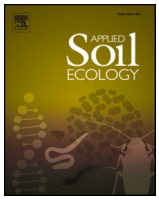
Applied Soil Ecology 228 (2026) 107549



Contents lists available at ScienceDirect

Applied Soil Ecology

journal homepage: www.elsevier.com/locate/apsoil



Research paper



Challenges of soil arthropod community assessments: Large site-related variation and hidden species-level diversity

María Fernández-Bravo [a,b,*], Florian Gschwend [a], Johanna Mayerhofer [a], Janine Moll-Mielewczik [c], Franco Widmer [a], Jürg Enkerli [a]

[a] Molecular Ecology, Agroscope, Reckenholzstrasse 191, Zürich, CH-8046, Switzerland
[b] Agricultural Entomology, Department of Agronomy, ETSIAM, University of Córdoba, Campus de Rabanales, International Excellence Agrifood Campus, CeiA3, C4 Celestino Mutis, Córdoba, 14071, Spain
[c] Soil Quality and Soil Use, Agroscope, Reckenholzstrasse 191, Zürich, CH-8046, Switzerland

ARTICLE INFO

Keywords:
Metabarcoding
COI
Soil fauna
Land-use type
Collembola
Arachnida
Insecta

ABSTRACT

Arthropods are essential for maintaining soil ecosystem functioning. However, their full diversity remains insufficiently explored due to their taxonomic complexity, the limitations of traditional methods such as morphotaxonomy, and constraints associated with DNA barcoding approaches. In this study, we applied COI-based DNA metabarcoding to survey soil arthropod richness and community composition at a single timepoint across 28 long-term monitoring sites in Switzerland representing three land-use types, i.e., arable land, permanent grassland, and forest. A total of 6464 arthropod amplicon sequence variants (ASVs) were identified, with forests exhibiting the highest richness, followed by permanent grasslands and arable land. Key environmental drivers of soil arthropod community composition included soil C:N ratio, mean annual temperature, and metabolic quotient. The arthropod ASVs were classified into 389 distinct taxa, including 214 species-level assignments. The maximum number of ASVs assigned to a single species was 180 for the collembolan *Lepidocyrtus cyaneus*. More detailed phylogenetic analyses of the genus *Lepidocyrtus*, containing 232 ASVs assigned to 4 species, revealed that i) ASVs assigned to a species and reference sequences annotated as that species were often distributed across several phylogenetic clusters, whereas in other cases phylogenetic clusters included sequences with different species-level assignments or taxonomic annotations, and ii) phylogenetic clusters were associated with specific land-use types or land-use type combinations, indicating potential hidden diversity. Two additional parthenogenetic collembolan species, *Folsomia candida* and *Parisotoma notabilis*, included more than 100 ASVs, revealing a high potential for hidden diversity. Together, our results highlight the potential of arthropod-specific metabarcoding combined with phylogenetic analyses to characterize hidden diversity in complex mesofaunal communities by linking ecological and phylogenetic information.

1. Introduction

Soil arthropods, which constitute a considerable fraction of the soil meso- and macrofauna, encompass various taxonomic groups, including springtails, mites, and insects (Andújar et al., 2022; Hermans et al., 2022). They contribute to important functions and ecosystem services in soil ecosystems, such as the decomposition of organic matter, mineralization of nutrients, and, due to their physical activity, they can improve soil aeration and support the dissemination of microorganisms (Delgado-Baquerizo et al., 2020; Sahdra et al., 2025). The total number of arthropod species in the soil remains uncertain, and the large proportion of undescribed species makes it difficult to fully characterize and understand their communities (Dopheide et al., 2019).

The assessment of soil arthropod communities is challenging due to their large taxonomic diversity. Soil arthropod communities may be assessed based on individual identification of organisms, which requires expert taxonomist knowledge or sequencing of a specific DNA marker (DNA barcoding) (Cook et al., 2010; Cristescu, 2014). These approaches are time-consuming and limit the number of samples that can be analyzed. The development of DNA metabarcoding, which refers to high-throughput DNA sequencing-based identifications of entire communities, has addressed some of these limitations in arthropod

* Corresponding author at: Molecular Ecology, Agroscope, Reckenholzstrasse 191, CH-8046, Zürich, Switzerland.
E-mail address: mariafernandezbravo@gmail.com (M. Fernández-Bravo).

https://doi.org/10.1016/j.apsoil.2026.107549
Received 27 January 2026; Received in revised form 9 October 2026; Accepted 13 October 2026
0929-1393/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

community studies, enabling the analysis of larger sample sets and revealing hidden diversity that traditional identification methods often overlook (Arribas et al., 2016; Braukmann et al., 2019; Porter and Hajibabaei, 2020; Young and Hebert, 2022). We define 'hidden diversity' as evolutionarily distinct lineages that are not yet recognized by current taxonomy. This includes cryptic diversity at the species level, defined as 'two or more distinct species classified as a single species' (Bickford et al., 2007), as well as undescribed species and molecular lineages (e.g., distinct BOLD BINs) that may represent independently evolving lineages and species hypotheses awaiting taxonomic recognition. Metabarcoding approaches are well established for assessing the soil microbial diversity of hundreds of samples simultaneously (Hermans et al., 2025). However, in contrast to soil microorganisms, soil arthropods typically occur at lower densities, are more heterogeneously distributed, and display greater mobility within soil ecosystems (Diekötter et al., 2010; Andújar et al., 2018; Oliverio et al., 2018). Therefore, larger soil volumes, or more commonly, indirect DNA extraction, which is the extraction of arthropod specimens from the soil matrix before DNA metabarcoding, are used to assess arthropod communities in bulk soils (Sapkota et al., 2025; Souto-Vilarós et al., 2025). Several extraction techniques have been developed for this purpose, such as Berlese-Tullgren and Macfadyen-type funnels, which efficiently separate the actively moving mesofauna from soil matrices (Macfadyen, 1953, 1961; Edwards, 1991; Lavelle, 1996). Furthermore, the life cycle of multicellular soil arthropods encompasses eggs, larval stages, and adults, resulting in highly variable body sizes and DNA contents, which poses an additional constraint. The highly variable body sizes, together with technical factors such as DNA extraction efficiency, mitochondrial copy number, and PCR amplification bias, complicate estimating soil arthropod abundances from bulk samples using metabarcoding and basically restrict the analyses to presence/absence assessments (Elbrecht and Leese, 2015).

The Cytochrome C Oxidase Subunit I (COI) sequence has proven to be a powerful DNA barcode for describing the hidden diversity of soil arthropods and detecting intra-specific variation (Andújar et al., 2018; Le Cadre et al., 2024). However, the actual number of hidden soil arthropod taxa remains unknown, with evidence suggesting that a substantial portion of soil biodiversity remains undescribed (Delgado-Baquerizo et al., 2020; Recuero et al., 2024). Metabarcoding studies frequently uncover sequences that cannot be matched to existing databases, underlining the extent of this hidden diversity (Carapelli et al., 2018; Kortmann et al., 2022). Certain groups of arthropods, mainly springtails (Collembola), mites (Acari), and some hyper-diverse insect orders like Coleoptera and Hymenoptera, are believed to harbor a higher proportion of undescribed species due to their small size, high morphological similarity, and soil-dwelling habits (Menta and Remelli, 2020; Simatupang and Rahardjo, 2023). Nevertheless, these taxa play crucial roles in soil ecosystems, which emphasizes the importance of characterizing their communities, ecological preferences, and genetic diversity to enable their protection and ensure sustainable soil management strategies (Remelli et al., 2024).

Land-use intensification is a major factor inducing biodiversity decline, which affects microbial, animal, and plant communities both above and below ground, leading to biotic homogenization (Gossner et al., 2016). While the effects of land-use intensification on soil microbial communities have been well documented, studies of soil arthropods have often focused on specific microarthropod groups, particularly Collembola and Acari, or on selected soil arthropod assemblages (Remelli et al., 2024; Galli et al., 2025). Broader assessments encompassing multiple taxonomic groups remain comparatively limited, although some recent studies have addressed whole-community patterns using broader taxonomic or metabarcoding approaches (Arribas et al., 2021; Hermans et al., 2022). The structure and composition of soil arthropod communities are driven by a combination of abiotic and biotic factors, which include, for instance, soil moisture, temperature, organic matter content, and land management (Blair et al., 1996; Gao et al., 2016). Anthropogenic disturbances, such as pesticide application and habitat fragmentation, can affect arthropod abundance and diversity, where springtails, acari, and ants serve as sensitive bioindicators (Hermans et al., 2022; Tóth et al., 2023). In vineyards, the transition from conventional to organic farming has been associated with increased soil arthropod richness, highlighting the positive role of specific management practices (Ghiglieno et al., 2021). The sensitivity of soil arthropods to disturbances emphasizes their potential as bioindicators for monitoring soil quality and ecosystem resilience (Blair et al., 1996; Remelli et al., 2024). However, the definition of bioindicators relies on the characterization of soil arthropod taxa and their associations with environmental factors, which we aimed to understand in greater depth in the current study.

In this study, we investigated COI-based ASV richness and community composition of soil arthropods at 28 long-term monitoring sites of the Swiss Soil Monitoring Network (NABO), covering the three land-use types (LUTs): arable land, permanent grassland, and forest. To better capture the high diversity of soil arthropod taxa, we applied arthropod-specific COI metabarcoding combined with an indirect DNA extraction, i.e., extraction of living organisms from the soil matrix prior to DNA extraction using a Macfadyen extractor. Our three main research goals were to: i) assess soil arthropod richness and community composition within and among sites and LUTs, ii) correlate soil arthropod richness and community composition with environmental factors, and iii) characterize taxonomic diversity, with a focus on selected taxa with possible hidden diversity.

2. Materials and methods

2.1. Sampling sites and sampling procedures

Samples were collected at 28 long-term monitoring sites of the Swiss soil monitoring network (NABO, www.nabo.ch), which are distributed across the country. These have been monitored for physical and chemical parameters since 1985, and biological parameters since 2012 (Gubler et al., 2015). The sites represent the three LUTs: arable land (9 sites), permanent grassland (9 sites), and forest (10 sites), which dominate and characterize the landscape in the lowlands and pre-alpine areas in Switzerland. Samples were collected after snowmelt and before the first fertilization in 2020, starting in Ticino (the Southern area of Switzerland) at the beginning of February and finishing in Grindelwald (the pre-alpine area) at the beginning of June. The temporally staggered sampling scheme was implemented to harmonize sampling time points, considering the vegetation status of individual sites, thereby minimizing phenological differences among sites despite differences in calendar date.

Two types of soil samples were collected from a georeferenced 10 × 10 m plot at each of the 28 sites: 1) For soil arthropod diversity assessments, seven soil cores of 8 cm diameter × 9 cm depth were taken from three edges of each plot (Supplementary Fig. S1). Soil cores were placed upside down into buckets (8.5 cm diameter × 9 cm depth) with a 7 cm hole at the bottom covered with a 5 mm metal mesh to extract all the living soil arthropods. Extractions were performed using a temperature and moisture gradient Macfadyen extractor (ecoTech, Bonn, Germany) by applying a stepwise increasing temperature gradient (Supplementary Table S1) from 20 to 60 °C over a period of 7 days (Macfadyen, 1953, 1961; Tresch et al., 2019). Arthropod samples were collected and stored in 70% EtOH at 4 °C until further use. Organisms larger than 0.5 cm were cut to this size before DNA extraction to avoid overrepresentation of larger organisms and to improve the detectability of smaller soil arthropods. A total of 196 arthropod samples were obtained from the 28 sites. The applied sampling and extraction approach primarily targets soil mesofauna communities. However, it may also include a broader range of soil-associated arthropods, such as taxa that are only temporarily associated with soil, e.g., during specific life stages, or surface-active organisms.

2) For soil physicochemical and microbial biomass measurements, three composite bulk soil samples were obtained. Each composite sample consisted of 25 bulk soil cores (2.5 cm diameter × 20 cm depth), which were taken evenly distributed within each plot following the standardized sampling protocol of the NABO (Gubler et al., 2015) (Supplementary Fig. S1). In total, 17 environmental factors were evaluated in this study (Supplementary Table S2). Physical, chemical, and biological parameters were measured in each composite sample, as indicated by Gubler et al. (2015) and Hug et al. (2018). Mean annual temperature and mean annual precipitation were determined based on grid data (1 km × 1 km) of the years 2010 to 2019 from MeteoSwiss, which is the Swiss Federal Office of Meteorology and Climatology (http://www.meteoswiss.admin.ch).

2.2. DNA extraction, PCR amplification, and Illumina sequencing

Arthropods stored in 70% EtOH were separated from remaining soil particles that accidentally fell into the collection tubes during the extraction, applying the flotation technique described by Ladell (1936). Briefly, samples were supplemented with a 25% NaCl solution as required, i.e., until arthropods started to float. The supernatant containing the floating arthropods was carefully filtered through 589 Schleicher and Schuell filter paper (Schleicher and Schuell, Dassel, Germany). Subsequently, arthropods retained on the filter paper were transferred to NucleoSpin Bead Tube Type D DNA extraction tubes (Machery & Nagel, Düren, Germany) using a brush. After filtration, arthropods were rinsed with 70% ethanol to remove residual NaCl and subsequently briefly lyophilized to remove remaining ethanol before DNA extraction. DNA was extracted from each lyophilized arthropod sample using the NucleoSpin DNA insect kit (Machery & Nagel, Düren, Germany). DNA concentration was determined using a Cary Eclipse Fluorescence Spectrophotometer (Varian, Inc. Palo Alto, California, USA) with PicoGreen fluorescent nucleic acid stain (Invitrogen, Carlsbad, California, USA).

A 421 bp fragment within the Folmer region of the Cytochrome C Oxidase Subunit I (COI) gene was amplified for metabarcoding using primer BF2 (5′-GCHCCHGAYATRGCHTTYCC-3′) and BR2 (5′-TCDGGRTGNCCRAARAAYCA-3′) (Elbrecht and Leese, 2017). Forward and reverse primers included adapter sequences CS1 (forward) and CS2 (reverse) at the 5′ ends of each corresponding primer to enable the subsequent addition of sample-specific barcodes during a second PCR (Fluidigm, South San Francisco, California, USA). DNA amplification was performed in 25 µl reaction volumes containing 25 ng genomic DNA, 1xGoTaq Flexi Buffer Colorless (Promega, Fitchburg, Wisconsin, USA), 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.5 µM of each primer, 0.6 mg ml^{-1} bovine serum albumin (Sigma-Aldrich, St. Louis, Missouri, USA) and 1.25 U of GoTaq G2 Hotstart DNA Polymerase (Promega). PCR cycling conditions consisted of 2 min denaturation at 95 °C followed by 35 cycles of 40 s denaturation at 94 °C, 40 s annealing at 50 °C, and 1 min elongation at 72 °C. Reactions were completed with a 10 min elongation step at 72 °C. The integrity, quality, and fragment length of the PCR products were verified on an agarose gel. Negative controls without DNA template were included in the first-round PCRs and showed no detectable amplification by gel electrophoresis. Four PCR replicates were conducted for each arthropod sample, which were pooled and sent for sequencing using the Illumina MiSeq platform at the Génome Québec Innovation Centre at McGill University (Montreal, Canada). Sample-specific barcodes were added to the PCR products during a second PCR using Fluidigm Access Array technology, enabling multiplex sequencing. Subsequently, PCR products were purified with AMPure XP beads (Beckman Coulter, Brea, California, USA) and paired-end sequencing was performed using Illumina MiSeq v3 (Illumina Inc., San Diego, California, USA). Raw sequence data are archived in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under the accession number PRJEB105415.

2.3. Sequence processing and taxonomic classification

Sequences were processed and classified using a customized pipeline mostly based on the open-source tool VSEARCH v2.13.3 (Rognes et al., 2016). The workflow consisted of: i) merging paired-end reads using VSEARCH's fastq_mergepairs without length cutoff; ii) removing primer sequences with Cutadapt 2.8 (Martin, 2011) allowing for 2 mismatches; iii) quality filtering with fastq_filter discarding reads with a maximum expected error greater than one; iv) sequence dereplication with derep_fulllength; and v) denoising sequences into amplicon sequence variants (ASVs) using cluster_unoise with minimum cluster size 4, all included in VSEARCH with a concurrent removal of chimeras with uchime3_denovo (Rognes et al., 2016).

For taxonomic classification of ASVs, sequences were compared to a curated COI reference database using the SINTAX algorithm implemented in VSEARCH (Rognes et al., 2016) with a confidence cutoff of 0.8. The COI reference database was constructed from publicly available Barcode of Life Database records from 2021 (BOLD: https://www.boldsystems.org), focusing on arthropod sequences, and processed following a standardized workflow including sequence trimming, length filtering, dereplication, and taxonomic annotation, based on established approaches for COI database curation (O'Rourke et al., 2020). ASVs assigned to the same taxon (identification at the lowest possible taxonomic level) were grouped to represent unique taxa at each taxonomic level (species, genus, family, order, class, or phylum level). ASVs that were not assigned to Arthropoda were removed from the dataset. Samples with a sequencing depth below 5000 reads were discarded using an empirically defined threshold based on the dataset to optimize the number of sequences per sample, while not removing too many samples. The sequencing depth of the remaining samples was standardized by rarefaction, i.e., by randomly subsampling to the lowest remaining number of sequences (6112). While rarefaction may reduce the detection of rare taxa, it was chosen as a conservative approach to standardize sampling effort across samples and avoid biases due to uneven sequencing depth. Due to variation in body size among arthropod species, which results in substantial differences in DNA yield, data were transformed to presence/absence for subsequent analyses (Elbrecht and Leese, 2015).

2.4. Statistical analysis

All analyses performed in R were conducted using R version 4.4.1 (R Core Team, 2024). ASV rarefaction curves, as well as the observed and asymptotic estimation of ASV richness at each site and LUT, were determined using the *iNEXT* R package (Hsieh et al., 2016). The estimated total ASV richness refers to the projected number of ASVs based on rarefaction curves and accumulation analyses. Based on the asymptotic estimation of each LUT, the percentage of observed ASV richness was calculated. The effect of LUT and/or site on soil arthropod ASV richness was analyzed using the non-parametric Kruskal-Wallis rank sum test (Kruskal and Wallis, 1952). The Benjamini & Hochberg adjustment method was applied to compare means of each site or LUT.

Nonmetric Multidimensional Scaling (NMDS) was used to represent soil arthropod community composition among samples and sites of the three LUTs. Analysis of similarity (ANOSIM) was used to estimate similarity of arthropod communities within sites, and among sites for each LUT. PERMANOVA was used to calculate the variation of community composition explained by LUT and site. NMDS ordination, ANOSIM, and PERMANOVA analyses were carried out based on Jaccard dissimilarity matrices using “metaMDS”, “anosim”, and “adonis2” functions, respectively, implemented in the *vegan* R package (Oksanen et al., 2018). Canonical Analysis of Principal Coordinates (CAP) was performed to evaluate the discrimination of LUTs based on soil arthropod community compositions. The analysis was conducted using the “CAPdiscrim” function in *BiodiversityR* package (Kindt and Coe, 2005), which applies linear discriminant analysis on PCoA scores derived from Jaccard

dissimilarities. We allowed the function to select the optimal number of PCoA axes (m), testing up to the default maximum (mmax = 10); the model maximizing cross-validation accuracy used 7 axes. Leave-one-out cross-validation was used to estimate prediction accuracy for LUT classification.

The relationship between environmental factors and soil arthropod communities was assessed using two complementary approaches. First, Spearman correlations were calculated in R to evaluate associations between environmental factors and ASV richness. Second, we employed distance-based linear modeling (DistLM; McArdle and Anderson, 2001) in PRIMER 7 + PERMANOVA (PRIMER-e, Plymouth Marine Laboratory, UK) to analyze how environmental factors correlate with community composition. DistLM allows for partitioning the variation of community composition based on environmental predictors, providing insights into the relative importance of each environmental factor through a marginal test, which assesses the independent contribution of each variable. For the model selection, a forward procedure and the Second-order Akaike Informative Criterion (AICc) were used. Venn diagrams to represent specific and shared ASVs and taxa were created with the function "venn. plot" included in the *VennDiagram* R package (Chen and Boutros, 2011). Analysis of ecological preferences of ASVs for the three LUTs was performed with the function "multipatt" and the association function "r.g" included in the *indicspecies* R package (de Cáceres and Legendre, 2009).

Sequences of ASVs belonging to the genus *Lepidocyrtus* Bourlet (Collembola: Entomobryidae) were phylogenetically analyzed to study the hidden diversity of this genus. ASV sequences were aligned using the ClustalW algorithm implemented in MEGA X (Kumar et al., 2018). Reference sequences were selected based on their availability in published studies and taxonomic reliability, prioritizing sequences with validated species identification (Zhang et al., 2019), and complemented with publicly available BOLD records. Only sequences originating from Europe were included to ensure geographic consistency with the study system. *Entomobrya nivalis* Schäffer (Collembola: Entomobryidae) (GenBank accession no. KM610125.1) was selected as the outgroup. The best-fit nucleotide substitution model was selected based on the lowest Bayesian Information Criterion (BIC) score using the model selection tool in MEGA X (Kumar et al., 2018). Maximum-likelihood (ML) phylogenetic trees were then constructed with 1000 bootstrap replicates. Phylogenetic clusters were conservatively defined based on tree topology, considering only well-supported and clearly separated clades with consistent branching structure, while excluding ambiguous or weakly supported groupings. Due to the high number of ASVs classified as *L. cyaneus* Tullberg or ASVs without species classification, independent ML trees were generated for these two sets of ASVs. A subset of ASVs was then selected to build a phylogenetic tree representing the entire genus. At least four ASVs were selected from each cluster in the *L. cyaneus* and the unclassified *Lepidocyrtus* trees. For the remaining *Lepidocyrtus* species, all ASVs were included in the compiled tree. Tree visualization and annotation were performed in R using the package *ggtree* (Yu et al., 2017).

Pairwise genetic distances were calculated (compiled tree) using the Kimura 2-parameter (K2P) model based on the COI sequence alignment used for phylogenetic reconstruction. Distances were computed among all sequences included in the compiled tree. To quantitatively assess the genetic structure of the phylogenetic clusters, pairwise comparisons were classified as either intra-cluster (between sequences belonging to the same cluster) or inter-cluster (between sequences belonging to different clusters).

3. Results

3.1. Sampling effort, sequencing depth, and sample similarities

A total of 3,989,866 COI sequences passing the quality-filtering were obtained from the 196 samples collected at the 28 sites representing the 3 LUTs: arable land, permanent grassland, and forest. One sample from arable land (site 20) yielded a sequence count below the quality threshold of 5000 sequences and was therefore discarded from further analyses. The final dataset thus included 195 samples, with an average of 20,348.4 ± 294.19 (mean ± SEM) sequences per sample. The sequences clustered into 8976 ASVs, of which 6464 (71.9%) containing 3,552,028 sequences (89.0%) were assigned to the phylum Arthropoda. The sequencing depth of individual samples was standardized by rarefaction (random subsampling to 6112 sequences, the lowest sequencing depth across samples). The transformed ASV table (presence/absence, Supplementary Table S3) was used for all the subsequent analyses. The sampling effort applied resulted in 43.0%, 45.6%, and 49.0% of the estimated maximal number of ASVs in arable land, permanent grassland, and forest, respectively (Supplementary Fig. S2).

3.2. Arthropod ASV richness and community composition

Arthropod ASV richness differed significantly among the three LUTs ($\chi^2 = 10.99$; p-value < 0.01; Kruskal-Wallis test; Fig. 1A). Pairwise tests (Wilcoxon rank sum test) revealed significant differences between the mean ASV richness (mean ± SEM) in arable land (82.7 ± 9.39 ASVs) and in forest (143.1 ± 11.92 ASVs), but not between arable land and permanent grassland (112.5 ± 9.63 ASVs) or permanent grassland and forest (Fig. 1A). Consequently, ASV richness was not LUT-specific, and a linear discriminant analysis in a leave-one-out cross-validation test was only able to predict LUTs with 50.3% accuracy, showing that ASV richness alone is insufficient to reliably distinguish between land-use types. Significant differences in arthropod ASV richness were also detected among sites ($\chi^2 = 112.08$; p-value < 0.0001; Kruskal-Wallis test), with site mean values ranging from 48.6 ± 12.78 to 215.1 ± 8.22 ASVs (mean ± SEM; Fig. 1B, Supplementary Table S4). However, ASV richness showed substantial variability among sites (Fig. 1B, Supplementary Table S4), with multiple significant differences detected among sites within and among LUTs.

To better understand the variation in soil arthropod communities, the composition across LUTs and sites was assessed. Non-metric multidimensional scaling (NMDS) ordination revealed clear clustering patterns among the three LUTs (Fig. 2A). PERMANOVA analysis showed that site-specific differences explained 37.3% of the variation in arthropod community composition (F = 3.65, p-value = 0.0001). In contrast, LUT explained only 4.6% (F = 4.82, p-value = 0.0001), confirming that community composition is more strongly influenced by local environmental conditions (site) than LUT. Canonical Analysis of Principal Coordinates (CAP), based on Jaccard dissimilarities and using 7 PCoA axes, evaluated the predictability of land-use types (LUTs) from community composition. The overall classification success was 89.3%, with high accuracy across LUTs: 100% for forest, 88.9% for arable land, and 77.8% for grassland. Three sites were misclassified, i.e., one arable land site (AL-3) was classified as grassland, and two grassland sites (G-19 and G-2) were classified as arable land and forest, respectively (Fig. 2A).

Forests harbored the highest number of exclusive ASVs (2687; 41.6%, Supplementary Fig. S3), while arable land harbored 1345 (20.8%) and permanent grassland 1281 exclusive ASVs (19.8%). A total of 131 ASVs (2.0%) were shared across all three LUTs, while 545 ASVs (8.4%), 312 ASVs (4.8%), and 163 ASVs (2.5%) were shared between arable land and grasslands, grasslands and forests, and forests and arable land, respectively. Communities of individual samples from the same site were more closely related than those from other sites, which was supported by "analysis of similarity" (ANOSIM) permutation tests of samples to their respective site for each LUT, revealing an R-value of 0.88 (p-value = 0.0001), 0.87 (p-value = 0.001), and 0.94 (p-value = 0.0001) for arable land, permanent grassland, and forest, respectively (Fig. 2B, C, and D). These R-values indicate strong dissimilarities between arthropod communities from different sites within each LUT. The closer the R-value is to 1, the more distinct the communities are between sites. The high R-values for arable land, permanent

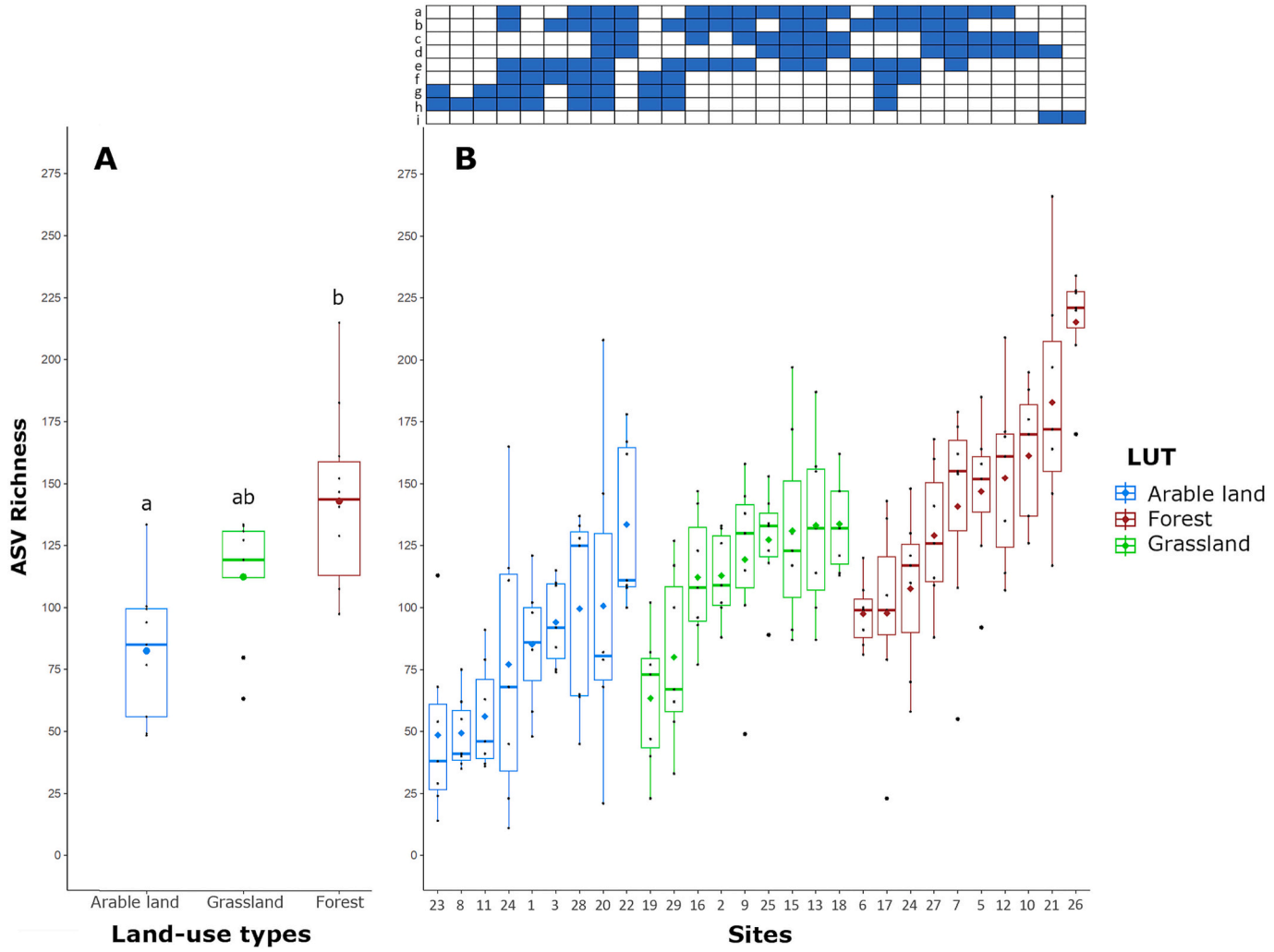


Fig. 1. Boxplots illustrating arthropod amplicon sequence variant (ASV) richness for A) the three land-use types (LUT): arable land (blue), permanent grassland (green), and forest (brown), and B) the 28 sites. The lower and upper boxes represent the first and third quartiles, respectively. The intersection represents the median, and the whiskers represent the ASV richness values within the first and fourth quartile ranges. Black dots represent values of A) the mean ASV richness of each site and B) the ASV richness of each sample per site ($n = 7$). Colored dots (A) and diamonds (B) represent mean ASV richness for each box. The blue grid code in panel B represents statistically significant differences in ASV richness among sites, determined based on pairwise comparisons using the Wilcoxon rank sum test with Benjamini-Hochberg p-value adjustment (p-value < 0.05). Each row in the grid corresponds to a significance class (a to i), grouping sites that do not differ significantly in ASV richness. Mean ± SEM of ASV richness for each site is detailed in Table S4.

grassland, and forest showed that soil arthropod communities exhibit site-specific compositions, with minimal overlap between different locations.

3.3. *Associations of environmental factors with soil arthropod richness and community composition*

Spearman's rank correlation analysis identified significant relationships between soil arthropod richness and four environmental factors, with C:N ratio, organic carbon, and total carbon showing positive correlations, while pH showed a negative correlation (Table 1). However, following Benjamini-Hochberg correction for multiple testing, only the correlations for C:N ratio ($p = 0.038$) and pH ($p = 0.040$) remained significant, whereas organic and total carbon lost statistical significance ($p = 0.135$ in both cases) (Table 1).

Nine out of the 17 evaluated environmental factors significantly explained the variation in soil arthropod community independently ($p < 0.05$, Table 1, DistLM marginal test). These significant factors (C:N ratio, mean annual temperature, metabolic quotient, elevation, bulk density, basal respiration, pH, organic carbon, and total carbon) each had a relatively low individual explanatory power, with explained variation ranging from 4.2% (total carbon) to 7.4% (C:N ratio) (Table 1, DistLM marginal test). Model selection using a forward procedure and the AICc as selection criterion revealed a model including only the C:N ratio ($R^2 = 7.4\%$).

The Kruskal-Wallis test (Supplementary Table S5) revealed significant differences in multiple environmental parameters among the three LUTs, with C:N ratio, pH, basal respiration, and metabolic quotient showing strong differentiation among forest, permanent grassland, and arable land. Specifically, C:N ratio, total carbon, and organic carbon were significantly lower in arable land compared to forest and grassland ($p < 0.05$), while pH was significantly higher in arable land compared to forest and grassland ($p < 0.05$). Bulk density, which negatively correlated with ASV richness, was significantly higher in arable land compared to forest and grassland. Additionally, mean annual temperature and mean annual precipitation also differed among LUTs, with arable land showing the lowest median values ($p < 0.05$).

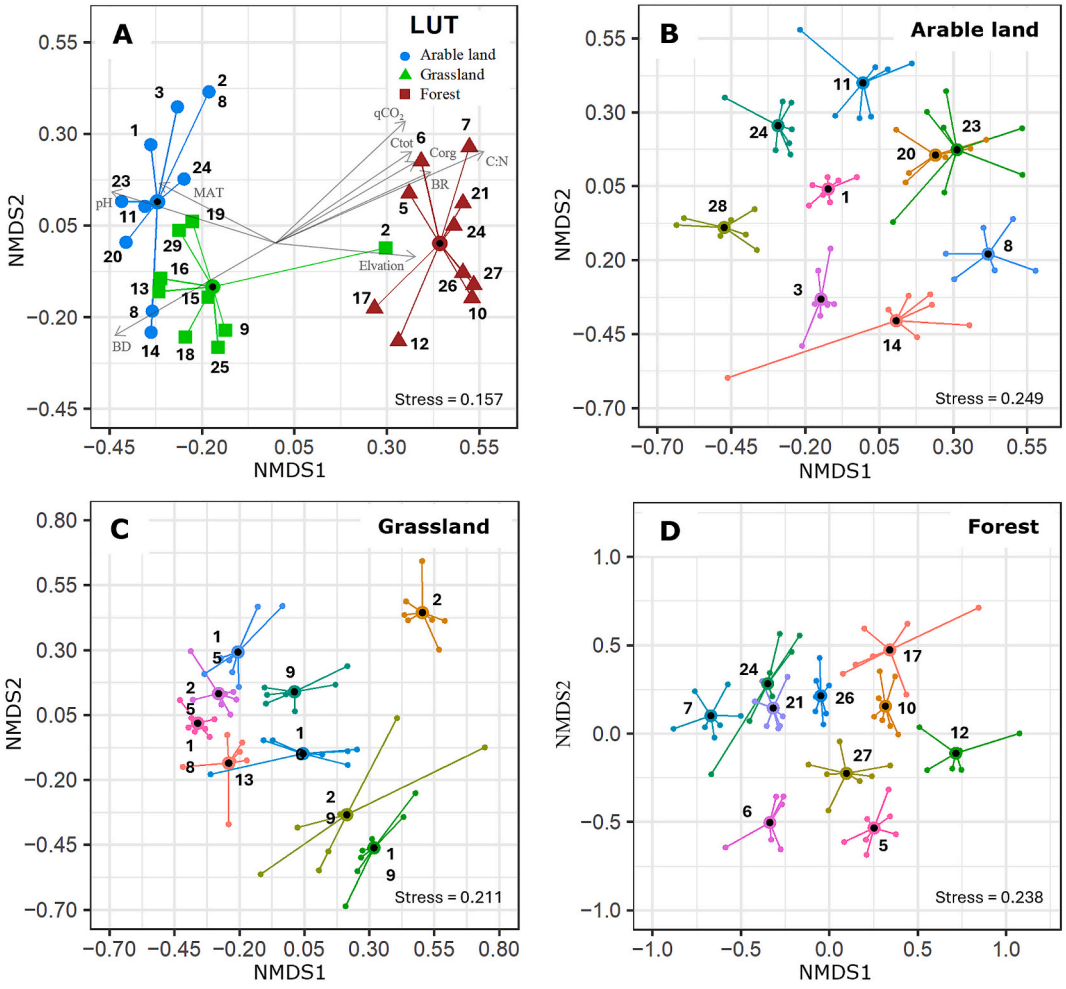


Fig. 2. Non-metric multidimensional scaling (NMDS) ordination based on binary (presence/absence) ASV data using a Jaccard dissimilarity matrix representing arthropod community composition: A) of each site of the three land-use types (LUTs); and the arthropod community composition of each sample and site for each LUT, i.e., B) arable land, C) grassland, and D) forest. In panel A, circles represent arable land sites, squares represent grassland sites, and triangles represent forest sites (site numbers indicated). Arrows indicate correlations of significant environmental factors to community compositions from Table 1. In panels B, C, and D, samples are shown as dots colored according to the sites (site numbers indicated). Environmental vectors are shown only for the NMDS including all sites (A). They were not fitted separately within individual LUTs (B–D) because of the limited number of sites and to avoid overinterpretation. Centroids in panels A–D are represented by circles with a black filling colored according to the LUT or the site. Lines connect samples of each land-use type (A) or site (B–D) with their corresponding calculated land-use type or site centroid.

3.4. Associations of ASVs to LUTs

Indicator analysis identified 69 ASVs as indicators for one of the three LUTs with a fidelity (phi) value above 0.6 (p-value < 0.01) (Supplementary Table S6). One indicator ASV was identified for arable land, whereas 14 and 54 were detected for permanent grassland and forest, respectively. In addition, 5 indicator ASVs were related to arable land and permanent grassland and 10 to permanent grassland and forest (Supplementary Table S6). Twelve of the fourteen grassland-indicator ASVs were assigned to Collembola, and more specifically nine of these were classified as the species *Lepidocyrtus cyaneus* (Collembola: Entomobryidae). In contrast, forest indicator ASVs included 25 ASVs classified to Arachnida, 20 to Collembola, 7 to Chilopoda, and 1 each to Insecta and Symphyla. Among the 25 ASVs classified to Arachnida, 12 were assigned to Sarcoptiformes, and 11 to Trombidiformes. Lower-level taxonomic classification failed in many cases, as 17 of the 25 ASVs remained unclassified at the genus level. All indicator ASVs of Collembola were classified to the genus level, including 16 assigned to the species *Parisotoma notabilis* Schäffer (Collembola: Isotomidae), as well as two assigned to the genus *Lepidocyrtus*.

3.5. Taxonomic resolution and hidden diversity

Of the 6464 arthropod ASVs, 4835 (74.8%) were classified to a lower level, i.e., to 10 classes, 29 orders, 136 families, 244 genera, or 214 species (Table 2, Supplementary Table S7). Collembola was the dominant class with 2137 ASVs, while Protura was represented by a single ASV (Table 2). A total of 389 unique taxa were identified (Supplementary Table S7). Among these, 125 taxa (32.1%) were represented by a single ASV, 253 taxa (65.0%) were represented by 2 to 98 ASVs, and eleven taxa (0.03%) were represented by more than 100 ASVs (5 Collembola, 2 Arachnida, 2 Insecta, 1 Chilopoda, and 1 unclassified arthropod). The three species-level taxa with the highest numbers of ASVs were *L. cyaneus* Tullberg (Collembola: Entomobryidae) with 180 ASVs, *Folsomia candida* Willem (Collembola: Isotomidae) with 149 ASVs, and *Parisotoma notabilis* with 130 ASVs (Supplementary Table S7).

Due to the large number of ASVs assigned to the species *L. cyaneus*, we selected the higher-ranking taxon, i.e., the genus *Lepidocyrtus* as an example to explore potential hidden diversity. In total, four species were identified within this genus, i.e., *L. cyaneus* (180 ASVs), *L. lanuginosus* Gmelin (7 ASVs), *L. lignorum* Fabricius (2 ASVs), and *L. ruber* Schott (1 ASV). In addition, 42 ASVs were only classified at the genus level. ASVs

Table 1
Correlations of environmental factors with ASV richness and community composition. Distance-based linear modeling (DistLM) marginal test was used to associate the 17 environmental factors with the variation of soil arthropod communities. Analyses were based on dissimilarities among the site centroids previously calculated from a Jaccard dissimilarity matrix. Environmental factors were ordered by decreasing percentage of explained variation of the marginal test of the DISTLM.

Environmental factor[a]	ASV richness Spearman correlation			Community composition DistLM (Marginal test)		
	rho	p-Value	p-Value after BH correction	Pseudo-F	p-Value by permutation	Explained variation [%]
C:N ratio	0.55	0.002	0.038*	2.1	0.001*	7.4
Mean annual temperature (°C)	−0.31	0.106	0.224	1.7	0.001*	6.1
qCO_2	0.29	0.133	0.251	1.7	0.002*	6.1
Elevation (m.a.s.l.)	0.33	0.082	0.200	1.7	0.004*	6.0
Bulk density (gDM/cm^3)	−0.34	0.074	0.200	1.6	0.002*	5.7
Basal respiration (mg CO^2-C kg^{-1} h^{-1})	0.36	0.057	0.194	1.5	0.002*	5.6
pH ($CaCl_2$)	−0.52	0.005	0.040*	1.3	0.026*	4.8
C_{org} (% [w./w.])	0.41	0.032	0.135	1.3	0.036*	4.7
C_{tot} (% [w./w.])	0.42	0.027	0.135	1.3	0.043*	4.7
FE_C (mg Cmic kg^{-1})	0.17	0.376	0.639	1.1	0.124	4.2
Soil skeleton (Vol%)	0.06	0.768	0.935	1.1	0.184	4.1
FE_N (mg Nmic kg^{-1})	0.07	0.734	0.935	1.1	0.216	4.0
Silt (% [w./w.])	−0.05	0.809	0.935	1.1	0.238	4.0
Sand (% [w./w.])	0.02	0.935	0.935	1.0	0.405	3.7
Mean annual precipitation (mm)	−0.03	0.892	0.935	1.0	0.445	3.7
N_{tot} (% [w./w.])	0.03	0.874	0.935	1.0	0.500	3.6
Clay (% [w./w.])	0.09	0.666	0.935	0.9	0.733	3.4

[a] C:N: ratio from C_{org}/N_{tot}; qCO_2: metabolic quotient (Basal respiration/FE_C); m.a.s.l.: Elevation in meters above sea level; DM: soil dry matter; C_{org}: Organic carbon; C_{tot}: Total carbon; FE_C: microbial biomass C (Chloroform-Fumigation-Extraction); Soil skeleton: Volume % of soil skeleton; FE_N: microbial biomass N (Chloroform-Fumigation-Extraction). *Indicates statistical significance ($p < 0.05$).

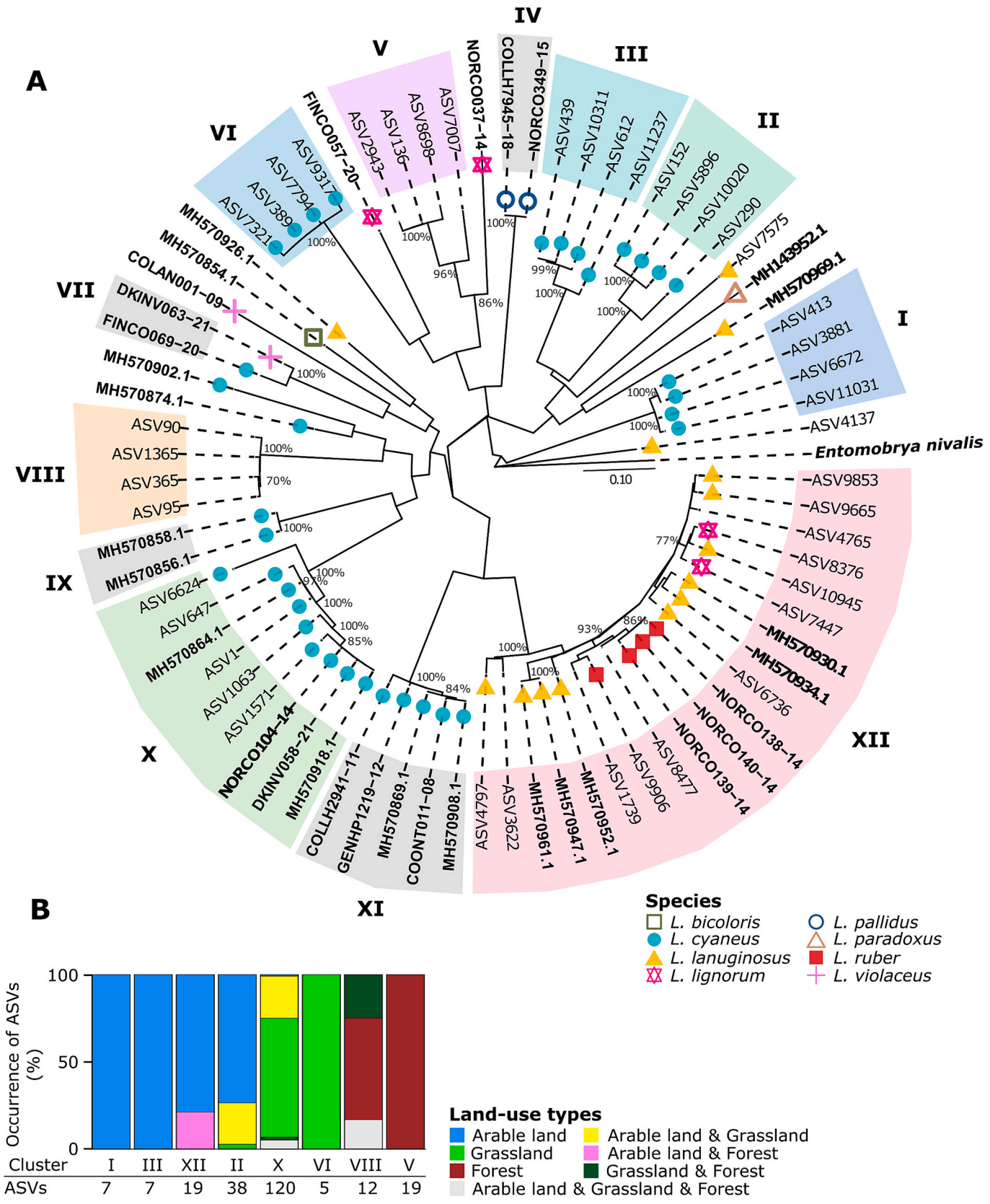
Table 2
Number of arthropod amplicon sequence variants (ASVs) assigned to taxa at different taxonomic levels. The table shows the distribution of unique taxa and ASVs within each arthropod class across different taxonomic levels (order, family, genus, and species). Classes were ordered by decreasing number of total assigned ASVs.

Class	Total assigned ASVs at Class level[a]	Order			Family			Genus			Species		
		No.[b]	Assigned ASVs	Unassigned ASVs	No.[b]	Assigned ASVs	Unassigned ASVs	No.[b]	Assigned ASVs	Unassigned ASVs	No.[b]	Assigned ASVs	Unassigned ASVs
Collembola	2137	4	2119	18	16	2001	136	46	1500	637	48	1075	1044
Insecta	1156	7	994	162	46	970	186	113	724	432	105	644	350
Arachnida	1078	5	1017	61	61	832	246	70	442	636	57	291	726
Chilopoda	218	2	217	1	5	98	120	5	98	120	11	98	119
Diplopoda	162	5	162	0	4	149	13	7	88	74	8	87	75
Symphyla	61	1	61	0	1	34	27	1	33	28	1	32	29
Diplura	16	1	14	2	1	13	3	1	10	6	1	10	4
Malacostraca	3	1	3	0	1	2	1	1	2	1	1	2	1
Ostracoda	3	1	3	0	0	0	3	0	0	3	0	0	3
Protura	1	1	1	0	1	1	0	0	0	1	0	0	1
Total	4835	29	4591	244	136	4100	735	244	2897	1938	214	2239	2352

[a] Values in the first column represent all ASVs assigned to each class. The following columns indicate how many of those ASVs could be further classified into orders, families, genera, or species within each class.
[b] No. = number of taxa (e.g., number of orders, families, genus, or species).

assigned to *L. cyaneus* were grouped into five phylogenetic clusters (Supplementary Fig. S4), while three clusters were identified among the unclassified *Lepidocyrtus*-ASVs (Supplementary Fig. S5). Combining four representative ASVs of each of these clusters with all ASVs assigned to the additional three species (*L. lanuginosus*, *L. lignorum* and *L. ruber*) and a reference set of identified sequences provided an overview of *Lepidocyrtus* diversity with twelve distinct clusters (Fig. 3A). Four of these, i.e., cluster VII, IX, X, and XI, included reference sequences assigned to *L. cyaneus*, and four additional clusters (I, II, III, and VI) were composed of ASVs assigned to *L. cyaneus* indicating the hidden diversity within this species. All ASVs that remained unclassified at the species level were included in the three phylogenetic clusters V, VIII, and XII. Eight of the twelve clusters corresponded to a single BOLD Barcode Index Number (BIN), three clusters to two BINs, and one cluster (XII) was associated with seven BINs (Supplementary Table S10). Cluster XII contained most *L. lanuginosus* reference sequences, five of the seven ASVs identified as *L. lanuginosus*, and two ASVs classified as *L. lignorum*, as well as a well-supported sub-cluster (bootstrap 93%) comprising both *L. ruber* reference sequences and the single ASV assigned to *L. ruber*. This composition indicates that cluster XII includes sequences assigned to multiple species and exhibits a complex internal structure compared to other clusters. To quantitatively support the phylogenetic clustering, pairwise genetic distances (K2P) were calculated and compared within and between clusters of the compiled tree in Fig. 3 (Supplementary Tables S8, S9). Overall, intra-cluster distances ($K2P < 0.1413$) were lower than inter-cluster ($K2P > 0.1844$) distances, indicating a clear genetic structure among clusters.

The LUT preference of the eight phylogenetic clusters containing ASVs from this study was identified based on the ASV occurrences. Of the five clusters containing ASVs assigned to *L. cyaneus*, two were composed of ASVs detected only in arable land (clusters I and III), one was composed of ASVs detected only in permanent grassland (cluster VI), and two were composed of ASVs predominantly found in either arable land (cluster II) or permanent grassland (cluster X) (Fig. 3B). Similarly, 84% of ASVs included in cluster XII combining *L. ruber*, most *L. lanuginosus* reference sequences and ASVs assigned to the two species



(caption on next page)

Fig. 3. Phylogenetic tree (A) and land-use type (LUT) distribution of *Lepidocyrtus* sequence clusters (B). A) Phylogenetic tree of ASVs and reference sequences of the genus *Lepidocyrtus*. The tree includes four representative ASVs from each cluster identified in the supplementary phylogenetic trees for *L. cyaneus* and ASVs identified as *Lepidocyrtus* without species assignment (Figs. S4 and S5) and ASVs assigned to *L. ruber* (1 ASV), *L. lanuginosus* (7 ASVs), and *L. lignorum* (2 ASVs). Reference sequences were retrieved from the BOLD database and NCBI (Zhang et al., 2019). Reference sequences are displayed in bold, and bootstrap support values >70% are indicated at the nodes. For ASVs with species-level taxonomic assignments, species names are indicated by different symbols and colors. No symbols are shown for ASVs classified only to genus level (*Lepidocyrtus* spp.). Species symbols therefore reflect the taxonomic assignments from the metabarcoding analysis and were not inferred from phylogenetic placement. The phylogeny was rooted with *Entomobrya nivalis* (Collembola: Entomobryidae; GenBank accession no. KM610125.1) as an outgroup. Twelve clusters with 100% bootstrap support were defined (I-XII). Clusters containing only reference sequences are shown with a grey background, whereas clusters containing ASVs (alone or mixed with references) are colored. Differences in color intensity (including paler shades) are used solely to visually distinguish clusters and do not represent additional biological information. B) Stacked bar plots show the relative proportions of ASVs in each cluster occurring at different land-use types (LUTs). Of the 232 ASVs assigned to the genus *Lepidocyrtus*, 230 were included, while two were excluded due to their short sequence length. Colors correspond to LUTs and LUT combinations as indicated.

as well as two ASVs classified as *L. lignorum* were detected mainly in arable land. Unclassified ASVs grouped into clusters V and VIII were detected exclusively or predominantly in forest soils, respectively. Therefore, ASVs assigned to the genus *Lepidocyrtus* were grouped into phylogenetic clusters, each with a specific LUT preference covering all three studied LUTs.

In addition to *Lepidocyrtus*, phylogenetic analyses were performed for the genera including the other two collembola species with more than 100 ASVs, i.e., *Folsomia* and *Parisotoma*. For *Folsomia* ASVs assigned to the three species *F. candida*, *F. quadrioculata*, and *F. sensibilis* were grouped into 10 clusters matching 13 BINs (Supplementary Figs. S6–S9, Supplementary Table S10). However, clusters VII, assigned to *F. quadrioculata*, and X, assigned to *F. sensibilis*, revealed a sequence identity of only 96.9% to BIN ABA1825 and BIN ACT9947, respectively, indicating a potential for so far unrecognized genetic lineages (Supplementary Table S10). *Parisotoma* sequences were classified into four phylogenetic clusters, which were all identified as *P. notabilis*, each representing a previously defined lineage (Supplementary Fig. S10, Supplementary Table S10). Clusters were associated with 2 to 4 BINs, indicating the existence of further genetic resolution.

4. Discussion

4.1. Associations of soil arthropod diversity to land-use types, sites, and environmental factors

To investigate soil arthropod diversity across Switzerland and to create a first reference set of their diversity at the Swiss long-term soil quality monitoring sites, we sampled 28 sites across the three different land-use types: arable land, permanent grassland, and forest, using arthropod-specific metabarcoding. Results revealed significant differences in soil arthropod richness and community composition among LUTs, with a gradual decline in ASV richness and number of land-use indicative ASVs from forest to permanent grassland and arable land. These patterns are in agreement with trends observed in studies comparing soil arthropod diversity across different LUTs, whether assessed through metabarcoding (Arribas et al., 2021) or morphological identification (Caruso et al., 2012; Wale and Yesuf, 2022), highlighting the robustness of this pattern across different methodological approaches. Wale and Yesuf (2022), who reported highest micro- and macro-arthropods abundance in Ethiopian forests as compared to permanent grasslands and cultivated areas, attributed this pattern to the greater structural complexity and stability of the litter layer in forests, which provides habitat and food resources for diverse arthropod groups. This explanation is also consistent with earlier reports showing that well-developed litter layers buffer microclimatic fluctuations and enhance arthropod diversity in forest ecosystems (Ashford et al., 2013).

Soil arthropod community compositions were most distinct between the forest and the other two land-use types, permanent grassland and arable land, which confirms previous studies showing stronger contrasts between forest and open habitats than between different agricultural systems (Arribas et al., 2021; Hermans et al., 2022; Galli et al., 2025). The composition of soil arthropod communities also differed significantly among sites, with site-specific factors explaining a higher proportion of variation (37.3%) than LUTs (4.6%). In agreement with our results, Arribas et al. (2021) and Zizka et al. (2020) found that soil arthropod communities were more similar at local scales (comparable to sites in our study) than at regional scales (between sites). Similarly, other authors have documented substantial variation in community composition not only among ecoregions, which are large ecological areas defined by climate, vegetation, and geography (Steinke et al., 2022), but also within the same LUT, for instance, among different forest stands (Hermans et al., 2022). Overall, the pattern observed in our study is consistent with evidence from other studies demonstrating that environmental heterogeneity and microhabitat structure shape mesofauna assemblages in combination with land-use related factors (Caruso et al., 2019; Zinger et al., 2019; Rota et al., 2020).

In our study, nine out of the seventeen environmental factors significantly explained differences in community composition, and among these C:N ratio, mean annual temperature, and the metabolic quotient were the three strongest drivers (Table 1; DistLM marginal test). The attempt to model arthropod community composition revealed C:N ratio as the only significant factor with an explained variation of 7.4%, and the inclusion of further factors did not add any significant portions of explained variation. This is similar to the results reported by Remelli et al. (2024), who also identified the C:N ratio, but in addition, soil organic matter content as key soil properties correlating to mesofaunal diversity across management regimes. Nevertheless, a substantial portion of variation remains unexplained, possibly reflecting effects related to other abiotic and biotic factors such as microhabitat configuration (Arribas et al., 2021), and biotic interactions with plants, animals, and other organisms (Tresch et al., 2019; Delgado-Baquerizo et al., 2020). However, sampling effort may not fully capture total species richness (Supplementary Fig. S2), although it remains sufficient to identify consistent ecological patterns across land-use types.

4.2. High-level taxonomic overview

Collembola represented the largest proportion of ASVs (33%), followed by Insecta with 17.9% containing notably Coleoptera and Diptera, and Arachnida with 16.7% containing notably the acarian orders Sarcoptiformes, Trombidiformes, and Mesostigmata (Table 2, Supplementary Table S7). Interestingly, the number of identified insect species was approximately double the number of identified Collembola and Arachnida species, most likely indicating an underrepresentation of the latter two groups (Table 2). Although insects were among the most dominantly detected groups, it is important to note that many of the insect taxa detected are not strictly permanent soil dwellers, but use the soil only during specific life stages, such as Tipulidae larvae (Diptera) or pupae of some Lepidoptera. Further, insect taxa are primarily surface- or litter-dwelling, such as Carabidae and Staphylinidae (Coleoptera), and are therefore associated with the soil only as surface-active fauna rather than true endogeic organisms (Andújar et al., 2018; Oliverio et al., 2018). In contrast, Collembola and Acari, which form the core of the detected soil mesofauna, are mostly soil-dwelling, and their large diversity and dominance are consistently reported across ecosystems and

methodologies, including both morphological and molecular approaches (Arribas et al., 2016; McGee et al., 2020; Hermans et al., 2022). These two groups represent the backbone of below-ground biodiversity studies and are among the best-characterized components of the soil fauna. However, recent molecular research has uncovered a remarkable level of hidden diversity and unresolved taxonomy, suggesting that their apparent dominance may conceal substantial genetic and ecological differentiation at finer scales (Porco et al., 2012; Mateos and Álvarez-Presas, 2022).

Other arthropod classes detected in this study were generally less prevalent but still widely distributed across samples. Diplopoda, Chilopoda, and Symphyla together represented approximately 10% of the total number of ASVs, with Symphyla being the least frequent and mainly occurring in forest soils. Their low representation may reflect both biological sensitivity to soil disturbance, as they typically inhabit undisturbed, organic-rich forest soils, and technical under-detection due to incomplete reference databases and low amplification efficiency (Bedano et al., 2006; Menta and Remelli, 2020; Remelli et al., 2024). Protura, Malacostraca, and Ostracoda exhibited even lower ASV counts; notably, only a single ASV was assigned to Protura, which is consistent with previous metabarcoding surveys where these small, poorly represented taxa are rarely recovered (Andújar et al., 2018; Oliverio et al., 2018; Sahdra et al., 2025). Morphology-based inventories often recover basal hexapod groups such as Diplura, Protura and Pauropoda, which remain poorly represented or difficult to detect in DNA-based datasets due to reference library gaps and methodological biases such as preferential amplification of certain target groups (Arribas et al., 2016; Oliverio et al., 2018). To avoid this constraint, we applied primers BF2/BR2 that have been designed to target a wide range of arthropod groups (Elbrecht and Leese, 2017), which was confirmed by an in silico PCR analysis (Supplementary Analysis S1). Our results underline the importance of integrative approaches combining molecular and morphological data, and of expanding reference databases with curated sequences from underrepresented taxa. Improving coverage for groups such as Protura, Symphyla, and Diplura will enhance the accuracy of DNA-based identifications and allow a more complete assessment of soil arthropod biodiversity across LUTs. DNA metabarcoding is a destructive approach, precluding morphological validation of the analyzed material. This restrains the use of DNA metabarcoding for improvement of coverage or expansion of reference databases. Efforts to increase database coverage mostly must rely on sequence data obtained from morphologically classified individuals.

4.3. Low-level taxonomic analysis: indications of hidden diversity

Lepidocyrtus is a cosmopolitan collembolan genus comprising nearly 250 described species, about 35 of which have been reported in Europe (Ratnasingham and Hebert, 2007; Mateos et al., 2021; Bellinger et al., 2025). However, public COI barcodes are currently available for only 30 species of the genus *Lepidocyrtus*, and several of these records are linked to uncertain or unverifiable identifications (Ratnasingham and Hebert, 2007; Bellinger et al., 2025). The observed association between phylogenetic structure and land-use type distribution suggests ecological differentiation among *Lepidocyrtus* lineages (Fig. 3; Supplementary Figs. S4–S5). Clusters restricted to a single LUT may reflect habitat-associated genetic lineages, whereas clusters occurring across multiple LUTs may indicate more generalist taxa or the presence of several closely related genetic lineages with broader ecological distributions.

Species-level interpretation of the phylogenetic clusters in *Lepidocyrtus* remained difficult, as cluster IV, containing two reference sequences of *L. pallidus* Reuter, was the only cluster exclusively represented by a single species. In contrast, other clusters either included sequences assigned to several species, or sequences assigned to the same nominal species were distributed across several clusters. This was particularly evident for *L. cyaneus*, for which ASVs and reference sequences occurred across eight phylogenetically distinct clusters. Similarly, the correspondence with BOLD BINs did not resolve the ambiguities related to species identification (Supplementary Table S10). This pattern suggests that the taxon currently identified as *L. cyaneus* represents a species complex composed of multiple phylogenetically and ecologically distinct lineages, which is in agreement with previous multi-locus studies (Winkler et al., 2020; Mateos et al., 2021). A similarly unclear taxonomic situation was observed for *L. lanuginosus*, for which two reference sequences, MH570926.1 and MH570969.1, both originating from the same study (Zhang et al., 2019) clustered apart from the main *L. lanuginosus* clade (cluster XII). In addition, ASV4137 and ASV7575, both classified as *L. lanuginosus*, fell into different phylogenetic branches, supporting the hypothesis that the species may be non-monophyletic. Interestingly, cluster XII, which contained most *L. lanuginosus* references, also included a strongly supported subclade of *L. ruber* reference sequences together with three of our ASVs (bootstrap 93%), suggesting possible historical misidentifications or incomplete lineage sorting between both taxa. Corresponding to our analyses, BIN ACG6072, which includes *L. ruber* reference sequences, does also include records annotated as *L. lignorum*. Together, this indicates presence of unresolved species boundaries within this COI-marker-based lineage complex. Cluster XII should therefore be interpreted as a broader mitochondrial lineage complex rather than a species-level unit. Detailed further analyses including additional molecular markers will be required to resolve the relationships among these nominal species. Sequences assigned to *L. lignorum* were scattered across the phylogenetic tree, adding to the inconsistency between the phylogenetic and taxonomic groupings. Recent molecular analyses of *L. ruber*, *L. lanuginosus*, and *L. lignorum* in Central Europe already suggested that their boundaries are blurred and that the three species may comprise several geographically or ecologically distinct entities (Winkler et al., 2020; Mateos and Álvarez-Presas, 2022). Similar inconsistencies between phylogenetic clustering and taxonomic assignments have been reported in previous studies, which describe non-monophyly, cryptic diversity, and unclear species delimitation within these taxa (Winkler et al., 2020; Mateos et al., 2021). Rather than contradicting the phylogenetic clustering approach, these results highlight the limitations of current taxonomic frameworks and reference databases for *Lepidocyrtus* and support the presence of hidden diversity within this genus.

In addition to the ASVs assigned to the four species *L. cyaneus*, *L. lanuginosus*, *L. ruber*, and *L. lignorum* discussed above, our dataset included two clusters (V and VIII) composed exclusively of unclassified ASVs. However, each cluster matched to a single BIN, i.e., cluster V to BIN AAI1222 identified as *Lepidocyrtus* sp. and cluster VIII to BIN AAI1224 identified as *L. lignorum* (Supplementary Table S10), further illustrating the incomplete correspondence between taxonomic assignments and molecular reference data. More generally, for ASVs lacking both a reliable taxonomic assignment and correspondence to an existing BOLD BIN, metabarcoding data alone do not allow us to distinguish between known but unbarcoded species, undescribed species, and previously unrecognized molecular lineages. Such ASVs should therefore be regarded as unresolved molecular diversity rather than evidence of novel species per se. Resolving their taxonomic status would require the recovery and morphological identification of corresponding specimens together with additional molecular evidence.

Besides *L. cyaneus*, two species were found to include more than 100 ASVs, i.e., *F. candida* with 149 ASVs, and *P. notabilis* with 130 ASVs. Analyses of ASVs assigned to *P. notabilis* indicated the existence of phylogenetic clusters corresponding to the well-described, intraspecific lineages L1, L2 and L4 (Porco et al., 2012). For *F. candida*, comparable evidence from natural soil populations is currently lacking.

Our results clearly illustrate the difficulty of species-level assignments in metabarcoding analyses when reference databases are incomplete or contain mislabeled sequences (Arribas et al., 2016; Porter and Hajibabaei, 2020). Our results also revealed cases where phylogenetic clusters reflect higher-level relationships rather than species-level

resolution. For instance, cluster XII grouped sequences assigned to *L. lanuginosus*, *L. ruber*, and *L. lignorum*, indicating that this cluster corresponds to a broader clade rather than a single taxonomic unit, as further supported by the overlap between intra- and inter-cluster K2P distances. On the one hand, application of longer or concatenated markers using Sanger sequencing or long-read high-throughput sequencing technologies could help resolve the narrow taxonomic boundaries within *Lepidocyrtus* and other unresolved genera and improve DNA barcoding databases (Winkler et al., 2020; Mateos et al., 2021). On the other hand, arthropod-specific COI metabarcoding coupled with phylogenetic analyses, even with short amplicons and under unresolved taxonomy, can effectively reveal the presence of hidden diversity in complex mesofauna communities.

5. Conclusions

Using arthropod-specific COI metabarcoding, we provide the first comprehensive overview of soil arthropod diversity across contrasting land-use types at Swiss soil quality monitoring sites. Our results revealed consistent patterns of decreasing richness and compositional complexity from forest to permanent grassland and arable land, confirming that forest ecosystems support the most diverse and distinct communities of soil arthropods. At the same time, site-specific factors explained a substantially larger share of community variation than land use, emphasizing the dominant role of local heterogeneity, microhabitat structure, and possibly stochastic processes in shaping soil mesofauna. Further, the C:N ratio was identified as the most important environmental factor explaining arthropod communities across land-use types. Across all sites, Collembola were the dominant and most diverse component of the soil mesofauna. Within Collembola, the genera *Lepidocyrtus*, *Folsomia* and *Parisotoma* were used to exemplify the extent of hidden diversity, revealed even with short amplicon sequencing. The genus *Lepidocyrtus* included several phylogenetic clusters with unresolved taxonomy, which were linked to specific land-use types, suggesting a clear habitat-associated differentiation among genetically distinct groups and ecological specialization within soil environments. Continued expansion of curated reference databases and integration of multilocus and morphological data will increase the resolution and reliability of taxonomic assignments and the value of metabarcoding for long-term soil biodiversity monitoring.

CRediT authorship contribution statement

María Fernández-Bravo: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Florian Gschwend:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Johanna Mayerhofer:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Janine Moll-Mielewczik:** Resources. **Franco Widmer:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Jürg Enkerli:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement no. 794526 as well as the Swiss Federal Office for the Environment (FOEN) within the frame of the project MiDiBo2. We thank the NABO field team at Agroscope for adapting their soil sampling campaign and providing the samples required for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apsoil.2026.107549.

Data availability

Raw DNA sequence data are archived in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under the accession number PRJEB105415. Metadata of the Swiss Soil Monitoring Network is available upon request by signing a usage agreement (www.nabo.ch/).

References

Andújar, C., Arribas, P., López, H., Arjona, Y., Pérez-Delgado, A., Oromí, P., Vogler, A.P., Emerson, B.C., 2022. Community assembly and metaphylogeography of soil biodiversity: insights from haplotype-level community DNA metabarcoding within an oceanic island. Mol. Ecol. 31, 4078–4094. https://doi.org/10.1111/mec.16560.

Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P., Emerson, B.C., 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Mol. Ecol. 27, 3968–3975. https://doi.org/10.1111/mec.14844.

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., Vogler, A.P., 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods Ecol. Evol. 7, 1071–1081. https://doi.org/10.1111/2041-210X.12557.

Arribas, P., Andújar, C., Salces-Castellano, A., Emerson, B.C., Vogler, A.P., 2021. The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. Mol. Ecol. 30, 48–61. https://doi.org/10.1111/mec.15591.

Ashford, O.S., Foster, W.A., Turner, B.L., Sayer, E.J., Sutcliffe, L., Tanner, E.V.J., 2013. Litter manipulation and the soil arthropod community in a lowland tropical rainforest. Soil Biol. Biochem. 62, 5–12. https://doi.org/10.1016/j.soilbio.2013.03.001.

Bedano, J.C., Cantú, M.P., Doucet, M.E., 2006. Soil springtails (Hexapoda: Collembola), symphylans and pauropods (Arthropoda: Myriapoda) under different management systems in agroecosystems of the subhumid Pampa (Argentina). Eur. J. Soil Biol. 42, 107–119. https://doi.org/10.1016/j.ejsobi.2005.11.004.

Bellinger, P., Christiansen, K., Janssens, F., 2025. Collembola.org: checklist of the collembola of the world. https://www.collembola.org. (Accessed December 2025).

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155. https://doi.org/10.1016/j.tree.2006.11.004.

Blair, J.M., Bohlen, P.J., Freckman, D.W., 1996. Soil invertebrates as indicators of soil quality. In: Dran, J.W., Jones, A.J. (Eds.), Methods for Assessing Soil, pp. 273–291. https://doi.org/10.2136/sssaspecpub49.c16.

Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E.V., Hebert, P.D.N., 2019. Metabarcoding a diverse arthropod mock community. Mol. Ecol. Resour. 19, 711–727. https://doi.org/10.1111/1755-0998.13008.

Carapelli, A., Soltani, A., Leo, C., Vitale, M., Amri, M., Jemàa, J.M. Ben, 2018. Cryptic diversity hidden within the leafminer genus *Liriomyza* (Diptera: Agromyzidae). Genes 9. https://doi.org/10.3390/genes9110554.

Caruso, T., Schaefer, I., Monson, F., Keith, A.M., 2019. Oribatid mites show how climate and latitudinal gradients in organic matter can drive large-scale biodiversity patterns of soil communities. J. Biogeogr. 46, 611–620. https://doi.org/10.1111/jbi.13501.

Caruso, T., Taormina, M., Migliorini, M., 2012. Relative role of deterministic and stochastic determinants of soil animal community: a spatially explicit analysis of oribatid mites. J. Anim. Ecol. 81, 214–221. https://doi.org/10.1111/j.1365-2656.2011.01886.x.

Chen, H., Boutros, P.C., 2011. *VennDiagram*: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35 [software]. https://doi.org/10.1186/1471-2105-12-35.

Cook, L.G., Edwards, R.D., Crisp, M.D., Hardy, N.B., 2010. Need morphology always be required for new species descriptions? Invertebr. Syst. 24, 322–326. https://doi.org/10.1071/IS10011.

Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. Trends Ecol. Evol. 29, 566–571. https://doi.org/10.1016/j.tree.2014.08.001.

de Cáceres, M., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90, 3566–3574. https://doi.org/10.1890/08-1823.1.

Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., Eldridge, D.J., Abades, S., Alfaro, F.D., Bastida, F., Berhe, A.A., Cutler, N.A., Gallardo, A., García-Velázquez, L., Hart, S.C., Hayes, P.E., He, J.-Z., Hseu, Z.-Y., Hu, H.-W., Kirchmair, M., Neuhauser, S., Pérez, C. A., Reed, S.C., Santos, F., Sullivan, B.W., Trivedi, P., Wang, J.-T., Weber-Grullon, L., Williams, M.A., Singh, B.K., 2020. Multiple elements of soil biodiversity drive ecosystem functions across biomes. Nat. Ecol. Evol. 4, 210–220. https://doi.org/10.1038/s41559-019-1084-y.

Diekötter, T., Wamser, S., Wolters, V., Birkhofer, K., 2010. Landscape and management effects on structure and function of soil arthropod communities in winter wheat. Agric. Ecosyst. Environ. 137, 108–112. https://doi.org/10.1016/j.agee.2010.01.008.

Dopheide, A., Tooman, L.K., Grosser, S., Agabiti, B., Rhode, B., Xie, D., Stevens, M.I., Nelson, N., Buckley, T.R., Drummond, A.J., Newcomb, R.D., 2019. Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. Ecol. Appl. 29, e01877. https://doi.org/10.1002/eap.1877.

Edwards, C.A., 1991. Methods for assessing populations of soil-inhabiting invertebrates: the assessment of populations of soil-inhabiting invertebrates. Agric. Ecosyst. Environ. 34, 145–176. https://doi.org/10.1016/0167-8809(91)90102-4.

Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS One 10. https://doi.org/10.1371/journal.pone.0130324.

Elbrecht, V., Leese, F., 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Front. Environ. Sci. 5, 11. https://doi.org/10.3389/fenvs.2017.00011.

Galli, L., Molyneux, T., Trombini, E., Zinni, M., 2025. Effects of land use on soil arthropod communities. Diversity 17, 247. https://doi.org/10.3390/d17040247.

Gao, M., Liu, D., Lin, L., Wu, D., 2016. The small-scale structure of a soil mite metacommunity. Eur. J. Soil Biol. 74, 69–75. https://doi.org/10.1016/j.ejsobi.2016.03.004.

Ghiglieno, I., Simonetto, A., Sperandio, G., Ventura, M., Gatti, F., Donna, P., Tonni, M., Valenti, L., Gilioli, G., 2021. Impact of environmental conditions and management on soil arthropod communities in vineyard ecosystems. Sustainability 13, 11999. https://doi.org/10.3390/su132111999.

Gossner, M.M., Lewinsohn, T.M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K., Renner, S.C., Sikorski, J., Wubet, T., Arndt, H., Baumgartner, V., Blaser, S., Blüthgen, N., Börschig, C., Buscot, F., DIekötter, T., Jorge, L.R., Jung, K., Keyel, A.C., Klein, A.M., Klemmer, S., Krauss, J., Lange, M., Müller, J., Overmann, J., Pašali, E., Penone, C., Perovic, D.J., Purschke, O., Schall, P., Socher, S.A., Sonnemann, I., Tschapka, M., Tscharntke, T., Türke, M., Venter, P.C., Weiner, C.N., Werner, M., Wolters, V., Wurst, S., Westphal, C., Fischer, M., Weisser, W.W., Allan, E., 2016. Land-use intensification causes multitrophic homogenization of grassland communities. Nature 540, 266–269. https://doi.org/10.1038/nature20575.

Gubler, A., Schwab, P., Wächter, D., Meuli, R.G., Keller, A., 2015. Ergebnisse der Nationalen Bodenbeobachtung (NABO) 1985–2009 (Bern, Switzerland).

Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M., Lear, G., 2025. Soil plant growth-promoting bacterial diversity changes with land use intensity and environmental conditions. Glob. Ecol. Biogeogr. 34, e70092. https://doi.org/10.1111/geb.70092.

Hermans, S.M., Lear, G., Buckley, T.R., Buckley, H.L., 2022. Environmental DNA sampling detects between-habitat variation in soil arthropod communities, but is a poor indicator of fine-scale spatial and seasonal variation. Ecol. Indic. 140, 109040. https://doi.org/10.1016/j.ecolind.2022.109040.

Hsieh, T.C., Ma, K.H., Chao, A., 2016. *iNEXT*: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. 7, 1451–1456 [software]. https://doi.org/10.1111/2041-210X.12613.

Hug, A.-S., Gubler, A., Gschwend, F., Widmer, F., Oberholzer, H.R., Frey, B., Meuli, R.G., 2018. NABObio – Bodenbiologie in der Nationalen Bodenbeobachtung. Agroscope Sci. 63, 1–55.

Kindt, R., Coe, R., 2005. Tree Diversity Analysis. A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies. World Agroforestry Centre (ICRAF).

Kortmann, M., Roth, N., Buse, J., Hilszczański, J., Jaworski, T., Morinière, J., Seidl, R., Thorn, S., Müller, J.C., 2022. Arthropod dark taxa provide new insights into diversity responses to bark beetle infestations. Ecol. Appl. 32. https://doi.org/10.1002/eap.2516.

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583. https://doi.org/10.2307/2280779.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Ladell, W.R.S., 1936. A new apparatus for separating insects and other arthropods from the soil. Ann. Appl. Biol. 23, 862–879. https://doi.org/10.1111/j.1744-7348.1936.tb06132.x.

Lavelle, P., 1996. Diversity of soil fauna and ecosystem function. Biol. Int. 33, 3–16.

Le Cadre, J., Klemp, F.L., Bálint, M., Scheu, S., Schaefer, I., 2024. Applicability and perspectives for DNA barcoding of soil invertebrates. PeerJ 12, e17709. https://doi.org/10.7717/peerj.17709.

Macfadyen, A., 1953. Notes on methods for the extraction of small soil arthropods. J. Anim. Ecol. 22, 65–77. https://doi.org/10.2307/1691.

Macfadyen, A., 1961. Improved funnel-type extractors for soil arthropods. J. Anim. Ecol. 30, 171–184. https://doi.org/10.2307/2120.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10 [software]. 10.14806/ej.17.1.200.

Mateos, E., Álvarez-Presas, M., 2022. Integrative taxonomy reveals three new species of European *Lepidocyrtus lignorum*-group (Collembola, Entomobryidae). Zootaxa 5100, 451–481. https://doi.org/10.11646/zootaxa.5100.4.1.

Mateos, E., Winkler, D., Riutort, M., Álvarez-Presas, M., 2021. New morphological and molecular data reveal an important underestimation of species diversity and indicate evolutionary patterns in European *Lepidocyrtus* (Collembola: Entomobryidae). Invertebr. Syst. 35, 471. https://doi.org/10.1071/IS20016.

McArdle, B., Anderson, M., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82, 290–297. https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2.

McGee, K.M., Porter, T.M., Wright, M., Hajibabaei, M., 2020. Drivers of tropical soil invertebrate community composition and richness across tropical secondary forests using DNA metasystematics. Sci. Rep. 10, 18429. https://doi.org/10.1038/s41598-020-75452-4.

Menta, C., Remelli, S., 2020. Soil health and arthropods: from complex system to worthwhile investigation. Insects 11, 54. https://doi.org/10.3390/insects11010054.

O'Rourke, D.R., Bokulich, N.A., Jusino, M.A., MacManes, M.D., Foster, J.T., 2020. A total crapshoot? Evaluating bioinformatic decisions in animal diet metabarcoding analyses. Ecol. Evol. 10, 9721–9739. https://doi.org/10.1002/ece3.6594.

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H., 2018. *Vegan*: community ecology package. R package version 2.4-6; RStudio. [software]. http://CRAN.R-project.org/package=vegan.

Oliverio, A.M., Gan, H., Wickings, K., Fierer, N., 2018. A DNA metabarcoding approach to characterize soil arthropod communities. Soil Biol. Biochem. 125, 37–43. https://doi.org/10.1016/j.soilbio.2018.06.026.

Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarzyński, D., Stevens, M.I., Jansen Van Vuuren, B., Deharveng, L., 2012. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebr. Syst. 26, 470–477. https://doi.org/10.1071/IS12026.

Porter, T.M., Hajibabaei, M., 2020. Putting COI metabarcoding in context: the utility of Exact Sequence Variants (ESVs) in biodiversity analysis. Front. Ecol. Evol. 8, 248. https://doi.org/10.3389/fevo.2020.00248.

R Core Team, 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ratnasingham, S., Hebert, P., 2007. BOLD: the barcode of life data system (www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Recuero, E., Etzler, F.E., Caterino, M.S., 2024. Most soil and litter arthropods are unidentifiable based on current DNA barcode reference libraries. Curr. Zool. 70, 637–646. https://doi.org/10.1093/cz/zoad051.

Remelli, S., Ghobari, H., Oliveira Filho, L.C.I., 2024. The role of soil mesofauna as indicators of sustainable ecosystem management plans. Front. Ecol. Evol. 12, 1400232. https://doi.org/10.3389/fevo.2024.1400232.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584.

Rota, N., Canedoli, C., Ferrè, C., Ficetola, G.F., Guerrieri, A., Padoa-Schioppa, E., 2020. Evaluation of soil biodiversity in alpine habitats through eDNA metabarcoding and relationships with environmental features. Forests 11, 738. https://doi.org/10.3390/F11070738.

Sahdra, A., Young, M.R., Phillips, L.A., Alavi, N., Lapen, D.R., Kerr, J.T., Schwarzfeld, M. D., 2025. Exploring the influence of natural features on soil mesofaunal communities in agricultural landscapes through DNA metabarcoding. Front. Sustain. Food Syst. 9, 1443215. https://doi.org/10.3389/fsufs.2025.1443215.

Sapkota, R., Buivydaitė, Ž., Lilja, M.A., Ellegaard-Jensen, L., Winding, A., Krogh, P.H., 2025. Evaluating DNA extraction methods for eDNA metabarcoding of soil invertebrate diversity. Eur. J. Soil Biol. 126, 103751. https://doi.org/10.1016/j.ejsobi.2025.103751.

Simatupang, R.A.S., Rahardjo, B.T., 2023. Diversity and abundance of Collembola in agroforestry land in UB forest. JTPP 4, 79–88. https://doi.org/10.21776/ub.jtpp.2023.004.2.4.

Souto-Vilarós, D., Navarro-Valencia, E., Zamora, A.C., Campusano, Y., Lamarre, G.P.A., Perez, F., Lopez, Y., Bobadilla, R., Ramírez Silva, J.A., Janda, M., deWaard, J.R., Leponce, M., Barrios, H., Basset, Y., 2025. Navigating the seven seas of arthropod collection protocols: metabarcoding arthropod diversity in a tropical forest. Methods Ecol. Evol. 16, 2395–2407. https://doi.org/10.1111/2041-210X.70135.

Steinke, D., DeWaard, S.L., Sones, J.E., Ivanova, N.V., Prosser, S.W.J., Perez, K., Braukmann, T.W.A., Milton, M., Zakharov, E.V., Dewaard, J.R., Ratnasingham, S., Hebert, P.D.N., 2022. Message in a bottle - metabarcoding enables biodiversity comparisons across ecoregions. Gigascience 11, 1–11. https://doi.org/10.1093/gigascience/giac040.

Tóth, Z., Dombos, M., Hornung, E., 2023. Urban soil quality deteriorates even with low heavy metal levels: an arthropod-based multi-indices approach. Ecol. Appl. 33. https://doi.org/10.1002/eap.2848.

Tresch, S., Frey, D., Le Bayon, R.C., Zanetta, A., Rasche, F., Fliessbach, A., Moretti, M., 2019. Litter decomposition driven by soil fauna, plant diversity and soil management in urban gardens. Sci. Total Environ. 658, 1614–1629. https://doi.org/10.1016/j.scitotenv.2018.12.235.

Wale, M., Yesuf, S., 2022. Abundance and diversity of soil arthropods in disturbed and undisturbed ecosystem in Western Amhara, Ethiopia. Int. J. Trop. Insect Sci. 42, 767–781. https://doi.org/10.1007/s42690-021-00600-w.

Winkler, D., Mateos, E., Traser, G., Lakatos, F., Tóth, V., 2020. New insight into the systematics of European *Lepidocyrtus* (Collembola: Entomobryidae) using molecular and morphological data. Insects 11, 302. https://doi.org/10.3390/insects11050302.

Young, M.R., Hebert, P.D.N., 2022. Unearthing soil arthropod diversity through DNA metabarcoding. PeerJ 10, 12845. https://doi.org/10.7717/peerj.12845.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T., 2017. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36 [software]. https://doi.org/10.1111/2041-210X.12628.

Zhang, B., Chen, T.W., Mateos, E., Scheu, S., Schaefer, I., 2019. DNA-based approaches uncover cryptic diversity in the European *Lepidocyrtus lanuginosus* species group

(Collembola: Entomobryidae). Invertebr. Syst. 33, 661–670. https://doi.org/10.1071/IS18068.

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., Chave, J., 2019. Body size determines soil community assembly in a tropical forest. Mol. Ecol. 28, 528–543. https://doi.org/10.1111/mec.14919.

Zizka, V.M.A., Geiger, M.F., Leese, F., 2020. DNA metabarcoding of stream invertebrates reveals spatio-temporal variation but consistent status class assessments in a natural and urban river. Ecol. Indic. 115, 106383. https://doi.org/10.1016/j.ecolind.2020.106383.