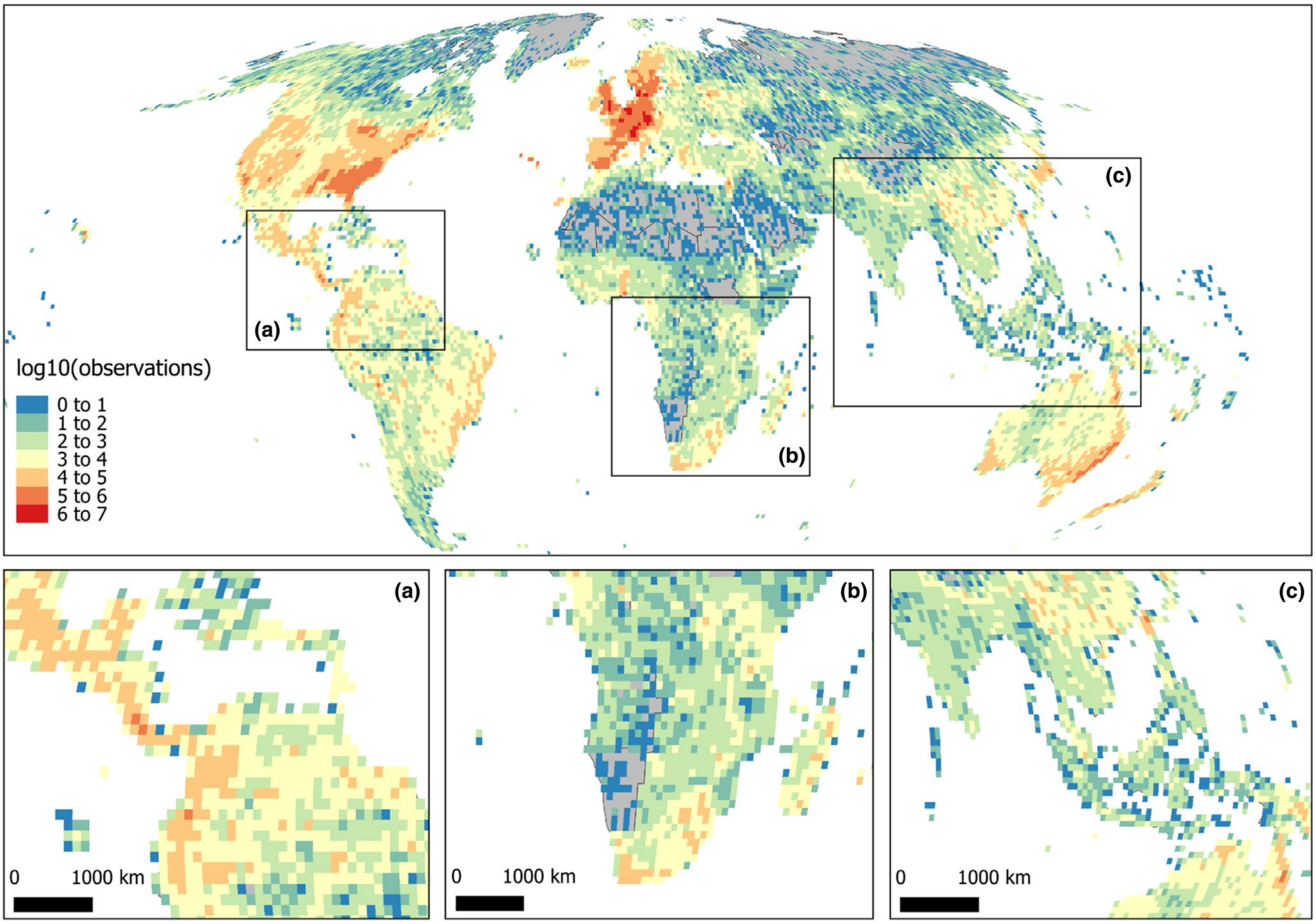


FIGURE 7 BIEN global sampling intensity maps. These maps illustrate the global distribution of sampling intensity for botanical observation records within the BIEN 4.2 db in each 100 km cell. Legend colours represent 5 classes using quantile values. Sampling intensity is represented as the density of records per unit area, highlighting regions with extensive data coverage and areas with sparse or incomplete sampling. The global sampling map (overview) integrates three primary data types: herbarium specimens, ecological plot surveys and trait observations. While these maps provide a broad overview, it is important to note some limitations. Maps underscore the uneven global distribution of plant biodiversity data, with sampling intensity concentrated in North America and parts of the Global North. These patterns reflect biases in historical data collection and highlight the need for increased sampling efforts in biodiversity-rich but underrepresented regions such as the Global South. Addressing these gaps is critical for achieving a more comprehensive understanding of global biodiversity patterns and improving the predictive capacity of biodiversity science. For example, ecological plots within BIEN 4.2 are overrepresented in certain regions such as Europe. Future integration with several rapidly growing plot and trait data networks (Cooper et al., 2024; Feng et al., 2022; ForestPlots.net et al., 2021; Gallagher et al., 2020; Kattge et al., 2020; Sabatini et al., 2021) will help fill in sampling gaps. Updates in BIEN 4.2 and beyond address these gaps, incorporating additional plot networks and refined geographic mapping to improve representation in undersurveyed regions.

3.5.4 | Use case 4—How many plant species are found in the UNESCO protected areas across the globe?

This use case addresses the challenge of accurately identifying native plant species within UNESCO World Heritage sites, correcting long-standing taxonomic and geographic errors in biodiversity data (Figure 8; see also Nic Lughadha et al., 2018). UNESCO World Heritage sites occur disproportionately in global biodiversity hotspots. They are believed to safeguard a significant fraction of global plant diversity while occupying less than 1% of Earth's surface (UNESCO International Union for Conservation of Nature, 2023). Quantifying their biodiversity contributions requires an open, transparent and reproducible workflow. Using BIEN, we partnered with UNESCO to assess the botanical diversity of these sites by integrating high-resolution, taxonomically standardized plant occurrence data with protected-area boundaries.

The workflow (available in Enquist et al. (2026a) in file 3_UNESCO_use_case.R) begins by querying BIEN's comprehensive database to retrieve plant occurrence records within UNESCO site boundaries. All taxonomic inconsistencies are resolved using the TNRS; geographic coordinates are validated through the GVS; and non-native species are flagged using the NSR. The resulting records are spatially overlaid with UNESCO site polygons, producing a clean, validated data set linking species occurrences directly to protected areas. This reproducible workflow can be readily adapted to other protected-area networks, allowing comparative biodiversity assessments across regions.

By integrating BIEN data with IUCN Red List estimates, we found that UNESCO World Heritage sites protect approximately 74,295 plant species, around 16% of global plant diversity (UNESCO International Union for Conservation of Nature, 2023). Of these, 2,130 species are classified as threatened, underscoring their critical conservation importance. However, if raw, unstandardized data were used (i.e. prior to TNRS correction), the estimated number of unique species would be inflated by roughly 52% (38,633 additional names) due to unresolved synonyms and misapplied taxa (UNESCO International Union for Conservation of Nature, 2023).

3.5.5 | Use case 5—Fetching and cleaning occurrence data for cactus distribution modelling

This use case highlights how BIEN's modular and reproducible ecosystem enables the development of high-quality, data-driven species distribution models. The workflow, Enquist et al. (2026c), developed by Pillet et al. (2022; Pillet et al. ms in prep) demonstrates BIEN's capacity to address global biodiversity challenges by automating the retrieval, cleaning and validation of occurrence data for the Cactaceae family—a taxonomically and ecologically diverse group often underrepresented in global data sets. The pipeline integrates multiple BIEN services, including RBIEN, TNRS, GNRS and NSR, along with the occTest package, to minimize manual curation and enhance data reliability (Figure 8). Specifically, it begins by querying global Cactaceae records from BIEN, then standardizes taxonomic names using the TNRS. Geographic validation is then performed through the GNRS and the NSR, which together flag occurrences located outside known native ranges.

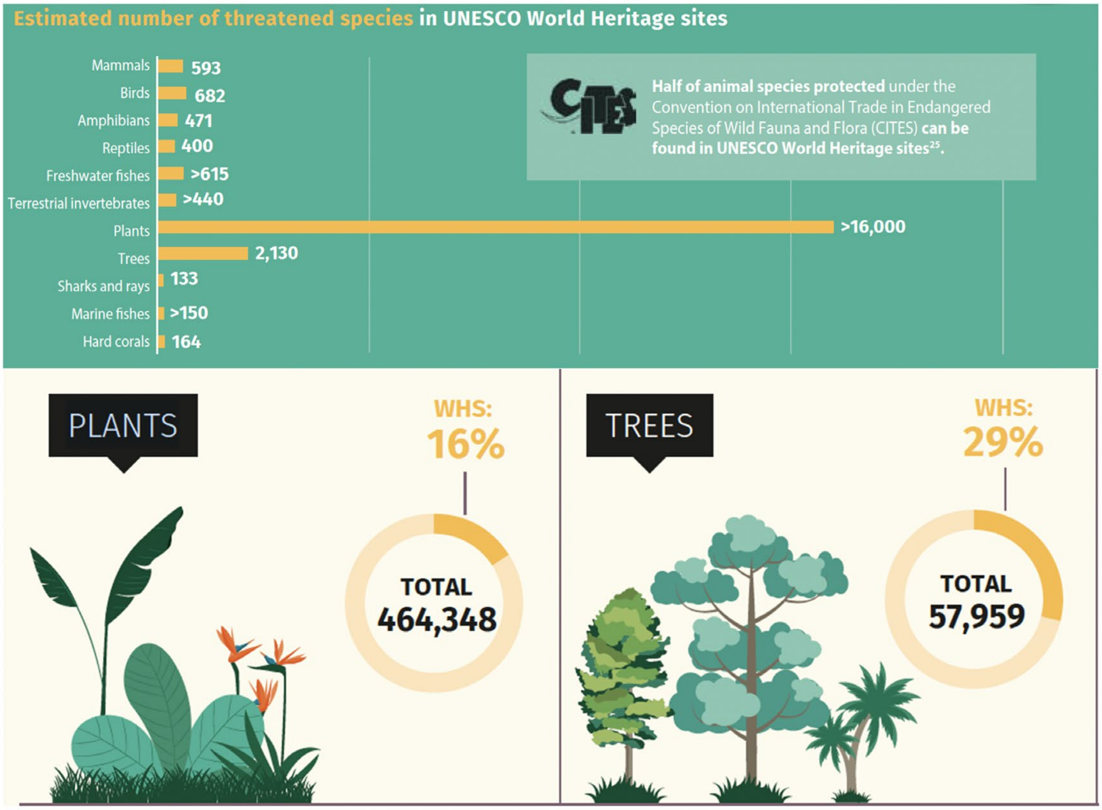
This use case illustrates how BIEN's ecosystem facilitates the development of high-quality species distribution models for Cactaceae (Pillet et al. ms in prep) while addressing challenges such as taxonomic inconsistencies, spatial errors and non-native observations (Figure 8). Despite filtering out low-quality records and species with sparse data, the resulting models produced robust and biologically consistent predictions of cactus species richness and distribution. This workflow underscores BIEN's role in enabling fully reproducible biodiversity modelling pipelines, in which users can combine standardized taxonomic, geographic and environmental validation tools to generate transparent, replicable data products for large-scale ecological research.

3.5.6 | Use case 6—Building your own integrated biodiversity data set by running your own data harmonization and scrubbing pipeline

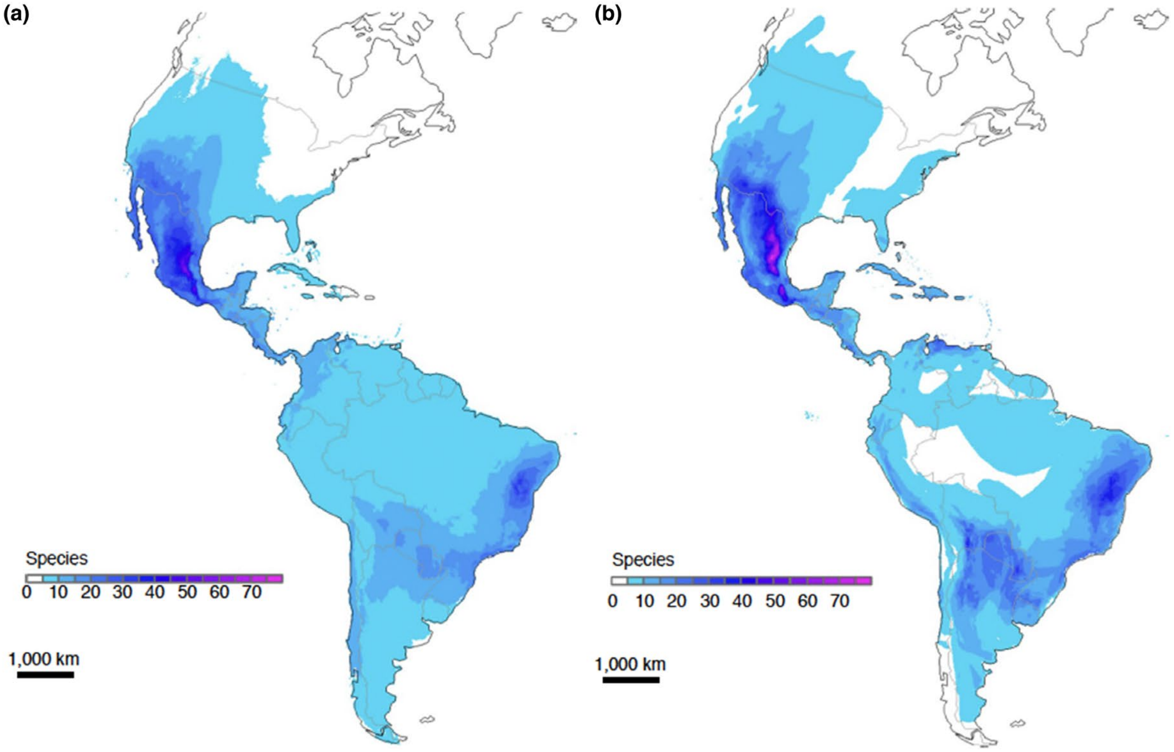
One of BIEN's most novel features is its modular architecture (Figure 2), which enables users to construct reproducible, customizable workflows tailored to their specific research goals. A data

FIGURE 8 (I) Contribution of UNESCO World Heritage Sites (WHS) to global biodiversity conservation. This figure from UNESCO-IUCN (2023) illustrates the estimated number of threatened species and the percentage of species protected within UNESCO World Heritage Sites (WHS), focusing on plants and trees. The WHS network protects an estimated 464,348 plant species (16% of global plant diversity), including 2130 threatened taxa, and safeguards 57,959 tree species, representing 29% of global tree diversity. The analysis leverages the BIEN database and workflow to integrate plant occurrence data with IUCN Red List assessments and UNESCO site boundaries through spatial overlays. Key BIEN tools (such as the Taxonomic Name Resolution Service (TNRS), Geographic Name Resolution Service (GNRS), Geocoordinate Validation Service (GVS) and Native Species Resolver (NSR)) enhance data accuracy by resolving taxonomic ambiguities, validating geographic data, and identifying native and introduced species. These steps reduce bias and errors, ensuring harmonized, reliable data sets. This figure highlights WHS as critical to global biodiversity conservation, protecting significant portions of Earth's plant and tree diversity, including many vulnerable taxa. It demonstrates how BIEN supports reproducible biodiversity science, improving data quality and representation for conservation planning and policy decisions. Data and Figure Source: BIEN database and UNESCO-IUCN (2023) (II) Comparison of stacked species distribution models predicted for current cactus species richness. (a), The average cactus species richness across all maps for North and South American cactus species using BIEN data and sources. (b), Cactus richness based on expert maps by the IUCN. Figure is from Pillet et al. (2022).

I


Estimated number of threatened species in UNESCO World Heritage sites
Mammals 593
Birds 682
Amphibians 471
Reptiles 400
Freshwater fishes >615
Terrestrial invertebrates >440
Plants >16,000
Trees 2,130
Sharks and rays 133
Marine fishes >150
Hard corals 164
CITES
Half of animal species protected under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) can be found in UNESCO World Heritage sites[25].
PLANTS
WHS: 16%
TOTAL 464,348
TREES
WHS: 29%
TOTAL 57,959

II


(a)
(b)
Species
0 10 20 30 40 50 60 70
1,000 km
Species
0 10 20 30 40 50 60 70
1,000 km

harmonization and scrubbing pipeline within the BIEN ecosystem typically proceeds through the following sequential steps:

1. (TNRS): The workflow begins by passing observation data through the TNRS, which standardizes scientific names by resolving inconsistencies, correcting spelling errors and updating outdated synonyms to accepted names. This ensures taxonomic consistency across data sets and reduces identification errors. TNRS also returns data-quality flags that can be carried forward into subsequent steps.
2. (GNRS): Next, the GNRS resolves and standardizes geographic place names (e.g. country, state, province). It corrects spelling errors and aligns locality information with known political boundaries, thereby enhancing spatial precision and enabling accurate cross-referencing among data sets.
3. (GVS): Observation data passing from the TNRS and GNRS are then further validated with the GVS, which flags geographic errors, identifies centroid-based inaccuracies and assesses coordinate precision. This ensures that spatial data are accurate, reproducible and suitable for spatial analyses.
4. (NSR): Data validated through TNRS, GNRS and GVS are then processed using the NSR, which determines whether occurrences represent native or introduced species within specific political divisions. By leveraging prior validation steps, NSR enables users to filter out cultivated or non-native observations for ecological and conservation analyses.
5. (OccTest Integration) As an optional final step, data can be further processed using the R package occTest, which identifies environmental outliers, duplicates, and records from unsuitable habitats. OccTest also provides a visualization of outputs to better scrutinize occurrence data. It integrates multiple BIEN services into a single automated pipeline and allows advanced users to customize modules for specific analytical needs.

The resulting data sets include all BIEN-derived flags and metadata augmentations (see Tables S1–S4), enabling researchers to subset and filter high-quality observation points and taxa for their specific analyses (see Figure 4). This flexible and reproducible workflow supports a wide range of research questions—from biogeographic modelling to conservation prioritization—while maintaining data integrity and provenance.

Users can deploy these tools through the BIEN web interface, R packages or APIs, creating automated workflows that yield harmonized taxonomic and geographic data, validated coordinates and native status flags. Whether leveraging pre-integrated tools like occTest or assembling bespoke pipelines, BIEN's modular system empowers researchers to clean, integrate and analyse biodiversity data with precision, transparency and reproducibility.

3.6 | BIEN as an open community

The ecosystem of BIEN data products, software products and services, and the people who have built, maintained and used these products, represents an open science community that magnifies and stimulates new research capabilities in biodiversity science. Building the BIEN ecosystem of data, tools and services required collaborative work from many people in biodiversity analytics and modelling, data systems design and development, software design and development, and computing systems operation and security. By following open science principles, BIEN's data and workflow can also be used more broadly in other biodiversity initiatives. For example, BIEN plant occurrences and range maps could be integrated with BON in a Box (developed by the Group on Earth Observations Biodiversity Observation Network, or GEO BON) to support biodiversity modelling and indicator calculation (Griffith et al., 2024). Nonetheless, significant challenges face future efforts by BIEN and other open communities (see Supporting Information S1). For BIEN to continue to impact biodiversity science, at a minimum, sustained investment in maintenance and operations will be required to address security and software changes needed for compatibility with the evolving software landscape in which BIEN is embedded (Downs et al., 2015; Hinsen, 2019).

4 | DISCUSSION

The Botanical Information and Ecology Network (BIEN) is an integrated, modular informatics framework for biodiversity science. By addressing key challenges in data integration, quality and accessibility, BIEN provides an open resource for large-scale, reproducible biodiversity research (see further discussion in Supporting Information S2). Its dual contributions—an integrated global plant biodiversity database and a suite of open-source tools—empower researchers to harmonize and validate data, conduct advanced analyses and generate insights into plant diversity and conservation.

4.1 | How the BIEN relates to other large-scale plant occurrence databases

Two key distinctions set the BIEN db apart from other large-scale biodiversity databases. First, BIEN integrates species occurrence data from multiple sources while ensuring high taxonomic resolution through its TNRS. For example, in comparison to the plant occurrence data of GBIF, BIEN provides broader taxonomic and spatial coverages (Feng et al., 2022; Figure 5). Also, GBIF used a comprehensive strategy to develop a taxonomic backbone that fits their aggregating purpose; in contrast, BIEN used TNRS to standardize the indexed taxonomy, which is expected to be of higher scientific rigour, evidenced by the broader usage of this approach by the community (Denelle et al., 2023; Kattge et al., 2020). Second, BIEN provides open-access tools and broad open-access availability of both species occurrence and vegetation plot data, facilitating biodiversity research at multiple scales. BIEN is among the largest and most comprehensive integrated plant biodiversity databases available. It aggregates millions of species occurrence records and includes harmonized vegetation plot data. Unlike many regional databases, BIEN offers globally standardized taxonomic and spatial data, ensuring comparability across studies. The BIEN workflow tools enable the integration and resolution of taxonomic and geographic inconsistencies, making it a valuable resource for ecological modelling, macroecology and conservation planning. By incorporating BIEN's TNRS, researchers can efficiently resolve taxonomic ambiguities, ensuring consistency in biodiversity data sets. Third, while there are several rapidly growing plot data databases/networks [e.g. sPlot (Sabatini et al., 2021)], the workflow developed by BIEN enables effective integration of plant occurrence, plot and trait data; the capacity to integrate different types of data will be very valuable in filling gaps in our global biodiversity knowledge in the next stage of biodiversity informatics (Feng et al., 2025; Hortal et al., 2015b).

4.2 | How BIEN relates to other global trait database efforts

In terms of the absolute total number of trait observation records, BIEN 4.2, with 25,932,454 records, *is the largest compilation of plant trait observations*. This is 10 million more records than the TRY global trait observation database (Kattge et al., 2011). The observation records within BIEN that contain trait data are all from (i) published open-source data sets (over ~800 published papers and books have contributed trait data observations in BIEN) or (ii) trait values that are extracted from other plant observation data sets (e.g. measurements of plant size, growth form from ecological plot surveys).

The BIEN trait data are well-suited to studies of intra- and interspecific variation in plant size because each individual has a large sample of plant size observations. The trait data within BIEN are a smaller subset of the number of different types of traits (BIEN has focused on a smaller number of core traits, 54 different traits) than in TRY (TRY contains around 2661 different traits). Furthermore, many plant trait observations in BIEN are derived from querying plot and survey data for measures of plant size (growth form, diameter at breast height [dbh], and height). Significantly, BIEN's data curation model and database schema differ from TRY in that they codify traceability of each trait observation to the original data provider (see criticism of the TRY database on data applicability and traceability in Augustine et al., 2024).

4.3 | BIEN as a model for future biodiversity database systems

BIEN's transparent, flexible architecture promotes collaboration, future development and Open Science, fostering innovation in the study of biodiversity patterns, species distributions and functional traits. By integrating herbarium, plot, citizen science and trait data into a geospatial framework, BIEN enables researchers to tackle complex ecological and conservation questions with unprecedented precision and scope.

While the BIEN database offers comprehensive coverage of global plant biodiversity, crucial sampling gaps remain (Figure 6). Large degrees of sampling heterogeneity exist. Important knowledge gaps in Asia and Africa undoubtedly limit our ability to assess the geographic distribution of plant diversity, to model species distributions and to estimate the total number of species. Addressing these biodiversity knowledge gaps will require addressing the representation of data collaborative networks (Pettorelli et al., 2021). Filling crucial gaps in the representation of researchers from the Global South persists while researchers from the Global North remain overrepresented. These imbalances exacerbate existing biodiversity knowledge gaps and limit global biodiversity knowledge (Park, Feng, et al., 2023).

Looking ahead, the BIEN workflow provides a model for advancing biodiversity science and biodiversity informatics and filling numerous knowledge and representation gaps. The BIEN workflow addresses the urgent challenges posed by biodiversity loss and climate change. By creating a globally integrated plant data observation network, BIEN establishes a foundational resource that underscores the critical role of plants in shaping ecosystems, regulating the climate and supporting life on Earth. However, plants are only one component of the intricate web of life and of ecosystem function.

Extending similar efforts to link with other major organism groups, such as fungi, invertebrates, insects, vertebrates and microbial communities, is essential for developing a comprehensive understanding of biodiversity and ecosystem dynamics. Expanding this framework to encompass diverse taxa would enable us to explore and quantify critical interactions, such as those between plants and their mutualists, decomposers and consumers (see the GLOBI data set of interactions, https://www.globalbioticinteractions.org/ (Poelen et al., 2014)). Geographic variation in the diversity of mycorrhizal associations (Mikryukov et al., 2023; Toussaint et al., 2020), pollinators (Olesen & Jordano, 2002), herbivores (Olff et al., 2002) and soil microbes (Buzzard et al., 2019) is among the most obvious 'next adjacent possibles' to link with. Such integration of diverse taxa and ecologies is crucial for understanding the cascading effects of biodiversity loss and alteration across trophic levels and their implications for ecosystem resilience, services and stability under global change.

AUTHOR CONTRIBUTIONS

Brian J. Enquist, Brad Boyle and Brian S. Maitner conceived the overarching goals and objectives and contributed to the design of the BIEN workflow and methodology. Brian J. Enquist, Brad Boyle, Brian S. Maitner, Daniel S. Park, Erica A. Newman, Michiel Pillet, Josep M. Serra-Diaz, Rethvick S. Y. Babu, George G. C. Barbosa, Nathan Casler, John C. Donoghue, Daniel Guaderrama, Rohith K. Sajja, Thomas L. P. Couvreur, Peter M. Jørgensen, Danilo M. Neves, Ary Oliveira-Filho, Robert K. Peet, Oliver Purschke, Hans ter Steege, Jens-Christian Svenning, Barbara Thiers, Cyrille Violle, Jan J. Wieringa and Susan K. Wiser contributed to data curation, integration and quality control by aggregating, validating, cleaning and standardizing data. Brian J. Enquist, Brad Boyle, Brian S. Maitner, Cory Merow, Xiao Feng,

Gabriel M. Moulatlet, Erica A. Newman, Daniel S. Park, Patrick R. Roehrdanz, Josep M. Serra-Diaz, John C. Donoghue, Matthew B. Jones, Peter M. Jørgensen, Nathan J. B. Kraft, Pablo A. Marquet, Brian J. Maitner, Naia Morueta-Holme, Naia Morueta-Holme, Danilo M. Neves, Robert K. Peet, Oliver Purschke, Brody Sandel, Mark Schildhauer, Irena Simova, Hans ter Steege, Jens-Christian Svenning, Barbara Thiers, Cyrille Violle and Susan K. Wiser participated in collaborative working groups that helped frame the ideas, workflows, biodiversity informatics, data integration and applications. Brad Boyle, Brian S. Maitner, Cory Merow, Josep M. Serra-Diaz, Rethvick S. Y. Babu, George G. C. Barbosa, Nathan Casler, John C. Donoghue, Daniel Guaderrama, Brody Sandel, Rohith K. Sajja and Matthew B. Jones developed software, maintained computational infrastructure and workflow automation. Brad Boyle, Xiao Feng and Cory Merow analysed and interpreted the integrated data to address key biodiversity knowledge gaps and evaluate the performance of the BIEN workflow. Brian S. Maitner, Daniel S. Park, Xiao Feng, Gabriel M. Moulatlet, Rethvick S. Y. Babu, Nathan Casler and Daniel Guaderrama created data visualizations, figures and other outputs to support data interpretation and communication of results. Brian J. Enquist, Brian S. Maitner, Brad Boyle and Erica A. Newman led the writing of the manuscript, including drafting sections on data methodology, integration, analysis and applications of the BIEN workflow. All authors contributed critically to the drafting and revision of the manuscript, providing feedback on sections related to their expertise. Brian J. Enquist, Cory Merow, Brian J. Maitner, Mark Schildhauer and Naia Morueta-Holme secured funding and institutional support for the development of the BIEN initiative and data infrastructure.

AFFILIATIONS

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA; [2]The Santa Fe Institute, Santa Fe, New Mexico, USA; [3]Department of Integrative Biology, University of South Florida, St. Petersburg, Florida, USA; [4]Eversource Energy Center and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA; [5]Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA; [6]Department of Biology, James Madison University, Harrisonburg, Virginia, USA; [7]Division of Life Sciences, Korea University, Seoul, Republic of Korea; [8]Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA; [9]Moore Center for Science, Conservation International, Arlington, Virginia, USA; [10]Botanical Institute of Barcelona, CSIC-CMCNB, Barcelona, Spain; [11]BlackSky, Seattle, Washington, USA; [12]Rare Plant Program, California Native Plant Society, Sacramento, California, USA; [13]Department of Geography and Environmental Sciences, San Bernardino Valley College, San Bernardino, California, USA; [14]DIADE, Université Montpellier, CIRAD, IRD, Montpellier, France; [15]Naturalis Biodiversity Center, Leiden, The Netherlands; [16]National Center for Ecological Analysis and Synthesis, NCEAS, University of California, Santa Barbara, California, USA; [17]St. Louis, Missouri, USA; [18]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, USA; [19]Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile; [20]Centro de Modelamiento Matemático (CMM), Universidad de Chile-IRL 2807 CNRS Beauchef 851, Santiago, Chile; [21]Centro de Cambio Global UC, Santiago, Chile; [22]School of Biology and Ecology, University of Maine, Orono, Maine, USA; [23]Mitchell Center for Sustainability Science, University of Maine, Orono, Maine, USA; [24]Data Science Institute, University of Arizona, Tucson, Arizona, USA; [25]Section for Biodiversity, Globe Institute, University of Copenhagen, København, Denmark; [26]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil; [27]School of Geography, University of Leeds, Leeds, UK; [28]Department of Biology, Santa Clara University, Santa Clara, California, USA; [29]Center for Theoretical Study, Charles University, Praha, Czech Republic; [30]Department of Ecology, Faculty of Science, Charles University, Praha, Czech Republic; [31]Quantitative Biodiversity Dynamics, Faculty of Science, Utrecht University, Utrecht, The Netherlands; [32]Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University, Aarhus, Denmark; [33]New York Botanical Garden, Bronx, New York, USA; [34]Denver Botanic Gardens, Denver, Colorado, USA; [35]CEFE, Université Montpellier, CNRS, EPHE, IRD, Montpellier, France and [36]Manaaki Whenua, Landcare Research, Lincoln, New Zealand

ACKNOWLEDGEMENTS

The BIEN working group would like to thank several key funding sources that rose to the challenge of supporting BIEN's long-term vision. Specifically, this work was first supported by the National Center for Ecological Analysis and Synthesis, a centre funded by NSF EF-0553768 at the University of California, Santa Barbara, and the State of California. Through the DataONE network, NCEAS hosts the BIEN database systems and provides systems, network and security support for the project. Additional support was provided to the BIEN working group by iPlant/CyVerse through NSF DBI-0735191. Through the years of BIEN development, the National Science Foundation continued to fund several components of BIEN. B.J.E. thanks the initial support from an NSF CAREER award. B.J. Enquist, B.J. Maitner and C. Merow were also supported by NSF ABI-1565118. B.J. Enquist and C. Merow were supported by NSF HDR-1934790. B.J. Enquist and P.R. Roehrdanz were supported by the Global Environment Facility SPARC project grant (GEF-5810). J.-C. Svenning and B.J. Enquist acknowledge support from the Center for Informatics Research on Complexity in Ecology (CIRCE), funded by the Aarhus University Research Foundation under the AU Ideas program. J.-C. Svenning also considers this work a contribution to his VILLUM Investigator project 'Biodiversity dynamics in a changing world' funded by VILLUM FONDEN (grant 16549) and Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), funded by Danish National Research Foundation (grant DNRF173). X. Feng, D.S. Park and E.A. Newman were supported by the University of Arizona Bridging Biodiversity and Conservation Science Programme. C. Merow acknowledges funding from NSF Grant DBI-1913673.

N. Morueta-Holme acknowledges support from the Independent Research Fund Denmark (grant ATEMPO 10.46540/2064-00091B) and from the VILLUM FOUNDATION (grant VIL53086). J.M. Serra-Diaz was supported by the Program RYC2022-035668-I, funded by MCIU/AEI/10.13039/501100011033 and FSE+. I. Simova was supported by the Czech Science Foundation (grant 20-29554X). D.S. Park was supported by a grant from Korean University. We would like to thank our dear and inspirational collaborator and friend, B. Boyle, and wish him all the best in retirement. Brad was among Alwyn Gentry's last PhD students. In many ways, BIEN's efforts were shaped, inspired and fuelled by Brad's enthusiasm for biodiversity and science, and his unique ability to bridge big data, databases, natural history and the intricacies of systematics and taxonomy. Lastly, we dedicate this BIEN effort to our friend, collaborator and coauthor John C. Donoghue II. He inspired and helped spearhead much of the earlier BIEN range models and the initial BIEN computational backend. We are deeply saddened by John's passing. His kind, loving, enthusiastic and ever-optimistic spirit will be missed.

We acknowledge the herbaria that contributed data to this work: A, AAH, AAS, AAU, ABD, ABH, ABN, ABRN, ABS, ACAD, ACOR, AD, ADW, AFS, AIMS, AJOU, AK, AKPM, AKU, ALCB, ALTA, ALU, AMD, AMDE, AMES, AMNH, AMO, ANA, ANGU, ANSP, AQP, ARAN, ARIZ, ARM, AS, ASDM, ASU, ATCC, AU, AUT, AWH, B, BA, BAA, BAB, BABY, BACP, BAF, BAFC, BAI, BAJ, BAK, BAL, BARC, BAS, BASBG, BBB, BBG, BBS, BC, BCB, BCF, BCN, BCRU, BCW, BDD, BDK, BEL, BEREA, BERN, BEX, BFT, BG, BH, BHCB, BHSC, BIO, BIRA, BIRM, BISH, BKF, BLA, BM, BNM, BNRH, BO, BOCH, BOG, BOL, BOLV, BON, BONN, BOON, BORH, BOTU, BOUM, BOZ, BPI, BR, BREM, BRI, BRIST, BRISTM, BRIT, BRIU, BRLU, BRM, BRNU, BRSN, BRU, BRWK, BRY, BSB, BSIP, BTN, BUF, BUL, BUT, C, CAM, CAMU, CAN, CANB, CAS, CATIE, CAY, CBG, CBM, CBS, CDBI, CEN, CEPEC, CESJ, CGE, CGG, CGMS, CHAM, CHAP, CHAS, CHI, CHL, CHR, CHRB, CHSC, CICY, CIIDIR, CINC, CLE, CLEMS, CLF, CLR, CM, CMC, CMM, CNH, CNPO, CNS, COA, COAH, COFC, COI, COL, COLO, CONC, COR, CORD, CP, CPAP, CPUN, CR, CRAI, CRP, CS, CSLA, CSU, CTES, CTESN, CU, CUVC, CUZ, CVRD, CYN, DAL, DAO, DAV, DBG, DBN, DCR, DEE, DES, DFS, DFSM, DGS, DHM, DLF, DMFS, DMNH, DMU, DNA, DOR, DPU, DR, DS, DSM, DSY, DUE, DUKE, DUSS, E, EA, EAC, EBH, ECON, ECU, EGHB, EIF, EIU, EKY, ELN, EMMA, ENCB, EPM, ER, ERA, ERE, ESA, ETH, EXR, F, FAA, FAU, FAUC, FB, FCO, FCQ, FEN, FH, FHI, FHO, FI, FLAS, FLOR, FM, FR, FRS, FRU, FSC, FSU, FTG, FUEL, FULD, FURB, G, GA, GAT, GB, GC, GDA, GENT, GEO, GES, GGO, GH, GI, GJO, GL, GLAM, GLM, GMDRC, GMNHJ, GOET, GOW, GRA, GRM, GSW, GUA, GUAM, GW, GZU, H, HA, HAC, HAL, HAM, HAMAB, HAMU, HAO, HAS, HAST, HASU, HAW, HB, HBG, HBR, HDD, HEID, HGI, HGM, HIB, HIP, HITBC, HIWNT, HLU, HME, HNT, HNUB, HO, HPL, HRCB, HRP, HSS, HSU, HTN, HU, HUA, HUAL, HUEFS, HUEM, HUFU, HUJ, HUSA, HUT, HWB, HXBH, HYO, IAA, IAC, IAN, IB, IBF, IBGE, IBK, IBSC, ICEL, ICESI, ICN, IEA, IFO, ILL, ILLS, INB, INEGI, INM, INPA, INV, IPA, IPRN, ISC, ISE, ISKW, ISL, ISTC, ISU, IZAC, JACA, JBAG, JBGP, JCT, JE, JUA, JYV, K, KAND, KATH, KCS, KGY, KHD, KIEL, KLE, KMN, KMNH, KOELN, KOR, KPABG, KPM, KR, KSP, KSTC, KSU, KTU, KU, KUN, KYO, L, LA, LAE, LAGU, LANC, LAU, LBG, LBV, LCN, LCR, LD, LDS, LE, LEA, LEB, LEI, LES, LI, LIL, LINN, LISC, LISE, LISI, LISU, LIV, LIVU, LL, LLN, LMA, LMJ, LMU, LOB, LOJA, LP, LPAG, LPB, LPD, LPS, LSR, LSU, LSUM, LTB, LTN, LTR, LU, LUA, LW, LYJB, LZ, M, MA, MACB, MACF, MAF, MAK, MAL, MAN, MANCH, MARY, MASS, MB, MBA, MBH, MBK, MBM, MBML, MCNS, MEL, MELU, MEN, MERL, MEXU, MFA, MFU, MG, MGC, MHA, MICH, MIL, MIN, MISS, MJG, MMMN, MNA, MNE, MNHM, MNHN, MO, MOR, MPU, MPUC, MSB, MSC, MSE, MSTR, MSUB, MSUN, MT, MTMG, MU, MUB, MUCV, MUR, MVFA, MVFQ, MVJB, MVM, MW, MY, N, NA, NAC, NAS, NCC, NCCBH, NCCE, NCCH, NCE, NCH, NCSC, NCU, NE, NEBC, NH, NHA, NHG, NHM, NHMC, NHT, NM, NMB, NMLU, NMNL, NMR, NMSU, NMW, NOT, NOU, NSPM, NSW, NT, NTS, NU, NUM, NWOSU, NY, O, OBI, OCLA, ODU, OKL, OKLA, OMG, OS, OSA, OSC, OSH, OULU, OWU, OXF, P, PACA, PAMP, PAR, PCU, PDD, PE, PEL, PERTH, PEUFR, PFC, PGM, PH, PI, PKDC, PLAT, PLH, PLYP, PMA, PMSP, PNG, PNH, POLL, POM, PORT, PR, PRC, PRE, PRU, PSU, PSY, PTH, PVNH, PY, QCA, QCNE, QFA, QK, QM, QMC, QRS, QUE, R, RAMM, RAS, RB, RBR, RDG, RDS, REG, RELC, RENO, RFA, RIOC, RM, RNG, RSA, RYU, S, SALA, SAM, SAN, SANT, SAPS, SASK, SAV, SBBG, SBT, SCAR, SCFS, SD, SDN, SDSU, SEL, SEV, SF, SFD, SFS, SFV, SGO, SHD, SHIN, SHYB, SI, SIM, SING, SIU, SJC, SJRP, SJSU, SLBI, SLSK, SMDB, SMF, SNM, SNUA, SOM, SP, SPF, SPN, SPSF, SQF, SRFA, SRGH, SSMF, STA, STDCM, STE, STI, STL, STR, STS, STU, SUN, SUU, SUVA, SVG, SWT, SZU, TAA, TAAM, TAES, TAI, TAIF, TALL, TAM, TAMU, TAN, TASH, TBI, TCD, TEF, TENN, TEPB, TEX, TFC, TI, TKPM, TNP, TNS, TO, TOGO, TOR, TOYA, TRA, TRH, TROM, TRT, TRTE, TRTS, TTN, TUB, TUCH, TULS, U, UADY, UAM, UB, UBA, UBC, UBT, UC, UCI, UCMM, UCNW, UCR, UCS, UCSA, UCSB, UCSC, UEA, UEC, UESC, UFG, UFMA, UFMT, UFP, UFRJ, UFRN, UFS, UGDA, UH, UI, UJAT, ULM, ULS, UME, UMO, UNA, UNB, UNCC, UNEX, UNM, UNR, UNSL, UNSW, UPCB, UPEI, UPNA, UPNG, UPP, UPS, US, USAS, USC, USCH, USF, USJ, USM, USNC, USP, USZ, UT, UTC, UTEP, UTLH, UU, UV, UVIC, UWO, V, VAL, VALD, VALPL, VDB, VEN, VIT, VMSL, VPI, VT, W, WAG, WAR, WARMS, WAT, WB, WCR, WCUH, WELT, WFU, WHN, WIES, WII, WIN, WIS, WM, WMNH, WOH, WOLL, WOS, WS, WSCO, WSY, WTU, WU, WXM, YA, YAMA, YK, YRK, Z, ZCM, ZMT, ZSS, ZT.

CONFLICT OF INTEREST STATEMENT

None.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70274.

DATA AVAILABILITY STATEMENT

Scripts for queries used in this manuscript are available at: https://github.com/bmaitner/BIEN_4.2_update_ms, https://github.com/shandongfx/2025_BIEN_sp, and https://github.com/mdpillet/DroughtForecasts, and have been permanently archived at https://doi.org/10.5281/zenodo.18391931 (Enquist et al., 2026a), https://

doi.org/10.5281/zenodo.18423054 (Enquist et al., 2026b), and https://doi.org/10.5281/zenodo.18423999 (Enquist et al., 2026c) (respectively).

STATEMENT OF INCLUSION

The BIEN initiative brings together a geographically diverse authorship team with expertise in plant ecology, biodiversity informatics, systematics, and conservation science. Our study aimed to integrate biodiversity data networks from global and regional databases of plant occurrences, traits and distributions to address critical knowledge gaps. Often, these data networks are organized by non-local researchers. Despite efforts to engage local researchers, particularly in biodiversity-rich regions where primary data were collected, these efforts were constrained by the scarcity of scientists working on large-scale biodiversity informatics in some areas, funding to support travel and involvement, and capacity constraints in forming new collaborations. Consequently, crucial gaps in the representation of researchers from the Global South persist. We acknowledge that our work is biased towards the perspectives and knowledge of researchers from the Global North. These imbalances exacerbate existing gaps in biodiversity knowledge. We sought to incorporate diverse perspectives by engaging with researchers at different career stages, citing relevant literature from regional experts and consulting local stakeholders to guide data interpretation. However, significant challenges remain in achieving broader inclusion, improving data integration and ensuring equitable access to biodiversity data. Moving forward, to close biodiversity knowledge gaps, it is essential to expand collaborations with scientists and institutions in underrepresented regions, promote capacity-building programmes and strengthen equitable access to biodiversity data for research and conservation planning.

ORCID

Brian J. Enquist https://orcid.org/0000-0002-6124-7096
Brian S. Maitner https://orcid.org/0000-0002-2118-9880
Xiao Feng https://orcid.org/0000-0003-4638-3927
Gabriel M. Moulatlet https://orcid.org/0000-0003-2571-1207
Erica A. Newman https://orcid.org/0000-0001-6433-8594
Daniel S. Park https://orcid.org/0000-0003-2783-530X
Josep M. Serra-Diaz https://orcid.org/0000-0003-1988-1154
Rethvick S. Y. Babu https://orcid.org/0009-0005-4965-5870
Nathan J. B. Kraft https://orcid.org/0000-0001-8867-7806
Brody Sandel https://orcid.org/0000-0003-2162-6902
Irena Simova https://orcid.org/0000-0002-9474-569X
Jens-Christian Svenning https://orcid.org/0000-0002-3415-0862
Cyrille Violle https://orcid.org/0000-0002-2471-9226

REFERENCES

Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature*, *530*(7588), 27–29.

Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial gaps in global biodiversity information and the role of citizen science. *Bioscience*, *66*(5), 393–400.

Anderson, C. B. (2018). Biodiversity monitoring, earth observations and the ecology of scale. *Ecology Letters*, *21*(10), 1572–1585.

Anderson-Teixeira, K. J., Davies, S. J., Bennett, A. C., Gonzalez-Akre, E. B., Muller-Landau, H. C., Wright, S. J., Abu Salim, K., Almeyda Zambrano, A. M., Alonso, A., Baltzer, J. L., Basset, Y., Bourg, N. A., Broadbent, E. N., Brockelman, W. Y., Bunyavejchewin, S., Burslem, D. F. R. P., Butt, N., Cao, M., Cardenas, D., … Zimmerman, J. (2015). CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Global Change Biology*, *21*(2), 528–549.

Augustine, S. P., Bailey-Marren, I., Charton, K. T., Kiel, N. G., & Peyton, M. S. (2024). Improper data practices erode the quality of global ecological databases and impede the progress of ecological research. *Global Change Biology*, *30*(1), e17116.

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., LaFrance, R., Ariño, A. H., & Guralnick, R. P. (2019). Research applications of primary biodiversity databases in the digital age. *PLoS One*, *14*(9), e0215794.

Barker, J. R., & MacIsaac, H. J. (2022). Species distribution models: Administrative boundary centroid occurrences require careful interpretation. *Ecological Modelling*, *472*, 110107.

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on earth. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6506–6511.

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, *8*(6), e1000385. https://doi.org/10.1371/journal.pbio.1000385

Bortolus, A. (2008). Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. *Ambio*, *37*(2), 114–118.

Bowler, D. E., Boyd, R. J., Callaghan, C. T., Robinson, R. A., Isaac, N. J. B., & Pocock, M. J. O. (2024). Treating gaps and biases in biodiversity data as a missing data problem. *Biological Reviews of the Cambridge Philosophical Society*, *100*, 50–67. https://doi.org/10.1111/brv.13127

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., McKay, S. J., Lowry, S., Freeland, C., Peet, R. K., & Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for automated standardization of plant names. *BMC Bioinformatics*, *14*, 16.

Boyle, B., & Maitner, B. (2021). *GNRS: Access the "Geographic Name Resolution Service"* (Version 0.3.4). https://CRAN.R-project.org/package=GNRS

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography* (2nd ed.). Sinauer Associates.

Buzzard, V., Michaletz, S. T., Deng, Y., He, Z., Ning, D., Shen, L., Tu, Q., Van Nostrand, J. D., Voordeckers, J. W., Wang, J., Weiser, M. D., Kaspari, M., Waide, R. B., Zhou, J., & Enquist, B. J. (2019). Continental scale structuring of forest and soil diversity via functional traits. *Nature Ecology & Evolution*, *3*(9), 1298–1308.

Cooper, D. L. M., Lewis, S. L., Sullivan, M. J. P., Prado, P. I., Ter Steege, H., Barbier, N., Slik, F., Sonké, B., Ewango, C. E. N., Adu-Bredu, S., Affum-Baffoe, K., de Aguiar, D. P. P., Ahuite Reategui, M. A., Aiba, S.-I., Albuquerque, B. W., de Almeida Matos, F. D., Alonso, A., Amani, C. A., do Amaral, D. D., … Zent, S. (2024). Consistent patterns of common species across tropical tree communities. *Nature*, *625*(7996), 728–734.

Dauby, G., Zaiss, R., Blach-Overgaard, A., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield, J., Droissart, V., Duarte, M. C., Engledow, H., Fadeur, G., Figueira, R., Gereau, R. E., Hardy, O. J., Harris, D. J., de Heij, J., Janssens, S., Klomberg, Y., … Couvreur, T. L. P. (2016). RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys*, *74*, 1–18.

Davis, C. C. (2023). The herbarium of the future. *Trends in Ecology & Evolution*, *38*(5), 412–423.

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*(3), 407–415.

Denelle, P., Weigelt, P., & Kreft, H. (2023). GIFT—An R package to access the global inventory of floras and traits. *Methods in Ecology and Evolution*, *14*(11), 2738–2748.

Diamond, J. M. (1975). Assembly of species communities. In M. L. Cody & J. M. Diamond (Eds.), *Ecology and evolution of communities* (pp. 342–444). Harvard University Press.

Downs, R. R., Lenhardt, W. C., Robinson, E., Davis, E., & Weber, N. (2015). Community recommendations for sustainable scientific software. *Journal of Open Research Software*, *3*(1), 11.

Echeverría-Londoño, S., Enquist, B. J., Neves, D. M., Violle, C., Boyle, B., Kraft, N. J. B., Maitner, B. S., McGill, B., Peet, R. K., Sandel, B., Smith, S. A., Svenning, J.-C., Wiser, S. K., & Kerkhoff, A. J. (2018). Plant functional diversity and the biogeography of biomes in north and South America. *Frontiers in Ecology and Evolution*, *6*, 219.

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science*, *289*(5488), 2312–2314.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57.

Engemann, K., Sandel, B., Boyle, B., Enquist, B. J., Jørgensen, P. M., Kattge, J., McGill, B. J., Morueta-Holme, N., Peet, R. K., Spencer, N. J., Violle, C., Wiser, S. K., & Svenning, J.-C. (2016). A plant growth form dataset for the New World. *Ecology*, *97*(11), 3243. https://doi.org/10.1002/ecy.1569

Enquist, B., & Boyle, B. (2012). SALVIAS—The SALVIAS vegetation inventory database. *Biodiversity & Ecology*, *4*, 288.

Enquist, B. J., Boyle, B., Maitner, B. S., Merow, C., Carlson, B. S., Feng, X., Moulatlet, G. M., Newman, E. A., Park, D. S., Pillet, M., Roehrdanz, P. R., Serra-Diaz, J. M., Babu, R. S. Y., Barbosa, G. G. C., Casler, N., Donoghue, J. C., Guaderrama, D., Sajja, R. K., Couvreur, T. L. P., … Wiser, S. K. (2026a). Summary of BIEN_4.2 Data, Tables, and Statistica (Version [v1.0]). *Zenodo*. https://doi.org/10.5281/zenodo.18391931

Doughty, C. E., Wolf, A., Morueta-Holme, N., Jørgensen, P. M., Sandel, B., Violle, C., Boyle, B., Kraft, N. J. B., Peet, R. K., Enquist, B. J., Svenning, J.-C., Blake, S., & Galetti, M. (2016). Megafauna extinction, tree species range reduction, and carbon storage in Amazonian forests. *Ecography*, *39*(2), 194–203.

Enquist, B. J., Boyle, B., Maitner, B. S., Merow, C., Carlson, B. S., Feng, X., Moulatlet, G. M., Newman, E. A., Park, D. S., Pillet, M., Roehrdanz, P. R., Serra-Diaz, J. M., Babu, R. S. Y., Barbosa, G. G. C., Casler, N., Donoghue, J. C., Guaderrama, D., Sajja, R. K., Couvreur, T. L. P., … Wiser, S. K. (2026b). The cumulative number of plant species recorded over time (Version [v1.0]). *Zenodo*. https://doi.org/10.5281/zenodo.18423054

Enquist, B. J., Boyle, B., Maitner, B. S., Merow, C., Carlson, B. S., Feng, X., Moulatlet, G. M., Newman, E. A., Park, D. S., Pillet, M., Roehrdanz, P. R., Serra-Diaz, J. M., Babu, R. S. Y., Barbosa, G. G. C., Casler, N., Donoghue, J. C., Guaderrama, D., Sajja, R. K., Couvreur, T. L. P., … Wiser, S. K. (2026c). BIEN—Cactaceae DroughtForecasts (Version [v1.0]). *Zenodo*. https://doi.org/10.5281/zenodo.18423999

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M., & Thiers, B. M. (2016). Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. PeerJ Preprints. https://peerj.com/preprints/2615.pdf

Enquist, B. J., Feng, X., Boyle, B., Maitner, B., Newman, E. A., Jørgensen, P. M., Roehrdanz, P. R., Thiers, B. M., Burger, J. R., Corlett, R. T., Couvreur, T. L. P., Dauby, G., Donoghue, J. C., Foden, W., Lovett, J. C., Marquet, P. A., Merow, C., Midgley, G., Morueta-Holme, N., … McGill, B. J. (2019). The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances*, *5*(11), eaaz0414.

Faurby, S., Eiserhardt, W. L., & Svenning, J.-C. (2016). Strong effects of variation in taxonomic opinion on diversification analyses. *Methods in Ecology and Evolution*, *7*(1), 4–13.

Feng, X., Enquist, B. J., Park, D. S., Boyle, B., Breshears, D. D., Gallagher, R. V., Lien, A., Newman, E. A., Burger, J. R., Maitner, B. S., Merow, C., Li, Y., Huynh, K. M., Ernst, K., Baldwin, E., Foden, W., Hannah, L., Jørgensen, P. M., Kraft, N. J. B., … Hurlbert, A. (2022). A review of the heterogeneous landscape of biodiversity databases: Opportunities and challenges for a synthesized biodiversity knowledge base. *Global Ecology and Biogeography: A Journal of Macroecology*, *31*, 1242–1260. https://doi.org/10.1111/geb.13497

Feng, X., Merow, C., Liu, Z., Park, D. S., Roehrdanz, P. R., Maitner, B., Newman, E. A., Boyle, B. L., Lien, A., Burger, J. R., Pires, M. M., Brando, P. M., Bush, M. B., McMichael, C. N. H., Neves, D. M., Nikolopoulos, E. I., Saleska, S. R., Hannah, L., Breshears, D. D., … Enquist, B. J. (2021). How deregulation, drought and increasing fire impact Amazonian biodiversity. *Nature*, *597*(7877), 516–521. https://doi.org/10.1038/s41586-021-03876-7

Feng, X., Rocha, T., Thammavong, H. T., Tulaiha, R., Chen, X., Xie, Y., & Park, D. S. (2024). GridDER: Grid detection and evaluation in R. *Ecological Informatics*, *79*, 102391.

Feng, X., Smith, A., Boyle, B., Chen, X., Enquist, B., Gallagher, R., Hammock, J., Ho, J., Lien, A., Maitner, B., Sokol, E., Soltis, P., Wenk, E., Willoughby, A., & Park, D. (2025). The next stage of biodiversity informatics: Community-driven synthesis and integration of biodiversity databases. *Bioscience*, *75*, 913–925. https://doi.org/10.1093/biosci/biaf112

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, *51*(2), 331–363.

Flower, C., Hodgson, W. C., Salywon, A. M., Maitner, B. S., Enquist, B. J., Peeples, M. A., & Blonder, B. (2021). Human food use increases plant geographical ranges in the Sonoran Desert. *Global Ecology and Biogeography*, *30*(7), 1461–1473.

ForestPlots.net, Blundo, C., Carilla, J., Grau, R., Malizia, A., Malizia, L., Osinaga-Acosta, O., Bird, M., Bradford, M., Catchpole, D., Ford, A., Graham, A., Hilbert, D., Kemp, J., Laurance, S., Laurance, W., Ishida, F. Y., Marshall, A., Waite, C., … Tran, H. D. (2021). Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation*, *260*, 108849.

Franz, N., Peet, R., & Weakley, A. (2006). 5 on the use of taxonomic concepts in support of biodiversity research and taxonomy. North, 61–84.

Franz, N. M., & Peet, R. K. (2009). Perspectives: Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity*, *7*(1), 5–20.

Funk, V. A., & Richardson, K. S. (2002). Systematic data in biodiversity studies: Use it or lose it. *Systematic Biology*, *51*(2), 303–316.

Gallagher, R. V., Allen, S. P., Govaerts, R., Rivers, M. C., Allen, A. P., Keith, D. A., Merow, C., Maitner, B., Butt, N., Auld, T. D., Enquist, B. J., Eiserhardt, W. L., Wright, I. J., Mifsud, J. C. O., Espinosa-Ruiz, S., Possingham, H., & Adams, V. M. (2023). Global shortfalls in threat assessments for endemic flora by country. *Plants, People, Planet*, *5*, 885–898. https://doi.org/10.1002/ppp3.10369

Gallagher, R. V., Falster, D. S., Maitner, B. S., Salguero-Gómez, R., Vandvik, V., Pearse, W. D., Schneider, F. D., Kattge, J., Poelen, J. H., Madin, J. S., Ankenbrand, M. J., Penone, C., Feng, X., Adams, V. M., Alroy, J., Andrew, S. C., Balk, M. A., Bland, L. M., Boyle, B. L., … Enquist, B. J. (2020). Open Science principles for accelerating trait-based science across the tree of life. *Nature Ecology & Evolution*, *4*(3), 294–303.

Goldsmith, G. R., Morueta-Holme, N., Sandel, B., Fitz, E. D., Fitz, S. D., Boyle, B., Casler, N., Engemann, K., Jørgensen, P. M., Kraft, N. J. B., McGill, B., Peet, R. K., Piel, W. H., Spencer, N., Svenning, J. C., Thiers, B. M., Violle, C., Wiser, S. K., & Enquist, B. J. (2016). Plant-O-Matic: a dynamic and mobile guide to all plants of the Americas. *Methods in Ecology and Evolution / British Ecological Society*, *7*(8), 960–965.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, *8*(1), 215.

Gray, A., Brandeis, T., Shaw, J., McWilliams, W., & Miles, P. (2012). Forest inventory and analysis database of The United States of America (FIA). *Biodiversity & Ecology*, *4*, 225–231.

Griffith, J., Lord, J.-M., Catchen, M., Arce-Plata, M., Galvez Bohorquez, M., Chandramohan, M., Diaz-Corzo, M. C., Gravel, D., Urbina Gonzalez, L. F., Gutiérrez, C., Helfenstein, I., Hoban, S., Kass, J., Laroque, G., Laikre, L., Leigh, D., Leung, B., Mastretta-Yanes, A., Millette, K., … Gonzalez, A. (2024). BON in a Box: An open and collaborative platform for biodiversity monitoring, indicator calculation, and reporting. EcoEvoRxiv. https://doi.org/10.32942/x2m320

Guralnick, R., & Van Cleve, J. (2005). Strengths and weaknesses of museum and national survey data sets for predicting regional species richness: Comparative and combined approaches. *Diversity and Distributions*, *11*(4), 349–359.

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., Michener, W. K., Mudge, J., Pourmokhtarian, A., Schildhauer, M. P., Woo, K. H., & Zimmerman, N. (2015). The Tao of open science for ecology. *Ecosphere*, *6*(7), art120.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162. https://doi.org/10.1890/120103

Hannah, L., Roehrdanz, P. R., Marquet, P. A., Enquist, B. J., Midgley, G., Foden, W., Lovett, J. C., Corlett, R. T., Corcoran, D., Butchart, S. H. M., Boyle, B., Feng, X., Maitner, B., Fajardo, J., McGill, B. J., Merow, C., Morueta-Holme, N., Newman, E. A., Park, D. S., & Svenning, J. (2020). 30% land conservation and climate action reduces tropical extinction risk by more than 50%. *Ecography*, *12*, 327.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

Heberling, J. M., Prather, L. A., & Tonsor, S. J. (2019). The changing uses of herbarium data in an era of global change: An overview using automated content analysis. *Bioscience*, *69*(10), 812–822.

Heinlein, R. A. (1966). *The moon is a harsh mistress*. G P Putnam's Sons.

Hinsen, K. (2019). Dealing with software collapse. *Computing in Science & Engineering*, *21*(3), 104–108.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015a). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015b). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), 539–551.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448.

Johnston, A., Fink, D., Reynolds, M. D., Hochachka, W. M., Sullivan, B. L., Bruns, N. E., Hallstein, E., Merrifield, M. S., Matsumoto, S., & Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications: A Publication of the Ecological Society of America*, *25*(7), 1749–1756.

Jung, M., Arnell, A., de Lamo, X., García-Rangel, S., Lewis, M., Mark, J., Merow, C., Miles, L., Ondo, I., Pironon, S., Ravilious, C., Rivers, M., Schepaschenko, D., Tallowin, O., van Soesbergen, A., Govaerts, R., Boyle, B. L., Enquist, B. J., Feng, X., & Visconti, P. (2021). Areas of global importance for conserving terrestrial biodiversity, carbon and water. *Nature Ecology & Evolution*, *5*(11), 1499–1509.

Kass, J. M., Pinilla-Buitrago, G. E., Paz, A., Johnson, B. A., Grisales-Betancur, V., Meenan, S. I., Attali, D., Broennimann, O., Galante, P. J., Maitner, B. S., Owens, H. L., Varela, S., Aiello-Lammens, M. E., Merow, C., Blair, M. E., & Anderson, R. P. (2023). *Wallace* 2: A shiny app for modeling species niches and distributions redesigned to facilitate expansion via module contributions. *Ecography*, *2023*, e06547. https://doi.org/10.1111/ecog.06547

Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C., & Anderson, R. P. (2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution*, *9*(4), 1151–1156.

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C, C., Aleixo, I., Ali, H., … Wirth, C. (2020). TRY plant trait database—enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., Cornelissen, J. H. C., Violle, C., Harrison, S. P., Van Bodegom, P. M., Reichstein, M., Enquist, B. J., Soudzilovskaia, N. a., Ackerly, D. D., Anand, M., … Wirth, C. (2011). TRY—A global database of plant traits. *Global Change Biology*, *17*(9), 2905–2935.

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., & Kreft, H. (2019). Biodiversity data integration-the significance of data resolution and domain. *PLoS Biology*, *17*(3), e3000183.

Kusumoto, B., Chao, A., Eiserhardt, W. L., Svenning, J.-C., Shiono, T., & Kubota, Y. (2023). Occurrence-based diversity estimation reveals macroecological and conservation knowledge gaps for global woody plants. *Science Advances*, *9*(40), eadh9719.

Lawton, J. H. (1993). Range, population abundance and conservation. *Trends in Ecology & Evolution*, *8*(11), 409–413.

Lessa, T., Stropp, J., Hortal, J., & Ladle, R. J. (2024). How taxonomic change influences forecasts of the Linnean shortfall (and what we can do about it)? *Journal of Biogeography*, *51*(8), 1365–1373.

Li, Y., Schuldt, A., Ebeling, A., Eisenhauer, N., Huang, Y., Albert, G., Albracht, C., Amyntas, A., Bonkowski, M., Bruelheide, H., Bröcher, M., Chesters, D., Chen, J., Chen, Y., Chen, J.-T., Ciobanu, M., Deng, X., Fornoff, F., Gleixner, G., … Liu, X. (2024). Plant diversity enhances ecosystem multifunctionality via multitrophic diversity. *Nature Ecology & Evolution*, *8*(11), 2037–2047.

Lien, A. M., Banki, O., Barik, S. K., Buckeridge, J. S., Christidis, L., Cigliano, M. M., Conix, S., Costello, M. J., Hobern, D., Kirk, P. M., Kroh, A., Montgomery, N., Nikolaeva, S., Orrell, T. M., Pyle, R. L., Raz, L., Thiele, K., Thomson, S. A., van Dijk, P. P., … Garnett, S. T. (2023). Widespread support for a global species list with a formal governance system. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(45), e2306899120.

Lu, M., & He, F. (2017). Estimating regional species richness: The case of China's vascular plant species. *Global Ecology and Biogeography*, *26*(7), 835–845.

Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1444), 711–719.

Maitner, B. (2023). *BIEN*: *Tools for accessing the botanical information and ecology network database* (Version 1.2.6). https://CRAN.R-project.org/package=BIEN

Maitner, B., & Boyle, B. (2024). *NSR*: *"Native Status Resolver"* (Version 0.1.0). https://CRAN.R-project.org/package=NSR

Maitner, B., Gallagher, R., Svenning, J.-C., Tietje, M., Wenk, E. H., & Eiserhardt, W. L. (2023). A global assessment of the Raunkiaeran shortfall in plants: Geographic biases in our knowledge of plant traits. *The New Phytologist*, *240*, 1345–1354. https://doi.org/10.1111/nph.18999

Maitner, B., Merow, C., Boyle, B., Feng, X., Babu, R. S. Y., & Enquist, B. (2025). *OpenRange*: *Code to access open access species range maps* (Version 0.0.1). https://CRAN.R-project.org/package=OpenRange

Maitner, B. S., Boyle, B., Casler, N., Condit, R., Donoghue, J., Durán, S. M., Guaderrama, D., Hinchliff, C. E., Jørgensen, P. M., Kraft, N. J. B., McGill, B., Merow, C., Morueta-Holme, N., Peet, R. K., Sandel, B., Schildhauer, M., Smith, S. A., Svenning, J.-C., Thiers, B., … Enquist, B. J. (2017). The BIEN R package: A tool to access the botanical information and ecology network (BIEN) database. *Methods in Ecology and Evolution/British Ecological Society*, *9*(2), 373–379.

Manikas, K. (2016). Revisiting software ecosystems research: A longitudinal literature study. *Journal of Systems and Software*, *117*, 84–103.

Marquet, P. A., Allen, A. P., Brown, J. H., Dunne, J. A., Enquist, B. J., Gillooly, J. F., Gowaty, P. A., Green, J. L., Harte, J., Hubbell, S. P., O'Dwyer, J., Okie, J. G., Ostling, A., Ritchie, M., Storch, D., & West, G. B. (2014). On theory in ecology. *Bioscience*, *64*(8), 701–710. https://doi.org/10.1093/biosci/biu098

McFadden, I. R., Sandel, B., Tsirogiannis, C., Morueta-Holme, N., Svenning, J.-C., Enquist, B. J., & Kraft, N. J. B. (2019). Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity. *Ecology Letters*, *22*, 1126–1135. https://doi.org/10.1111/ele.13269

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, *19*(8), 992–1006.

Mikryukov, V., Dulya, O., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., Prylutskyi, O., Delgado-Baquerizo, M., Maestre, F. T., Nilsson, H., Pärn, J., Öpik, M., Moora, M., Zobel, M., Espenberg, M., Mander, Ü., Khalid, A. N., Corrales, A., Agan, A., … Tedersoo, L. (2023). Connecting the multiple dimensions of global soil fungal diversity. *Science Advances*, *9*(48), eadj8016.

Morueta-Holme, N., Enquist, B. J., McGill, B. J., Boyle, B., Jørgensen, P. M., Ott, J. E., Peet, R. K., Símová, I., Sloat, L. L., Thiers, B., Violle, C., Wiser, S. K., Dolins, S., Donoghue, J. C., Kraft, N. J. B., Regetz, J., Schildhauer, M., Spencer, N., & Svenning, J.-C. (2013). Habitat area and climate stability determine geographical variation in plant species range sizes. *Ecology Letters*, *16*(12), 1446–1454.

Morueta-Holme, N., & Svenning, J.-C. (2018). Geography of plants in the new world: Humboldt's relevance in the age of big data. *Annals of the Missouri Botanical Garden*, *103*(3), 315–329.

Moulatlet, G. M., Merow, C., Maitner, B., Boyle, B., Feng, X., Frazier, A. E., Hinojo-Hinojo, C., Newman, E. A., Roehrdanz, P. R., Song, L., Villalobos, F., Marquet, P. A., Svenning, J.-C., & Enquist, B. J. (2025). General Laws of biodiversity: Climatic niches predict plant range size and ecological dominance globally. *Proceedings of the National Academy of Sciences*, *122*, e2517585122. https://doi.org/10.1073/pnas.2517585122

Moura, M. R., & Jetz, W. (2021). Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nature Ecology & Evolution*, *5*(5), 631–639.

Neves, D. M., Kerkhoff, A. J., Echeverría-Londoño, S., Merow, C., Morueta-Holme, N., Peet, R. K., Sandel, B., Svenning, J.-C., Wiser, S. K., & Enquist, B. J. (2021). The adaptive challenge of extreme conditions shapes evolutionary diversity of plant assemblages at continental scales. *Proceedings of the National Academy of Sciences*, *118*(37), e2021132118.

Nic Lughadha, E., Walker, B. E., Canteiro, C., Chadburn, H., Davis, A. P., Hargreaves, S., Lucas, E. J., Schuiteman, A., Williams, E., Bachman, S. P., Baines, D., Barker, A., Budden, A. P., Carretero, J., Clarkson, J. J., Roberts, A., & Rivers, M. C. (2018). The use and misuse of herbarium specimens in evaluating plant extinction risks. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1763), 20170402.

Nielsen, M. (2020). *Reinventing discovery* (2nd ed.). Princeton University Press.

Olesen, J. M., & Jordano, P. (2002). Geographic patterns in plant-pollinator mutualistic networks. *Ecology*, *83*(9), 2416.

Olff, H., Ritchie, M. E., & Prins, H. H. T. (2002). Global environmental controls of diversity in large herbivores. *Nature*, *415*(6874), 901–904.

Owens, H. L., Merow, C., Maitner, B. S., Kass, J. M., Barve, V., & Guralnick, R. P. (2021). occCite: Tools for querying and managing large biodiversity occurrence datasets. *Ecography*, *44*(8), 1228–1235.

Panter, C. T., Clegg, R. L., Moat, J., Bachman, S. P., Klitgård, B. B., & White, R. L. (2020). To clean or not to clean: Cleaning open-source data improves extinction risk assessments for threatened plant species. *Conservation Science and Practice*, *2*(12), e311. https://doi.org/10.1111/csp2.311

Park, D. S., & Davis, C. C. (2017). Implications and alternatives of assigning climate data to geographical centroids. *Journal of Biogeography*, *44*(10), 2188–2198.

Park, D. S., Feng, X., Akiyama, S., Ardiyani, M., Avendaño, N., Barina, Z., Bärtschi, B., Belgrano, M., Betancur, J., Bijmoer, R., Bogaerts, A., Cano, A., Danihelka, J., Garg, A., Giblin, D. E., Gogoi, R., Guggisberg, A., Hyvärinen, M., James, S. A., … Davis, C. C. (2023). The colonial legacy of herbaria. *Nature Human Behaviour*, *7*(7), 1059–1068.

Park, D. S., Xie, Y., Thammavong, H. T., Tulaiha, R., & Feng, X. (2023). Artificial hotspot occurrence inventory (AHOI). *Journal of Biogeography*, *50*(2), 441–449.

Peet, R., Lee, M., Boyle, F., Wentworth, T., Schafale, M., & Weakley, A. (2012). Vegetation-plot database of the Carolina vegetation survey. *Biodiversity & Ecology*, *4*, 243–253.

Peet, R. K., Lee, M. T., Jennings, M. D., & Faber-Langendoen, D. (2012). VegBank: a permanent, open-access archive for vegetation plot data. *Biodiversity and Ecology = Biodiversitat Und Okologie*, *4*, 233–241.

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*(6060), 1226–1227.

Pettorelli, N., Barlow, J., Nuñez, M. A., Rader, R., Stephens, P. A., Pinfield, T., & Newton, E. (2021). How international journals can support ecology from the global south. *Journal of Applied Ecology*, *58*(1), 4–8.

Pillet, M., Goettsch, B., Merow, C., Maitner, B., Feng, X., Roehrdanz, P. R., & Enquist, B. J. (2022). Elevated extinction risk of cacti under climate change. *Nature Plants*, *8*(4), 366–372.

Pironon, S., Ondo, I., Diazgranados, M., Allkin, R., Baquero, A. C., Cámara-Leret, R., Canteiro, C., Dennehy-Carr, Z., Govaerts, R., Hargreaves, S., Hudson, A. J., Lemmens, R., Milliken, W., Nesbitt, M., Patmore, K., Schmelzer, G., Turner, R. M., van Andel, T. R., Ulian, T., & Willis, K. J. (2024). The global distribution of plants used by humans. *Science*, *383*(6680), 293–297.

Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, *24*, 148–159.

Postgis Project Steering Committee, & Others. (2018). *PostGIS*, *spatial and geographic objects for postgreSQL*. https://postgis.net

PostgreSQL Global Development Group. (2024). *PostgreSQL* (Version 15.8). https://www.postgresql.org/

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Reyes-García, V., Cámara-Leret, R., Halpern, B. S., O'Hara, C., Renard, D., Zafra-Calvo, N., & Díaz, S. (2023). Biocultural vulnerability exposes threats of culturally important species. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(2), e2217303120.

Ronquillo, C., Stropp, J., Medina, N. G., & Hortal, J. (2023). Exploring the impact of data curation criteria on the observed geographical distribution of mosses. *Ecology and Evolution*, *13*(12), e10786.

Rovero, F., & Ahumada, J. (2017). The tropical ecology, assessment and monitoring (TEAM) network: An early warning system for tropical rain forests. *Science of the Total Environment*, *574*, 914–923.

Rushton, S. P., Ormerod, S. J., & Kerby, G. (2004). New paradigms for modelling species distributions? *The Journal of Applied Ecology*, *41*(2), 193–200.

Sabatini, F. M., Lenoir, J., Hattab, T., Arnst, E. A., Chytrý, M., Dengler, J., De Ruffray, P., Hennekens, S. M., Jandt, U., Jansen, F., Jiménez-Alfaro, B., Kattge, J., Levesley, A., Pillar, V. D., Purschke, O., Sandel, B., Sultana, F., Aavik, T., Aćić, S., … Bruelheide, H. (2021). sPlotOpen—An environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography: A Journal of Macroecology*, *30*(9), 1740–1764.

Salguero-Gómez, R., Jackson, J., & Gascoigne, S. J. L. (2021). Four key challenges in the open-data revolution. *The Journal of Animal Ecology*, *90*(9), 2000–2004.

Schimper, A. F. W. (1903). *Plant geography upon a physiological basis*.

Schuldt, A., Ebeling, A., Kunz, M., Staab, M., Guimarães-Steinicke, C., Bachmann, D., Buchmann, N., Durka, W., Fichtner, A., Fornoff, F., Härdtle, W., Hertzog, L. R., Klein, A.-M., Roscher, C., Schaller, J., von Oheimb, G., Weigelt, A., Weisser, W., Wirth, C., … Eisenhauer, N. (2019). Multiple plant diversity components drive consumer communities across ecosystems. *Nature Communications*, *10*(1), 1460.

Serra-Diaz, J. M., Borderieux, J., Maitner, B., Boonman, C. C. F., Park, D., Guo, W.-Y., Callebaut, A., Enquist, B. J., Svenning, J.-C., & Merow, C. (2024). occTest: An integrated approach for quality control of species occurrence data. *Global Ecology and Biogeography: A Journal of Macroecology*, *33*(7), e13847. https://doi.org/10.1111/geb.13847

Soberón, J., & Peterson, A. T. (2004). Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1444), 689–698.

Soley-Guardia, M., Alvarado-Serrano, D. F., & Anderson, R. P. (2024). Top ten hazards to avoid when modeling species distributions: a didactic guide of assumptions, problems, and recommendations. *Ecography*, *2024*(4), e06852. https://doi.org/10.1111/ecog.06852

Soltis, P. S. (2017). Digitization of herbaria enables novel research. *American Journal of Botany*, *104*(9), 1281–1284.

Stropp, J., Pereira, A. S. S., Emilio, T., Meyer, L., Trad, R. J., Alves-Martins, F., Ladle, R. J., & Hortal, J. (2025). The impact of taxonomic change on the Amazonian palm flora. *Proceedings. Biological sciences*, *292*(2050), 20242738.

Toussaint, A., Bueno, G., Davison, J., Moora, M., Tedersoo, L., Zobel, M., Öpik, M., & Pärtel, M. (2020). Asymmetric patterns of global diversity among plants and mycorrhizal fungi. *Journal of Vegetation Science: Official Organ of the International Association for Vegetation Science*, *31*(2), 355–366.

UNESCO International Union for Conservation of Nature. (2023). *World heritage: a unique contribution to biodiversity conservation*. International Union for Conservation of Nature.

Weigelt, P., König, C., & Kreft, H. (2020). GIFT—A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography*, *47*(1), 16–43.

Weiher, E., & Keddy, P. (1999). Assembly rules as general constraints on community composition. In *Ecological assembly rules: Perspectives, advances, retreats* (pp. 251–271). Cambridge University Press.

Weiher, E., & Keddy, P. (2001). *Ecological assembly rules: Perspectives, advances, retreats*. University Press.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1. Overview of BIEN's principal applications and their corresponding R packages or core repositories.

Table S2. BIEN technical infrastructure: service layers and components.

Table S3. TNRS Data Dictionary.

Table S4. GNRS Data Dictionary.

Table S5. NSR Data Dictionary.

Table S6. GVS Data Dictionary.

Table S7. Trait names within the BIEN db and the number of trait observations associated with each trait.

Table S8. A listing of 19 published research studies that have used RBIEN and their associated links to open source code.

Figure S1. BIEN global plot and survey sampling intensity maps.

Appendix 1. R Code to generate BIEN annotation statistics.

Appendix 2. BIEN Trait Observation References.

How to cite this article: Enquist, B. J., Boyle, B., Maitner, B. S., Merow, C., Carlson, B. S., Feng, X., Moulatlet, G. M., Newman, E. A., Park, D. S., Pillet, M., Roehrdanz, P. R., Serra-Diaz, J. M., Babu, R. S. Y., Barbosa, G. G. C., Casler, N., Donoghue, J. C. II, Guaderrama, D., Sajja, R. K., Couvreur, T. L. P., … Wiser, S. K. (2026). BIEN: A biodiversity informatics ecosystem advancing open and reproducible workflows for plant observation, plot and trait data. *Methods in Ecology and Evolution*, *00*, 1–29. https://doi.org/10.1111/2041-210x.70274