COI metabarcoding with a curated reference database and optimized protocol provides a reliable species-level diversity assessment of tardigrades

Bartłomiej Surmacz[1]*, Matteo Vecchi[1], Diego Fontaneto[2,3], Krystian Budzik[1], Jakub Godziek[1], Yelyzaveta Matsko[4,5], Daniel Stec[1]*

[1] *Institute of Systematics and Evolution of Animals of the Polish Academy of Sciences, Sławkowska 17, 31-016 Kraków, Poland*

[2] *Molecular Ecology Group (MEG), National Research Council of Italy, Water Research Institute (CNR-IRSA), Largo Tonolli 50, 28922 Verbania Pallanza, Italy*

[3] *National Biodiversity Future Center (NBFC), Piazza Marina 61, 90133 Palermo, Italy*

[4] *Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia, Bankowa 9, 40-007 Katowice, Poland*

[5] *Doctoral School at the University of Silesia in Katowice, Bankowa 14, 40-007 Katowice, Poland*

Correspondence: bartek9865@gmail.com; daniel_stec@interia.eu;

Abstract

DNA metabarcoding is revolutionizing biodiversity research by providing rapid and efficient ways of collecting species occurrence data. However, it has not yet been effectively applied to many taxonomic groups, mainly due to a significant lack of reference sequences and dedicated protocols. One such group is the tardigrades - a charismatic phylum of microinvertebrates known for their extremophilic and cryptobiotic capabilities. In this study, we provide the first curated database of 3,194 tardigrade COI sequences sourced from public databases and supplemented with newly produced barcodes. We demonstrate tardigrade metabarcoding in action with optimized PCR primers and a sample processing protocol using 78 samples collected in Poland and Italy. The metabarcoding revealed the presence of more than a hundred operational taxonomic units classified as Tardigrada, representing 23 genera. We compared the metabarcoding results with a morphological survey, which revealed the presence of the same genera, but a lower number of species-level taxa identified morphologically. We observed congruent patterns of tardigrade species richness and taxonomic composition between metabarcoding and morphological surveys. The metabarcoding had a higher discriminatory power, revealing cryptic diversity, and distinguishing species belonging to taxonomically challenging species complexes. By combining metabarcoding with morphological study we were able to find rare taxa, including novel

biogeographic records and putative species new to science, showing also that this approach can be extremely powerful and effective in meiofauna research.

Keywords: Tardigrada; biodiversity; invertebrates; DNA barcoding; meiofauna; metabarcoding;

Introduction

Recognizing biodiversity is crucial for understanding Earth's ecosystems as well as for preserving and conserving life (Norris et al. 2020, Jaureguiberry et al. 2022). In times of biodiversity crisis and rapid environmental changes, efficient methods of exploring and monitoring the vanishing diversity of life are needed. The traditional approaches to biodiversity research, which rely on morphological identifications, are time-consuming, require high taxonomic expertise, and frequently miss cryptic or morphologically similar species (Bickford et al. 2007). The advancements in DNA metabarcoding and the use of environmental DNA (eDNA) have revolutionized biodiversity studies (Taberlet et al. 2012, Deiner et al. 2017, Ficetola & Taberlet 2023). These molecular techniques allow for rapid, large-scale identification of species from environmental samples, providing a powerful means to address the challenges in biodiversity research (Creer et al. 2016, Pawlowski et al. 2018, Taberlet et al. 2018). The use of DNA metabarcoding has a relatively recent history but has rapidly transformed the field of biodiversity assessment. Initially developed in the early 2000s, DNA barcoding aimed to identify species using short, standardized genetic markers (Hebert et al. 2003, Taberlet et al. 2018). This approach laid the groundwork for metabarcoding, which involves the simultaneous identification of multiple species from mixed environmental samples (Taberlet et al. 2012). For metabarcoding of animals, the most widely used marker is cytochrome c oxidase subunit I (COI), a mitochondrial protein-coding gene that has been proposed as the standard species barcode for metazoans (Hebert et al. 2003). Compared to more conservative ribosomal DNA markers, COI's great advantage is the higher taxonomic discriminatory power, usually allowing for accurate assignment of sequences at species-level resolution (Hebert et al. 2003, Tang et al. 2012, Giebner et al. 2020). However, while COI metabarcoding has significantly increased our ability to efficiently detect and monitor animal diversity, it stays not without challenges. Firstly, many taxonomic groups suffer from low coverage of reference sequences and a lack of curated reference databases (Keck et al. 2023, Ficetola & Taberlet 2023). Thus, when using general databases like NCBI GenBank, reference sequences are not extensively curated, and incoherent species assignment may lead to inaccurate species identifications leading to unreliable biodiversity assessments (Porter & Hajibabaei 2018, Fleming 2023). Secondly, due to the higher variability of the COI gene compared to ribosomal regions, the PCR primers should be carefully chosen for the group of interest to avoid amplification bias and off-target amplification (Clarke et al. 2014, Corse et al. 2019). Indeed, well-developed reference libraries and verified protocols exist for some groups, like fish (Kasmi et al. 2023) or insects (Magoga et al. 2022), but for many others, the use of metabarcoding for species-level taxonomic resolution is still extremely limited (Porter & Hajibabaei 2018, Fleming 2023, Macher et al. 2024).

An example of a group for which metabarcoding protocols are yet not sufficiently developed is the phylum Tardigrada. Tardigrades, also known as water bears or moss piglets, are microscopic, water-dwelling invertebrates known for their remarkable resilience to extreme conditions (Nelson et al. 2015).

They can be found in various environments worldwide, including terrestrial habitats such as mosses, lichens, soil, and leaf litter, as well as aquatic environments like marine sediments and freshwater ecosystems (Nelson et al. 2015). The phylum Tardigrada comprises more than 1400 described species (Degma & Guidetti 2024) and previous estimates considered them as a species-poor group (Bartels et al. 2016). However, recent empirical studies underlined the existence of a large portion of undescribed species diversity, mostly due to the presence of cryptic or pseudocryptic speciation, which is a common phenomenon in this group (e.g. Stec et al. 2018, Guidetti et al. 2019, Morek et al. 2021). It suggests that tardigrades are quite a diverse group, poorly understood in terms of their ecology since even their basic biodiversity inventories are extremely challenging and laborious. The biodiversity research on tardigrades is usually achieved by an integrative taxonomy approach, which combines morphological and genetic data to disentangle taxonomic obstacles and cryptic species complexes (e.g. Guidetti et al. 2019, Stec et al. 2021, Stec et al. 2022, Stec 2023, Bertolani et al. 2023). Such an approach is powerful in tardigrades, but time and labour-intensive. Therefore, the premises of DNA metabarcoding in rapid, accurate, and large-scale tardigrade biodiversity research are vast. So far, DNA metabarcoding aimed at tardigrades has been used in five studies which shed light on their diversity, community composition, and trophic structure. Arakawa (2020) performed simultaneous metabarcoding of eukaryote and prokaryote community structures in tardigrade habitats, using 18S rRNA and 16S rRNA metabarcoding of mosses. He et al. (2024) and Pust et al. (2024) used 18S rRNA metabarcoding to investigate soil tardigrades in Australia and Denmark, finding environmental correlates with tardigrade distributions and validating the use of 18S rRNA marker for the classification of tardigrades (usually to higher taxonomic levels than species), even with incomplete reference databases. Zawierucha et al. (2022) utilized metabarcoding of 18S rRNA to characterize microeukaryotic communities of cryoconite holes and their trophic relations. The only attempt to use COI metabarcoding to study tardigrades was the paper by Topstad et al. (2021), in which the authors designed two COI primer sets and tested them together with 18S rRNA metabarcoding and morphological identification. The authors attempted to overcome the limitations of poor reference libraries and possible primer bias by using two markers and two independent COI primer sets to obtain complementary results. However, even in the combined dataset of the two COI amplicons, some morphologically identified taxa were still missing in the metabarcoding results. They concluded that the COI data gave reliable community profiles only when supplemented by 18S rRNA, also revealing higher species richness than traditional morphology-based methods.

Our study further develops the progress achieved by Topstad et al. (2021) by addressing the abovementioned issues in COI metabarcoding to enhance high-throughput biodiversity research on tardigrades. We created the first curated database comprising all available tardigrade COI sequences sourced from NCBI GenBank and the database BOLD (Ratnasingham & Hebert 2007). Then, we optimized the PCR primers and sample processing protocol developed for tardigrade metabarcoding. We implemented our protocol to analyze tardigrade communities of 78 moss samples collected from six

sampling sites across Poland and Italy. Finally, we compared the metabarcoding results to traditional morphological identification to demonstrate the reliability of the primers and protocols we developed for future surveys of tardigrade diversity using COI metabarcoding.

Materials and Methods

COI curated database

We downloaded all the available tardigrade COI sequences from NCBI Genbank using the query string "txid42241[Organism:exp] AND mitochondrion[filter]" (last search date 29/07/2024), and all the COI sequences from BOLD (Ratnasingham & Hebert 2007) labelled as Tardigrada. All sequences were initially manually screened to ensure they contained only COI, as the GenBank search string can sometimes return other mitochondrial genes. The sequences were then aligned using MAFFT (Katoh et al. 2002, Katoh & Toh 2008) with the options --globalpair and --adjustdirection. A phylogenetic tree was constructed from the alignment of all sequences in the database using IQ-TREE (Nguyen et al. 2015) with the GTR+I+G model and codon-based partitioning. Using the resulting tree as a guide, along with information from the publications involved in the generation and reanalysis of the sequences, the database entries were manually curated to reflect the current tardigrade taxonomy and nomenclature (Degma & Guidetti 2024). This curation process included the elimination of pseudogenes and the harmonization of sequence naming conventions. To assign species-level ranks to unclassified sequences or undescribed species, a temporary nomenclature based on the method described by Padial et al. (2010) was applied. Specifically, this nomenclature combines the genus name with the GenBank or BOLD accession number of a representative sequence chosen for that taxon. When morphological resemblance (e.g., cf., aff.) was indicated in the original sequence metadata, this information was retained. For example, "*Macrobiotus* cf. *sapiens* OK662997" indicates a putative species resembling *Macrobiotus sapiens*, with sequence OK662997 selected as the reference for that taxon. Problematic sequences, such as those that were unclassified or contained frameshift mutations, were excluded from the main database but are provided in a separate table available alongside the main database. The complete database is provided in Excel format as well as FASTA file in Supplementary Materials 1. Additional details on rank names are included within the database file itself.

Primer optimization

As an initial choice, we selected the only COI primers used for tardigrade metabarcoding, used by Topstad et al. (2021). We aligned the primer sequences to our curated reference database to assess whether primer bias could have influenced their results. We observed that mismatches in the forward primer region coincided with the failure to detect the genus *Pseudechiniscus*. To address this issue, we

modified the primer by adding ambiguities to two positions (Table 1). In our study, we followed the same naming convention as Topstad et al. (2021): 'COIA' refers to regions amplified using BF2_TardF_2 + BR2, and 'COIB' refers to amplicons generated using BF2_TardF_2 + TardR.

Sample collection and processing

We utilized a collection of 78 moss samples collected in two European countries 39 in Italy and 39 in Poland, to perform a test of our metabarcoding approach. Within each country, three localities representing several environments were chosen (Table 2, Figure 1A), with 13 samples collected in each locality. The Italian samples were collected in 2021, while the Polish samples were collected in 2023 and the size of each moss sample was about 15 cm in diameter. After collection, the samples were placed in paper envelopes, transferred to the laboratory, and stored at room temperature until processing. For detailed information about the samples, please see Supplementary Materials 2.S1. Before processing, each moss sample was carefully fragmented by hand and soaked in tap water overnight in a 0.5-liter plastic beaker. The next day, the sample was vigorously shaken and mixed with a metal spatula, and the water from the beaker was poured through a set of metal sieves: 500 µm, 250 µm, and 36 µm. The procedure of filling the beaker containing moss with tap water, mixing, and pouring the water through the sieves was repeated twice. Then, each upper sieve was carefully washed with tap water so that all fine sediment and tardigrades were caught in the finest sieve. Next, the sediment from the finest sieve was split into three different portions for different analyses (Figure 1B). Portions A and B were prepared by transferring approximately 0.25 g of sediment with a small metal spatula to a 1.5 ml Eppendorf tube. The exact weight of each sample was measured using a laboratory scale. Portion C comprised all remaining sediment left in the finest sieve, which was washed into a 50 ml falcon tube using a wash bottle. All portions were labelled and stored in a freezer until further processing. Between processing of different samples, gloves were changed, and all equipment and the sink were sterilized with 20% bleach. The small metal spatula used for weighing portions A and B was sterilized with alcohol and flame.

Samples preparation for metabarcoding

Sediment portions A and B were used for whole community metabarcoding analysis. Portion A was used directly as sediment, while from portion B, tardigrades and eggs were manually sorted to prepare clean pooled samples, to improve the yield of tardigrade reads. The tubes containing portion B were thawed, and distilled water was added to them. Using a sterile plastic Pasteur pipette, the content of the tube was mixed and transferred into two glass Petri dishes (Ø 7 cm) with engraved lines (7 mm distance between each). Samples in each Petri dish were further diluted and evenly spread using a washing bottle and distilled water. These prepared Petri dishes with diluted sediment were examined under a stereomicroscope, and all tardigrade specimens (animals, eggs, exuviae) were collected into embryological dishes using glass Pasteur pipettes or Irwin loops. For all portions B, division into two

Petri dishes always allowed for good visual condition of the sample, not different from the standard faunistic tardigrade studies. Next, all specimens extracted from portion B were counted and transferred to new Eppendorf tubes using a glass Pasteur pipette with the smallest possible amount of water and stored in the freezer until further processing. If no tardigrades were found in portion B of a given sample, it was marked as empty and the sample was later excluded for morphological survey (details below). The manual extraction was performed with gloves, all Eppendorf tubes were sterile, and plastic, as well as glass pipettes, were changed between different samples. Glass embryo dishes and engraved Petri dishes were sterilized with 20% bleach and washed between examinations of subsequent samples. For detailed information and data from sample processing, please see Supplementary Materials 2.S1.

DNA extraction and library preparation

DNA was extracted from portions A and B using the DNeasy® PowerSoil® Pro Kit (Qiagen) according to the manufacturer's protocol, with a modification: before mechanical homogenization, 10 µl of Proteinase K (concentration 20 mg/ml; A&A Biotechnology) were added to the tubes containing the lysis buffer, zirconium beads, and sediment/pooled tardigrades sample. The samples were then incubated in a thermoshaker at 56 °C with 700 RPM for 30 minutes. After this step, all subsequent procedures followed the original protocol. An extraction blank sample was also included. The final DNA was eluted with 100 µl of elution buffer and stored in the freezer. The libraries were performed using a two-step PCR method. In the first reaction (3 minutes of initial denaturation at 95°C, 35 cycles of 30s of denaturation at 95°C, 30s of annealing at 55°C, and 30s of elongation at 72°C, followed by 5 minutes of final elongation at 72°C), region-specific primers with Illumina overhangs amplified the target regions. The second, indexing PCR amplified the product from the first reaction using primer sets containing flow-cell binding domains and unique indices Nextera XT Index Kit (FC-131-1001/FC-131-1002) following the protocol (Illumina 2013). The same protocol was used for COIA and COIB amplicons (Table 1). For five samples of portion A (samples V1.1 – V.1.5), the COIB libraries were performed in triplicates to check the repeatability of metabarcoding. Purified and pooled libraries were sequenced across two Illumina Novaseq 6000 lanes in 2x250 bp runs, multiplexed with other samples. Library preparation and sequencing were performed by a commercial provider (IGA Technology, Udine, Italy).

Bioinformatic analysis

The raw reads were processed by the custom pipeline based on vsearch (Rognes et al. 2016). Briefly, the paired reads were assembled into contigs and length and quality-filtered using PEAR v0.9.11(Zhang et al. 2014). Then, the primer sequences were trimmed using Cutadapt 4.6 (Martin 2011) with default parameters, and contigs without both primers or of incorrect length after trimming (419-424 for COIA and 439-445 for COIB) were discarded. Then the trimmed contigs were dereplicated, denoised, and

screened for chimeras using the USEARCH- UCHIME algorithm (Edgar et al. 2011). The resulting denoised zero-radius Operational Taxonomic Units were clustered into OTUs with a 97% similarity threshold. While some tardigrade species show a higher intraspecific COI divergence (e.g., Morek et al. 2019), for which the OTU richness may not be equal to species richness, the 97% threshold between intra- and interspecific variability is still recommended in tardigrade barcoding to prevent over-merging (Cesari et al. 2013) and the potential OTUs belonging to problematic taxa can be identified afterward. To remove low-abundance artifacts possibly derived from abundant sequences by index switching we applied 'OTU %' filtering, removing all read counts within a sample, which correspond to OTU that contribute to less than 0.05% of the sum of all reads of a given OTU (Drake et al. 2022) and OTUs with less than four reads. The OTUs' representative sequences were translated into amino acids using R package 'Biostrings' (Pages et al. 2013) using translation table 5 and those containing stop codons were removed from the analysis. To classify the tardigrade sequences, as well as the other eukaryotes we used the 'MIDORI2' reference database (v. GB260 Leray et al. 2022), replacing the sequences classified as Tardigrada with our curated tardigrade reference database. The taxonomy was assigned by classifying the representative sequences of each OTU using the 'BLASTN' method (Altschul et al. 1990, Camacho et al. 2009) with parameters *evalue* = 0.00001, *pident* = 75. The loose threshold of 75% sequence identity was set to find potential tardigrade sequences lacking reference barcodes, to manually choose some potential OTUs as a target for the "Reference database supplementation" step (e.g., when an abundant OTU with the best BLAST hit matching Tardigrada with 75-95% sequence identity was present in a given sample, we aimed to supplement the reference database by barcoding an individual representing the potential tardigrade OTU from the sample). While the 75% sequence identity threshold was used in COI metabarcoding studies (Edgar 2010, Beng et al. 2016) and in our case allowed to recover several tardigrade sequences, to reduce potential false positives in the final OTU tables, we classified the OTUs to phylum level using the commonly accepted threshold of 85% sequence similarity (Clarke et al. 2021, Macher et al. 2024). The sequences with less than 85% identity to a reference were discarded from the analyses. When the similarity to the known tardigrade sequences in the BLASTN search was ≥ 97%, we assigned the taxonomy to the species level, otherwise, the classification was made to the genus level, since it falls within the COI sequence divergence level observed within tardigrade genera (e.g., Morek et al. 2019, Grobys et al. 2020, Stec et al. 2020a). The classification process was done twice: first, using the curated database, and second, after supplementing the reference database with new barcode sequences generated in this study. Finally, we removed OTUs corresponding to putative nuclear paralogs (NUMTs). While commonly used methods, apart from read distribution, rely on the presence of well-annotated reference sequences (Andújar et al. 2021), we anticipated finding a high sequence diversity and many OTUs without exact references in tardigrade metabarcode data, due to a known lack of reference sequences (Topstad et al. 2021). Therefore, we adopted a conservative approach, removing OTUs that consistently co-occurred in samples with at least 10 times more abundant OTUs that had a higher identity to the reference sequence and the same best BLAST match. We identified six such OTUs

in the COIA dataset and three in the COIB dataset. To assess the repeatability of metabarcoding, we calculated the mean percentage of tardigrade taxa identified across all technical replicates. The complete bioinformatics pipeline is provided in Supplementary Material 3.

Morphological survey

In order to compare taxa recovered by metabarcoding analysis and by traditional morphological analysis we manually extracted tardigrade specimens from Portion C of the sediment. Falcons with sediment were thawed at room temperature. To standardize the effort made to examine each sample, similarly to the case of Portion B, two engraved Petri dishes with diluted and evenly spread sediment enabling good vision of the sample were prepared and checked for tardigrades under a stereomicroscope. In order to standardize the procedure, we set the following maximum limits for specimens' extraction: first 50 animals, first 20 eggs, and first 10 exuviae. The two Petri dishes prepared for each sample were examined until reaching these limits. Then, permanent microscope slides were prepared by mounting tardigrade specimens in the Hoyer's medium and securing them with a cover slide. Slides were dried in a heater at 50 °C for five days and after that secured with nail polish. For morphological identification, slides were examined under a Leica DMLB light microscope with phase contrast (PCM), equipped with a digital camera to identify tardigrades to the lowest possible taxonomic level. Plastic and glass Pasteur pipettes were changed between examination of subsequent samples, while Petri and embryo dishes were sterilized with 20% bleach and cleaned. Presence/absence data for morphological operational taxonomic units (morpho OTU table) found in morphological survey is provided in Supplementary Materials 2.S2.

Reference database supplementation (barcoding of selected taxa)

After reviewing the results from metabarcoding and morphological analyses, we selected several sequences of putative tardigrade origin that lacked reference sequences – specifically, sequences with the highest similarity to known tardigrade sequences but with only 75-95% similarity, making it difficult to assign them to a species or even genus level. Based on the metabarcoding results, we attempted to locate individuals matching these sequences in order to provide new COI reference barcodes. In addition to these inferred sequences, we also aimed to collect species belonging to uncommon taxa, or genera and species groups that were not available in the curated reference dataset. We focused on identifying specimens from portion C that corresponded to putative tardigrade genera with the highest similarity to the metabarcoding OTUs found in these samples. Prior to DNA extraction, each specimen was photo-vouchered under a Leica DMLB light microscope. DNA extraction, COI amplification, and sequencing were conducted following the methods of Stec et al. (2020b), using the LCO1490-JJ and HCO2198-JJ primers from Astrin and Stüben (2008). Sequencing was performed on an ABI 3130xl sequencer at Genomed (Warsaw, Poland). Sequences were processed using BioEdit ver. 7.2.5 (Hall 1999) and subsequently submitted to GenBank. Before submission, all COI sequences were translated into protein sequences using MEGA11 (Tamura et al. 2021) to check for pseudogenes. In addition to the samples

investigated in this project, we also produced COI sequences from tardigrades extracted from unrelated samples to increase the diversity and coverage of the database. The new sequences included in the database belong to *Fractonotus verrucosus* (Richters, 1900), *Bertolanius volubilis* (Durante Pasa & Maucci, 1975), *Bertolanius weglarskae* (Dastych, 1972), *Thulinius* aff. *augusti*, *Grevenius* sp., and *Hypsibius* spp. A table listing the taxa selected for generating new barcodes, based on preliminary metabarcoding and morphological surveys along with our justification for the chosen targets, is provided in Supplementary Materials 2.S5. Additionally, a list of all new COI barcodes generated in this study, along with their GenBank accession numbers, is available in Supplementary Materials 2.S7.

Morphological validation of selected taxa found primarily by metabarcoding

To check for putative false positives of taxa recovered by metabarcoding but absent in our morphological investigations, we re-examined selected samples containing such taxa in the metabarcoding results. The re-examination procedure was similar to protocol used for the examination of Portion C, with no petri-dish limits and with the only criterion being the discovery of at least one specimen of the taxon in question. We selected 14 such cases for re-examination. Detailed information about the samples examined in this test is provided in Supplementary Materials 2.S6.

Statistical analysis

The effect of sequencing depth on metabarcoding tardigrade OTU richness was investigated through the analysis of rarefaction curves. The relationships between species richness obtained from the morphological survey and metabarcoding OTU richness were tested using generalized linear models (GLM) with negative binomial distribution and log link function. The relationships between the estimated number of tardigrades per gram of sediment and the percentage of reads classified as Tardigrada in proportion to the total number of reads per sample in Portion A were tested using a GLM with a binomial distribution and logit link function. The estimated tardigrade abundances were $\log_{10}(x+1)$ transformed before analysis due to the high skewness of the values. The models were run separately for COIA and COIB datasets. Differences between the number of OTUs detected by metabarcoding and morphological species richness were tested using the Wilcoxon rank sum test. Differences in community composition between sampling localities, as observed in the morphological and metabarcoding surveys, were tested using PERMANOVA with Jaccard distances, implemented via the *adonis* function in the R package vegan, using 999 permutations (Oksanen et al. 2013). All calculations were performed using R version 4.3.2 (R Core Team 2022).

Data deposition

The raw sequencing reads are deposited in the NCBI Short Read Archive under accession number PRJNA1135541. The raw OTU tables for COIA and COIB are provided in Supplementary Materials 2.S3 and 2.S4, respectively. All supplementary materials associated with this study are also stored in

FigShare repository (https://doi.org/10.6084/m9.figshare.27048157.v1), and the reference database is available as version 1 of the *Tardi-COI* database stored at the GitHub repository https://github.com/bsurmacz/Tardi-COI. The newly obtained barcode references are deposited in GenBank with accession numbers PQ140616- PQ140659.

Results

Curated reference database

The final curated database comprises a total of 3,238 COI sequences from 616 putative species-level taxa. Among these putative species, 293 are assigned to named species, representing nearly 20% of all nominal taxa currently recognized within Tardigrada. Most of the sequences belong to Parachela (1,632), followed by Echiniscoidea (1,367), Apochela (223), and Arthrotardigrada (11). The database is heavily biased toward limnoterrestrial taxa (Figure 2), with some instances where more putative species were identified than the total number of species currently described within certain families (e.g., Calohypsibiidae, Acutuncidae, Adorybiotidae, Richtersiusidae, Milnesiidae, Echiniscoididae). However, there are also 13 families for which no COI references are available, including seven marine families (Anisonychidae, Archechiniscidae, Batillipedidae, Coronarctidae, Neostygarctidae, Renaudarctidae, Stygarctidae) and six limnoterrestrial groups (Carphanidae, Microhypsibiidae, Hexapodibiidae, and the taxa *Mixibius*, *Thalerius konradi* Dastych, 2009, and *Necopinatum mirabile* Pilato, 1971 as three *incertae sedis*).

Metabarcoding results

In total, the filtered dataset included 9,167,186 reads for COIA and 2,810,654 reads for COIB. Supplementing the reference database with newly obtained barcodes for selected taxa increased the percentage of reads classified as tardigrades by approximately 6% in the COIA dataset, while for COIB, the difference before and after supplementation was less than 1%. In terms of the number of OTUs recovered after supplementation with new barcodes, the increase was similar, allowing for the classification of 20 more OTUs in the COIA dataset and 18 more OTUs in the COIB dataset. Finally, 30.22% and 85.81% of reads were classified as tardigrades, respectively (see details in Table 3). The blank samples tested negative for tardigrades. Analysis of the replicates indicated that despite some replicates being outliers, the consistency among PCR replicates of metabarcoding for Portion A was high: on average, 71.9% of tardigrade taxa were recovered by all three replicates, 4.4% in two replicates, and 23.6% were found in only one replicate (see details in Supplementary Materials 2.S8). The COIA marker showed a lower proportion of tardigrade reads compared to COIB, but had an overall higher OTU richness (Table 3). Although it could be linked to a higher coverage: on average 79,280 (SD = 62,586) reads in the COIA libraries of portion A, compared to 6,288 (SD = 8,734) in COIB, the analysis

of rarefaction curves showed that the depth of sequencing was sufficient for both markers (Supplementary Materials 2.S10). Additionally, we observed that COIB consistently does not recover several taxa that are abundant in the COIA dataset (e.g., *Macrobiotus ripperi* Stec, Vecchi & Michalczyk, 2021, *Rammazzottius subanomalus* (Biserov, 1985)*, Cornechiniscus lobatus* (Ramazzotti, 1943)) suggesting possible primer bias. The amplification success of libraries from portion B (the one from extracted tardigrades) was highly variable: in two samples of COIA and 38 samples of COIB, in the final dataset (assembled reads of correct length containing primer sequences) we obtained less than 100 reads per sample. Due to the incompleteness of the data and the potential artifacts caused by cross-contamination or index hopping between libraries, we decided not to use the data from Portion B in the analyses of tardigrade distributions.

Morphological validation of metabarcoding

Among the 14 cases of taxa recovered exclusively by metabarcoding and selected for morphological re-examination to check for false positives, we confirmed the presence of 11 taxa (Supplementary Materials 2.S6). This allowed us to add five genera and six species to the morphological dataset. After supplementing the morphological dataset with these new observations, all the genera identified by metabarcoding were also supported by morphological evidence (Figure 3).

Morphology vs. metabarcoding – assessment of the methods

The morphological survey indicated the presence of 36 distinct tardigrade OTUs, among which 11 were identified only at the genus or species complex/morphogroup level. The richness of morpho-OTUs per sample ranged from zero to nine, with an average of two OTUs per sample. The final species-level identification was achieved for 31 taxa in total, including those found later during the morphological validation step (Figure 3).

During tardigrade extraction from Portion B, no tardigrade specimens were found in 11 out of 78 samples. In the 67 positive samples, tardigrade densities varied, ranging from 4 to 4,417 specimens per gram of sediment, with an average of 366 specimens per gram of sediment per sample.

The metabarcoding analysis recovered 104 tardigrade OTUs (Table 3), of which 61 were classified at the species level after applying a 97% similarity threshold. These 61 OTUs represented 57 distinct species recognized within the curated database, due to the presence of species with high intraspecific variability, such as for example *Milnesium tardigradum* Doyère, 1840 and *Paramacrobiotus bifrons* Brandoli, Cesari, Massa, Vecchi, Rebecchi & Guidetti, 2024 (Morek et al. 2019, Brandoli et al. 2024).

Among the 11 putatively empty samples, in which no tardigrades were found during the initial examination under a stereomicroscope (Portion B), the metabarcoding results from Portion A showed that five samples were negative for tardigrades, with four of these corresponding to samples that were initially marked as empty (Figure 4 and 5).

A GLM analysis indicated a positive relationship between the estimated number of tardigrades per gram of sediment and the percentage of tardigrade reads in both COIA ($p=0.029$) and COIB ($p<0.001$)(Figure 6; Supplementary Materials 2.S9). Additionally, a positive relationship was found between the number of morphological OTUs per sample and the number of metabarcoding OTUs found in both COIA (Nagelkerke's $R^2=0.444$, $p<0.001$) and COIB (Nagelkerke's $R^2=0.283$, $p<0.001$) (Figure 6; Supplementary Materials 2.S9). The number of OTUs identified by COIA was higher than those revealed by morphology (Wilcoxon rank sum test, n=78, W=1673, $p<0.001$), whereas no significant difference was observed between the number of OTUs identified by COIB and morphology (Wilcoxon rank sum test, n=78, W=2822.5, $p=0.431$).

Tardigrade faunas comparison and novel discoveries

Tardigrade communities revealed by metabarcoding were distinct between the Polish and Italian sampling sites for both the COIA (PERMANOVA, $p=0.001$) and COIB datasets (PERMANOVA, $p=0.001$). Among the 57 species identified by metabarcoding, nine were found exclusively in Polish samples, 24 were found only in Italian samples, and 24 species were present in both areas. The communities also exhibited distinct faunal compositions between localities within each country, both for COIA (PERMANOVA, $p=0.003$ for Polish samples; $p=0.001$ for Italian samples) and COIB (PERMANOVA, $p=0.011$ for Polish samples; $p=0.001$ for Italian samples).

By combining DNA metabarcoding with morphological investigation, this study revealed several novel findings regarding tardigrade diversity and distributions. Notably, *Calohypsibius* sp. discovered in an Italian sample represents a species new to science, which needs to be described. *Echiniscus virginicus* Riggin, 1962, found in Italy, constitutes the first record of this taxon in the country and the first record in the Palearctic region. Other new records for Italy include: *Crenubiotus ruhesteini* Guidetti, Schill, Giovannini, Massa, Goldoni, Ebel, Förschler, Rebecchi & Cesari, 2021; *Mesobiotus mandalori* Erdmann, Kosicki, Kayastha, Mioduchowska & Kaczmarek, 2024; and *Mesocrista revelata* Gąsiorek, Stec, Morek, Zawierucha, Kaczmarek, Lachowska-Cierlik & Michalczyk, 2016. New record for Poland is *Mesobiotus nikolaevae* Tumanov, 2018.

Moreover, our metabarcoding results provide evidence for undiscovered putative cryptic or pseudo-cryptic lineages within *Hypsibius* (10 putative species), the *Paramacrobiotus richtersi* complex (one putative species not barcoded before), the *Macrobiotus polonicus* complex (one putative species not barcoded before), and the *Macrobiotus pallari* complex (one putative species not barcoded before). Finally, *Macrobiotus sottilei* Pilato, Kiosya, Lisi & Sabella, 2012, found in this study in southern Poland, are genetically distinct from the population previously found in northern Poland, representing an example of potential cryptic taxa.

Discussion

Our study provides a detailed protocol for COI metabarcoding of limnoterrestrial tardigrades, which, together with the curated reference database, enabled the first large-scale metabarcoding survey of tardigrades at species-level resolution. The DNA extracted from a portion of the sieved substrate from the tardigrade microhabitat was found to accurately represent the tardigrade fauna present in a sample, better than when starting from extracted tardigrades from the same sample. We observed that species richness and community composition obtained through metabarcoding were congruent with results from classical morphological surveys, with metabarcoding allowing us to detect more species (Figure 3B). In many cases, these additional species found by metabarcoding were confirmed through re-examination of the samples, leading to novel discoveries of tardigrade diversity and distributions. This explicitly demonstrates that metabarcoding can be a powerful tool for detecting also rare taxa, including potentially new species to science.

The primer set developed in this study (COIA) showed no evident taxonomic bias, as all the genera identified through morphology were also detected in the metabarcoding data (Figure 3A). However, while being more universal, this fragment was less specific to tardigrades compared to COIB. This outcome is not surprising, given that the reverse primer for the COIB marker (TardR) was designed specifically for tardigrades in times when genetic data for this group was extremely limited (Guil & Giribet 2009). Therefore, it is in line with the theoretical prediction that a more conservative primer will have increased specificity, but if the given DNA fragment is more variable in studied organisms only a specific portion of them can be targeted during the amplification. Nevertheless, apart from samples with very abundant tardigrade populations, in most COIA libraries, less than 10% of reads were classified as Tardigrada (Figure 6A). This information should be considered for planning future tardigrade metabarcoding experiments using COIA primer set. Due to the decreasing costs of sequencing, the lower specificity of the primer set to tardigrades can be easily overcome by a higher sequencing depth, avoiding the taxonomical bias caused by more selective primers. According to logical prediction, metabarcoding of samples composed of manually extracted tardigrades (portion B) efficiently increased the percent of tardigrade reads (Table 3). However, it brought difficulties in library preparation and the risk of cross-contamination due to extremely low input DNA concentrations, which sometimes constitute an important issue in tardigrade DNA studies (Arakawa et al. 2016, Koutsovoulos et al. 2016, Surmacz et al. 2024). In our study, metabarcoding of DNA extracted from sieved sediment using the COIA primer set proved to be an accurate and time-efficient method for revealing tardigrade communities that are consistent with morphological observations. The higher number of OTUs identified by metabarcoding likely reflects issues related to rare species or cryptic species complexes rather than artifacts or the presence of extracellular DNA, since we were able to confirm the presence

of rare taxa identified by metabarcoding by post-hoc morphological validation in 11 out of 14 tested cases (Supplementary Materials 2.S6).

Despite the extensive collection and curation of tardigrade barcodes in the reference database, the absence of reference barcodes for certain taxa still limits the accuracy of tardigrade metabarcoding. We partially addressed this issue by supplementing the database with local sequences from individuals identified in the study, focusing on poorly studied genera and taxa designated after reviewing the initial metabarcoding data. This approach proved to be efficient in increasing the reliability of the metabarcoding results by more confident OTU classification and is a common improvement implemented when studying such challenging groups of animals as those included within meiofauna (e.g. Macher et al. 2024).

The final patterns of community composition revealed by metabarcoding, such as the distinctiveness of regional fauna and the presence of widespread species, align with the current understanding of tardigrade biogeography (Morek et al. 2019, Gąsiorek 2023). Cosmopolitan species such as *Echiniscus testudo* (Doyère, 1840), *Macrobiotus hufelandi* C.A.S. Schultze, and *Milnesium inceptum* were found in both regions, while other species showed rather regional preferences (e.g., *Macrobiotus sandrae* Bertolani & Rebecchi, 1993 and *Paramacrobiotus bifrons*). Additionally, we discovered several new tardigrade records for various countries, including the first occurrence of a species in the Palearctic region (*Echiniscus virginicus*). Interestingly, the distribution of this species modeled in Gąsiorek et al. (2019a) indicated some European localities that could potentially constitute suitable habitats, and our findings (metabarcoding results validated by morphological identification of *E. virginicus* in the sample) empirically confirmed it. All this indicates that tardigrade species distributions are complex and still poorly explored, and the metabarcoding protocol developed in our study is an effective and powerful method for their investigation.

Curated database and local reference barcodes enhance COI metabarcoding of tardigrades

The curation of the COI database revealed a significant imbalance in the presence of COI sequences in GenBank and BOLD, with a strong bias toward limnoterrestrial tardigrades (Figure 1). This imbalance could be attributed to several factors, including the lower number of researchers working on marine tardigrades, their small size, low population densities, and the technical challenges associated with sampling, DNA extraction, and sequencing. It is obvious that to improve the comprehensiveness and completeness of the database, increased sampling efforts are still necessary.

Another important issue concerns misidentification: some sequences in the public databases have been classified into different genera or even classes than what is suggested by their phylogenetic placement. For example, the unpublished sequence KC344344, deposited as *Claxtonia wendti* (Dastych, 1984), has 98.96% identity with *Nebularmis reticulatus* (Murray, 1905) indicating a possible misidentification, as these genera may appear similar to non-taxonomists. Additionally, sequence NOTAR328-19 from

BOLD is registered as *Pseudechiniscus* sp., but when searched against the BOLD database, it returns a 100% similarity with *Minibiotus* sp. This particular sequence is linked to a photo-voucher that clearly shows an Echiniscidae, confirming a reliable identification in BOLD, and thus suggesting a possible contamination during the sequencing process rather than a misidentification of the animal. Another important issue identified during the curation of the database is the presence of many sequences in public databases where neither the original authors nor the community have updated the final taxon names as they appeared in the published papers reporting these DNA data. Without extensive literature knowledge and thorough searches, followed by manual curation of the taxonomy for these sequence entries, their usefulness stays significantly limited. Intrageneric misclassification is also common, often due to the presence of cryptic and pseudocryptic species complexes. A prime example of such a hopeless situation is *Milnesium tardigradum* that for many years was considered the only species present in the genus. Only relatively recently, we learned that this group comprises much more species of which many are morphologically extremely similar (Morek & Michalczyk 2020, Morek et al. 2021). Thus, this situation has led to a proliferation of sequences labelled as *M. tardigradum* that actually belong to other taxa of the genus *Milnesium*. These highlighted issues underscore the importance of manual database curation and ongoing maintenance by researchers with taxonomic expertise in the group of interest, as this is later crucial for providing reliable reference sequences for effective and confident taxa identification, classification and delimitation.

Combining metabarcoding with morphological surveys accelerates meiofauna research

The traditional morphological approach to studying meiofauna is time-consuming and often impractical due to the scarcity of trained taxonomists. This challenge is particularly acute for several tardigrade groups that exhibit significant morphological stasis or cryptic characteristics, which hinder biodiversity assessment using traditional morphological methods (e.g., Stec et al. 2018, Guidetti et al. 2019, Morek et al. 2021). Moreover, classical faunistic studies face additional challenges related to (*i*) ontogenetic variability (e.g., Surmacz et al. 2019, Gąsiorek et al. 2019b, Surmacz et al. 2020), (*ii*) high intraspecific morphological variability within animals or their eggs (e.g., Stec et al. 2016, Brandoli et al. 2024, Stec 2024), and (*iii*) often the limited number of specimens extracted from samples. This third challenge is especially relevant for tardigrade groups that lay ornamented eggs freely in the environment. For these groups, egg morphology often provides crucial information for accurate species identification (Guidetti 2024), yet eggs are frequently absent in samples or easily overlooked. The high labor costs of classical faunistic studies largely stem from the need to secure a sufficient number of specimens for detailed and careful morphological examination.

On the other hand, metabarcoding offers a potentially more advantageous approach since it requires only the DNA of the specimen for detection, capturing all ontogenetic stages or morphological variants that contain DNA. Previous tardigrade metabarcoding studies (Arakawa 2020, Topstad et al. 2021,

Zawierucha et al. 2022, He et al. 2024, Pust et al. 2024), along with the results presented here by us, confirm the benefits of metabarcoding. However, they also highlight several methodological issues, with the most significant being the lack of reference sequences and methodological biases. Our curated COI database of tardigrade sequences clearly shows that the majority of nominal tardigrade species are not barcoded and that marine taxa are severely underrepresented in the database. Although substantial progress has been made since the advent of DNA barcoding (Hebert et al. 2003, DeSalle & Goldstein 2019), considerable gaps remain in DNA reference libraries for many other invertebrates as well (e.g., Jażdżewska et al. 2021, Csabai et al. 2023, Macher et al. 2024, Torres et al. 2024). The approach used in our study involves employing DNA metabarcoding first, followed by traditional morphological methods to refine and validate findings as needed. Using this combined approach, we identified at least six new species records for Italy and Poland, discovered at least one species new to science, and provided COI barcodes for 23 previously unbarcoded taxa. These discoveries would have been extremely time-consuming or even impossible using traditional morphological methods alone.

Our results also highlighted an ongoing challenge in tardigrade taxonomy: the significant level of cryptic and pseudocryptic diversity (e.g., Stec et al. 2018, Guidetti et al. 2019, Morek et al. 2021, Bertolani et al. 2023). In our dataset, we observed ten putative *Hypsibius* species among the OTUs recovered by metabarcoding, whereas only three recognizable morphospecies were identified by light microscopy. This finding suggests that the group is understudied and supports the presence of cryptic lineages, as recently suggested (Zawierucha et al. 2020). Despite recent extensive studies on several problematic species complexes in tardigrades (those challenging due to extreme morphological similarities) some still require further attention. Morphological analysis allowed us to tentatively identify *Paramacrobiotus gadabouti* Kayastha, Stec, Mioduchowska & Kaczmarek, 2023, which belong to the *Parmacrobiotus richtersi* species complex. However, metabarcoding (supported also by newly obtained barcodes) confirmed the presence of other genetically distinct, unidentified species, despite two recent extensive revisions of this group (Guidetti et al. 2019, Stec et al. 2020a). Similarly, for the *Macrobiotus polonicus* and *Macrobiotus pallarii* complexes, which were the subjects of taxonomic revisions based on several European populations by Bertolani et al. (2023) and Stec et al. (2021), our analyses identified two species in each group. In both cases, one taxon was the nominal species (*Macrobiotus dolosus* Bertolani, Cesari, Giovannini, Rebecchi, Guidetti, Kaczmarek & Pilato, 2022, and *Macrobiotus ripperi*), while the other was an unidentified and previously unbarcoded putative species. Finally, morphological analysis in this study assigned several Polish populations to *Macrobiotus sottilei* Pilato, Kiosya, Lisi & Sabella, 2012, as they perfectly matched the original description by Pilato et al. (2012) and the recently found population of this species from northern Poland studied by Kiosya et al. (2021). However, the northern and southern Polish populations constitute genetically separate taxa, with 24% divergence in uncorrected COI distances, and can be considered (pseudo)cryptic species.

Conclusions

The results of our study demonstrate that COI metabarcoding is an efficient method for inventorying tardigrade diversity. The curated database represents a significant step toward the widespread use of metabarcoding in tardigrade research, which could mark a breakthrough in ecological studies of these organisms as well as their biogeography. The biggest challenge in COI metabarcoding of tardigrades remains the lack of reference barcodes for many taxa. However, even with an incomplete reference database, it is still possible to reliably assign most of the tardigrade OTUs to the genus level, as demonstrated in our study. The metabarcoding repeatability and species discovery rate (especially for low-abundance taxa) could be improved by using multiple PCR replicates per sample. Nonetheless, our study showed that using a simple protocol of sediment metabarcoding with highly degenerate primers and a single reaction per sample provides a good estimate of tardigrade communities, largely congruent with the traditional morphological approach.

Although our research was conducted in one of the most well-explored regions of the world for tardigrade fauna (Poland and Italy), where there is a long tradition of studies on these animals, many taxa were still missing reference barcodes. Therefore, increased research efforts are needed to document tardigrade diversity and expand barcode data, especially from poorly explored regions, as well as to barcode underrepresented limnoterrestrial tardigrade families (e.g., Microhypsibidae, Hexapodibidae, Isohypsibidae) and all marine tardigrade taxa. A collective effort by researchers worldwide will be crucial for developing reliable high-throughput inventories, which will help us understand the patterns of distribution and diversity of the enigmatic phylum Tardigrada. Our study design, which improves the results of metabarcoding from environmental samples by their morphological re-examination and providing new reference barcodes, can be also adapted for advancing metabarcoding protocols for other microinvertebrate groups.

Acknowledgments

We are grateful to Szymon Grzesik (Jagiellonian University, Poland) for his kind help with moss collection and sorting tardigrades from samples. For the purpose of Open Access, the authors have applied a CC-BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission. This research was funded by the Preludium (Polish part of the study) and Sonatina (Italian part of the study) programmes of the National Science Centre, Poland (grant no. 2022/45/N/NZ8/01992 to BS and 2022/44/C/NZ8/00050 to DS, respectively. DF is supported by the National Biodiversity Future Centre (NBFC) funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, "Dalla ricerca all'impresa", Investimento 1.4, Project CN00000033.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of molecular biology* **215**, 403–410.

Andújar C, Creedy TJ, Arribas P, López H, Salces-Castellano A, Pérez-Delgado AJ, Vogler AP, Emerson BC (2021). Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. *Molecular Ecology Resources* **21**, 1772–1787.

Arakawa K (2020). Simultaneous Metabarcoding of Eukaryotes and Prokaryotes to Elucidate the Community Structures within Tardigrade Microhabitats. *Diversity* **12**, 110.

Arakawa K, Yoshida Y, Tomita M (2016). Genome sequencing of a single tardigrade *Hypsibius dujardini* individual. *Scientific Data* **3**, 1–7.

Astrin JJ, Stüben PE (2008). Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). *Invertebrate Systematics* **22**, 503–522.

Bartels PJ, Apodaca JJ, Mora C, Nelson DR (2016). A global biodiversity estimate of a poorly known taxon: phylum Tardigrada. *Zoological Journal of the Linnean Society* **178**, 737–746.

Beng KC, Tomlinson KW, Shen XH, Surget-Groba Y, Hughes AC, Corlett RT, Slik JF (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports* **6**, 24965.

Bertolani R, Cesari M, Giovannini I, Rebecchi L, Guidetti R, Kaczmarek Ł, Pilato G (2023). The *Macrobiotus persimilis-polonicus* complex (Eutardigrada, Macrobiotidae), another example of problematic species identification, with the description of four new species. *Organisms Diversity & Evolution* **23**, 329–368.

619 Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I (2007). Cryptic
620 species as a window on diversity and conservation. *Trends in ecology & evolution* **22**, 148–155.

621 Brandoli S, Cesari M, Massa E, Vecchi M, Rebecchi L, Guidetti R (2024). Diverse eggs, diverse species?
622 Production of two egg morphotypes in *Paramacrobiotus bifrons*, a new eutardigrade species within
623 the *areolatus* group. *The European Zoological Journal* **91**, 274–297.

624 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009). BLAST+:
625 architecture and applications. *BMC bioinformatics* **10**, 1–9.

626 Cesari M, Guidetti R, Rebecchi L, Giovannini I, Bertolani R (2013). A DNA barcoding approach in the
627 study of tardigrades. *Journal of Limnology* **72**, 182–198.

628 Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014). Environmental metabarcodes for insects: in silico
629 PCR reveals potential for taxonomic bias. *Molecular ecology resources* **14**, 1160–1170.

630 Clarke LJ, Suter L, Deagle BE, Polanowski AM, Terauds A, Johnstone GJ, Stark JS (2021).
631 Environmental DNA metabarcoding for monitoring metazoan biodiversity in Antarctic nearshore
632 ecosystems. *PeerJ* **9**, e12458.

633 Corse E, Tougard C, Archambaud-Suard G, Agnèse JF, Messu Mandeng FD, Bilong Bilong CF, Duneau
634 D, Zinger L, Chappaz R, Xu CCY, Meglécz E, Dubut V (2019). One-locus-several-primers: A
635 strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecology*
636 *and Evolution* **9**, 4603–4620.

637 Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM (2016). The
638 ecologist's field guide to sequence-based identification of biodiversity. *Methods in ecology and*
639 *evolution* **7**, 1008–1018.

640 Csabai Z, Čiamporová-Zaťovičová Z, Boda P, Čiampor JrF (2023). 50%, not great, not terrible: Pan-
641 European gap-analysis shows the real status of the DNA barcode reference libraries in two aquatic
642 invertebrate groups and points the way ahead. *Science of the Total Environment* **863**, 160922.

643 Degma P, Guidetti R (2024). Actual checklist of Tardigrada species. DOI: 10.25431/11380_1178608.

644 Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista
645 I, Lodge DM, Vere N, Pfrender ME, Bernatchez L (2017). Environmental DNA metabarcoding:
646 Transforming how we survey animal and plant communities. *Molecular ecology* **26**, 5872–5895.

647 DeSalle R, Goldstein P (2019). Review and interpretation of trends in DNA barcoding. *Frontiers in*
648 *Ecology and Evolution* **7**, 302.

649 Drake LE, Cuff JP, Young RE, Marchbank A, Chadwick EA, Symondson WO (2022). An assessment of
650 minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary
651 metabarcoding data. *Methods in Ecology and Evolution* **13**, 694–710.

652 Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**,
653 2460–2461.

654 Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed
655 of chimera detection. *Bioinformatics* **27**, 2194–2200.

656 Elbrecht V, Leese F (2017). Validation and development of COI metabarcoding primers for freshwater
657 macroinvertebrate bioassessment. *Frontiers in Environmental Science* **5**, 11.

658 Ficetola GF, Taberlet P (2023). Towards exhaustive community ecology via DNA
659 metabarcoding. *Molecular Ecology* **32**, 6320–6329.

660 Fleming JF (2023). The wealth of shared resources: Improving molecular taxonomy using eDNA and
661 public databases. *Zoologica Scripta* **52**, 226–234.

662 Gąsiorek P (2023). Catch me if you can, or how paradigms of tardigrade biogeography evolved from
663 cosmopolitism to 'localism'. *Zoological Journal of the Linnean Society* **202**, zlad191.

664 Gąsiorek P, Blagden B, Michalczyk Ł (2019b). Towards a better understanding of echiniscid
665 intraspecific variability: A redescription of *Nebularmis reticulatus* (Murray, 1905) (Heterotardigrada:
666 Echiniscoidea). *Zoologischer Anzeiger* **283**, 242–255.

667 Gąsiorek P, Jackson KJ, Meyer HA, Zając K, Nelson DR, Kristensen RM, Michalczyk Ł (2019a).
668 *Echiniscus virginicus* complex: the first case of pseudocryptic allopatry and pantropical distribution
669 in tardigrades. *Biological Journal of the Linnean Society* **128**, 789–805.

670 Giebner H, Langen K, Bourlat SJ, Kukowka S, Mayer C, Astrin JJ, Misof B, Fonseca VG (2020).
671 Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding
672 approaches using 18S and COI genes. *Molecular Ecology Resources* **20**, 1333–1345.

673 Grobys D, Roszkowska M, Gawlak M, Kmita H, Kepel A, Kepel M, Parnikoza I, Bartylak T, Kaczmarek
674 Ł (2020). High diversity in the *Pseudechiniscus suillus–facettalis* complex (Heterotardigrada:
675 Echiniscidae) with remarks on the morphology of the genus *Pseudechiniscus*. *Zoological Journal of*
676 *the Linnean Society* **188**, 733–752.

677 Guidetti R (2024). Evolution of egg deposition strategies, exaptations of exuvia, and thanatochresis in
678 tardigrades. *Organisms Diversity & Evolution* 1–10. https://doi.org/10.1007/s13127-024-00642-1

679 Guidetti R, Cesari M, Bertolani R, Altiero T, Rebecchi L (2019). High diversity in species, reproductive
680 modes and distribution within the *Paramacrobiotus richtersi* complex (Eutardigrada,
681 Macrobiotidae). *Zoological Letters* **5**, 1.

682 Guil N, Giribet G (2009). Fine scale population structure in the *Echiniscus blumi-canadensis* series
683 (Heterotardigrada, Tardigrada) in an Iberian mountain range—When morphology fails to explain
684 genetic structure. *Molecular Phylogenetics and Evolution* **51**, 606–613.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

He ZY, Hu HW, Nguyen BAT, Chen QL, Weatherley A, Nash M, Bi L, Wu K, He JZ (2024). Distribution of soil tardigrades as revealed by molecular identification across a large-scale area of Australia. *Soil Biology and Biochemistry* 109506.

Hebert PD, Cywinska A, Ball SL, DeWaard JR (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 313–321.

Ilumina (2013). 16s metagenomic sequencing library preparation. Illumina: San Diego, CA, USA, 21.

Jaureguiberry P, Titeux N, Wiemers M, Bowler DE, Coscieme L, Golden AS, Guerra C, Jacob U, Takahashi Y, Settele J, Diaz S, Molnar Z, Purvis A (2022). The direct drivers of recent global anthropogenic biodiversity loss. *Science advances* **8**, eabm9982.

Jażdżewska AM, Tandberg AHS, Horton T, Brix S (2021). Global gap-analysis of amphipod barcode library. *PeerJ* **9**, e12352.

Kasmi Y, Eschbach E, Hanel R (2023). Mare-MAGE curated reference database of fish mitochondrial genes. *BMC Genom Data* **24**, 18.

Katoh K, Toh H (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics* **9**, 286–298.

Katoh K, Misawa K, Kuma KI, Miyata T (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059–3066.

Keck F, Couton M, Altermatt F (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources* **23**, 742–755.

Kiosya Y, Pogwizd J, Matsko Y, Vecchi M, Stec D (2021). Phylogenetic position of two *Macrobiotus* species with a revisional note on *Macrobiotus sottilei* Pilato, Kiosya, Lisi & Sabella, 2012 (Tardigrada: Eutardigrada: Macrobiotidae). *Zootaxa* **4933**, 113–135.

Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National Academy of Sciences* **113**, 5053–5058.

Leray M, Knowlton N, Machida RJ (2022). MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. *Environmental DNA* **4**, 894–907.

Macher JN, Martínez A, Çakir S, Cholley PE, Christoforou E, Curini Galletti M, Gallen L, Garcia-Cobo M, Jondelius U, Jong D, Leasi F, Lemke M, Rubio Lopez I, Sanchez N, Vinter Sorensen M, Todaro MA, Renema W, Fontaneto D (2024). Enhancing metabarcoding efficiency and ecological insights

through integrated taxonomy and DNA reference barcoding: A case study on beach meiofauna. *Molecular Ecology Resources* e13997.

Magoga G, Forni G, Brunetti M, Meral A, Spada A, De Biase A, Montagna M (2022). Curation of a reference database of COI sequences for insect identification through DNA metabarcoding: COins. *Database*, **2022**, baac055.

Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.

Morek W, Michalczyk Ł (2020). First extensive multilocus phylogeny of the genus *Milnesium* Doyère, 1840 (Tardigrada) reveals no congruence between genetic markers and morphological traits. *Zoological Journal of the Linnean Society* **188**, 681–693.

Morek W, Stec D, Gąsiorek P, Surmacz B, Michalczyk Ł (2019). *Milnesium tardigradum* Doyère, 1840: The first integrative study of interpopulation variability in a tardigrade species. *Journal of Zoological Systematics and Evolutionary Research* **57**, 1–23.

Morek W, Surmacz B, López-López A, Michalczyk Ł (2021). "Everything is not everywhere": time-calibrated phylogeography of the genus *Milnesium* (Tardigrada). *Molecular Ecology* **30**, 3590–3609.

Nelson DR, Guidetti R, Rebecchi L (2015). Phylum Tardigrada. In: Ecology and General Biology: Vol. 1: Thorp and Covich's Freshwater Invertebrates (4th Edition): 347-380 (chapter 17).

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**, 268–274.

Norris K, Terry A, Hansford JP, Turvey ST (2020). Biodiversity conservation and the earth system: mind the gap. *Trends in Ecology & Evolution* **35**, 919–926.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Wagner H (2013). vegan: Community Ecology Package. R package version 2.0-10. URL: http://CRAN. R-project. org/package= vegan.

Padial JM, Miralles A, De la Riva I, Vences M (2010). The integrative future of taxonomy. *Frontiers in zoology* **7**, 1–14.

Pages H, Aboyoun P, Gentleman R, DebRoy S, Pages MH, DataImport D, XStringSet-io R (2013). Package 'Biostrings'. Bioconductor, 18129.

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero, A et al. (2018). The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment* **637**, 1295–1310.

Pilato G, Kiosya Y, Lisi O, Sabella G (2012). New records of Eutardigrada from Belarus with the description of three new species. *Zootaxa* **3179**, 39–60.

Porter TM, Hajibabaei M (2018). Over 2.5 million COI sequences in GenBank and growing. *PLoS ONE*, *13*, e0204175.

Pust FL, Frøslev TG, Kristensen RM, Møbjerg N (2024). Environmental DNA metabarcoding of Danish soil samples reveals new insight into the hidden diversity of eutardigrades in Denmark. Zoological *Journal of the Linnean Society* **200**, 20–33.

R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org

Ratnasingham S, Hebert PD (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes* **7**, 355–364.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.

Stec D (2023). Integrative taxonomy helps to revise systematics and questions the purported cosmopolitan nature of the type species within the genus *Diaforobiotus* (Eutardigrada: Richtersiusidae). *Organisms Diversity & Evolution* 23, 309–328.

Stec D (2024). Integrative taxonomy supports two new species of *Macrobiotus* (Tardigrada: Eutardigrada: Macrobiotidae) allowing further discussion on the genus phylogeny. *European Journal of Taxonomy* 930, 79–123.

Stec D, Kristensen RM, Michalczyk Ł (2020b). An integrative description of *Minibiotus ioculator* sp. nov. from the Republic of South Africa with notes on *Minibiotus pentannulatus* Londoño et al., 2017 (Tardigrada: Macrobiotidae). *Zoologischer Anzeiger* **286**, 117–134.

Stec D, Krzywański Ł, Zawierucha K, Michalczyk Ł (2020a). Untangling systematics of the *Paramacrobiotus areolatus* species complex by an integrative redescription of the nominal species for the group, with multilocus phylogeny and species delineation in the genus *Paramacrobiotus*. *Zoological Journal of the Linnean Society* **188**, 694–716.

Stec D, Morek W, Gąsiorek P, Michalczyk Ł (2018). Unmasking hidden species diversity within the *Ramazzottius oberhaeuseri* complex, with an integrative redescription of the nominal species for the family Ramazzottiidae (Tardigrada: Eutardigrada: Parachela). *Systematics and Biodiversity* **16**, 357–376.

Stec D, Morek W, Gąsiorek P, Kaczmarek Ł, Michalczyk Ł (2016). Determinants and taxonomic consequences of extreme egg shell variability in *Ramazzottius subanomalus* (Biserov, 1985) (Tardigrada). *Zootaxa* **4208**, 176–188.

Stec D, Vecchi M, Dudziak M, Bartels PJ, Calhim S, Michalczyk Ł (2021) Integrative taxonomy resolves species identities within the *Macrobiotus pallarii* complex (Eutardigrada: Macrobiotidae). *Zoological Letters* **7**, 9.

Stec D, Vončina K, Kristensen RM, Michalczyk Ł (2022). The *Macrobiotus ariekammensis* species complex provides evidence for parallel evolution of claw elongation in macrobiotid tardigrades. *Zoological Journal of the Linnean Society* **195**, 1067–1099.

Surmacz B, Morek W, Michalczyk Ł (2019). What if multiple claw configurations are present in a sample: a case study with the description of *Milnesium pseudotardigradum* sp. nov. with unique developmental variability. *Zoological Studies* **58**, 32.

Surmacz B, Morek W, Michalczyk Ł (2020). What to do when ontogenetic tracking is unavailable: a morphometric method to classify instars in *Milnesium* (Tardigrada). *Zoological Journal of the Linnean Society* **188**, 797–808.

Surmacz B, Stec D, Prus-Frankowska M, Buczek M, Michalczyk Ł, Łukasik P (2024). Pinpointing the microbiota of tardigrades: What is really there? *Environmental Biology* **26**, e16659.

Taberlet P, Bonin A, Zinger L, Coissac E (2018). Environmental DNA: For biodiversity research and monitoring. *Oxford University Press*.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012). Environmental DNA. *Molecular ecology* **2**, 11789–1793.

Tamura K, Stecher G, Kumar S (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution* **38**, 3022–3027.

Tang CQ, Leasi F, Obertegger U, Kieneke A, Barraclough TG, Fontaneto D (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences* **109**, 16208–16212.

Topstad L, Guidetti R, Majaneva M, Ekrem T (2021). Multi-marker DNA metabarcoding reflects tardigrade diversity in different habitats. *Genome* **64**, 161–164.

Torres AF, Wangensteen OS, Renema W, Meyer CP, Fontanilla IKC, Todd JA (2024). Global species hotspots and COI barcoding cold spots of marine Gastropoda. *Biodiversity and Conservation* 33, 2925–2947.

Zawierucha K, Buda J, Jaromerska T, Janko K, Gąsiorek P (2020). Integrative approach reveals new species of water bears (*Pilatobius*, *Grevenius*, and *Acutuncus*) from Arctic cryoconite holes, with the discovery of hidden lineages of *Hypsibius*. *Zoologischer Anzeiger* **289**, 141–165.

814 Zawierucha K, Trzebny A, Buda J, Bagshaw E, Franzetti A, Dabert M, Ambrosini R (2022). Trophic
815 and symbiotic links between obligate-glacier water bears (Tardigrada) and cryoconite
816 microorganisms. *Plos one* **17**, e0262039.

817 Zhang J, Kobert K, Flouri T, Stamatakis A (2014). PEAR: a fast and accurate Illumina Paired-End reAd
818 mergeR. *Bioinformatics* **30**, 614–620.

819 **Tables**

820 **Table 1.** Primers used in this study. Bolded "NS" in the forward primer replaced "KC" present in the
821 original primer from Topstad et al. (2021).

Primer name	Source	Direction	Sequence (5' – 3')
BF2_TardF_2	modified from Topstad et al. (2021) (BF2_TardF)	Forward	GCNCCNGAYATR**NS**NTTYCC
BR2	Elbrecht and Leese (2017)	Reverse	TCDGGRTGNCCRAARAAYCA
TardR	Guil and Giribet (2009)	Reverse	GGWARAATHARAATATADAC

822

823 **Table 2.** Information about sampling sites.

Locality	**Geographic name**	**Coordinates**		**Description**
K1	Tyniec, Kraków, Poland	50.019°N	19.801°E	suburban area covered by limestone rocks and anthropogenic surfaces
K2	Przegorzały, Kraków, Poland	50.038°N	19.887°E	remnants of near-natural riparian forest surrounded by the city
K3	Kraków, Poland	50.065°N	19.949°E	urbanized city centre
V1	Mergozzo, Italy	45.940°N	8.440°E	tourist trail in deciduous forest
V2	Cavandone, Verbania, Italy	45.944°N	8.526°E	rural area surrounded by deciduous forests
V3	Pallanza, Verbania, Italy	45.926°N	8.565°E	urbanized town centre

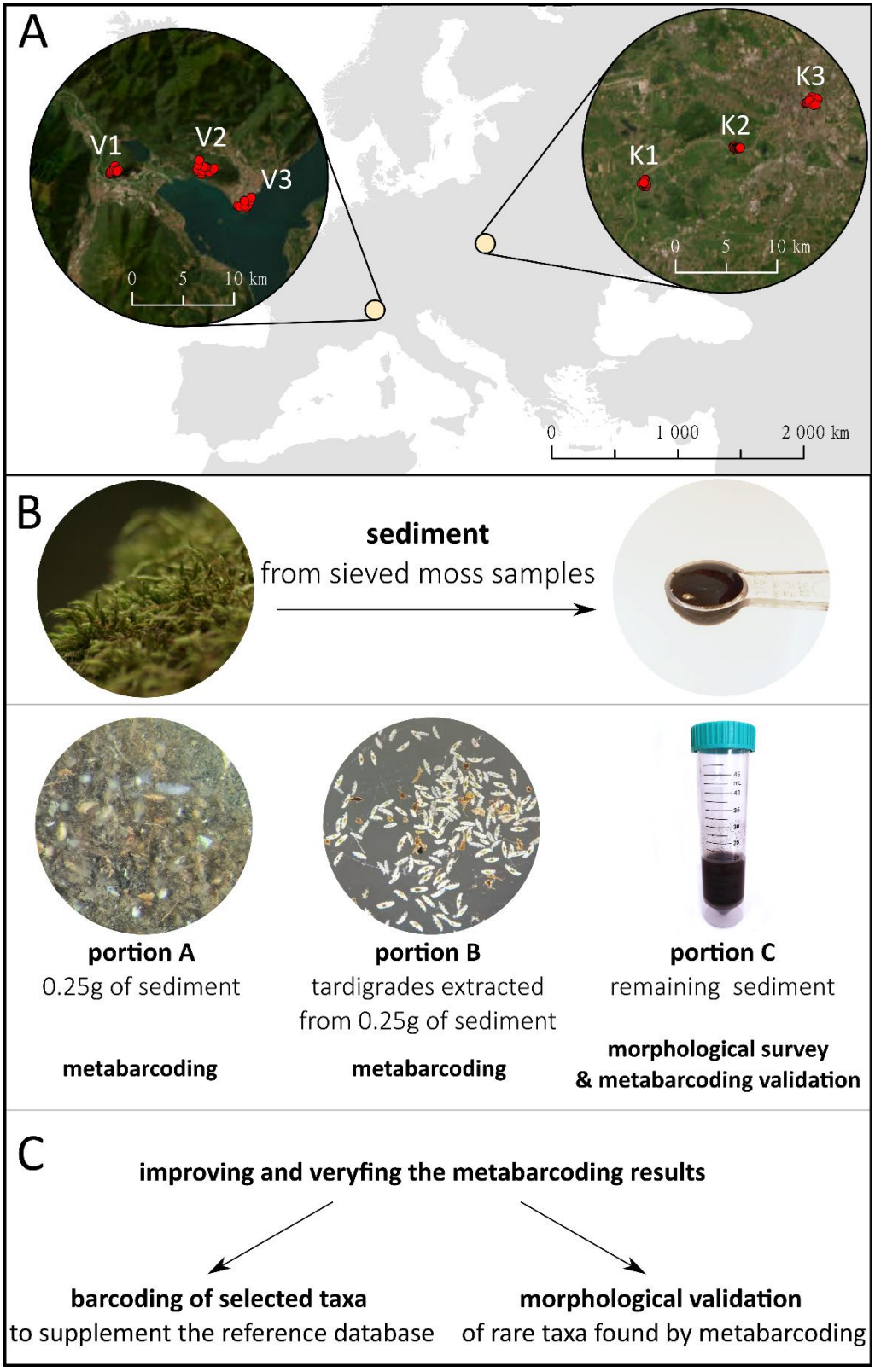
824

825 **Table 3.** Summary of metabarcoding results.

		Before supplementing the reference database		After supplementing the reference database	
		COIA	COIB	COIA	COIB
% of tardigrade reads	Portion A	10.60%	35.17%	11.31%	38.16%
	Portion B	53.40%	97.00%	69.41%	97.33%
Number of tardigrade OTUs	Portion A	81	49	100	62
	Portion B	80	48	99	65
	Total (A+B)	84	56	104	74

826

827 **Figures**



828

829 **Figure 1.** Sampling sites (A) and overview of sample processing methods (B and C). Background
830 satellite imagery source: Earthstar Geographics.

831

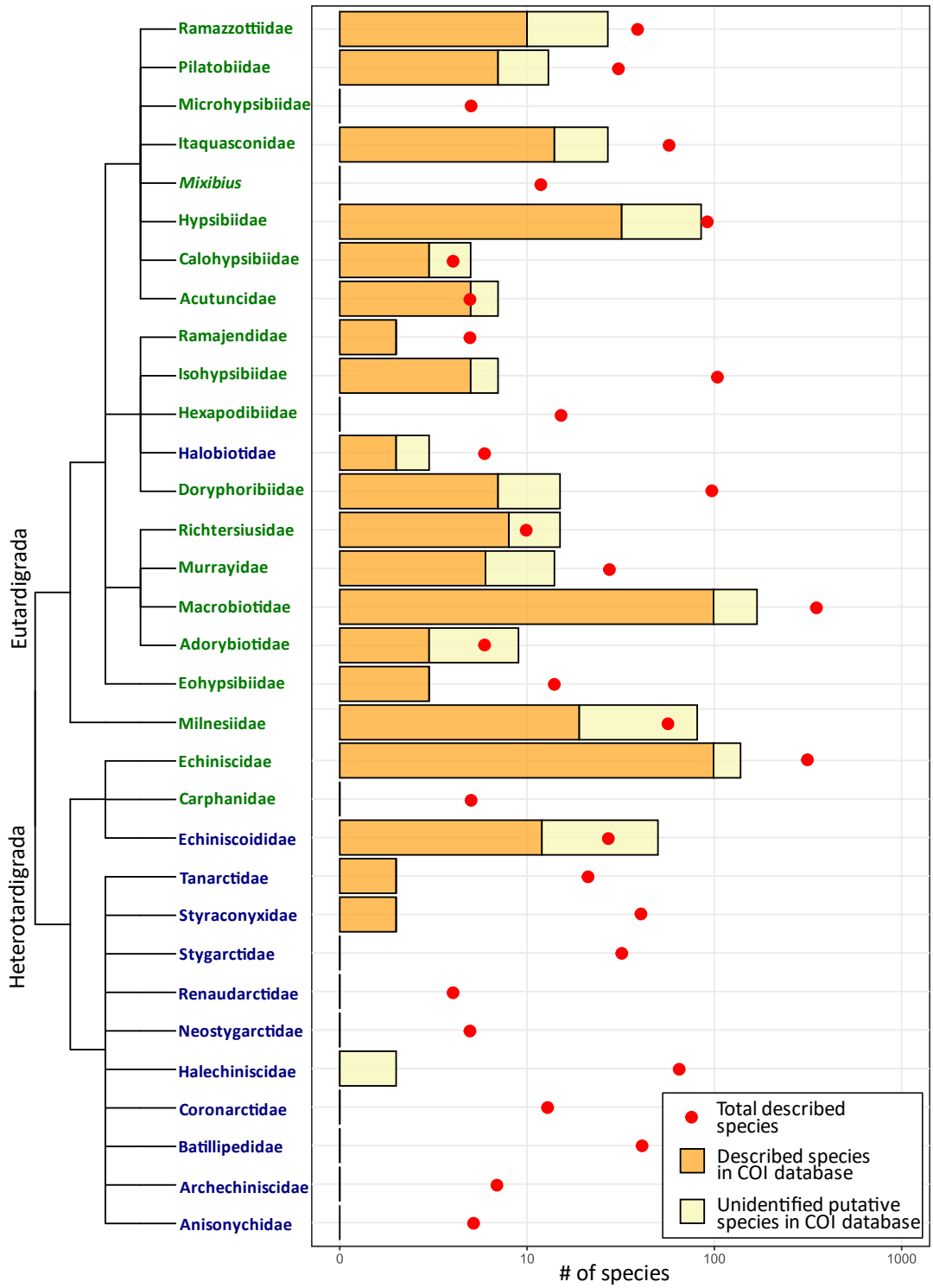


Figure 2. The number of putative species-level taxa in the COI curated database aggregated at the family level. Families marked in blue comprise marine taxa, while those marked in green comprise limnoterrestrial taxa. The genus *Mixibius* is presented alongside families as it is *incertae sedis* and contains 11 species. Other *incertae sedis* taxa (*Thalerius konradi* and *Necopinatum mirabile*) and fossil taxa are not included. The total number of described species was obtained from Degma and Guidetti (2024).

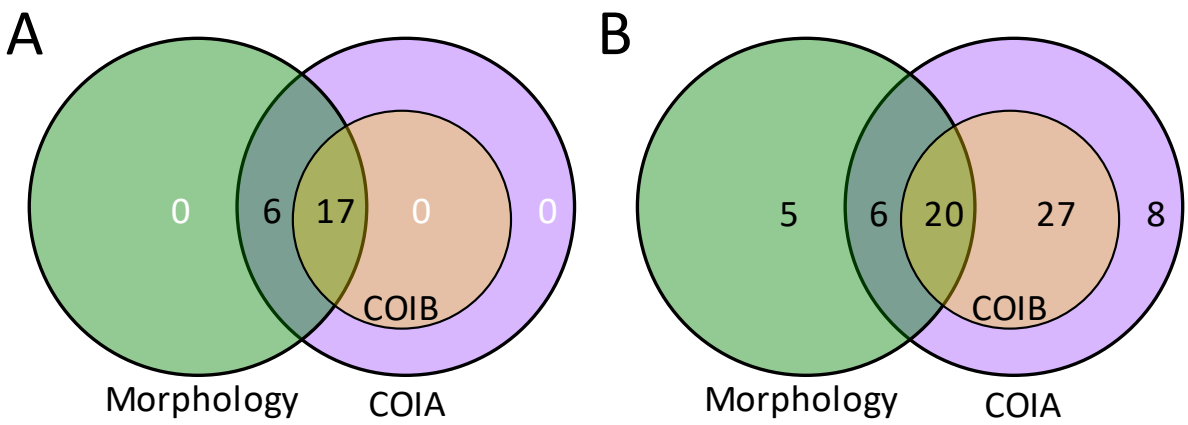
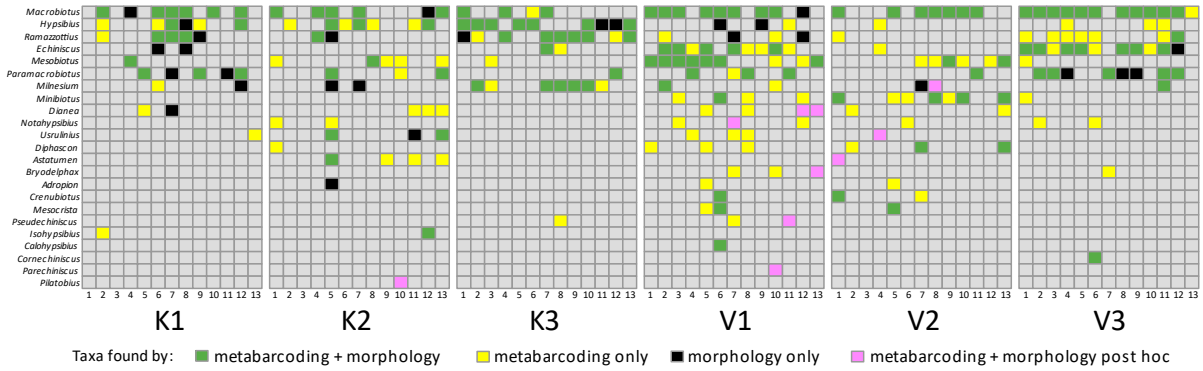


Figure 3. Comparison of the number of tardigrade genera (A) and species (B) identified using each method, including taxa found during the post hoc morphological validation step.



845

846 **Figure 4.** Comparison of tardigrade genera found by morphological survey and metabarcoding of
847 portion A (COIA + COIB) in the six studied localities.

848

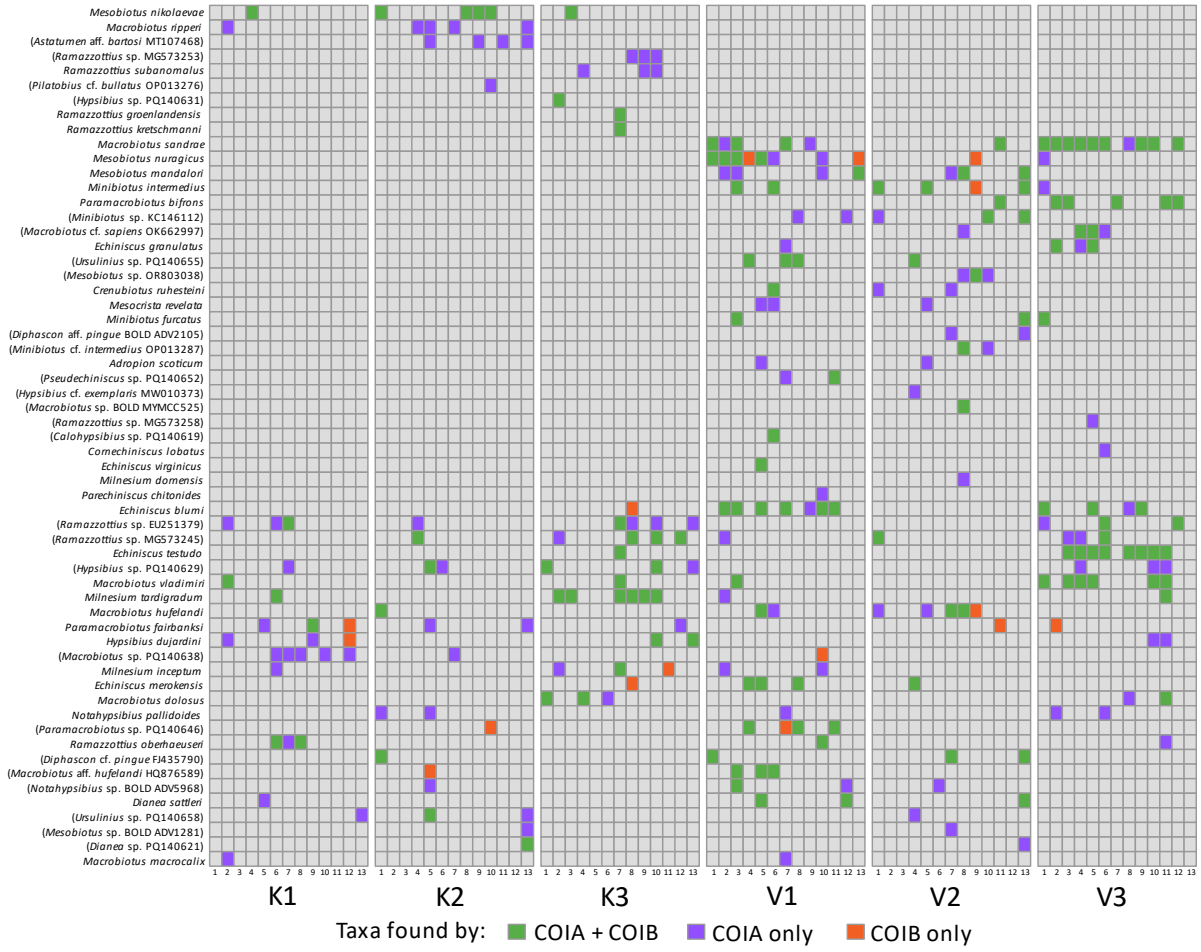
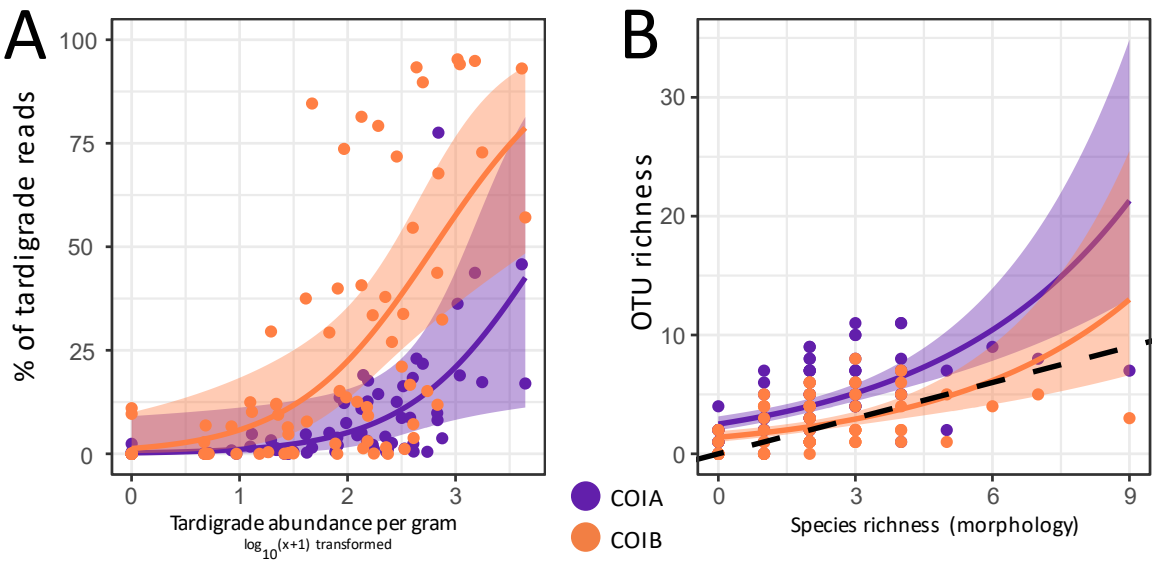


Figure 5. Tardigrade species (OTUs with at least 97% sequence identity to reference) recovered by metabarcoding of portion A in the six studied localities.



853

854 **Figure 6.** Comparison of metabarcoding results from Portion A and the morphological survey: (A)
855 relationship between the estimated abundance of tardigrades per gram of sediment and the percentage
856 of reads classified as tardigrades, and (B) relationship between morphological species richness and the
857 number of tardigrade OTUs recovered from metabarcoding. Solid lines indicate regression lines with
858 95% confidence intervals from GLM, while the dashed line in panel B represents identical values.

859

860

861

Supplementary Materials (https://doi.org/10.6084/m9.figshare.27048157.v1)

Supplementary Materials 1. Curated database of tardigrade COI sequences.

Supplementary Materials 2. Supplementary tables for "COI metabarcoding with a curated reference database and optimized protocol provides a reliable species-level diversity assessment of tardigrades": (S1) sample information, (S2) Morpho-OTU table, (S3) COIA-OTU table, (S4) COIB-OTU table, (S5) targets for new barcodes, (S6) morphological validation, (S7) additional barcodes, (S8) replicates in metabarcoding, (S9) GLM results, (S10) rarefaction curves.

Supplementary Materials 3. Bioinformatics pipeline used in this study.