Multimodal data integration to model, predict, and understand changes in plant biodiversity: a systematic review

EMILCE SOLEDAD MARTINEZ[1,2], EVA TEJADA-GUTIÉRREZ[1], ALBERT SORRIBAS[1], JORDI MATEO-FORNES[3], FRANCESC SOLSONA[3], RAQUEL DEFACIO[2], RUI ALVES[1,*]

1 Departament de Ciencies Mèdiques Bàsiques, Universitat de Lleida & IRBLleida, Lleida, Spain.

2 Instituto Nacional de Tecnología Agropecuaria, Banco Activo de Germoplasma, Pergamino, Argentina.

3 Departament de d'Enginyeria Informàtica i Disseny Digital, Universitat de Lleida & INSPIRES, Lleida, Spain.

* Corresponding author e-mail: rui.alves@udl.cat

Running head: Multimodal data integration in biodiversity

Abstract

The integration of multimodal data to analyze, model, and predict changes in plant biodiversity is critical for addressing global conservation challenges. This systematic review examines the current landscape of plant biodiversity data, focusing on the identification, classification, and evaluation of key open-access data sources and integration methodologies. We highlight the strengths and limitations of major biodiversity platforms, emphasizing their contributions to species occurrence, trait data, taxonomic checklists, and environmental variables. The review also explores computational approaches for multimodal data integration, including metadata standardization frameworks (e.g., Darwin Core) and advanced modeling techniques like Species Distribution Models (SDMs) and deep learning methods. Additionally, we assess environmental drivers of biodiversity change—such as climate variables, land-use patterns, and elevation gradients—and the tools available for their analysis. Despite significant advancements in biodiversity informatics, challenges persist in achieving interoperability across datasets, addressing spatial and temporal biases, and integrating remote sensing with *in situ* observations. By identifying gaps in current methodologies and proposing pathways for improvement, this review contributes to advancing biodiversity monitoring strategies. These efforts align with global conservation goals outlined by the Convention on Biological Diversity and the United Nations Sustainable Development Goal 15. Ultimately, the findings underscore the importance of harmonized data integration frameworks to enhance predictive modeling capabilities and inform effective conservation policies.

Keywords: Biodiversity, Multimodal data, Databases, Data integration, Big data, Smart data, Mathematical ecology

1 Introduction

Biodiversity includes the full spectrum of life on Earth, encompassing genetic, species, and ecosystem diversity (Convention on Biological Diversity, 2011). Its conservation is crucial for sustaining ecosystems, ensuring food security, and protecting human health. Forests are vital to global biodiversity, serving as habitats that support diverse life forms (Pontius *et al.*, 2020). However, recent studies highlight an imbalance in research focus, with greater attention given to animal diversity while plant biodiversity hotspots in developing countries remain understudied (Wani *et al.*, 2024). This knowledge gap, along with the threats facing plant biodiversity, hampers the development of effective conservation strategies, particularly in the face of climate change and human activity (FAO, 2022).

High-quality, comprehensive global plant biodiversity data are essential for developing adequate biodiversity conservation strategies. Although plant biodiversity data have been collected for over a century (Tejada-Gutierrez *et al.*, 2024), satellite technologies have only recently enabled large-scale monitoring (Chen *et al.*, 2021; Cohen *et al.*, 2020; Khare *et al.*, 2021; Negassa *et al.*, 2020; Xu *et al.*, 2022). Many of these datasets, now openly accessible through online platforms (Saran *et al.*, 2022), provide diverse information, including plant phenotypes, geodiversity, climate variables, and molecular sequences (Enquist *et al.*, 2016; Fernández *et al.*, 2020; Keppel *et al.*, 2021). These data are generated using multimodal acquisition methods, ranging from *in situ* measurements to remote sensing. Yet, their heterogeneity and diverse formats complicate access, integration, and interpretation (Williamson & Leonelli, 2023).

Effective multimodal data integration enhances biodiversity change detection, refines existing databases, and informs conservation strategies to mitigate biodiversity loss (Cai *et al.*, 2023;

Wäldchen & Mäder, 2017). Computational approaches are poised to play an increasingly important role in that integration, improving data usability for biodiversity monitoring and conservation (Ganjdanesh *et al.*, 2022; Saran *et al.*, 2022; Schellenberger Costa *et al.*, 2023).

This paper presents a comprehensive review of the current state of plant biodiversity data, emphasizing the identification, classification, and critical evaluation of major data sources and integration methodologies employed to study spatial and temporal patterns of plant biodiversity. We pay special attention to open-access tools and resources that support biodiversity research, with an analysis of their strengths and limitations to identify areas requiring further development. By advancing strategies for biodiversity monitoring, this work contributes to the goals of the Convention on Biological Diversity (Pörtner *et al.*, 2021; Robinson, 2020; United Nations, 1992) and supports the achievement of the United Nations Sustainable Development Goal 15 (United Nations, 2018).

2 Selection of web platforms

As a first step to identify plant biodiversity platforms, we searched Web of Science, Scopus, and Google Scholar using combination of the keywords biodiversity, databases, data, platforms, and informatics. We applied specific filters in each portal and refined the results using additional criteria (see Supplementary Table S1 for search strategy details). From this process, we selected 21 review articles, including also a recent publication describing the ForestForward platform (Tejada-Gutiérrez *et al.*, 2024), developed in our research group and directly aligned with the objectives of this study. To identify and organize the biodiversity databases (DBs) mentioned in those articles, we used Google's NotebookLM (2024). We identified 107 web platforms containing biodiversity data

(Supplementary Table S2). Figure 1 summarizes the protocol and criteria we used to select the platforms to include in our analysis, which are described below.

First, we applied a selection criterion based on taxonomic coverage. We selected only platforms with information across all biological kingdoms or fully covering the *Plantae* kingdom. Second, we applied a geographic resolution criterion, selecting only platforms containing DBs with global coverage, and excluding DBs limited to a specific region or country. Third, we included only platforms cited in more than one article. Fourth, we prioritized DBs focused on species occurrence, taxonomy, and traits, unless they were recent (less than one year), contained valuable data on species occurrence, taxonomy, traits, or employed innovative approaches, such as time series analysis. Fifth, we revised accessibility of the platforms that passed the first four filters (n= 13, marked in Supplementary Table S1), to ensure that the website was active at the time of access (May 2025). After applying all the selection criteria, we include 12 DBs in our sample (see Supplementary Table S2).

3 Plant Biodiversity platforms

In this section, we review the selected biodiversity data portals, and the information they contain. We also briefly describe how each portal integrates data and infrastructure provided by the other portals. Table 1 provides a summary of the data available in each portal, while Figure 2 summarizes the interactions and dependencies between portals.

The Global Biodiversity Information Facility (GBIF) is the leading portal for biodiversity and climate change research (Heberling *et al.*, 2021; Saran *et al.*, 2022) and provides approximately 538 million plant occurrence species records, organized in thousands of individual datasets (accessed on: May 2025). Temporal coverage is from 1600s to the present. GBIF integrates data from multiple sources such as Tropicos and World Flora Online (WFO), and supports portals such as Map of Life (MOL), Catalogue of Life (CoL), and Encyclopedia of Life (EOL) through open standards such as the Darwin Core Standard (DwC) (Wieczorek *et al.*, 2012). The individual datasets in GBIF lack uniformity and more than half of the records are incomplete, for example lacking information about either genus/species, date of registration, or location (Tejada-Gutierrez *et al.*, 2024).

The Yale Center for Biodiversity and Global Change hosts the MOL portal, which provides 44 plant datasets (accessed on: May 2025). The interface of the portal organizes data by species richness, rarity, and includes alien species checklist for some plant species, with temporal coverage from the 1500s to present. MOL collaborates with the International Union for Conservation of Nature Red List of threatened Species, Global Register of introduced and invasive species (GRIIS), and GBIF (Jetz *et al.*, 2012).

The CoL portal provides a global taxonomic checklist (TC) with accepted names and synonyms, hosting approximately 2 million living and 147,000 extinct species (accessed on: May 2025). Temporal coverage varies by datasets, but species names are kept up to date. CoL is an essential source that was designed specifically to link global taxonomic information, without considering how to integrate geographic technologies. Its data is not organized to enable spatial analyses of biodiversity (Saran *et al.*, 2022). CoL integrates taxonomy data from the Plants of the World Flora Online (POWO) and WFO (Bánki *et al.*, 2018). It also provides access to additional information such as GBIF distribution data, literature (Biodiversity Heritage Library- BHL), and trait data (Hobern *et al.*, 2021).

BioTIME provides time series of species abundances organized by biomes. BioTIME provides over 12 million records on a global scale (accessed on: May 2025), with a temporal coverage from 1800s to 2016. BioTIME principally integrates data from GBIF, Ocean Biodiversity Information System (OBIS), and Ecological Data Wiki (Dornelas *et al.*, 2018).

WFO provides approximately 1.6 million plant names, covering a temporal scale from 1700s to present (accessed on: May 2025). WFO's stated goal is to support Target 1 of the Convention on Biological Diversity (2012). WFO collaborates closely with GBIF, CoL, POWO, and Taxonomic Expert Networks (Borsch *et al.*, 2020).

Plants of the World Online (POWO) is hosted by the Royal Botanic Gardens Kew and provides approximately 1.4 million plant names (accessed on: May 2025), with a temporal coverage encompassing the past 250 years (Saran *et al.*, 2022; Zuanny *et al.*, 2024). POWO updates its data weekly, integrating data from World Checklist of Vascular Plants (WCVP), GBIF, CoL, and WFO.

The Missouri Botanical Garden hosts Tropicos, which provides approximately 1.33 million scientific names (accessed on: May 2025), with a temporal coverage from 1700s to present. Specimen data is accessible through the GBIF database (Teisher & Stimmel, 2024).

The ForestForward portal provides approximately 172 million records, with a temporal coverage from 1600s to the present (access on: May 2025). ForestForward integrates data from GBIF, using a semi-automated Extract, Transform, and Load (ETL) process to curate data. This ensures data quality for spatial-temporal analysis of plant biodiversity (Tejada-Gutierrez *et al.*, 2024).

The Smithsonian Institution of National Museum of Natural History hosts EOL, which provides trait data on animals, plants, fungi, protists, and bacteria of approximately 2 million species, covering a temporal scale from 1600s to present (Baker *et al.*, 2014; Parr *et al.*, 2014) (accessed on: May 2025). Plant traits include ecological adaptations, propagations methods, multimedia information, morphological and physiological characteristics. EOL collaborates with portals such as BHL, Barcode of Life, CoL, and GBIF. The portal depends on GBIF for occurrence data (Saran *et al.*, 2022).

The Max Planck Institute for Biogeochemistry in Jena hosts the Plant Trait Database (TRY), which provides approximately 15 million trait records (accessed on: May 2025), including morphological, physiological and phenological traits (Kattge *et al.*, 2011, 2020). TRY integrates data from sources such as Life History Traits of the Northwest European Flora (LEDA), Global Plant Trait Network (GlopNet), Trait Database of the German Flora (BiolFlor), and collaborates with The Global Inventory of Floras and Traits (GIFT) and the Botanical Information and Ecology Network (BIEN) (Kattge *et al.*, 2020).

The GIFT is a global repository of plants with approximately 360,000 species across 3,627 geographic regions (accessed on: May 2025). Trait data include information about morphology, phenology, and physiology. GIFT collaborates with Global Naturalized Alien Flora (GloNAF), TRY, and MOL (Weigelt *et al.*, 2019).

The BIEN portal provides approximately 81 million records, 900,000 of which contain trait data about morphological characteristics (Maitner *et al.*, 2017). The last DBs version, BIEN 4, expanded ecological plot observations by 23% (accessed on: May 2025). BIEN 4 integrates datasets from sources such as GBIF, The Royal Botanical Garden of Sydney, and NeoTropTree (Enquist *et al.*, 2016).

4 Biodiversity data integration: approaches and challenges

The vast and heterogeneous nature of biodiversity datasets complicates their integration across portals and computational systems. Estimates suggest that data preparation consumes approximately 80% of research time due to these complexities (Huber *et al.*, 2021). This underscores the need for efficient computational methods for multimodal data integration (Ganjdanesh *et al.*, 2022). This section examines the primary strategies used by biodiversity portals to integrate diverse data types from heterogeneous sources.

4.1 Metadata standardization

Biodiversity databases employ a limited but interconnected set of approaches to standardize and integrate data, with metadata standardization playing a critical role. Most portals adopt the Darwin Core (DwC) standards, developed by the Biodiversity Information Standards (TDWG), to ensure consistency in key data fields such as species names, locations, and observation dates. This standardization facilitates seamless data exchange across portals (Borsch *et al.*, 2020; Enquist *et al.*, 2016; Govaerts *et al.*, 2021; Hobern *et al.*, 2021; Parr *et al.*, 2016; Weigelt *et al.*, 2019; Wieczorek *et al.*, 2012).

Some portals also support alternative metadata standards, including CoL Data Package (COLDP), Annual Checklist Exchange Format (ACEF), Ecological Metadata Language (EML), VegX, and Humboldt Core (Döring & Ower, 2019; Hobern *et al.*, 2021; Guralnick *et al.*, 2018; Jones *et al.* 2019; Jetz *et al.*, 2012; Weigelt *et al.*, 2019; Wieczorek *et al.*, 2012) (Supplementary Table S3). Additionally, stable unique identifiers—such as Digital Object Identifiers (DOIs) and WFO IDs—link species occurrences to images, morphological data, and passport information (Döring & Ower, 2019; Hobern *et al.*, 2021; Palese *et al.*, 2019; Steidinger *et al.*, 2019).

Despite these efforts, integration workflows vary across portals, leading to inconsistencies, data duplication, and omissions. These challenges hinder the incorporation of datasets, particularly those containing images, species descriptors, or specific trait data (Reichman *et al.*, 2011; Saran *et al.*, 2022). Although workflow standardization and data interchange are improving, inconsistencies in data quality and structure remain significant obstacles.

4.2 Primary data integration

Scientific names serve as the primary key for integrating biodiversity data. However, taxonomic nomenclature changes over time, leading to issues such as homonyms, synonyms, and ambiguous spellings (Boyle *et al.*, 2013; Feng *et al.*, 2022; Franz *et al.*, 2008). To address these challenges, biodiversity portals employ different taxonomic resolution approaches (Table 2).

Taxonomic Backbone Systems (TBS) integrate name-based information from multiple sources, providing a unified framework for taxonomic searches, cross-referencing, and data consistency. Some portals, such as GBIF, WFO, and POWO, use TBS. GBIF's Backbone Taxonomy incorporates all names in GBIF datasets, primarily referencing the CoL. WFO combines a taxonomic backbone, supported by Botalista, with active curation by **Taxonomic Expert Networks (TENs)** (Borsch *et al.*, 2020; Palese *et al.*, 2019). POWO relies on the World Checklist of Vascular Plants (WCVP) as its taxonomic backbone, which is also integrated into CoL and GBIF (Govaerts *et al.*, 2021).

Checklist Approaches (CA) focus on regional species or specific taxonomic groups, minimizing interconnectivity between DBs (Feng *et al.*, 2022). Some portals use **Taxonomic Name Resolution Services (TNRS)** to resolve nomenclature conflicts. Portals such as BIEN, TRY, and GIFT implement TNRS, integrating global and regional checklists, including WCVP, WFO, Tropicos, USDA's Plants Database, the Global Compositae Checklist, and the International Legume Database and Information Service (Boyle *et al.*, 2013; Denelle *et al.*, 2023; Kattge *et al.*, 2020; Maitner *et al.*, 2018; Weigelt *et al.*, 2019).

Some portals collaborate with taxonomic experts and integrate multiple checklists for comprehensive data management. Platforms such as CoL, Tropicos, and WFO serve as primary checklists within major databases like GBIF, EOL, MOL, and TNRS (Boyle *et al.*, 2013; Hobern *et al.*, 2021; Jetz *et al.*, 2012; Parr *et al.*, 2016). Additionally, CoL and GBIF developed **ChecklistBank** (https://www.checklistbank.org/), an open portal that integrates taxonomic checklists (Hobern *et al.*, 2021).

Some biodiversity portals implement custom integration workflows tailored to specific research needs. MOL overlays geographical information to link species distributions with environmental data (Jetz *et al.*, 2012). BioTIME employs a structured table system with five levels (records, sample, plot, site, and study) to connect species abundance data with spatial and temporal information. Typographic errors and misspellings in species names are corrected during curation process (CP) (Dornelas *et al.*, 2018). ForestForward utilizes an automated ETL process with Python scripts to integrate thousands of GBIF datasets (Tejada-Gutierrez *et al.*, 2024).

In general, and despite their methodological differences, most portals use semi-automated or fully automated processes to remove duplicate and incomplete records, standardize taxonomic data, and curate missing or ambiguous information (Döring & Ower, 2019; Hobern *et al*., 2021; Tejada-Gutierrez *et al.*, 2024). Many also rely on data and standards derived from GBIF, reinforcing its central role in biodiversity data integration.

4.3 Trait Data Standardization

Another key challenge in biodiversity data integration is the standardization of trait data. Variability in measurement units, methodologies, and trait terminology complicates interoperability and often exceeds the capabilities of standards such as DwC (Deans *et al.*, 2015; Feng *et al.*, 2022; Wieczorek *et al.*, 2012). To address this, biodiversity portals typically adopt either Ontology-Based Integration or Standardized Trait Terminology (Table 3).

For example, EOL's TraitBank employs DwC metadata standard (MS) and its packaging format DwC-A to organize and share data. Additionally, TraitBank integrates ontologies to describe traits in a structured and interoperable way. These include the Plant Trait Ontology (TO), Phenotypic Quality Ontology (PATO), and the Environment Ontology (ENVO), and other vocabularies from Open Biological and Biomedical Ontologies (OBO) (Baker *et al.*, 2014; Parr *et al.*, 2016). TRY enhances interoperability using the Extensible Observation Ontology (OBOE) schema and trait vocabulary (TV) and protocols (P) as the Thesaurus of Plant Characteristics (TOP) and the Plant Trait Handbook (Garnier *et al.*, 2017; Pérez-Harguindeguy *et al.*, 2013). Similarly, GFIT and BIEN integrate with TRY to ensure consistency in trait data representation (Kattge *et al.*, 2020) (Suppelementary Table S3).

Effective multimodal data integration requires a multidisciplinary approach, as biodiversity datasets encompass complex variables, including species occurrences, traits, environmental parameters, and remote sensing information. Addressing these standardization challenges is crucial for advancing biodiversity analysis, which remains constrained by data inconsistencies and limited interoperability (Enquist *et al.*, 2019; Gotelli, 2008).

5 Multimodal Data in Biodiversity Research

The plant-related information contained in biodiversity databases is essential for documenting and analyzing how that biodiversity changes. However, identifying the factors that correlate with or drive these changes requires integrating additional multimodal datasets, including genetic, morphological, climatic, soil, land-use, and ecological parameters (O'Connor *et al.*, 2020; Soltis &

Soltis, 2016). This section reviews key environmental and climate-related data sources used to assess and predict biodiversity changes.

5.1 Environmental drivers of biodiversity change

Environmental variables, such as temperature, precipitation, and solar radiation, play a crucial role in shaping biodiversity by influencing species fitness, adaptation, and evolution (Guan *et al.*, 2023; McGaughran *et al.*, 2021; Segovia, 2023; Zhang *et al.*, 2018; Zheng *et al.*, 2024). Temperature fluctuations, particularly extreme highs and lows, impact key physiological processes, including photosynthesis, respiration, germination, and growth (Minoli *et al.*, 2019; Najeeb *et al.*, 2019; Pimentel, 2022). Freezing temperatures can induce stress, affecting cellular survival, gene expression, and overall plant development (Guan *et al.*, 2023). Meanwhile, precipitation patterns influence plant productivity, biome distribution, and species diversity, especially in arid environments, highlighting the importance of spatial resolution in biodiversity studies (Korell *et al.*, 2021).

Land-use and land-cover changes are also major drivers of biodiversity loss and carbon stock alterations (O'Connor *et al.*, 2020). Deforestation accelerates greenhouse gas emissions and disrupts ecosystems, often leading to declines in native species and increased prevalence of invasive species, particularly when forests are converted to agricultural land (Calvin *et al.*, 2023; Montràs-Janer *et al.*, 2024). Conversely, climate change itself influences land-use patterns as communities adapt to shifting environmental conditions (Calvin *et al.*, 2023; IPCC, 2018; Pörtner *et al.*, 2021). Given their predictive power in climate change research, greenhouse gas emissions are among the most frequently analyzed environmental parameters (Calvin *et al.*, 2023; Conradi *et al.*, 2024; Isbell *et al.*, 2023).

Altitude significantly affects climate variables such as temperature, precipitation, and wind patterns, shaping vegetation distribution and biodiversity across different elevation zones (Umair *et al.*, 2023). Variations in elevation, slope, and land relief create distinct biomes, each supporting different plant communities (Hakkenberg *et al.*, 2023). Research has identified three primary elevation-related biodiversity patterns: (1) a low-elevation diversity plateau, (2) an intermediate peak in mid-elevation zones, and (3) a steady decline in diversity at high elevations (Umair *et al.*, 2023). Microclimate and biome-level assessments are critical for understanding spatial changes in plant biodiversity.

5.2 Data Sources and Tools for Environmental Analysis

Environmental data are available from multiple sources with varying temporal and spatial resolutions. Climate Data Records (CDRs)—derived from satellite observations, climatological stations, and radar systems—offer long-term, homogeneous datasets for detecting climate trends (National Research Council, 2004). Remote sensing technologies provide crucial data on land-use, land-cover, and topographic changes over time (O'Connor *et al.*, 2020). Table 4 summarizes the primary sources for climate and environmental data used in biodiversity and climate change research (Al-Zu'bi *et al.*, 2022; Bedair *et al.*, 2023; Lundstad *et al.*, 2023).

Predictive tools such as the Köppen-Geiger climate classification and Shared Socioeconomic Pathways (SSPs)—developed within the Coupled Model Intercomparison Project (CMIP)—remain underutilized in biodiversity studies. These frameworks offer significant potential for integrating geographical, ecological, and climate change information to forecast biodiversity changes (Beck *et al.*, 2018, 2023; Calvin *et al.*, 2023). Leveraging these predictive approaches could enhance the accuracy of biodiversity projections and improve conservation strategies.

6 Plant biodiversity data analysis: approaches and challenges

6.1 Measuring Biodiversity

Large-scale biodiversity assessments often rely on diversity indices that condense complex ecological information into a single value (Table 5). Biodiversity is typically characterized by two key components: **species richness**, which represents the total number of species in a community, and **evenness**, which describes the relative distribution of individuals among species (Enquist *et al.*, 2019; Gotelli, 2008; Robinson, 2020).

Species richness is a fundamental but challenging metric, as identifying all species within a community is often impractical, and its value is highly dependent on sample size (Enquist *et al.*, 2019; Hurlbert, 1971). To account for this, indices such as **Margalef** (1958) and **Menhinick** (1954) estimate richness by relating the number of species to the total number of individuals observed. Other indices, including **Shannon-Wiener** (Shannon & Weaver, 1949) and **Simpson** (1949)**,** integrate both richness and evenness to provide a more comprehensive assessment of community structure. Changes in these indices can be used to evaluate how environmental factors, such as land-use modifications, impact biodiversity (IPCC, 2018).

Beyond species diversity, understanding species composition is crucial for assessing ecosystem health. For instance, invasive species frequently establish themselves in areas altered by climate change, outcompeting native species and reducing local biodiversity (Calvin *et al.*, 2023; Segan *et al.*, 2016; Wang *et al.*, 2019). Monitoring the abundance of invasive species is essential for developing conservation policies aimed at protecting native biodiversity and managing vulnerable ecosystems (Bellard *et al.*, 2014; Costello, 2023; de Lima *et al.*, 2020; Guo *et al.*, 2023). However, research on invasive species management has declined over the past two decades (Heberling *et al.*, 2021). Additionally, Enquist *et al.* (2019) reported that approximately **36.5% of the 435,000 known plant species are classified as rare**, underscoring the need to incorporate species composition into biodiversity assessments. Addressing these challenges requires access to high-quality, reliable, and traceable biodiversity data.

Portals such as **ForestForward**, which integrates and standardizes over 4,000 plant datasets from GBIF, are efforts in the right direction to improve biodiversity data quality and accessibility. By curating original datasets and removing low-quality entries with incomplete information, ForestForward enhances data reliability for biodiversity research (Tejada-Gutierrez *et al.*, 2024). However, like many similar platforms, it lacks comprehensive environmental variable datasets necessary for understanding the factors influencing biodiversity patterns (Heberling *et al.*, 2021). Expanding such portals to include environmental data would significantly improve their utility for biodiversity analysis and conservation planning.

6.2 Understanding Changes in Biodiversity

Understanding the effects of climate change on plant biodiversity is essential for modeling future scenarios and assessing the impact of human activities on ecosystems (Bellard *et al.*, 2014; Mantyka-Pringle *et al.*, 2015). Advancing biodiversity research requires a comprehensive analysis of the **spatial-temporal dynamics** embedded in ecological data (Saran *et al.*, 2022). However, few biodiversity portals provide analytical tools for data exploration and modeling due to **heterogeneous data sources, biases, and methodological constraints**, often necessitating custom analytical pipelines. This section reviews key methodologies for analyzing spatial-temporal changes in plant biodiversity (Figure 3). Some portals have developed specialized tools for biodiversity trend analysis. POWO's 'expowo' R package enables biodiversity assessments across taxonomic levels above genus or within specific geographic regions but is restricted to POWO datasets (Zuanny *et al.*, 2024). Similarly, ForestGEO's CTFS R package facilitates the analysis of tree and stem measurements from forest dynamics plots (Anderson-Teixeira *et al.*, 2015; Condit, 2012; Davies *et al.*, 2021).

Integrative approaches often combine multiple data sources and analytical techniques. Wani *et al.* (2024) analyzed plant biodiversity in the Indian Himalayan Region (IHR) by aggregating data from the Botanical Survey of India, eFlora of India, the India Biodiversity Portal, and Flowers of India. They applied Sørensen's similarity index ('vegan' R package by Oksanen *et al.*, 2022) to compare species richness across states and used Chi-square tests on Pearson's contingency tables ('vcd' R package by Meyer *et al.*, 2020) to assess trait distribution between the Himalaya and Indo-Burma biodiversity hotspots. Additionally, global plant growth trends were analyzed using logistic regression (nnte' R package by Venables & Ripley, 2002) with trait data from GFIT (Weigelt *et al.*, 2019).

Species Distribution Models (SDMs) are widely used to predict biodiversity patterns based on environmental factors. The use of GBIF data is frequent in SDMs to characterize species occurrence, identify conservation areas, and assess climate change impacts (Araújo *et al.*, 2019; Elith & Leathwick, 2009; Ferraille *et al.*, 2023; Jetz *et al.*, 2012, 2019). However, SDMs are susceptible to geographical and environmental biases, as biodiversity data is often collected opportunistically (Baker *et al.*, 2022; Botella *et al.*, 2020).

To address SDM limitations, the 'ibis.iSDM' R package (Jung, 2023) integrates multiple data sources into unified models. Deep Neural Networks (DNNs) implemented using PyTorch (Paszke *et al.*, 2019; Van Rossum & Drake, 2009), have demonstrated higher predictive accuracy than traditional SDMs (Brun *et al.*, 2024). Probabilistic modeling approaches offer alternative solutions, using inhomogeneous Poisson process (IPP) models for presence-only data and Bernoulli generalized linear models (GLM) for presence/absence data to mitigate spatial sampling bias (Fithian *et al.*, 2014).

Bayesian mixed-effects models (Johnson *et al.*, 2024) address hierarchical and correlative non-independence (e.g., temporal, spatial, and phylogenetic autocorrelations) using integrated nested Laplace approximation (INLA). This approach is similar to the methodologies employed by Browning *et al.* (2022) and Dinnage *et al.* (2020) in their biodiversity studies. Alternative techniques include simple linear regression for long-term trends and Generalized Additive Models (GAMs) for nonlinear short-term patterns**.** Temporal correlations are managed through ARIMA models**,** LOESS smoothing, splines, and piecewise regression (Dornelas *et al.*, 2013). Zou *et al.* (2024) demonstrated the effectiveness of GAMs in controlling environmental covariates in forest leaf type distribution analyses.

In general, correlative models tend to more accurately predict biodiversity trends. Modern biodiversity models increasingly incorporate remote sensing data to map biodiversity based on forest structure attributes (Hakkenberg *et al.*, 2023). Generalized Linear Mixed Models (GLMMs) and Generalized Least Squares (GLS) (‘MASS’ and ‘nlme’ R packages) have been applied to analyze structural drivers of alpha, beta, and gamma diversity (Pinheiro & Bates, 2022; Venables & Ripley, 2002).

A comprehensive approach to biodiversity modeling is demonstrated by Steidinger *et al.* (2019), who evaluated global mutualistic symbioses between forest trees and soil microorganisms. Their study integrated environmental variables, functional traits, remote sensing, and forest plot data from The Global Forest Biodiversity Initiative (GFBI), employing Random Forest algorithms for predictive modeling. These techniques enhance biodiversity research by leveraging large-scale datasets to infer ecological relationships.

Despite these advances, challenges persist in integrating remote sensing, satellite imagery, and environmental datasets. Most studies focus on species-level patterns, often neglecting ecosystem functions and species interactions (Heberling *et al.*, 2021). Future research should emphasize multi-source data integration and interdisciplinary modeling approaches to improve biodiversity predictions and conservation strategies.

7 Final considerations

The integration of multimodal data offers a promising approach to model, predict, and understand plant biodiversity changes over time and geography (Dornelas *et al.*, 2013; Paul & Shanta Kumar, 2020; Wauchope *et al.*, 2021). The use of environmental variables could provide valuable insights into species abundance trends, ecosystem dynamics, and responses to climate change, but the lack of high-resolution, long-term, and harmonized datasets implied significant challenges (Balzan *et al.*, 2018; Beck *et al.*, 2023).

Several challenges present the integration of remote sensing data with biodiversity data. These include technological difficulties in accessing heterogeneous data with unified protocols, discrepancies and uncertainties between remote sensing and *in situ* data, spatial misalignment in biodiversity datasets, and the discontinuity of remote sensing data. These issues complicate the prediction of biodiversity patterns from these datasets (Guralnick & Neufeld, 2005; Hakkenberg *et al.*, 2023; Mo *et al.*, 2023; Saran *et al.*, 2022). Addressing these integration challenges is essential for leveraging the full potential of remote sensing and geographic information systems in advancing biodiversity research.

Additionally, technical incompatibility, incomplete or outdated datasets, and discrepancies in spatial and temporal resolutions complicate the synthesis of biodiversity data across portals (Huber *et al.*, 2021; IPBES, 2019). Robust computational tools and standardized protocols are necessary to solve interoperability across databases, ensuring high-quality data integration (Halbritter *et al.*, 2024; Heberling *et al.*, 2021; Latz *et al.*, 2024). By addressing these challenges, researchers can analyze biodiversity patterns and predict changes with improved precision, using quality data from sources such as ForestForward.

Integrating and standardizing historical datasets remains important for gaining insights into the evolution of biodiversity over the past century, even if remote sensing and real-time satellite data from programs like Copernicus Sentinel is increasingly being used to monitor biodiversity. Even though the classical datasets are sparser, they register information over a very prolonged time period that is absent from remote sensing data. In addition, the accuracy of observations made with the newer technologies is harder to validate and depends on several factors: training data quality, canopy coverage - open landscapes with small plants are harder to monitor-, and habitat size, among others (Paz *et al.*, 2022; Schweiger & Laliberté, 2022).

Modern approaches base on artificial intelligence, such as Large Language Models, have been implemented in medicine to address the limitations of unstructured data (Choi *et al.*, 2024; Shool *et al.*, 2025). Similarly, Physics-Informed Graph Neural Networks have been implemented to integrate and analyze spatiotemporal environmental data with discontinuous records (Liu & Pyrcz, 2023). Recently, ecological studies adopted these approaches to study species distribution and natural phenomena using images (Espitalier *et al.*, 2025; Taccari *et al.*, 2024).

In conclusion, quantifying and predicting plant diversity changes through time, contributes to identifying causes that affect the species diversity and to detect critical points to diversity conservation (Johnson *et al.*, 2017; Reba & Seto, 2020; Tittensor *et al.*, 2014). Understanding how plant diversity responds to environmental parameters is essential for developing effective conservation strategies, managing ecosystems, and attaining environmental and human health (FAO, 2021; Humphreys *et al.*, 2019). Combining modern tools with traditional methods for multimodal data integration can enhance our understanding and prediction of changes in plant biodiversity on a large scale.

Author Contributions

Emilce Soledad Martinez: conceptualization, investigation, visualization, writing - original draft preparation, writing – review and ending. **Eva Tejada-Gutiérrez**: writing – review and ending. **Jordi Mateo-Fornes**: writing – review and ending. Francesc Solsona: writing – review and ending, **Raquel Defacio**: writing – review and ending. **Rui Alves**: conceptualization, funding acquisition, investigation, project administration, resources, supervision, visualization, writing - original draft preparation, writing – review and ending.

Acknowledgments

We acknowledge 2021 SGR 1353 research group accreditation by Generalitat de Catalunya.

Competing Interests

The authors declare no competing interests.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Emilce S. Martinez https://orcid.org/0009-0004-5995-5538

Eva Tejada-Gutiérrez https://orcid.org/0009-0000-7884-1369

Albert Sorribas https://orcid.org/0000-0002-7407-4075

Jordi Mateo-Fornes https://orcid.org/0000-0002-1660-0380

Francesc Solsona https://orcid.org/0000-0002-4830-9184

Raquel Defacio https://orcid.org/0000-0003-2670-3065

Rui Alves https://orcid.org/0000-0002-8112-5184

References

Al-Zu'bi, M., Dejene, S. W., Hounkpè, J., Kupika, O. L., Lwasa, S., Mbenge, M., Mwongera, C., Ouedraogo, N. S., & Touré, N. D. E. (2022). African perspectives on climate change research. *Nature Climate Change*, *12*(12), 1078–1084. https://doi.org/10.1038/s41558-022-01519-x

Anderson-Teixeira, K. J., Davies, S. J., Bennett, A. C., Gonzalez-Akre, E. B., Muller-Landau, H. C., Joseph Wright, S., 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. https://doi.org/10.1111/GCB.12712

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1). https://doi.org/10.1126/SCIADV.AAT4858

Baker, D. J., Maclean, I. M. D., Goodall, M., & Gaston, K. J. (2022). Correlations between spatial sampling biases and environmental niches affect species distribution models. *Global Ecology and Biogeography*, *31*(6), 1038–1050. https://doi.org/10.1111/GEB.13491

Baker, E., Rycroft, S., & Smith, V. S. (2014). Linking multiple biodiversity informatics platforms with Darwin Core Archives. *Biodiversity Data Journal 2: E1039*, *2*(1), e1039-. https://doi.org/10.3897/BDJ.2.E1039

Balzan, M. V., Potschin-Young, M., & Haines-Young, R. (2018). Island ecosystem services: Insights from a literature review on case-study island ecosystem services and future prospects.

International Journal of Biodiversity Science, Ecosystem Services & Management, *14*(1), 71–90. https://doi.org/10.1080/21513732.2018.1439103

Bánki, O., Döring, M., Holleman, A., & Addink, W. (2018). Catalogue of Life Plus: Innovating the CoL systems as a foundation for a clearinghouse for names and taxonomy. *Biodiversity Information Science and Standards*, *2*, e26922. https://doi.org/10.3897/biss.2.26922

Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., & Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, *10*(1), 724. https://doi.org/10.1038/s41597-023-02549-6

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data 2018 5:1*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2018.214

Bedair, H., Alghariani, M. S., Omar, E., Anibaba, Q. A., Remon, M., Bornman, C., Kiboi, S. K., Rady, H. A., Salifu, A.-M. A., Ghosh, S., Guuroh, R. T., Sanou, L., & Alzain, H. M. (2023). Global Warming Status in the African Continent: Sources, Challenges, Policies, and Future Direction. *International Journal of Environmental Research*, *17*(3), 45. https://doi.org/10.1007/s41742-023-00534-w

Bellard, C., Leclerc, C., Leroy, B., Bakkenes, M., Veloz, S., Thuiller, W., & Courchamp, F. (2014). Vulnerability of biodiversity hotspots to global change. *Global Ecology and Biogeography*, *23*(12), 1376–1386. https://doi.org/10.1111/geb.12228

Blair, J., Gwiazdowski, R., Borrelli, A., Hotchkiss, M., Park, C., Perrett, G., & Hanner, R. (2020). Towards a catalogue of biodiversity databases: An ontological case study. *Biodiversity Data Journal*, *8*, e32765. https://doi.org/10.3897/BDJ.8.E32765

Borsch, T., Berendsohn, W., Dalcin, E., Delmas, M., Demissew, S., Elliott, A., Fritsch, P., Fuchs, A., Geltman, D., Güner, A., Haevermans, T., Knapp, S., Roux, M. M. le, Loizeau, P. A., Miller, C., Miller, J., Miller, J. T., Palese, R., Paton, A., … Zamora, N. (2020). World Flora Online: Placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. *TAXON*, *69*(6), 1311–1341. https://doi.org/10.1002/TAX.12373

Botella, C., Joly, A., Monestiez, P., Bonnet, P., & Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. *PLOS ONE*, *15*(5), e0232078. https://doi.org/10.1371/JOURNAL.PONE.0232078

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*(1), 1–15. https://doi.org/10.1186/1471-2105-14-16/TABLES/5

Browning, E., Freeman, R., Boughey, K. L., Isaac, N. J. B., & Jones, K. E. (2022). Accounting for spatial autocorrelation and environment are important to derive robust bat population trends from citizen science data. *Ecological Indicators*, *136*. https://doi.org/10.1016/J.ECOLIND.2022.108719

Brun, P., Karger, D. N., Zurell, D., Descombes, P., de Witte, L. C., de Lutio, R., Wegner, J. D., & Zimmermann, N. E. (2024). Multispecies deep learning using citizen science data produces more informative plant community models. *Nature Communications*, *15*(1), 4421. https://doi.org/10.1038/s41467-024-48559-9

Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., van Kleunen, M., Pergl, J., Pyšek, P., Stein, A., Winter, M., Barcelona, J. F., Fuentes, N., Inderjit, Karger, D. N., Kartesz, J., Kuprijanov, A., Nishino, M., Nickrent, D., … Weigelt, P. (2023). Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist*, *237*(4), 1432–1445. https://doi.org/10.1111/NPH.18533

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W. L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Ha, M. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* Intergovernmental Panel on Climate Change. https://doi.org/10.59327/IPCC/AR6-9789291691647

Chen, N., Tsendbazar, N.-E., Hamunyela, E., Verbesselt, J., & Herold, M. (2021). Sub-annual tropical forest disturbance monitoring using harmonized Landsat and Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102386. https://doi.org/10.1016/j.jag.2021.102386

Choi, I., Long, Q., & Getzen, E. (2024). *Filling the gaps: Leveraging large language models for temporal harmonization of clinical text across multiple medical visits for clinical prediction*. https://doi.org/10.1101/2024.05.06.24306959

Cohen, W. B., Healey, S. P., Yang, Z., Zhu, Z., & Gorelick, N. (2020). Diversity of algorithm and spectral band inputs improves landsat monitoring of forest disturbance. *Remote Sensing*, *12*(10). https://doi.org/10.3390/rs12101673

Condit, R. (2012). *CTFS R Package.* http://ctfs.si.edu/Public/CTFSRPackage

Conradi, T., Eggli, U., Kreft, H., Schweiger, A. H., Weigelt, P., & Higgins, S. I. (2024). Reassessment of the risks of climate change for terrestrial ecosystems. *Nature Ecology & Evolution*, 1–13. https://doi.org/10.1038/s41559-024-02333-8

Convention on Biological Diversity. (2011). *Convention on Biological Diversity: Text and annexes* (p. Article 2, Use of terms, 4.). Secretariat of the Convention on Biological Diversity, United Nations Environment Programme. https://www.cbd.int/convention/articles?a=cbd-02

Costello, M. J., M. M. Vale, W. Kiessling, S. Maharaj, J. Price, and G. H. Talukdar. (2023). Biodiversity Hotspots. In D. C. R. [H.-O. Pörtner M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. (Ed.), *Climate Change 2022 – Impacts, Adaptation and Vulnerability* (pp. 2123–2162). Cambridge University Press. https://doi.org/10.1017/9781009325844.018

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P. S., Baker, P. J., Baker, M. E., Baltzer, J. L., Basset, Y., Bissiengou, P., Bohlman, S., Bourg, N. A., Brockelman, W. Y., Bunyavejchewin, S., … Zuleta, D. (2021). ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biological Conservation*, *253*, 108907. https://doi.org/10.1016/j.biocon.2020.108907

de Lima, R. A. F., Oliveira, A. A., Pitta, G. R., de Gasper, A. L., Vibrans, A. C., Chave, J., ter Steege, H., & Prado, P. I. (2020). The erosion of biodiversity and biomass in the Atlantic Forest biodiversity hotspot. *Nature Communications*, *11*(1), 1–16. https://doi.org/10.1038/s41467-020-20217-w

de Melo, P. H. A., Bystriakova, N., Lucas, E., & Monro, A. K. (2024). A new R package to parse plant species occurrence records into unique collection events efficiently reduces data redundancy. *Scientific Reports*, *14*(1), 5450. https://doi.org/10.1038/s41598-024-56158-3

Deans, A. R., Lewis, S. E., Huala, E., Anzaldo, S. S., Ashburner, M., Balhoff, J. P., Blackburn, D. C., Blake, J. A., Burleigh, J. G., Chanet, B., Cooper, L. D., Courtot, M., Csösz, S., Cui, H., Dahdul, W., Das, S., Dececchi, T. A., Dettai, A., Diogo, R., … Mabee, P. (2015). Finding Our Way through Phenotypes. *PLOS Biology*, *13*(1), e1002033. https://doi.org/10.1371/JOURNAL.PBIO.1002033

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. https://doi.org/10.1111/2041-210X.14213

Dinnage, R., Skeels, A., & Cardillo, M. (2020). Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proceedings of the Royal Society B*, *287*(1926). https://doi.org/10.1098/RSPB.2019.2817

Döring, M., & Ower, G. (2019). The Catalogue of Life Data Package—A new format for exchanging nomenclatural and taxonomic information. *Biodiversity Information Science and Standards*, *3*. https://doi.org/10.3897/biss.3.38771

Dornelas, M., Antão, L. H., Moyes, F., Bates, A. E., Magurran, A. E., Adam, D., Akhmetzhanova, A. A., Appeltans, W., Arcos, J. M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A. H., Barbosa, M., Barreto, T. E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., … Zettler, M. L. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, *27*(7), 760–786. https://doi.org/10.1111/geb.12729

Dornelas, M., Magurran, A. E., Buckland, S. T., Chao, A., Chazdon, R. L., Colwell, R. K., Curtis, T., Gaston, K. J., Gotelli, N. J., Kosnik, M. A., McGill, B., McCune, J. L., Morlon, H., Mumby, P. J., Øvreås, L., Studeny, A., & Vellend, M. (2013). Quantifying temporal change in biodiversity: Challenges and opportunities. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1750), 20121931. https://doi.org/10.1098/rspb.2012.1931

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*, *289*, 2312–2314.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(Volume 40, 2009), 677–697. https://doi.org/10.1146/ANNUREV.ECOLSYS.110308.120159/1

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. https://doi.org/10.1126/sciadv.aaz0414

Enquist, B. J., Schildhauer, M., Thiers, B. M., Steere, L., & Svenning, C. (2016). *Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity Short title: Botanical Information and Ecology Network (BIEN) Project Leaders*. https://doi.org/10.7287/peerj.preprints.2615v2

Espitalier, V., Lombardo, J.-C., Goëau, H., Botella, C., Høye, T. T., Dyrmann, M., Bonnet, P., & Joly, A. (2025). Adapting a global plant identification model to detect invasive alien plant species in high-resolution road side images. *Ecological Informatics*, *89*, 103129. https://doi.org/10.1016/j.ecoinf.2025.103129

FAO. (2021). *Programa de la FAO de respuesta y recuperación ante la pandemia COVID-19—Prevenir la próxima pandemia zoonótica*. FAO. https://doi.org/10.4060/cb0301es

FAO. (2022). El estado de los bosques del mundo 2022, Vías forestales hacia la recuperación verde y la creación de economías inclusivas, resilientes y sostenibles. In *El estado de los bosques del mundo 2022*.

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., … López-Hoffman, L. (2022). A review of the heterogeneous landscape of biodiversity databases: Opportunities and challenges for a synthesized biodiversity knowledge base. *Global Ecology and Biogeography*, *31*(7), 1242–1260. https://doi.org/10.1111/GEB.13497

Fernández, N., Ferrier, S., Navarro, L. M., & Pereira, H. M. (2020). Essential biodiversity variables: Integrating in-situ observations and remote sensing through modeling. *Remote Sensing of Plant Biodiversity*, 485–501. https://doi.org/10.1007/978-3-030-33157-3_18/FIGURES/5

Ferraille, T., Kerbiriou, C., Bigard, C., Claireau, F., & Thompson, J. D. (2023). Integrating biodiversity assessments into local conservation planning: The importance of assessing suitable data sources. *Peer Community in Ecology*. https://doi.org/10.1101/2023.05.09.539999

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2014). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, *6*(4), 424. https://doi.org/10.1111/2041-210X.12242

Franz, N. M., Peet, R. K., & Weakley, A. S. (2008). On the use of taxonomic concepts in support of biodiversity research and taxonomy. *The New Taxonomy*, 63–86. https://doi.org/10.1201/9781420008562

Ganjdanesh, A., Zhang, J., Yan, S., Chen, W., & Huang, H. (2022). Multimodal Genotype and Phenotype Data Integration to Improve Partial Data-Based Longitudinal Prediction. *Journal of Computational Biology*, *29*(12), 1324–1345. https://doi.org/10.1089/cmb.2022.0378

Garnier, E., Stahl, U., Laporte, M. A., Kattge, J., Mougenot, I., Kühn, I., Laporte, B., Amiaud, B., Ahrestani, F. S., Bönisch, G., Bunker, D. E., Cornelissen, J. H. C., Díaz, S., Enquist, B. J., Gachet, S., Jaureguiberry, P., Kleyer, M., Lavorel, S., Maicher, L., … Klotz, S. (2017).

Towards a thesaurus of plant characteristics: An ecological contribution. *Journal of Ecology*, *105*(2), 298–309. https://doi.org/10.1111/1365-2745.12698

Google. (2024). NotebookLM [AI document analysis tool]. https://notebooklm.google.com/

Gotelli, N. J. (2008). A Primer of Ecology. In *Google Docs*. Sinauer Associates, Inc. https://drive.google.com/file/d/1OgmuQQJIdbJfSwb1LG7t93eDWu9_QUum/view?usp=drive_open&edoph=true&usp=embed_facebook

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*, 215. https://doi.org/10.1038/s41597-021-00997-6

Guan, Y., Hwarari, D., Korboe, H. M., Ahmad, B., Cao, Y., Movahedi, A., & Yang, L. (2023). Low temperature stress-induced perception and molecular signaling pathways in plants. *Environmental and Experimental Botany*, *207*, 105190. https://doi.org/10.1016/j.envexpbot.2022.105190

Guo, W.-Y., Serra-Diaz, J. M., Eiserhardt, W. L., Maitner, B. S., Merow, C., Violle, C., Pound, M. J., Sun, M., Slik, F., Blach-Overgaard, A., Enquist, B. J., & Svenning, J.-C. (2023). Climate change and land use threaten global hotspots of phylogenetic endemism for trees. *Nature Communications*, *14*(1), 6950. https://doi.org/10.1038/s41467-023-42671-y

Guralnick, R., Walls, R., & Jetz, W. (2018). Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography*, *41*(5), 713–725. https://doi.org/10.1111/ecog.02942

Guralnick, R. P., & Neufeld, D. (2005). Challenges Building Online GIS Services to Support Global Biodiversity Mapping and Analysis: Lessons from the Mountain and Plains Database and Informatics project. *Biodiversity Informatics*, *2*(0), 56–69. https://doi.org/10.17161/BI.V2I0.19

Guralnick, R., Walls, R., & Jetz, W. (2018). Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography, 41(5), 713–725. https://doi.org/10.1111/ecog.02942

Hakkenberg, C. R., Atkins, J. W., Brodie, J. F., Burns, P., Cushman, S., Jantz, P., Kaszta, Z., Quinn, C. A., Rose, M. D., & Goetz, S. J. (2023). Inferring alpha, beta, and gamma plant diversity across biomes with GEDI spaceborne lidar. *Environmental Research: Ecology*, *2*(3), 035005. https://doi.org/10.1088/2752-664X/acffcd

Halbritter, A. H., Vandvik, V., Cotner, S. H., Farfan-Rios, W., Maitner, B. S., Michaletz, S. T., Oliveras Menor, I., Telford, R. J., Ccahuana, A., Cruz, R., Sallo-Bravo, J., Santos-Andrade, P. E., Vilca-Bustamante, L. L., Castorena, M., Chacón-Labella, J., Christiansen, C. T., Duran, S. M., Egelkraut, D. D., Gya, R., … Enquist, B. J. (2024). Plant trait and vegetation data along a 1314 m elevation gradient with fire history in Puna grasslands, Perú. *Scientific Data*, *11*(1), 225. https://doi.org/10.1038/s41597-024-02980-3

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/SUPPL_FILE/PNAS.2018093118.SAPP.PDF

Hobern, D., Barik, S. K., Christidis, L., T.Garnett, S., Kirk, P., Orrell, T. M., Pape, T., Pyle, R. L., Thiele, K. R., Zachos, F. E., & Bánki, O. (2021). Towards a global list of accepted species VI: The Catalogue of Life checklist. *Organisms Diversity and Evolution*, *21*(4), 677–690. https://doi.org/10.1007/s13127-021-00516-w

Huber, R., D’Onofrio, C., Devaraju, A., Klump, J., Loescher, H. W., Kindermann, S., Guru, S., Grant, M., Morris, B., Wyborn, L., Evans, B., Goldfarb, D., Genazzio, M. A., Ren, X., Magagna, B., Thiemann, H., & Stocker, M. (2021). Integrating data and analysis technologies within leading environmental research infrastructures: Challenges and approaches. *Ecological Informatics*, *61*. https://doi.org/10.1016/j.ecoinf.2021.101245

Humphreys, A. M., Govaerts, R., Ficinski, S. Z., Nic Lughadha, E., & Vorontsova, M. S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature Ecology \& Evolution*, *3*(7), 1043–1047. https://doi.org/10.1038/s41559-019-0906-2

Hurlbert, S. H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. Ecology, 52(4), 577–586. https://doi.org/10.2307/1934145

IPBES. (2019). Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity. In *IPBES secretariat*.

IPCC. (2018). Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. In V. P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor & T. W. Masson-Delmotte (Eds.), *Global Warming of 1.5°C* (pp. 3–24). Cambridge University Press. https://www.cambridge.org/core/product/identifier/9781009157940%23prf2/type/book_part

Isbell, F., Balvanera, P., Mori, A. S., He, J. S., Bullock, J. M., Regmi, G. R., Seabloom, E. W., Ferrier, S., Sala, O. E., Guerrero-Ramírez, N. R., Tavella, J., Larkin, D. J., Schmid, B., Outhwaite, C. L., Pramual, P., Borer, E. T., Loreau, M., Omotoriogun, T. C., Obura, D. O., … Palmer, M. S. (2023). Expert perspectives on global biodiversity loss and its drivers and impacts on people. *Frontiers in Ecology and the Environment*, *21*(2), 94–103. https://doi.org/10.1002/FEE.2536

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 2019 3:4*, *3*(4), 539–551. https://doi.org/10.1038/s41559-019-0826-1

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology & Evolution*, *27*(3), 151–159. https://doi.org/10.1016/J.TREE.2011.09.007

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science*, *356*(6335), 270–275. https://doi.org/10.1126/science.aam9317

Johnson, T. F., Beckerman, A. P., Childs, D. Z., Webb, T. J., Evans, K. L., Griffiths, C. A., Capdevila, P., Clements, C. F., Besson, M., Gregory, R. D., Thomas, G. H., Delmas, E., & Freckleton, R. P. (2024). Revealing uncertainty in the status of biodiversity change. *Nature*, *628*(8009), 788–794. https://doi.org/10.1038/s41586-024-07236-z

Jones, M. B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T., Earl, S., & Chong, S. (2019). *Ecological Metadata Language (EML)*.

Jung, M. (2023). An integrated species distribution modelling framework for heterogeneous biodiversity data. *Ecological Informatics*, *76*. https://doi.org/10.1016/j.ecoinf.2023.102127

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., C, C. A., Aleixo, I., Ali, H., … Wirth, C. (2020). TRY plant trait database –

enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/10.1111/GCB.14904

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. https://doi.org/10.1111/J.1365-2486.2011.02451.X

Keppel, G., Craven, D., Weigelt, P., Smith, S. A., Sande, M. T. van der, Sandel, B., Levin, S. C., Kreft, H., & Knight, T. M. (2021). Synthesizing tree biodiversity data to understand global patterns and processes of vegetation. *Journal of Vegetation Science*, *32*(3). https://doi.org/10.1111/jvs.13021

Khare, S., Latifi, H., & Rossi, S. (2021). A 15-year spatio-temporal analysis of plant β-diversity using Landsat time series derived Rao's Q index. *Ecological Indicators*, *121*. https://doi.org/10.1016/j.ecolind.2020.107105

Korell, L., Auge, H., Chase, J. M., Harpole, W. S., & Knight, T. M. (2021). Responses of plant diversity to precipitation change are strongest at local spatial scales and in drylands. *Nature Communications*, *12*(1), 2489. https://doi.org/10.1038/s41467-021-22766-0

Latz, M. A. C., Andersson, A., Brugel, S., Hedblom, M., Jurdzinski, K. T., Karlson, B., Lindh, M., Lycken, J., Torstensson, A., & Andersson, A. F. (2024). A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea. *Scientific Data*, *11*(1), 18. https://doi.org/10.1038/s41597-023-02825-5

Liu, W., & Pyrcz, M. J. (2023). Physics-informed graph neural network for spatial-temporal production forecasting. *Geoenergy Science and Engineering*, *223*, 211486. https://doi.org/10.1016/j.geoen.2023.211486

Lundstad, E., Brugnara, Y., Pappert, D., Kopp, J., Samakinwa, E., Hürzeler, A., Andersson, A., Chimani, B., Cornes, R., Demarée, G., Filipiak, J., Gates, L., Ives, G. L., Jones, J. M., Jourdain, S., Kiss, A., Nicholson, S. E., Przybylak, R., Jones, P., … Brönnimann, S. (2023). The global historical climate database HCLIM. *Scientific Data*, *10*(1), 44. https://doi.org/10.1038/s41597-022-01919-w

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, *2*(3), 279–296. https://doi.org/10.1016/J.ECOINF.2007.05.004

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. (2018). The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution*, *9*(2), 373–379. https://doi.org/10.1111/2041-210X.12861

Mantyka-Pringle, C. S., Visconti, P., Di Marco, M., Martin, T. G., Rondinini, C., & Rhodes, J. R. (2015). Climate change modifies risk of global biodiversity loss due to land-cover change. *Biological Conservation*, *187*, 103–111. https://doi.org/10.1016/j.biocon.2015.04.016

Margalef, D. R. (1958). Information Theory in Ecology. *General Systematics*, *3*, 36–71.

McGaughran, A., Laver, R., & Fraser, C. (2021). Evolutionary Responses to Warming. *Trends in Ecology & Evolution*, *36*(7), 591–600. https://doi.org/10.1016/j.tree.2021.02.014

Menhinick, E. F. (1964). A Comparison of Some Species-Individuals Diversity Indices Applied to Samples of Field Insects. *Ecology*, *45*(4), 859–861. https://doi.org/10.2307/1934933

Meyer, D., Zeileis, A., Hornik, K., & Friendly, M. (2020). Visualizing Categorical Data [R package vcd version 1.4-10]. *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.VCD

Minoli, S., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Zabel, F., Dury, M., Folberth, C., François, L., Hank, T., Jacquemin, I., Liu, W., Olin, S., & Pugh, T. A. M. (2019). Global Response Patterns of Major Rainfed Crops to Adaptation by Maintaining Current Growing Periods and Irrigation. *Earth's Future*, *7*(12), 1464–1480. https://doi.org/10.1029/2018EF001130

Mo, L., Zohner, C. M., Reich, P. B., Liang, J., de Miguel, S., Nabuurs, G.-J., Renner, S. S., van den Hoogen, J., Araza, A., Herold, M., Mirzagholi, L., Ma, H., Averill, C., Phillips, O. L., Gamarra, J. G. P., Hordijk, I., Routh, D., Abegg, M., Adou Yao, Y. C., … Crowther, T. W. (2023). Integrated global assessment of the natural forest carbon potential. *Nature*, *624*(7990), Article 7990. https://doi.org/10.1038/s41586-023-06723-z

Montràs-Janer, T., Suggitt, A. J., Fox, R., Jönsson, M., Martay, B., Roy, D. B., Walker, K. J., & Auffret, A. G. (2024). Anthropogenic climate and land-use change drive short- and long-term biodiversity shifts across taxa. *Nature Ecology & Evolution*, *8*(4), 739–751. https://doi.org/10.1038/s41559-024-02326-7

Najeeb, U., Tan, D. K. Y., Sarwar, M., & Ali, S. (2019). Adaptation of Crops to Warmer Climates: Morphological and Physiological Mechanisms. In *Sustainable Solutions for Food Security* (pp. 27–50). Springer International Publishing. https://doi.org/10.1007/978-3-319-77878-5_2

National Research Council. (2004). *Climate Data Records from Environmental Satellites*. National Academies Press. https://doi.org/10.17226/10944

Negassa, M. D., Mallie, D. T., & Gemeda, D. O. (2020). Forest cover change detection using Geographic Information Systems and remote sensing techniques: A spatio-temporal study on Komto Protected forest priority area, East Wollega Zone, Ethiopia. *Environmental Systems Research*, *9*(1), 1–14. https://doi.org/10.1186/s40068-020-0163-z

O'Connor, B., Bojinski, S., Röösli, C., & Schaepman, M. E. (2020). Monitoring global changes in biodiversity and climate essential as ecological crisis intensifies. *Ecological Informatics*, *55*. https://doi.org/10.1016/j.ecoinf.2019.101033

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Borman, T. (2022). Community Ecology Package [R package vegan version 2.6-2]. *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.VEGAN

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Palese, R., Boillat, C., & Loizeau, P.-A. (2019). World Flora Online (WFO)—Quality control workflow for an evolving taxonomic backbone. *Biodiversity Information Science and Standards 3: E35307*, *3*, e35307-. https://doi.org/10.3897/BISS.3.35307

Parr, C. S., Schulz, K. S., Hammock, J., Wilson, N., Leary, P., Rice, J., & Corrigan, R. J. (2016). TraitBank: Practical semantics for organism attribute data. *Semantic Web*, *7*(6), 577–588. https://doi.org/10.3233/SW-150190

Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., Hammock, J. A., Goddard, A., Rice, J., Studer, M., Holmes, J. T. G., & Corrigan, R. J. (2014). The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal*, *2*(1). https://doi.org/10.3897/BDJ.2.E1079

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury Google, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Xamla, A. K., Yang, E., Devito, Z., Raison Nabla, M., Tejani, A., Chilamkurthy, S., Ai, Q., Steiner, B., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, *32*.

Paul, R., & Shanta Kumar, T. (2020). *A Machine Learning Approach to Biodiversity Time Series Analysis* (SSRN Scholarly Paper 3520735). https://doi.org/10.2139/ssrn.3520735

Paz, A., Silva, T. S., & Carnaval, A. C. (2022). A framework for near-real time monitoring of diversity patterns based on indirect remote sensing, with an application in the Brazilian Atlantic rainforest. PeerJ, 10, e13534. https://doi.org/10.7717/peerj.13534

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M. S., Cornwell, W. K., Craine, J. M., Gurvich, D. E., Urcelay, C., Veneklaas, E. J., Reich, P. B., Poorter, L., Wright, I. J., Ray, P., Enrico, L., Pausas, J. G., De Vos, A. C., … Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, *61*(3), 167–234. https://doi.org/10.1071/BT12225

Pimentel, C. (2022). Plant Responses to High-Temperature Stress. *Archives of Agriculture Research and Technology (AART)*, *3*(3), 1–2. https://doi.org/10.54026/AART/1043

Pinheiro, J., & Bates, D. (2022). nlme: Linear and nonlinear mixed effects models R package version 3.1–157. *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.PACKAGE.NLME

Pontius, J., Schaberg, P., & Hanavan, R. P. (2020). Remote sensing for early, detailed, and accurate detection of forest disturbance and decline for protection of biodiversity. *Remote Sensing of Plant Biodiversity*, 121–154. https://doi.org/10.1007/978-3-030-33157-3_6/FIGURES/13

Pörtner, H. O., Scholes, R. J., Agard, J., Archer, E., Arneth, A., Bai, X., Barnes, D., Burrows, M., Strassburg, B., Chan, L., Cheung, W. L., Diamond, S., Donatti, C., Duarte, C., Eisenhauer, N., Foden, W., Gasalla, M. A., Handa, C., Hickler, T., … Ngo, H. T. (2021). IPBES-IPCC CO-SPONSORED WORKSHOP BIODIVERSITY AND CLIMATE CHANGE WORKSHOP REPORT. *IPBES and IPCC*. https://doi.org/10.5281/zenodo.4782538

Reba, M., & Seto, K. C. (2020). A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change. *Remote Sensing of Environment*, *242*, 111739. https://doi.org/10.1016/j.rse.2020.111739

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, *331*(6018), 703–705. https://doi.org/10.1126/SCIENCE.1197962

Robinson, S. (2020). A richness index for baselining climate change adaptations in small island developing states. *Environmental and Sustainability Indicators*, *8*, 100065. https://doi.org/10.1016/j.indic.2020.100065

Saran, S., Chaudhary, S. K., Singh, P., Tiwari, A., & Kumar, V. (2022). A comprehensive review on biodiversity information portals. *Biodiversity and Conservation*, *31*(5–6), 1445–1468. https://doi.org/10.1007/s10531-022-02420-x

Schellenberger Costa, D., Boehnisch, G., Freiberg, M., Govaerts, R., Grenié, M., Hassler, M., Kattge, J., Muellner-Riehl, A. N., Rojas Andrés, B. M., Winter, M., Watson, M., Zizka, A., & Wirth, C. (2023). The big four of plant taxonomy – a comparison of global checklists of vascular plant names. *New Phytologist*, *240*(4), 1687–1702. https://doi.org/10.1111/nph.18961

Segan, D. B., Murray, K. A., & Watson, J. E. M. (2016). A global assessment of current and future biodiversity vulnerability to habitat loss–climate change interactions. *Global Ecology and Conservation*, *5*, 12–21. https://doi.org/10.1016/j.gecco.2015.11.002

Segovia, R. A. (2023). Temperature predicts maximum tree-species richness and water availability and frost shape the residual variation. *Ecology*, *104*(5). https://doi.org/10.1002/ecy.4000

Shool, S., Adimi, S., Saboori Amleshi, R., Bitaraf, E., Golpira, R., & Tara, M. (2025). A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Medical Informatics and Decision Making*, *25*(1), 117. https://doi.org/10.1186/s12911-025-02954-4

Soltis, D. E., & Soltis, P. S. (2016). Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Diversity*, *38*(6), 264–270. https://doi.org/10.1016/j.pld.2016.12.001

Steidinger, B. S., Crowther, T. W., Liang, J., Van Nuland, M. E., Werner, G. D. A., Reich, P. B., Nabuurs, G. J., de-Miguel, S., Zhou, M., Picard, N., Herault, B., Zhao, X., Zhang, C., Routh, D., & Peay, K. G. (2019). Author Correction: Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature*, *571*(7765), E8–E8. https://doi.org/10.1038/s41586-019-1342-9

Taccari, M. L., Wang, H., Nuttall, J., Chen, X., & Jimack, P. K. (2024). Spatial-temporal graph neural networks for groundwater data. *Scientific Reports*, *14*(1), 24564. https://doi.org/10.1038/s41598-024-75385-2

Teisher, J., & Stimmel, H. (2024). Tropicos MO Specimen Data. *Missouri Botanical Garden.*, *Occurrence dataset*. https://doi.org/10.15468/HJA69F

Tejada-Gutierrez, E. L., Mateo-Fornés, J., Solsona, F., & Alves, R. (2024). ForestForward: Visualizing and accessing integrated world forest data from the last fifty years. *bioRxiv*, 2024.12.05.626943. https://doi.org/10.1101/2024.12.05.626943

Tittensor, D. P., Walpole, M., Hill, S. L. L., Boyce, D. G., Britten, G. L., Burgess, N. D., Butchart, S. H. M., Leadley, P. W., Regan, E. C., Alkemade, R., Baumung, R., Bellard, C., Bouwman, L., Bowles-Newark, N. J., Chenery, A. M., Cheung, W. W. L., Christensen, V., Cooper, H. D., Crowther, A. R., … Ye, Y. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science (New York, N.Y.)*, *346*(6206), 241–244. https://doi.org/10.1126/science.1257484

Umair, M., Hu, X., Cheng, Q., Ali, S., & Ni, J. (2023). Distribution patterns of fern species richness along elevations the Tibetan Plateau in China: Regional differences and effects of climate change variables. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1178603

United Nations. (1992). *Convention on Biological Diversity (text with annexes).* United Nations.

United Nations. (2018). *The 2030 Agenda and the Sustainable Development Goals An opportunity for Latin America and the Caribbean Goals, Targets and Global Indicators*. www.issuu.com/publicacionescepal/stacks

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. https://doi.org/10.1007/978-0-387-21706-2

Wäldchen, J., & Mäder, P. (2017). Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. *Archives of Computational Methods in Engineering 2017 25:2*, *25*(2), 507–543. https://doi.org/10.1007/S11831-016-9206-Z

Wang, C.-J., Li, Q.-F., & Wan, J.-Z. (2019). Potential invasive plant expansion in global ecoregions under climate change. *PeerJ*, *7*, e6479. https://doi.org/10.7717/peerj.6479

Wani, S. A., Khuroo, A. A., Zaffar, N., Rafiqi, S., Farooq, I., Afzal, S., & Rashid, I. (2024). Data synthesis for biodiversity science: A database on plant diversity of the Indian Himalayan Region. *Biodiversity and Conservation*. https://doi.org/10.1007/s10531-024-02784-2

Wauchope, H. S., Amano, T., Geldmann, J., Johnston, A., Simmons, B. I., Sutherland, W. J., & Jones, J. P. G. (2021). Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution*, *36*(3), 196–205. https://doi.org/10.1016/j.tree.2020.11.001

Weigelt, P., König, C., & Kreft, Holger. (2019). GIFT-A Global Inventory of Floras and Traits for macroecology and biogeography. *Journal of Biogeography*, *47*, 16–43. https://doi.org/10.1111/jbi.13623

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE*, *7*(1), e29715. https://doi.org/10.1371/JOURNAL.PONE.0029715

Williamson, H. F., & Leonelli, S. (2023). Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. In *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. https://doi.org/10.1007/978-3-031-13276-6

Xu, L., Herold, M., Tsendbazar, N.-E., Masiliūnas, D., Li, L., Lesiv, M., Fritz, S., & Verbesselt, J. (2022). Time series analysis for global land cover change monitoring: A comparison across sensors. *Remote Sensing of Environment*, *217*, 112905. https://doi.org/10.1016/j.rse.2022.112905

Zhang, Q.-G., Lu, H.-S., & Buckling, A. (2018). Temperature drives diversification in a model adaptive radiation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1886), 20181515. https://doi.org/10.1098/rspb.2018.1515

Zheng, J., Guo, N., Huang, Y., Guo, X., & Wagner, A. (2024). High temperature delays and low temperature accelerates evolution of a new protein phenotype. *Nature Communications*, *15*(1), 2495. https://doi.org/10.1038/s41467-024-46332-6

Zou, Y., Zohner, C. M., Averill, C., Ma, H., Merder, J., Berdugo, M., Bialic-Murphy, L., Mo, L., Brun, P., Zimmermann, N. E., Liang, J., de-Miguel, S., Nabuurs, G.-J., Reich, P. B., Niinements, U., Dahlgren, J., Kändler, G., Ratcliffe, S., Ruiz-Benito, P., … Crowther, T. W. (2024). Positive feedbacks and alternative stable states in forest leaf types. *Nature Communications*, *15*(1), 4658. https://doi.org/10.1038/s41467-024-48676-5

Zuanny, D. C., Vilela, B., Moonlight, P. W., Särkinen, T. E., & Cardoso, D. (2024). expowo: An R package for mining global plant diversity and distribution data. *Applications in Plant Sciences*. https://doi.org/10.1002/aps3.11609

Figure legend

Figure 1. Database selection flowchart, adapted from the PRISMA statement flowchart (Page *et al.*, 2021).

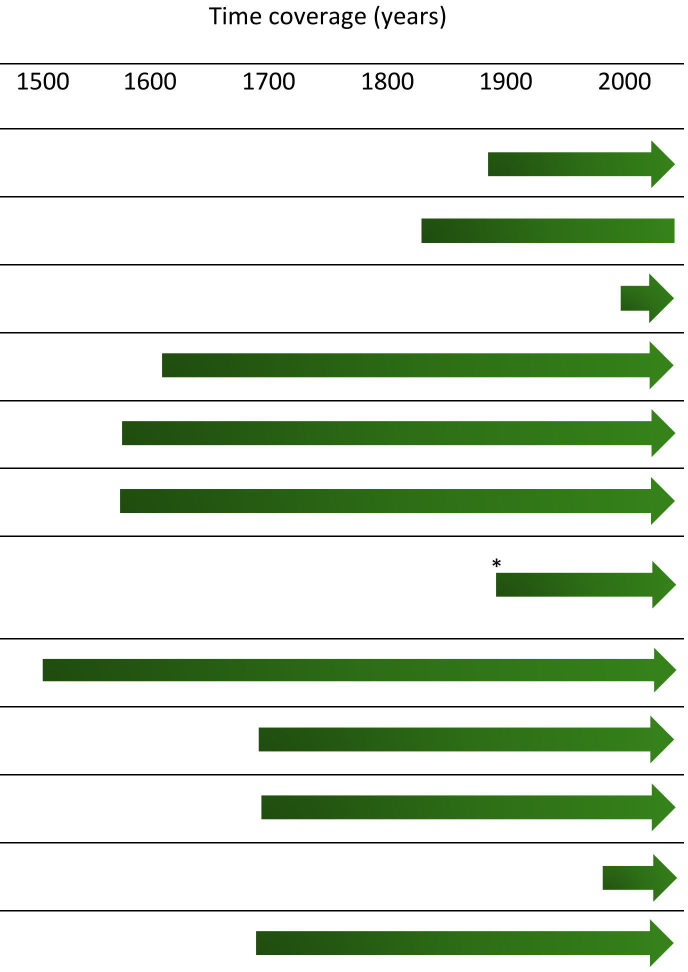
Figure 2. Interconnectedness of biodiversity platforms.

Figure 3. Biodiversity analysis methodology and tools. EV= environmental variables, INLA= integrated nested Laplace approximation, IPP= inhomogeneous Poisson process model, DNNs= Deep Neural Networks, SDM= Species Distribution Modeling, iSDM= integrative Species Distribution Modeling.

Tables

Table 1. Open access platforms with plant data on a global scale.

Portal	Taxon	Data type	Download: Reg	Download: Options	Download: Format
BIEN	*Plantae*	O, Tr, G, F, ED	yes	API**	NS
BioTIME	all	Ts	yes	D	CSV
CoL	all	TC	no	API, ColDP, DwC-A	CSV, TSV, TextTree
EOL	all	T, Tr, R	no	API	JSON, CSV
ForestForward	*Plantae*	O, G	no	D	CSV
GBIF	all	*ED*, O, *T*, *G*, S, I	yes	API, DwC-A	CSV
GIFT	*Plantae*	O, TC, G, Tr, L, E, B, S, Sc	no	API**	NS
MOL	all	O, TC, G, M	yes	D	CSV
POWO	*Plantae*	TC, *G*, D, R, L, I	no	DwC-A	CSV
Tropicos	*Plantae*	TC, O, G, R, I	yes	API, DwC-A	CSV
TRY	*Plantae*	Tr, T, E, ED, G, R	yes	D	CSV, XLSX, PDF, PNG, TXT
WFO	*Plantae*	TC, S, D, I, G, R	no	API, DwC-A	CSV, TXT



O= species occurrences, T= taxonomy, TC= Taxonomic checklist, G= geographic distribution, S= species status, I= images, D= descriptions, R= references, L= lifeform, Tr= trait data, E= environmental variables, ED= event date, F= phylogenies, Ts= time series of species abundances, B= bioclimatic data, Sc= Socio-economy data, M= distribution maps of specie richness, rarity, and alien species, Reg= registration, API= application programming interface, D= directly from the portal, CoLDP= CoL data package, DwC-A= Darwin Core Archive, NS= not specified. *Species names derived from heterogeneous resources and published over a timespan of about one hundred years (Weigelt *et al.*, 2019). **Required: BIEN R package (Enquist *et al.*, 2016) and GIFT R package (Denelle *et al.*, 2023). Italicized letters indicate incompleteness of information.

974

Table 2. Standardization Frameworks and Taxonomic Conflict Resolution Mechanisms in Global Biodiversity Databases.

Portal	Standardization/Integration Framework	Taxonomic/Nomenclature conflict resolution	Checklist/Tool used	Reference
BIEN	DwC, VegX	TNRS	WFO, Tropicos	Enquist *et al.* (2016)
BioTIME	Table system (csv, SQL)	CP	-	Dornelas *et al.* (2018)
CoL	COLDP, DwC-A, ACEF, Unique Identifiers (DOIs)	Multiple CA	WCVP, WFO, ChecklistBank	Döring & Ower (2019; Hobern *et al.* (2021)
EOL	DwC, DwC-A, DwC- M, DC	Multiple CA	CoL, Tropicos, WFO	Parr *et al.* (2016)
FF	ETL	-	GBIF	Tejada-Gutiérrez *et al.* (2024)
GBIF	DwC, EML	TBS	CoL, WCVP, Tropicos, WFO	Wieczorek *et al.* (2012)
GIFT	DwC, HC	TNRS	WCVP, WFO, Tropicos	Denelle *et al.* (2023); Weigelt *et al.* (2019)
MOL	DwC, EML, HC	CA	CoL	Jetz *et al.* (2012); Wieczorek *et al.* (2012)
POWO	DwC	TBS	CoL, GBIF	Govaerts *et al.* (2021)
TRY	EAV, OBOE	TNRS	WFO, Tropicos	Kattge *et al.* (2020)
WFO	DwC, Unique Identifiers (WFO ID), DwC-A, Dublin Core	TBS, TENs	Botalista	Borsch *et al.* (2020); Palese *et al.* (2019)

ACEF= Annual Checklist Exchange Format, CA= Checklist Approach, COLDP= CoL Data Package, CP= curation process, DC= Dublin core, DwC= Darwin core, DwC- A= Darwin core archive, DwC-M= Darwin core MeasurementOrFact extension, EML= Ecological Metadata Language, ETL= Extract, Transform, Load processes, EAV= Entity-Attribute- Value, FF= ForestForward, HC= Humboldt Core, VegX= Vegetation data exchange.

Table 3. Integration of trait data standards and ontologies in global biodiversity databases.

Integration Standards/ Ontologies	Key Features	Portal	References
OBOE, TOP, PTH	MS, TV, P	TRY, BIEN, GFIT	Garnier *et al.* (2017); Pérez-Harguindeguy *et al.* (2013); Kattge *et al.* (2020)
OBO (TO, PATO, ENVO)	SVS	EOL (TraitBank)	Baker *et al.* (2014) ; Parr *et al.* (2016)

ENVO= Environment Ontology, MS= Metadata Schema, OBO= Open Biological and Biomedical Ontologies, OBOE= Extensible Observation Ontology, PTH= Plant Trait Handbook, PATO= Phenotypic Quality Ontology, P = Protocols, SVS= Structured Vocabulary Systems, TO= Plant Trait Ontology, TOP= Thesaurus of Plant Characteristics, TV= Trait Vocabulary.

Table 4. Sources for environmental data.

Source	Data	Resolution (km)	Web site	Time period (years) 1800 1900 2000
GloH2O	pp, Tº, W, sH, Lc	10	https://www.gloh2o.org	1980 →
NOAA	Tº, pp, w, p, g	25- 250	https://www.noaa.gov/	1880 →
NASA	Tº, pp, g, ae, p, Lu, Lc, tp	25- 100 *	https://www.earthdata.nasa.gov/	1880 →
ECMWF	Tº, pp, w, r, g, H, Lc	10	https://cds.climate.copernicus.eu/#!/home	1950 →
WorldClim	Tº, pp, a, tp, B	1- 5	https://www.worldclim.org/	1950 – 2000**
Climate Explorer	Tº, pp, p, W, g	250 *	https://climexp.knmi.nl/selectfield_obs2.cgi?id=someone@somewhere	1850 →
ESA	SI	0.01-0.06	https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2	2015 →
CHELSA	Tº, pp, B	1	https://chelsa-climate.org/	1979 – 2013
USGS	SI	0.03	https://www.usgs.gov/landsat-missions/landsat-data-access	1972 →

GloH2O= Köppen-Geiger Maps, NOAA= National Centers for Environmental Information, NASA= National Aeronautics and Space Administration, ECMWF= European Centre for Medium-Range Weather Forecasts (The Copernicus programme), ESA= European Space Agency- Sentinel 2 (Copernicus programme), USGS= United State Geological Survey- Landsat (NASA), CHELSA= Climatologies at high resolutions for earth's land surface areas.

Pp= precipitation, Tº = air temperature, w= wind, W= water levels/sea levels, sH= soil moisture, Lc= land-cover, Lu= land-use, p= atmospheric pressure, g= greenhouse gases (CO2, CH4), ae = aerosols, tp= topography, r= radiation, H= humidity, a= altitude, B= bioclimatic variables, SI= satellite images.

* Varies by dataset, ** Projections up to 2100.

Table 5. Summary of common diversity indices

Index Name	Equation	Key Features	Reference
Margalef	$R_M = \frac{(S-1)}{ln\,(n)}$	Corrects richness by sample size using the natural log of total individuals.	Margalef (1958)
Menhinick	$R_{Mn} = \frac{S}{\sqrt{n}}$	Relates number of species to the square root of total individuals.	Menhinick (1964)
Shannon-Wiener	$H' = -\sum_{i=1}^{S} (p_i\, ln\, ln\, p_i)$	Consider both richness and evenness. Sensitive to rare species.	Shannon & Weaver (1949)
Simpson	$D_{Si} = \sum_{i=1}^{S} p_i^2$	Probability that two individuals randomly selected belong to the same species.	Simpson (1949)

S= number of species, n= number of individuals per species, p_i = proportion of individuals of species *i.*

Figure 1

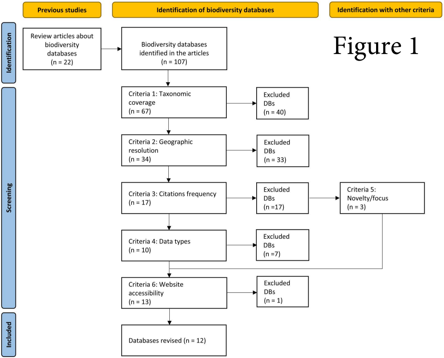

Previous studies
Identification of biodiversity databases
Identification with other criteria
Identification
Screening
Included
Review articles about biodiversity databases (n = 22)
Biodiversity databases identified in the articles (n = 107)
Criteria 1: Taxonomic coverage (n = 67)
Excluded DBs (n = 40)
Criteria 2: Geographic resolution (n = 34)
Excluded DBs (n = 33)
Criteria 3: Citations frequency (n = 17)
Excluded DBs (n =17)
Criteria 5: Novelty/focus (n = 3)
Criteria 4: Data types (n = 10)
Excluded DBs (n =7)
Criteria 6: Website accessibility (n = 13)
Excluded DBs (n = 1)
Databases revised (n = 12)

Figure 2

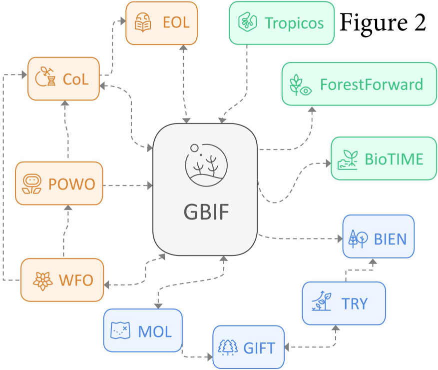
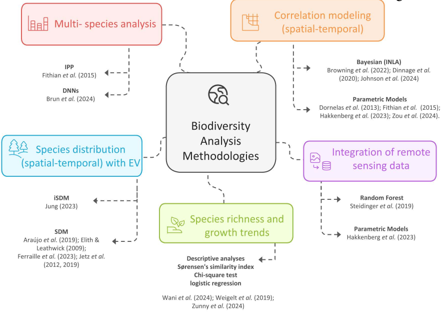

EOL
Tropicos
CoL
ForestForward
BioTIME
POWO
GBIF
BIEN
WFO
TRY
MOL
GIFT

Figure 3


Biodiversity Analysis Methodologies
Multi- species analysis
IPP
Fithian *et al.* (2015)
DNNs
Brun *et al.* (2024)
Correlation modeling (spatial-temporal)
Bayesian (INLA)
Browning *et al.* (2022); Dinnage *et al.* (2020); Johnson *et al.* (2024)
Parametric Models
Dornelas *et al.* (2013); Fithian *et al.* (2015); Hakkenberg *et al.* (2023); Zou *et al.* (2024).
Species distribution (spatial-temporal) with EV
iSDM
Jung (2023)
SDM
Araújo *et al.* (2019); Elith & Leathwick (2009); Ferraille *et al.* (2023); Jetz *et al.* (2012, 2019)
Integration of remote sensing data
Random Forest
Steidinger *et al.* (2019)
Parametric Models
Hakkenberg *et al.* (2023)
Species richness and growth trends
Descriptive analyses
Sørensen's similarity index
Chi-square test
logistic regression
Wani *et al.* (2024); Weigelt *et al.* (2019); Zunny *et al.* (2024)