METRIN-KG: A knowledge graph integrating plant metabolites, traits and biotic interactions

Disha Tandon[1*+], Tarcisio Mendes De Farias[2*], Pierre-Marie Allard[2,3*], Emmanuel Defossez[1*]

[1]Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland

[2]SIB Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland

[3]Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland

[*]Corresponding authors

[+]Current address: European Bioinformatics Institute (EMBL-EBI), CB10 1SD Hinxton, UK

Correspondence to:

Disha Tandon: dtandon@ebi.ac.uk

Tarcisio Mendes De Farias: tarcisio.mendes@sib.swiss,

Pierre-Marie Allard: pierre-marie.allard@unifr.ch,

Emmanuel Defossez: emmanuel.defossez@unine.ch

Abstract

Background

In recent years, biodiversity data management has emerged as a critical pillar in global conservation efforts. Today, the ability to efficiently collect, structure, and analyze biodiversity data is central to breakthroughs in conservation, drug development, disease monitoring, ecological forecasting, and agri-tech innovation. However, due to the vastness and heterogeneity of biodiversity data, it is often confined to databases for specific research areas in isolated formats and disconnected from other relevant resources. Crucial components of such data in kingdom Plantae comprise of metabolomes - the vast array of compounds produced by plants; traits - measurable characteristics of plants that influence their growth, survival, and reproduction, and that affect ecosystem processes; and biotic interactions - relationships of plants with other living organisms, affecting the ecosystem functions.

Results

In this work, we present METRIN-KG (MEtabolomes, TRaits, and INteractions-Knowledge Graph) a powerful data structure simplifying the integration of diverse and heterogeneous data resources such as plant metabolomes, traits and biotic interactions.

Conclusions

The proposed knowledge graph provides an interface to interactively search for data relating plant metabolomes, traits and interactions. This, in turn, will facilitate development of research questions in life-sciences. In this context, we provide representative case studies on how to frame queries that can be used to search for relevant data in the knowledge graph.

Keywords: Knowledge Graphs, Plant Metabolomes, Plant Traits, Biotic Interactions

Context

All species of our planet are connected by means of shared resources for sustenance, response to environmental effects, and multi-level biotic and abiotic interactions. An in-depth understanding of this intricate and multi-scale network of interactions enables researchers from natural sciences, chemistry, microbiology, ecology, plant biology and climate change to address critical questions in ecology, biodiversity conservation, agriculture and human health. As referenced by studies before [1–3], there are seven main shortfalls of biodiversity knowledge. Amongst these shortfalls, Raunkiæran and Eltonian, emphasize *lack of connected knowledge on species' traits and interactions, as well as their corresponding relation to ecological functions*. The chemistry of life, by governing biotic interactions, resource access, environmental adaptation and individual phenotype, provides a key link to building the mechanistic network required to understand species' processes underlying ecosystem functioning [4].

The metabolome, a key component of the chemistry of life, refers to all metabolites forming the substrates or products of enzymatic reactions in an organism. To a certain extent, the metabolome provides a bridge between the ecosystem functioning and the contextual information of species' states across multiple scales (spatial, temporal and environmental) [5–7]. In the Plantae kingdom, the sheer magnitude of metabolite diversity produced by plants is estimated to be between 1.5 and 25.7 million collectively for 400,000 plants [8]. This tremendous chemical diversity poses significant challenges for metabolomic data analysis [9,10].

It has been shown that metabolomes can provide a valuable proxy for the estimation of the plant functions, thus revealing the cause-and-effect relationship between external environmental factors and plant fitness & functions [9]. However, exploring the metabolome structure across spatial, temporal and environmental scales requires extensive studies to unravel the interplay of hundreds of compound classes and their potential biological and ecological functions. Since experimental approaches alone cannot provide such comprehensive datasets, aggregating data across multiple studies that combine metabolome and physiological or environmental data appears as a

necessary solution. Such information can include functional data like traits, co-dependence data like interactions, and omics-based data like genomics as well as transcriptomics.

Plant ecology focuses on how an organism's traits influence its interactions with and in response to environmental conditions throughout its life cycle. Traits like plant height, seed mass, leaf area, leaf carbon, nitrogen and phosphorus contents have been used to define plant processes across taxonomic (tree of life) [11–13], spatial (e.g.: subalpine regions) [14], environmental (e.g.: weather changes) [15], and temporal (e.g.: reproduction during the life of plant or life history) scales [15–17]. However, the mechanistic understanding behind the effect of traits on plant fitness and functioning is not clear [18–21] because of the unexplained or ambiguous relation of traits and variation in ecosystem functioning.

Plant biotic interactions, including relationships with other plants, fungi, bacteria, and soil organisms, have been shown to influence ecosystem structure and function, ultimately shaping biodiversity patterns [22–24]. As sessile organisms, plant biotic interactions also rely heavily on chemical mediation, involving both volatile and non-volatile compounds to compensate for their lack of mobility. Such interactions, particularly those involving insects, such as defense mechanisms or pollinator attraction, have been closely linked to the diversity of plant metabolites through evolutionary processes [25–29].

In recent years, efforts have been made to include chemistry into the functional traits framework to resolve plant functions in ecosystems [30,31]. Many studies have tried to combine the traits with specific compounds or classes of compounds [30–34], biotic interactions with traits [35–38], and biotic interactions with metabolomes [39–42]. In addition, other studies have combined the three to decipher ecosystem functioning, for instance, plant traits, their interactions with soil biota and related chemodiversity [43,44], effect of insect herbivory on plant traits and their secondary metabolite concentrations [45–47], links between plant root traits to nutrient foraging and chemodiversity as well as their symbiotic relationship with mycorrhizal fungi [48–50], allelopathic interactions and nutrient mobilization during intercropping in agriculture [51,52], plant root traits,

exudates (metabolites produced by roots) and its interactions with rhizosphere microbiome [53], links between biotic interactions, elevation gradients and metabolomics [54], climate-induced plant host shift in insects [55], plant microbe interactions and chemical defense mechanisms [56], and finally, insect herbivory and defense mechanisms facilitated by secondary metabolites [57–59]. Moreover, several reviews published during the last decade, have stressed the importance of combining plant traits, interactions and metabolomes [33,60–67]. The main challenge in studies that combine these three components is the high dimensionality of each, coupled with limited resources available to manage such complexity.

The experimental approaches commonly used in functional ecology are powerful tools for exploring and describing specific mechanisms. However, they remain limited when it comes to disentangling complex processes operating at larger scales. Linking chemical pathways, compound classes, or molecular structures to ecological functions remains a major challenge in ecological research. Recent development of databases and advances in data science offer a new potential to reach this goal. A data-driven approach is required to first map the existing knowledge in all three areas followed by the development of research hypotheses. Building on the few previous studies in this direction, we explored various datasets and databases related to traits, interactions, and metabolomics, which ultimately inspired the core idea of this study.

Much information on the heterogeneity of metabolites biosynthesized by organisms lies locked in isolated tabular formats, excel sheets and pdf documents (as discussed in [68]) and few databases provide comprehensive information on metabolites or natural products [68–72]. The Earth Metabolome Initiative (EMI) [73] and its pilot project Digital Botanical Gardens Initiative (DBGI) [74], were launched in 2022. They aim to document metabolic content for all known species on Earth (initiating point being botanical gardens living collections), following the Findable, Accessible, Interoperable, Reusable (FAIR) guidelines [75]. Under the EMI umbrella, comes the Experimental Natural Products Knowledge Graph (ENPKG) [69], a published resource of metabolomes from 1600 tropical plant extracts. ENPKG uses a sample-centric approach comprising semantic annotations to structure large, heterogeneous metabolomics datasets into

knowledge graphs. It also enables harmonization of experimental data with publicly available resources. However, no ecological metadata was integrated in ENPKG.

Like metabolite data, multi-level interaction data also lies locked in the guise of pairwise correlation metrics in peer-reviewed research papers. While Global Biotic Interactions (GloBI) [76] and tools like BiotXplorer [77] have succeeded in collecting this information in a FAIR resource [78], they allow only partial multi-taxonomic level interaction mapping and are limited to pairwise interactions. Similarly, data on plant trait heterogeneity is limited to few resources like TRY [79,80] and global plant trait network [81].

In addition, there are individual databases listing organism traits [79,82], metabolite-pathway relations of limited number of plant species [83] as well as microbes [84,85], pairwise interaction maps [76], medicinal herbs [86] and food [87]. However, there is no single resource available that combines knowledge across organisms, detailing their traits, interactions, and their complete/partial metabolomes, all of which are important for the in-depth understanding of the complex network of life.

In this Data Note, we present the first efforts in combining publicly available data on plant traits, interactions and metabolome from peer-reviewed research and databases in MEtabolomes, TRaits, and INteractions-Knowledge Graph (METRIN-KG). We have linked enriched metabolome datasets behind ENPKG [69] with high dimensional data on plant traits from TRY database [79], pairwise interaction data from GloBI [76], and annotated data on natural products from LOTUS [68]. To ensure semantic interoperability, we have used ontologies for knowledge representation like the Earth Metabolome Ontology [88,89], and the ENPKG ontology for natural product-specific concepts [69]. We used Ontop [90,91] and Python rdflib library [92] to materialize the Resource Description Framework (RDF) triples. We further implemented a SPARQL Protocol and RDF Query Language (SPARQL) editor to query METRIN-KG and retrieve results.

We anticipate METRIN-KG to be useful for interactively searching related information on traits, interactions and metabolome, thus guiding and inspiring the development of future research questions in the fields of ecology, biodiversity conservation, agriculture and human health. Moreover, we believe this information could also be valuable to a

wider audience beyond researchers, such as policymakers and public health professionals, by enhancing their understanding of the broader implications of research within their respective fields.

In the following sections we provide details on how the datasets used in METRIN-KG were retrieved, structured and linked. We also provide a brief overview of the construction of the Earth Metabolome Ontology, discuss potential reuse of METRIN-KG, and present representative queries for exploring the knowledge graph.

Methods

Data retrieval from TRY database

For this study, the pilot dataset was retrieved from the TRY website by requesting data for 41 traits based on the most used functional trait categories in plant physiology studies [30] - plant height, seed mass, leaf area, leaf carbon, nitrogen, and phosphorus contents (**Table-1**). Scientific names of the species from this dataset were mapped to Wikidata as listed in the section 'Taxonomy mapping'. The retrieved data is archived publicly in file 'TRYdb_40340.txt.gz' at https://doi.org/10.5281/zenodo.16874620.

Table-1: List of traits used for retrieving data from TRY database

Trait identifier in TRY	Trait Name
1	Leaf area
3108	Leaf area (in case of compound leaves: leaf, petiole excluded)
3109	Leaf area (in case of compound leaves: leaflet, petiole excluded)
3110	Leaf area (in case of compound leaves: leaf, petiole included)
3111	Leaf area (in case of compound leaves: leaflet, petiole

	included)
3112	Leaf area (in case of compound leaves: leaf, undefined if petiole in- or excluded)
3113	Leaf area (in case of compound leaves: leaflet, undefined if petiole is in- or excluded)
3114	Leaf area (in case of compound leaves undefined if leaf or leaflet, undefined if petiole is in- or excluded)
11	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA)
3115	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA): petiole excluded
3116	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA): petiole included
3117	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA): undefined if petiole is in- or excluded)
3085	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA) of leaf lamina
3086	Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA) petiole, rhachis and midrib excluded
3106	Plant height vegetative
3107	Plant height generative
26	Seed dry mass
3660	Seed fresh mass
3836	Seed mass per fruit
4	Stem specific density (SSD, stem dry mass per stem fresh volume) or wood density
3452	Stem specific density (SSD, stem dry mass per stem fresh

	volume) or wood density: heartwood
3764	Stem specific density (SSD, stem dry mass per stem fresh volume) or wood density: stem
13	Leaf carbon (C) content per leaf dry mass
14	Leaf nitrogen (N) content per leaf dry mass
15	Leaf phosphorus (P) content per leaf dry mass
12	Leaf lifespan (longevity)
663	Leaf respiration rate in light per dry mass
512	Leaf respiration rate in light per leaf area
664	Leaf respiration rate in light per leaf respiration rate in the dark
2943	Leaf respiration rate in the dark as a fraction of photosynthetic carboxylation capacity (Vcmax)
665	Leaf respiration rate in the dark minus respiration in light, mass based
54	Leaf respiration rate in the dark per leaf area
41	Leaf respiration rate in the dark per leaf dry mass
500	Leaf respiration rate in the dark per leaf nitrogen (N) content
272	Leaf respiration rate in the dark temperature dependence
53	Photosynthesis rate per leaf area
974	Photosynthesis rate per leaf area: transition to TPU limited photosynthesis
40	Photosynthesis rate per leaf dry mass
135	Photosynthesis rate per leaf nitrogen (N) content (photosynthetic nitrogen use efficiency, PNUE)
134	Photosynthesis rate per leaf transpiration (photosynthetic water use efficiency: WUE)

3128	Photosynthesis rate per stomatal conductance

Abbreviations - SLA: Specific Leaf Area; LMA: Leaf dry Mass per Area; SSD: Stem Specific Density; TPU: Triose Phosphate Utilisation; PNUE: Photosynthetic Nitrogen Use Efficiency; WUE: Water Use Efficiency

Data retrieval from GloBI

The compressed stable release of interaction data was downloaded from the Zenodo archive of GloBI, version from January 2025 [93], representing 'species interactions tabulated as pairwise interactions in a zipped tab-separated values format.' (as described on the GloBI website). The creators of GloBI mention that 'included taxonomic names are not interpreted, but included as documented in their sources'. The taxonomic identifiers and names were mapped to Wikidata as listed in the section 'Taxonomy mapping'. The data is available publicly in file 'verbatim-interactions.tsv.gz' at https://zenodo.org/records/14640564.

Taxonomy mapping

GloBI collates interaction information from standardized data from numerous sources including online repositories and projects (e.g.: Encyclopedia of Life) as well as data entered by individual projects through GloBI's dataset-template [94]. The full list of data sources is updated on its website [95]. The combination of these two approaches resulted in taxonomic identifiers from around 15 taxonomy databases (**Supplementary Table 1**) to be present in GloBI. The initial taxonomic mappings from these 15 resources as well as their scientific names to Wikidata identifiers were done using QLever [96] SPARQL user interface (UI) for Wikidata [97]. SPARQL queries used for mapping are as follows:

a) Query for mapping Wikidata identifiers to ones from 15 taxonomy databases (**Supplementary Table 1**)

```
`PREFIX wdt: <http://www.wikidata.org/prop/direct/>

PREFIX wd: <http://www.wikidata.org/entity/>

SELECT ?WdID ?eol ?gbif ?ncbi ?ott ?itis ?irmng ?col ?nbn ?worms ?bold ?plazi ?apni ?msw3 ?iNat ?eppo ?WdName WHERE {

    ?WdID wdt:P31 wd:Q16521;

            wdt:P225 ?WdName .

    OPTIONAL { ?WdID wdt:P9157 ?ott . }

    OPTIONAL { ?WdID wdt:P685 ?ncbi . }

    OPTIONAL { ?WdID wdt:P846 ?gbif . }

    OPTIONAL { ?WdID wdt:P830 ?eol . }

    OPTIONAL { ?WdID wdt:P815 ?itis . }

    OPTIONAL { ?WdID wdt:P5055 ?irmng . }

    OPTIONAL { ?WdID wdt:P10585 ?col . }

    OPTIONAL { ?WdID wdt:P3240 ?nbn . }

    OPTIONAL { ?WdID wdt:P850 ?worms . }

    OPTIONAL { ?WdID wdt:P3606 ?bold . }

    OPTIONAL { ?WdID wdt:P1992 ?plazi . }

    OPTIONAL { ?WdID wdt:P5984 ?apni . }

    OPTIONAL { ?WdID wdt:P959 ?msw3 . }

    OPTIONAL { ?WdID wdt:P3151 ?iNat . }

    OPTIONAL { ?WdID wdt:P3031 ?eppo . }

  }`
```

b) Query for retrieving Wikidata identifiers and their lineage

```
`PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
SELECT ?WdID ?WdName ?hTax ?hTaxName ?hTaxRank WHERE {
    ?WdID wdt:P31 wd:Q16521;
        wdt:P225 ?WdName ;
        wdt:P171* ?hTax .
    ?hTax wdt:P225 ?hTaxName ;
        wdt:P105 ?hTaxRank .
}`
```

The retrieved data is available publicly in 'wdTax' files at https://doi.org/10.5281/zenodo.16874620.

For the TRY database, the scientific names were directly matched to ones obtained from Wikidata, to obtain corresponding Wikidata identifiers.

Metadata mapping

GloBI provides limited mappings of life stages, and body parts to Uber-Anatomy Ontology (UBERON), Plant Ontology (PO), Environment Ontology (ENVO), Gene Ontology (GO; body parts only) and Phenotype and Trait Ontology (PATO; body parts only) for organisms involved in interactions (**Supplementary Table 2**). For biological sex, the names are provided in GloBI, but no specific mapping to existing ontologies.

To semantically match input terms (life stages, body parts and biological sex) with ontology concepts, we developed a pipeline using ontology parsing and sentence embeddings. We utilized the Python Owlready2 library version-0.47 [98] to load a suite

of biomedical and environmental ontologies, namely the ones originally present in GloBI - UBERON, PO, ENVO, GO, PATO, as well as others including Fungal Anatomy Ontology (FAO), Hymenoptera Anatomy Ontology (HAO), BRENDA Tissue Ontology (BTO), National Cancer Institute Thesaurus (NCIT), Agronomy Ontology (AO), Ascomycete Phenotype Ontology (APO), Biological Spatial Ontology (BSPO), Cephalopod Ontology (CEPH), Collembola Anatomy Ontology (CLAO), Dictyostelium Discoideum Anatomy (DDANAT), Ecology Core Ontology (ECOCORE), Experimental Factor Ontology (EFO), Foundational Model of Anatomy (FMA), ontology for plant gall phenotypes (GallOnt), Infectious Disease Ontology (IDO), Infectious Disease Ontology for Malaria (IDOMAL), Ontology for Biomedical Investigations (OBI), Ontology for MIcroRNA Target Prediction (OMIT), Pathogen-Host Interaction Phenotype Ontology (PHIPO), Porifera Ontology (PORO), Mosquito gross anatomy ontology (TGMA), Chemical Entities of Biological Interest (ChEBI), Cell Ontology (CL), Flora Phenotype Ontology (FLOPO), Food Ontology (FOODON) from their publicly accessible Uniform Resource Locators (URL) (see also **Supplementary Table 2**). For life stages and body parts, the full suite was considered, whereas for biological sex, only UBERON and PATO were considered.

From each ontology, we extracted class labels and associated synonyms (including exact, broad, and related synonyms where available) as candidate terms for matching. Synonyms were collected from Web Ontology Language (OWL) annotation properties such as `hasExactSynonym`, `hasBroadSynonym`, and `hasRelatedSynonym`. Each candidate term was paired with its primary label and class Internationalized Resource Identifier (IRI) for traceability.

To compute semantic similarity, we used the pre-trained `all-MiniLM-L6-v2` model from Python SentenceTransformers library version-3.3.1 [99]. This model is a lightweight transformer trained to generate 384-dimensional dense vector embeddings that reflect the semantic content of short text spans. It is particularly suited for tasks such as semantic textual similarity and clustering.

All ontology terms (labels and synonyms) were encoded into dense vector representations using this model. Similarly, each user-provided input term for unmapped

life stages, body parts and biological sex was independently embedded into the same vector space. Term embeddings were computed using the model's `encode()` method with `convert_to_tensor=True`, which produced PyTorch-compatible tensors suitable for high-performance vector operations.

Cosine similarity was computed between each input term vector and all ontology term vectors using the `util.pytorch_cos_sim` function in SentenceTransformers. For each input term, the ontology term with the highest cosine similarity score was identified as the top candidate match. Cosine similarity is a common metric for comparing the orientation, rather than magnitude of two vectors. Given two vectors A and B, cosine similarity is defined as:

$$cosine_similarity(A, B) = \frac{A.B}{|A|.|B|}$$

Values range from -1 to 1, where 1 indicates identical direction (i.e., maximum similarity), and 0 indicates orthogonality (no similarity). For each input term, we computed its similarity against all ontology terms. The term with the highest similarity score was selected as the top candidate match.

All results were output to a comma-separated values (CSV) file, including the input term, matched ontology label, its primary label, the ontology class IRI, and the similarity score. While the automated process provided high-quality initial suggestions, all matches were manually reviewed and corrected to ensure semantic appropriateness. Corrections were informed by domain-specific knowledge and contextual relevance, particularly in cases where terms had multiple meanings or where high similarity scores did not guarantee ontological alignment.

This two-stage process, automated semantic matching followed by manual curation, ensured both scalability and accuracy in aligning non-standardized input terms with controlled ontology concepts.

The above workflow was also run for matching the units of measure for trait data from TRY to the units vocabulary provided by Quantities, Units, Dimensions, and Types (QUDT 2.1 schema) [100], followed by manual correction.

The mapped data is archived in the folders 'globi' and 'trydb' under processed data at https://doi.org/10.5281/zenodo.16874620.

Ontology

EMI [73] is a global effort to profile the metabolic content of all currently known species on our planet. Here, knowledge representation plays a key role to correctly capture, contextualize and structure the vast amount of chemical diversity data that have been and will be generated in the upcoming years. Consequently, this will facilitate data (re)use and interoperability. To accurately represent the EMI knowledge, we explored several general-purpose and domain-specific ontologies to design a framework to describe chemical compounds (e.g., natural products) and their related data such as geolocation, provenance, organism sample metadata, and organism interactions. The EMI ontology reuses and repurposes, where applicable, several other ontologies beyond the life sciences such as World Wide Web Consortium's (W3C) Sensor, Observation, Sample, and Actuator (SOSA) ontology [101], that was primarily designed for other applications including the Open Geospatial Consortium (OGC) use cases. OGC is a consortium that aims to improve access to geospatial and location information [102]. Furthermore, semantic reconciliation powered by biocuration is at the core of our proposed framework. For example, our approach can accommodate and harmonize different organismal, chemical and material sample taxonomies as well as vocabularies such as the Relation Ontology (RO) to define interactions between organisms (e.g., "has pathogen"). The other ontologies reused in EMI are those of Simple Knowledge Organization System (SKOS) [103], standard units of measures from QUDT [100], geo-locations from the basic geo (World Geodetic System 1984 (WGS84) lat/long) vocabulary [104], and natural product-specific concepts from the ENPKG ontology. The EMI ontology documentation is available online at https://w3id.org/emi. The code and a tutorial to build a knowledge graph with this ontology are available at https://github.com/earth-metabolome-initiative/earth_metabolome_ontology. **Figure-1** shows the EMI ontology schema.

Knowledge graph construction

Based on the EMI ontology as a data schema (**Figure-1**), we built a knowledge graph (KG) that semantically enriches, integrates and interoperates 3 data sources - enriched metabolome datasets behind the ENPKG, trait data from TRY-db and interaction data from GloBI. We developed the KG by combining subgraphs in two stages:

a) We applied the Ontop tool [90,105] for developing the subgraph from enriched datasets (metabolite annotation, molecular networks and taxonomical resolution results) originally used to build the ENPKG graph [106,107]. Ontop is a virtual knowledge graph system where SPARQL queries are translated into Structured Query Language (SQL) queries based on the predefined mappings. Ontop also provides means to materialize a KG according to these mappings. Therefore, to build the graph, we first created one table for each tabular file in a relational database with a simplistic data schema. We loaded the files in this database through mysql version-8.2 [108]. Finally, to construct the KG, we defined mappings between the relational schema and the EMI ontology using the Ontop mapping language [90]. With these mappings (https://github.com/earth-metabolome-initiative/earth_metabolome_ontology/blob/main/ontop_config/emi-v1/emi-v1.obda), the EMI ontology [88] and the relational database, we materialized the RDF triples with Ontop to compose the KG.

b) For developing the subgraphs from TRY database and GloBI, we used Python rdflib library version-7.0.0 [92,109]. To construct them, we used the tab-separated value (TSV) files obtained for each of these two resources. These tables were then connected through Wikidata identifiers (**Figure-2**) for taxonomy, whenever they were available as indicated in section 'Taxonomy mapping'. We defined mappings between the relational schema and the EMI ontology using rdflib's inherent capability to map ontology elements.

The process for incorporating mappings to metadata for GloBI (source/target taxonomy, life stage renamed to developmental stage, body part renamed to

anatomical entity, and biological sex) and TRY database (taxonomy and trait's units of measure) was performed before generating the RDF triples.

The RDF files can be downloaded at https://doi.org/10.5281/zenodo.16874620 in the folder 'KG' under processed data.

Indexing the knowledge graph and implementing the SPARQL end-point

We used Qlever [96], a SPARQL engine, to index our knowledge graph by implementing the steps described in qlever-control [110]. A SPARQL editor to query the indexed graph is available at https://kg.earthmetabolome.org/metrin/. This endpoint can also be accessed through https://kg.earthmetabolome.org/metrin/api for programmatic purposes. This interface includes several examples covering specific use-cases for the three datasets (ENPKG metabolites, TRY-db traits and GloBI interactions). Some of these use-cases are presented as case studies in the section 'Data re-use and case studies'.

To provide the SPARQL editor user-interface for the indexed graph, we implemented qlever-ui [111,112] with few changes. To improve the user experience when navigating entities in the knowledge graph, we extended the qlever-ui interface to support fallback representations for non-available or missing web pages. This was achieved by incorporating SPARQL DESCRIBE queries into the frontend of the UI. The modified interface attempts to retrieve a minimal RDF-based summary of the requested entity when the original resource is not accessible (e.g., due to missing data in the knowledge graph). When such a condition is met, the interface automatically issues a SPARQL DESCRIBE query for the corresponding IRI. The result is rendered in a simplified RDF triple view, providing context about the entity based on available data in the backend knowledge base. The implementation is available via a fork of the official qlever-ui repository here: https://github.com/earth-metabolome-initiative/qlever-ui.

Data Structure

Mapping of TRY data

The original structure of the TRY datasets is modelled as plant species mapping to trait and non-trait data as well as related metadata like information on studies from which the data was obtained, as shown in **Figure-3**. The column names depicted in the figure are taken directly from the CSV table retrieved from TRY, the description of which is provided in **Supplementary Table 3**. Columns retained in the knowledge graph are indicated in the table and the figure. Overall, 20,272,589 records for 70,748 unique plant species were retrieved in a TSV format. (**Supplementary Table 4**).

The dataset was refined to a minimal form by removing columns that contained redundant or ambiguous information. For example, author first and last names were excluded, as this information is already available in the retained 'Reference' field. Likewise, the 'OriginalName' column duplicated data present in 'DataName' and was therefore omitted. 'Replicates' is also an entity mentioned in 'DataName' and therefore not retained. The TRY database includes standardized values for certain traits in the columns 'ErrorRisk', 'StdValue', 'StdValueStr', 'OrigUncertainityStr', 'UncertaintyName', 'ValueKindName', and 'RelUncertainityPercent'. However, these were excluded due to their limited coverage across the dataset. This selective retention helped minimize redundancy and ensured consistency within the constructed knowledge graph. The metrics of the dataset following this refinement are as follows:

A) 65,672 unique species names ('AccSpeciesName'; see **Supplementary Table 4**) out of 70,748 from TRY database were mapped to Wikidata (see section Taxonomy mapping).
B) Overall 1,826,877 traits & 17,214,311 non-trait (e.g.: number of replicates, latitude, longitude) records, data values and corresponding units

(**Supplementary Table 4**) were retained after mapping 65,672 species to Wikidata identifiers.

Query to retrieve metrics: https://kg.earthmetabolome.org/metrin/9lkgto

Mapping of GloBI data

The original structure and column descriptions of the GloBI data are described in **Figure-4** and **Supplementary Table 5**. The column descriptions are also provided on the github account of GloBI as depicted on its website [113]. Columns retained in the knowledge graph are indicated in the table and the figure. Lineage information was not retained, as it can be retrieved via federated queries through Wikidata. Physiological stage data was omitted due to the absence of reliable mappings to established ontologies or controlled vocabularies. The 'eventDateTime' field was removed owing to ambiguity - specifically, it was not clear whether it referred to the timing of the interaction event or the time of observation.

Overall, 20,480,925 records were obtained from 1,747,254 unique taxonomic identifiers and/or names (**Supplementary Table 4**), out of which 609,087 were mapped to 337,293 unique Wikidata identifiers. Overall, 12,872,681 GloBI records for only the mapped taxonomic identifiers were retained.

Query for number of Wikidata identifiers: https://kg.earthmetabolome.org/metrin/Pbt0pN

Query for number of records: https://kg.earthmetabolome.org/metrin/nMswqb

At the time of writing this manuscript, 1867 body part names were mapped to 996 ontology terms. 621 life stage names were mapped to 227 ontology terms. 57 biological sex names were mapped to 10 ontology terms.

Data re-use and case studies

Potential re-use and expansion of data

The current state of the knowledge graph provides a solid foundation, and its utility can be significantly enhanced by incorporating additional data sources and refining its structure.

a) Data re-use for studies in life sciences research

Researchers can query METRIN-KG to identify potential relationships between plant traits, interactions, and the presence or bioactivity of specific natural products. For example, one could query for plants with specific traits known to be associated with defense mechanisms and then explore the natural products they produce. The integrated data can be used to build predictive models for natural product occurrence or bioactivity based on plant traits and ecological context. This information can help prioritize plant species or natural product classes with a higher likelihood of possessing desired bioactivities. In addition, by analyzing trait-interaction-metabolome associations, researchers can prioritize plant species or ecological contexts that are more likely to yield novel or interesting natural products. Investigating such ecological roles of natural products can provide clues about their potential mechanisms of action and possible drug targets.

METRIN-KG can also facilitate the investigation of how natural products mediate species interactions. For instance, one could explore how specific compounds are associated with pollination or defense against herbivores in plants with particular traits. The understanding of the evolutionary pressures that shape the diversity of natural products can be investigated by linking them with the respective host traits and interactions.

In the context of potential data re-use for studies in life sciences research, we aimed to retrieve data from METRIN-KG on a few known questions in plant ecology, agriculture and biodiversity. **Table-2** lists the questions and the respective links to the SPARQL queries. All example queries are considered by ExpasyGPT [114], an AI-driven tool that allows for querying knowledge graphs in natural language. This tool facilitates querying METRIN-KG in plain English as well as various other life-science databases. In other words, ExpasyGPT aids in constructing SPARQL queries in response to user's text-format questions. A tutorial to build queries using ExpasyGPT is provided here: https://github.com/earth-metabolome-initiative/metrin-kg/wiki/ExpasyGPT-for-METRIN%E2%80%90KG.

Table-2: Examples of SPARQL queries for METRIN-KG

Query#	Title	Subject	Direct link to queries	Versioned github link to queries
1 (CS1)	List interactions of all species which have an IUCN status of near threatened.	Conservation science	https://kg.earthmetabolome.org/metrin/Yl2y7K	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/011.ttl
2 (CS1)	List traits of all species which have an IUCN status of near threatened.	Conservation science	https://kg.earthmetabolome.org/metrin/dmjpx0	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/012.ttl
3 (CS1)	List all metabolites produced by species with near threatened IUCN status and with values available (or greater than a specific value) for trait 'Seed dry mass'.	Conservation science	https://kg.earthmetabolome.org/metrin/IIXa2u	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/013.ttl
4 (CS2)	List traits (and their values) of plants producing Diterpenoids	Functional ecology	https://kg.earthmetabolome.org/metrin/ShXRT0	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/014.ttl

5 (CS3)	Natural producers (and their interactions that might be useful in agriculture) of onopordopicrin (wd:Q27107580), which might exhibit antimicrobial and cytotoxic activities, especially against human-derived macrophages and against epidermoid carcinoma cells.	Human health	https://kg.earthmetabolome.org/metrin/u5gt2c	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/020.ttl
6 (CS4)	List links between host-parasite and host-allelopathic interactions, where host is an agricultural crop, allelopath is usually a plant that inhibits growth of parasites which are usually insects. Inhibition happens by root/stem/leaf exudates of the allelopath.	Sustainable agriculture	https://kg.earthmetabolome.org/metrin/kiYpYO	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/018.ttl
7 (CS5)	Retrieve data of 4 traits in leaf economics spectrum	Theoretical ecology	https://kg.earthmetabolome.org/metrin/h0hO3L	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/022.ttl
8 (CS5)	Retrieve metabolites of organisms with data available from 4 traits in leaf economics spectrum	Theoretical ecology	https://kg.earthmetabolome.org/metrin/inTU5f	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/023.ttl
9	List of possible natural locations of plants, which produce senkyunolide which is a thalide	Human health	https://kg.earthmetabolome.org/metrin/QgmRFC	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/016.ttl

	for cerebral disorders (https://doi.org/10.3390/molecules28083636)			
10	List metabolites of plants that interact with plant parasite moth Orgyia postica	Functional ecology	https://kg.earthmetabolome.org/metrin/Ayl4WM	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/017.ttl
11	A list of interactions depicting connections between parasitoids harmful for insects living as parasites on plants.	Functional ecology	https://kg.earthmetabolome.org/metrin/1DsKNN	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/019.ttl
12	List possible interactions of plants that can produce Norhyocyamine, a plant secondary metabolite	Functional ecology	https://kg.earthmetabolome.org/metrin/i3VQB0	https://github.com/sib-swiss/sparql-examples/blob/master/examples/emi/021.ttl

Abbreviations - CS: Case Study; IUCN: International Union for Conservation of Nature

b) User-friendly approach to add queries as examples within METRIN-KG

We devised a method for the users of METRIN-KG to propose a query and incorporate it within the examples (e.g. queries listed in **Table-2**) using the SPARQL query-editor proposed in [115]. Through the SPARQL query UI for METRIN-KG provided by the Swiss Institute of Bioinformatics (https://sib-swiss.github.io/sparql-editor/metrin-kg), users can write a query and propose for it to be included in the examples list by clicking on 'Save the current query as example' tab. Once the query has been proposed, a github repository issue will be created on https://github.com/sib-swiss/sparql-examples under the user's github handle, allowing us to review it for inclusion in the examples. Once the review and corrections are complete, the query will be available on both SPARQL query UIs : https://kg.earthmetabolome.org/metrin/ and https://sib-swiss.github.io/sparql-editor/metrin-kg. While the examples listed in **Table-2**

provide several complex queries, this option allows the users to provide context-specific queries useful for building research questions.

A tutorial for contributing queries to METRIN-KG is provided here: https://github.com/earth-metabolome-initiative/metrin-kg/wiki/Contribute-queries-to-METRIN%E2%80%90KG.

c) Adding data or federating over knowledge graphs presenting datasets from other publicly available datasets

Incorporating environmental data (e.g., climate data and soil data) could allow for the analysis of how environmental factors influence traits, interactions, and natural product production and distribution. Including data on the timing of biological events (e.g., flowering times) could add a temporal dimension to the analysis of interactions and natural product occurrence. In addition, incorporating image data (e.g., plant morphology) and spectral data (e.g., metabolomic profiles) could provide richer characterizations of the entities within the knowledge graph. Integrating broader ecological context, such as community composition and ecosystem dynamics, could provide a more holistic understanding of the relationships. Where ethically and legally appropriate, incorporating curated traditional knowledge related to plant uses and natural products could add valuable perspectives. Moreover, including data from other natural product databases (e.g.: PubChem [70], ChEMBL [71,116]) could broaden the coverage of chemical compounds and their properties.

d) Enhancing Ontological Structure and Semantics:

For developing more granular relationship types, moving beyond simple pairwise interactions to include more specific relationship types (e.g., pollination, herbivory, symbiosis) with associated properties (e.g., strength, specificity) could be useful. The first efforts in this direction are taken by GloBI [76] and BiotXplorer [77]. However, they are limited to pairwise interactions and not connected to other types of data, which our knowledge graph provides. A useful step in this direction could be to implement

semantic reasoning and hence infer new relationships and knowledge that are not explicitly stated in the data.

METRIN-KG holds the potential to be a central, semantically integrated resource for accessing and querying diverse datasets related to natural products. Through its common framework and vocabulary, it can facilitate knowledge sharing across various subfields in areas of ecology, human health and agriculture.

Case studies

Out of the 12 examples listed in **Table-2**, we built 5 case study (CS) summaries for 8 as listed below.

CS1: List traits, interactions and metabolites of all species which have an International Union for Conservation of Nature (IUCN) status of near threatened.

Subject - Conservation science

(https://kg.earthmetabolome.org/metrin/dmjpx0,

https://kg.earthmetabolome.org/metrin/Yl2y7K,

https://kg.earthmetabolome.org/metrin/IIXa2u)

Some studies have addressed the need to identify the threat status of plants by studying their functional traits [117–119]. Moreover, one study has suggested approaches for conservation of threatened plant species using functional trait patterns [120]. Some other studies have used predictive modelling to enhance knowledge of interacting species pairs (e.g.: predator-prey interactions) [121,122] in the context of their habitat and traits. In addition, some researchers have suggested and reviewed ecological restoration strategies for threatened ecosystems based on functional traits and plant-animal interactions [123–125].

Keeping such broad studies in mind, we developed 3 SPARQL queries retrieving the traits, interactions and metabolites of plant species with IUCN status of near threatened. Such queries would aid the researchers in collecting data relevant to near threatened plant species (in some cases, insects) and develop strategies to model patterns specific to climate change, thereby developing ecological restoration plans.

While these 3 queries can be combined into one, we kept them separate to improve efficiency and retain modularity.

CS2: List traits and their values for plants producing diterpenoids

Subject - Functional ecology

(https://kg.earthmetabolome.org/metrin/ShXRT0)

Recent studies have utilized trait data to understand the implications of diterpenoids producing invasive plant species and their allelopathic relationships with agricultural crops [126–131]. In the context of agriculture, there have been reviews of plant traits and their effects on insect herbivory [132], crop resilience (rice) of plants producing diterpenoids [133–135] and the implications of drought stress on diterpenoid producing plants and their traits [136]. In the context of biodiversity conservation, there have been studies on plant traits and their metabolic profile [137]. In addition, there has been research on plant interactions with microbial communities of plants where leaf surface has been reported to contain diterpenoids (tobacco and phyllosphere) [138].

Considering such diverse research on diterpenoids-producing plants or where diterpenoids are known to be extracted from plant surfaces, we developed a simple query to list traits of plants which have been known to produce or harbor diterpenoids.

CS3: Natural producers and their biotic interactions that might be useful in agriculture of onopordopicrin.

Subject - Human health

(https://kg.earthmetabolome.org/metrin/u5gt2c)

Onopordopicrin is a metabolite produced by plants like *Arctium lappa* and genus *Shangwua.* It has been shown to potentially exhibit antimicrobial and cytotoxic activities, especially against human-derived macrophages and against epidermoid carcinoma cells [139]. There is less scientific evidence to support these claims and research in this area has only recently been enhanced [140–143], making onopordopicrin an appropriate and timely subject for prospective research studies in human health and pharmacology.

To specifically target the research in this area, we developed a query to retrieve the natural producers of onopordopicrin and their biotic interactions.

CS4: Allelopathy interactions in push-pull agriculture

Subject - Sustainable agriculture

(https://kg.earthmetabolome.org/metrin/kiYpYO)

Studies on allelopathy have advanced our knowledge of plant interactions with other organisms, particularly in agriculture [144–146] and invasion ecology [126]. For instance, species of genera *Desmodium*, through their root and stem exudates, may protect crops of maize (*Zea mays*) and sorghum (*Sorghum*) against attack from stemborer arthropods and weed *Striga hermonthica* [147,148]. Besides focussed research [149], there are few studies exploring the allelopathic properties and the plant metabolites responsible for them.

We developed a query to target such interactions and the metabolites produced by the protective allelopathic plants.

CS5: Retrieve traits and metabolites for organisms for which traits from Leaf Economic Spectrum (LES) are available

Subject - Theoretical ecology

(https://kg.earthmetabolome.org/metrin/h0hO3L,

https://kg.earthmetabolome.org/metrin/inTU5f)

Leaf Economic Spectrum is a foundational concept in plant ecology, describing coordinated variation in leaf traits across species, typically forming a trade-off between resource-acquisitive and resource-conservative strategies [81,150]. Resource-acquisitive species have traits like high specific leaf area (SLA), high nitrogen content, short leaf lifespan, which are optimized for rapid growth in resource-rich environments. Resource-conservative species show traits like low SLA, low nutrient content, long leaf lifespan, which are better suited for resource-poor environments with conservative growth. These relationships are global, cross-taxonomic, and continuous, and they help predict ecosystem functions, species distributions, and responses to environmental changes.

We developed a query to retrieve values for 4 traits relevant for LES - specific leaf area, leaf dry matter content, leaf nitrogen content, and lead phosphorus content. Further, we developed another query to retrieve metabolites produced by species with data for four LES traits available in the knowledge graph.

Like CS1, these 2 queries can be combined into one, but we kept them separate for efficiency and modularity reasons.

At the time of writing this manuscript, the results of our queries reflect the state of METRIN-KG and the federated sources available. As these underlying datasets are continuously updated and expanded, the query outputs presented here may evolve over time.

Conclusions

We provide METRIN-KG aiming to integrate seemingly disconnected databases of plant metabolomes, plant traits and interactions. To facilitate exploration of this resource, we provide a way to search for those connections through a user interface. We aspire to

draw the attention of researchers in areas of drug-discovery, ecology, human health, biodiversity conservation and agriculture to the power of knowledge graphs in integrating life-sciences datasets. We wish to increase awareness amongst researchers to include as varied data as possible, so as to approach research questions from multiple perspectives. In this context, we provide representative case studies to retrieve data for different research projects. We plan to extend METRIN-KG to incorporate the full breadth of metabolome data from plants in the near future. In the distant long-term, we plan to extend it, adding more information on metabolite data for other kingdoms beyond Plantae, while updating the trait and interaction data from TRY and GloBI respectively as and when their new versions are released. We hope that our efforts will encourage other researchers to use and contribute to this resource.

Availability of source code

All data supporting this work are available at the github repositories of Earth Metabolome Initiative for the knowledge graph development (https://github.com/earth-metabolome-initiative/metrin-kg) and the earth metabolome ontology (https://github.com/earth-metabolome-initiative/earth_metabolome_ontology).

Data availability

All input and output files used in mapping taxonomy, metadata, ontology and knowledge graph construction as well as RDF files are available on Zenodo (https://doi.org/10.5281/zenodo.16874620).

Abbreviations

AO: Agronomy Ontology; APO: Ascomycete Phenotype Ontology; BRENDA: Braunschweig Enzyme Database; BSPO: Biological Spatial Ontology; BTO: BRENDA Tissue Ontology; CEPH: Cephalopod Ontology; ChEBI: Chemical Entities of Biological Interest; CL: Cell Ontology; CLAO: Collembola Anatomy Ontology; CS: Case Study; CSV: Comma-Separated Values; DBGI: Digital Botanical Gardens Initiative; DDANAT: Dictyostelium Discoideum Anatomy; ECOCORE: Ecology Core Ontology; EFO: Experimental Factor Ontology; EMI: Earth Metabolome Initiative; ENPKG: Experimental Natural Products Knowledge Graph; ENVO: Environment Ontology; FAIR: Findable, Accessible, Interoperable and Reusable; FAO: Fungal Anatomy Ontology; FLOPO: Flora Phenotype Ontology; FMA: Foundational Model of Anatomy; FOODON: Food Ontology; GallOnt: ontology for plant gall phenotypes; GloBI: Global Biotic Interactions; GO: Gene Ontology; HAO: Hymenoptera Anatomy Ontology; IDO: Infectious Disease Ontology; IDOMAL: Infectious Disease Ontology for Malaria; IRI: Internationalized Resource Identifier; IUCN: International Union for Conservation of Nature; KG: Knowledge Graph; LES: Leaf Economic Spectrum; LMA: Leaf dry mass per area; LOTUS: *not an acronym*; METRIN-KG: MEtabolomes, TRaits, and INteractions-Knowledge Graph; NCIT: National Cancer Institute Thesaurus; OBI: Ontology for Biomedical Investigations; OGC: Open Geospatial Consortium; OMIT: Ontology for MIcroRNA Target Prediction; OWL: Web Ontology Language; PATO: Phenotype and Trait Ontology; PHIPO: Pathogen-Host Interaction Phenotype Ontology; PNUE: Photosynthetic Nitrogen Use Efficiency; PO: Plant Ontology; PORO: Porifera Ontology; QUDT: Quantities, Units, Dimensions, and Types; RDF: Resource Description Framework; RO: Relation Ontology; SKOS: Simple Knowledge Organization System; SLA: Specific Leaf Area; SOSA: Sensor, Observation, Sample, and Actuator; SPARQL: SPARQL Protocol and RDF Query Language; SLA: Specific Leaf Area; SSD: Stem Specific Density; SQL: Structured Query Language; TGMA: Mosquito gross anatomy ontology; TPU: Triose Phosphate Utilisation; TRY: *not an acronym*; TSV: Tab-Separated Values; UBERON: Uber-Anatomy Ontology; UI: User Interface; URL: Uniform Resource

Locators; W3C: World Wide Web Consortium; WGS84: World Geodetic System 1984; WUE: Water Use Efficiency

Competing Interests

The authors declare that they have no competing interests

Funding

This project is supported by Swiss Open Research Data Grants (CHORD) in Open Science I, a program coordinated by Swissuniversities. In addition, it is also supported by three Swiss National Science Foundation Grants (Anticipating the Chemistry of Life - IC00I0-227830, MetaDiv 315230_215724, MetaboLinkAI 10.002.786). This work was also partially supported by Horizon Europe, project MICROBES-4-CLIMATE 101131818.

Authors' Contributions

D.T., P.M.A., T.M.F. and E.D. conceptualized the study. D.T. and T.M.F curated the data, designed the methodology and developed the knowledge graph. T.M.F. designed and developed the Earth Metabolome Ontology with inputs from D.T. and P.M.A.. D.T. and T.M.F. prepared the visualization. D.T. and E.D. identified the example case-studies relevant to the knowledge graph. T.M.F and P.M.A. provided computing resources for the project. P.M.A, T.M.F. and E.D. acquired the funding for the project. D.T. and T.M.F. wrote the original draft. All authors reviewed and edited the manuscript.

Acknowledgements

The authors thank members of Earth Metabolome Initiative consortium for their useful comments and suggestions on the project.

Figure legends

Figure-1: Snapshot of the data schema of the Earth Metabolome Initiative (EMI) ontology. Classes are represented in round-edged rectangles. Entity relationships are represented by directional arrows.

Figure-2: Schema of relations between different entities of integrated datasets of metabolome, traits and interactions. Entities mapped are represented in bold and the corresponding relationship is represented by directional arrows.

Figure-3: Original structure of traits dataset retrieved from TRY database. Purple-shaded rectangles represent columns retained for METRIN-KG. Gray-shaded rectangles represent columns not retained for METRIN-KG. Column relationships are represented by directional arrows.

Figure-4: Original structure of interactions dataset retrieved from GloBI. Purple-shaded rectangles represent columns retained for METRIN-KG. Gray-shaded rectangles represent columns not retained for METRIN-KG. Column relationships are represented by directional arrows.

Table legends

Table-1: List of traits used for retrieving data from TRY database

Table-2: Examples of SPARQL queries for METRIN-KG

Supplementary Table legends

Supplementary Table 1: List of taxonomy databases used in GloBI

Supplementary Table 2: List of ontologies used in METRIN-KG

Supplementary Table 3: Column descriptions of data retrieved from TRY database. Column names and descriptions were retained from the files provided by TRY database managers

Supplementary Table 4: Metrics for TRY database and GloBI

Supplementary Table 5: Column descriptions of data retrieved from GloBI. Column names and descriptions obtained from (https://github.com/globalbioticinteractions/template-dataset#data-format-and-dictionary)

References

1. Hortal J, Bello F de, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu Rev Ecol Evol Syst*. Annual Reviews; 2015; doi: 10.1146/annurev-ecolsys-112414-054400.

2. Cardoso P, Erwin T, Borges P, New T. The seven impediments in invertebrate conservation and how to overcome them. *Biol Conserv*. 2011; doi: 10.1016/j.biocon.2011.07.024.

3. Pollock LJ, Kitzes J, Beery S, Gaynor KM, Jarzyna MA, Mac Aodha O, et al.. Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nat Rev Biodivers*. Nature Publishing Group; 2025; doi: 10.1038/s44358-025-00022-3.

4. Walker TWN, Alexander JM, Allard P-M, Baines O, Baldy V, Bardgett RD, et al.. Functional Traits 2.0: The power of the metabolome for ecology. *J Ecol*. 2022; doi: 10.1111/1365-2745.13826.

5. Hilker M. New Synthesis: Parallels Between Biodiversity and Chemodiversity. *J Chem Ecol*. 2014; doi: 10.1007/s10886-014-0402-8.

6. Schuman MC, Baldwin IT. The Layers of Plant Responses to Insect Herbivores. *Annu Rev Entomol*. Annual Reviews; 2016; doi: 10.1146/annurev-ento-010715-023851.

7. Wink M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*. 2003; doi: 10.1016/S0031-9422(03)00300-5.

8. Hart CE, Gadiya Y, Kind T, Krettler CA, Gaetz M, Misra BB, et al.. Defining the limits of plant chemical space: challenges and estimations. *GigaScience*. 2025; doi: 10.1093/gigascience/giaf033.

9. van Dam NM, van der Meijden E. A Role for Metabolomics in Plant Ecology. *Annu Plant Rev Online*. John Wiley & Sons, Ltd; 2011; doi: 10.1002/9781444339956.ch4.

10. Weckwerth W. Metabolomics in Systems Biology. *Annu Rev Plant Biol*. Annual

Reviews; 2003; doi: 10.1146/annurev.arplant.54.031902.135014.

11. Echeverri A, Karp DS, Naidoo R, Tobias JA, Zhao J, Chan KMA. Can avian functional traits predict cultural ecosystem services? *People Nat*. 2020; doi: 10.1002/pan3.10058.

12. Wong MKL, Guénard B, Lewis OT. Trait-based ecology of terrestrial arthropods. *Biol Rev*. 2019; doi: 10.1111/brv.12488.

13. Lundgren EJ, Schowanek SD, Rowan J, Middleton O, Pedersen RØ, Wallach AD, et al.. Functional traits of the world's late Quaternary large-bodied avian and mammalian herbivores. *Sci Data*. Nature Publishing Group; 2021; doi: 10.1038/s41597-020-00788-5.

14. Kunin WE, Vergeer P, Kenta T, Davey MP, Burke T, Ian Woodward F, et al.. Variation at range margins across multiple spatial scales: environmental temperature, population genetics and metabolomic phenotype. *Proc R Soc B Biol Sci*. 2009; doi: 10.1098/rspb.2008.1767.

15. Dwyer JM, Hobbs RJ, Mayfield MM. Specific leaf area responses to environmental gradients through space and time. *Ecology*. 2014; doi: 10.1890/13-0412.1.

16. Adler PB, Salguero-Gómez R, Compagnoni A, Hsu JS, Ray-Mukherjee J, Mbeau-Ache C, et al.. Functional traits explain variation in plant life history strategies. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 2014; doi: 10.1073/pnas.1315179111.

17. Pistón N, de Bello F, Dias ATC, Götzenberger L, Rosado BHP, de Mattos EA, et al.. Multidimensional ecological analyses demonstrate how interactions between functional traits shape fitness and life history strategies. *J Ecol*. 2019; doi: 10.1111/1365-2745.13190.

18. Walker TWN, Weckwerth W, Bragazza L, Fragner L, Forde BG, Ostle NJ, et al.. Plastic and genetic responses of a common sedge to warming have contrasting effects on carbon cycle processes. *Ecol Lett*. 2019; doi: 10.1111/ele.13178.

19. van der Plas F, Schröder-Georgi T, Weigelt A, Barry K, Meyer S, Alzate A, et al.. Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. *Nat Ecol Evol*. Nature Publishing Group; 2020; doi: 10.1038/s41559-020-01316-9.

20. Firn J, McGree JM, Harvey E, Flores-Moreno H, Schütz M, Buckley YM, et al.. Leaf nutrients, not specific leaf area, are consistent indicators of elevated nutrient inputs. *Nat Ecol Evol*. Nature Publishing Group; 2019; doi: 10.1038/s41559-018-0790-1.

21. Laughlin DC, Gremer JR, Adler PB, Mitchell RM, Moore MM. The Net Effect of Functional Traits on Fitness. *Trends Ecol Evol*. Elsevier; 2020; doi: 10.1016/j.tree.2020.07.010.

22. Eisenhauer N, Bonfante P, Buscot F, Cesarz S, Guerra C, Heintz-Buschart A, et al.. Biotic Interactions as Mediators of Context-Dependent Biodiversity-Ecosystem Functioning Relationships. *Res Ideas Outcomes*. Pensoft Publishers; 2022; doi: 10.3897/rio.8.e85873.

23. Maestre FT, Bowker MA, Escolar C, Puche MD, Soliveres S, Maltez-Mouro S, et al.. Do biotic interactions modulate ecosystem functioning along stress gradients? Insights from semi-arid plant and biological soil crust communities. *Philos Trans R Soc B Biol Sci*. 2010; doi: 10.1098/rstb.2010.0016.

24. Jayaramaiah RH, Egidi E, Macdonald CA, Singh BK. Linking biodiversity and biotic interactions to ecosystem functioning. *J Sustain Agric Environ*. 2024; doi: 10.1002/sae2.12119.

25. de Vries S, Feussner I. Biotic interactions, evolutionary forces and the pan-plant specialized metabolism. *Philos Trans R Soc B Biol Sci*. Royal Society; 2024; doi: 10.1098/rstb.2023.0362.

26. Defossez E, Pitteloud C, Descombes P, Glauser G, Allard P-M, Walker TWN, et al.. Spatial and evolutionary predictability of phytochemical diversity. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 2021; doi:

10.1073/pnas.2013344118.

27. Agrawal AA, Fishbein M, Halitschke R, Hastings AP, Rabosky DL, Rasmann S. Evidence for adaptive radiation from a phylogenetic study of plant defenses. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 2009; doi: 10.1073/pnas.0904862106.

28. Ehrlich PR, Raven PH. Butterflies and Plants: A Study in Coevolution. *Evolution*. [Society for the Study of Evolution, Wiley]; 1964; doi: 10.2307/2406212.

29. Bennett RN, Wallsgrove RM. Secondary metabolites in plant defence mechanisms. *New Phytol*. 1994; doi: 10.1111/j.1469-8137.1994.tb02968.x.

30. Walker TWN, Schrodt F, Allard P-M, Defossez E, Jassey VEJ, Schuman MC, et al.. Leaf metabolic traits reveal hidden dimensions of plant form and function. *Sci Adv*. American Association for the Advancement of Science; 2023; doi: 10.1126/sciadv.adi4029.

31. Deng P, Yin R, Wang H, Chen L, Cao X, Xu X. Comparative analyses of functional traits based on metabolome and economic traits variation of Bletilla striata: Contribution of intercropping. *Front Plant Sci*. Frontiers; 2023; doi: 10.3389/fpls.2023.1147076.

32. Wei J, Wang A, Li R, Qu H, Jia Z. Metabolome-wide association studies for agronomic traits of rice. *Heredity*. Nature Publishing Group; 2018; doi: 10.1038/s41437-017-0032-3.

33. Delory BM, Callaway RM, Semchenko M. A trait-based framework linking the soil metabolome to plant–soil feedbacks. *New Phytol*. 2024; doi: 10.1111/nph.19490.

34. Shi T, Zhu A, Jia J, Hu X, Chen J, Liu W, et al.. Metabolomics analysis and metabolite‐agronomic trait associations using kernels of wheat (Triticum aestivum) recombinant inbred lines. *Plant J*. 2020; doi: 10.1111/tpj.14727.

35. Green SJ, Brookson CB, Hardy NA, Crowder LB. Trait-based approaches to global change ecology: moving from description to prediction. *Proc R Soc B Biol Sci*. Royal

Society; 2022; doi: 10.1098/rspb.2022.0071.

36. Walter HE, Pagel J, Cooksley H, Neu A, Schleuning M, Schurr FM. Effects of biotic interactions on plant fecundity depend on spatial and functional structure of communities and time since disturbance. *J Ecol*. 2023; doi: 10.1111/1365-2745.14018.

37. Pueyo Y, Kéfi S, Díaz-Sierra R, Alados CL, Rietkerk M. The role of reproductive plant traits and biotic interactions in the dynamics of semi-arid plant communities. *Theor Popul Biol*. 2010; doi: 10.1016/j.tpb.2010.09.001.

38. Gaüzère P, O'Connor L, Botella C, Poggiato G, Münkemüller T, Pollock LJ, et al.. The diversity of biotic interactions complements functional and phylogenetic facets of biodiversity. *Curr Biol*. 2022; doi: 10.1016/j.cub.2022.03.009.

39. Tenenboim H, Brotman Y. Omic Relief for the Biotically Stressed: Metabolomics of Plant Biotic Interactions. *Trends Plant Sci*. 2016; doi: 10.1016/j.tplants.2016.04.009.

40. Gupta S, Schillaci M, Roessner U. Metabolomics as an emerging tool to study plant–microbe interactions. *Emerg Top Life Sci*. 2022; doi: 10.1042/ETLS20210262.

41. Henderson D, Sedio BE, Tello JS, Cayola L, Fuentes AF, Alvestegui B, et al.. Ecological metabolomics of tropical tree communities across an elevational gradient: Implications for chemically-mediated biotic interactions and species diversity. bioRxiv. 2023; doi: 10.1101/2023.10.04.560880.

42. Maag D, Erb M, Glauser G. Metabolomics in plant–herbivore interactions: challenges and applications. *Entomol Exp Appl*. 2015; doi: 10.1111/eea.12336.

43. Majumdar S, Kaur H, Rinella MJ, Kundu A, Vadassery J, Erbilgin N, et al.. Synergistic effects of canopy chemistry and autogenic soil biota on a global invader. *J Ecol*. 2023; doi: 10.1111/1365-2745.14113.

44. Semchenko M, Nettan S, Sepp A, Zhang Q, Abakumova M, Davison J, et al.. Soil biota and chemical interactions promote co-existence in co-evolved grassland communities. *J Ecol*. 2019; doi: 10.1111/1365-2745.13220.

45. Burghardt KT, Bradford MA, Schmitz OJ. Acceleration or deceleration of litter decomposition by herbivory depends on nutrient availability through intraspecific differences in induced plant resistance traits. *J Ecol*. 2018; doi: 10.1111/1365-2745.13002.

46. Heinen R, Biere A, Bezemer TM. Plant traits shape soil legacy effects on individual plant–insect interactions. *Oikos*. 2020; doi: 10.1111/oik.06812.

47. De Long JR, Heinen R, Hannula SE, Jongen R, Steinauer K, Bezemer TM. Plant-litter-soil feedbacks in common grass species are slightly negative and only marginally modified by litter exposed to insect herbivory. *Plant Soil*. 2023; doi: 10.1007/s11104-022-05590-3.

48. Eissenstat DM, Kucharski JM, Zadworny M, Adams TS, Koide RT. Linking root traits to nutrient foraging in arbuscular mycorrhizal trees in a temperate forest. *New Phytol*. 2015; doi: 10.1111/nph.13451.

49. Stiblíková P, Klimeš A, Cahill JF, Koubek T, Weiser M. Interspecific differences in root foraging precision cannot be directly inferred from species' mycorrhizal status or fine root economics. *Oikos*. 2023; doi: 10.1111/oik.08995.

50. Xia M, Valverde-Barrantes OJ, Suseela V, Blackwood CB, Tharayil N. Coordination between compound-specific chemistry and morphology in plant roots aligns with ancestral mycorrhizal association in woody angiosperms. *New Phytol*. 2021; doi: 10.1111/nph.17561.

51. Kong C-H, Zhang S-Z, Li Y-H, Xia Z-C, Yang X-F, Meiners SJ, et al.. Plant neighbor detection and allelochemical response are driven by root-secreted signaling chemicals. *Nat Commun*. 2018; doi: 10.1038/s41467-018-06429-1.

52. Li L, Li S-M, Sun J-H, Zhou L-L, Bao X-G, Zhang H-G, et al.. Diversity enhances agricultural productivity via rhizosphere phosphorus facilitation on phosphorus-deficient soils. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 2007; doi: 10.1073/pnas.0704591104.

53. Steinauer K, Thakur MP, Emilia Hannula S, Weinhold A, Uthe H, van Dam NM, et al.. Root exudates and rhizosphere microbiomes jointly determine temporal shifts in plant-soil feedbacks. *Plant Cell Environ*. 2023; doi: 10.1111/pce.14570.

54. Henderson D, Tello JS, Cayola L, Fuentes AF, Alvestegui B, Muchhala N, et al.. Testing the role of biotic interactions in shaping elevational diversity gradients: An ecological metabolomics approach. *Ecology*. 2025; doi: 10.1002/ecy.70069.

55. Bovay B, Descombes P, Chittaro Y, Glauser G, Nomoto H, Rasmann S. Adapting to change: Exploring the consequences of climate-induced host plant shifts in two specialist Lepidoptera species. *Ecol Evol*. 2024; doi: 10.1002/ece3.11596.

56. Sierra AM, Meléndez O, Bethancourt R, Bethancourt A, Rodríguez-Castro L, López CA, et al.. Leaf Endophytes Relationship with Host Metabolome Expression in Tropical Gymnosperms. *J Chem Ecol*. 2024; doi: 10.1007/s10886-024-01511-z.

57. Mleziva AD, Ngumbi EN. Comparative analysis of defensive secondary metabolites in wild teosinte and cultivated maize under flooding and herbivory stress. *Physiol Plant*. 2024; doi: 10.1111/ppl.14216.

58. Gallon ME, Muchoney ND, Smilanich AM. Viral Infection Induces Changes to the Metabolome, Immune Response and Development of a Generalist Insect Herbivore. *J Chem Ecol*. 2024; doi: 10.1007/s10886-024-01472-3.

59. Contreras-Cornejo HA, Schmoll M, Esquivel-Ayala BA, González-Esquivel CE, Rocha-Ramírez V, Larsen J. Mechanisms for plant growth promotion activated by *Trichoderma* in natural and managed terrestrial ecosystems. *Microbiol Res*. 2024; doi: 10.1016/j.micres.2024.127621.

60. Yasmin F, Cowie AE, Zerbe P. Understanding the chemical language mediating maize immunity and environmental adaptation. *New Phytol*. 2024; doi: 10.1111/nph.20000.

61. Ehlers BK, Berg MP, Staudt M, Holmstrup M, Glasius M, Ellers J, et al.. Plant Secondary Compounds in Soil and Their Role in Belowground Species Interactions.

Trends Ecol Evol. Elsevier; 2020; doi: 10.1016/j.tree.2020.04.001.

62. Delory BM, Delaplace P, Fauconnier M-L, du Jardin P. Root-emitted volatile organic compounds: can they mediate belowground plant-plant interactions? *Plant Soil*. 2016; doi: 10.1007/s11104-016-2823-3.

63. Semchenko M, Barry KE, de Vries FT, Mommer L, Moora M, Maciá-Vicente JG. Deciphering the role of specialist and generalist plant–microbial interactions as drivers of plant–soil feedback. *New Phytol*. 2022; doi: 10.1111/nph.18118.

64. Ninkovic V, Markovic D, Rensing M. Plant volatiles as cues and signals in plant communication. *Plant Cell Environ*. 2021; doi: 10.1111/pce.13910.

65. Moore BD, Andrew RL, Külheim C, Foley WJ. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytol*. 2014; doi: 10.1111/nph.12526.

66. Bilas RD, Bretman A, Bennett T. Friends, neighbours and enemies: an overview of the communal and social biology of plants. *Plant Cell Environ*. 2021; doi: 10.1111/pce.13965.

67. Wang N-Q, Kong C-H, Wang P, Meiners SJ. Root exudate signals in plant–plant interactions. *Plant Cell Environ*. 2021; doi: 10.1111/pce.13892.

68. Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, et al.. The LOTUS initiative for open knowledge management in natural products research. Donoso DA, Akhmanova A, Tapley Hoyt C, editors. *eLife*. eLife Sciences Publications, Ltd; 2022; doi: 10.7554/eLife.70780.

69. Gaudry A, Pagni M, Mehl F, Moretti S, Quiros-Guerrero L-M, Cappelletti L, et al.. A Sample-Centric and Knowledge-Driven Computational Framework for Natural Products Drug Discovery. *ACS Cent Sci*. American Chemical Society; 2024; doi: 10.1021/acscentsci.3c00800.

70. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al.. PubChem 2025 update.

Nucleic Acids Res. 2025; doi: 10.1093/nar/gkae1059.

71. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al.. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res*. 2024; doi: 10.1093/nar/gkad1004.

72. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, et al.. KNApSAcK Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research. *Plant Cell Physiol*. 2012; doi: 10.1093/pcp/pcr165.

73. The Earth Metabolome Initiative. https://www.earthmetabolome.org/ (2023) Accessed 2025 Jun 4.

74. The Digital Botanical Gardens Initiative. https://www.dbgi.org/ (2022) Accessed 2025 Jun 4.

75. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al.. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. Nature Publishing Group; 2016; doi: 10.1038/sdata.2016.18.

76. Poelen JH, Simons JD, Mungall CJ. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecol Inform*. 2014; doi: 10.1016/j.ecoinf.2014.08.005.

77. BiotXplorer. https://biotxplorer.sibils.org/ Accessed 2025 Jun 5.

78. Page R. Towards a biodiversity knowledge graph. *Res Ideas Outcomes*. Pensoft Publishers; 2016; doi: 10.3897/rio.2.e8767.

79. Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, et al.. TRY plant trait database – enhanced coverage and open access. *Glob Change Biol*. 2020; doi: 10.1111/gcb.14904.

80. Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, et al.. TRY – a global database of plant traits. *Glob Change Biol*. 2011; doi:

10.1111/j.1365-2486.2011.02451.x.

81. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, et al.. The worldwide leaf economics spectrum. *Nature*. Nature Publishing Group; 2004; doi: 10.1038/nature02403.

82. Nordt B, Hensen I, Bucher SF, Freiberg M, Primack RB, Stevens A-D, et al.. The PhenObs initiative: A standardised protocol for monitoring phenological responses to climate change using herbaceous plant species in botanical gardens. *Funct Ecol*. 2021; doi: 10.1111/1365-2435.13747.

83. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al.. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res*. 2020; doi: 10.1093/nar/gkz862.

84. Karp PD, Paley S, Caspi R, Kothari A, Krummenacker M, Midford PE, et al.. The EcoCyc Database (2023). *EcoSal Plus*. American Society for Microbiology; 2023; doi: 10.1128/ecosalplus.esp-0002-2023.

85. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al.. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform*. 2019; doi: 10.1093/bib/bbx085.

86. Singh K, Maurya H, Singh P, Panda P, Behera AK, Jamal A, et al.. DISPEL: database for ascertaining the best medicinal plants to cure human diseases. *Database*. 2023; doi: 10.1093/database/baad073.

87. Youn J, Li F, Simmons G, Kim S, Tagkopoulos I. FoodAtlas: Automated Knowledge Extraction of Food and Chemicals from Literature. bioRxiv. 2024; doi: 10.1101/2024.05.16.594596.

88. Earth Metabolome Ontology (2024). https://github.com/earth-metabolome-initiative/earth_metabolome_ontology Accessed 2025 July 16.

89. The Earth Metabolome Initiative (EMI) ontology. https://w3id.org/emi (2023) Accessed 2025 Aug 16.

90. Calvanese D, Lanti D, Mendes De Farias T, Mosca A, Xiao G. Accessing scientific data through knowledge graphs with Ontop. *Patterns*. 2021; doi: 10.1016/j.patter.2021.100346.

91. Calvanese D, Cogrel B, Komla-Ebri S, Kontchakov R, Lanti D, Rezk M, et al.. Ontop: Answering SPARQL queries over relational databases. *Semantic Web*. SAGE Publications; 2016; doi: 10.3233/SW-160217.

92. RDFlib. https://rdflib.dev/ Accessed 2025 Jun 6.

93. GloBI community (2025). Global Biotic Interactions: Interpreted Data Products (Version 0.8). Zenodo; doi: 10.5281/zenodo.14640564

94. Global Biotic Interactions: globalbioticinteractions. https://github.com/globalbioticinteractions Accessed 2025 Jun 4.

95. Global Biotic Interactions: datasets. https://www.globalbioticinteractions.org/datasets Accessed 2025 Jun 4.

96. Bast H, Buchhold B. QLever: A Query Engine for Efficient SPARQL+Text Search. *Proc 2017 ACM Conf Inf Knowl Manag*. New York, NY, USA: Association for Computing Machinery; doi: 10.1145/3132847.3132921

97. The QLever SPARQL engine: fast, scalable, with autocompletion and text search. https://qlever.cs.uni-freiburg.de/wikidata Accessed 2025 Jun 4.

98. Lamy J-B. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. *Artif Intell Med*. 2017; doi: 10.1016/j.artmed.2017.07.002.

99. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv. 2019; doi: 10.48550/arXiv.1908.10084.

100. FAIRsharing Team. FAIRsharing record for: Quantities, Units, Dimensions and Types. FAIRsharing (2015); doi: 10.25504/FAIRSHARING.D3PQW7

101. Janowicz K, Haller A, Cox SJD, Le Phuoc D, Lefrançois M. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *J Web Semant*. 2019; doi: 10.1016/j.websem.2018.06.003.

102. Advancing Geospatial Standards and Technology | OGC. Open Geospatial Consort. https://www.ogc.org/ Accessed 2025 Jun 4.

103. SKOS Simple Knowledge Organization System Namespace Document - HTML Variant, 18 August 2009 Recommendation Edition. https://www.w3.org/2009/08/skos-reference/skos.html Accessed 2025 Jun 4.

104. W3C Semantic Web Interest Group: Basic Geo (WGS84 lat/long) Vocabulary. https://www.w3.org/2003/01/geo/ Accessed 2025 Jun 4.

105. Xiao G, Lanti D, Kontchakov R, Komla-Ebri S, Güzel-Kalaycı E, Ding L, et al.. The Virtual Knowledge Graph System Ontop (Extended Abstract).

106. Allard P-M, Gaudry A. Input and enriched files for the pf1600 dataset - ENPKG. 2024 Zenodo; doi: 10.5281/zenodo.10827917.

107. Allard P-M, Gaudry A, Quirós-Guerrero L-M, Rutz A, Dounoue-Kubo M, Walker TWN, et al.. Open and reusable annotated mass spectrometry dataset of a chemodiverse collection of 1,600 plant extracts. *GigaScience*. 2023; doi: 10.1093/gigascience/giac124.

108. MySQL : MySQL 8.2 Release Notes. https://dev.mysql.com/doc/relnotes/mysql/8.2/en/ Accessed 2025 Jun 6.

109. Krech D, Grimnes GAa, Higgins G, Hees J, Aucamp I, Lindström N, et al.. RDFLib. Zenodo. 2023; doi: 10.5281/zenodo.8206632.

110. ad-freiburg/qlever-control (2022). University of Freiburg: Algorithms and Data

Structures Group; https://github.com/ad-freiburg/qlever-control. Accessed 2025 Jun 4.

111. Bast H, Kalmbach J, Klumpp T, Kramer F, Schnelle N. Efficient SPARQL Autocompletion via SPARQL. arXiv. 2021; doi: 10.48550/arXiv.2104.14595.

112. ad-freiburg/qlever-ui (2018). University of Freiburg: Algorithms and Data Structures Group; https://github.com/ad-freiburg/qlever-ui. Accessed 2025 Jun 4.

113. globalbioticinteractions/template-dataset (2014). globalbioticinteractions; https://github.com/globalbioticinteractions/template-dataset. Accessed 2025 Jun 4.

114. Emonet V, Bolleman J, Duvaud S, Farias TM de, Sima AC. LLM-based SPARQL Query Generation from Natural Language over Federated Knowledge Graphs. arXiv. 2025; doi: 10.48550/arXiv.2410.06062.

115. Emonet V, Sima A-C, Farias TM de. A user-friendly SPARQL query editor powered by lightweight metadata. arXiv. 2025; doi: 10.48550/arXiv.2503.02688.

116. Jupp S, Malone J, Bolleman J, Brandizi M, Davies M, Garcia L, et al.. The EBI RDF platform: linked open data for the life sciences. *Bioinformatics*. 2014; doi: 10.1093/bioinformatics/btt765.

117. Calleja JA, Domènech G, Sáez L, Lara F, Garilleti R, Albertos B. Extinction risk of threatened and non-threatened mosses: Reproductive and ecological patterns. *Glob Ecol Conserv*. 2022; doi: 10.1016/j.gecco.2022.e02254.

118. Gürlek S, Araújo AC, Brummitt N. Predicting the Threat Status of Mosses Using Functional Traits. *Plants*. Multidisciplinary Digital Publishing Institute; 2024; doi: 10.3390/plants13152019.

119. Junaedi DI, Nasution T, Putri DM, Iryadi R, Lestari R, Kurniawan V, et al.. Threatened exotic species of botanical gardens: application of trait-based naturalized species risk scoring assessment. *South Afr J Bot*. 29. 2022; doi: 10.1016/j.sajb.2022.11.046

120. Álvarez-Yépiz JC, Búrquez A, Martínez-Yrízar A, Dovciak M. A trait-based approach to the conservation of threatened plant species. *Oryx*. 2019; doi: 10.1017/S003060531800087X.

121. Hill JL, Grisnik M, Hanscom RJ, Sukumaran J, Higham TE, Clark RW. The past, present, and future of predator–prey interactions in a warming world: Using species distribution modeling to forecast ectotherm–endotherm niche overlap. *Ecol Evol*. 2024; doi: 10.1002/ece3.11067.

122. Stelling-Wood TP, Poore AGB, Hughes AR, Everett JD, Gribben PE. Habitat traits and predation interact to drive abundance and body size patterns in associated fauna. *Ecol Evol*. 2023; doi: 10.1002/ece3.10771.

123. Ray K, Basak SK, Giri CK, Kotal HN, Mandal A, Chatterjee K, et al.. Ecological restoration at pilot-scale employing site-specific rationales for small-patch degraded mangroves in Indian Sundarbans. *Sci Rep*. Nature Publishing Group; 2024; doi: 10.1038/s41598-024-63281-8.

124. Mendes SB, Nogales M, Vargas P, Olesen JM, Marrero P, Romero J, et al.. Climb forest, climb: diverse disperser communities are key to assist plants tracking climate change on altitudinal gradients. *New Phytol*. 2025; doi: 10.1111/nph.20300.

125. Flickinger HD, Dukes JS. A Review of Theory: Comparing Invasion Ecology and Climate Change-Induced Range Shifting. *Glob Change Biol*. 2024; doi: 10.1111/gcb.17612.

126. Wang X, Cao Y, Jin Y, Sun L, Tang F, Dong L. Ecophysiological Trade-Off Strategies of Three Gramineous Crops in Response to Root Extracts of Phytolacca americana. *Plants*. Multidisciplinary Digital Publishing Institute; 2024; doi: 10.3390/plants13213026.

127. De La Peña R, Sattely ES. Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat Chem Biol*. Nature Publishing Group; 2021; doi: 10.1038/s41589-020-00669-3.

128. Knoch E, Kovács J, Deiber S, Tomita K, Shanmuganathan R, Serra Serra N, et al.. Transcriptional response of a target plant to benzoxazinoid and diterpene allelochemicals highlights commonalities in detoxification. *BMC Plant Biol*. 2022; doi: 10.1186/s12870-022-03780-w.

129. Lu X, Zhang J, Brown B, Li R, Rodríguez-Romero J, Berasategui A, et al.. Inferring Roles in Defense from Metabolic Allocation of Rice Diterpenoids. *Plant Cell*. 2018; doi: 10.1105/tpc.18.00205.

130. Zhou S, Zhang R, Wang Q, Zhu J, Zhou J, Sun Y, et al.. OsbHLH5 Synergically Regulates Phenolamide and Diterpenoid Phytoalexins Involved in the Defense of Rice Against Pathogens. *Int J Mol Sci*. Multidisciplinary Digital Publishing Institute; 2024; doi: 10.3390/ijms252212152.

131. Vela F, Anese S, Varela RM, Torres A, Molinillo JMG, Macías FA. Bioactive Diterpenes from the Brazilian Native Plant (Moquiniastrum pulchrum) and Their Application in Weed Control. *Molecules*. Multidisciplinary Digital Publishing Institute; 2021; doi: 10.3390/molecules26154632.

132. Maraia H, Charles-Dominique T, Tomlinson KW, Staver AC, Jorge LR, Gélin U, et al.. Substantial Insect Herbivory in a South African Savanna-Forest Mosaic: A Neglected Topic. *Ecol Evol*. 2024; doi: 10.1002/ece3.70466.

133. Zhao Y, Hu J, Zhou Z, Li L, Zhang X, He Y, et al.. Biofortified Rice Provides Rich Sakuranetin in Endosperm. *Rice*. 2024; doi: 10.1186/s12284-024-00697-w.

134. Khatibi SMH, Dimaano NG, Veliz E, Sundaresan V, Ali J. Exploring and exploiting the rice phytobiome to tackle climate change challenges. *Plant Commun*. Elsevier; 2024; doi: 10.1016/j.xplc.2024.101078.

135. Bian S, Li Z, Song S, Zhang X, Shang J, Wang W, et al.. Enhancing Crop Resilience: Insights from Labdane-Related Diterpenoid Phytoalexin Research in Rice (Oryza sativa L.). *Curr Issues Mol Biol*. Multidisciplinary Digital Publishing Institute; 2024; doi: 10.3390/cimb46090634.

136. Wang X, Li X, Zhao W, Hou X, Dong S. Current views of drought research: experimental methods, adaptation mechanisms and regulatory strategies. *Front Plant Sci*. Frontiers; 2024; doi: 10.3389/fpls.2024.1371895.

137. Yang H, Ji S, Wu D, Zhu M, Lv G. Effects of Root–Root Interactions on the Physiological Characteristics of Haloxylon ammodendron Seedlings. *Plants*. Multidisciplinary Digital Publishing Institute; 2024; doi: 10.3390/plants13050683.

138. Shi Y, He Y, Zheng Y, Liu X, Wang S, Xiong T, et al.. Characteristics of the phyllosphere microbial community and its relationship with major aroma precursors during the tobacco maturation process. *Front Plant Sci*. Frontiers; 2024; doi: 10.3389/fpls.2024.1346154.

139. Chain FE, Romano E, Leyton P, Paipa C, Catalán CAN, Fortuna M, et al.. Vibrational and structural study of onopordopicrin based on the FTIR spectrum and DFT calculations. *Spectrochim Acta A Mol Biomol Spectrosc*. 2015; doi: 10.1016/j.saa.2015.05.072.

140. Suzuki M, Iwasaki A, Suenaga K, Kato-Noguchi H. Phytotoxic activity of crop residues from Burdock and an active substance. *J Environ Sci Health Part B*. Taylor & Francis; 2019; doi: 10.1080/03601234.2019.1636600.

141. El Khatib N, Morel S, Hugon G, Rapior S, Carnac G, Saint N. Identification of a Sesquiterpene Lactone from Arctium lappa Leaves with Antioxidant Activity in Primary Human Muscle Cells. *Molecules*. Multidisciplinary Digital Publishing Institute; 2021; doi: 10.3390/molecules26051328.

142. Zhang J, Zheng Z-Q, Xu Q, Li Y, Gao K, Fang J. Onopordopicrin from the new genus Shangwua as a novel thioredoxin reductase inhibitor to induce oxidative stress-mediated tumor cell apoptosis. *J Enzyme Inhib Med Chem*. Taylor & Francis; 2021; doi: 10.1080/14756366.2021.1899169.

143. Maeta A, Okamoto Y, Ishikawa H, Matsunaga T, Takahashi K. Japanese Leaf Burdock Extract Inhibits Adipocyte Differentiation in 3T3-L1 Cells. *Plant Foods Hum*

Nutr. 2025; doi: 10.1007/s11130-024-01257-9.

144. Jalloh AA, Khamis FM, Yusuf AA, Subramanian S, Mutyambai DM. Long-term push–pull cropping system shifts soil and maize-root microbiome diversity paving way to resilient farming system. *BMC Microbiol*. 2024; doi: 10.1186/s12866-024-03238-z.

145. Czarnobai De Jorge B, Koßmann A, Hummel HE, Gross J. Evaluation of a push-and-pull strategy using volatiles of host and non-host plants for the management of pear psyllids in organic farming. *Front Plant Sci*. Frontiers; 2024; doi: 10.3389/fpls.2024.1375495.

146. Khan ZR, Chiliswa P, Ampong-Nyarko K, Smart LE, Polaszek A, Wandera J, et al.. Utilisation of Wild Gramineous Plants for Management of Cereal Stemborers in Africa. *Int J Trop Insect Sci*. 1997; doi: 10.1017/S1742758400022268.

147. Midega CAO, Wasonga CJ, Hooper AM, Pickett JA, Khan ZR. Drought-tolerant *Desmodium* species effectively suppress parasitic striga weed and improve cereal grain yields in western Kenya. *Crop Prot*. 2017; doi: 10.1016/j.cropro.2017.03.018.

148. Khan ZR, Midega CAO, Amudavi DM, Hassanali A, Pickett JA. On-farm evaluation of the ‘push–pull’ technology for the control of stemborers and striga weed on maize in western Kenya. *Field Crops Res*. 2008; doi: 10.1016/j.fcr.2007.12.002.

149. Hooper AM, Caulfield JC, Hao B, Pickett JA, Midega CAO, Khan ZR. Isolation and identification of *Desmodium* root exudates from drought tolerant species used as intercrops against *Striga hermonthica*. *Phytochemistry*. 2015; doi: 10.1016/j.phytochem.2015.06.026.

150. Reich PB, Walters MB, Ellsworth DS. From tropics to tundra: Global convergence in plant functioning. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 1997; doi: 10.1073/pnas.94.25.13730.

Figures

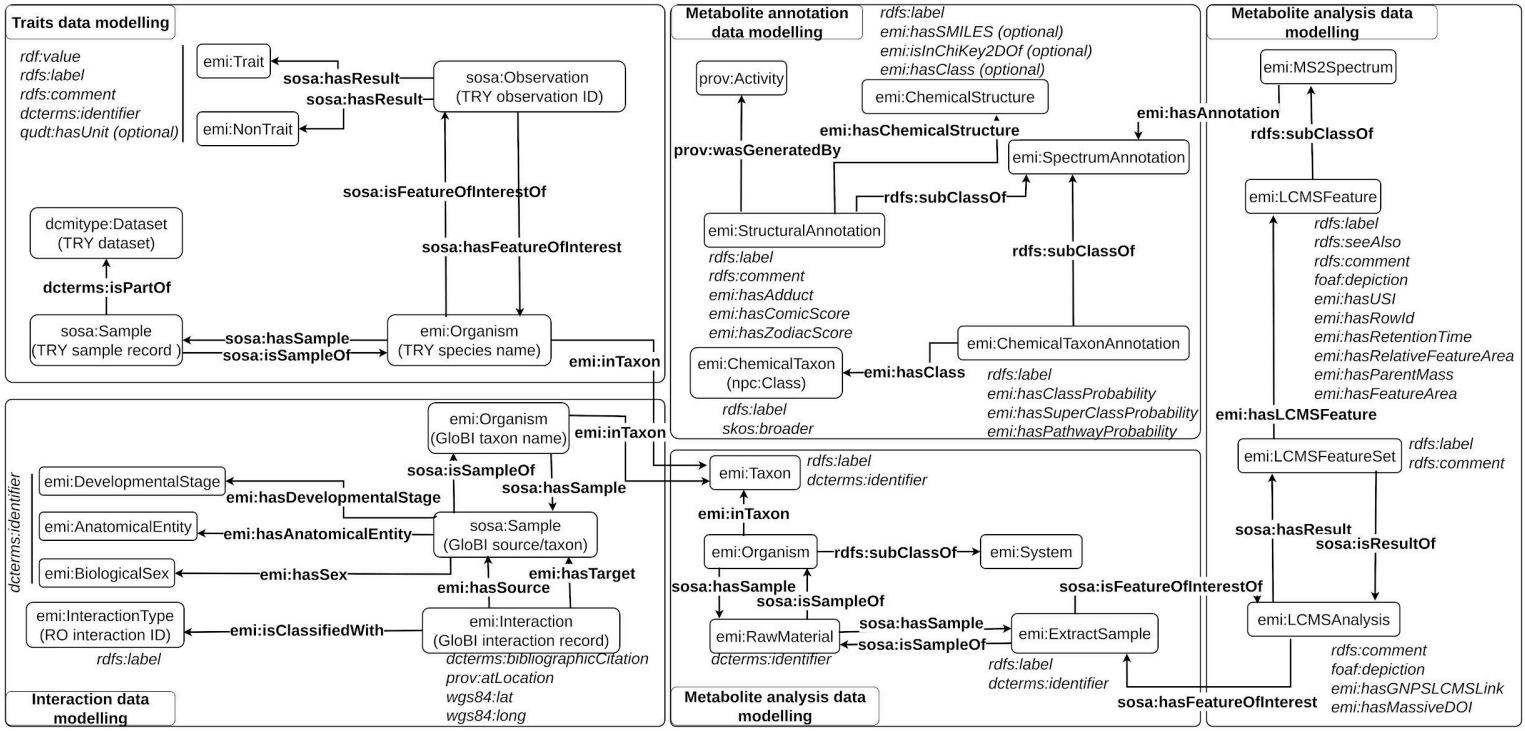

Figure-1: Snapshot of the data schema of the Earth Metabolome Initiative (EMI) ontology. Classes are represented in round-edged rectangles. Entity relationships are represented by directional arrows.

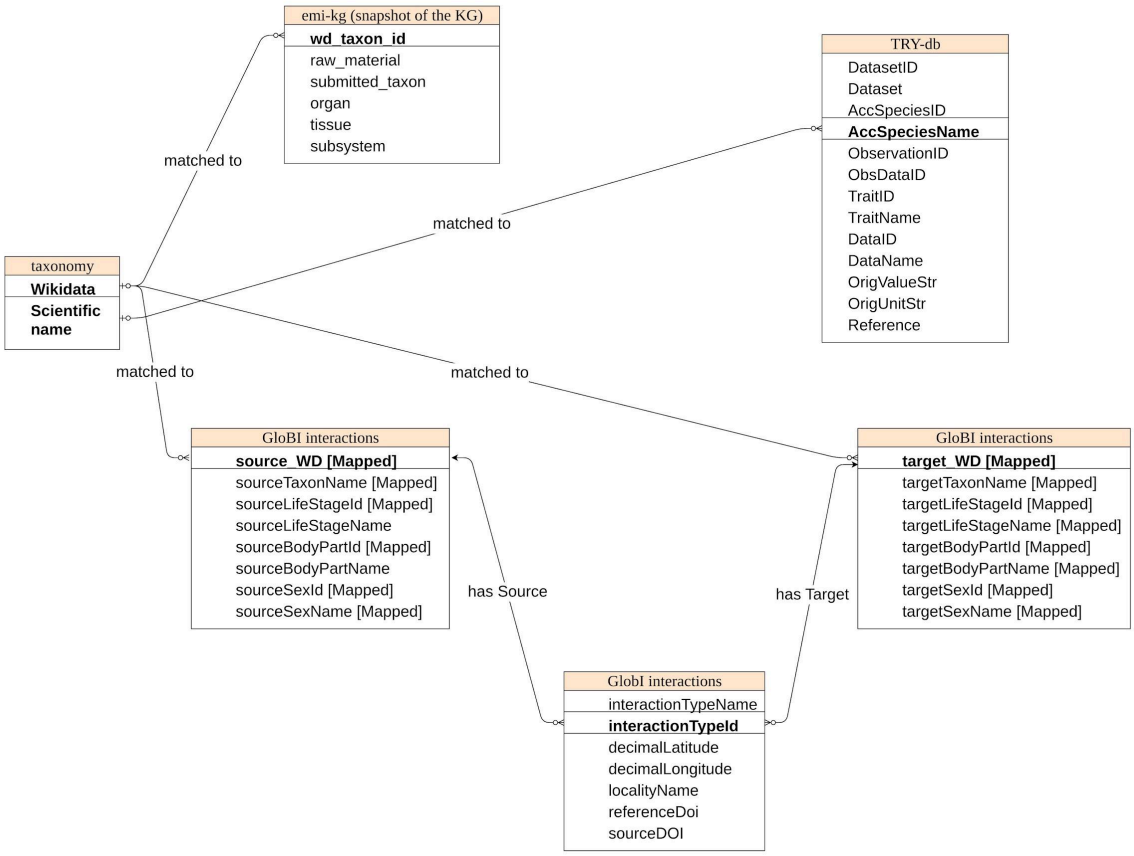

Figure-2: Schema of relations between different entities of integrated datasets of metabolome, traits and interactions. Entities mapped are represented in bold and the corresponding relationship is represented by directional arrows.

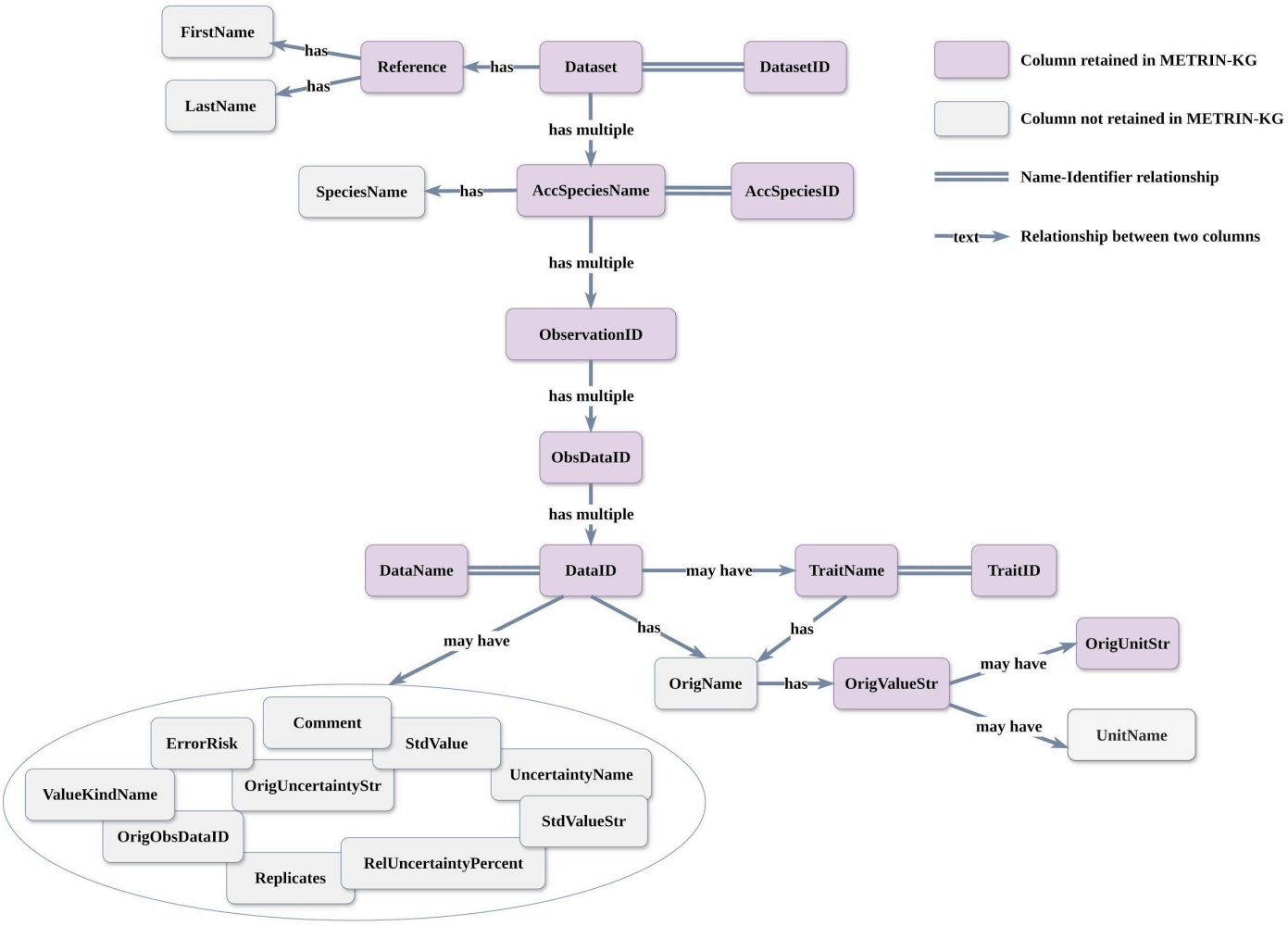

Figure-3: Original structure of traits dataset retrieved from TRY database. Purple-shaded rectangles represent columns retained for METRIN-KG. Gray-shaded rectangles represent columns not retained for METRIN-KG. Column relationships are represented by directional arrows.

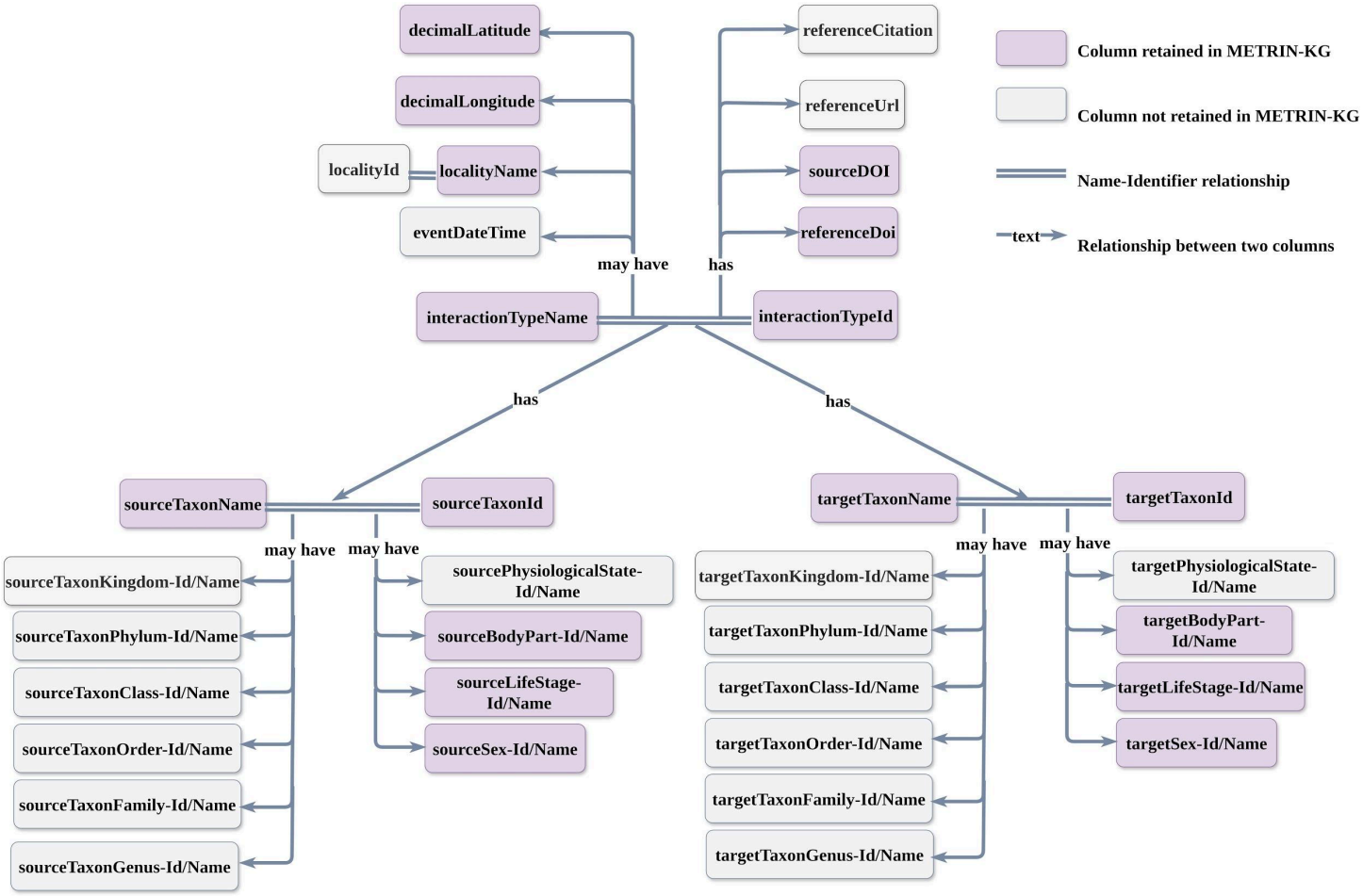

Figure-4: Original structure of interactions dataset retrieved from GloBI. Purple-shaded rectangles represent columns retained for METRIN-KG. Gray-shaded rectangles represent columns not retained for METRIN-KG. Column relationships are represented by directional arrows.