Soil Organisms 98 (2) 2026

pp. 161–172
DOI 10.25674/496

RESEARCH ARTICLE

# Assessing DNA barcode sequences for Collembola species discrimination

**Ulrich Burkhardt[1]*, Susanne Eckert[2], Christian Bluhm[3] & Jürgen Zipperle[2]**

[1] Soil Organism Research, 02826 Görlitz, Germany
[2] Baden-Württemberg State Institute for the Environment (LUBW), 76231 Karlsruhe, Germany
[3] Department of Soil and Environment, Forest Research Institute Baden-Württemberg, 79100 Freiburg, Germany
* Corresponding author, email: soil-organism-research@gmx.net

Received 18 January 2026 | Accepted 5 May 2026 | Published online 1 August 2026

## Abstract

When dealing with small and highly abundant soil organisms such as Collembola, studies on e.g., population genetics or monitoring of soil conditions can significantly be accelerated by molecular-based high-throughput DNA barcoding methods. However, a prerequisite for the success of such approaches is a sufficient completeness of the reference databases. We used Collembola as an example to examine the current status of completeness of the two most important public repositories for DNA barcode sequences, the Barcode of Life DataSystems (BOLD) and GenBank, and the current accuracy of Collembola sequence identification. As part of the surveys for the Biological Monitoring Network of the federal state of Baden-Württemberg, we analysed how many of the Collembola species found in Baden-Württemberg so far have DNA barcodes. Additional sequence data were generated from currently obtained Collembola samples in order to close identified gaps in the barcode reference libraries. We then used these new sequences to determine how accurately they are assigned to a respective taxon by the identification systems of the reference databases, from species level to higher taxonomic ranks, or whether there are discrepancies. The results showed that the probability of correct identification in the reference databases is related to the number of individuals stored and identified to species level, but in most cases the accuracy of matches seemed to be less related to the sheer number of sequence data of the respective taxon already deposited but rather to the accuracy of morphological determination of the respective top matches. We give recommendations for users of such databases to avoid ambiguous entries when uploading their own sequences and to improve the quality of their sequence data through curation and annotation.

**Keywords** Springtails | COI DNA barcode | specimen identification | Germany

## 1 Introduction

Soils are a habitat for a vast variety of animals and microorganisms (Anthony et al. 2023), and its biological activity is the fundamental basis of most soil functions essential to human well-being, such as food production and water filtration (Lavelle et al. 2006, Wall 2012, Anthony et al. 2023). However, our knowledge on the diversity in our soils is far from being complete, and the need to deepen our knowledge of these insufficiently researched biotic communities has been emphasized for years (e.g. Wagg et al. 2014, Griffiths et al. 2016, Orgiazzi et al. 2016, and Gonzalez et al. 2023).

In soil, Collembola (springtails) are one of the major functional groups due to their role in, e.g. nutrient cycling, or as indicators of ecosystem health (Mebes

© 2026 by the author(s), licensed by Senckenberg Gesellschaft für Naturforschung. This article is published under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

SENCKENBERG
NATURE RESEARCH

ISSN 1864-6417 · eISSN 2509-9523
www.soil-organisms.org

& Filser 1998, Brussaard et al. 2007, Bispo et al. 2009). However, conventional morphological and molecular approaches for recording soil fauna are in many cases insufficient for the characterization of such complex mesofauna communities (Arribas et al. 2016). Collembola are conventionally identified to the species level by preparation on slides, followed by microscopic examination of their morphological features; therefore, a community assessment of this group based on morphological methods is very time-consuming as it has to process up to 50,000-100,000 specimens per m² (Bardgett & van der Putten 2014) and renders inoperable when dealing with damaged or juvenile specimens. During the last decade the number of Collembola known to science has raised from 8,500 in 2014 to 9,600 in 2024 (Bellinger et al. 1996-2024), i.e. about 100 springtail species per year are either newly described or rearranged as valid species. Both facts pose challenges for species identification and require additional tools for specimen identification to be included. The usage of digital images to record distinctly coloured Collembola has proven promising (e.g. van Bezouw et al. 2022), however, an identification based on photography alone bears severe risks, as emphasized by e.g. Ceriaco et al. (2016). At the molecular level, COI DNA barcoding, using a part of the 5'-sequence of the mitochondrial DNA cytochrome c oxidase subunit I gene (mt COI) has been implemented as efficient technique for Collembola identification (Hogg & Hebert 2004), including immature stages or aberrant specimens among colour pattern species. However, the potential of a single genetic loci to yield data that are sufficient in taxonomic scope to recognize many species lineages has been questioned (Lorenz et al. 2005, Rubinoff 2006, Whitworth et al. 2007, Buhay 2009). Integrative taxonomic approaches have been proposed to overcome flaws of single method approaches (Schlick-Steiner et al. 2010, Dejaco et al. 2012) as well as the taxonomic impediment spanning between requirements of biodiversity assessment and a growing deficit of taxonomists (e.g. Rougerie et al., 2009; Porco et al. 2012, Engel et al. 2021). Meanwhile it is possible to construct genome assemblies from single soil invertebrate specimens (Collins et al. 2023). Although advances in DNA barcoding are highly promising for future ease and standardisation of soil organism monitoring purposes, it has also been stressed that current barcode reference libraries are far from completion, which makes it difficult to effectively identify little-known taxon groups. Besides, inappropriate reference libraries increase the risk of false positive and false negative results, regardless of the morphological quality of the submitted identification data (e.g. Jänsch et al. 2023, Recuero et al. 2024, Ristok et al. 2025).

It is therefore desirable to know how many of the morphologically defined species already have barcode sequences available, and to what extent widespread and common Collembola species are already present in barcode reference libraries, including separable genetic lineages. This would allow to identify gaps in our knowledge and determine whether certain families or genera need to be studied more intensively in order to either improve the coverage of libraries or identify reasons for unexpected heterogeneity in genetic lineages. Future efforts should initially focus on closing the gaps in the Collembola checklist, prioritising those species that are found more frequently in such samplings.

Our aim was first to analyse the degree of completion of such barcode reference databases for Collembola in the federal state of Baden-Württemberg, Germany, and in a second step to produce additional sequence data from common soil-dwelling Collembola to fill detected gaps in barcoding reference libraries. Baden-Württemberg was chosen because studies on soil biodiversity have been conducted there for decades based on its cross-media environmental monitoring (see following paragraph for details), which makes it the federal state in Germany with the most detailed knowledge on Collembola biodiversity. As a third step, we used these newly obtained sequences to determine how accurately they are assigned to a respective taxon by the identification systems of both BOLD and GenBank reference sequence databases, from species level upwards to higher taxonomical ranks, or whether there is a mismatch. Finally, we estimated whether a large number of sequences already stored for a given species correlates with a high degree of accuracy for this species in the aforementioned identification engines.

## 2 Study area

### 2.1 Biological Monitoring Network by LUBW Baden-Württemberg

Sampling material was primarily taken from 11 forest locations that are part of the Biological Monitoring Network (BMN). This BMN program, conducted by the Baden-Württemberg State Institute for the Environment (LUBW), serves as an essential tool for analysing interactions between the environmental media of soil, water, and air. The primary objective of this program is to develop a comprehensive understanding of the impacts of anthropogenic activities on the environment, facilitating the early detection and response to ecological changes. Originally established to address pressing issues like acid rain and forest dieback, the BMN has since evolved into

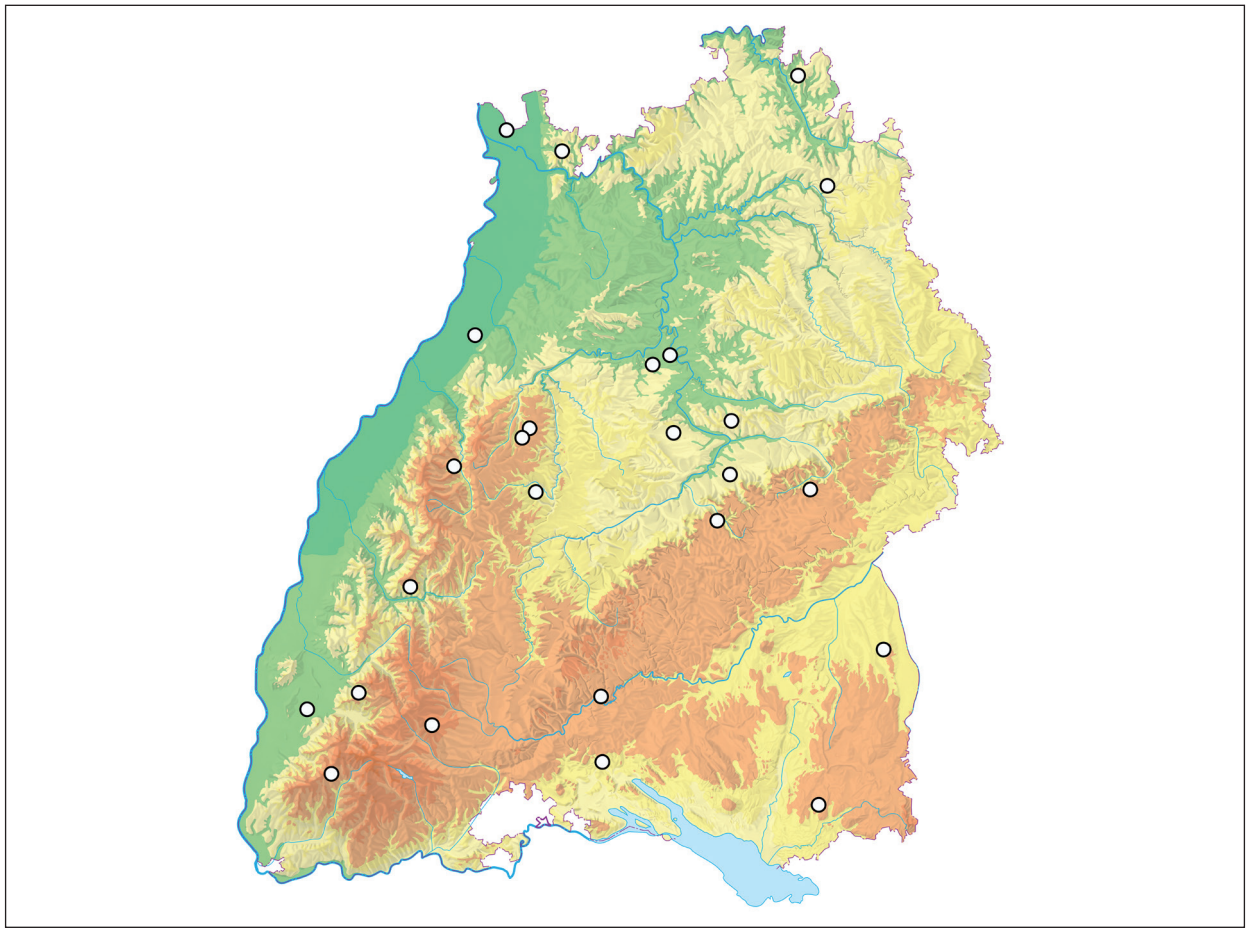

**Figure 1.** Location of sampling sites within Baden-Württemberg (Germany). Map source: LGL, www.lgl-bw.de

a comprehensive monitoring program that systematically examines selected areas in Baden-Württemberg since 1984. Currently, the BMN encompasses 21 forest and 18 grassland permanent monitoring sites (Fig. 1) distributed across the state, allowing for the study of diverse natural environments - from the warm Rhine plain to cool high altitudes of the Black Forest.

The forest areas, in particular, are characterized by a large number of recorded parameters that enable long-term observations of the environment. These long-term studies are crucial for detecting environmental changes, assessing the efficacy of environmental policies, and providing early warning signals for emerging ecological threats. Through its integrative and interdisciplinary approach, the BMN generates comprehensive environmental data, which not only reveals current pollution levels but also identifies potential future environmental challenges (LUBW 2025).

In order to increase the number of species, specimens from 21 additional forest or grassland sites situated in Baden-Württemberg were included, sampled between 2002 and 2020. Of these sites, 7 yielded sampling material of the Forest Research Institute Baden-Württemberg, Freiburg, Germany (FVA), and further material came from 14 sites of the first author's tissue sample collection (UBMA). Details on sites, habitat characteristics, sampling date, collection number and coordinates are given in Tab. S1a.

## 3 Materials and methods

To determine which common taxa occur most frequently in the forest sites of Baden-Württemberg, a list of Collembola species currently known from forest or grassland sites in Baden-Württemberg was compiled based on a query with the soil organism data warehouse Edaphobase (Burkhardt et al. 2014, Russell et al. 2024), and a second query was made from BOLD Systems database (Ratnasingham & Hebert 2007) for mt COI barcodes available from Collembola sampled in Germany. Based on both queries, we selected for DNA analysis Collembola taxa that were (i) present with more than 50 records, hence assumed to be more frequently found in soils, and (ii) so far represented in the BOLD database with few or no COI barcode

entries, or (iii) common but polytypic species, known to contain several genetic lineages, e.g. *Parisotoma notabilis*, *Ceratophysella denticulata* (Porco et al. 2012), *Isotomiella minor*, or *Folsomia quadrioculata* (Saltzwedel et al. 2016). From the resulting list, we searched for taxa contained therein in fresh soil samples provided by the LUBW that had been taken in spring and autumn of 2020 and 2021 from 11 forest sites of the BMN (LUBW 2025) and extracted with a MacFadyen high gradient apparatus (Macfadyen 1961).

Springtails were collected in undiluted ethylene glycol solution (CAS 107-21-1) with 3 drops of detergent (EcoTop by Stockmeier Chemie, Bielefeld, Germany), preserved in a 95/5 mixture (950 ml absolute ethanol + 50 ml propylene glycol), and stored at 6°C +/-2°C until further processing. Individuals were washed and sorted in 70% ethanol, and from species with distinct colour pattern a picture was taken prior to DNA extraction. Up to 11 individuals per taxon and site were selected for DNA extraction under 10-fold magnification, punctured with a needle, separately transferred to a 1.5ml reaction tube containing 150 µl TNES buffer (Sunnucks & Hales 1996) and 3 µl proteinase K, and incubated at 56°C on a thermal shaker (Ditabis MHR13) at 300 rpm for 8-12 hours. The cuticle was then removed and washed in 30% ethanol to remove remaining buffer salts. Morphological vouchers were prepared on slides in Hüther's E65 mixture (Hüther 1993), and determined under a Leica DM 2500 DIC microscope according to the respective keys of (Bretfeld 1999), Deharveng (1982), Dunger & Schlitt (2011), Fjellberg (1998, 2007), Jordana (2012), Kovac & Palacios-Vargas (2008), Mari Mutt (1980), Mateos (2008, 2011; 2012), Mateos & Petersen (2012), Mateos & Winkler (2018), Palissa (1964), Parimuchova & Kovac (2016), Pasnik & Weiner (2017), Pomorski (1990, 1998), Potapov (2001), Schneider & D'Haese (2013), Simon Benito & Palacios-Vargas (2007, 2008), Smolis & Deharveng (2017), Stach (1960), Thibaud et al. (2004), Weiner & Fiera (2020), and Winkler et al. (2021). Voucher specimens and DNA vouchers are stored at the Senckenberg Museum of Natural History Görlitz (SMNG). For details on allocation see Tab. S1.

For the COI barcoding region we used either standard primers according to Folmer (1994) or the Glom primer mix (Resch et al. 2014). PCR runs were performed according to Roithmeier et al. (2018). PCR reactions were performed in 10 µl reaction volumes containing 10-15 ng DNA, 0.12-0.15 µM primer each, 1 x PCR buffer (Nippon Genetics, Düren, Germany), with 1.5 mM $MgCl_2$, 0.2 mM 4x dNTP and 0.2 U Taq polymerase. PCR products were inspected on a 1% agarose gel, successful (i.e. single band) amplifications were purified with ExoSap-IT (Applied Biosystems by Thermo Fisher Scientific). DNA concentration was adjusted to the specifications of the sequencing companies, MicroSynth (Göttingen, Germany), Eurofins (Cologne, Germany) or StarSeq (Mainz, Germany). All obtained sequences were aligned manually in ClustalX ver. 1.83 (Chenna et al. 2003) and checked with blastn for contamination (Altschul et al. 1990).

For estimating the respective ability of both GenBank and BOLD databases to assign our barcode sequences to taxa, we deliberately used the default settings of both identification programmes as far as possible, in order to approximate a search that a normal user of the databases would also perform: In BOLD v4, we used the search algorithm of the Identification System (Ratnasingham & Hebert 2007) to compare each COI sequence with the corresponding collection of barcode sequences with the All Barcode Records option in BOLD. In GenBank, we used the megablast program to search for similar sequences in the core nucleotide database. The search results were evaluated according to the top match result presented by the respective database search. Species level match was assigned if a record with the same taxonomic name had the best match statistics (with highest similarity in percent, in GenBank the ranking is based on the lowest E value, according to default settings). A mere genus name as top match was classified as genus match, a family name as family match, an order level (e.g. 'Symphypleona sp.' for a globular springtail) as order match. Different taxa with the same top match statistics were classified as ambiguous match. Top matches with a correct genus but incorrect species name were classified as species mismatch, and fully incorrect assignments were classified as mismatch.

## 4 Results

The Edaphobase data warehouse currently lists 290 Collembola species known to occur in forests, grassland, arable land and near-natural areas of Baden-Württemberg (Edaphobase 2025, Tab. S1b). 76 of these species were present with more than 50 datasets, and for 55 of these at least one COI-sequence from sampling sites within Germany is available in the BOLD database. Sequences for 110 of these species are also available in GenBank. However, it is often not possible to assign those data to specific collection sites within a country, as DNA sequence uploads to GenBank rarely include any metadata apart from species name, sample number, and the laboratory that performed sequencing. Of those 290 species, 36% (105 species) are represented by at least one COI sequence in both BOLD and GenBank, 1.7% only in GenBank, and 26% only in the BOLD database. It should be noted that most of this COI data

originates from a single BOLD data upload (Burkhardt & LUBW 2025, dx.doi.org/10.5883/DS-LUBWG22) that raised the number of taxa present with one or more COI barcodes from 26 to 55 (Burkhardt et al. unpubl.). Currently the most frequently represented species in the BOLD database are *Entomobrya nivalis* with 53 COI sequence entries, *Isotomiella paraminor* (25 entries), *Supraphorura furcifera* (24) *Protaphorura aurantiaca* (21), *Ceratophysella denticulata* (20), *Pseudosinella alba* and *Parisotoma notabilis* (18 entries each).

Our molecular sequencing yielded mt COI sequences from 302 springtail specimens, belonging to 15 Collembola families and 68 species (plus one species group, Tab. S1c). Of these sequences, 138 came from the BMN areas, 106 from FVA areas and further 58 from UBMA areas. COI sequences varied in length, ranging from 520–658 bp due to variation in amplification and read quality. Data are available on BOLD in the public dataset dx.doi.org/10.5883/DS-BW19C21.

From our sequences amplified, 31 and 32% of all specimens matched to public sequences on species level in BOLD and GenBank, respectively (Fig. 2), and 17 and 12% at least to genus level. BOLD had a higher hit rate for correct family level identifications and lower numbers of both species mismatches and total mismatches, but also delivered 2% of ambiguous matches, especially for specimens from the genus *Isotomurus* (see Tab. S1c for details). From the 68 species tested in this study, 14 species were not present so far with a COI barcode sequence in BOLD, and 12 not in GenBank. This applied to the species *Folsomia spinosa, Friesea truncata, Hymenaphorura arantiana, Hypogastrura monticola, Isotomiella paraminor, Megalothorax willemi* (not in BOLD), *Mesaphorura tenuisensillata, Neelides minutus, Protaphorura campata* (not in BOLD), *P. glebata, P. quadriocellata, P. subarmata, Pseudosinella huetheri* and *Stenognathellus denisi*. Besides, we found in our samples *Protaphorura humata*, a species currently not regarded as valid species in the recent key by Parimuchova & Kovac (2016) – an opinion not shared by the first author.

Comparing the degree of identification accuracy to the amount of sequence data present in BOLD for the respective taxon indicated low correlation ($n = 69$, Pearson's $r = 0.15$, $p = 0.21$). Of our Collembola species, 22 were present with more than 25 barcode sequences per taxon, and 17 of these reached species level or higher in match rank (Tab. 1). GenBank often yielded a family level match (or higher) when queried for *Folsomia* spp. (n=18), *Heteromurus nitidus* (n=14), *Isotomiella paraminor* (n=12), *Tomocerus vulgaris* (n=7), *Paratullbergia callipygos* (n=5), and *Pygmarrrhopalites pygmaeus* (order level, n=5). Mismatches in GenBank queries most often occurred in the genera *Pseudosinella* spp. (n=17), *Mesaphorura* spp. (n=5), and *Stenognathellus denisi* (n=4).

Mismatches in BOLD most frequently occurred in queries for *Pseudosinella huetheri* (n=9),) and *Megalothorax minimus* (n=3); mere family level matches

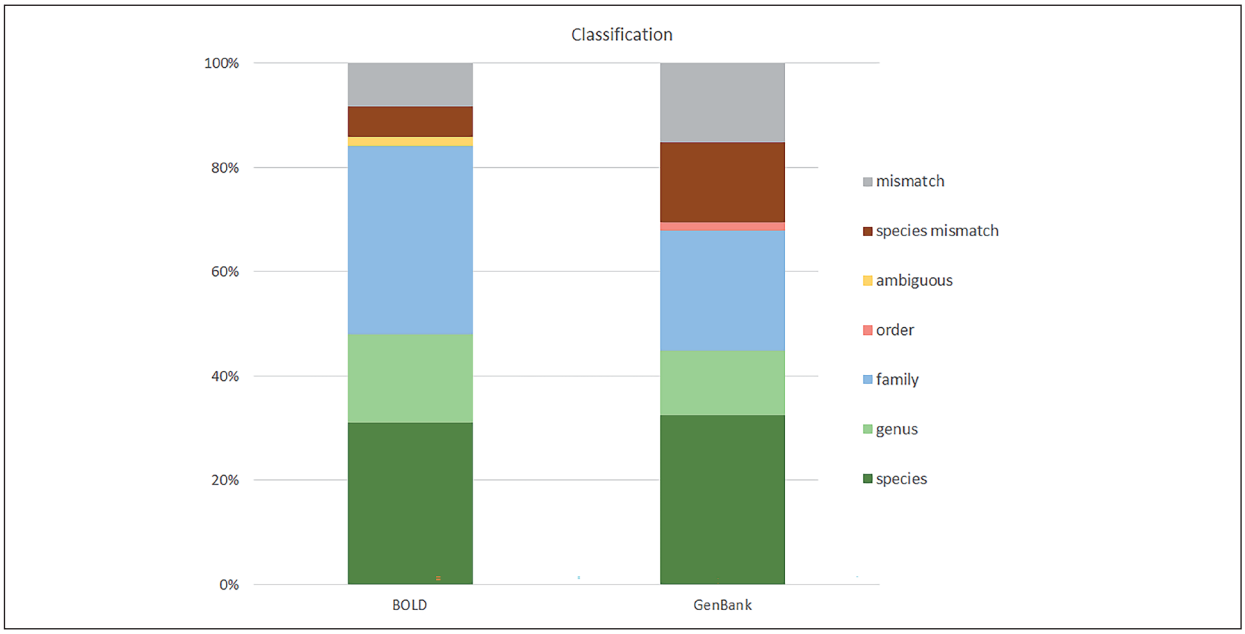


**Figure 2**. Classification of mt COI barcode sequences using BOLD and GenBank data, based on assessment of respective top match specificity in the respective database: species level match (green), genus level match (light green), family level match (blue), order rank match (red), ambiguous match (yellow) with high percentage matches but suggesting different taxa, species mismatches (i.e. correct genus but incorrect species name, brown), suggesting congeneric taxa, and mismatches (i.e. fully incorrect assignments, grey).

**Table 1.** Taxa most affected by imprecise classifications or mismatch levels (above genus or species).

| Taxon | BOLD query result % | | GenBank query result % | |
|---|---|---|---|---|
| *Folsomia* (several sp.) | Family level match | 59 | Family level match | 53 |
| *Friesea mirabilis / F. truncata* | Family level match | 100 | Family level match | 25 |
| *Heteromurus nitidus* | Family or mismatch | 100 | Family level match | 100 |
| *Isotomiella minor* | Family level match | 80 | [species match] | 100 |
| *Isotomiella paraminor* | Family or mismatch | 100 | Family or mismatch | 100 |
| *Megalothorax minimus* | Family or mismatch | 66 | Family or mismatch | 66 |
| *Mesaphorura* (several sp.) | Family level match | 66 | mismatch | 83 |
| *Paratullbergia callipygos* | [Genus level match] | 80 | Family level match | 100 |
| *Protaphorura* (several sp.) | Family level match | 41 | mismatch | 5 |
| *Pseudisotoma sensibilis* | Family level match | 100 | Family or mismatch | 100 |
| *Pseudosinella alba* | Family level match | 100 | mismatch | 100 |
| *Pseudosinella huetheri* | mismatch | 100 | mismatch | 100 |
| *Pygmarrhopalites pygmaeus* | Family level match | 100 | Order level match | 100 |
| *Stenognathellus denisi* | Family or mismatch | 100 | mismatch | 100 |
| *Tomocerus minor* | Family level match | 38 | Species mismatch | 38 |
| *Tomocerus vulgaris* | Family level match | 88 | Family level match | 88 |

occurred in *Folsomia* spp. (n=20), *Protaphorura* spp. (n=18), *Isotomiella* spp. (n=16), *Tomocerus* spp. (n=10), *Pseudosinella* alba (n=8), *Pygmarrhopalites pygmaeus* (n=5) *Heteromurus nitidus* (n=5), Pseudisotoma sensibilis (n=4), *Friesea* spp. (n=4).

## 5 Discussion

Previous research has successfully demonstrated the suitability of genetically based identification methods for soil organisms such as Collembola, so that a corresponding reference database with DNA barcoding sequences, extracted from deposited individual specimens, can make a backbone for future genetically based monitoring measures (Hendrich et al. 2014, Astrin et al. 2016, Geiger et al. 2016, Hardulak et al. 2020). The need to uncover the astounding diversity of soil arthropods through such tools that yield faster results than laborious morphology-based identification was emphasized (e.g. Packer et al. 2009, Hamilton et al. 2009, Young et al. 2021, Young & Hebert 2022) just as the importance of such tools for e.g. understanding of ecology, or conservation purposes (Gostel & Kress 2022). However, our results also underline the persisting need to achieve, for improved assignment success, better coverage in the reference libraries of both taxonomic levels finer than mere order or family (Young et al., 2021), and of divergent genetic lineages (Young et al. 2021, Ahmed et al. 2022, Antil et al. 2023).

Our results indicate good identification results, i.e. exact matches, among Collembola so far to occur primarily with widespread and morphologically easily recognisable taxa without similar sister species. Good sequence similarity with reference to several taxa ('biased') occurred in the genera *Isotomurus* and *Lepidocyrtus*. *Isotomurus* is identified in current keys (Potapov 2001, Carapelli et al. 2001) on the basis of colour patterns, albeit specimens frequently occur in transitional patterns, or pale forms, which may seriously compromise proper identification. *Lepidocyrtus*, on the other hand, is currently being recognised and processed as a systematically difficult species complex (Mateos 2008, 2012, Mateos et al. 2018, Mateos & Álvarez-Presas 2022), hence the latest findings are often not yet represented in the current keys and confusion may occur in such taxonomically altering species groups, as shown e.g. in van Bezouw et al. (2022). The same applies to species-rich genera that are only correctly recognised here at the genus or family level, such as *Folsomia*, *Friesea*, *Pseudosinella* or *Protaphorura*, as well as to classifications where the sequence is available in the database with high accuracy, but has so far been linked merely to a rough taxon classification, such as *Entomobrya*, *Mesaphorura*, *Neelides*, or *Pygmarrhopalites*. Those genera include frequently encountered species, which, however, pose difficulties to species level identification, requiring up-to-date determination keys and good microscopic equipment. This applies in particular to taxa whose latest available keys are not yet generally used, or to those which are still a subject of ongoing research due to recognised genetic

diversity, or to rather rarely encountered genera lacking up-to-date keys, such as *Pseudosinella*. Here in particular, non-destructively working barcoding projects should be able to create significant improvement and simplification for future research with regard to metabarcoding and eDNA monitoring by providing sequence data from vouchered specimens, either identified by taxonomic experts (Meiklejohn et al. 2019) or with information on the determination literature used (Meier 2017). Besides, integrating morphological and molecular methods by non-destructive extraction methods and subsequent storage of the morphological voucher in a reference collection facilitates critical verification of the determination in case of sequence discrepancies, or re-determination in cases of subsequent taxon splitting (Young et al. 2019, Buckner et al. 2021). On the other hand, such an approach remains time-consuming and expensive as it involves processing individual specimens (Borisenko et al. 2009). Besides, flawed morphological-based identification, or 'bad taxonomy' as defined in Bortolus (2008), bears a high risk to produce mixed up sequence data, thus blurring existing genetic differences between taxonomic units and hampering the detection of cryptic diversity.

We also found that a large number of stored barcode sequences in many - albeit not all - cases can improve the accuracy of the identification engine relying on these sequences. In most cases the accuracy of matches seems to be less related to the sheer number of sequence data of the respective taxon already deposited in BOLD but rather to the accuracy of morphological determination of the respective top matches. Apart from drawbacks due to taxonomic overrepresentation (Chorlton 2024), mass uploads of sequences of high quality but with unspecific taxonomic labelling therefore appear rather counterproductive for the assignment success of such reference databases. On the other hand, some seemingly easy-to-recognise species such as *Heteromurus nitidus*, *Protaphorura aurantiaca, Tomocerus vulgaris*, *T. minor,* or *Pseudisotoma sensibilis* were found to contain high genetic variability within COI, indicating that these taxa should be scrutinized for existence of possible cryptic taxa.

The barcode gap concept, i.e. the tenet that more DNA barcode variation is found across species than within species, was found to fail quite often in common soil-living microarthropod groups such as Oribatida and Collembola (Phillips et al. 2022, Le Cadre et al. 2024). As such, a wider sampling on more sites per taxon and additional use of secondary barcoding gene regions seems necessary to improve determination reliability in (meta) barcoding studies, as well as thorough morphological identification prior to database upload. Furthermore, when occurrences of species groups with persistent gaps in taxonomic knowledge are examined based solely on incomplete reference databases (Pedersen et al. 2015, Schenekar et al. 2020), there is an increased risk that best match analyses, such as those offered by gene databases, will yield matches with species whose occurrence in the respective area is highly unlikely based on their known distribution, but which are displayed as supposed matches because the suggested species are already represented by numerous sequences in the gene database (Jänsch et al. 2023). Similar to the recommended measures for suspected sample contamination (Thomsen & Willerslev 2015), new observations for a given region based solely on molecular detection methods should be scrutinised by comparing them with existing regional taxon checklists as well as verification methods based on methodologically independent replicates before the occurrence of the respective taxon is considered positive. Due to the uncertainties that still exist within barcoding methodologies, it is necessary to critically examine such DNA-based identifications in terms of integrative taxonomy (Schlick-Steiner et al. 2010) and compare them with biogeographical and ecological knowledge in order to verify the plausibility of a species' occurrence (Pentinsaari et al. 2020).

We thus recommend that when assessing a molecular barcoding database tool for specimen identification, users should keep in mind that they themselves can contribute to database improvement

- by making their own morphological and ecological data publicly available, too, after completion of their studies, for example on portals such as Edaphobase;
- by enriching also their molecular sequence data prior to upload for future data reuse (Pasquetto et al. 2017, Tenopir et al. 2020, Gotelli et al. 2021) by adding to BOLD, e.g., EUNIS-readable habitat information, sample coordinates, country information, collection methods and phenology data;
- by increasing data reliability, namely by also providing details on the used identification keys, since species misidentification can severely affect taxonomic certainty in biodiversity databases (Chorlton 2024, Coca-de-la-Iglesia et al. 2024).

Besides, for determination of the Collembola fauna in Germany somewhat outdated keys are still in use for the identification of several groups (e.g. Gisin 1960), or keys are used that were designed for other faunal regions, thus being unable to fully cover the Central European species spectrum (e.g. Fjellberg 1998; 2007; Hopkin 2007). Using instead current online available determination keys as provided by Bellinger et al. (1996-2024) can help to bridge existing gaps and quickly include newly described species. A complete edition of determination

keys for the Palaearctic Collembola has not yet been fully realised (Dunger 1994) but, as available so far, should act as the basis for the current state of working, and be supplemented by current group keys for individual genera or species groups available in literature, as was shown to be practicable in the present study.

## Online Supplemetary Material

**S1a Table**. Sampling site accompanying data
**S1b Table**. List of Collembola observed in Baden-Württemberg (Germany) forests, grassland and arable land, based on a data query in Edaphobase data warehouse (https://portal.edaphobase.org, acc. 2025-07-18)
**S1c Table**. BOLD and GenBank output information used for statistical analysis for the 302 tested Collembola specimens

## Acknowledgements

Sample material from the Forest Research Institute Baden-Württemberg, Department of Soil and Environment (FVA), was part of the project 'monitoring of forest soil fauna', funded by: 'Sonderprogramm zur Stärkung der biologischen Vielfalt in Baden-Württemberg'. The collection of data in this study was partially funded by the German Federal Ministry of Research, Technology and Space as part of the GBOL I project (www.bolgermany.de). The first author wishes to thank Senckenberg Nature Research Society for kind permission to use their infrastructure during the preparation of this study. We thank two anonymous referees for their helpful suggestions and contributions to this manuscript.

## References

Ahmed, S., M. Ibrahim, C. Nantasenamat, M. F. Nisar, A. A. Malik, R. Waheed, M. Z. Ahmed, S. C. Ojha, M. K. Alam & S. Khan (2022). Pragmatic Applications and Universality of DNA Barcoding for Substantial Organisms at Species Level. A Review to Explore a Way Forward. *BioMed Research International*, 2022(2), 1–19.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers & D. J. Lipman (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.

Anthony, M. A., S. F. Bender & M. G. A. van der Heijden (2023). Enumerating soil biodiversity. *Proceedings of the National Academy of Sciences,* 120(33).

Antil, S., J. S. Abraham, S. Sripoorna, S. Maurya, J. Dagar, S. Makhija, P. Bhagat, R. Gupta, U. Sood, R. Lal & R. Toteja (2023). DNA barcoding, an effective tool for species identification. A review. *Molecular Biology Reports*, 50(1), 761–775.

Arribas, P., C. Andújar, K. Hopkins, M. Shepherd, A. P. Vogler & D. Yu (2016). Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, 7 (9), 1071–1081.

Astrin, J. J., H. Höfer, J. Spelda, J. Holstein, S. Bayer, L. Hendrich, B. A. Huber, K.-H. Kielhorn, H.-J. Krammer, M. Lemke, J. C. Monje, J. Morinière, B. Rulik, M. Petersen, H. Janssen, C. Muster & M. Kuntner (2016). Towards a DNA Barcode Reference Database for Spiders and Harvestmen of Germany. *PLOS ONE*, 11(9). e0162624.

Bardgett, R. D. & W. H. van der Putten (2014). Belowground biodiversity and ecosystem functioning. *Nature,* 515, 505–511.

Bellinger, P. F., K. Christiansen & F. Janssens (1996-2024). Checklist of the Collembola of the World. https://www.collembola.org/.

Bispo, A., D. Cluzeau, R. Creamer, M. Dombos, U. Graefe, P. Krogh, J. Sousa, G. Peres, M. Rutgers, A. Winding & J. Römbke (2009). Indicators for monitoring soil biodiversity. *Integrated environmental assessment and management*, 5(4), 717–719.

Borisenko, A. V., J. E. Sones & P. D. N. Hebert (2009). The front-end logistics of DNA barcoding. *Challenges and prospects. Molecular Ecology Resources*, 9 Suppl s1, 27–34.

Bortolus, A. (2008). Error Cascades in the Biological Sciences. The unwanted consequences of using bad taxonomy in ecology. AMBIO: *A Journal of the Human Environment*, 37(2), 114–118.

Bretfeld, G. (1999). Synopses on Palaearctic Collembola Part II: Symphypleona. *Abhandlungen und Berichte des Naturkundemuseums Görlitz,* 71(1), 1–318.

Brussaard, L., P. C. de Ruiter & G. G. Brown (2007). Soil biodiversity for agricultural sustainability. *Agriculture, Ecosystems & Environment*, 121(3), 233–244.

Buckner, J. C., R. C. Sanders, B. C. Faircloth & P. Chakrabarty (2021). *The critical importance of vouchers in genomics. eLife*, 10.

Buhay, J. E. (2009). "COI-like" Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. *Journal of Crustacean Biology*, 29(1), 96–110.

Burkhardt, U., D. J. Russell, P. Decker, M. Döhler, H. Höfer, S. Lesch, S. Rick, J. Römbke, C. Trog, J. Vorwald, E. Wurst & W. E. R. Xylander (2014). The Edaphobase Project of GBIF-Germany - A new online soil-zoological data warehouse. *Applied Soil Ecology*, 83, 3–12.

Carapelli, A., F. Frati, P. P. Fanciulli & R. Dallai (2001). Taxonomic revision of 14 south-western European species of *Isotomurus* (Collembola, Isotomidae), with description

of four new species and the designation of the neotype for *I.palustris*. *Zoologica Scripta*, 30(2), 115–143.

Ceríaco, L. M. P., E. E. Gutiérrez & A. Dubois (2016). Photography-based taxonomy is inadequate, unnecessary, and potentially harmful for biological sciences. *Zootaxa*, 4196(3), zootaxa.4196.3.9.

Chorlton, S. D. (2024). Ten common issues with reference sequence databases and how to mitigate them. *Frontiers in bioinformatics*, 4(1278228).

Coca-de-la-Iglesia, M., A. Gallego-Narbón, A. Alonso & V. Valcárcel (2024). High rate of species misidentification reduces the taxonomic certainty of European biodiversity databases of ivies (*Hedera* L.). *Scientific Reports*,14(1), 4876.

Collins, G., C. Schneider, L. L. Boštjančić, U. Burkhardt, A. Christian, P. Decker, I. Ebersberger, K. Hohberg, O. Lecompte, D. Merges, H. Muelbaier, J. Romahn, J. Römbke, C. Rutz, R. Schmelz, A. Schmidt, K. Theissinger, R. Veres, R. Lehmitz, M. Pfenninger & M. Bálint (2023). The MetaInvert soil invertebrate genome resource provides insights into below-ground biodiversity and evolution. *Communications Biology*, 6(1), 1241.

Deharveng, L. (1982). Clé de détermination des genres de Neanurinae (Collemboles) d'Europe et de la région méditerranéenne, avec description de deux nouveaux genres. *Travaux du Laboratoire d'Ecobiologie des Arthropododes édaphiques, Toulouse* 3, 7–13.

Dejaco, T., T. Arthofer, H. D. Sheets, K. Moder, B. Thaler-Knoflach, E. Christian, L. F. Mendes, B. C. Schlick-Steiner & F. M. Steiner (2012). A toolbox for integrative species delimitation in *Machilis* jumping bristletails (Microcoryphia: Machilidae). *Zoologischer Anzeiger,* 251(4), 307–316.

Dunger, W. (1994). Synopses on Palaearctic Collembola: Introduction. *Abhandlungen und Berichte des Naturkundemuseums Görlitz,* 68(3), 1–4.

Dunger, W. & B. Schlitt (2011). Synopses on Palaearctic Collembola Part VI-1: Tullbergiidae. *Soil Organisms*, 83(1), 1–168.

Edaphobase (2025). Edaphobase data warehouse for soil organisms. Accessed via the Edaphobase Data-Query Portal. https://portal.edaphobase.org.

Engel, M., L. Ceríaco, G. Daniel, P. Dellapé, I. Löbl, M. Marinov, R. Reis, M. Young, A. Dubois, I. Agarwal, P. Lehmann A., M. Alvarado, N. Alvarez, F. Andreone, K. Araujo-Vieira, J. Ascher, D. Baêta, D. Baldo, S. Bandeira & C. Kusamba (2021). The taxonomic impediment. A shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean Society*, 193, 381–387.

Fjellberg, A. (1998). The Collembola of Fennoscandia and Denmark. Part 1. Poduromorpha. *Fauna Entomologica Scandinavica,* 35, 1–184.

Fjellberg, A. (2007). The Collembola of Fennoscandia and Denmark. Part 2. Entomobryomorpha and Symphypleona. *Fauna Entomologica Scandinavica*, 42, 1–266.

Folmer, O., M. Black, W. Hoeh, R. Lutz & R. Vrijenhoek (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* 3(5), 294–299.

Geiger, M. F., J. J. Astrin, T. Borsch, U. Burkhardt, P. Grobe, R. Hand, A. Hausmann, K. Hohberg, L. Krogmann, M. Lutz, J. C. Monje, B. Misof, J. Moriniere, K. F. Müller, S. Pietsch, D. Quandt, B. Rulik, M. Scholler, W. Traunspurger, G. Haszprunar & W. Wägele (2016). How to tackle the molecular species inventory for an industrialized nation - lessons from the first phase of the German Barcode of Life initiative GBOL (2012-2015). *Genome*, 59(9), 661–670.

Gisin, H. (1960). *Collembolenfauna Europas*. Museum D'Histoire Naturelle, Genève, 1–311.

Gonzalez, A., P. Vihervaara, P. Balvanera, A. E. Bates, E. Bayraktarov, P. J. Bellingham, A. Bruder, J. Campbell, M. D. Catchen, J. Cavender-Bares et al (2023). A global biodiversity observing system to unite monitoring and guide action. *Nature Ecology & Evolution*, 7(12), 1947–1952.

Gostel, M. R. & W. J. Kress (2022). The Expanding Role of DNA Barcodes. Indispensable Tools for Ecology, Evolution, and Conservation. *Diversity*, 14(3), 213.

Gotelli, N. J., D. B. Booher, M. C. Urban, W. Ulrich, A. V. Suarez, D. K. Skelly, D. J. Russell, R. J. Rowe, M. Rothendler, N. Rios, S. M. Rehan, G. Ni, C. S. Moreau, A. E. Magurran, F. A. M. Jones, G. R. Graves, C. Fiera, U. Burkhardt & R. B. Primack (2021). Estimating species relative abundances from museum records. *Methods in Ecology and Evolution*, 14(2), 431–443.

Griffiths, B. S., J. Römbke, R. M. Schmelz, A. Scheffczyk, J. H. Faber, J. Bloem, G. Pérès, D. Cluzeau, A. Chabbi, M. Suhadolc et al. (2016). Selecting cost effective and policy-relevant biological indicators for European monitoring of soil biodiversity and ecosystem function. *Ecological Indicators,* 69, 213–223.

Hamilton, H. C., M. S. Strickland, K. Wickings, M. A. Bradford & N. Fierer (2009). Surveying soil faunal communities using a direct molecular approach. *Zoologischer Anzeiger*, 41(6), 1311–1314.

Hardulak, L. A., J. Morinière, A. Hausmann, L. Hendrich, S. Schmidt, D. Doczkal, J. Müller, P. D. N. Hebert & G. Haszprunar (2020). DNA metabarcoding for biodiversity monitoring in a national park. Screening for invasive and pest species. *Molecular Ecology Resources*.

Hendrich, L., J. Moriniere, G. Haszprunar, P. D. N. Hebert, A. Hausmann, F. Köhler & M. Balke (2014). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*.

Hogg, I. D. & P. D. N. Hebert (2004). Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology*, 82(5), 749–754.

Hopkin, S. P. (2007). A Key to the Collembola (springtails) of Britain and Ireland. *Field Studies Council, Totnes / Devon,* 252 pp.

Hüther, W. (1993). Sammel- und Präparationsmethoden für Bodenarthropoden. *Abhandlungen und Berichte des Naturkundemuseums Görlitz,* 66 (6), 1–11.

Jänsch, S., A. Scheffczyk, J. Römbke, V. Rojo, J. Vierna, A. Vizcaino, T. Natal da Luz, D. Alves, P. Martins, S. Mendes, L. Scopel, J. P. Sousa, P. H. Krogh, R. Sapkota & R. Schmelz (2023). Bewertung der Biologischen Vielfalt mittels DNA-Extraktion aus Bodenproben von Dauerbeobachtungsflächen (BDF). *UBA-Texte* 142/2023. https://www.umweltbundesamt.de/publikationen/bewertung-der-biologischen-vielfalt-mittels-dna.

Jordana, R. (2012). Synopses on Palaearctic Collembola 7/1: Capbryinae and Entomobryini. *Soil Organisms*, 84(1), 1–390.

Kovac, L. & J. G. Palacios-Vargas (2008). Redescription of *Isotomiella alulu* and *I. delamarei* (Collembola: Isotomidae) with notes on the systematics of the genus and new records from the Neotropics. *Zootaxa*, 1825, 1–17.

Lavelle, P., T. Decaens, M. Aubert, S. Barot, M. Blouin, F. Bureau, P. Margerie, P. Mora & J. P. Rossi (2006). Soil invertebrates and ecosystem services. *European Journal of Soil Biology*, 42 (Supplement1), S3-S15.

Le Cadre, J., F. L. Klemp, M. Bálint, S. Scheu & I. Schaefer (2024). Applicability and perspectives for DNA barcoding of soil invertebrates. *PeerJ*, 12: e17709.

Lorenz, J. G., W. E. Jackson, J. C. Beck & R. Hanner (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 360: 1869–1878.

LUBW (2025). Baden-Wuerttemberg State Institute for the Environment, dept. 23 (LUBW). *Medienübergreifende Umweltbeobachtung.* https://www.lubw.baden-wuerttemberg.de/medienuebergreifende-umweltbeobachtung.

Macfadyen, A. (1961). Improved funnel-type extractors for soil arthropods. *Journal of Animal Ecology*, 30(1), 171–184.

Mari Mutt, J. A. (1980). A revision of *Heteromurus* s. str. (Insecta: Collembola: Entomobryidae). *Transactions of the Illinois State Academy of Sciences*, 72(3), 29–50.

Mateos, E. (2008). The European *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). *Zootaxa* 1769: 35–59.

Mateos, E. (2011). New *Lepidocyrtus* Bourlet, 1839 taxa from Greece (Collembola: Entomobryidae). Zootaxa, 3108, 25–40.

Mateos, E. (2012). The European *Lepidocyrtus lanuginosus* group (Collembola: Entomobryidae), definition and description of a new species from Spain. *Zootaxa*, 3570, 69–81.

Mateos, E. & M. Álvarez-Presas (2022). Integrative taxonomy reveals three new species of European *Lepidocyrtus lignorum*-group (Collembola, Entomobryidae). *Zootaxa*, 5100(4), 451–481.

Mateos, E., P. Escuer, G. Buşmachiu, M. Riutort & M. Álvarez-Presas (2018). Untangling Lepidocyrtus (Collembola, Entomobryidae). New molecular data shed light on the relationships of the European groups. *Invertebrate Systematics*.

Mateos, E. & H. Petersen (2012). Definition of the European *Lepidocyrtus curvicollis* group (Collembola: Entomobryidae) with description of a new species from Sardinia (Italy). *Zootaxa*, 3237, 51–62.

Mateos, E. & D. Winkler (2018). New data clarifying the taxonomy of European members of the *Lepidocyrtus pallidus–serbicus* group (Collembola, Entomobryidae). *Zootaxa*, 4429(3).

Mebes, K. H. & J. Filser (1998). Does the species composition of Collembola affect nitrogen turnover? *Applied Soil Ecology*, 9(1-3), 241–247.

Meier, R. (2017). Citation of taxonomic publications. The why, when, what and what not. *Systematic Entomology,* 42(2), 301–304.

Meiklejohn, K. A., N. Damaso & J. M. Robertson (2019). Assessment of BOLD and GenBank - Their accuracy and reliability for the identification of biological materials. *PLOS ONE*, 14(6), e0217084.

Orgiazzi, A., R. Bardgett, E. Barrios, V. Behan-Pelletier, M. Briones, J.-L. Chotte, G. B. Deyn, P. Eggleton, N. Fierer, T. D. Fraser, K. Hedlund, S. Jeffery, N. Johnson, A. Jones, E. Kandeler, N. Kaneko, P. Lavelle, P. Lemanceau, L. Miko & D. Wall (eds) (2016). *Global Soil Biodiversity Atlas*. Luxembourg,176 pp.

Packer, L., J. Gibbs, C. Sheffield & R. Hanner (2009). DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, 9, 42–50.

Palissa, A. (1964). Collembola. In: Brohmer, P., P. Ehrmann & G. Ulmer (eds), Die Tierwelt Mitteleuropas, Band 4 Insekten 1. Teil, 1. Abschnitt. *Verlag von Quelle & Meyer*, Leipzig, 2–299.

Parimuchova, A. & Ľ. Kovac (2016). A new cave species of the genus Protaphorura Absolon, 1901 (Collembola, Onychiuridae) from the Western Carpathians (Slovakia) with critical comments to the Palaearctic representatives of the genus. *Zootaxa*, 4098(2), 254–272.

Pasnik, G. & W. M. Weiner (2017). First phylogenetic analysis of the tribe Oligaphorurini (Collembola. Onychiuridae) inferred from morphological data, with implications for generic classification. *Organisms Diversity & Evolution*, 17(3), 619–631.

Pasquetto, I. V., B. M. Randles & C. L. Borgman (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16, 8.

Pedersen, M. W., S. Overballe-Petersen, L. Ermini, C. D. Sarkissian, J. Haile, M. Hellstrom, J. Spens, P. F. Thomsen, K. Bohmann, E. Cappellini, I. B. Schnell, N. A. Wales, C. Carøe, P. F. Campos, A. M. Z. Schmidt, M. T. P. Gilbert, A.

J. Hansen, L. Orlando & E. Willerslev (2015). Ancient and modern environmental DNA. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences*, 370(1660), 20130383.

Pentinsaari, M., S. Ratnasingham, S. E. Miller & P. D. N. Hebert (2020). BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries? *PLOS ONE* 15(4), e0231814.

Phillips, J. D., D. J. Gillis & R. H. Hanner (2022). Lack of Statistical Rigor in DNA Barcoding Likely Invalidates the Presence of a True Species' Barcode Gap. *Frontiers in Ecology and Evolution*, 10, 44.

Pomorski, R. J. (1990). Morphological-systematic studies on the variability of pseudocelli and some morphological characters in *Onychiurus* of the *"armatus*-group*"* (Collembola, Onychiuridae) Part II. On synonyms within the *''armatus*-group'', with special reference to diagnostic characters. *Annales Zoologici*, 43, 535–575.

Pomorski, R. J. (1998). Onychiurinae of Poland (Collembola: Onychiuridae), Genus - International Journal of Invertebrate Taxonomy (Supplement). *Polish Taxonomical Society*, Wroclaw, 1–201.

Pomorski, R. J. (2001). Review of the genus *Hymenaphorura* Bagnall, 1948 with description of eleven new species from North America and Siberia (Collembola: Onychiuridae). *Insect Systematics and Evolution*, 32(4), 439–474.

Porco, D., A. Bedos, P. Greenslade, C. Janion, D. Skarzynski, M. I. Stevens, B. J. van Vuuren & L. Deharveng (2012). Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebrate Systematics*, 26, 470–477.

Porco, D., L. Deharveng, T. Decaens, S. James, P. Stoev, D. Walter & P. D. N. Hebert (2012). *DNA barcoding soil invertebrates: lighting up the underground*. Book of Abstracts XVI ICSZ, 60.

Potapov, M. B. (2001). Synopses on Palaearctic Collembola Part III: Isotomidae. *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 73(2), 1–603.

Ratnasingham, S. & P. D. N. Hebert (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, **7**(3), 355–364.

Recuero, E., F. E. Etzler & M. S. Caterino (2024). Most soil and litter arthropods are unidentifiable based on current DNA barcode reference libraries. *Current Zoology,* 70(5), 637–646.

Resch, M. C., J. Shrubovych, D. Bartel, N. U. Szucsich, G. Timelthaler, Y. Bu, M. Walzl & G. Pass (2014). Where Taxonomy Based on Subtle Morphological Differences Is Perfectly Mirrored by Huge Genetic Distances: DNA Barcoding in Protura (Hexapoda). *PLOS ONE*, 9(3), e90653.

Ristok, C., D. Babin, B. Bertkowski, B. Burkhard, J. Filser, K. Hohberg, J. Kleemann, S. Kolb, R. Lehmitz, M. C. Rillig, J. Römbke, L. Ruess, S. Scheu, N. Scheunemann, A. Schmidt, B. Steinhoff-Knopp, N. Eisenhauer, C. C. Tebbe & W. E. R. Xylander (2025). Towards a comprehensive assessment of soil biodiversity in Germany: status quo, challenges, and policy implications. *Soil Organisms*, 97(2), 143–157

Roithmeier, O., U. Burkhardt, E. Daghighi & J. Filser (2018). *Desoria trispinata* (MacGillivray, 1896), a promising model Collembola species to study biological invasions in soil communities. *Pedobiologia,* 67, 45–56.

Rougerie, R., T. Decaens, L. Deharveng, D. Porco, S. W. James, C. H. Chang, B. Richard, M. Potapov, Y. Suhardjono & P. D. N. Hebert (2009). DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira*, 44(8), 789–802.

Rubinoff, D. (2006). Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, 20(4), 1026–1033.

Russell, D. J., E. Naudts, N. A. Soudzilovskaia, M.J.I. Briones, M. Çakır, E. Conti, J. Cortet, C. Fiera, D. H. Kutuzovic, M. Hedde, K. Hohberg, D. Indjic, P. H. Krogh, R. Lehmitz, S. Lesch, Z. Marjanovic, C. Mulder, L. Mumladze, M. Murvanidze, S. Rick, M. Roß-Nickoll, J. Schlaghamerský, O. Schmidt, O. Shelef, M. Suhadolc, M. Tsiafouli, A. Winding, A. Zaytsev & A. Potapov (2024). Edaphobase 2.0. Advanced international data warehouse for collating and using soil biodiversity datasets. *Applied Soil Ecology*, 204(3), 105710.

Saltzwedel, H. von, S. Scheu & I. Schaefer (2016). Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. *BMC Evolutionary Biology*, 16(1), 69.

Schenekar, T., M. Schletterer, L. A. Lecaudey & S. J. Weiss (2020). Reference databases, primer choice, and assay sensitivity for environmental metabarcoding. Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. *River Research and Applications*, 36(7), 1004–1013.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian & R. H. Crozier (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology*, 55, 421–438.

Schneider, C. & C. A. D'Haese (2013). Morphological and molecular insights on *Megalothorax*: the largest Neelipleona genus revisited (Collembola). *Invertebrate Systematics*, 27, 317–364.

Simon Benito, J. C. & J. G. Palacios-Vargas (2007). New species of *Pseudosinella* (Collembola: Entomobryidae) from Iberian Peninsula. *Zootaxa*, 1479, 9–19.

Simon Benito, J. C. & J. G. Palacios-Vargas (2008). Four new *Pseudosinella* (Collembola: Entomobryidae) from La Rioja, Spain. *Zootaxa*, 1728, 59–68.

Smolis, A. & L. Deharveng (2017). *Neanura judithae* n. sp. from Polish Carpathians, with an updated and illustrated key to all species of the genus Neanura MacGillivray, 1893 (Collembola, Neanuridae). *Zoosystema,* 39(1), 37–47.

Stach, J. (1960). Materialien zur Kenntnis der Collembolenfauna Afghanistans [Contribution à l'étude de la faune d'Afghanistan]. *Acta Zoologica Cracoviensia*, 5(12), 507–582.

Sunnucks, P. & D. F. Hales (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the Genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, 13(3), 510–524.

Tenopir, C., N. M. Rice, S. Allard, L. Baird, J. Borycz, L. Christian, B. Grant, R. Olendorf & R. J. Sandusky (2020). Data sharing, management, use, and reuse. Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3), e0229003.

Thibaud, J.-M., H. J. Schulz, M. M. da Gama & W. Dunger (2004). Synopses on Palaearctic Collembola Part IV: Hypogastruridae. *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 75(2), 1–287.

Thomsen, P. F. & E. Willerslev (2015). Environmental DNA. An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18.

van Bezouw, R., J. McCulloch, F. Janssens & M. Berg (2022). An emended description of *Sminthurinus lawrencei* Gisin 1963, with notes on the identification of black *Sminthurinus* species. *Soil Organisms*, 94(3), 127–138.

Wagg, C., F. Bender, F. Widmer & M. G.-A. van der Heijden (2014). Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences of the United States of America,* 111, 5266–5270.

Wall, D. H. (ed) (2012). Soil ecology and ecosystem services. *Oxford University Press*, Oxford, 406 pp.

Weiner, W. M. & C. Fiera (2020). Two new species of *Hymenaphorura* Bagnall, 1948 (Collembola, Onychiuridae) from Romania and an updated key to the genus. *ZooKeys*, 960, 17–38.

Whitworth, T. L., R. D. Dawson, H. Magalon & E. Baudry (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera. Calliphoridae). *Proceedings of the Royal Society B-Biological Sciences*, 274(1619), 1731–1739.

Winkler, D., M. T. Németh & C. Fiera (2021). A new species of *Pseudosinella* Schäffer, 1897 (Collembola, Entomobryidae) from Hungary and Romania, with redescription of the related species *Pseudosinella huetheri* Stomp, 1971. *ZooKeys*, 1063, 121–137.

Young, M. R., J. R. deWaard & P. D. N. Hebert (2021). DNA barcodes enable higher taxonomic assignments in the Acari. *Scientific Reports*, 11(1), 2963.

Young, M. R. & P. D. N. Hebert (2022). Unearthing soil arthropod diversity through DNA metabarcoding. *PeerJ*, 10 (6822), e12845.

Young, M. R., M. L. Moraza, E. Ueckermann, D. Heylen, L. F. Baardsen, J. F. Lima-Barbero, S. Gal, E. Gavish-Regev, Y. Gottlieb, L. Roy, E. Recht, M. El Adouzi & E. Palevsky (2019). Linking morphological and molecular taxonomy for the identification of poultry house, soil, and nest dwelling mites in the Western Palearctic. *Scientific Reports*, 9(1), 5784.