The World Spider Trait database (WST): a centralised global open repository for curated data on spider traits

OXFORD UNIVERSITY PRESS | DATABASE

The World Spider Trait database (WST): a centralised global open repository for curated data on spider traits

Journal:	*DATABASE*
Manuscript ID	DATABASE-2021-0081.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	13-Aug-2021
Complete List of Authors:	Pekar, Stano; Masaryk University, Department of Botany and Zoology Wolff , Jonas ; University of Greifswald Department and Museum of Zoology Černecká , Ľudmila ; Institute of Forest Ecology SAS Birkhofer , Klaus ; Brandenburg University of Technology Cottbus-Senftenberg, Department of Ecology Mammola , Stefano ; Finnish Museum of Natural History Lowe , Elizabeth; Macquarie University, Department of Biological Sciences Fukushima , Caroline; Finnish Museum of Natural History, Laboratory for Integrative Biodiversity Research Herberstein , Marie; Macquarie University, Department of Biological Sciences Kučera , Adam ; Masaryk University, Department of Botany and Zoology Buzatto , Bruno ; Macquarie University, Department of Biological Sciences Djoudi , El Aziz ; Brandenburg University of Technology Cottbus-Senftenberg, Department of Ecology Domenech , Marc ; Universitat de Barcelona, Department of Evolutionary Biology, Ecology and Environmental Sciences & Biodiversity Research Institute Enciso , Alison ; 10 Fundación Protectora Ambiental Planadas Tolima (FUPAPT Piñanez Espejo, Yolanda ; Instituto de Biología Subtropical Febles , Sara ; Grupo de Investigaciones Entomológicas de Tenerife García , Luis ; Universidad de la República Uruguay Gonçalves-Souza , Thiago ; Federal Rural University of Pernambuco, Department of Biology, Ecological Synthesis and Biodiversity Conservation Lab Isaia , Marco ; University of Turin, Department of Life Sciences and Systems Biology Lafage , Denis ; Université de Rennes 1 Líznarová , Eva ; Masaryk University, Department of Botany and Zoology Macías-Hernández , Nuria; Finnish Museum of Natural History, Laboratory for Integrative Biodiversity Research Magalhães , Ivan ; Museo Argentino de Ciencias Naturales Bernardino Rivadavia

Malumbres-Olarte , Jagoba; Finnish Museum of Natural History, Laboratory for Integrative Biodiversity Research
Michálek , Ondřej ; Masaryk University, Department of Botany and Zoology, Faculty of Science
Michalik , Peter ; University of Greifswald Department and Museum of Zoology
Michalko , Radek ; Mendel University in Brno Faculty of Forestry and Wood Technology, Department of Forest Ecology
Milano , Filippo ; University of Turin, Department of Life Sciences and Systems Biology
Munévar , Ana ; Instituto de Biología Subtropical
Nentwig , Wolfgang ; University of Bern, Institute of Ecology and Evolution
Nicolosi , Giuseppe ; University of Turin, Department of Life Sciences and Systems Biology
Painting , Christina; University of Waikato School of Science
Pétillon , Julien ; Université de Rennes 1
Piano , Elena ; University of Turin, Department of Life Sciences and Systems Biology
Privet , Kaïna ; Université de Rennes 1
Ramírez , Martín ; Museo Argentino de Ciencias Naturales Bernardino Rivadavia
Ramos , Cândida ; Finnish Museum of Natural History, Laboratory for Integrative Biodiversity Research
Řezáč , Milan ; Crop Research Institute
Ridel , Aurélien ; Université de Rennes 1
Růžička , Vlastimil ; Institute of Entomology Biology Centre Czech Academy of Sciences
Santos , Irene ; Grupo de Investigaciones Entomológicas de Tenerife (GIET)
Sentenská , Lenka ; Masaryk University, Department of Botany and Zoology
Walker , Leilani ; Auckland War Memorial Museum, Natural Sciences
Wierucka , Kaja ; Macquarie University, Department of Biological Sciences
Zurita , Gustavo ; Instituto de Biología Subtropical
Cardoso , Pedro ; Finnish Museum of Natural History

Keywords: Araneae, trait, ecology, database

The World Spider Trait database (WST): a centralised global open repository for curated data on spider traits

Stano Pekár[1], Jonas Wolff[2,3], Ľudmila Černecká[4], Klaus Birkhofer[5], Stefano Mammola[6,7], Elizabeth C. Lowe[3], Caroline S. Fukushima[6], Marie E. Herberstein[3], Adam Kučera[1], Bruno Buzatto[3,8], El Aziz Djoudi[5], Marc Domenech[9], Alison Vanesa Enciso[10], Yolanda M. G. Piñanez Espejo[11], Sara Febles[12], Luis F. García[13], Thiago Gonçalves-Souza[14], Marco Isaia[15], Denis Lafage[16], Eva Líznarová[1], Nuria Macías-Hernández[6,17], Ivan Magalhães[18], Jagoba Malumbres-Olarte[6,19], Ondřej Michálek[1], Peter Michalik[2], Radek Michalko[20], Filippo Milano[15], Ana Munévar[11], Wolfgang Nentwig[21], Giuseppe Nicolosi[15], Christina J. Painting[22], Julien Pétillon[16], Elena Piano[15], Kaïna Privet[16], Martín J. Ramírez[18], Cândida Ramos[6], Milan Řezáč[23], Aurélien Ridel[16], Vlastimil Růžička[24], Irene Santos[12,25], Lenka Sentenská[1], Leilani Walker[26], Kaja Wierucka[3,27], Gustavo Andres Zurita[11], Pedro Cardoso[6]

[1] Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czechia

[2] Zoological Institute and Museum, University of Greifswald, Loitzer Str. 26, 17489, Greifswald, Germany

[3] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

[4] Institute of Forest Ecology, Slovak Academy of Sciences, Zvolen 960 01, Slovak Republic

[5] Department of Ecology, Brandenburg University of Technology Cottbus-Senftenberg, 03046 Cottbus, Germany

[6] Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History LUOMUS, University of Helsinki, Pohjoinen Rautatiekatu 13, 00014 Helsinki, Finland

[7] Molecular Ecology Group (MEG), Water Research Institute (IRSA), National Research Council (CNR), Corso Tonolli, 50, 28922 Pallanza, Italy

[8] School of Biological Sciences, University of Western Australia, Crawley, WA 6009, Australia

[9] Department of Evolutionary Biology, Ecology and Environmental Sciences & Biodiversity Research Institute (IRBio), Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Spain

[10] Fundación Protectora Ambiental Planadas Tolima (FUPAPT), Tolima, Colombia

[11] Instituto de Biología Subtropical (UNAM-CONICET), Puerto Iguazú, Argentina

[12] Grupo de Investigaciones Entomológicas de Tenerife (GIET), La Laguna, Canary Islands, Spain

[13] Centro Universitario Regional del Este, Universidad de la República, Treinta y Tres, Uruguay

[14] Department of Biology, Ecological Synthesis and Biodiversity Conservation Lab, Federal Rural University of Pernambuco, Dom Manuel de Medeiros, s/n, Dois Irmãos - CEP, Recife, PE, Brazil

[15] Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina, 13, 10123 Turin, Italy

[16] UMR CNRS 6553 ECOBIO, Université de Rennes 1, Rennes, France

[17] Departamento de Biología Animal, Edafología y Geología, Universidad de La Laguna, Tenerife, Spain

[18] Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" – CONICET, Av. Ángel Gallardo 470, C1405DJR, Buenos Aires, Argentina

[19] CE3C – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores, Angra do Heroísmo, Azores, Portugal

[20] Department of Forest Ecology, Faculty of Forestry and Wood Technology, Mendel University in Brno, Brno, Czech Republic

[21] Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[22] Te Aka Mātuatua School of Science, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
[23] Crop Research Institute, Drnovská 507, CZ-16106 Prague 6, Czechia
[24] Biology Centre, Czech Academy of Sciences, Institute of Entomology, Branišovská 31, 370 05 České Budějovice, Czechia
[25] Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología (IPNA-CSIC), 38206, La Laguna, Tenerife, Canary Islands, Spain
[26] Natural Sciences, Auckland War Memorial Museum, Parnell, Auckland 1010, New Zealand
[27] Division of Ecology and Biodiversity, School of Biological Sciences, the University of Hong Kong, Hong Kong

Corresponding author(s): Stano Pekár (pekar@sci.muni.cz)

Abstract
Spiders are a highly diversified group of arthropods, play an important role in terrestrial ecosystems as ubiquitous predators, making them a suitable group to test a variety of eco-evolutionary hypotheses. For this purpose, knowledge of a diverse range of species traits is required. Until now, data on spider traits have been scattered across thousands of publications produced for over two centuries and written in diverse languages. To facilitate access to such data we developed an online database for archiving and accessing spider traits at a global scale. The database has been designed to accommodate a great variety of traits (e.g., ecological, behavioural, morphological) measured at individual, species or higher taxonomic levels. Records are accompanied by extensive metadata (e.g., location, method). The database is curated by an expert team, regularly updated and open to any user. A future goal of the growing database is to include all published and unpublished data on spider traits provided by experts worldwide, and to facilitate broad cross-taxon assays in functional ecology and comparative biology.

Database URL: https://spidertraits.sci.muni.cz/

Introduction
With almost 50,000 species described to date (1), spiders are among the most diverse orders of terrestrial arthropods (2). Spiders rank among the most dominant arthropod predators in a huge variety of ecosystems, and therefore provide important ecosystem services, such as biological control (3,4) and bio-indication (5). They are also potentially an important source of molecules to be used in new biotechnologies and human medicine (6,7). In addition to these uses, spiders provide suitable models to test a breadth of ecological and evolutionary hypotheses (8-10).

Successful use of spiders for research and environmental assessments is based on knowledge of traits (morphological, ecological, physiological or behavioural characteristics), which characterise responses to environmental conditions and both change and define the effects of spiders on ecosystem functioning (10). Assembling trait values for species in a community is laborious, because for some traits and species this information either does not exist or is not easily available, as it is hidden in old publications (often not in English), unpublished records, technical reports, or even field notes. Although difficult to access, the data available are extensive, as research on spiders has covered a huge diversity of topics for over 200 years (11). Data on spider traits continues to be generated on a daily basis, most of it being used in individual publications or retained in unpublished datasets. Trait data are stored in different places and forms, and most data that originated before the use of personal computers are only available from printed publications. More recently, collected data are often stored in digital form in different repositories (from personal computers to data archive servers), but it is often difficult to compile and standardize datasets with different formats, and completeness of metadata, which are necessary for leveraging data for common purposes, as pointed out in the concept of Essential Biodiversity Variables (12, 13).

Trait databases already exist for a number of taxonomic groups, such as plants (14), corals (15), reptiles (16), copepods (17), and ground beetles (18), with a similar aim to accumulate and organize available data in a single repository. The success of such databases can be seen in their frequent use by many scholars (19). A general database of spider traits has not yet been developed. However, a range of spider traits can currently be found in several online resources, for example, the body size of European species (20), cytogenetic data (21), protein toxins of spiders (22), habitat and phenology of British (http://srs.britishspiders.org.uk/) and Czech spiders (http://arachnobaze.cz/), and various traits of ground-dwelling spiders (https://portail.betsi.cnrs.fr).

A single database that accommodates all trait data would enable scientists to investigate spiders more effectively and to perform large-scale comparative analyses (23-29). A trait-based approach has the advantage that some investigations (e.g. bio-indication) can be performed even when the taxonomic identity is missing or inaccurate (using morphospecies, for example) (30). Using trait, instead of taxonomic information, also allows for a comparison of community patterns and responses across regions with different species pools (32). For these purposes, it is important that trait data are available in appropriate quality and quantity, and have broad taxon and regional coverage. Overcoming these barriers will foster collaboration among arachnologists and other researchers that aim for multi-taxa analyses (24, 31-33).

Recently, Lowe et al. (10) initiated the establishment of a centralised database that aims to cover all spider traits and store data in a single place under FAIR (findable, accessible, interoperable, reusable) principles (34). Lowe et al. (10) built the foundation of such a database by detailed coverage of the trait definition, their standardisation, input data types, database governance, geographical coverage, accessibility, quality control, and sustainability. Furthermore, Lowe et al. (10) recognised that the unification of the trait records can only be accomplished by careful examination of the data during the validation procedure.

Following the initiative (10), here we present a curated global database that follows the FAIR principles and hosts a variety of traits recorded for spiders (Fig. 1). With the potential to grow indefinitely, we have already collected data for more than 7,000 spider taxa so far. The database has two main goals: 1) to collect and curate trait data on spiders from different sources, either (un)published or to be published in the future; and 2) to provide public access to these data under a CC BY licence, facilitating their widespread use by researchers.

Methods

Definitions

We adopted a broad definition of traits for inclusion in our database: any measurable phenotypic (i.e., morphological, ecological, physiological, behavioural, etc.) characteristic of an individual or taxon. This may also include ‘pure’ (heritable) traits (35), as well as the response to environmental conditions or a treatment (36, 37). Traits can be either quantitative (continuous, integers, proportions) or categorical (qualitative, binary, and ordinal). Trait values can represent individual-level measurements (single observation) to higher taxonomic (species-, genus-, family-) level measurements (aggregates), often recorded as a statistic (mean, median, minimum, maximum). We do not consider descriptive molecular data (such as DNA or protein sequences) or faunistic records to be traits, unless these contain reference to some trait (e.g., habitat type), as these have already established repositories such as GenBank® or the Global Biodiversity Information Facility (GBIF).

The definition of specific traits (including units for numerical traits or eligible values for categorical traits) was adopted from widely used definitions in a variety of published papers on spiders. To achieve semantic interoperability, each trait is described by standardized terms (Table S1). Two types of ontologies, describing the process of data collection and the traits themselves, were implemented during development of the database structure, as suggested by Kissling et al. (12). The process of measurement, i.e. details of data collection, is provided as metadata and the trait measured is given in the main table (see below).

To increase the interoperability of this database with other databases, the next step in the update of the database will be setting up an expert team to develop ontologies, detailed vocabularies, and a hierarchical structure for all traits. Some traits thus might be re-defined. This will not affect the current content but will prepare space for a harmonised collection of future data.

Database structure

We developed an online application and architecture called the World Spider Trait database, currently in version 1.0 (https://spidertraits.sci.muni.cz/), to store and retrieve trait data on spider species (Fig. 2). The database is able to accommodate traits measured at any taxonomic level. As many trait values show variation (phenotypic plasticity) as a response to varying conditions, each trait record can be accompanied by extensive metadata, describing the conditions under which it was measured (such as treatment, sampling method, geographic location, habitat, date). The database was built to meet the FAIR principles: it is available at a public domain under an open-access licence in a machine-readable format. This is enhanced by comprehensive online search options, and export capabilities.

The database has multi-layered structure. It is composed of a main table (Fig. 1) including five mandatory variables, namely (1) Original species name (taxon name as reported in the original source), (2) Trait abbreviation (unique abbreviation of each trait), (3) Trait value (measured value of a trait), (4) Method abbreviation (unique abbreviation of each method used to measure a trait), and (5) Reference abbreviation (unique abbreviation of each source). Several other variables are optional, namely WSC LSID (unique taxon identifier taken from the World Spider Catalog), Trait category (see beow), Trait name, Trait description, Trait data type, Trait unit, Measure (type of the measured value), Life stage (ontogenetic stage), Sex, Frequency (relative frequency of occurrence), Sample size (total number of observations per record), Treatment (treatment conditions), Method name (see below), Method description, Location abbreviation (unique identifier of a location), Latitude, Longitude, Altitude, Locality (the name or description of the place), Country, Habitat (habitat type according to a local classification), Microhabitat, Date, Note (any note related to a record), Row link (unique identifier of related measurements), and Reference (full reference). For a detailed description of each variable and examples see Table 1.

In the backend of the application, there are five additional metadata tables (extensions) which provide auxiliary information: (1) Taxa, (2) Locations, (3) Traits, (4) Methods, and (5) References. The Taxa table includes valid species or subspecies name, genus, family, LSID

(Taxonomic identifier automatically retrieved from the World Spider Catalog (1), taxonomic authority and year. The content of this table is automatically updated on a weekly basis from the spider nomenclature information available in the World Spider Catalog (1), which contains valid Latin names and synonyms. Morpho-species do not have valid species names, thus higher level categories (e.g. genus) are used, optionally accompanied by additional information provided by the uploader in the Note field. The Locations table includes country code, country name, locality name, coordinates, and its abbreviation. The Traits table contains Trait name, Category, Description, Data type, Unit, and its abbreviation. The Methods table includes method name, description and its abbreviation. References table includes full reference and its abbreviation. For more details see Table 1.

We defined 175 traits that are currently grouped into 12 categories according to the discipline (Anatomy; Biomechanics; Communication; Cytology; Defence; Ecology; Life-History; Morphology; Morphometry; Physiology; Predation; Reproduction) (Table S1). Information on the way a trait was measured is described in the Methods table. The provision of this metadata is mandatory during upload to ensure comparability of data. The Methods list includes field collection techniques, as well as laboratory methodologies. Currently, there are 37 methods defined (Table S2). The included pre-defined traits, categories, and methods are meant to cover the majority of traits and methodologies in spider research. However, the architecture of the database is flexible enough that further traits, categories and methods can be added in the future, to accommodate new trait types and novel methodologies.

This database is hosted, developed, and maintained at the Department of Botany and Zoology of Masaryk University in collaboration with the University IT centre. It is connected to the World Spider Catalog (1), administered and curated by the core team members (Fig. 2).

<u>Data upload procedure</u>

Upon collection the data must be harmonised. Before a dataset can be submitted to the database, the data must be in a valid format (for a detailed description see https://github.com/oookoook/spider-trait-database/blob/master/docs/template.md). For this purpose, we developed a MS Excel spreadsheet (Template) that should fit the great majority of trait types with predefined columns. The spreadsheet was designed to enable easy data manipulation by classical statistical software, such as R (38). The template can be downloaded from the World Spider Trait database webpage (https://spidertraits.sci.muni.cz/contribute). It contains 31 columns, some of which are mandatory so they must be filled with appropriate numerical or character values. Eligible values for all columns can be found in the header of each variable, in the List of Traits (Table S1), and List of Methods (Table S2). If the input trait or method is not already defined, the contributor should provide all of the following information to create a new trait or method: Trait category, trait name, trait description, trait data type, and trait unit in the case of missing traits, or method name and method description in the case of missing methods. Similarly, for references, the contributor either provides an abbreviation of a reference if it is in List of References or a full reference. Unpublished data are referenced as personal observations.

The data in the template then needs to be saved either as an .xls(x) or a comma-delimited .csv file, and the file should be encoded as UTF-8 to assure compatibility with special (regional) characters. Once the template is uploaded the contributor must approve it using the tools within the web application.

Software used

The code of the web application is stored at GitHub (https://github.com/oookoook/spider-trait-database) and is available under the GNU GPL v 3.0. The phylogenetic tree was produced using functions within ape package (39) within R (38).

Results and Discussion

Data Records

Integration of data from different sources was based on standardisation and harmonisation. This involved conversion of trait values to comparable units/trait, use of controlled vocabulary in the definition of traits, standardisation of eligible character values, and use of single spreadsheet format. Each record was accompanied by licence information and the original source.

Currently, both published (from more than 1,000 publications) and unpublished data from diverse study designs (both descriptive and experimental) are included in the database, with the citation of the original source. So far 70 datasets have been contributed, with a total number of more than 221,000 records belonging to more than 7,500 taxa. Of these, 40 datasets (34.1% of records) are unlocked (i.e., freely accessible without user registration). The remainder (i.e., embargoed datasets) are previously unpublished data compilations and can be viewed and downloaded by registered users only, to ensure applicable authorship credits (see 'Usage Notes'). Registration and data usage is free under a CC BY licence. Embargoed data compilations may eventually become unlocked (e.g., once these have been used in published studies).

Geographical coverage of the database is global, but currently there are more records from Europe and South America than from other continents (Fig. 3)—a typical bias in biodiversity research (40). Data on taxa from North America, Africa, and Asia are represented by very few records. The great majority of records available now come from Europe. Specifically, 20 datasets (66.1% of records) concern European species. This includes data on body size (2024 species), light & moisture preferences (1949 species), guild classification (1017 species), and conservation status (1557 species). In terms of traits, anatomical, behavioural, and physiological data are largely missing.

As for the taxonomic coverage, of 129 known spider families (1) only two (Euctenizidae and Penestomidae) have no records in the database so far (Fig. 4). Several families (e.g., Gnaphosidae, Lycosidae, Salticidae, Sicariidae, Theridiidae) each have data for more than 40% of the 138 traits, but 38 families still have fewer than 5% of all traits covered. As for the number of records per family, most records come from the most speciose families, namely Linyphiidae, followed by Lycosidae, Theridiidae, and Salticidae (Fig. 5A). Because not every trait has been measured for every taxon, the taxon × trait matrix is highly incomplete (2.82% completeness, Fig. 5B). This is to be expected for a highly diverse and severely understudied taxonomic order. With respect to sex/stage, there are 33.6% records for adult males, 55.8% adult females, and 8.6 % for juveniles.

The content of the database reflects real historical differences among geographic areas and disciplines. The database thus can be used to identify gaps and help to prioritise future areas for investigation to achieve more complete sets of records. To fill these gaps we plan to encourage contributions from specific areas, traits, and trait categories in the future. This can include collection of data from other repositories, extraction of data from publications, and archiving currently produced data. We will also ask curators of specialised spider trait databases to provide their data to be centrally stored here. Since many funders and journals now require data to be made publicly available, the database can be used as a permanent data archive option (an alternative to, e.g., Dryad or Figshare), provided that each contributed dataset meets the standards of the database format, which allows efficient reuse and synthesis. Each dataset obtains a unique URL and, in near future, it will be associated with a DOI provided by DataCite. In the future we expect to mainly gather data on new traits and new taxa, and would like to encourage colleagues to contribute their datasets of both published and unpublished data. A coordinated effort is needed to achieve this goal.

To promote the process of data collection, we invite arachnologists to download the template and use it for data storage on their personal computers. At the same time, we ask arachnologists to get used to the vocabulary of the database, adopt definition of the traits that are used here (or suggest alternatives), and develop protocols that follow the same standards. This will markedly enhance integration of their datasets into the database.

Data Validation

Validation is performed at several steps during submission, in order to retain only high quality records.

Firstly, a contributor is advised to search through the current database content in order to ensure that such (exact) data are not already included for the taxon/taxa under investigation. It is also useful at this point to check whether the proposed trait(s) and method(s) are already defined. Contributors become eligible to upload their dataset after requesting registration from the administrator.

To upload a new dataset, a contributor must specify the name of the dataset, their full name and email address. In addition, a contributor can specify the authors of the dataset, author emails, mark whether the data can be immediately accessed or are under an embargo, and add any note. Then, the dataset sheet is created and the contributor is able to upload the data. The data is then imported to the temporary cache. During the upload process the web application checks the presence of eligible values in the following variables (Original name, Trait abbreviation, Value, Measure, Sex, Life stage, Frequency, Sample size, Method abbreviation, Latitude, Longitude, Altitude, Country, Date, Reference) and identifies duplicate records. Invalid records are highlighted to facilitate corrections. The taxonomy check includes existence of the name and match with a current valid name according to the World Spider Catalog (1).

At this stage, the contributor can view the dataset and must edit invalid cells in order to comply with the database requirements. Editing is done using the web application tools. When the contributor completes all changes and the dataset is valid, it can be sent to the administrator or editor for review. The contributor can include a message to the editor when submitting the dataset for review, in which the contributor can explain any problems they had encountered while editing the dataset.

The administrator or editor is informed of a new dataset submission by an email. The dataset enters a second validation phase, which can only be done by the administrator or editor. The administrator or editor must add new trait(s) and method(s) to the database, check for additional errors, such as extreme (unlikely) values of traits (e.g., resulting from typos, wrong digit separator, etc.), imprecise definition of new traits and methods, or an incorrect format of references. Once the dataset is validated by the administrator or editor it is published in the database. This means that all the data are transferred from the temporary import cache to the main database and becomes available to the general public, unless embargoed. If the administrator or editor observes any problems, the dataset is rejected and sent back to the contributor with an email containing a description of the problem(s) to be fixed. Any dataset can be *post-hoc* corrected/altered by the administrator or editor without contributors' consent.

<u>Data Usage</u>

A user can view the whole content of the database using the Data Explorer within the online application. In the Data Explorer, the user can apply filters (Family, Genus, Species, Original name, Trait category, Trait, Method, Location, Country, Dataset, References, Row links) to display selected content. The result can be displayed in a spreadsheet or in bar figure window. Selected data can then be downloaded in a .csv or .xlsx format. If the selected data contain data from datasets under embargo the user is given a warning. In order to download embargoed data the user has to send a request to the administrator or editor, who will then contact the dataset authors. All data with or without embargo can be download only after receiving login data.

In addition, the database provides an Application Programming Interface (API) to allow access to data via web platforms or software. An R package, named spidR (41), with few easy-to-use functions that allow downloading and pre-processing data from the database, is now available. Resulting data frames can then be analysed with a variety of tools available in R (38). Access of the embargoed data via API requires login as well.

As the trait value data can be a mixture of various statistics, it is important that the user checks the 'Measure' variable of each record and adopts appropriate procedures prior to analysis. Furthermore, due to inherent variation in most trait values, the user must consider conditions (such as habitat, altitude, treatment, etc.) under which it was measured. Not all conditions (e.g. hunger state, mating status, etc.) are recorded in the auxiliary variables, thus the user is strongly advised to study the original publication.

A number of traits included in this database are candidates of Essential Biodiversity Variables proposed by the Group on Earth Observations Biodiversity Observation Network (12, 13). The traits are recorded with many metadata, and thus allow quantification of intra-specific variation with respect to environmental conditions, space and time. These traits can be of societal relevance, as they can be used in study spread of invasive species or biodiversity change.

Although the use of data is free, users are strongly encouraged to contribute their data, particularly if they have not contributed yet, following the simple 'first give, then take' principle. Only by these means will the database grow in quantity and frequency of use.

Contained data are publicly available under a Creative Commons Attribution license (CC BY 4.0), so that anyone can use received data under the condition of appropriate citation of this publication. In the case of datasets that have not been published and are under embargo, the user must agree with the dataset contributor on the conditions of use. Typically, this should include citation (URL or DOI) of the specific dataset, in addition to the database citation.

Acknowledgements

We would like to thank different Masaryk University bodies, namely MUNI Press and IT Centre, for their overall support, and Pentti Tuomikoski Grant from the Finnish Museum of Natural History for financial support of the core group meeting. Furthermore, we wish to thank N. Scharff, E. Soto, and W. Fannes for contributing data. MŘ was supported by the Czech Ministry of Agriculture (MZe RO0418).

Author contributions

SP: co-developed the application, contributed and validated data, wrote the manuscript; KB, SM: involved in designing the database principles and structure, contributed data, revised the manuscript; AK: developed the application, revised the manuscript; JW, PC: involved in designing the database principles and structure, coordinated data acquisition, contributed data, revised the manuscript; LČ: coordinated data acquisition, contributed and validated data, revised the manuscript; CSF, ECL, MEH: contributed to designing the database principles, revised the manuscript; BAB, LC, AED, MD, AVE, YMGPE, SF, LFG, TGS, MI, EL, NMH, IM, JMO, OM, PM, RM, FM, AM, WN, GN, CJP, EP, MJR, CR, MŘ, AR, DL, KP, JP, VR, IS, LS, MS, LW, KW, GAZ: contributed data, revised the manuscript.

Competing interests

All authors have no conflict of interest.

References

1 World Spider Catalog. (2021) *World Spider Catalog*. Version 22.0. Natural History Museum Bern, http://wsc.nmbe.ch, doi: 10.24436/2

2 Dimitrov,D. and Hormiga,G. (2021) Spider diversification through space and time. *Annu. Rev. Entomol.*, **66**, 225–241.

3 Nyffeler, M. and Birkhofer, K. (2017) An estimated 400–800 million tons of prey are annually killed by the global spider community. *Sci. Nat.*, **104(3-4)**, 30.

4 Michalko,R., Pekár,S., Duľa,M. and Entling,M.H. (2019) Global patterns in the biocontrol efficacy of spiders: A meta-analysis. *Global Ecol. Biogeogr.*, **28(9)**, 1366–1378.

5 Cardoso,P., Arnedo,M.A., Triantis,K.A. et al. (2010) Drivers of diversity in Macaronesian spiders and the role of species extinctions. *J. Biogeogr.*, **37**, 1034–1046.

6 Heim,M., Keerl,D. and Scheibel,T. (2009) Spider silk: from soluble protein to extraordinary fiber. *Angew. Chem. Int. Edit.*, **48(20)**, 3584–3596.

7 Matavel,A., Estrada,G. and De Marco Almeida,F. (2016) Spider Venom and Drug Discovery: A Review. In: Gopalakrishnakone P., Corzo G., de Lima M., Diego-García E. (eds) Spider Venoms. Toxinology. Springer, Dordrecht, pp. 273–292.

8 Liu,J., May-Collado,J.L., Pekár,S. et al. (2016) A revised and dated phylogeny of cobweb spiders (Araneae, Araneoidea, Theridiidae): A predatory Cretaceous lineage diversifying in the era of the ants. *Mol. Phylogenet. Evol.,* **94(B)**, 658–675.

9 Pekár,S., Coddington,J.A. and Blackledge,T. (2012) Evolution of stenophagy in spiders (Araneae): evidence based on the comparative analysis of spider diets. *Evolution*, **66(3)**, 776–806.

10 Lowe,E.C., Wolff,J.O., Aceves-Aparicio,A. et al. (2020) Towards establishment of a centralized spider traits database. *J. Arachnol.*, **48(2)**, 103-109.

11 Mammola,S., Michalik,P., Hebets,E.A. et al. (2017) Record breaking achievements by spiders and the scientists who study them. *PeerJ*, **5**, e3972.

12 Kissling,W.D., Walls,R., Bowser,A., et al. (2018a) Towards global data products of Essential Biodiversity Variables on species traits. *Nat. Ecol. Evol*., **2(10)**, 1531-1540.

13 Kissling,W.D., Ahumada,J.A., Bowser,A., et al. (2018b) Building essential biodiversity variables (EBV s) of species distribution and abundance at a global scale. *Biol. Rev.,* **93(1)**, 600-625.

14 Kattge,J., Diaz,S., Lavorel,S., et al. (2011) TRY–a global database of plant traits. *Glob. Change Biol.,* **17(9)**, 2905-2935.

15 Madin,J.S., Anderson,K.D., Andreasen,M.H., et al. (2016) The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Sci. Data*, **3(1)**, 1-22.

16 Grimm, A., Ramírez, A. M. P., Moulherat, et al. (2014) Life-history trait database of European reptile species. *Nat. Conserv.*, **9**, 45.

17 Brun, P., Payne, M. R. and Kiørboe, T. (2017) A trait database for marine copepods. *Earth Syst. Sci. Data*, **9(1)**, 99-113.

18 Homburg, K., Homburg, N., Schäfer, F., et al. (2014) Carabids.org–a dynamic online database of ground beetle species traits (Coleoptera, Carabidae). *Insect Conserv. Divers.*, **7(3)**, 195–205.

19 Kattge,J., Bönisch,G., Díaz,S., et al. (2020) TRY plant trait database–enhanced coverage and open access. *Glob. Change Biol.*, **26(1)**, 119-188.

20 Nentwig,W., Blick,T., Bosmans, R., et al. (2021) *Spiders of Europe*. Version 2.2021. https://www.araneae.nmbe.ch, https://doi.org/10.24436/1

21 Araujo,D., Schneider,M.C., Paula-Neto,E. et al. (2021) *The spider cytogenetic database.* www.arthropodacytogenetics.bio.br/spiderdatabase.

22 Wood,D.L., Miljenović,T., Cai,S., et al. (2009) ArachnoServer: a database of protein toxins from spiders. *BMC Genomics*, **10,** 375.

23 Cardoso,P., Pekár,S., Jocqué,R. et al. (2011) Global patterns of guild composition and functional diversity of spiders. *PloS One*, **6(6)**, e21710.

24 Birkhofer,K., Wolters,V. and Diekötter,T. (2014) Grassy margins along organically managed cereal fields foster trait diversity and taxonomic distinctness of arthropod communities. *Insect Conserv. Divers.*, **7(3)**, 274–287.

25 Wolff, J.O., Nentwig,W. and Gorb, S.N. (2013) The great silk alternative: multiple co-evolution of web loss and sticky hairs in spiders. *PLoS One*, **8(5)**, e62682.

26 Michalko,R. and Pekár,S. (2016) Different hunting strategies of generalist predators result in functional differences. *Oecologia*, **181(4)**, 1187–1197.

27 Pekár,S. (2012) Spiders (Araneae) in the pesticide world: an ecotoxicological review. *Pest Manag. Sci.*, **68(11)**, 1438–1446.

28 Pekár,S. (2014) Comparative analysis of primary defences in spiders (Araneae). *J. Anim. Ecol.*, **83(4)**, 779–790.

29 Pekár,S. and Toft,S. (2015) Trophic specialisation in a predatory group: the case of prey-specialised spiders (Araneae). *Biol. Rev.*, **90(3)**, 744–761.

30 Moretti,M., Dias,A.T., De Bello,F., et al. (2017) Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. *Func. Ecol.*, **31(3)**, 558-567.

31 Gallagher, R.V., Falster, D.S., Maitner, B.S., et al. (2020) Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecol. Evol.*, **4**, 294–303.

32 Birkhofer, K., Smith, H. G., Weisser, W. W., et al. (2015) Land-use effects on the functional distinctness of arthropod communities. *Ecography*, **38(9)**, 889–900.

33 Birkhofer,K., Gossner,M.M., Diekötter,T., et al. (2017) Land-use type and intensity differentially filter traits in above-and below-ground arthropod communities. *J. Anim. Ecol.*, **86(3)**, 511–520.

34 Wilkinson,M.D., Dumontier,M., Aalbersberg, I.J., et al. (2016) Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.

35 Wong,M.K., Guénard,B. and Lewis,O.T. (2019) Trait-based ecology of terrestrial arthropods. *Biol. Rev.*, **94(3)**, 999–1022.

36 McGill, B. J., Enquist, B. J., Weiher, E. and Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol.*, **21(4)**, 178-185 .

37 Schneider,F.D., Fichtmueller,D., Gossner,M.M., et al. (2019) Towards an ecological trait-data standard. *Methods Ecol Evol.*, **10(12)**, 2006–2019.

38 R Core Team. (2020) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

39 Paradis, E. and Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35,** 526–528.

40 Titley,M.A., Snaddon,J.L. and Turner,E.C. (2017) Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PloS One*, **12(12)**, e0189577

41 Cardoso, P. (2021) ARAKNO: ARAchnid KNowledge Online. R package. https://cran.rstudio.com/web/packages/arakno/index.html version 1.1.0. https://CRAN.R-project.org/package=arakno

42 Wheeler,W.C., Coddington,J.A., Crowley,L.M., et al. (2017) The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics*, **33(6)**, 574–616.

Table 1. Content of the template file. For each variable there is its name, description and eligible values. Mandatory variables are indicated by an asterisk (*). Eligible values are predefined only for some variables. Examples are given in parenthesis.

Variable name	**Description**	**Eligible values or examples**
WSC LSID	Taxonomic identifier (URN) from the World Spider Catalog	(urn:lsid:nmbe.ch:spidersp:033381)
Original name*	Taxon name as reported in the original source	(Linyphiidae, *Zodarion* sp., *Pimoa rupicola*)
Trait abbreviation*	Abbreviation (see Table S1)	(indu)
Value*	Measured value of a trait	(110)
Measure	Type of the measured value	Single observation, mean, median, min, max, description.
Sex	Sex	Female, male, both, unknown.
Life stage	Ontogenetic stage	Egg, larva, juvenile, adult, all.
Frequency	Relative frequency of occurrence	(0.43)
Sample size	Total number of observations per record	(45)
Treatment	Treatment and conditions at which it was measured	(Effect of a pesticide, type of prey, wavelength, temperature)
Method abbreviation*	Abbreviation (see Table S2)	(ptf)
Latitude	The geographic latitude (in decimal degrees or other widely used formats)	(45.74, -37.22285)
Longitude	The geographic longitude (in decimal degrees or other widely used formats)	(102.478922, -0.4767)
Altitude	Altitude of the location (above sea level in meters)	(567)
Locality	The name or description of the place	(Municipality of Helsinki, small hill close to the river, Mount Fuji)
Country	The standard code for the country	According to ISO 3166 (CZ, IT, BR, CZE)
Habitat	Habitat type according to a local classification, such as EUNIS	(Pine forest, grassland, cave)
Microhabitat	Microhabitat type	(Under stones, ground, canopy)
Date	The date-time or interval	(1963-03-08T14:07-0600, 2009-02-20T08:40Z, 2018-08-29T15:19 - 3:19pm, 1906-06, 1971)
Note	Any note related to information provided	(Habitat classification, experimental procedure)
Row link	Unique identifier marking related data (same individuals)	(a1)
Reference*	Full reference of the published or unpublished data	(Journal: Elias DO, Hebets EA, Hoy RR & Mason AC. 2005. Seismic signals are crucial for male mating success in a visual specialist jumping spider (Araneae: Salticidae). Animal Behaviour 69(4): 931–938. Book: Preston-Mafham R. 1990. The Book

	of Spiders and Scorpions. London, Quantum Books. Book Chapter: Nentwig W. 1987. The prey of spiders. In Nentwig W (Ed.), Ecophysiology of Spiders. Berlin, Springer-Verlag, pp. 249–263. Website: Nentwig W, Blick T, Bosmans R, Gloor D, Hänggi A, Kropf C (2021) Spiders of Europe. Online at https://www.araneae.nmbe.ch. Unpublished: Michalko R, pers. comm.

Fig. 1. A scheme of the database structure. There is the main table connected to five metadata tables. * marks mandatory variables. Examples of trait categories are given on the right. Photos: S. Pekár.

Fig. 2. The scheme of the World Spider Trait database application, depicting the role of contributing bodies and the frontpage of the webpage (https://spidertraits.sci.muni.cz/, accessed on March 5th, 2021). WSC stands for World Spider Catalog, MUNI stands for Masaryk University.

Fig. 3. Geographic coverage of the data currently in the database. Orange points represent geo-referenced records, while blue points are country centroids (for records that do not have a geographical reference). There are records from 70 countries and 479 locations. The map was created using Google Maps.

Fig. 4. Trait coverage mapped on the tree. The tree is on the family level (composed of 121 families) with the proportion of the total number of traits (orange) displayed as pie charts (the fuller the pie the more traits). The tree was constructed based on the most recent phylogeny of spiders (42). Five families (Hexurellidae, Mecicobothriidae, Megahexuridae, Microhexuridae, Myrmecicultoridae) were omitted because their position in the tree is not known.

Fig. 5. Quantitative content of the database. A. Number of records (logarithmically transformed) for each family included in the database, arranged alphabetically. B. The taxon by trait matrix representing the completeness. The most complete traits include body length (64% of taxa), followed by cephalothorax length (23%), and cephalothorax width (19%). Dots represent logarithmically transformed number of records per taxon. Taxon includes one of the following: subspecies, species, genus, or family.

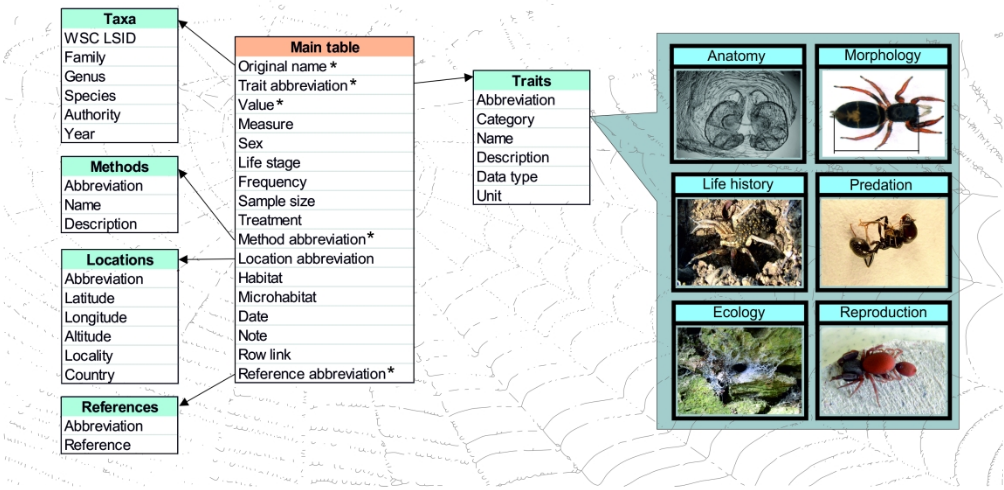

figure 1

241x130mm (300 x 300 DPI)

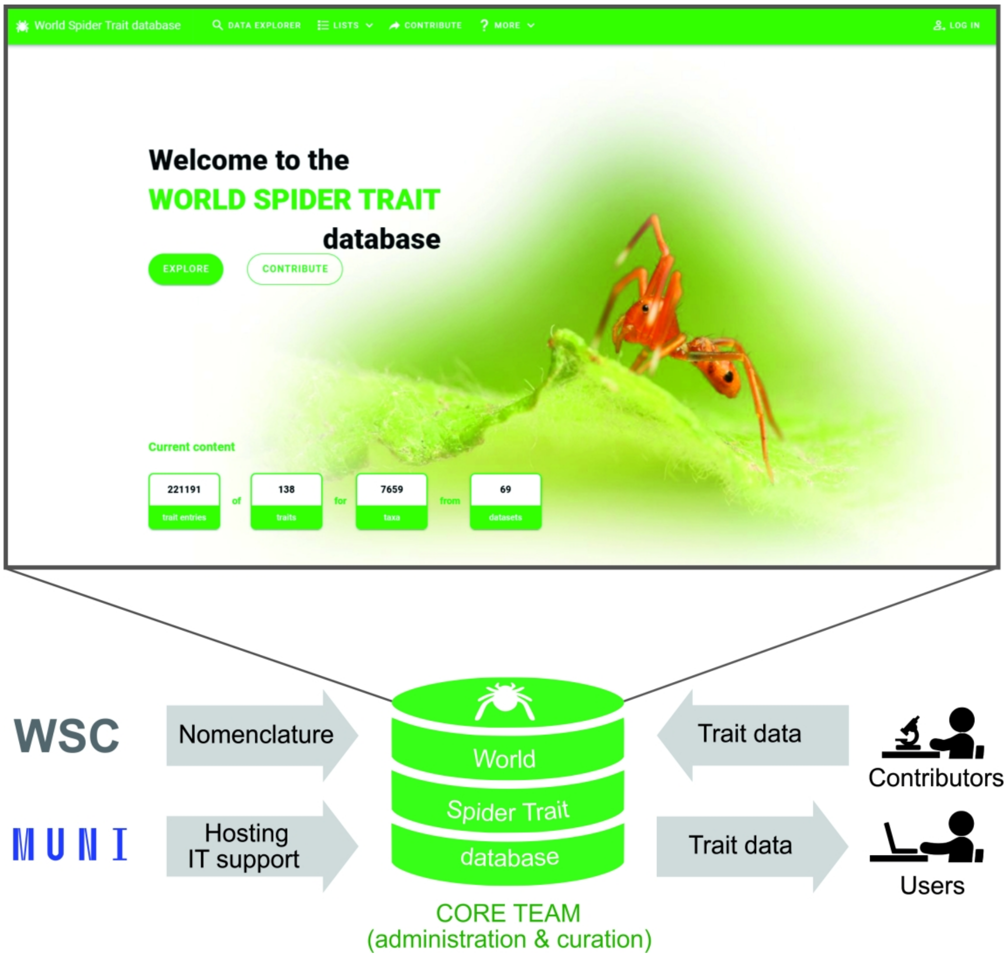

figure 2

231x219mm (300 x 300 DPI)

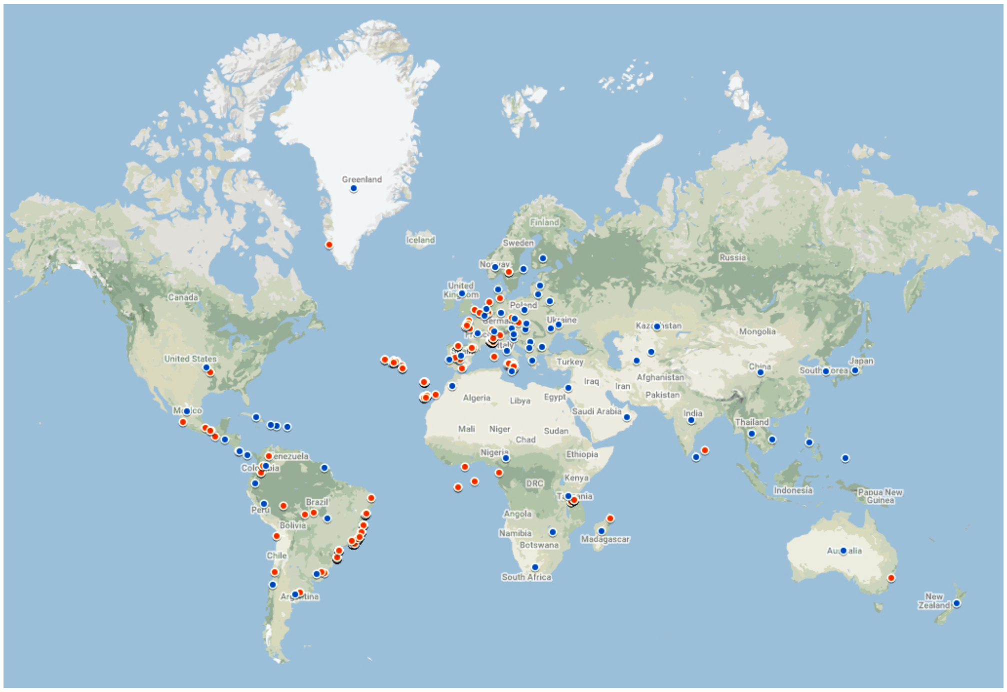

figure 3

685x469mm (38 x 38 DPI)

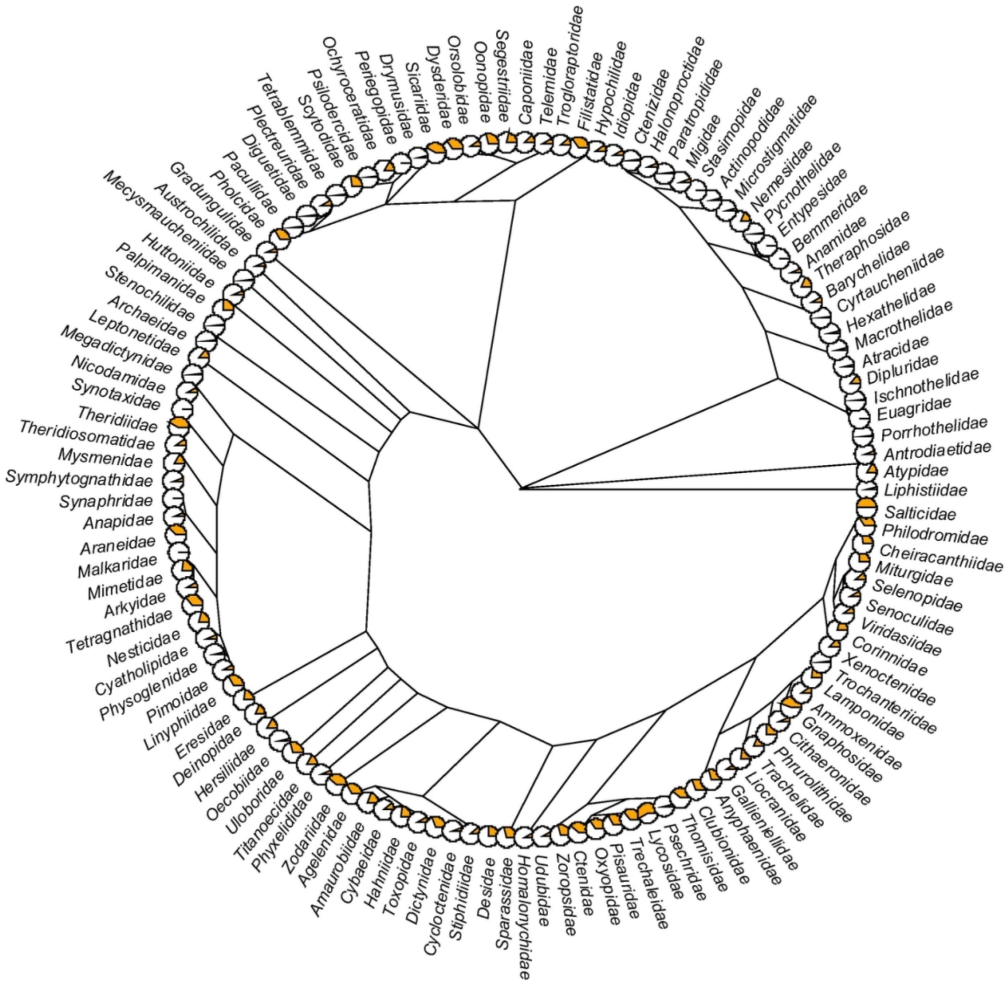

figure 4

118x115mm (300 x 300 DPI)

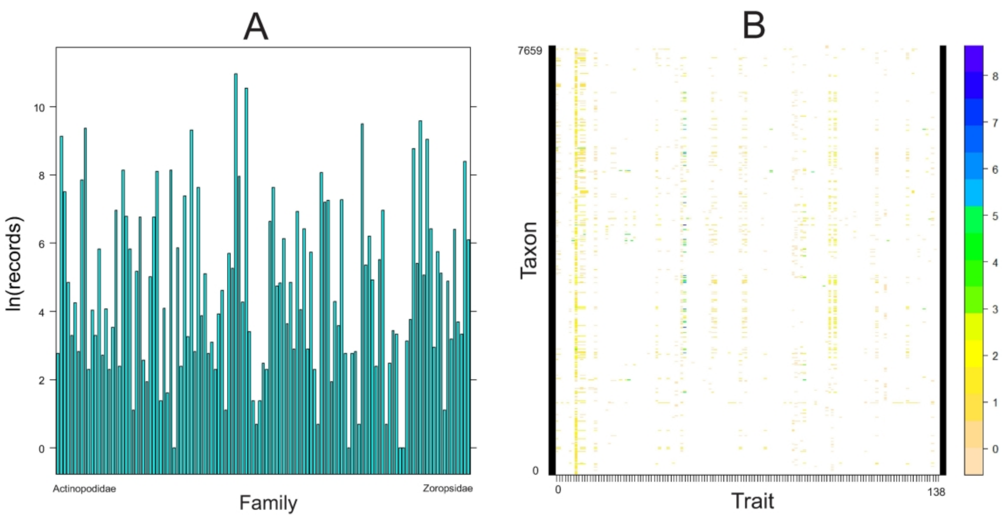

figure 5

224x113mm (300 x 300 DPI)

Table S1. List of traits, their description, types of data, units and eligible values, arranged alphabetically within trait categories.

Abbr.	Category / Name	Description	Data type	Unit	Eligible values
	Anatomy				
brsi	Brain size	Volume of central nervous system	real number	mm^3	
cuth	Cuticle thickness	Thickness of a cuticle on a body part	real number	mm	
nugl	Number of silk glands	Number of any silk glands per spinneret	integer		
prsy	Posterior respiratory system	Type of posterior respiratory system	character		Booklungs; tubular tracheae; lamella; absent
scle	Sclerotisation	Enhanced sclerotisation of prosoma	character		
sigl	Silk gland size	Volume of any of the silk glands	real number	mm^3	
stfo	Sperm transfer form	State in which sperm is transferred to the female	character		
vgsi	Venom gland size	Volume of a venom gland	real number	mm^3	
	Biomechanics				
adhe	Web adhesion	Adhesion of capture thread	real number	MPa	
cspd	Climbing speed	Climbing speed (moving on a slope)	real number	cm/s	
rspd	Running speed	Running speed (moving horizontally)	real number	cm/s	
stra	Silk strain	Engineering strain of silk	real number	mm/mm	
stre	Silk strength	Engineering strength of silk	real number	Mpa	
toug	Silk toughness	Toughness of silk	real number	MPa	
	Communication				
coop	Colouration of opisthosoma	Reflectance of dorsal side of opisthosoma at a certain wavelength (specify in Treatment variable)	real number	%	
copr	Colouration of prosoma	Reflectance of dorsal side of prosoma at a certain wavelength (specified in Treatment variable)	real number	%	
fred	Dominant frequency	The peak frequency of the sound produced	real number	Hz	
freq	Lower frequency range	The minimum frequency of the sound produced	real number	Hz	
freu	Upper frequency range	The maximum frequency of the sound produced	real number	Hz	
soun	Sound production	Mechanism of sound production	character		Drumming; stridulation

sour	Sound source	Organs used to produce sound	character		Chelicera/pedipalp; chelicerae; prosoma/abdomen; leg/leg
	Cytology				
chrn	Chromosome number	Diploid number of chromosomes	integer		
chrs	Sex chromosome system	The sex-chromosome system	character		X0; XX0
chrt	Chromosome morphology	Type of chromosomes according to the position of the centromere	character		Acrocentric; holocentric; metacentric
	Defence				
modl	Model	Model imitated by species using camouflage and mimicry	character		Ant; beetle; wasp; snail; twig; branch; flower; etc.
prde	Primary defence	A strategy used prior to being detected by a predator	character		Cryptic (background matching); Cryptic (disruptive coloration); Cryptic (countershading); Aposematic; Batesian mimicry; Camouflage; Müllerian mimicry
pred	Predator	Taxonomical classification of a predator	character		
retr	Retreat	Type of a retreat used to avoid predation	character		On web; in grass; under bark; silk sac; burrow; other
sede	Secondary defence	A strategy used after being detected by a predator	character		Death feigning; rapid escape; threatening posture; dazzle camouflage; startle; chemical deterrents; colour change; sound production
	Ecology				

balo	Ballooning	Developmental stage that disperses by ballooning	character		Juvenile; adult
circ	Circadian activity	Hours of a day when the species is active, i.e. foraging, mating, web-building	character		Diurnal; nocturnal; 1-24
disp	Dispersal time	Months at which dispersal occurs	character		January; February; March; April; May; June; July; August; September; October; November; December
girl	Global IUCN Red List category	Category of risk according to the global IUCN Red List guidelines	character		
habi	IUCN habitat	Habitat type according to the global IUCN classification	character		Forest; Savanna; Shrubland; Grassland; Wetlands; Rocky areas; Caves and Subterranean Habitats; Desert; Freshwater; Coastal; Urban, Agricultural; Other
halo	Habitat local CZ	Habitat type according to Czech habitat classification according to Chytrý M, Kučera T, Kočí M, Grulich V & Lustyk P. 2010. Habitat Catalogue of the Czech Republic. 2nd ed. Praha: Agentura ochrany přírody a krajiny ČR.	character		
lepr	Legal protection	Legal protection in national or subnational legislations	character		
ligh	Light	Light gradient according to Entling W, Schmidt MH, Bacher S, Brandl R & Nentwig W. 2007. Niche properties of Central European spiders: Shading, moisture and the evolution of the habitat niche. Global Ecology and Biogeography 16(4): 440–448.	real number		
lig2	Light 2	Light gradient according to Buchar J & Růžička V. 2002. Catalogue of spiders of the Czech Republic. Praha, Peres.	character		
mdl	Regional non-IUCN Red List category	Category of risk according to the regional guidelines (non-IUCN)	character		
miha	Microhabitat	Habitat where species occurs	character		among stones; bare ground; herbs; bushes;

					foliage; litter
moi1	Moisture 1	Moisture gradient according to Entling W, Schmidt MH, Bacher S, Brandl R & Nentwig W. 2007. Niche properties of Central European spiders: Shading, moisture and the evolution of the habitat niche. Global Ecology and Biogeography 16(4): 440–448.	real number		
moi2	Moisture 2	Moisture gradient according to Buchar J & Růžička V. 2002. Catalogue of spiders of the Czech Republic. Praha, Peres. Quantification of preference, where 1=preferred value, 0.5=primary value, 0.1=marginal value, can be given in Frequency variable.	character		Dry; semi-humid; humid
ovws	Overwintering stage	Developmental stage that overwinters	character		Egg; larva; juvenile; adult
phen	Phenology	Months at which adult stage occurs	character		January; February; March; April; May; June; July; August; September; October; November; December
rasi	Range size	Area of the species distribution range	real number	km^2	
regl	Regional IUCN Red List category	Category of risk according to the regional IUCN guidelines	character		
soci	Social degree	Degree of sociality	character		Solitary; subsocial; colonial; quasisocial; social
strt	Stratum	Horizontal stratum occupied	character		Underground; ground; herb layer; shrub layer; tree trunks; canopy; wall
suaf	Subterranean affinity	Degree of subterranean affinity	character		Troglobiont; troglophile
urha	Urban habitat	Affinity for urban habitats: % of urban habitats (i.e. impervious surfaces) in a buffer of 1600 m of radius around the sampling point	real number	%	
	Life History				
indu	instar duration	Number of days spent in a certain ontogenetic stage (egg, larva, or instar) at a certain temperature (specified in Treatment variable)	integer	days	
inst	Number of instars	Total number of instars, beginning with the first free instar and ending with the adult stage	integer		

lonv	Longevity	Number of days from hatching to death	integer	days	
mort	Mortality	Mortality either natural or due to any treatment	real number	%	
sexr	Sex ratio	Number of males divided by the number of females	real number		
surv	Survival	Proportion of surviving individuals	real number	%	
	Morphology				
crib	Cribellum	Presence of functional cribellum and calamistrum (e.g., present)	character		
ente	Entelegyne	Presence of epigyne in females (e.g., present)			
eyes	Eye number	Total number of eyes	integer		
flat	Body flattening	Significantly flattened body as an adaptation to shelter in crevices	character		
nusp	Spine number	Number of erectable spines (macrosetae) on the prolateral side of leg I (that play a role in the formation of the capture basket)	integer		
scoa	Scopula area	Area of scopula hairs on leg segments	real number	mm^2	
scod	Scopula density	Number of scopula hairs per area on a leg segment	real number	mm^2	
spin	Spinnerets	Total number of functional spinnerets	integer		
	Morphometry				
abhe	Abdomen height	Opisthosoma height at highest point	real number	mm	
able	Abdomen length	Opisthosoma length from anterior to posterior along longitudinal axis (excl. petiole and spinnerets)	real number	mm	
abwi	Abdomen width	Opisthosoma width at widest point	real number	mm	
aled	ALE	Diameter of one anterior lateral eye	real number	mm	
alsl	Spinneret ALS	Total length of anterior lateral spinneret (from base to tip)	real number	mm	
amed	AME	Diameter of one anterior median eye	real number	mm	
bodm	Body mass	Body mass (in a normal nutritional condition)	real number	g	
bole	Body length	Total body length (from carapace frontal, excl. chelicerae, to opisthosoma posterior, excl. spinnerets)	real number	mm	
cehe	Cephalothorax height	Height of prosoma at the highest point (from sternum most ventral to carapace most dorsal)	real number	mm	
cele	Cephalothorax length	Length of prosoma (carapace) along the longitudinal body axis	real number	mm	
cewe	Cephalothorax width	Width of prosoma (carapace) at the widest point	real number	mm	
chle	Chelicerae basal	Length of cheliceral base segment (paturon) along external margin	real number	mm	

	part (paturon) length				
cox1	Coxa I length	Coxa length of leg I	real number	mm	
cox2	Coxa II length	Coxa length of leg II	real number	mm	
cox3	Coxa III length	Coxa length of leg III	real number	mm	
cox4	Coxa IV length	Coxa length of leg IV	real number	mm	
criw	Cribellum width	Width of cribellum or colulus	real number	mm	
ctar	Claw tuft area	Projected area of adhesive foot pad (claw tuft) on leg IV	real number	mm^2	
ctde	Claw tuft density	Density of adhesive foot pad (claw tuft) on leg IV, i.e. number of tenant setae per area unit	integer		
eggs	Egg size	Diameter of an egg	real number	mm	
eggv	Egg volume	Volume of an egg	real number	mm^3	
epaw	Epigyne anterior plate width	Width of anterior border of epigyne plate	real number	mm	
epcw	Epigyne central plate width	Width of central border of epigyne plate	real number	mm	
eple	Epigyne length	Length of epigynal plate	real number	mm	
eppw	Epigyne posterior plate width	Width of posterior border of epigyne plate	real number	mm	
eyew	Eye region width	maximum width of eye region	real number	mm	
fale	Fang length	Cheliceral fang length from base articulation to the tip (measured along the median arc)	real number	mm	
fem1	Femur I length	Femur length of leg I (measured between condyles)	real number	mm	
fem2	Femur II length	Femur length of leg II (measured between condyles)	real number	mm	
fem3	Femur III length	Femur length of leg III (measured between condyles)	real number	mm	
fem4	Femur IV length	Femur length of leg IV (measured between condyles)	real number	mm	
l1le	Leg I length	Total length of one leg from the first (front) leg pair, excluding coxa and trochanter	real number	mm	
l2le	Leg II length	Total length of one leg from the second leg pair, excluding coxa and trochanter	real number	mm	
l3le	Leg III length	Total length of one leg from the third leg pair, excluding coxa and trochanter	real number	mm	

l4le	Leg IV length	Total length of one leg from the fourth leg pair, excluding coxa and trochanter	real number	mm	
met1	Metatarsus I length	Metatarsus length of leg I (measured between condyles)	real number	mm	
met2	Metatarsus II length	Metatarsus length of leg II (measured between condyles)	real number	mm	
met3	Metatarsus III length	Metatarsus length of leg III (measured between condyles)	real number	mm	
met4	Metatarsus IV length	Metatarsus length of leg IV (measured between condyles)	real number	mm	
ocdi	Ocular distance	Sum of diameters of one side of the caparace eyes (1 ALE, 1 PLE, 1 PME, 1 AME)	real number	mm	
pat1	Patella I length	Patella length of leg I (measured between condyles)	real number	mm	
pat2	Patella II length	Patella length of leg II (measured between condyles)	real number	mm	
pat3	Patella III length	Patella length of leg III (measured between condyles)	real number	mm	
pat4	Patella IV length	Patella length of leg IV (measured between condyles)	real number	mm	
pled	PLE	Diameter of one posterior median eye	real number	mm	
plsl	Spinneret PLS	Total length of posterior lateral spinneret (from base to tip)	real number	mm	
pmed	PME	Diameter of one posterior lateral eye	real number	mm	
pmsl	Spinneret PMS	Total length of posterior median spinneret (from base to tip)	real number	mm	
ptal	Palpal tarsus length	Length of palpal tarsus in males	real number	mm	
ptwi	Palpal tarsus width	Width of male palpal tarsus	real number	mm	
scoc	Scopula cover	Relative area of the prolateral side of leg I segment(s) covered with hairy adhesive pad (scopula), excluding claw tufts	real number	mm^2	
stle	Sternum length	Width of sternum at widest point	real number	mm	
stwi	Sternum width	Length of sternum along the longitudinal axis	real number	mm	
tale	Tegular apophysis length	Length of tegular apophysis on male bulbus	real number	mm	
tar1	Tarsus I length	Tarsus length of leg I (measured between condyles)	real number	mm	
tar2	Tarsus II length	Tarsus length of leg II (measured between condyles)	real number	mm	
tar3	Tarsus III length	Tarsus length of leg III (measured between condyles)	real number	mm	
tar4	Tarsus IV length	Tarsus length of leg IV (measured between condyles)	real number	mm	
tawi	Tegular apophysis width	Width of tegular apophysis on male bulbus	real number	mm	

tib1	Tibia I length	Tibia length of leg I (measured between condyles)	real number	mm	
tib2	Tibia II length	Tibia length of leg II (measured between condyles)	real number	mm	
tib3	Tibia III length	Tibia length of leg III (measured between condyles)	real number	mm	
tib4	Tibia IV length	Tibia length of leg IV (measured between condyles)	real number	mm	
tro1	Trochanter I length	Trochanter length of leg I (measured between condyles)	real number	mm	
tro2	Trochanter II length	Trochanter length of leg II (measured between condyles)	real number	mm	
tro3	Trochanter III length	Trochanter length of leg III (measured between condyles)	real number	mm	
tro4	Trochanter IV length	Trochanter length of leg IV (measured between condyles)	real number	mm	
	Physiology				
gluc	Glucose content	Amount of glucose in a wet mass	real number	µg/mg	
ld50	Venom toxicity	LD50 of venom on Drosophila prey	real number	nl venom/ mg fly	
pydr	Drought tolerance	Relative humidity the spider can tolerate	real number	%	
pytl	Lower thermal limit	Temperature limit at which growth occurs	real number	°C	
pymr	Resting metabolic rate	Oxygen consumption per time when inactive	real number	W	
pysb	Submerging time	Time of surviving under water	real number	h	
pytu	Upper thermal limit	Temperature limit at which growth occurs	real number	°C	
prot	Protein content	Amount of proteins in a wet mass	real number	µg/mg	
trig	Triglyceride content	Amount of triglycerides in a wet body mass	real number	µg/mg	
	Predation				
cons	Consumption time	Time spent consuming certain prey (specified in Treatment variable)	real number	h	
guil	Hunting guild	Ecological hunting guild according to Cardoso P, Pekár S, Jocqué R & Coddington JA 2011. Global patterns of guild composition and functional diversity of spiders. PloS One 6(6): e21710.	character		Sensing web weavers; sheet web weavers; space web weavers; orb web weavers; specialists; ambush hunters; ground hunters; other hunters.
klep	Kleptoparasitism	Occurrence of kleptoparasitism	character		
nich	Trophic niche breadth	Levin's standardised index of niche breadth according to Hurlbert SH 1978. The measurement of niche overlap and some relatives. Ecology	real number		

		59(1): 67-77.			
para	Paralysis latency	Time between attack and prey immobilisation	real number	min	
prdi	Prey diversity	Shannon-Weaver index of diversity of captured prey as a measure of niche breadth	real number		
prek	Overkilling	Proportion of prey items killed but not consumed	real number	%	
prec	Prey capture	Mode of prey capture	character		Bite-and-release; grab-and-hold; wrapping; throwing silk; other
prem	Satiation	Number of prey items killed and consumed per certain time interval (specified in Treatment variable)	integer		
preo	Prey order	Taxonomic order of an organism the spider preys on	character		
prey	Prey stage	Developmental stage of prey organism	character		Egg; larva/caterpillar; pupa; imago
prsi	Prey size	Prey size (total body length)	real number	mm	
stsp	Strike speed	Time to complete a predatory strike (start of strike to first bite)	real number		
weba	Web area	Size of web projected in a 2-dimensional space	real number	cm^2	
webb	Web building	Use of a web for prey capture (not a retreat)	character		
webd	Web diameter	Linear dimension of a web	real number	cm	
webt	Web type	Type of capture web	character		Orb web; cob web with gum-foot lines; sheet web; canopy web; space web; open tube; tube with trap door; tube with signalling lines; single line; other
webv	Web volume	3-dimensional size of a web	real number	cm^3	
	Reproduction				
coco	Coersive copulation	Presence of coercive mating indicated by causing injuries to the other sex	character		
codi	Cocoon diameter	Maximum diameter of the cocoon	real number		
coty	Courtship type	Sensual modality used during courtship (verbal description)	character		
codu	Courtship duration	Time from starting the courtship to the beginning of copulation	real number	min	

duma	Duration of mating	Total mating time	real number	min	
eggm	Eggsac mass	Weight of an eggsac	real number	g	
eggn	Number of eggs/sac	Number of eggs in a clutch (eggsac) / eggsac order	integer		
egsn	Number of eggsacs	Total number of eggsacs produced by a female during her life	integer		
eplu	Epigyne plugging	Mode of blocking access to the female epigyne	character		Excretion; embolus; none; other
fert	Fertility	Number of hatched offspring	integer		
maca	Maternal care	Extent of maternal care	character		None; guarding egg sac; guarding egg sac and spiderlings
maph	Matriphagy	Presence of matriphagy (i.e., offspring consuming tissue of their mother)	character		
mapo	Mating position	Type of a mating position according to Foelix R F. 2011. Biology of Spiders. 3rd ed. New York: Oxford University Press.	character		Type 1; type 2; type 3
nuin	No. of insertions	Total number of insertions during copulation	integer		
nupa	No. of partners	Total number of mated partners	integer		
ovip	Oviposition	Time to oviposition (following the first mating)	real number	days	
sexc	Sexual cannibalism	Presence of sexual cannibalism and the sex of cannibal	character		Female; male

Table S2. List of Methods. For each method there is an abbreviation, name and a short description.

Abbreviation	Method name	Description
bea	Beating	Capture by beating over net
cit	Citizen Science	Observation collected through citizen science
col	Colorimetry	Concentration assessment of a chemical compound in a homogenate
dis	Dissection	Obtained using dissection
exp	Expert-Base	Assessed based on expert opinion
fie	Field Observation	Observation performed in nature
fot	Photoeclector trapping	Capture by photoeclector
fun	Functional Response	Experiment of functional response
gut	Gut-Content Analysis	Molecular analysis of gut content
han	Hand Collection	Capture by individual hand sampling
kar	Karyology	Karyology on dissected tissue
lab	Laboratory Observation	Observation performed under laboratory conditions
mal	Malaise Trapping	Capture by Malaise traps
mic	Microscopic Measurement	Measurement done under microscope or in micro-photographs
mor	Morphometry	Length determination based on microscopy
mov	Movement Measurement	Measurements done using video-tracking software (e.g., Ethovision)
mul	Multiple data analysis	Analysis of results of former multiple studies
na	Not available	This information is not available
olf	Olactometry	Measurement done using olfactometer
pan	Yellow Pan Trapping	Capture by yellow pan traps
pho	Photographic Analysis	Analysis of photographs
pro	Protein content	Measurement of protein content using Bradford's method
ptf	Pitfall Trapping	Capture by pitfall traps
res	Respirometry	Measurement done using respirometer
she	Shelter Trapping	Capture by shelters (e.g. bark bands)
sie	Sieving	Capture by sieving
sou	Sound Recording	Sound recorded by a recorder
spe	Spectrophotometric Measurement	Measurement done using spectrophotometer
suc	Suction trapping	Capture by a suction trap placed in the air
swe	Sweeping	Capture by sweeping net

tem	Transmission electron microscopy	Transmission electron microscopy using standard protocol for chemically fixed samples
the	Thermometry	Measurement done using temperature controlled chamber
tox	Toxicology	Toxicology bioassays
vac	G-VAC sampling	Capture by sucking up device.
ven	Venom potency test	Test of venom potency using a standardized protocol (specified in trait or notes)
web	Web Analysis	Analysis of the web content
wei	Weighing	Weighing on a lab scale (i.e. analytical balance)