Abstract

1: Molecular-based methods offer valuable opportunities for assessing soil biodiversity in diverse ecosystems. However, their reliability and large-scale applicability depend on developing protocols and establishing high-quality, curated local reference databases.

2: We focused on soil macroinvertebrates and the assessment of metabarcoding steps, namely the sample decontamination process and the efficiency of taxa recovery, by comparing taxonomy, barcoding, and metabarcoding data. Twenty-four sites were sampled, making up 216 pitfall traps. Following established procedures, specimens were morphologically classified to the lowest taxonomic level. Individuals were then pooled and went through the metabarcoding process. We have developed three experiments: 1) Impact of sample decontamination, where half of the samples were washed with sodium hypochlorite, a commonly used decontaminant. 2) The carabid specimens were identified at the species level by experts, which allowed the building of a carabid mock community and curated reference database for this taxon. The species recovery efficiency was achieved by comparing individual barcoding with metabarcoding samples spiked with DNA from the mock community.

3: Metabarcoding provided a more comprehensive identification of taxonomic orders compared to morphological methods, achieving a higher level of taxonomic resolution. The decontamination process showed no significant impact on OTU richness, indicating it may not be essential. Discrepancies in the recovery of carabid species were noted across the three methods.

4: DNA metabarcoding is a promising technique for macroinvertebrate assessment regarding time, efficiency, and costs, yet reaching greater depth in taxonomic resolution. The common decontamination step is not crucial for soil macrofauna metabarcoding accuracy, and its removal reduces the time and effort required. Here, we demonstrated the potential use of integrative methodologies for robust and rapid biodiversity assessments. Additionally, we state that the reliability of molecular methods is highly dependent on the high quality and availability of well-curated barcode reference databases.

Code and Data Review Statement

All data, code, and bioinformatic pipelines used in this study are available on Dryad through the link DOI: 10.5061/dryad.g1jwstr28. All carabid sequence identifiers are also available in Supplementary Materials B.

Key-words:

Carabidae; Metabarcoding; Methodological Comparison; Mock Community; Soil Macrofauna; Taxonomy

Introduction

Soil is the foundation of terrestrial ecosystems; it supports multiple functions fundamental for ecosystems' balanced functioning. Moreover, the complexity and modularity of the soil system provide different ecological niches supporting an astonishing range of biodiversity (FAO, 2020; Stolte et al., 2016). Indeed, soil is likely one of the most biodiverse environments in terrestrial ecosystems, holding nearly a quarter of Earth's species(Arribas et al., 2021; Decaëns et al., 2006; FAO, 2020; Guerra et al., 2021), consequently being considered a global reservoir of biodiversity (Arribas et al., 2021; FAO, 2020; Guerra et al., 2021).

The critical role of soil in terrestrial ecosystems is widely recognized by organizations, public entities, and governments (European Commission, 2021). In Europe, 60 to 70% of soils are considered unhealthy, with ongoing degradation processes causing significant damage each year (European Commission, 2021). This has prompted an urgent call for soil preservation and restoration efforts, alongside the development of new policies. Soil biodiversity, encompassing the variety of life belowground from genes to communities, is essential for maintaining soil functions and the ecosystem services they provide (FAO, 2020). However, biodiversity loss, a widespread issue across ecosystems, also affects soils, posing a significant threat to European landscapes (European Commission, 2021; Jeffery et al., 2010) . Beyond sustaining soil health, soil biodiversity

has the potential to enhance ecosystem functionality, enabling more efficient use of pesticides and fertilizers while reducing management costs (European Commission, 2021). Establishing fast and reliable monitoring schemes is therefore crucial to assess the impacts of management practices

Biodiversity loss has been widely reported across various ecosystems, and soils are not immune to this trend (Jeffery et al., 2010). Soil biodiversity can be defined as the variety of life belowground, from genes to communities; these organisms play crucial roles in maintaining and supporting soil functions and, therefore, the goods and services provided by soils (FAO, 2020). Indeed, soil biodiversity loss is a threat European soils are facing (European Commission, 2021). Soil biodiversity not only contributes to the health of soils, but it may also be used to improve the ecosystem functionality in a way that allows more efficient use of pesticides and fertilizers, cutting down management costs (European Commission, 2021). Accordingly, setting fast and reliable monitoring schemes is of utmost importance to understand the impact of changes in management practices.

Despite this, there is a clear taxonomic bias in biodiversity research. Soil biodiversity remains understudied compared to other taxa, such as aboveground organisms (Bardgett & Van Der Putten, 2014; Cameron et al., 2018; Decaëns, 2010; Wall, 2012). Explicitly, the classes Arachnida and Insecta contain the least represented species groups in the Global Biodiversity Information Facility (GBIF) database when compared to the number of known species (Troudet et al., 2017). More broadly, however, research on most soil-associated taxa remains sparse, highlighting a significant gap in our understanding of these critical species (Phillips et al., 2020; Salis et al., 2024). Despite

this bias, most classes over and underrepresented in 1950 remained the same in 2016 (Troudet et al., 2017), further complicating soil conservation (Cameron et al., 2018).

Assessing soil biodiversity through morphological-based methodologies faces several challenges: it is financially and time-consuming, soils are highly spatially heterogeneous (FAO, 2020; Jeffery et al., 2010), and demands specialized taxonomic expertise (Arribas et al., 2022; Watts et al., 2019). Besides, entomological specialists are becoming increasingly scarce (Cao et al., 2016; Valdecasas & Camacho, 2003), adding to the already recognised challenges. Cryptic species, larval stages, and damaged individuals can also pose identification difficulties, potentially leading to erroneous results or the inability to be morphologically identified. Furthermore, the lack of a global consensus on sampling and methodological approaches to assess soil biodiversity (Cameron et al., 2018; Guerra et al., 2022) delays or impairs data set comparison (Cameron et al., 2018).

Molecular techniques, such as DNA metabarcoding, are becoming promising methodological alternatives for biodiversity assessments across different ecosystems, including freshwater (Baselga et al., 2013; Bista et al., 2018) marine (Fonseca et al., 2010; Leray & Knowlton, 2015), terrestrial (Arjona et al., 2022; Ji et al., 2013; Martoni et al., 2023; Mata et al., 2021; Watts et al., 2019) and paleoenvironments (Cao et al., 2020).

Metabarcoding can offer rapid, consistent, and cost-effective soil diversity assessments, potentially increasing, significantly the amount of biodiversity data while lessening reliance on diminishing taxonomic expertise (Ji et al., 2013). DNA metabarcoding combines DNA barcoding and high-throughput DNA sequencing (HTS), allowing the sequencing of multiple species in a sample. This allows the simultaneous identification

of several individuals, thereby efficiently reducing the time needed to evaluate the communities present in a given sampling site.

This study addresses the methodological challenges of using molecular methods to assess soil biodiversity, focusing on soil macrofauna in central Portugal. We hypothesize that combining DNA metabarcoding with traditional morphotaxonomy methods can provide a more comprehensive, rapid, and cost-effective approach to assessing soil biodiversity, specifically for soil macrofauna. This study aims to establish and evaluate protocols for collecting, preserving, and analyzing soil macrofauna by comparing and complementing NGS (Next Generation Sequencing) methods, barcoding, and metabarcoding of whole-organism community DNA (wocDNA) with morpho-taxonomy. Arthropod wocDNA metabarcoding is currently widely used to assess arthropod communities (Arribas et al., 2022), yet a harmonized and established protocol is still lacking. Additionally, this research highlights the crucial contribution of traditional taxonomy in creating, curating and expanding reference databases, which are essential for the reliability and precision of metabarcoding results. By integrating these approaches, the study aims to validate the effectiveness and reliability of the methodologies, considering the cost-benefit relationships of sample decontamination and the using of mock communities for the validation of the protocol. This integrative method aims to enhance biodiversity monitoring and conservation efforts, by leveraging the strengths of both molecular techniques and traditional taxonomic expertise.

Sample collection, handling, and storage processes for molecular-based approaches are fundamental and must be carefully considered to avoid DNA contamination and ensure

DNA preservation (Arribas et al., 2022; Liu et al., 2020). Moreover, sample decontamination is a common step before DNA extraction in metabarcoding protocols. It is used to minimize contamination between samples by removing exogenous DNA that might be present in bulk samples, increasing methodological reliability (Hausmann et al., 2021; Jüds et al., 2023). Despite its importance, sample decontamination is time-consuming and can potentially damage the targeted DNA content. In this study, we evaluated the cost-benefit relationship of decontamination by applying a bleaching process, allowing us to assess its impact on DNA integrity and contamination levels. Additionally, a mock community of carabid species was added to selected samples to validate the recovery efficiency and reliability of the metabarcoding pathway. Carabid individuals were morphologically identified, barcoded, and pooled, with their DNA incorporated into pre-selected metabarcoding samples.

This study not only enriches the local reference database of cytochrome oxidase I (COI) mitochondrial gene sequences for carabid species in Portugal but also highlights the importance of integrating molecular techniques to enhance our understanding and conservation of soil ecosystems. Establishing accurate, reliable methods for soil fauna assessment is essential to fill knowledge gaps and improve biodiversity monitoring of soil communities and populations, which still rely heavily on existing taxonomic expertise.

Materials and Methods

Sample Collection and Processing

This study was developed as part of the CULTIVAR project in Idanha-a-Nova, Portugal. The sampling design included 24 soil communities monitoring plots for containing nine pitfall traps arranged in a grid of three parallel transects of three traps each. This layout resulted in a total of 216 pitfall traps across all plots. Pitfall traps were deployed at the end of the fall season, from late November to early December 2022, and remained in the field for 13 days to 17 days. Pitfall traps were filled with 0.3 L of ethylene glycol, and lids were used to minimize rainwater accumulation and preservative dilution and to maintain specimen integrity. After collected, the samples were stored in 96% ethanol at room temperature until further analysis. Samples were sorted under a low-power microscope, where organisms were counted and identified to the lowest possible taxonomic level based on morphological traits. Identification was based on available taxonomic expertise and literature (Quigley & Madge, 1988).

Establishment of a Local COI Barcode Reference Database for Carabid Species Identification

Specimens belonging to the Carabidae family were identified at the species level based on morphological traits (Aguiar & Serrano, 2012). To prevent cross-contamination among samples during specimen handling, all tools, including tweezers, were thoroughly cleaned and flame-sterilized before each use. A local barcode reference database for the cytochrome oxidase subunit 1 (COI) was generated for at least one individual representative of each identified Carabid morphospecies using the 710 bp

COI sequence amplified by the Folmer primers, LCO1490: 5'-GGTCAACAAATCATAAAGATATTGG-3' and HC02198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' (Folmer et al., 1994). For this, individual specimens were removed from 96% ethanol, let dry for 10 minutes on tissue paper, and then bleached with a 1 % sodium hypochlorite (NaOCl) solution by gently shaking the specimen in a 2ml tube for three minutes (Palmer-Young et al., 2019). Subsequently, the bleach was discarded, and individuals washed in distilled water three times for one minute each (Hausmann et al., 2021). DNA extraction was done using a Qiagen DNeasy® Blood & Tissue Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions, with some adaptations. Briefly, small individuals (<1cm) underwent a non-destructive approach where the whole organism was immersed in a digestion buffer overnight. For bigger specimens (>1cm), DNA extraction was done using 1-3 appendices on one side per specimen. These were removed and macerated. DNA quality check was done using a Nanodrop and quantified using Qubit (Thermo Fisher Scientific, Waltham, MA, USA).

For the PCR, 10 µl of Supreme NZYTaq II 2x Green Master Mix (NZYtech, Portugal), 1 µl of each primer (10 nM) and at least 10 ng/µl DNA template were used in each reaction in a total volume of 20 µl per reaction. Cycling conditions consisted of an initial denaturation at 95 ° C for 3 min and five cycles of denaturation at 95 ° C for 30 s, annealing at 46 ° C for 30 s, and extension at 72 ° C for 45 s, followed by 32 cycles of denaturation at 95 ° C for 30 s, annealing at 51 ° C for 30 s, and extension at 72 ° C for with a final elongation at 72 ° C for 10 min. The PCR products were checked on a 1% agarose gel to verify the amplification success and purified with ExoSAP-IT™ Express

(Applied Biosystems by Thermo Fisher Scientific, MA, USA). Sanger sequencing of the purified PCR products was performed (Eurofins, Germany). COI barcode sequences were used as a query in a search against the BOLD database (Barcode of Life Data System: https://boldsystems.org/) and taxonomically assigned to the corresponding species identification where possible. Sequences with no taxonomic assignment on BOLD were also analyzed through blast (Altschul et al., 1990) using the NCBI (National Centre for Biotechnology Information: https://www.ncbi.nlm.nih.gov/) database.

Bulk Sample Preparation, DNA Extraction, and Library Preparation for COI Metabarcoding

Morphotyped specimens from each previously sorted sample were transferred to a Falcon tube, resulting in 191 bulk samples. When handling specimens, all materials, such as tweezers and spatulas, were cleaned with sodium hypochlorite (3%) and 96% ethanol. Before DNA extraction, half of the bulk samples were decontaminated with a 3% sodium hypochlorite solution for one minute to reduce exogenous DNA and were washed three times with distilled water for 1 minute (Hausmann et al., 2021). Bulk samples were then air-dried overnight in an incubator at 56º C. According to the amount of biological content, one to four glass beads of 8 mm diameter were added to each bulk sample to turn the biological material into a fine powder through a homogenizing process using the Bullet Blender 50-DX homogenizer (Next Advance, NY, USA) for at least 15 min. DNA extraction was performed using the E.Z.N.A.® Tissue DNA Kit (Omega Bio-Tek, GE, USA), following an adapted protocol (Mata et al., 2021). DNA extractions were performed using a maximum of 70 mg per sample, when possible 70 mg of

homogenized bulk insect powder. Ants (Formicidae) and Carabid beetles (Carabidae) did not take part in the metabarcoding extraction and amplification process as the first family was required for a different project, while the second was used to make the mock sample used to evaluate the efficiency of the metabarcoding methodology. For the PCR, BF3 (CCHGAYATRGCHTTYCCHCG) and BR2 (TCDGGRTGNCCRAARAAYCA) primer pair with illumina adapters was used, which amplifies the 418 bp amplicon fragment of the cytochrome oxidase I (COI) mitochondrial gene (Elbrecht et al., 2019). A two-step PCR protocol was followed for Illumina library preparation. The first PCR comprised 5 µl of Qiagen Multiplex Master Mix (QIAGEN, Hilden, Germany), 0.3 µl of each 10 nM primer, 3.4 µl of H_2O, and 1 µl of template DNA, previously diluted 1:100. Cycling conditions consisted of initial denaturation at 95 ° C for 15 min, followed by 35 cycles of denaturation at 95 ° C for 30 s, annealing at 45 ° C for 30 s, and extension at 72 ° C for 30 s, with final elongation at 60 ° C for 10 min. PCR products were checked on 2 % agarose gel to verify the amplification success. All samples produced visible amplification bands, while extraction blanks and PCR negative controls showed no signs of contamination. PCR products were diluted 1:4 and a second PCR was performed to incorporate the 7-bp-long identification indexes and Illumina P5 and P7 Illumina sequencing adapters. PCR conditions were similar to the first PCR, except that 7 µl of Kapa HiFi Hot Start mix (Roche, Basel, Switzerland) was used, as well as 0.7 µl of each 10 nM indexing primer. Cycling conditions consisted of initial denaturation at 95 ° C for 3 min, followed by eight cycles of denaturation at 95 ° C for 30 s, annealing at 55 ° C for 30 s, and extension at 72 ° C for 30 s, with a final elongation at 72 ° C for 5 min. The left-side of the PCR products was purified using Agencourt AMPure XP beads (Beckman

Coulter, Brea, CA, United States), quantified using Nanodrop, and diluted to 20 nM. Purified and normalized PCR products were further pooled into a single library and quantified using qPCR with KAPA Library Quant Kit qPCR Mix (Roche, Basel, Switzerland). The final library was sequenced on an Illumina Novaseq Platform at Novogene (Cambridge, UK). The dual PCR steps and conditions are illustrated in the Supplementary Materials A Fig. 1.

Mock Community Assembly and Evaluation of OTU Recovery Efficiency

Extracted DNA from each representative carabid species, 31 morphospecies, were pooled in equimolar proportions after measuring with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). This mock community sample served as a control to validate the metabarcoding workflow, particularly in assessing primer performance, taxonomic resolution, and estimation of OTU richness. This community was designed to evaluate the efficiency of metabarcoding pipelines in the recovery efficiency of known taxa sequences by comparing the expected morphospecies diversity with the observed OTU richness, we assessed the accuracy of taxonomic recovery within this pipeline.

DNA was pooled to simulate varying proportions, 10%, 25%, and 50% of the mock sample within high (n = 15) and low-diversity samples (n = 15), to assess the impact of sample diversity complexity. Each treatment combination was replicated five times, resulting in 30 total samples. Additionally, untreated samples without mock community additions were included as controls for comparison. This design allowed us to evaluate

the impact of sample diversity and complexity on the recovery of taxonomic richness and community composition through the metabarcoding pipeline.

Bioinformatic Pipeline

The following bioinformatic pipeline is schematized in the Supplementary Materials A, Fig. 2 and the bioinformatic script can be found as a supplementary material. OBITools 4 (Boyer et al., 2016) was used for the sequence processing, coupled with VSEARCH (Rognes et al., 2016, p. 20) and LULU (Frøslev et al., 2017) for denoising. With the "obipairing" command, paired-end reads were aligned and merged (sequences that failed to overlap were removed). Surviving reads were assigned to samples, and primer sequences were removed using the "obimultiplex" command, allowing four mismatches in expected primer sequences. The "obiuniq" command merged reads that were collapsed into haplotypes, and singletons (haplotypes with only one read per sample) were removed using "obigrep". Then, each remaining haplotype was combined into a single file and once more dereplicated using the command "--derep_fulllength". The "--cluster_unoise" was used to remove PCR and sequencing error sequences, and "--uchime3_denovo" was used to exclude possible chimeric sequences. The remaining sequences were then grouped by a 99% similarity criterion using "--cluster_size" to define operational taxonomic units (OTUs). Putative NUMTs (nuclear copies of the mitochondrial COI) were then identified and removed using LULU (Frøslev et al., 2017), a fundamental step to reduce the number of pseudogene copies that would otherwise be present in the final dataset, thus wrongly inflating the number of OTUs. All remaining OTUs were identified by comparing the representative OTUs of each cluster against

online databases (BOLD and NCBI) through Boldigger (Buchner & Leese, 2020). Only sequences of targeted taxa were kept and analyzed.

Comparative Analysis of Taxonomic Identification Methods and Statistical Validation in Soil Macrofauna Metabarcoding

Although some groups were identified with lower taxonomic levels through morphology, all statistical analyses were conducted at the order level for the identified morphotypes. Larvae and other unidentifiable specimens were excluded from further analysis due to the inherent challenges in achieving accurate identification. For the OTUs, sequences matching animals not part of soil macrofauna or with insufficient taxonomic information were also excluded. OTUs with taxonomic assignments above Class level and with >85% similarity were retained for further analysis. Data resulting from DNA extraction blanks and PCR negatives were assessed for contaminations (Supplementary Materials A, Fig. 3). Two-sample t-tests were conducted to evaluate the comparability of sample richness between decontaminated and non-decontaminated samples. The relationship between the log-transformed abundance and number of OTU reads across the arthropod orders (Supplementary Materials A, Fig. 4) was also assessed. Carabid species-level taxonomic assignments were used for methodological comparisons between traditional, barcoding, and metabarcoding methods. A barcode gap of 7% was determined (Supplementary Materials A, Fig. 5) by merging COI barcode sequences from over 500 carabid nucleotide sequences within the 640-650 bp range obtained from NCBI (Supplementary Materials B). Sequences underwent editing in

MEGA 11.0. 10 (Tamura et al., 2021) and alignment using the ClustalW tool (Thompson et al., 1994). Species partitions were proposed based on genetic distances calculated using the Assemble Species by Automatic Partitioning (ASAP) species delimitation (Puillandre et al., 2021). The barcode gap was identified using the Kimura-2 (K80) parameter model with default settings to estimate genetic distances (Kimura, 1980), considering both intra- and inter-specific distances. Carabid species-level identifications in molecular-based approaches were considered only for similarity values exceeding 93%.

For the COI barcode reference database for carabid species, we used the criteria developed by Conrado et al. (2023), to delimit the Operational Taxonomic Units (OTUs) and morphospecies (see Supplementary Materials Table 1). Here, species where both morphological and molecular data was available were defined as Integrated Operational Taxonomic Units (IOTUs) and species identification that relied only on molecular were separated as Molecular Operational Taxonomic Units (MOTUs). This classification allowed the creation of a Venn diagram for comparing the three methods used. All statistical analyses and data visualization were performed in R Version 2023.06.0 + 421 using packages' iNEXT', v3.0.0. (Hsieh et al., 2016), for diversity estimates, ggplot2, v3.4.2 (Wickham et al., 2024) and ggpubr v0.6.0 (Kassambara, 2023), for data visualization, and ggVennDiagram v1.12 (Gao et al., 2021, 2024) for the creation of a Venn diagram.

Results

Sample Collection and Specimens Identification

A total of 9,418 individuals were collected, and 13 morpho-groups identified (Fig. 1, A). The metabarcoding analysis identified 877 Operational Taxonomic Units (OTUs) belonging to 26 different taxonomic orders. The OTUs identified spanned across a diverse range of orders, including Araneae, Archaeognatha, Blattodea, Chordeumatida, Coleoptera, Crassiclitellata, Dermaptera, Embioptera, Enchytraeida, Hemiptera, Hymenoptera, Isopoda, Julida, Lithobiomorpha, Opiliones, Orthoptera, Polydesmida, Pseudoscorpiones, Pulmonata, Rhabditida, Scolopendromorpha, Scutigeromorpha, Stylommatophora, Tricladida, and Zygentoma (Fig. 1, B).

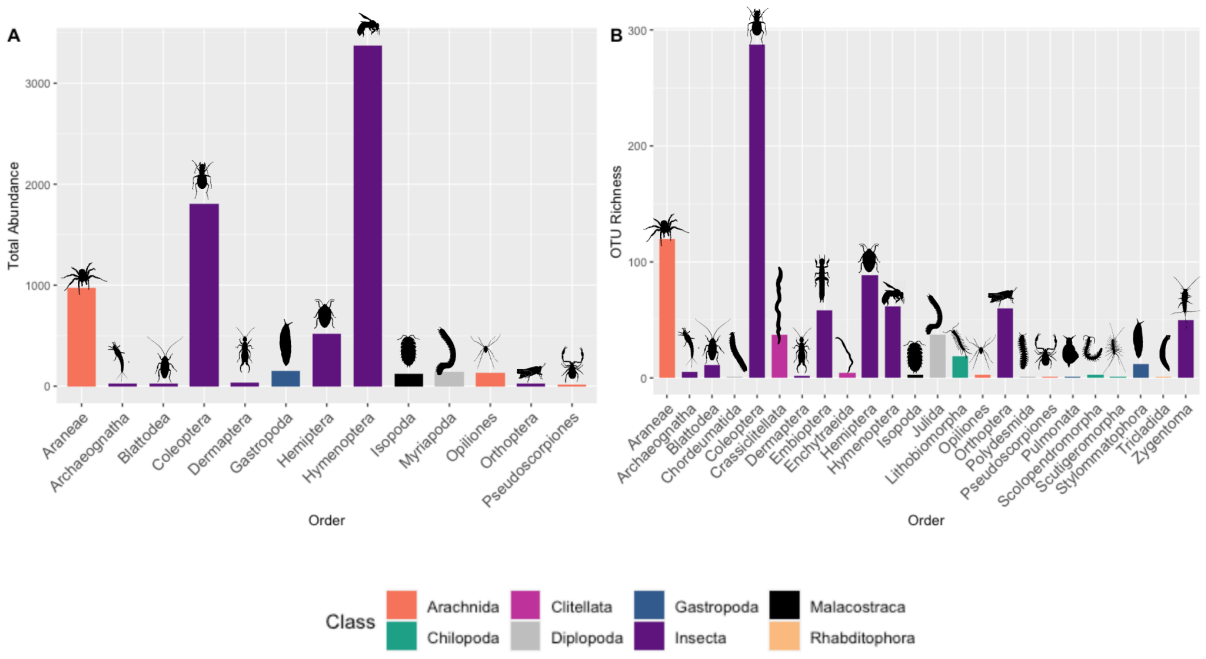

Fig. 1: Abundance and OTU richness of arthropod orders across different taxonomic classes. (A) Total abundance of morphologically identified individuals (lowest taxonomic level possible) (B) OTU richness by order, taxonomic assignment was performed on 07/12/2023. Orders are categorized by their respective classes, indicated by color coding. Silhouettes extracted from PhyloPic (phylopic.org).

Evaluating Decontamination of Bulk Samples

The sodium hypochlorite (NaOCl) decontamination method did not show significant differences in OTU richness between decontaminated and non-decontaminated samples (Fig. 2, A). Additionally, rarefaction curves for both sets of samples followed similar trends, indicating comparable levels of sequencing depth and species richness (Fig. 2, B).

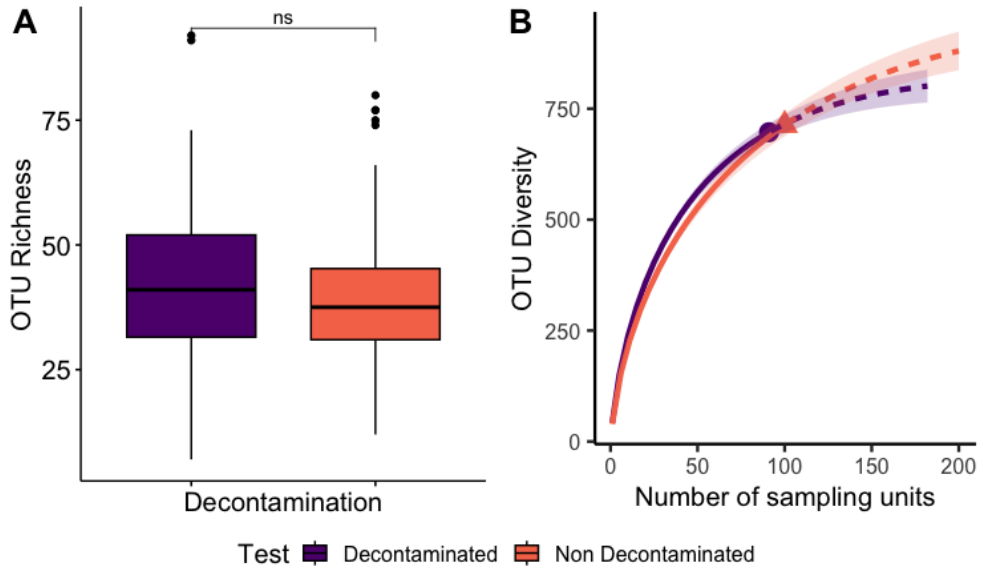

Fig. 2. Comparison of OTU richness and diversity in decontaminated (purple) vs. non-decontaminated (orange) samples: A) Boxplots comparing OTU richness between decontaminated and non-decontaminated samples and analyzed using a two-sample t-test (ns = non-significant); B) Rarefaction curves of OTU richness plot, based on an incidence matrix. Solid lines represent interpolation, dotted lines represent extrapolation, and shaded areas represent 95 % confidence intervals.

Evaluation of OTU Recovery Efficiency

Morphologically, 660 carabid individuals were identified and classified into 31 morphospecies, the corresponding barcodes constituting the mock community. As expected, OTU richness varied with the addition of mock sample (Supplementary Materials A, Fig. 6, A). In high- and low-diversity samples, OTU richness increased proportionally with the addition of mock community DNA regardless of the concentration added (Supplementary Materials A, Fig. 6, A), suggesting a strong ability to recover taxonomic diversity through metabarcoding. However, in low-diversity samples, richness estimates were less consistent and lower overall (Supplementary Materials A, Fig. 6, A). The mock community generated 83 OTUs, of which 65 (77%) OTUs belonged to the order Coleoptera, with 34 different OTUs belonging to the Carabidae family (Supplementary Materials A, Fig. 6, B), consistent with the expected morphospecies composition. This consistency demonstrates the high resolution of metabarcoding for identifying dominant taxa within the sample.

Carabid Species Level Recovery Efficiency

From the 31 morphospecies identified (Fig. 3, A), 52 barcoded specimens generated 24 (63%) COI sequences with a taxonomic assignment, while the metabarcoded mock community generated 18 (47%) carabid sequences with a taxonomic assignment, Fig. 3, A. Only sequences according to the pre-established bad code gap (Supplementary Materials A, Fig. 3) were used for comparison, six sequences exhibited identity

percentages lower than 93% and were excluded from consideration. Fig. 3, A illustrates an overlap in species identification across the three methods of 39% (n = 38). Traditional taxonomy detected 13 species exclusively (34%), while barcoding and metabarcoding had five (15%) and zero (0%) unique identifications, respectively. Notably, two species (5%) were shared between taxonomy and barcoding but not found in metabarcoding, highlighting differences in detection capabilities. At the genus level, Fig. 3, B, with 20 genera considered, most genera (12, or 60%) were identified by all three methods. Traditional taxonomy identified 2 unique genera (10%), barcoding and metabarcoding shared the only identification missing on the method based on traditional taxonomy.

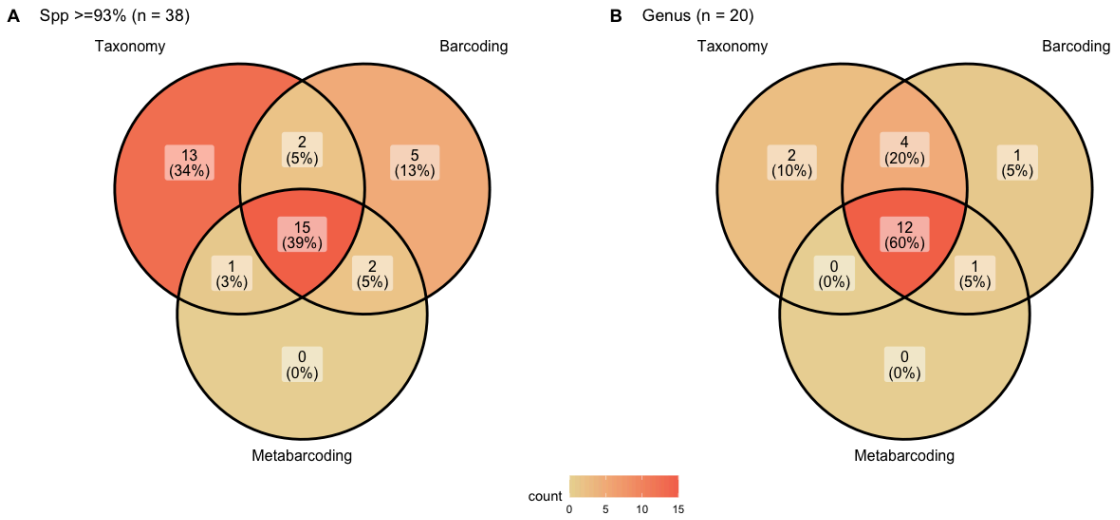

Fig. 3: Comparison of identification methods at species and genus levels for carabid morphospecies using taxonomy, barcoding, and metabarcoding methods. (A) Species-level analysis (n=38) following the criteria established by Conrado et al. (2023), with a barcode gap threshold of ≤7% of homology (Supplementary Materials A, Fig. 5). A

total of 38 species were identified, of which 18 were defined as Integrated Operational Taxonomic Units (IOTUs), six as Molecular Operational Taxonomic Units (MOTUs) and 13 as morphospecies (see Supplementary Materials A, Table 1). Concordance among all three methods used was observed in 15 IOTUs (39%), while 34% of the species were only identified morphologically. (B) Genus-level analysis (n = 20) of carabid taxa showed that 60% of the genera were identified by the three methods and taxonomy identifies 90% of the present genera. Two genus, *Acupalpus* and *Lionepha*, were exclusively identified by molecular methods.

Discussion

Value in the Decontamination effect on DNA metabarcoding samples

Our findings highlight key considerations for optimizing biodiversity research methodologies and ensuring the reliability of molecular analyses. By assessing decontamination efficiency, we provide insights into methodological approaches for molecular studies. Prior research indicates that sodium hypochlorite bleaching does not compromise COI barcoding success, making it effective for decontaminating bulk DNA samples by removing external contaminants and reducing cross-sample contamination (Hausmann et al., 2021; Jüds et al., 2023). This procedure can be time-consuming and, as our results indicate, there is no significant difference in OTU richness between decontaminated and non-decontaminated samples (Fig. 2, A), as further evidenced by similar rarefaction curves (Fig. 2, B). This suggests that decontamination did not alter

the overall taxonomic composition and, therefore, may not be essential for ensuring data reliability when appropriate sampling designs and careful sample handling are in place.

Our results suggest that while decontamination can be valuable (Hausmann et al., 2021), it may not always be essential, especially when robust sampling and handling protocols are in place. Researchers should consider decontamination based on specific research objectives, available resources, and sample-handling practices. The research question, resource availability, and other practical considerations should guide the decision to include decontamination. Additionally, it's essential to recognize the limitations of decontamination; its effectiveness can vary with target taxa, environmental conditions, and specific protocols, underscoring the need for careful evaluation in molecular ecology studies. To further control contamination sources, incorporating negative controls throughout the sampling process is recommended (Liu et al., 2020).

The importance of using mock communities

For each sample where the mock community was added, regardless of concentration, the OTU richness increased as expected. (Supplementary Materials A, Fig. 6, A). This result contributes to validating metabarcoding as a reliable method to recover OTUs regardless of their input proportions. The mock community consisting of 31 barcoded morphospecies of Carabidae identified morphologically, provided a robust benchmark

for assessing the efficiency of metabarcoding pipelines. The observed OTU richness and taxonomic composition was expected; 83 OTUs were identified in the mock community sample, particularly for Carabidae (Supplementary Materials A, Fig. 6, B). Taxa belonging to other coleoptera families or the other orders present may be due to secondary amplification, such as gut content. This suggests that metabarcoding is an effective tool for recovering dominant taxa.

Metabarcoding of the mock community sample assigned 34 OTUs to Carabidae family (Supplementary Materials A, Fig. 6, B), slightly exceeding the expected number of 31 morphospecies. The overestimation observed in assigning OTUs can be attributed to a confluence of factors: the complexities of intraspecific variation, the unpredictable nature of sequencing artifacts, the nuances of secondary amplification, and the biases that may emerge from bioinformatic pipelines(Keck et al., 2023). Using a mock community of Carabidae to validate our metabarcoding pipeline demonstrated that while both barcoding and metabarcoding techniques identified carabid species, certain inconsistencies arose, underscoring the value of integrating multiple methods for accurate species identification (Janzen et al., 2005; Salis et al., 2024).

Due to the filtering of OTUs according to the barcode gap (Supplementary Materials A, Fig. 3), of the 34 OTUs, only 18 remained for the methodological comparison. This reinforces that the effectiveness of DNA metabarcoding relies on the quality, comprehensiveness and accuracy of DNA sequence databases, as taxonomic sequence assignment requires matching unknown DNA sequences against reference databases (Keck et al., 2023; Taberlet et al., 2018). Our findings highlight the importance of

integrative approaches to address the limitations of each method, allowing for more reliable species assignments (Conrado et al., 2023; Cristescu, 2014; Janzen et al., 2005; Salis et al., 2024).

The use of integrative approaches

Collecting biodiversity data on soil macrofauna remains challenging due to the lack of standardized sampling procedures and difficulties in taxonomic assignment (Cameron et al., 2018; Guerra et al., 2022). In our study, sampling across the 24 sites yielded a substantial dataset comprising 9,418 individuals representing a diverse array of taxa.

Morphological classification sorted these specimens into 13 orders (Fig. 1, A), providing a baseline understanding of the macroscopic biodiversity within the study area. Importantly, our metabarcoding approach significantly enhanced taxonomic resolution, increasing to 24 orders as a result of 868 operational taxonomic units (OTUs) identified. As seen in other recent studies, molecular techniques often reveal more diversity at the same taxa level than morphological approaches alone (Arjona et al., 2022; Mata et al., 2021), making metabarcoding an invaluable tool for comprehensive and practical assessments.

Metabarcoding not only allowed a finer-scale taxonomic resolution but also uncovers cryptic diversity that might have been overlooked through traditional morphological methods alone, such as cryptic species complexes (Salis et al., 2024), found in different soil invertebrates, for example in beetles (Mitchell et al., 2020; Pérez-Delgado et al.,

2022) earthworms (King et al., 2008) and pseudoscorpiones (Pérez-Delgado et al., 2022). This enhanced resolution underscores the value of metabarcoding as a tool combined with traditional taxonomy and barcoding, enabling more thorough and accurate biodiversity assessments (Conrado et al., 2023; Cristescu, 2014; Janzen et al., 2005; Salis et al., 2024) and revealing hidden layers of soil biodiversity essential for conservation and ecosystem management .

The importance of comprehensive and reliable reference databases

Molecular reference database platforms such as Bold Systems and NCBI, provide valuable resources for DNA metabarcoding. However, species-level assignment through sequence databases reliability remains a topic of discussion (Leray et al., 2020; Locatelli et al., 2020). Challenges are particularly pronounced for less-researched invertebrate taxa (Leray et al., 2020), whereas assignments at the genus or higher taxonomic levels are generally more reliable (Leray et al., 2019). Our results reflect this trend. As shown in Fig. 3, the strong intersection at the genus level (60%) implies a higher concordance between methods for genus identification compared to species (39%), accordingly a higher rate of assignment discrepancies was found at the species level compared to the genus level (Fig. 3).

These discrepancies may be caused by mismatches between global reference sequences and the local species compositions of the studied ecosystem, which can impair molecular identifications, as a sequence can only be reliably assigned to a

species if its sequence has been previously vouchered (Arjona et al., 2022). Therefore, biodiversity research should not rely exclusively on reference databases for molecular identification of species and these databases must be continuously updated and curated (Keck et al., 2023). In our study, the barcoding assessment of Carabidae revealed cases where sequences did not match any known references, likely due to genetic divergence or underrepresentation in existing databases (Arjona et al., 2022; Keck et al., 2023). The development of high-quality reference databases, particularly local barcode reference libraries based on established taxonomic knowledge, is essential to increase the reliability of molecular approaches (Cristescu, 2014; Janzen et al., 2005; Keck et al., 2023; Salis et al., 2024). For example, updating the BOLD database with our carabid voucher sequences is expected to reduce the proportion of species identified solely through morphological methods and increase congruence across methods. This is projected to shift our Venn diagram results (Fig. 3, A) from the current 34% of species identified only by taxonomy to nearly zero while boosting the overlap across methods.

Creating local barcode references involves collecting specimens from the study area, accurately identifying them using dichotomous keys and trained experts, generating DNA sequences, and comparing these sequences to other verified specimens. Voucher specimens should provide DNA barcodes, key morphological traits, and a bibliography for identification (Cao et al., 2016; Collins & Cruickshank, 2013). For instance, after uploading our carabid barcodes, we anticipate increased alignment across methods, illustrating the critical role of local barcode references for soil macrofauna.

Conclusions and Recommendations

This study highlights molecular approaches' advancements for soil macrofauna biodiversity assessments, particularly DNA barcoding and metabarcoding. We can overcome current limitations by fostering integrative research of both molecular and morphological taxonomic methodologies and developing robust reference libraries, improving data accuracy, and deepening our understanding of complex ecosystems. This incorporation offers significant potential for advancing global and regional biodiversity research and ecological monitoring as researchers can bridge taxonomic gaps, increase resolution, and achieve a deeper understanding of local biodiversity. However, it is critical to establish robust local barcode references.

As our findings demonstrate, while traditional morphological classification provides a foundational understanding of biodiversity, molecular methods significantly enhance taxonomic resolution, document regional species composition and genetic diversity, uncover cryptic diversity, ultimately advancing biodiversity research, ecological monitoring, and our understanding of local biodiversity. Additionally we emphasize the need to continuously update, review and expand global databases, ensuring the accuracy and reliability of molecular identifications, as one of the major challenges in molecular biodiversity research lies in the limitations of reference databases. Curating species-level annotations will require a substantial scientific effort and close interdisciplinary collaboration, involving molecular ecologists and experts from each taxonomic group. For future research, we recommend:

1. Enhancing Local Reference Libraries: Invest in developing and curating local barcode references that capture species' genetic diversity at regional or even local scales. This step is essential for accurate species assignments, particularly for soil invertebrates with high cryptic diversity.
2. Integrative Taxonomic Approaches: Employ both morphological and molecular methods to maximize taxonomic resolution, enabling researchers to bridge gaps in species identification and reveal cryptic taxa.
3. Refining Decontamination Protocols: Investigate the effects of decontamination across different taxa, environmental contexts, and sample types to optimize protocols for molecular studies. Employing negative controls throughout the sampling and laboratory components is also advised to control contamination sources.
4. Database Development and Maintenance: Prioritize the systematic updating of reference databases with voucher-based sequences to enhance the reliability of molecular identifications.
5. Focus on relationships between demographic parameters such as biomass and density with read counts might provide insights in population dynamics

Ultimately, combining molecular and traditional taxonomic methods offers a promising way to address current limitations by fostering integrative research methodologies, developing robust reference libraries, and improving data accuracy. This approach has the potential to enhance biodiversity research, deepen our understanding of species interactions and environmental health, improve existing ecosystem management practices, and optimize conservation strategies.

References

Aguiar, C. A. S., & Serrano, A. R. M. (2012). *Coleópteros carabídeos (Coleoptera, Carabidae) de Portugal Continental: Chaves para a sua identificação*. Sociedade Portuguesa de Entomologia.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arjona, Y., Arribas, P., Salces‐Castellano, A., López, H., Emerson, B. C., & Andújar, C. (2022). Metabarcoding for biodiversity inventory blind spots: A test case using the beetle fauna of an insular cloud forest. *Molecular Ecology*. https://doi.org/10.1111/mec.16716

Arribas, P., Andújar, C., Bohmann, K., deWaard, J. R., Economo, E. P., Elbrecht, V., Geisen, S., Goberna, M., Krehenwinkel, H., Novotny, V., Zinger, L., Creedy, T. J., Meramveliotakis, E., Noguerales, V., Overcast, I., Morlon, H., Papadopoulou, A., Vogler, A. P., & Emerson, B. C. (2022). Toward global integration of biodiversity big data: A harmonized metabarcode data generation module for terrestrial arthropods. *GigaScience*, *11*, 1–12. https://doi.org/10.1093/GIGASCIENCE/GIAC065

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. *Molecular Ecology*, *30*(1), 48–61.

https://doi.org/10.1111/mec.15591

Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, *515*(7528), 505–511. https://doi.org/10.1038/nature13855

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nat. Commun.* https://doi.org/10.1038/ncomms2881

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, *18*(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Buchner, D., & Leese, F. (2020). BOLDigger – a Python package to identify and organise sequences with the Barcode of Life Data systems. *Metabarcoding and Metagenomics*, *4*, e53535. https://doi.org/10.3897/mbmg.4.53535

Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Delgado-Baquerizo, M., Ferlian, O., Fierer, N., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., … Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature Ecology and Evolution*, *2*(7), 1042–1043. https://doi.org/10.1038/s41559-018-0573-8

Cao, X., Liu, J., Chen, J., Zheng, G., Kuntner, M., & Agnarsson, I. (2016). Rapid dissemination of taxonomic discoveries based on DNA barcoding and morphology.

Scientific Reports, *6*(1), Article 1. https://doi.org/10.1038/srep37066

Cao, X., Xu, X., Bian, R., Wang, Y., Yu, H., Xu, Y., Duan, G., Bi, L., Chen, P., Gao, S., Wang, J., Peng, J., & Qu, J. (2020). Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities. *Water Research*, *183*, 116077. https://doi.org/10.1016/j.watres.2020.116077

Collins, R. A., & Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, *13*(6), 969–975. https://doi.org/10.1111/1755-0998.12046

Conrado, A. C., Demetrio, W. C., Stanton, D. W. G., Bartz, M. L. C., James, S. W., Santos, A., da Silva, E., Ferreira, T., Acioli, A. N. S., Ferreira, A. C., Maia, L. S., Silva, T. A. C., Lavelle, P., Velasquez, E., Tapia-Coral, S. C., Muniz, A. W., Segalla, R. F., Decaëns, T., Nadolny, H. S., … Cunha, L. (2023). Amazonian earthworm biodiversity is heavily impacted by ancient and recent human disturbance. *Science of The Total Environment*, *895*, 165087. https://doi.org/10.1016/j.scitotenv.2023.165087

Cristescu, M. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, *29*. https://doi.org/10.1016/j.tree.2014.08.001

Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and Biogeography*, *19*(3), 287–302. https://doi.org/10.1111/j.1466-8238.2009.00517.x

Decaëns, T., Jiménez, J. J., Gioia, C., Measey, G. J., & Lavelle, P. (2006). The values of soil animals for conservation biology. *European Journal of Soil Biology*, *42*, S23–S38. https://doi.org/10.1016/j.ejsobi.2006.07.001

Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, *2019*(10).

https://doi.org/10.7717/peerj.7745

European Commission. (2021). *EU Soil Strategy for 2030: Reaping the benefits of healthy soils for people, food, nature and climate*. European Comission. https://www.eea.europa.eu/data-and-maps/dashboards/land-take-statistics#tab-based-on-data

FAO (with ITPS, GSBI, CBD, & EC). (2020). *State of knowledge of soil biodiversity—Status, challenges and potentialities*. FAO. https://doi.org/10.4060/cb1928en

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

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., Packer, M., Blaxter, M. L., Lambshead, P. J. D., Thomas, W. K., & Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, *1*(1), Article 1. https://doi.org/10.1038/ncomms1095

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-01312-x

Gao, C.-H., Chen, C., Akyol, T., Dusa, A., Yu, G., Cao, B., & Cai, P. (2024). ggVennDiagram: Intuitive Venn diagram software extended. *iMeta*, *3*(1), e177. https://doi.org/10.1002/imt2.177

Gao, C.-H., Yu, G., & Cai, P. (2021). ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.706907

Guerra, C. A., Bardgett, R. D., Caon, L., Crowther, T. W., Delgado-Baquerizo, M., Montanarella, L., Navarro, L. M., Orgiazzi, A., Singh, B. K., Tedersoo, L., Vargas-Rojas, R., Briones, M. J. I., Buscot, F., Cameron, E. K., Cesarz, S., Chatzinotas, A., Cowan, D. A., Djukic, I., Van Den Hoogen, J., … Eisenhauer, N. (2021). Tracking, targeting, and conserving soil biodiversity: A monitoring and indicator system can inform policy. *Science*, *371*(6526), 239–241. https://doi.org/10.1126/science.abd7926

Guerra, C. A., Berdugo, M., Eldridge, D. J., Eisenhauer, N., Singh, B. K., Cui, H., Abades, S., Alfaro, F. D., Bamigboye, A. R., Bastida, F., Blanco-Pastor, J. L., de los Ríos, A., Durán, J., Grebenc, T., Illán, J. G., Liu, Y.-R., Makhalanyane, T. P., Mamet, S., Molina-Montenegro, M. A., … Delgado-Baquerizo, M. (2022). Global hotspots for soil nature conservation. *Nature*, 1–6. https://doi.org/10.1038/s41586-022-05292-x

Hausmann, A., Höcherl, A., Niessner, A., Zakharov, E., Dolynskyj, S., Hausmann, J. D., Höcherl, A., Zakharov, A., Dolynskyi, E., Diller, J., & Dolynskyi, S. (2021). Accurate decontamination of insects from bulk samples does not affect DNA sequencing success. *Spixana*, *44*(1), 71–76. https://doi.org/10.5883/DS-PANLARVA

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

Janzen, D. H., Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., & Hebert, P. D. N. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1835–1845. https://doi.org/10.1098/rstb.2005.1715

Jeffery, S., Gardi, C., Jones, A., Montanarella, M., Marmo, Miko, L., Ritz, K., Pérès, G., Römbke, J., & H. van der Putten, W. (2010). *European atlas of soil biodiversity.*

European Commission. Joint Research Centre. Publications Office of the European Union. https://doi.org/10.2788/94222

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Jüds, M., Heidemann, K., Eitzinger, B., & Scheu, S. (2023). Testing the effectiveness of different wash protocols to remove body surface contaminants in invertebrate food web studies. *PeerJ*, *11*, e16018. https://doi.org/10.7717/peerj.16018

Kassambara, A. (2023). *ggpubr: "ggplot2" Based Publication Ready Plots* (Version R package version 0.6.0) [Computer software]. https://rpkgs.datanovia.com/ggpubr/

Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources*, *23*(4), 742–755. https://doi.org/10.1111/1755-0998.13746

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences*, *116*(45), 22651–22656.

https://doi.org/10.1073/pnas.1911714116

Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B. N., & Machida, R. J. (2020). Reply to Locatelli et al.: Evaluating species-level accuracy of GenBank metazoan sequences will require experts' effort in each group. *Proceedings of the National Academy of Sciences*, *117*(51), 32213–32214. https://doi.org/10.1073/pnas.2019903117

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, *45*(3), 373–385. https://doi.org/10.1111/EEN.12831

Locatelli, N. S., McIntyre, P. B., Therkildsen, N. O., & Baetscher, D. S. (2020). GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. *Proceedings of the National Academy of Sciences*, *117*(51), 32211–32212. https://doi.org/10.1073/pnas.2007421117

Martoni, F., Smith, R. L., Piper, A. M., Nancarrow, N., Aftab, M., Trebicki, P., Kimber, R. B. E., Rodoni, B. C., & Blacket, M. J. (2023). Non-destructive insect metabarcoding as a surveillance tool for the Australian grains industry: A first trial for the iMapPESTS smart trap. *Metabarcoding and Metagenomics*, *7*. https://doi.org/10.3897/mbmg.7.95650

Mata, V. A., Ferreira, S., Campos, R. M., Silva, L. P., Veríssimo, J., Corley, M. F. V., & Beja, P. (2021). Efficient assessment of nocturnal flying insect communities by combining automatic light traps and DNA metabarcoding. *Environmental DNA*, *3*(2), 398–408. https://doi.org/10.1002/edn3.125

Mitchell, A., Moeseneder, C. H., & Hutchinson, P. M. (2020). Hiding in plain sight: DNA barcoding suggests cryptic species in all 'well-known' Australian flower beetles (Scarabaeidae: Cetoniinae). *PeerJ*, *8*, e9348. https://doi.org/10.7717/peerj.9348

Palmer-Young, E. C., Ngor, L., Burciaga Nevarez, R., Rothman, J. A., Raffel, T. R., &

McFrederick, Q. S. (2019). Temperature dependence of parasitic infection and gut bacterial communities in bumble bees. *Environmental Microbiology*, *21*(12), 4706–4723. https://doi.org/10.1111/1462-2920.14805

Pérez-Delgado, A. J., Arribas, P., Hernando, C., López, H., Arjona, Y., Suárez-Ramos, D., Emerson, B. C., & Andújar, C. (2022). Hidden island endemic species and their implications for cryptic speciation within soil arthropods. *Journal of Biogeography*, *49*(7), 1367–1380. https://doi.org/10.1111/jbi.14388

Phillips, H. R. P., Heintz‐Buschart, A., & Eisenhauer, N. (2020). Putting soil invertebrate diversity on the map. *Molecular Ecology*, *29*(4), 655–657. https://doi.org/10.1111/mec.15371

Puillandre, N., Brouillet, S., & Achaz, G. (2021). ASAP: Assemble species by automatic partitioning. *Molecular Ecology Resources*, *21*(2), 609–620. https://doi.org/10.1111/1755-0998.13281

Quigley, M., & Madge, D. (1988). *Land Invertebrates*. Oxford Basil Blackwell Hanitat Field Guides.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Salis, R., Sunde, J., Gubonin, N., Franzén, M., & Forsman, A. (2024). Performance of DNA metabarcoding, standard barcoding and morphological approaches in the identification of insect biodiversity. *Molecular Ecology Resources*, *24*(8), e14018. https://doi.org/10.1111/1755-0998.14018

Stolte, J., Tesfai, M., Øygarden, L., Kværnø, S., Keizer, J., Verheijen, F., Panagos, P., Ballabio, C., Hessel, R., & European Commission. Joint Research Centre. Institute for Environment and Sustainability. (2016). *Soil threats in Europe.* Publications Office.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), 9132. https://doi.org/10.1038/s41598-017-09084-6

Valdecasas, A. G., & Camacho, A. I. (2003). Conservation to the rescue of taxonomy. *Biodiversity & Conservation*, *12*(6), 1113–1117. https://doi.org/10.1023/A:1023082606162

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

Watts, C., Dopheide, A., Holdaway, R., Davis, C., Wood, J., Thornburrow, D., & Dickie, I. A. (2019). DNA metabarcoding as a tool for invertebrate community monitoring: A case study comparison with conventional techniques. *Austral Entomology*, *58*(3), 675–686. https://doi.org/10.1111/aen.12384

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Brand, T. van den, Posit, & PBC. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.5.1) [Computer software]. https://cran.r-project.org/web/packages/ggplot2/index.html