



OPEN

DATA DESCRIPTOR

Functional trait dataset of European groundwater Amphipoda: Niphargidae and Typhlogammaridae

Ester Premate & Cene Fišer

Groundwater represents a vast, but mostly hidden and inaccessible ecosystem. Although often overlooked in freshwater research, groundwater organisms form a significant part of freshwater biodiversity, whereas their functions are crucial in different ecosystem processes. Knowledge on functional traits is generally lacking for most groundwater species worldwide, yet European groundwater amphipods, particularly the family Niphargidae, are an exception. They are well-researched and used as a model system in ecological and evolutionary studies. We focused on this group to assemble a first functional trait dataset dedicated to groundwater species. We gathered data for eight morphological functional traits quantified through 27 measurements for 1123 individuals which represent 180 species and 314 MOTUs. Besides functional trait data, every entry is accompanied with locality information, including habitat type, and DNA sequences if available. The structure of the dataset and data processing information provided along enable wide applicability and extension to other amphipod taxa. When coupled with phylogeny, the dataset may further enhance different aspects of groundwater research, including biodiversity patterns, community assembly processes, and trait evolution.

Background & Summary

Functional traits have become an important tool in basic and applied ecological research. Among other, they provide an insight into functional roles of organisms in their environments, diversity of functions and ecological processes within ecosystems, and community response to environmental change^{1–3}. In the era of the broad usage of functional traits in ecology, the constant production of trait data across different taxa and ecosystems prompted the organization of these data in publicly available datasets and databases. In this respect, freshwater organisms and ecosystems received a considerable amount of attention worldwide^{4–7}. Yet, freshwater's hidden realm, the groundwater, has been overlooked, and functional traits of groundwater invertebrates largely unexplored compared to their epigeal counterparts^{8–10}.

Groundwater represents a vast, globally widespread ecosystem which stores by far the largest portion of available freshwater^{11–14}. Groundwater invertebrates significantly contribute to the overall biodiversity of freshwater ecosystems¹⁵ and play vital roles in ecosystem processes such as nutrient cycling and bioturbation. Groundwater provides a number of ecosystem services, from supporting terrestrial and epigeal freshwater ecosystems to the provision of drinking water¹², all of which depend on the functional traits of groundwater biota. Exploration of functional traits and gathering trait data of groundwater invertebrates is thus critical for the advancement in our understanding of groundwater ecological processes, ecosystem services, and responses to environmental change^{8,9,16,17}.

Functional trait data for few groundwater invertebrate species are included in publicly available databases^{6,7}, but are limited to only few taxa and often incomplete. Additionally, functional trait entries of groundwater organisms are usually very general and wide to fit in the databases focusing on freshwater invertebrates, meaning that they might not carry sufficient information to study processes within the groundwater itself (e.g. the value for certain functional trait would be the same across all groundwater taxa). Recently, Hose *et al.*⁸ provided a general overview of expected functional trait states in groundwater invertebrates, but showed that available knowledge is scarce. Reasons for this knowledge gap primarily lie in habitat inaccessibility, but also in difficulties associated with rearing and studying groundwater invertebrates in laboratory conditions¹⁸. Both hamper

University of Ljubljana, Biotechnical Faculty, Department of Biology, SubBioLab, Jamnikarjeva 101, SI-1000, Ljubljana, Slovenia. ✉e-mail: ester.premate@bf.uni-lj.si

in-situ observations and experimental testing of traits' functionality. Nevertheless, a long history of research of groundwater amphipods paved the way to the organization of the first functional trait dataset dedicated to subterranean species.

Amphipods are an important macroinvertebrate group in groundwater, both in terms of abundances and species richness¹⁹. They inhabit all continents except Antarctica, and they even survived the Pleistocene glaciation in Iceland thermal waters²⁰. In the Western Palearctic, amphipods comprise about one third of all subterranean crustaceans²¹. The most common and studied genus among them is *Niphargus*, which is found across different groundwater habitats. It is the largest freshwater amphipod genus^{19,22} and an important model system in ecology and evolution^{23–25}. Previous laboratory and comparative studies have identified several functional traits related to habitat selection, locomotion, feeding, reproduction, and defense against predators^{24–29}. Studies that explicitly tested functionality of different traits provided baseline knowledge of functional traits for groundwater amphipods. We compiled functional trait data available from already published and unpublished datasets and collected new data for missing and still undescribed species.

With the Functional trait dataset of European groundwater Amphipoda: Niphargidae and Typhlogammaridae, we aim to facilitate the use of functional trait data in groundwater research. Crustaceans are globally well-represented in groundwater communities, dominating in terms of species richness³⁰. As such, they are suitable model systems for tackling various ecological and evolutionary questions through the usage of functional traits, especially when combined with phylogenetic, ecological, and biogeographical data. The present dataset combines all: individual-level trait data, ecology, geographical location, and cytochrome oxidase I (COI) gene sequences, and can thus be readily used to assess e.g. local and regional biodiversity patterns and community assembly processes, as well as to study the evolution of functional traits. At the same time, it is the first functional trait dataset of such extent and completeness for both subterranean fauna and amphipods. It serves as a basis for future publicly available online database, which will gather functional trait data for subterranean amphipods and offer the possibility to update and contribute data, as well as to extend it to other freshwater and marine amphipod taxa.

Methods

Selection of functional traits. Functional traits are any traits influencing organismal performance^{3,31}, and can be morphological, behavioral, physiological, or life-history. Most of the trait types, except morphological, are difficult to measure in groundwater invertebrates, as their collection, rearing, or observation in their natural habitat is at least limited, if not impossible. They are also less abundant than invertebrates in other freshwater ecosystems and reaching adequate sample size is often hindered¹⁸. We focused on morphological traits that have been linked to a specific function in prior studies. We provide a review of these traits and their corresponding functions below and in Table 1.

We acquired 27 morphological measurements collectively representing eight morphological functional traits: body size, body shape, locomotory apparatus, antennae, gnathopods, defense spines, uropods, and number and size of eggs. *Body size* is a fundamental trait related to species' biology. It affects the organisms' metabolic rate, ability to disperse, their locomotion, and microhabitat selection^{24,32}. Similarly to body size, *body shape* relates to locomotion and microhabitat selection^{24,27}, but also determines the shape of the ventral channel, which controls the velocity of water flow bringing oxygen to the gills and creating propulsion for movement³³ (Fig. 1b). In females, the shape of the ventral channel is also connected to the space available for marsupium²⁶.

We further measured morphological traits for *locomotory*, *sensory*, *feeding*, and *defensive* functions. Pereopod length (trunk appendage, see Fig. 1) relates to movement speed in *Niphargus*. Species with longer pereopods move faster than species with shorter pereopods²⁷. Antennae play an important role in amphipod chemoreception³³. To account for sensory function, we measured the length of both pairs of antennae. Further, the first two pairs of amphipod pereopods, the gnathopods, are modified for grabbing food particles, guarding females, and grooming^{34–36}. In *Niphargus*, their size and shape relates to feeding habits^{25,28}. Species with larger gnathopods tend to occupy higher trophic positions and feed as predators, whereas species with smaller gnathopods occupy lower trophic positions and feed as detritivores²⁸. Gnathopod shape determines how broadly the last two articles, propodus and dactylus, open, and defines the maximum size of the food particle that the animal can grab³⁷ (Fig. 1c). The reach depends on so-called palmar angle of propodus. Propodi with smaller angles enable grabbing relatively larger food particles and are generally more triangle-shaped, whereas propodi with larger angles enable grabbing of only smaller food particles and are square-shaped. Finally, amphipods may express anti-predatory traits^{38,39}. In *Niphargus*, pleonal spines play a defensive role and are present in several species²⁹ (Fig. 1d).

Besides abovementioned functional traits, we also included data specific for some species and individuals related to sexual selection and fecundity, respectively. Several species of *Niphargus* exhibit sexual dimorphism. Previous studies have identified sexually dimorphic traits, including body size and relative lengths of abdominal appendages, the uropods⁴⁰. In species exhibiting sexual dimorphism, males are larger than females and have longer uropods⁴⁰ (Fig. 1e), which are presumably also involved in locomotion. Lastly, female body size and shape relates to brood size, with larger amphipods having more eggs^{26,41}. We counted number of eggs and measured their diameter (whole clutch except when some of the eggs were damaged) in females that carried eggs at the time of their collection.

Samples acquisition and ecological characterization. Most of the samples were collected during field work of researchers from the Department of Biology, Biotechnical Faculty, University of Ljubljana, between 1950s and 2022 (Fig. 2). Samples were collected from different aquatic habitats, such as springs, caves, brooks and river interstitial, and artificial habitats, such as wells and tunnels, using adequate sampling techniques according to habitat type. Animals were mostly collected using different water nets, but also with the usage of other methods, as e.g.

	Functional trait	Function - general	Function - detailed	Measurements	References
A	Body size	Complex	Affects metabolic rate, dispersal ability, locomotion, microhabitat selection. Sexually dimorphic trait in some species.	Body length	24,32
A, B	Body shape including ventral channel shape	Complex	Overall body shape affects locomotion and microhabitat selection. The depth of the ventral channel determines the velocity of water flow which brings oxygen to the gills and creates propulsion when moving. It also determines the maximum size of the marsupium.	Coxa II depth Coxa III depth Pereopod V basis width Pereopod VI basis width Pereopod VII basis width	24,26,33
A	Locomotory apparatus	Locomotion	Length of the appendages involved in locomotion relates to movement speed.	Pereopod V length Pereopod VI length Pereopod VII length	27
A	Antennae	Sensory	Antennae carry numerous chemoreceptors and are involved in chemoreception.	Antenna I length Antenna II length	33
C	Gnathopods	Feeding	Gnathopods are used for grabbing food particles and grooming. Their size and shape relates to feeding habits and trophic position.	Gnathopod I carpus length Gnathopod II carpus length Gnathopod I propodus length Gnathopod I propodus width Gnathopod I propodus diagonal Gnathopod II propodus length Gnathopod II propodus width Gnathopod II propodus diagonal	28,34,35
D	Spines	Defensive	Defensive, anti-predatory trait in some species.	Presence/absence of spines	29,38,39
E	Uropods	Presumably involved in locomotion	Sexually dimorphic trait in some species.	Uropod I basis length Uropod I endopodite length Uropod I exopodite length Uropod III exopodite proximal article length Uropod III exopodite distal article length	37,40
	Number and size of eggs	Reproductive	Egg number per single brood affects overall fecundity of a female. Egg size corresponds to yolk	Egg number Egg diameter	26

Table 1. Overview of functional traits included in Functional trait dataset of European Groundwater Amphipoda: Niphargidae and Typhlogammaridae. Letters in the first column correspond to Fig. 1 panels where the measurements and functional traits are presented.

Karaman-Chapuis and Bou-Rouch sondes⁴² or water traps. The sampling has been most intensive in Slovenia and Dinarides, a karstic massif in the SE Europe (Fig. 3), for two reasons. First, Dinaric subterranean habitats have traditionally been at focus of many Slovenian and foreign researchers due to extremely high diversity of subterranean species^{43,44}. Second, *Niphargus* diversity peaks in the Dinarides^{45,46}, and sampling campaigns were often dedicated to collection of *Niphargus* for different taxonomic, ecological, and evolutionary studies.

Many ecological studies benefit from inclusion of species' preferred habitat. Groundwater habitats are heterogeneous and differ in environmental parameters, such as isolation from surface, connectivity, pore size, water flow velocity, and chemistry, all of which affect different aspects of species' ecology^{45,47}. A habitat preference can be assigned to a species from multiple records and analysis where species most frequently occurs. We included locality type where an individual was collected whenever possible to allow inference of species habitat and ecological preferences based on occurrence frequencies in different locality types. We acknowledge that epigeal freshwater and groundwater habitats may be classified in a variety of ways using different parameters as a key criteria. In this dataset, we followed general locality types as in SubBioDB, an internal database on subterranean fauna⁴⁸ (e.g. cave, spring, river interstitial), which do not include microhabitat details (e.g. cave puddles, cave streams) and are not restricted to only groundwater habitats. This way, we avoided potential drawbacks of microhabitats due to different levels of accuracy, and at the same time enabled inclusion of other amphipod taxa from other freshwater habitats.

Specimen selection, preparation, and measurement. The dataset consists of two sets of morphological measurements. The first set gathers data obtained from older samples and existing microscopic slides. Some of these data were already published, but scattered across many publications. We compiled and unified these individual published and internal datasets for the whole distribution range of Niphargidae. The second set represents morphological data acquired in recent years to cover as many species and undescribed species (discovered through the usage of molecular methods, hereafter referred to as MOTUs: Molecular Operational Taxonomic Units) as possible following the latest available phylogeny⁴⁹. In this second set, we focused on the most well-sampled area, Slovenia and the Dinarides, due to the availability of the material. In the second set, we

a *Niphargus* functional model

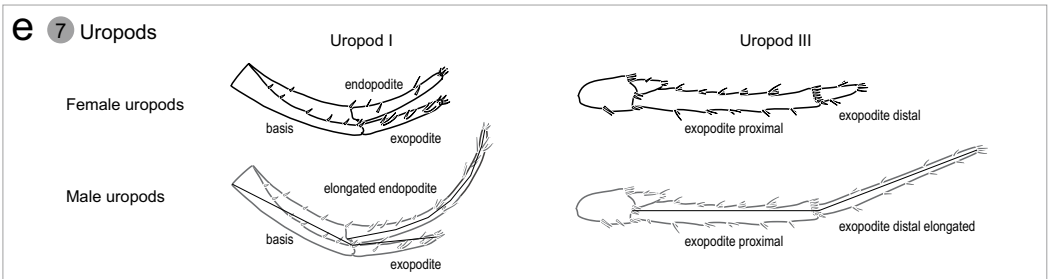
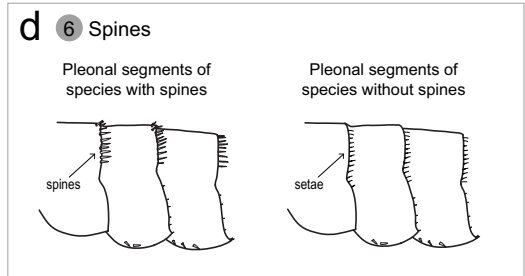
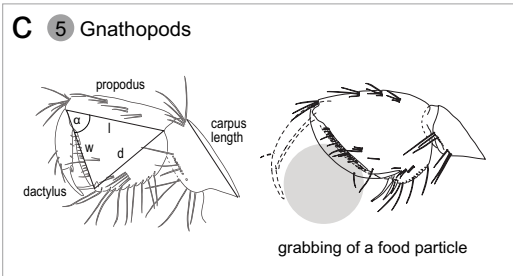
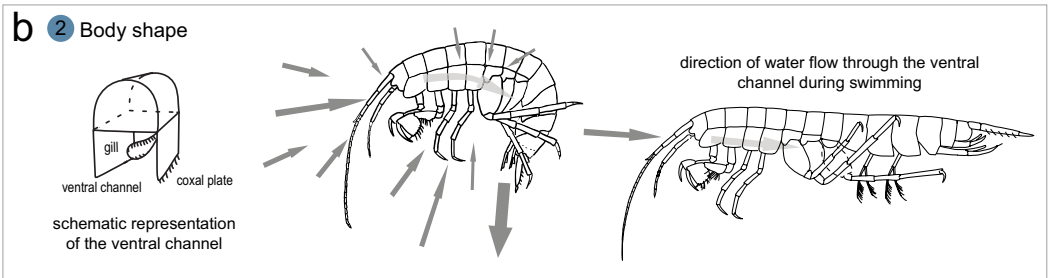
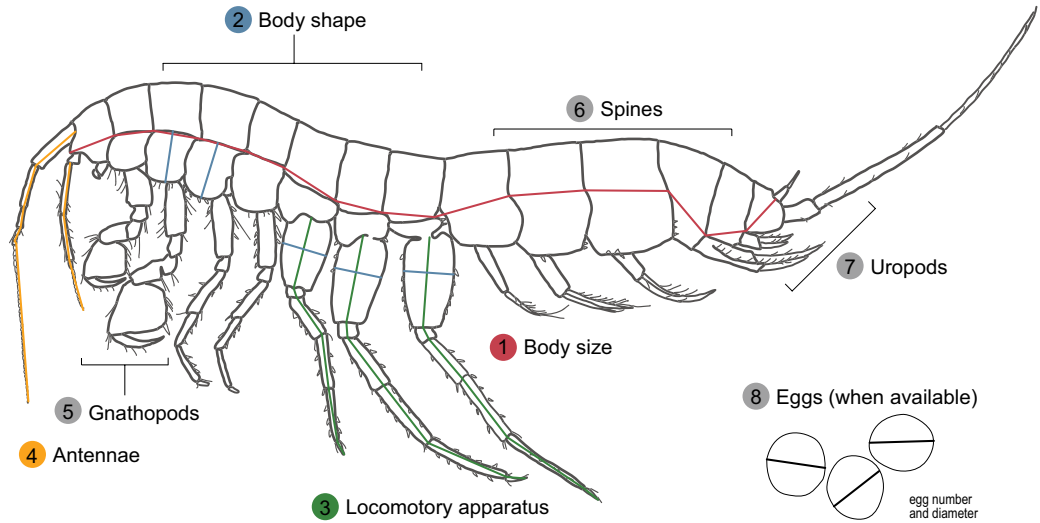


Fig. 1 *Niphargus* functional model. An example of *Niphargus* specimen (a) and measured morphological functional traits, excluding egg diameter. Lines represent measurements of traits according to the landmarks³⁷. Panels b–e detail four functional traits. Measurements of gnathopods and uropods are shown on panels c and e, respectively.

selected the individuals with best-preserved morphology, and preferentially used those that already had available DNA sequences or additionally sequenced them in the case of uncertain morphological diagnosis.

In the first set, the collected individuals were mounted on microscopic slides. They were treated in a hot solution of 10% KOH, rinsed with HCl and distilled water, and stained using different pigments. After treatment, they were dissected and put on the slides in a gelatine glycerol medium (Merck) and measured under stereomicroscope. Contrary to the first set of measurements, animals of the second set were not mounted on permanent

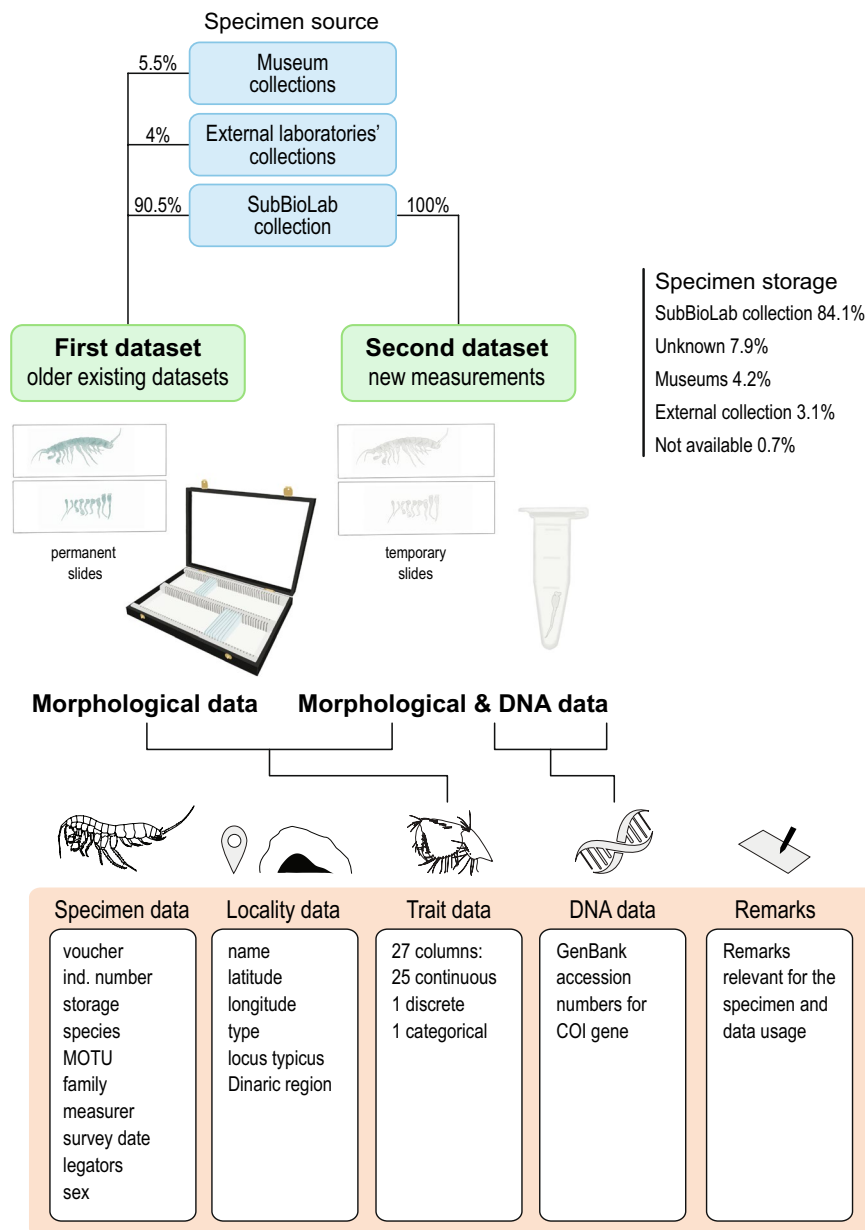


Fig. 2 Overview of data collection and dataset structure for both families, Niphargidae and Typhlogammaridae. Most of the specimen and locality data originates from SubBioDB⁴⁸.

microscopic slides, but rather prepared temporarily to preserve the animal for other analyses and DNA extraction. These animals were stored in collection in 96% ethanol at -20°C . Prior measurements, we transferred them to glycerol for 1–2 days and dissected them. We put them on the slides without any additional treatment by using glycerol as a medium, measured them under stereomicroscope, and stored them back in 96% ethanol at -20°C . The reasons behind such preservation of the animals are their rarity and effort associated with obtaining additional samples for other analyses. We assessed potential differences and errors due to different measurement techniques, which turned out as negligible.

When measuring morphological traits, we followed the standard landmarks for morphological measurements of *Niphargus*³⁷ (Fig. 1). We used two stereomicroscopes with mounted digital cameras and corresponding software: Olympus SZX9 coupled with ColorViewIII camera and cellSens Entry software, and Leica M165C coupled with Leica Flexacam C1 camera and LASX software. The magnification used depended on the specimen size. For specimens measured within the second set, images are stored and measurements repeatable directly on the images. They can also be exported in optional formats and resolutions of up to 4000×3000 pixels.

The identifier issue resolved: individual versus MOTU versus species. The dataset is constructed at an individual level that allows wide reusability of the data and analytic flexibility. It also establishes a stable link between an individual and its measurable attributes, including morphometric data, distribution, DNA sequence,

and ecology⁵⁰. The dataset allows easy addition of new data, making it volatile and reusable in different analytic frameworks. With addition of new samples, morphometric data can be reused, and, for example, estimations of the mean trait value and its variation can be refined. Similarly, addition of new individual-level ecological data may provide additional insights into species ecology. Such refinements would not be possible if data were entered as mean values of the populations or species. Hence, the addition of records will make the dataset reusable for analyses at three hierarchical levels: among-individuals, between-populations, and between-species.

The main reason for individual-level organization of the dataset is the inherent nature of taxonomy, i.e. species hypothesis^{51,52}. Addition of new data in the dataset may yield new species hypotheses, i.e., an individual can be assigned to another species after revision. If addition of new records yields a modified taxonomic structure of the dataset, e.g., species A is split into species A and B, individuals with barcodes are reassigned to new species and biological attributes automatically follow the revised taxonomic structure without needing a revision⁵¹. This flexibility does not hamper analyses relying on population or species level, as we provide customized scripts for estimation of mean values, or population-level traits, such as sexual dimorphism, as discussed under the Usage Notes.

Individuals in our dataset are labelled using voucher system from the SubBioDB⁴⁸. In separate columns, we assigned MOTU and/or species names to each individual (except rare cases, see Data Records). To keep the link with previous publications, we followed already established labelling system of MOTU names^{46,49}. This three label-system might seem redundant, however, we find it useful to comply with requirements of contemporary ecological and evolutionary research that relies on MOTUs, biological conservation that depends on Linnean taxonomy, and technical management of the dataset, where individuals as biological attribute bearers can be reassigned to revised taxonomic structure at minimal effort and minimal chance for introduction of errors or data loss^{50,51,53}.

Data Records

The dataset is available for download from figshare⁵⁴ as a Microsoft Excel (.xlsx) file. The first sheet, named Data, contains all raw data, and the second sheet, named Description, contains descriptions of columns and values similarly as in Supplementary Table 1.

The dataset includes data related to an individual and data related to a locality where it was collected (Fig. 2). The first 16 columns summarize metadata about the measured individual and locality. They include voucher and/or individual number, species and MOTU identities, family, storage, measurer, survey date, legators, and sex, and columns providing data on the locality include name, geographical coordinates, and locality type. Additional two columns carry information on whether the locality is a type locality of a species and whether it lies within Dinaric region or not. With the extension of the dataset, especially by addition of records outside the Dinaric region, the last column can be transformed into “biogeographical region”.

The morphological data are contained in 27 columns and include 25 columns with morphometric (continuous) functional trait data and two columns with discrete or categorical morphological functional trait data. Lastly, we included information on COI sequences and remarks important for the usage of the dataset. An overview of the column contents is given in Supplementary Table 1.

The dataset includes measures of 1123 individuals. They belong to two families (1092 to Niphargidae and 31 to Typhlogammaridae), 180 species and 314 MOTUs. We identified species based either on morphological data or DNA sequence, while we assigned MOTUs to individuals where there was at least one sequence from a locality where individuals were sampled. For an additional 51 individuals, we assigned MOTUs even less stringently, when the measured individuals were from type locality or when morphology was studied in detail, although not connected with a DNA sequence. We entered this information accordingly in the “remarks” column. In the case of undescribed species, we assigned “sp.” as a species name, and entered MOTU information in the MOTU column. Only five individuals remain without both species and MOTU identities. This less stringent assignment of MOTUs increases sample sizes per MOTU in downstream analyses and secures reusability of the available data.

When obtaining data in the second set of measurements (see previous section and Fig. 2), we aimed to measure at least two adult specimens per MOTU when enough material was available. This approach captured at least some intra-MOTU variability, but also enabled more accurate estimates of missing data (e.g. when one specimen is missing one trait and the second one another trait, both can be imputed by using data from the closest specimen in the phylogeny; see Usage notes). An overview of number of specimens per MOTU and number of MOTUs is given in Fig. 4.

The specimens included in the dataset were collected at 389 different localities spread across the whole range of both families²³, but with most of the localities clustered in Slovenia and in the Dinaric region (Fig. 3a,b). Most of the localities are by their type caves (49.4%) followed by springs (20.6%), and wells (8.5%). 11.8% and 9.8% of the localities are of other or unknown types, respectively (locality types of Niphargidae shown on Fig. 3e).

Technical Validation

We constructed as detailed and as complete functional trait dataset for subterranean amphipods as possible in terms of reliable species and MOTU identification, measurement accuracy and completeness, spatial coverage, and phylogenetic coverage. The link between the morphological trait and its function was tested in eight independent studies (see Table 1 for references).

Phenotypic identity of the specimens was assessed by experts (CF). When species identification was uncertain or in the case of cryptic species, we also examined genetic identity by sequencing the COI or multiple genes. Nine-hundred and five (80.6%) specimens have MOTU identification. A large part of the dataset thus provides the most detailed phenotypic and genotypic identity of the specimens.

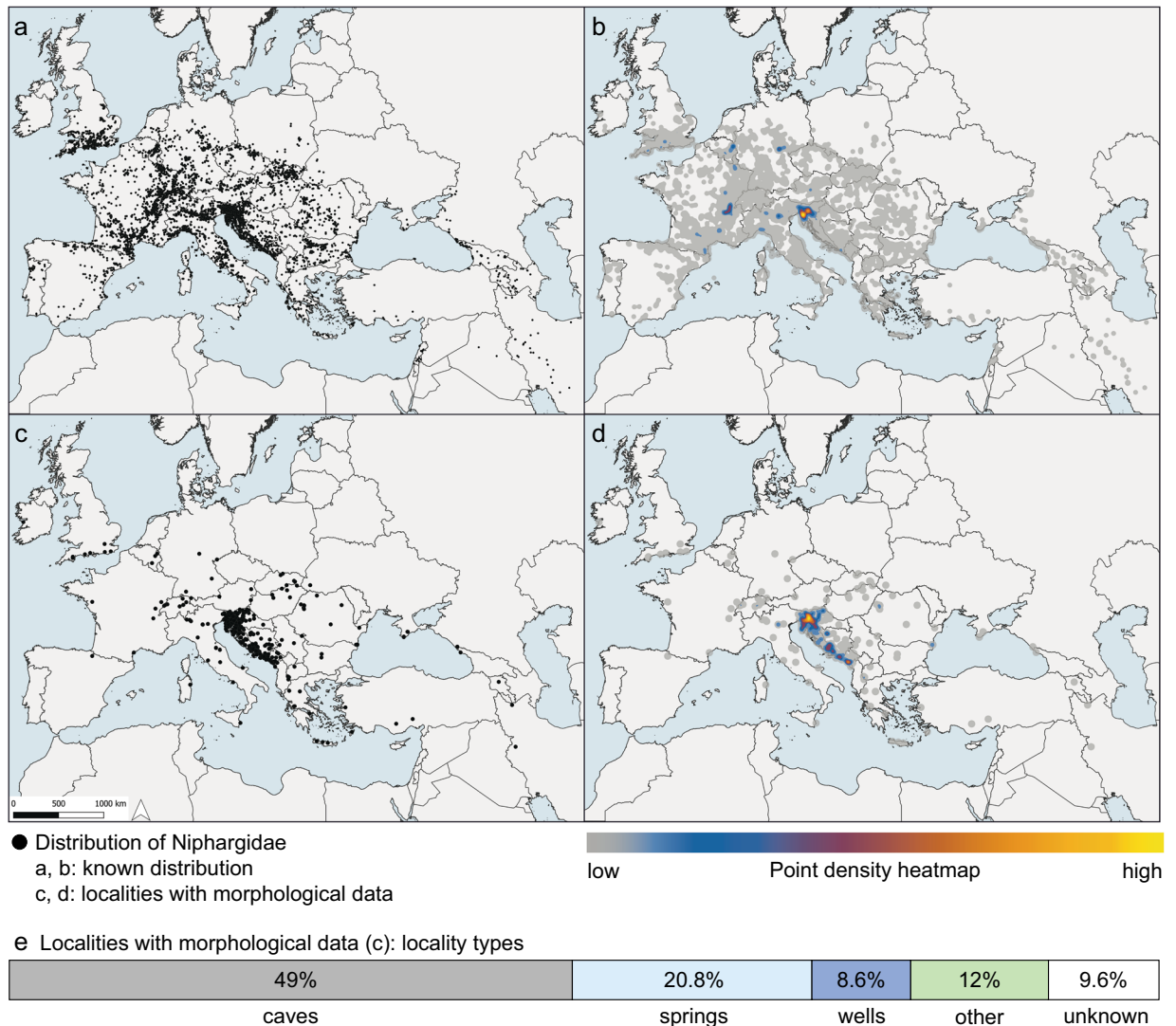


Fig. 3 Distribution of Niphargidae. (a and b) Currently known distribution of Niphargidae^{21,46,48} and localities with morphological data (c,d). Panels a and c show individual data points, and panels b and d the same data as a heatmap. The highest density of Niphargidae localities is in Slovenia (panel b), whereas density of localities with morphological data extends from Slovenia across Dinarides (panel d). Panel e summarizes types of the localities with morphological data. Type “other” includes different freshwater habitats (interstitial, brook, river, lake, puddle, wetland) and artificial habitats (tunnel, ditch, borehole).

Morphometric data were obtained mostly by two researchers (CF, first dataset, and EP, second dataset) with minor contributions of others, hence minimizing the error related to individual measuring differences. As body length influences all other continuous and discrete traits and possible downstream analyses which would account for individual’s body length, we measured it three times and included the mean value of the three measurements to the dataset.

Among Niphargidae, the amount of missing data does not exceed 14% per continuous trait measurable in both sexes, except for uropod measurements (41–50%). Most of other continuous traits have less than 5% of missing data (13 traits out of 19), while 6 have 5–14%. We provide an example workflow to deal with missing data in Usage Notes. The related R code is available from figshare⁵⁴.

The dataset covers well both currently known spatial distribution of Niphargidae throughout the Western Palearctic and the latest phylogeny. The largest share of the data is available for the Dinaric region and Slovenia (Fig. 3c,d). Phylogeny includes 561 MOTUs, and 300 (53,5%) out of those also have morphological data. When pruned to the Dinaric region (215 MOTUs), which represented the focal area in obtaining the second set of measurements, the share of MOTUs with morphological data increases to 86% (Fig. 4a).

To validate the possibility of wider usage of selected traits across different groundwater amphipod taxa, we additionally measured individuals belonging to another groundwater amphipod family present in Europe and distributed in the Dinarides, Typhlogammaridae. We measured specimens belonging to three genera: *Metohia*, *Accubogammarus*, and *Typhlogammarus*. We proved that measurement of the functional traits included in our dataset is applicable beyond Niphargidae, which opens a possibility for future broadening of the dataset.

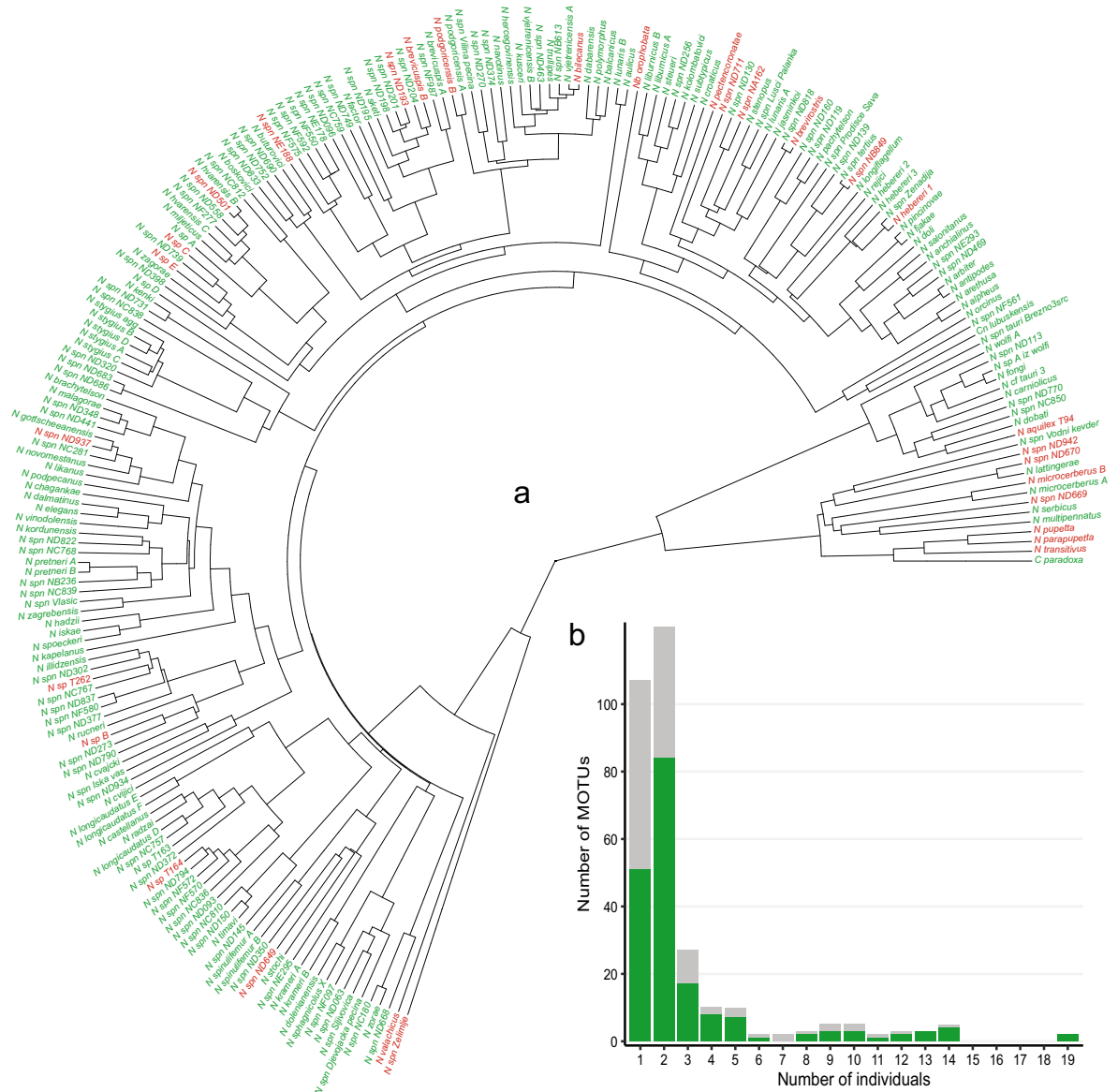


Fig. 4 Dinaric Niphargidae phylogeny and overview of number of individuals measured per MOTU. **(a)** phylogenetic tree^{49,54} pruned to MOTUs distributed in the Dinaric region. Tips colored according to data availability (green: available morphological data, red: missing morphological data). In the most species-rich and sampled region, the share of MOTUs with missing data is low. **(b)** Numbers of individuals measured per MOTU colored by distribution (grey: whole dataset, green: limited to MOTUs distributed in the Dinaric region). Most MOTUs are represented with at least two specimens.

Usage Notes

Usage Notes refer to the usage of data for the family Niphargidae and were so far not applied in Typhlogammaridae. For all sections, we provide R code to repeat the calculations and imputation of missing data. It is available for download from figshare⁵⁴. The R script was written in R version 4.2.2⁵⁵ using R Studio version 2022.12.0.353⁵⁶ and packages phytools 1.2.0⁵⁷, ape 5.6-2⁵⁸, PVR 0.3⁵⁹ and missForest 1.5⁶⁰.

Quantification of gnathopod size and shape. Gnathopod size and shape relate to *Niphargus*' feeding habits and can be quantified using three raw measurements of the sixth gnathopod article, propodus^{25,28}. Gnathopod size is quantified as gnathopod perimeter, a sum of propodus length (pl), width (pw), and diagonal (pd) (marked l, w, d in Fig. 1c; Eq. (1)). Gnathopod shape is defined by the angle α between the propodus length and width (Fig. 1c). This angle can be retrieved using the same three raw measurements, propodus length, width, and diagonal, and cosine theorem (Eq. (2)).

$$\text{Gnathopod perimeter} = pl + pw + pd \quad (1)$$

$$\cos\alpha = \frac{pw^2 + pl^2 - pd^2}{2 \cdot pl \cdot pw} \quad (2)$$

Sexually dimorphic traits. The dataset is structured on an individual level, and as such does not provide population-, MOTU-, or species-level traits, as e.g. the presence of sexual dimorphism, which occurs in several Niphargidae species⁴⁰. However, this information can be obtained for 75 species included in the dataset using measurements of body length, uropod I endopodite and exopodite, and uropod III proximal and distal articles of exopodite (Fig. 1e). Males of sexually dimorphic species are larger than females and have relatively longer uropod I endopodite and exopodite of uropod III. To assess differences between males and females, the length of uropod I endopodite can be compared against the length of uropod I exopodite, while the length of uropod III exopodite distal article can be compared against the length of uropod III exopodite proximal article. The ratios of uropod I endopodite: uropod I exopodite and uropod III exopodite distal article: uropod III exopodite proximal article are larger in males.

Imputation of missing data. Trait datasets often contain missing values. To some extent, this can be avoided during data collection, but not entirely. Missing values can cause issues in downstream analyses and can be handled in many ways. The simplest and most rigid approach is to remove entries with missing data, but this may heavily affect the size and strength of the dataset. On the other hand, missing values can be replaced with imputed values through several different computational approaches which differ mostly in computational background and ability to incorporate phylogenetic information^{61,62}. As our dataset contains missing values, we here provide an example workflow for their imputation.

We imputed missing values similarly as in Penone *et al.*⁶¹ and Debastiani *et al.*⁶³ using missForest package⁶⁰ which uses Random Forest algorithms. We used both trait data and phylogenetic information to impute missing values, as the latter usually improves estimations⁶¹. We imputed missing values in continuous traits, excluding uropod and egg traits which were not measured in many individuals, keeping the traits on a level of single individual. To incorporate phylogenetic information, we followed the approach of Penone *et al.*⁶¹ and Debastiani *et al.*⁶³ by first transforming phylogenetic distance matrix into orthogonal vectors. We selected the number of eigenvectors (first 40) and added them to the trait dataset.

Imputed missing values should be handled carefully and checked for their accuracy⁶². We evaluated the accuracy of estimated values in two ways. First, we checked the out-of-box imputation errors (NRMSE, normalized root mean squared error) returned by the Random Forest algorithm for each trait separately. Second, we calculated the ratios between trait value and body length for observed and imputed data for each MOTU and summarized them in a table where they can be easily compared and checked for any outliers or errors.

Code availability

The R code used for summaries of the Data Records and for creating the plots in Fig. 4, as well as the R code for the analyses described in the Usage Notes are available for download from figshare⁵⁴. We did not use any other custom code during data collection.

Received: 19 September 2023; Accepted: 29 January 2024;

Published online: 10 February 2024

References

- Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail. *Funct. Ecol.* **16**, 545–556 (2002).
- Mouillot, D., Graham, N. A. J., Villéger, S., Mason, N. W. H. & Bellwood, D. R. A functional approach reveals community responses to disturbances. *Trends Ecol. Evol.* **28**, 167–177 (2013).
- McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21**, 178–185 (2006).
- Schmera, D., Heino, J., Podani, J., Erős, T. & Dolédec, S. Functional diversity: a review of methodology and current knowledge in freshwater macroinvertebrate research. *Hydrobiologia* **787**, 27–44 (2017).
- Vieira, B. N. K. M. *et al.* A Database of Lotic Invertebrate Traits for North America. *Director* 19 (2006).
- Fišer, C., Brancelj, A., Yoshizawa, M., Mammola, S. & Fišer, Ž. Dissolving morphological and behavioral traits of groundwater animals into a functional phenotype. in *Groundwater Ecology and Evolution* (eds. Malard, F., Griebler, C. & Retaux, S.) 415–438 (Academic Press, 2023).
- Sarremejane, R. *et al.* DISPERSSE, a trait database to assess the dispersal potential of European aquatic macroinvertebrates. *Sci. Data* **7**, 1–9 (2020).
- Hose, G. C. *et al.* Invertebrate traits, diversity and the vulnerability of groundwater ecosystems. *Funct. Ecol.* 0–3, <https://doi.org/10.1111/1365-2435.14125> (2022).
- Gibert, J. *et al.* Assessing and conserving groundwater biodiversity: Synthesis and perspectives. *Freshw. Biol.* **54**, 930–941 (2009).
- Fišer, C., Brancelj, A., Yoshizawa, M., Mammola, S. & Fišer, Ž. Dissolving morphological and behavioral traits of groundwater animals into a functional phenotype. in *Groundwater Ecology and Evolution* (eds. Malard, F., Griebler, C. & Retaux, S.) 415–438 (Academic Press, 2023).
- Danielopol, D. L., Griebler, C., Gunatilaka, A. & Notenboom, J. Present state and future prospects for groundwater ecosystems. *Environ. Conserv.* **30**, 104–130 (2003).
- Griebler, C. & Avramov, M. Groundwater ecosystem services: A review. *Freshw. Sci.* **34**, 355–367 (2015).
- Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. & Cardenas, M. B. The global volume and distribution of modern groundwater. *Nat. Geosci.* **9**, 161–164 (2016).
- Ferguson, G. *et al.* Crustal Groundwater Volumes Greater Than Previously Thought. *Geophys. Res. Lett.* **48**, 1–9 (2021).
- Boulton, A. J., Fenwick, G. D., Hancock, P. J. & Harvey, M. S. Biodiversity, functional roles and ecosystem services of groundwater invertebrates. *Invertebr. Syst.* **22**, 103–116 (2008).

16. Descloux, S., Datry, T. & Usseglio-Polatera, P. Trait-based structure of invertebrates along a gradient of sediment colmation: Benthos versus hyporheos responses. *Sci. Total Environ.* **466–467**, 265–276 (2014).
17. Di Lorenzo, T., Fiasca, B., Di, M., Cifoni, M. & Galassi, D. M. P. Taxonomic and functional trait variation along a gradient of ammonium contamination in the hyporheic zone of a Mediterranean stream. *Ecol. Indic.* **132**, 108268 (2021).
18. Mammola, S., Lunghi, E., Bilandzija, H. & Cardoso, P. Collecting eco-evolutionary data in the dark: Impediments to subterranean research and how to overcome them. *Ecol. Evol.* **228**, 1–16 (2021).
19. Väinölä, R. *et al.* Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia* **595**, 241–255 (2008).
20. Kristjánsson, B. K. & Svavarsson, J. Subglacial refugia in Iceland enabled groundwater amphipods to survive glaciations. *Am. Nat.* **170**, 292–296 (2007).
21. Zagmajster, M. *et al.* Geographic variation in range size and beta diversity of groundwater crustaceans: Insights from habitats with low thermal seasonality. *Glob. Ecol. Biogeogr.* **23**, 1135–1145 (2014).
22. Horton, T. *et al.* *World Amphipoda Database*. <http://www.marinespecies.org/amphipoda> (2022).
23. Borko, Š., Trontelj, P., Seehausen, O., Moškrič, A. & Fišer, C. A subterranean adaptive radiation of amphipods in Europe. *Nat. Commun.* **12**, 1–12 (2021).
24. Trontelj, P., Blejec, A. & Fišer, C. Ecomorphological convergence of cave communities. *Evolution (N. Y.)* **66**, 3852–3865 (2012).
25. Fišer, C., Delić, T., Luštrik, R., Zagmajster, M. & Altermatt, F. Niches within a niche: ecological differentiation of subterranean amphipods across Europe's interstitial waters. *Ecography (Cop.)* **42**, 1212–1223 (2019).
26. Fišer, C., Zagmajster, M. & Zakšek, V. Coevolution of life history traits and morphology in female subterranean amphipods. *Oikos* **122**, 770–778 (2013).
27. Kralj-Fišer, S. *et al.* The interplay between habitat use, morphology and locomotion in subterranean crustaceans of the genus *Niphargus*. *Zoology* **139** (2020).
28. Premate, E. *et al.* Cave amphipods reveal co-variation between morphology and trophic niche in a low-productivity environment. *Freshw. Biol.* **66**, 1876–1888 (2021).
29. Premate, E., Zagmajster, M. & Fišer, C. Inferring predator–prey interaction in the subterranean environment: a case study from Dinaric caves. *Sci. Rep.* **11**, 1–9 (2021).
30. Danielopol, D. L., Pospisil, P. & Rouch, R. Biodiversity in groundwater: A large-scale view. *Trends Ecol. Evol.* **15**, 223–224 (2000).
31. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
32. Peters, R. H. The Ecological Implications of Body Size. *The Ecological Implications of Body Size* <https://doi.org/10.1017/cbo9780511608551> (1983).
33. Dahl, E. The Amphipod Functional Model and Its Bearing upon Systematics and Phylogeny. *Zool. Scr.* **6**, 221–228 (1978).
34. Arndt, C. E., Brandt, A. & Berge, J. Mouthpart-Atlas of Arctic Sympagic Amphipods—Trophic Niche Separation Based on Mouthpart Morphology and Feeding. *Ecology. J. Crustac. Biol.* **25**, 401–412 (2005).
35. Holmquist, J. The functional morphology of gnathopods: Importance in grooming, and variation with regard to habitat, in Talitroidean Amphipods. *J. Crustac. Biol.* **2**, 159–179 (1982).
36. Borowsky, B. The Use of the Males' Gnathopods During Precopulation in Some Gammaridean Amphipods. *Crustaceana* **47**, 245–250 (1984).
37. Fišer, C., Trontelj, P., Luštrik, R. & Sket, B. Toward a unified taxonomy of *Niphargus* (Crustacea: Amphipoda): a review of morphological variability. *Zootaxa* **2061**, 1–22 (2009).
38. Bollache, L. I., Kaldonski, N., Troussard, J. P., Lagrue, C. & Rigaud, T. Spines and behaviour as defences against fish predators in an invasive freshwater amphipod. *Anim. Behav.* **72**, 627–633 (2006).
39. Copilaș-Ciocianu, D., Borza, P. & Petrussek, A. Extensive variation in the morphological anti-predator defense mechanism of *Gammarus roeselii* Gervais, 1835 (Crustacea: Amphipoda). *Freshw. Sci.* **39**, 47–55 (2020).
40. Fišer, C., Bininda-Emonds, O. R. P., Blejec, A. & Sket, B. Can heterochrony help explain the high morphological diversity within the genus *Niphargus* (Crustacea: Amphipoda)? *Org. Divers. Evol.* **8**, 146–162 (2008).
41. Sainte-Marie, B. A review of the reproductive bionomics of aquatic gammaridean amphipods: variation of life history traits with latitude, depth, salinity and superfamily. *Hydrobiologia* **223**, 189–227 (1991).
42. Bou, C. & Rouch, R. Un nouveau champ de recherches sur la faune aquatique souterraine. *Comptes rendus l'Académie des Sci.* **265**, 369–370 (1967).
43. Culver, D. C. & Pipan, T. *The Biology of Caves and Other Subterranean Habitats*. (Oxford University Press, 2009).
44. Culver, D. C. & Sket, B. Hotspots of subterranean biodiversity in caves and wells. *J. Cave Karst Stud.* **62**, 11–17 (2000).
45. Fišer, C., Blejec, A. & Trontelj, P. Niche-based mechanisms operating within extreme habitats: A case study of subterranean amphipod communities. *Biol. Lett.* **8**, 578–581 (2012).
46. Borko, Š., Altermatt, F., Zagmajster, M. & Fišer, C. A hotspot of groundwater amphipod diversity on a crossroad of evolutionary radiations. *Divers. Distrib.* 1–13, <https://doi.org/10.1111/ddi.13500> (2022).
47. Borko, Š., Premate, E., Zagmajster, M. & Fišer, C. Determinants of range sizes pinpoint vulnerability of groundwater species to climate change: A case study on subterranean amphipods from the Dinarides. 1–8, <https://doi.org/10.1002/aqc.3941> (2023).
48. SubBioDB. Subterranean Fauna Database. Research group for speleobiology, Biotechnical faculty, University of Ljubljana. <https://db.subbio.net/> (2023)
49. Delić, T. *et al.* Evolutionary origin of morphologically cryptic species imprints co-occurrence and sympatry patterns. *bioRxiv* <https://doi.org/10.1101/2023.09.13.557531> (2023).
50. Malard, F. *et al.* GOTIT: A laboratory application software for optimizing multi-criteria species-based research. *Methods Ecol. Evol.* **11**, 159–167 (2020).
51. Fišer, C. Collaborative databasing should be encouraged. *Trends Ecol. Evol.* **34**, 184–185 (2019).
52. Thiele, K. & Yeates, D. Tension arises from duality at the heart of taxonomy. *Nature* **419**, 337 (2002).
53. Fišer, C., Robinson, C. T. & Malard, F. Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* **27**, 613–635 (2018).
54. Premate, E. & Fišer, C. Functional trait dataset of European groundwater Amphipoda: Niphargidae and Typhlogammaridae, *figshare*, <https://doi.org/10.6084/m9.figshare.c.6841665.v1> (2024).
55. R Development Core Team. A language and environment for statistical computing. (2023).
56. Team, R. RStudio: Integrated Development Environment for R. (2023).
57. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
58. Didier, G., Heibl, C. & Jones, B. Package 'ape': Analyses of Phylogenetics and Evolution. (2019).
59. Santos, T. PVR: Phylogenetic Eigenvectors Regression and Phylogenetic Signal-Representation Curve. (2018).
60. Stekhoven, D. J. & Bühlmann, P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
61. Penone, C. *et al.* Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods Ecol. Evol.* **5**, 961–970 (2014).
62. Johnson, T. F., Isaac, N. J. B., Paviolo, A. & González-Suárez, M. Handling missing values in trait data. *Glob. Ecol. Biogeogr.* **30**, 51–62 (2021).
63. Debastiani, V. J., Bastazini, V. A. G. & Pillar, V. D. Using phylogenetic information to impute missing functional trait values in ecological databases. *Ecol. Inform.* **63**, (2021).

Acknowledgements

We thank many colleagues who helped us in the field and contributed the samples listed in the dataset. We specially thank Roman Alther, Florian Altermatt, Gergely Balazs, Špela Borko, Teo Delić, Branko Jalžić, Florian Malard, Slavko Polak, Behare Rexhepi, Boris Sket, Peter Trontelj, Rudi Verovnik and Maja Zgmajster. The study was funded by Slovenian Research Agency through core funding (Program P1-0184), project J1-2464, PhD grant to EP and University foundation of eng. Milan Lenarčič scholarship to EP. The research was partly funded by Biodiversa+, the European Biodiversity Partnership under the 2021-2022 BiodivProtect joint call for research proposals, co-funded by the European Commission (GA no. 101052342) and with the funding organizations Ministry of Universities and Research (Italy), Agencia Estatal de Investigación—Fundación Biodiversidad (Spain), Fundo Regional para a Ciência e Tecnologia (Portugal), Suomen Akatemia—Ministry of the Environment (Finland), Belgian Science Policy Office (Belgium), Agence Nationale de la Recherche (France), Deutsche Forschungsgemeinschaft e.V. –BMBF-VDI/ VDE INNOVATION + TECHNIK GMBH (Germany), Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (Switzerland), Fonds zur Förderung der Wissenschaftlichen Forschung (Austria), Ministry of Higher Education, Science and Innovation (Slovenia) and the Executive Agency for Higher Education, Research, Development and Innovation Funding (Romania). The equipment used was purchased for the project “Development of research infrastructure for the international competitiveness of the Slovenian RRI space –RI-SI-LifeWatch”. The operation is co-financed by the Republic of Slovenia, Ministry of Education, Science and Sport and the European Union from the European Regional Development Fund.

Author contributions

Ester Premate designed the study, collected data, prepared R scripts, and prepared the first version of the manuscript. Cene Fišer designed the study, collected data, and contributed to writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03020-w>.

Correspondence and requests for materials should be addressed to E.P.

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

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



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

© The Author(s) 2024