RESOURCE ARTICLE

Enhancing metabarcoding efficiency and ecological insights through integrated taxonomy and DNA reference barcoding: A case study on beach meiofauna

Jan-Niklas Macher[1,2] | Alejandro Martínez[3] | Sude Çakir[1] | Pierre-Etienne Cholley[1] | Eleni Christoforou[4,5] | Marco Curini Galletti[6,7] | Lotte van Galen[1] | Marta García-Cobo[8] | Ulf Jondelius[9] | Daphne de Jong[1] | Francesca Leasi[10] | Michael Lemke[1] | Iñigo Rubio Lopez[3,11,12] | Nuria Sánchez[8] | Martin Vinther Sørensen[13] | M. Antonio Todaro[7,14] | Willem Renema[1,15] | Diego Fontaneto[3,7]

[1]Naturalis Biodiversity Center, Marine Biodiversity, Leiden, The Netherlands

[2]Department of Environmental Biology, Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands

[3]National Research Council of Italy (CNR), Water Research Institute (IRSA), Verbania Pallanza, Italy

[4]Cyprus Marine and Maritime Institute, CMMI House, Larnaca, Cyprus

[5]Department of Chemical Engineering, Cyprus University of Technology, Limassol, Cyprus

[6]Dipartimento di Medicina Veterinaria, Università di Sassari, Sassari, Italy

[7]National Biodiversity Future Center (NBFC), Palermo, Italy

[8]Department of Biodiversity, Ecology and Evolution, Universidad Complutense de Madrid, Madrid, Spain

[9]Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden

[10]Department of Biology, Geology, and Environmental Science, University of Tennessee at Chattanooga, Chattanooga, Tennessee, USA

[11]Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic

[12]Department of Ecology, Faculty of Science, Charles University, Prague 2, Czech Republic

[13]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

[14]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

[15]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Jan-Niklas Macher, Naturalis Biodiversity Center, Marine Biodiversity, Darwinweg 2, 2333 CR Leiden, The Netherlands.
Email: jan.macher@naturalis.nl

Abstract

Molecular techniques like metabarcoding, while promising for exploring diversity of communities, are often impeded by the lack of reference DNA sequences available for taxonomic annotation. Our study explores the benefits of combining targeted DNA barcoding and morphological taxonomy to improve metabarcoding efficiency, using beach meiofauna as a case study. Beaches are globally important ecosystems and are inhabited by meiofauna, microscopic animals living in the interstitial space between the sand grains, which play a key role in coastal biodiversity and ecosystem dynamics. However, research on meiofauna faces challenges due to limited taxonomic expertise and sparse sampling. We generated 775 new cytochrome c oxidase I DNA barcodes from meiofauna specimens collected along the Netherlands' west coast and combined them with the NCBI GenBank database. We analysed alpha and beta diversity in 561 metabarcoding samples from 24

North Sea beaches, a region extensively studied for meiofauna, using both the enriched reference database and the NCBI database without the additional reference barcodes. Our results show a 2.5-fold increase in sequence annotation and a doubling of species-level Operational Taxonomic Units (OTUs) identification when annotating the metabarcoding data with the enhanced database. Additionally, our analyses revealed a bell-shaped curve of OTU richness across the intertidal zone, aligning more closely with morphological analysis patterns, and more defined community dissimilarity patterns between supralittoral and intertidal sites. Our research highlights the importance of expanding molecular reference databases and combining morphological taxonomy with molecular techniques for biodiversity assessments, ultimately improving our understanding of coastal ecosystems.

KEYWORDS
community ecology, DNA barcoding, invertebrates, Molecular reference database

1 | INTRODUCTION

Metabarcoding has become a standard technique for the study of biological communities and a valuable tool for understanding biodiversity and ecology of species and communities (Ficetola & Taberlet, 2023; Taberlet et al., 2012). Despite technological advances, the efficiency of metabarcoding studies and their usability for ecological analyses depends on the completeness of reference sequence databases. These databases, such as NCBI GenBank (Benson et al., 2012) and Barcode of Life Data System (BOLD; Ratnasingham & Hebert, 2007) contain reference DNA barcodes (Hebert et al., 2003) generated from individual, taxonomically identified specimens that serve as reference against which the metabarcoding sequences can be compared and identified. The availability of such references greatly varies across taxonomic groups and ecosystems, remaining particularly poor for species-rich but understudied taxa and regions (Keck et al., 2023; Leasi et al., 2018; Leray & Knowlton, 2015; Weigand & Macher, 2018). A promising approach to mitigate this issue lies in the targeted collection, identification, and sequencing of a high number of specimens, followed by their addition to molecular reference databases. This method, sometimes implemented through taxonomic expert workshops, has been successfully employed across diverse taxonomic groups and ecosystems (Behrens-Chapuis et al., 2021; Creedy et al., 2020; De Souza Amorim et al., 2023; Dugal et al., 2022; Emerson et al., 2023; Zhou et al., 2011). Since well-studied and well-described taxa are important for the identification and classification of species, and for evaluating the utility of reference databases, involving taxonomic experts in the build-up and improvement of molecular reference databases is of utmost importance.

Using beach meiofauna as a case study, we investigate the effect of targeted sampling and expert identification on improving metabarcoding efficiency by increasing the completeness of reference libraries. Beach meiofauna comprises small animals inhabiting the interstitial spaces between sand grains. They form highly diverse communities, providing important ecosystem services and are valuable for the understanding of biodiversity and ecosystem dynamics of coastal areas (Felix et al., 2016; Schratzberger & Ingels, 2018). Beaches are one of the Earth's most dynamic environments, which is threatened by sea level rise and other human impacts (Lansu et al., 2024; Schlacher et al., 2007). Studying meiofauna is challenging due to their generally microscopic size and the scarcity of taxonomists for most of the world's geographic regions and taxonomic groups (Leasi & Norenburg, 2014; Zeppilli et al., 2015). This taxonomic impediment means many beach meiofauna species remain undescribed even in intensively sampled areas (Curini-Galletti et al., 2012; Martinez et al., 2023). Advancements in DNA metabarcoding have drastically improved our ability to detect species in complex communities (Ficetola & Taberlet, 2023), and after pioneering studies demonstrating the use of DNA metabarcoding for marine meiofauna (Creer et al., 2010; Fonseca et al., 2010), the technique now allows for more detailed studies and gaining a better understanding of the impact of environmental factors and stressors on meiofauna (Atherton & Jondelius, 2020; Gielings et al., 2021; Martínez et al., 2020).

Here we introduce a dataset comprising 775 newly generated cytochrome c oxidase I reference sequences ('DNA barcodes') derived from meiofauna specimens collected along the Netherlands' west coast. This region, a part of the southern North Sea, is one of the best-studied areas in marine meiofaunal research (Germán Rodríguez, 2004; Gray & Rieger, 1971; Kotwicki et al., 2005). By analysing a total of 576 metabarcoding samples, we study the influence of enhanced reference databases on inferred species richness and community composition patterns, both on a local (across the intertidal zone) and regional (spanning 650km of the North Sea Coast) scales. By demonstrating major improvements in identification of meiofauna in metabarcoding data through reference database enhancement, our study highlights the critical need for combining taxonomic expertise with reference sequencing even in well-studied areas. Due to the massive improvement demonstrated by the use of a local reference library in one of the best-studied areas in the world, the improvement that can be obtained in poorly studied, species-rich areas will be much stronger. We seek to inspire similar initiatives in biodiversity research and advocate

for combining traditional taxonomic methods and modern genetic techniques to improve the understanding of biodiversity in important, but often overlooked taxonomic groups and ecosystems.

2 | MATERIALS AND METHODS

2.1 | Sampling for reference barcoding

Meiofauna specimens were collected during two taxonomic workshops held at Naturalis Biodiversity Center in Leiden, the Netherlands, in May/June 2022 and July/August 2023. Samples were collected from 20 locations along the Dutch West Coast, by either taking sediment cores to a depth of 10 cm, by filtering coastal groundwater and sand through dug holes, or by scraping hard substrates, depending on the targeted taxonomic group. The samples were transported to the Naturalis laboratory for meiofauna extraction using decantation through a 40 μm sieve after anesthetization with isosmotic $MgCl_2$ (Somerfield & Warwick, 2013). For a detailed list of all sample locations and sample types, see Table S1. After extraction, meiofauna specimens were sorted and pictured alive using stereo and light microscopes. They were identified to the lowest possible taxonomic rank using a combination of identification keys for the respective taxa (see Schmidt-Rhaesa, 2020 and references therein for an overview) and the expertise of multiple taxonomic experts. Specimens were transferred to PCR plates and submitted to the DNA extraction pipeline described below.

2.2 | DNA extraction and amplification for reference barcoding

The DNA extraction process for meiofauna specimens was performed using the Macherey-Nagel (Düren, Germany) NucleoSpin tissue kit on the KingFisher (Waltham, USA) robotic platform, following the manufacturer's protocol. After extraction, PCRs were performed with Geller COI primers (Geller et al., 2013), targeting the 658 base-pair-long Folmer fragment of the mitochondrial cytochrome c oxidase I gene, which is the commonly used DNA barcode for animals (Folmer et al., 1994; Hebert et al., 2003). The protocol was as follows: Each PCR reaction contained 10.2 μL of MiliQ water, 4 μL of 5× PCR buffer (Qiagen; Hilden, Germany), 0.8 μL of 10 mg/mL BSA (Promega, Madison, Wisconsin, United States), 1 μL of 10 picomolar/μL primers, 0.4 μL of 2.5 mM dNTPs, 0.4 μL of 5 U/μL Phire II Taq polymerase (Thermo Fisher, Waltham, Massachusetts, United States), and 1 μL template DNA. PCR started with an initial denaturation of 30 s at 98°C, followed by 35 cycles of 5 s denaturation at 98°C, 10 s annealing at 50°C, 15 s elongation at 72°C, and a final extension of 5 min at 72°C. Each PCR included a negative control using Milli-Q water (Merck; Rahway, New Jersey, United States) instead of template DNA. We cleaned the PCR products using AmPure magnetic beads (Brea,

California, United States) with a ratio of 0.9:1. Subsequently, a second PCR was performed to individually label the samples, with 2.5 μL of ONT barcode primers, 5 μL of LongAmp Taq 2× master mix (New England Biolabs, Ipswich, Massachusetts, United States), and 2.5 μL of the PCR product as template. The PCR protocol for the second amplification was as follows: an initial denaturation of 3 min at 95°C, followed by 15 cycles of 15 s denaturation at 95°C, 15 s of annealing at 65°C, 50 s of elongation, and a final extension step of 3 min at 65°C. The PCR success was checked using the TapeStation platform. The samples were then pooled at equimolar concentrations to achieve a final concentration of approximately 200 femtomolar. This was followed by a purification step using a 0.7:1 bead clean-up, targeting amplicons of 700 bp length. Finally, the purified DNA pools were eluted in 11 μL of nuclease-free water, and their concentration was quantified using the Tape station (D5000 kit).

2.3 | Sequencing and bioinformatics for reference barcoding

Sequencing was conducted using the Oxford Nanopore GridION sequencer on two FLO-MIN112 flow cells, with the SQK-NBD112.24 sequencing kit. The Basecalling was done with MinKNOW (v23.04.5), the run duration was set to 72H, and super accuracy basecalling was selected. The demultiplexing was performed with Guppy barcoder (v6.4.6). The consensus calling consisted of several steps combined together in a Snakemake (Mölder et al., 2021) pipeline: First, the reads (containing primers at both ends) were filtered by size ($>= 558$, $<= 758$) and quality ($>= 10$), and then reoriented with Cutadapt (v4.5, max error rate 20%, 80% coverage), which also removed flanking sequences. Then consensus sequences were generated using NGSpeciesID v0.3.0 (Sahlin et al., 2021) with Medaka polishing (v1.8.0, r104_e81_sup_g5015 model). A final round of primer sequence trimming was performed with Cutadapt. Following this, multi-fasta files containing consensus sequences were written by using a custom script. Quality control and visualization of the processed FASTQ files were conducted using NanoPlot (De Coster et al., 2018) and MultiQC (Ewels et al., 2016). All resulting sequences underwent manual curation in Geneious Prime (version 2023.2) and were searched against existing references in NCBI GenBank using BLASTn (Ye et al., 2006). All scripts used for processing of Nanopore data are available on GitHub: https://gitlab.com/arise-biodiversity/sequencing/arise -barcoding-pipeline/-/tree/1a2fa544615be54dccb7136ca20d36 64ba85d467.

2.4 | Sampling and environmental variable measurement for metabarcoding

We collected meiofauna samples for metabarcoding from 24 sandy beaches along the Dutch and German Coast during the summers of
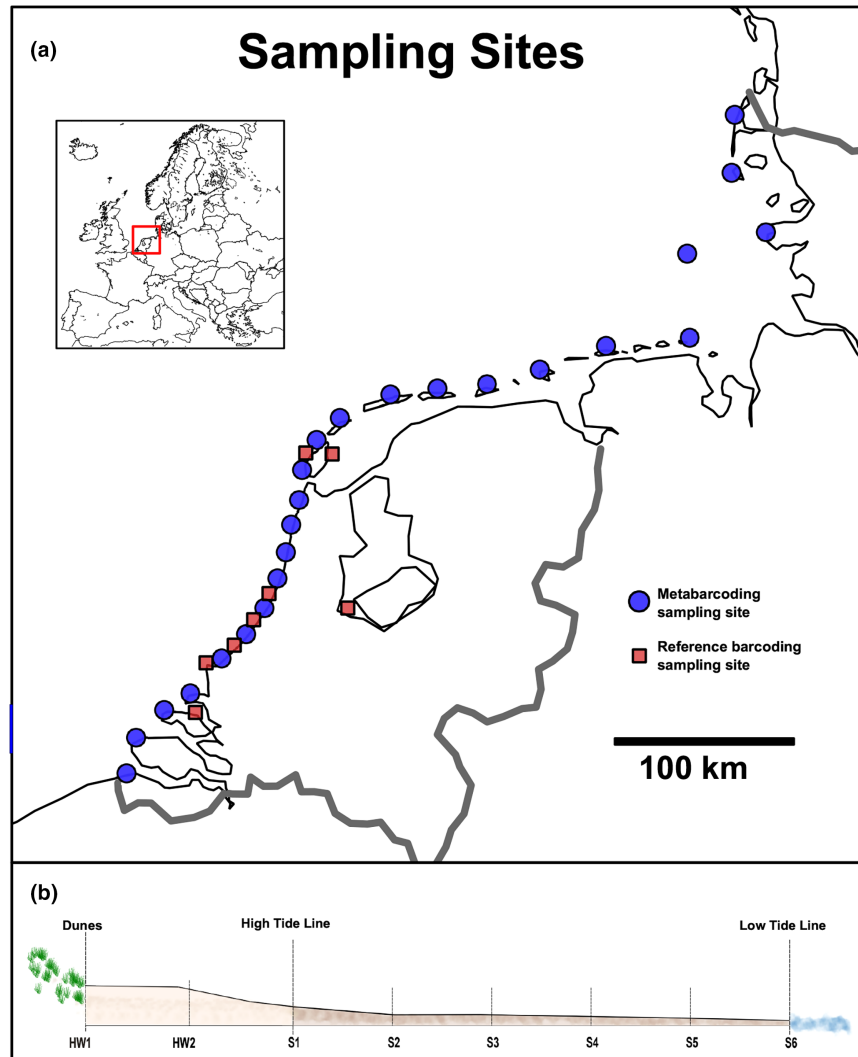
(a) Sampling Sites

Metabarcoding sampling site

Reference barcoding sampling site

100 km

(b)

Dunes — High Tide Line — Low Tide Line

HW1 HW2 S1 S2 S3 S4 S5 S6

FIGURE 1 (a) Map showing the location of the 24 beaches sampled for metabarcoding (blue circles) and locations sampled for reference barcoding (red squares). Note that reference barcoding sites in close proximity to each other are not shown separately due to the map scale. The mini map shows the locations of the sampling area in Europe. (b) Schematic view of a beach showing the location of the eight sampling points per transect from dunes to the low tide line.

2021 and 2022. See Figure 1 for a map and Table S2 for the coordinates of the sampled beaches. Sampling was conducted at maximum low tide. At each beach, we sampled along three parallel transects, with eight sampling sites each. The first sampling site was located at the foot of the dunes, the second midway between the dunes and the high-tide line, and the remaining six samples were evenly spaced from the high-tide line to the low-tide line. Following McLachlan's classification (McLachlan et al., 2018), most sampled beaches were tide-modified. Two beaches were wave-dominated (Relative Tide Range Index (RTR) of <3), and two beaches were tide-dominated (RTR >10). At each sampling site along the transects, two sediment cores were collected using sterile plastic syringes: one with a 5 cm diameter and 10 cm length (approximate volume of 200 mL), and another core with 1 cm diameter and 10 cm length (approximate volume of 8 mL). The smaller core was immediately transferred into a 50 mL Falcon tube, while the larger core was stored in a sterile 1-litre plastic bottle. On the beach, we extracted meiofauna from the larger core using the $MgCl_2$ decantation method, by adding 500 mL of isosmotic $MgCl_2$ solution to the sediment. After a 5-min incubation, the sediment-$MgCl_2$ mixture was swirled ten times, and the supernatant containing meiofauna was decanted through a 1 mm and

40 μm sieve cascade, a common practice in beach meiofauna studies (Castro et al., 2021; Haenel et al., 2017; Martínez et al., 2020). The meiofauna retained on the 40 μm sieve was then rinsed into sterile 15 mL Falcon tubes and preserved in 10 mL of 96% ethanol. All samples were subsequently transported to the Naturalis Biodiversity Centre laboratory and stored at −20°C until processing. The sediment from the smaller core was dried for grain size measurement using a LS13320 Particle Size Analyser (Beckman-Coulter, USA).

2.5 | DNA extraction, amplification and sequencing for metabarcoding

We extracted DNA from dried meiofauna samples after evaporating the ethanol at 50°C overnight in a sterile warming cabinet and transferring the dried samples to 2 mL Eppendorf tubes. DNA extraction was performed using the Macherey Nagel NucleoSpin Soil kit (Macherey Nagel, Düren, Germany) following the standard protocol including bead beating, but with an additional overnight Proteinase K digestion step (50 μL 250 μg/mL ProtK, Thermo Fisher Scientific, Waltham, USA) added to the lysis buffer provided with

the kit to improve cell lysis, as done in previous studies on meiofauna (Martínez et al., 2020; Weigand & Macher, 2018).

For community metabarcoding, we amplified meiofauna DNA using a two-step PCR protocol with the widely-used LerayXT primers targeting a 313 base pair region of the mitochondrial cytochrome c oxidase I (COI) gene of a broad range of Eukaryota (Leray et al., 2013; Wangensteen et al., 2018). The first PCR reaction contained 11.7 μL mQ water, 2 μL Qiagen CL buffer (10×; Qiagen, Hilden, Germany), 0.4 μL MgCl$_2$ (25 mM; Qiagen), 0.8 μL Bovine Serum Albumine (BSA, 10 mg/mL), 0.4 μL dNTPs (2.5 mM), 0.2 μL Qiagen Taq (5 U/μL), 1 μL of each nextera-tailed primer (10 pmol/μL), and 2.5 μL of DNA template. PCR amplification involved an initial denaturation at 96°C for 3 min, followed by 30 cycles of denaturation for 15 s at 96°C, annealing at 50°C for 30 s, and extension for 40 s at 72°C, concluding with a final extension at 72°C for 5 min. We processed six negative controls (Milli-Q water, Merck, Kenilworth, USA) alongside the samples to check for potential contamination. After the first PCR, samples were cleaned with AMPure beads (Beckman Coulter, Brea, United States) at a 0.9:1 ratio according to the protocol to remove short fragments and primer dimers. The second PCR involved amplification with individually tagged primers, following the same protocol as above and using the PCR product from the first PCR as the template, but reducing the PCR cycle number to 10. We measured DNA concentrations using the Fragment Analyser (Agilent Technologies, Santa Clara, CA, USA) with the High Sensitivity Kit and pooled samples equimolarly. The final library was cleaned with AMPure beads as described above and sent for sequencing on three Illumina MiSeq runs (paired-end, 300 bp read length) at Baseclear (Leiden, The Netherlands).

2.6 | Bioinformatic processing of metabarcoding data

The raw metabarcoding reads were processed using APSCALE (Buchner et al., 2022) with the following settings: maximum differences in percentage: 20; minimum overlap: 50, minimum read sequence length: 310 bp; maximum read length: 316 bp, minimum size to pool: 20 sequences. Sequences were clustered into both ESVs (setting: alpha 1, minimum 20 sequences) and OTUs, the latter with a sequence similarity threshold of 97%. To account for potential low-level contamination or tag jumping common on Illumina platforms (Schnell et al., 2015), we removed OTUs with an abundance of <0.03% of reads per sample, and OTUs that were present in less than six out of 561 samples (<1% occurrence). The taxonomic assignment was first performed using NCBI GenBank expanded with reference COI barcodes generated as part of taxonomic expert workshops in Leiden. Taxonomic ranks were assigned to OTUs using established identity thresholds: >97%: species, >95%: genus, >90%: family, >85% order (Macher et al., 2023). OTUs that were assigned with less than 85% identity to a reference or identified as non-meiofauna taxa were excluded from further analyses. The three replicates per tidal level per beach were merged into one composite sample to account for potential variability within tidal levels, resulting in 190 composite samples.

2.7 | Analysing the increase in annotation efficiency by enhancing the reference database

We evaluated the efficiency of taxonomic assignment by calculating the percentage of reads and the number of Operational Taxonomic Units (OTUs) assigned to species, genus, family, order, and class when using NCBI GenBank for identification, and when enhancing the database with the newly generated DNA barcodes. For this, we downloaded NCBI GenBank, conducted a nucleotide blast (blastn) for annotation, and then a second blastn annotation with the enhanced database. All analyses were performed on the Naturalis Biodiversity Center High Performance Computing Cluster.

2.8 | OTU richness and community similarity across the intertidal zone

We calculated the OTU richness for each tidal level, both for the dataset annotated only with the NCBI database, and for the dataset annotated with the enhanced database. We created point plots showing the increase in OTU number per taxonomic group and tidal level using the ggplot2 package in R, and used paired t-tests to test for significance in difference of OTU numbers. Further, we visualized the total number of OTUs at each tidal level using Cumming estimation plots using the 'dabestr' package, and used estimation statistics (Ho et al., 2019) to assess the mean difference between OTU numbers between the two datasets. We calculated the proportion of OTU numbers per taxonomic group and tidal level and computed stacked bar plots using ggplot2 in R. To analyse differences in taxonomic composition between the datasets at each tidal level, we applied chi-square tests, incorporating a Monte Carlo simulation for p-value estimation (simulate.p.value = TRUE) to address the small sample sizes and low expected frequencies.

We analysed the distribution of OTUs across the beach transect from dunes to the low tide line, both for the dataset annotated only with NCBI GenBank and the dataset annotated with the enhanced database. Tidal levels were ordered categorically from dune level (HW1) to low water level (S6) to reflect the natural gradient. To quantify the relationship between OTU richness and both tidal levels and the Relative Tide Range Index (RTR), we employed generalized linear mixed-effects models (GLMMs). This approach enabled us to assess the statistical significance of linear and potentially nonlinear relationships affecting OTU richness across tidal levels and RTR, while also accounting for potential within-beach correlation in OTU richness. We tested for overdispersion and to ensure the model represented the underlying data structure and OTU count variability, we selected Poisson

distribution for the dataset annotated with NCBI only, and a negative binomial distribution for the dataset annotated with the enhanced database.

We calculated beta diversity using the Jaccard index, based on presence-absence data, and used the adonis2 PERMANOVA implemented in the 'vegan' package in R to assess the influence of the tidal level, that is, the distance from the low tide, and the Relative Tide Range Index (RTR), that is, the beach type, on community composition, using 'Beach' as a stratifying factor. Further, we calculated Non-metric Multidimensional Scaling plots (NMDS) for visualization. Subsequently, we used a Mantel test to assess the correlation and significance of the similarity between these datasets.

2.9 | Impact of proximity to reference barcoding sites on meiofauna OTU richness in metabarcoding data

We tested whether the distance between sites sampled for reference barcoding during the taxonomic workshops and the sites sampled for metabarcoding significantly influenced the increase in the number of OTUs identified as meiofauna in our metabarcoding dataset. To do this, we employed a linear modelling approach in R. We first calculated the geographic distances of metabarcoding sampling sites to the nearest site sampled for reference barcoding using Haversine straight-line distances with the R library 'geosphere' (Hijmans et al., 2017). Then, we constructed linear models with the lm function in R to explore the relationship between the percentage increase in OTU richness and the distance to the nearest reference site. This was repeated for all eight tidal levels.

3 | RESULTS

We structure the results into two main sections. First, we report the striking difference in taxonomic classification when enhancing the reference library with the newly generated COI barcodes. Second, we examine the impact of this enhanced reference library on ecological interpretations when analysing the same metabarcoding dataset.

3.1 | Reference DNA barcoding

We sequenced 775 mitochondrial cytochrome c oxidase I (COI) DNA barcodes of meiofaunal specimens following two taxonomic workshops in Leiden in the years 2022 and 2023. All specimens were identified at least on class level, with the exception of Platyhelminthes, which were classified at least to the subphylum level. For 643 specimens (82.97%), identification was achieved at least to the order level. Additionally, 490 specimens (63.23%) were identified to at least the family level, 383 specimens (49.42%) to the genus level, and 164 specimens (21.16%) to the species level. The

majority of barcoded specimens were Copepoda (143), followed by Chromadorea and Enoplea (Nematoda), with 137 and 116 specimens respectively. Rhabditophora (Platyhelminthes) and Polychaeta were present with 80 and 75 specimens, respectively, followed by Gastrotricha (63 specimens), Acoela (41), Clitellata (30), Rotifera (28), Arachnida (20), Collembola (14), Eutardigrada (7), Heterotardigrada (7), Palaeonemertea (4), Ostracoda (3), Branchiopoda (1), and Gnathostomulida (1). In addition, we found and sequenced four specimens of Malacostraca, and one Asteroidea larva. See Table S3 for the complete taxonomic list of barcoded specimens and COI barcode sequences. The specimen list is also available as a GBIF dataset: https://doi.org/10.15468/gemfv4.

3.2 | Increase in annotation efficiency of metabarcoding data by enhancing the reference database

From the 561 samples analysed through metabarcoding, we retained 14,822,456 sequences and 566 OTUs after bioinformatic processing and quality filtering. Merging the three biological replicates per tidal level resulted in 190 composite samples. Using only NCBI GenBank for annotation, 4,633,286 sequences in 114 OTUs were assigned to meiofaunal taxa at least at the phylum level. Using the enhanced database, 11,361,563 sequences in 188 OTUs were identified as meiofauna at least at the phylum level. This corresponds to an increase of 145% in reads and an increase of 65% in OTUs annotated to meiofauna at least at the phylum level. The majority of non-meiofaunal reads were assigned to Bacillariophyta (2,652,322 reads, corresponding to 76.64% of non-meiofaunal reads). The number of OTUs assigned to order level increased by 63%, from 115 to 188. On the family level, we found an increase of 160%, from 53 to 138 OTUs. The number of OTUs assigned to at least genus level increased by 202%, from 40 to 121, and the number of OTUs annotated to species level increased by 205%, from 36 to 110.

The OTU number increased for all taxonomic groups except Branchiopoda (Arthropoda) and Pilidiophora (Nemertea), with the strongest increase observed in Copepoda (Arthropoda), Chromadorea and Enoplea (both: Nematoda). Commonly, we found the strongest increase in number of OTUs in the middle to lower intertidal zone (sampling sites S3 to S6), with the exception of Clitellata (Annelida), Arachnida, and Collembola (both: Arthropoda), for which we found the strongest increase in OTU numbers in the supralittoral zone (sampling sites HW1, HW2) and the upper intertidal zone (sampling sites S1, S2). Heterotardigrada OTUs were only found in the dataset annotated with the enhanced database (Figure 2). See Table S4 for paired t-test results.

3.3 | Taxonomic composition in metabarcoding data

Comparing the dataset annotated with only NCBI reference barcodes and the enhanced database, we found a consistent distribution
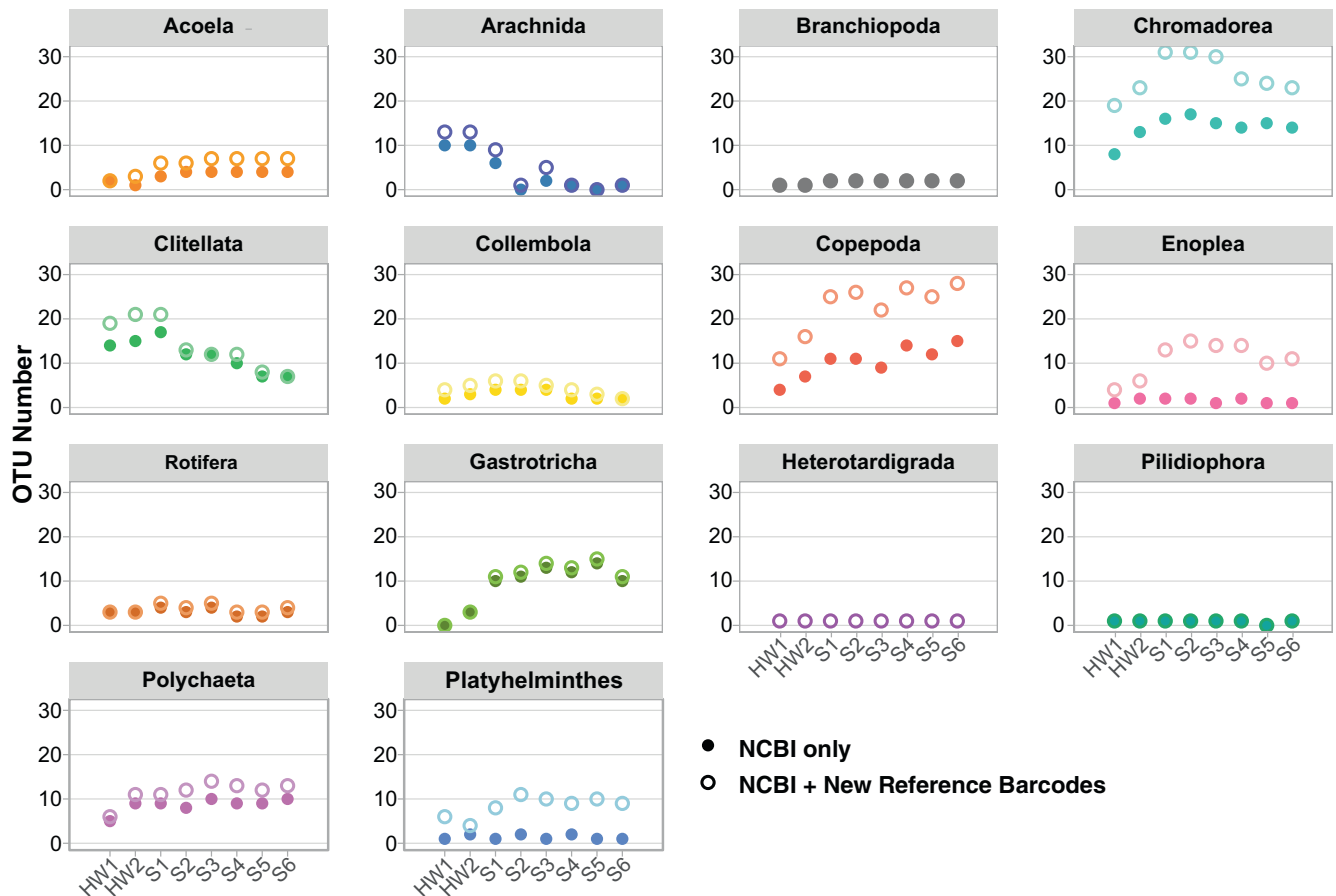
FIGURE 2 Increase of OTU number by taxonomic group and tidal level. Supralittoral sampling sites are labelled HW1 and HW2, and intertidal sampling sites are labelled S1 to S6. Filled circles indicate the number of OTUs identified using only NCBI GenBank for annotation of metabarcoding data, and open circles indicate the number of OTUs identified when combining NCBI GenBank with the new reference barcodes.

of OTU proportions per taxonomic group across the intertidal zone. Enoplea (Nematoda) exhibited an increased proportion across all tidal levels when annotated with the expanded reference dataset, and a similar trend was observed for Platyhelminthes. Additionally, Heterotardigrada, while present in low proportions, was exclusively identified in the dataset annotated with the enhanced database (Figure 3). However, chi-square tests showed that overall proportions of taxonomic groups per tidal level did not differ significantly between the datasets (see Table S5 for results).

3.4 | OTU richness along beach transects

The enhancement of the molecular reference database significantly improved the assignment of Operational Taxonomic Units (OTUs) to meiofauna taxa for all tidal levels. The most notable increase was observed in tidal levels S3, S4, and S5, while the lowest increase was found for tidal levels HW1 and HW2. See Figure 4 for graphical representation and Table 1 for statistical results.

The analysis of OTU richness across beach transects, from dunes to the low tide line, revealed distinct patterns when comparing the

datasets annotated only with NCBI GenBank and those annotated with the enhanced database.

For the dataset annotated only with NCBI GenBank references, the GLMM revealed significant linear ($p=.002$) and quadratic ($p=.003$) effects of tidal levels on OTU numbers, indicating a non-linear distribution of OTU richness across the tidal gradient. The positive linear term suggests an increase in OTU numbers towards the middle and lower intertidal level, and the negative quadratic term indicates that this increase peaks at intermediate levels, rather than continuing linearly across the entire gradient. The random effects component attributed to beach location accounted for a variance of 0.037 (standard deviation: 0.1928), and the model's AIC and BIC were 1044.2 and 1076.7, respectively.

The analysis of the dataset annotated the enhanced database showed a more pronounced non-linear (bell-shaped) pattern in OTU richness across tidal levels, with highly significant linear ($p<.001$) and quadratic ($p<.001$) relationships. This model accounted for a greater proportion of variability in OTU numbers, with a variance of 0.045 (standard deviation: 0.212) attributed to beach location, and the model's AIC and BIC were 1278.8 and 1314.5, respectively. The influence of the Relative Tide Range Index (RTR) on OTU richness
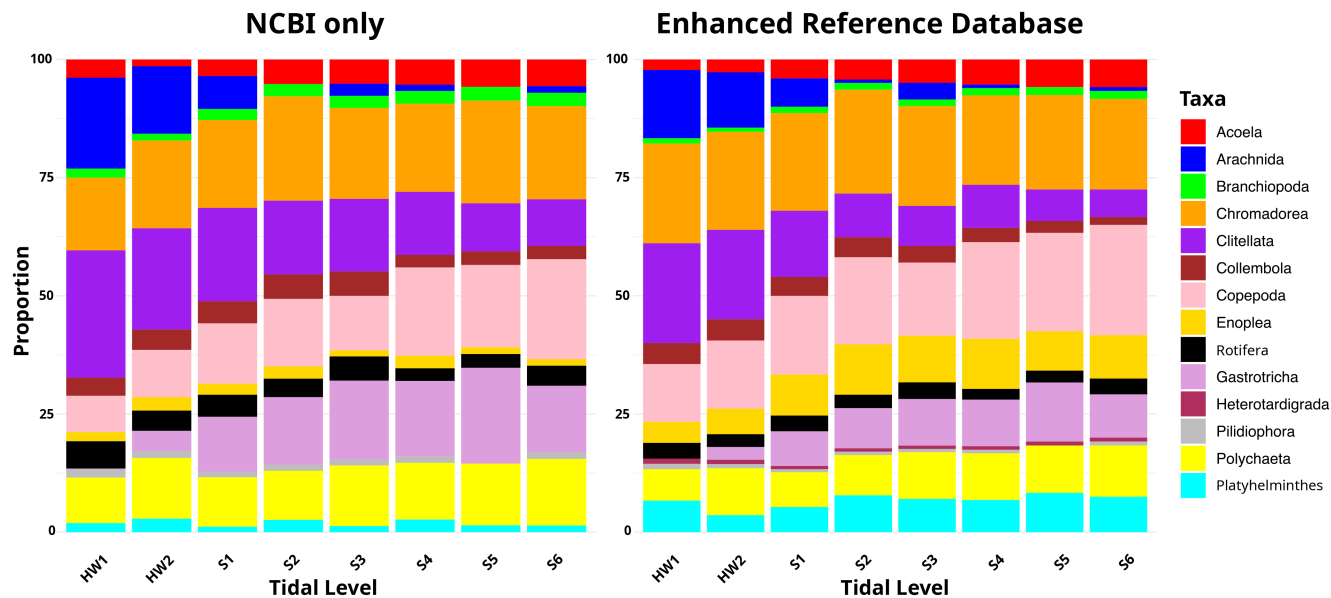
FIGURE 3 Stacked bar plots showing the proportions of meiofauna OTUs per taxonomic group across the beach transect. The left plot represents the dataset annotated using only NCBI reference barcodes and the right plot represents the dataset annotated with the enhanced database. Each bar represents a tidal level (HW: Sublittoral sites, S: Intertidal sites), with the vertical stacking indicating the proportion of OTUs.
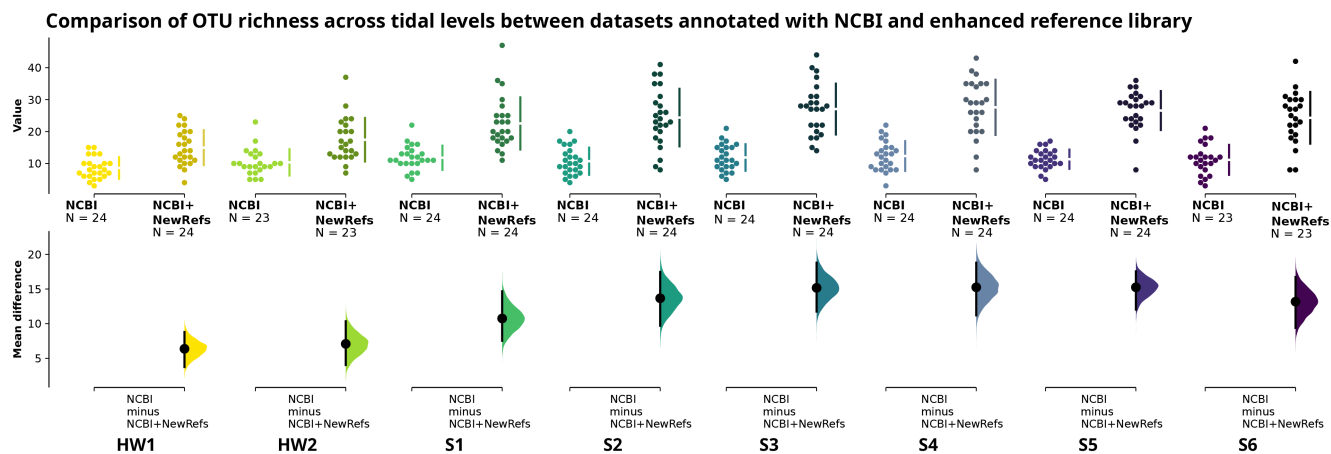

FIGURE 4 Cumming estimation plot showing the mean differences for OTU richness in the eight analysed tidal levels, between the dataset annotated with only NCBI reference data, and annotated with the enhanced database. The raw data is plotted on the upper axes; each mean difference is plotted on the lower axes as a bootstrap sampling distribution. Mean differences are depicted as dots; 95% confidence intervals are indicated by the ends of the vertical error bars.

was not statistically significant in either dataset ($p > .05$), indicating that within the context of these models, RTR did not play a significant role in shaping OTU richness.

3.5 | Community composition

The analysis of community composition using Jaccard dissimilarity showed that in the dataset annotated only with NCBI data, the tidal level significantly influenced community composition ($R^2 = .127$, $p < .001$). However, the R^2 value of .127 indicates that the explained variance is limited. The RTR, characterizing beach type, also showed a significant but smaller effect ($R^2 = .019$, $p = .001$). The variability attributed to different beaches was significant ($R^2 = .273$, $p < .001$) and had the highest explanatory value. A large proportion of variability in community composition remained unexplained (Residual $R^2 = .581$).

For the dataset annotated with the enhanced database, we found a more substantial influence of the tidal level on community composition ($R^2 = .155$, $p < .001$). The effect of RTR remained similar to that observed in the NCBI-only dataset ($R^2 = .018$, $p < .001$). The variability attributed to different beaches was slightly more pronounced in this dataset ($R^2 = .28$, $p < .001$). The residual variability in this dataset was slightly lower (Residual $R^2 = .547$), indicating that

TABLE 1 Unpaired mean differences, 95% confidence intervals (CI), and p values from two-sided permutation t-tests comparing OTU richness per tidal levels between the dataset annotated with NCBI data only, and the dataset annotated with the enhanced database.

Tidal level	Unpaired mean difference	95% CI	p Value
HW1	6.38	3.75, 8.79	<.001
HW2	7.09	4.04, 10.3	<.001
S1	10.8	7.54, 14.7	<.001
S2	13.7	9.71, 17.5	<.001
S3	15.2	11.7, 18.8	<.001
S4	15.2	11.2, 18.8	<.001
S5	15.2	12.0, 17.5	<.001
S6	13.2	9.39, 16.7	<.001

Note: Results are based on 5000 bootstrap samples to assess the statistical significance of observed differences, with p values reflecting the probability of observing the effect sizes under the null hypothesis of no difference between groups.

TABLE 2 Summary of adonis2 PERMANOVA results comparing the effects of tidal level, relative tide range (RTR), and beach on inferred community composition using only NCBI and NCBI + new reference barcodes for annotation of metabarcoding data.

Test	df	SumOfSqs	R^2	F	Pr (>F)
NCBI only					
Tidal level	7	9.715	.127	4.9756	<0.001
RTR	1	1.420	.019	5.0896	<0.001
Beach	22	20.867	.273	3.4004	<0.001
Residual	159	44.350	.581		
NCBI+new references					
Tidal level	7	11.222	.155	6.4542	<0.001
RTR	1	1.283	.018	5.1640	<0.001
Beach	22	20.187	.28	3.6943	<0.001
Residual	159	39.493	.547		

the inclusion of new reference barcodes slightly improved the model's explanatory power. See Table 2.

The Non-metric Multidimensional Scaling (NMDS) plots further underline these findings. In both datasets, samples collected from the highest tidal levels (HW1 and HW2) loosely cluster together, distinguishing them from samples taken from the middle to lower intertidal zones (S3–S6). Samples from the upper intertidal or swash zone (S1 and S2) cluster between and within these two groups. The NMDS plot based on the dataset annotated with the enhanced database shows a clearer separation between the supralittoral sampling sites and the intertidal communities (see Figure 5).

The Mantel test results showed a strong and statistically significant correlation between the two distance matrices, with a Mantel statistic (r) of .7848 and a p-value of .001.

3.6 | Influence of distance from reference barcoding sites on the increase of identified meiofauna OTUs in metabarcoding data

We found negative correlations between the increase in identified meiofauna OTUs and the distance of the sampling site to a site sampled for reference barcoding. This was true for sampling sites in the tidal levels S6 (coefficient=−0.0005, p=.02), S4 (coefficient=−0.0002, p=.04), S3 (coefficient=−0.0003, p=.05), S2 (coefficient=−0.0004, p=.01) and S1 (coefficient=−0.0003, p=.01). We observed a negative, but non-significant, trend for tidal level S5 (coefficient=−0.0003, p=.06). No correlations were found in the higher tidal levels HW1 (coefficient=0.00003, p=.6) and HW2 (coefficient=−0.000001, p=.98) (Figure 6).

4 | DISCUSSION

We tested the impact of enriching molecular reference databases on the efficiency of metabarcoding analyses, using beach meiofauna, an important yet frequently overlooked component of coastal biodiversity, as a case study. Our findings reveal an impressive improvement in meiofauna identification across all taxonomic levels following two targeted reference sampling campaigns conducted with taxonomic experts. The inclusion of a local reference library enabled us to significantly enhance metabarcoding efficiency. This improvement occurred in one of the world's most thoroughly researched areas for beach meiofauna, and we expect that the impact in other regions and habitats will be even greater.

4.1 | Improved metabarcoding efficiency through enhanced molecular reference databases

We show that with a local reference library, 11,361,563 instead of just 4,633,286 sequences could be annotated to meiofauna, an increase of 6,728,277 sequences (145%) that would otherwise remain unassigned. On the genus and species level, reference barcoding increased the number of identified OTUs by over 200%. We identified more OTUs at all tidal levels, from dunes to the low tide line, and showed that patterns of community dissimilarity between tidal levels become more clear after adding new reference barcodes. Spatial analysis revealed that the enhanced database improved OTU assignment across all tidal levels. Our results of a clear separation in community composition between supralittoral areas of the beach and areas in the lower intertidal zone, with samples from the swash zone around the high tide line falling in between, are in line with previous findings on beach meiofauna (Pereira et al., 2017). Our results show, however, that this pattern was more evident after adding new reference barcodes. We show that enhancing molecular reference databases significantly boosts metabarcoding efficiency for meiofauna across various ecological niches and geographic regions. Our results
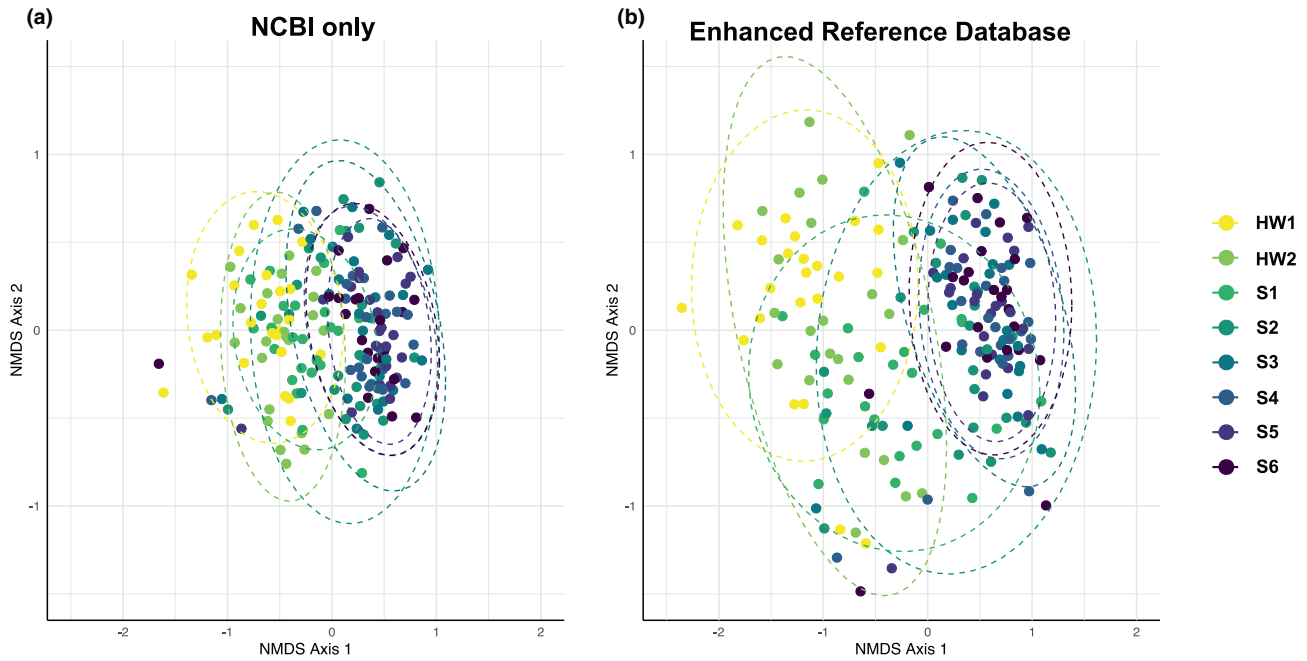
FIGURE 5 Non-metric Multidimensional Scaling (NMDS) ordination plots comparing community compositions across different tidal levels, based on Jaccard dissimilarity. (a) Dataset annotated with NCBI reference data; (b) Dataset annotated with the enhanced database. Points represent individual samples, coloured by their respective tidal level. Groups are outlined by 95% confidence ellipses.
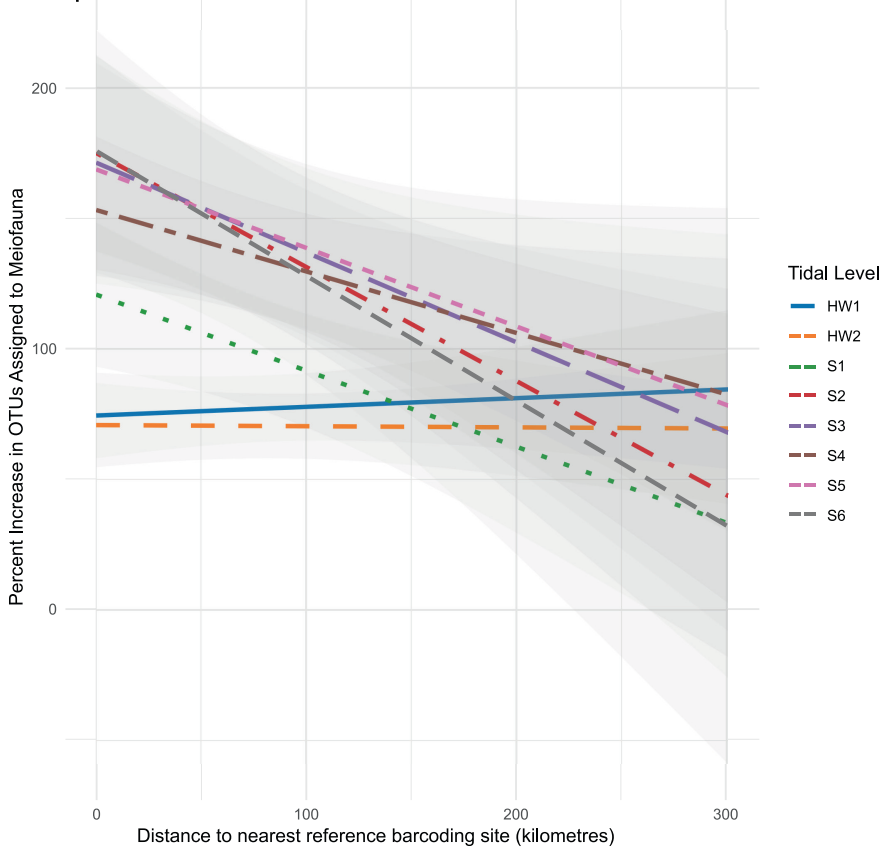

FIGURE 6 Linear relationship between the distance to the nearest reference site with newly generated reference barcodes (in kilometres) and the percent increase in Operational Taxonomic Units (OTUs) assigned to meiofauna. Linear regression lines in different colours and shaded confidence intervals show the trend for each tidal level.

are in line with findings from soil research, where the complexity and diversity of faunal communities are a significant obstacle to accurate taxonomic annotation and subsequent ecological analysis

(Recuero et al., 2023). Enriching reference libraries with local and well-annotated sequences has been shown to dramatically improve the resolution and accuracy of community composition analyses

(Caterino & Recuero, 2024; Maggia et al., 2021). The advantages of enriched reference databases are further documented for diverse ecosystems and taxonomic groups (Garlasché et al., 2023; Gold et al., 2021; Kjærandsen, 2022; Kocher et al., 2017; Magoga et al., 2022), but our results show a substantial improvement even in one of the most extensively studied areas for beach meiofauna. This underscores the massive benefit of local reference database enhancement on metabarcoding studies, underlining its value regardless of the prior level of research intensity.

4.2 | Combining morphological and molecular techniques

Our study emphasizes the need for continuous efforts in combining traditional taxonomic identification and DNA barcoding, especially targeting under-represented regions and taxa, as the lack of taxonomic expertise and sampling leaves many beach meiofauna species undescribed (Curini-Galletti et al., 2012, 2023; Martinez et al., 2023; Martínez et al., 2019). This hampers more detailed ecological studies that would facilitate the understanding of beach biodiversity and ecosystem processes, a task urgently needed due to increasing pressure on the globally important ecosystems (Lansu et al., 2024; Schlacher et al., 2007).

Combining morphological identification with molecular methods is increasingly feasible due to the development of rapid DNA barcoding pipelines relying on Oxford Nanopore technology (Srivathsan et al., 2019, 2023), potentially allowing the production of molecular references in field locations (Chang et al., 2020; Marin et al., 2022), while taxonomic experts are sorting and identifying specimens. An increased identification efficiency for metabarcoding data will significantly contribute to more precise ecological results (Faria et al., 2018), which will facilitate the use of molecular methods in monitoring (Aylagas et al., 2016), but also species detection (Giribet et al., 2023), and can therefore in turn facilitate taxonomic work on meiofauna.

4.3 | Change in inferred ecological patterns

We found that enhancing the reference database resulted in an increased identified OTU richness across all tidal levels, with the most significant increase observed in the middle intertidal zone. This led to a pattern of OTU richness peaking in the middle intertidal zone, which was more pronounced when annotating metabarcoding data with the NCBI reference database. This pattern, characterized by the highest richness in the middle intertidal zone and lower richness towards both the upper intertidal and supralittoral zones as well as the lower intertidal zone, thereby forming a bell-shaped diversity curve, has been previously described in studies based on morphological analyses of meiofauna (Armonies & Reise, 2000; Gingold et al., 2010; Maria et al., 2013). Thus, an improved reference database might reveal meiofauna richness patterns that align more

closely with those identified through thorough, but time-intensive and expertise-dependent studies on morphology.

We found that the inclusion of new reference barcodes for annotating metabarcoding data increased the community dissimilarity between tidal levels, and led to a more distinct inferred community composition of the intertidal communities as opposed to communities from the supralittoral and the swash zone. This is likely because the addition of new reference barcodes allows for both the identification of more widespread meiofauna OTUs, which could not be identified with less complete reference databases, and rare OTUs found in only a few sites. For example, OTU 1 in our dataset, present in 119 samples and with >2 million reads, was only identified as a meiofaunal copepod after adding the new reference barcodes, and was unassigned based on annotation with NCBI. The meiofauna community composition in the middle to lower intertidal zones of the studied beaches may be more similar, potentially due to similar ecological conditions across these areas. In contrast, the higher variability between communities from the supralittoral levels and the swash zone may indicate these areas' exposure to more fluctuating environmental conditions or disturbances, leading to reduced community stability. This is plausible as the supralittoral zone of beaches is often heavily influenced by human activities, such as tourism-related activities including trampling and beach driving, which have been shown to have major impacts on meiofauna communities (Gheskiere et al., 2005; Martínez et al., 2020; Pereira et al., 2017).

4.4 | Limitations and biases in reference barcoding

Despite the major improvements in databases and metabarcoding efficiency, our study has limitations. The focus on a limited geographic region (here: the Netherlands' west coast) for reference barcoding can introduce bias in the molecular reference database, as shown by our analyses on the correlation of increase in annotation efficiency with distance to the nearest sampling site used for reference barcoding. However, existing reference databases are almost always biased towards well-studied taxonomic groups and regions, and are thereby inherently biased (Weigand et al., 2019), a fact that is usually not tested due to the lack of comparison data from the same study. Our analyses show that the existing reference data in NCBI GenBank allows identifying the general patterns in meiofauna diversity in the study area, but the additional reference barcodes allow identifying more taxa and detect clearer patterns of alpha and beta diversity across the intertidal zone. In addition, the 'local' reference library provided significant improvements along 650km of coastline and up to 300km away from the sites where the organisms were obtained for reference barcoding. Furthermore, primer bias can hamper the amplification of both new reference barcodes and taxa in metabarcoding datasets. This is a known phenomenon that affects several taxonomic groups that are numerous in beach meiofauna communities, such as Nematoda (Ren et al., 2023) and Platyhelminthes (Balsamo et al., 2020). However, the highly degenerate LerayXT primers allow for amplification of a wide range of

meiofaunal taxa (Macher et al., 2022), and we show that targeted reference barcode sequencing of individual meiofaunal specimens on Oxford Nanopore platforms can overcome this impediment, as we successfully amplified reference barcodes for all major meiofauna taxa. We acknowledge, however, that the high degeneracy of the primers can also lead to co-amplification of a high number of non-target taxa, which can make metabarcoding based on the mitochondrial COI gene less efficient, and other markers such as 18S rRNA can be an appropriate choice for targeting taxonomically diverse communities (Collins et al., 2019; Taberlet et al., 2018).

4.5 | Future collaborative efforts for reference database expansion

Future research efforts should target both unexplored areas and taxa to enhance the taxonomic and geographical scope of reference databases, but also revisit areas that have been studied, as we unexpectedly found that COI reference barcodes for beach meiofauna in the North Sea region are rare. This gap is surprising because the North Sea has been the focus of extensive meiofauna studies (Armonies, 2018; Reise & Ax, 1979; Vincx et al., 1990). However, in general, only a limited number of studies on meiofauna incorporate molecular techniques, and there is a tendency to use the 18S rRNA marker due to its higher success rate in amplification, even though it offers less specificity in species identification (Gielings et al., 2021).

Our work demonstrates the feasibility of using COI barcoding and metabarcoding for a diverse range of beach meiofauna. The mitochondrial COI gene is the standard molecular barcode for most animals, and often offers superior species-level identification compared to 18S rRNA (Fontaneto et al., 2015; Giebner et al., 2020; Tang et al., 2012). Improving the 18S rRNA reference library is important, but might not be as helpful in delimiting meiofaunal species due to the conserved nature and thereby lower variability in this marker. It is therefore desirable to increase the availability of COI barcodes for meiofauna, as this will facilitate species-level molecular studies and enable more direct comparisons with macrofauna and other invertebrate groups, thanks to the use of the same amplified marker gene and gene region.

We advocate for the replication of our approach of combining taxonomic expert workshops, conducting reference barcoding, and applying metabarcoding to study not only meiofauna but also other understudied taxonomic groups. Future efforts should focus on increasing the number of reference specimens that are identified at the species level. Additionally, integrating environmental data with metabarcoding results will offer insights into the ecological factors influencing meiofauna distribution and diversity.

5 | CONCLUSION

Our study highlights the benefits of enhancing molecular reference databases for DNA metabarcoding analyses through integrated taxonomy and DNA barcoding. This approach significantly improves species identification and biodiversity assessment and shows the need for a collaborative effort in merging traditional taxonomic with molecular approaches. We advocate for continued efforts to build comprehensive, globally representative reference databases for meiofauna, which will be fundamental to advancing our understanding of coastal biodiversity and ecosystem dynamics.

DATA AVAILABILITY STATEMENT
All raw data is available in the NCBI Sequence Read Archive (SRA), under BioProject number PRJNA1081920, and in figshare, doi: 10.6084/m9.figshare.26144068. The reference specimen list is also available as a GBIF dataset: https://doi.org/10.15468/gemfv4.

ORCID
Jan-Niklas Macher https://orcid.org/0000-0003-3010-7522
Francesca Leasi https://orcid.org/0000-0001-6995-0934
Iñigo Rubio Lopez https://orcid.org/0000-0001-7316-7942

REFERENCES
Armonies, W. (2018). Uncharted biodiversity in the marine benthos: The void of the smallish with description of ten new platyhelminth taxa from the well-studied North Sea. *Helgoland Marine Research*, *72*(1), 1–29.
Armonies, W., & Reise, K. (2000). Faunal diversity across a sandy shore. *Marine Ecology Progress Series*, *196*, 49–57.
Atherton, S., & Jondelius, U. (2020). Biodiversity between sand grains: Meiofauna composition across southern and western Sweden assessed by metabarcoding. *Biodiversity Data Journal*, *8*, e51813.
Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science*, *3*, 96. https://doi.org/10.3389/fmars.2016.00096

Balsamo, M., Artois, T., Smith, J. P. S., 3rd, Todaro, M. A., Guidi, L., Leander, B. S., & Van Steenkiste, N. W. L. (2020). The curious and neglected soft-bodied meiofauna: Rouphozoa (Gastrotricha and Platyhelminthes). *Hydrobiologia*, *847*(12), 2613–2644.

Behrens-Chapuis, S., Herder, F., & Geiger, M. F. (2021). Adding DNA barcoding to stream monitoring protocols - What's the additional value and congruence between morphological and molecular identification approaches? *PLoS One*, *16*(1), e0244598.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42.

Buchner, D., Macher, T.-H., & Leese, F. (2022). APSCALE: Advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. *Bioinformatics*, *38*(20), 4817–4819.

Castro, L. R., Meyer, R. S., Shapiro, B., Shirazi, S., Cutler, S., Lagos, A. M., & Quiroga, S. Y. (2021). Metabarcoding meiofauna biodiversity assessment in four beaches of northern Colombia: Effects of sampling protocols and primer choice. *Hydrobiologia*, *848*(15), 3407–3426.

Caterino, M. S., & Recuero, E. (2024). Shedding light on dark taxa in sky-Island Appalachian leaf litter: Assessing patterns of endemicity using large-scale, voucher-based barcoding. *Insect Conservation and Diversity/Royal Entomological Society of London*, *17*(1), 16–30.

Chang, J. J. M., Ip, Y. C. A., Ng, C. S. L., & Huang, D. (2020). Takeaways from Mobile DNA barcoding with BentoLab and MinION. *Genes*, *11*(10), 1121.

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., Genner, M. J., & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*, 1985–2001. https://doi.org/10.1111/2041-210X.13276

Creedy, T. J., Norman, H., Tang, C. Q., Qing Chin, K., Andujar, C., Arribas, P., O'Connor, R. S., Carvell, C., Notton, D. G., & Vogler, A. P. (2020). A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. *Molecular Ecology Resources*, *20*(1), 40–53.

Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-Davis, R. M., Sung, W., Power, D. M., Packer, M., Carvalho, G. R., Blaxter, M. L., Lambshead, P. J. D., & Thomas, W. K. (2010). Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology*, *19*(Suppl 1), 4–20.

Curini-Galletti, M., Artois, T., Delogu, V., De Smet, W. H., Fontaneto, D., Jondelius, U., Leasi, F., Martínez, A., Meyer-Wachsmuth, I., Nilsson, K. S., Tongiorgi, P., Worsaae, K., & Todaro, M. A. (2012). Patterns of diversity in soft-bodied meiofauna: Dispersal ability and body size matter. *PLoS One*, *7*(3), e33801.

Curini-Galletti, M., Fontaneto, D., & Martinez, A. (2023). Diversity of Platyhelminthes Proseriata in Western Mediterranean sandy beaches: A database of species occurrences and traits. *Biogeographia – The Journal of Integrative Biogeography*, *38*(2). https://doi.org/10.21426/B638261295

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669.

De Souza Amorim, D., Oliveira, S. S., Balbi, M. I. P., Ang, Y., Yeo, D., Srivathsan, A., & Meier, R. (2023). An integrative taxonomic treatment of the Mycetophilidae (Diptera: Bibionomorpha) from Singapore reveals 115 new species on 730km2. *bioRxiv*. https://doi.org/10.1101/2023.09.02.555672

Dugal, L., Thomas, L., Wilkinson, S. P., Richards, Z. T., Alexander, J. B., Adam, A. A. S., Kennington, W. J., Jarman, S., Ryan, N. M., Bunce, M., & Gilmour, J. P. (2022). Coral monitoring in northwest Australia with environmental DNA metabarcoding using a curated reference database for optimized detection. *Environmental DNA*, *4*(1), 63–76.

Emerson, B. C., Borges, P. A. V., Cardoso, P., Convey, P., deWaard, J. R., Economo, E. P., Gillespie, R. G., Kennedy, S., Krehenwinkel, H., Meier, R., Roderick, G. K., Strasberg, D., Thébaud, C., Traveset, A., Creedy, T. J., Meramveliotakis, E., Noguerales, V., Overcast, I., Morlon, H., … Andújar, C. (2023). Collective and harmonized high throughput barcoding of insular arthropod biodiversity: Toward a genomic observatories network for islands. *Molecular Ecology*, *32*(23), 6161–6176.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

Faria, L. C., Di Domenico, M., Andrade, S. C. S., Santos, M. C. D., Fonseca, G., Zanol, J., & Amaral, A. C. Z. (2018). The use of metabarcoding for meiofauna ecological patterns assessment. *Marine Environmental Research*, *140*, 160–168.

Felix, G., Marenzi, R. C., Polette, M., & Netto, S. A. (2016). Landscape visual quality and meiofauna biodiversity on Sandy beaches. *Environmental Management*, *58*(4), 682–693.

Ficetola, G. F., & Taberlet, P. (2023). Towards exhaustive community ecology via DNA metabarcoding. *Molecular Ecology*, *32*(23), 6320–6329.

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. *Molecular Marine Biology and Biotechnology*, *3*(5), 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*(7), 98.

Fontaneto, D., Flot, J.-F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity: A Journal of the Senckenberg Research Institute/Senckenberg Forschungsinstitut Und Naturmudeum*, *45*(3), 433–451.

Garlasché, G., Borgomaneiro, G., Piscia, R., Manca, M., Eckert, E. M., & Fontaneto, D. (2023). Metabarcoding to monitor the crustacean zooplankton of a lake improves when using a reference DNA library from local samples. *Journal of Limnology*, *82*. https://doi.org/10.4081/jlimnol.2023.2087

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861.

Germán Rodríguez, J. (2004). Community structure of intertidal meiofauna along a gradient of morphodynamic states on an exposed North Sea beach. *Sarsia*, *89*(1), 22–32.

Gheskiere, T., Vincx, M., Weslawski, J. M., Scapini, F., & Degraer, S. (2005). Meiofauna as descriptor of tourism-induced changes at sandy beaches. *Marine Environmental Research*, *60*(2), 245–265.

Giebner, H., Langen, K., Bourlat, S. J., Kukowka, S., Mayer, C., Astrin, J. J., Misof, B., & Fonseca, V. G. (2020). Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. *Molecular Ecology Resources*, *20*(5), 1333–1345.

Gielings, R., Fais, M., Fontaneto, D., Creer, S., Costa, F. O., Renema, W., & Macher, J.-N. (2021). DNA metabarcoding methods for the study of marine benthic meiofauna: A review. *Frontiers in Marine Science*, *8*, 730063. https://doi.org/10.3389/fmars.2021.730063

Gingold, R., Mundo-Ocampo, M., Holovachov, O., & Rocha-Olivares, A. (2010). The role of habitat heterogeneity in structuring the community of intertidal free-living marine nematodes. *Marine Biology*, *157*(8), 1741–1753.

Giribet, G., Wangensteen, O. S., Garcés-Pastor, S., Møller, P. R., & Worsaae, K. (2023). Using eDNA to find micrognathozoa. *Current Biology*, *33*(14), R756–R757.

Gold, Z., Curd, E. E., Goodwin, K. D., Choi, E. S., Frable, B. W., Thompson, A. R., Walker, H. J., Jr., Burton, R. S., Kacev, D., Martz, L. D., & Barber, P. H. (2021). Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. *Molecular Ecology Resources*, *21*(7), 2546–2564.

Gray, J. S., & Rieger, R. M. (1971). A quantitative study of the meiofauna of an exposed Sandy Beach, at Robin Hood's bay, Yorkshire. *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom, 51*(1), 1–19.

Haenel, Q., Holovachov, O., Jondelius, U., Sundberg, P., & Bourlat, S. (2017). NGS-based biodiversity and community structure analysis of meiofaunal eukaryotes in shell sand from Hållö Island, Smögen, and soft mud from Gullmarn Fjord, Sweden. *Biodiversity Data Journal, 5*, e12731.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences, 270*(1512), 313–321.

Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2017). Package "geosphere". *Spherical Trigonometry, 1*(7), 1–45.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Data analysis with estimation graphics. *Nature Methods, 16*(7), 565–566.

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.

Kjærandsen, J. (2022). Current state of DNA barcoding of Sciaroidea (Diptera)-highlighting the need to build the reference library. *Insects, 13*(2), 142. https://doi.org/10.3390/insects13020147

Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A.-L., & Murienne, J. (2017). Vector soup: High-throughput identification of neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources, 17*(2), 172–182.

Kotwicki, L., Troch, M. D., Urban-Malinga, B., Gheskiere, T., & Węslawski, J. M. (2005). Horizontal and vertical distribution of meiofauna on sandy beaches of the North Sea (The Netherlands, Belgium, France). *Helgoland Marine Research, 59*(4), 255–264.

Lansu, E. M., Reijers, V. C., Höfer, S., Luijendijk, A., Rietkerk, M., Wassen, M. J., Lammerts, E. J., & van der Heide, T. (2024). A global analysis of how human infrastructure squeezes sandy coasts. *Nature Communications, 15*(1), 432.

Leasi, F., & Norenburg, J. L. (2014). The necessity of DNA taxonomy to reveal cryptic diversity and spatial distribution of meiofauna, with a focus on Nemertea. *PLoS One, 9*(8), e104385.

Leasi, F., Sevigny, J. L., Laflamme, E. M., Artois, T., Curini-Galletti, M., de Jesus Navarrete, A., Di Domenico, M., Goetz, F., Hall, J. A., Hochberg, R., Jörger, K. M., Jondelius, U., Todaro, M. A., Wirshing, H. H., Norenburg, J. L., & Thomas, W. K. (2018). Biodiversity estimates and ecological interpretations of meiofaunal communities are biased by the taxonomic approach. *Communications Biology, 1*, 112.

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 of the United States of America, 112*(7), 2076–2081.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology, 10*(1), 1–14.

Macher, J.-N., Bloska, D. M., Holzmann, M., Girard, E. B., Pawlowski, J., & Renema, W. (2022). Mitochondrial cytochrome c oxidase subunit I (COI) metabarcoding of foraminifera communities using taxon-specific primers. *PeerJ, 10*, e13952.

Macher, T.-H., Schütz, R., Yildiz, A., Beermann, A. J., & Leese, F. (2023). Evaluating five primer pairs for environmental DNA metabarcoding of central European fish species based on mock communities. *Metabarcoding and Metagenomics, 7*, e103856.

Maggia, M.-E., Decaëns, T., Lapied, E., Dupont, L., Roy, V., Schimann, H., Orivel, J., Murienne, J., Baraloto, C., Cottenie, K., & Steinke, D. (2021). At each site its diversity: DNA barcoding reveals remarkable earthworm diversity in neotropical rainforests of French Guiana. *Applied Soil Ecology: A Section of Agriculture, Ecosystems & Environment, 164*, 103932.

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: The Journal of Biological Databases and Curation, 2022*, baac055.

Maria, T. F., Esteves, A. M., Vanaverbeke, J., & Vanreusel, A. (2013). Is nematode colonisation in the presence of Scolelepis in tropical sandy-beach sediment similar to the colonisation process in temperate sandy beaches? *Brazilian Journal of Biology, 73*(1), 19–28.

Marin, C., Marco-Jiménez, F., Martínez-Priego, L., De Marco-Romero, G., Soriano-Chirona, V., Lorenzo-Rebenaque, L., & D'Auria, G. (2022). Rapid Oxford nanopore technologies MinION sequencing workflow for campylobacter jejuni identification in broilers on site—A proof-of-concept study. *Animals, 12*(16), 2065.

Martínez, A., Di Domenico, M., Leasi, F., Curini-Galletti, M., Todaro, M. A., Zotto, M. D., Gobert, S., Artois, T., Norenburg, J., Jörger, K. M., Núñez, J., Fontaneto, D., & Worsaae, K. (2019). Patterns of diversity and endemism of soft-bodied meiofauna in an oceanic Island, Lanzarote, Canary Islands. *Marine Biodiversity: A Journal of the Senckenberg Research Institute/Senckenberg Forschungsinstitut Und Naturmudeum, 49*(5), 2033–2055.

Martínez, A., Eckert, E. M., Artois, T., Careddu, G., Casu, M., Curini-Galletti, M., Gazale, V., Gobert, S., Ivanenko, V. N., Jondelius, U., Marzano, M., Pesole, G., Zanello, A., Todaro, M. A., & Fontaneto, D. (2020). Human access impacts biodiversity of microscopic animals in sandy beaches. *Communications Biology, 3*(1), 1–9.

Martinez, A., Fontaneto, D., & Curini-Galletti, M. (2023). The swash zone selects functionally specialized assemblages of beach interstitial meiofauna (Platyhelminthes, Proseriata). *Authorea Preprints.* https://doi.org/10.22541/au.169777988.89244524/v1

McLachlan, A., Defeo, O., & Short, A. D. (2018). Characterising sandy beaches into major types and states: Implications for ecologists and managers. *Estuarine, Coastal and Shelf Science, 215*, 152–160.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research, 10*, 33.

Pereira, T. J., Gingold, R., Villegas, A. D. M., & Rocha-Olivares, A. (2017). Patterns of spatial variation of meiofauna in Sandy beaches of northwestern Mexico with contrasting levels of disturbance. *Thalassas: An International Journal of Marine Sciences, 34*(1), 53–63.

Ratnasingham, S., & Hebert, P. D. N. (2007). Bold: The barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes, 7*(3), 355–364.

Recuero, E., Etzler, F. E., & Caterino, M. S. (2023). Most soil and litter arthropods are unidentifiable based on current DNA barcode reference libraries. *Current Zoology*, zoad051.

Reise, K., & Ax, P. (1979). A meiofaunal "thiobios" limited to the anaerobic sulfide system of marine sand does not exist. *Marine Biology, 54*(3), 225–237.

Ren, Y., Porazinska, D. L., Ma, Q., Liu, S., Li, H., & Qing, X. (2023). A single degenerated primer significantly improves COX1 barcoding performance in soil nematode community profiling. *Soil Ecology Letters, 6*(2), 1–14.

Sahlin, K., Lim, M. C. W., & Prost, S. (2021). NGSpeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecology and Evolution, 11*(3), 1392–1398.

Schlacher, T. A., Dugan, J., Schoeman, D. S., Lastra, M., Jones, A., Scapini, F., McLachlan, A., & Defeo, O. (2007). Sandy beaches at the brink. *Diversity and Distributions, 13*(5), 556–560.

Schmidt-Rhaesa, A. (2020). *Guide to the identification of marine meiofauna.* Verlag Dr. Friedrich Pfeil.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – Reducing sequence-to-sample misidentifications

in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303.

Schratzberger, M., & Ingels, J. (2018). Meiofauna matters: The roles of meiofauna in benthic ecosystems. *Journal of Experimental Marine Biology and Ecology*, *502*, 12–25.

Somerfield, P. J., & Warwick, R. M. (2013). Meiofauna techniques. In *Methods for the study of marine benthos* (pp. 253–284). John Wiley & Sons, Ltd.

Srivathsan, A., Feng, V., Suárez, D., Emerson, B., & Meier, R. (2023). ONTbarcoder 2.0: Rapid species discovery and identification with real-time barcoding facilitated by Oxford nanopore R10.4. *Cladistics: The International Journal of the Willi Hennig Society*, *40*, 192–203. https://doi.org/10.1111/cla.12566

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W. T., Kutty, S. N., Kurina, O., & Meier, R. (2019). Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, *17*(1), 96.

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

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050.

Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & 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 of the United States of America*, *109*(40), 16208–16212.

Vincx, M., Meire, P., & Heip, C. (1990). The distribution of nematodes communities in the southern bight of the North Sea. *Cahiers de Biologie Marine*, *31*, 107–129.

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, *6*, e4705.

Weigand, A. M., & Macher, J.-N. (2018). A DNA metabarcoding protocol for hyporheic freshwater meiofauna: Evaluating highly degenerate COI primers and replication strategy. *Metabarcoding and Metagenomics*, *2*, e26869.

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., … Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *The Science of the Total Environment*, *678*, 499–524.

Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: Improvements for better sequence analysis. *Nucleic Acids Research*, *34*(suppl_2), W6–W9.

Zeppilli, D., Sarrazin, J., Leduc, D., Arbizu, P. M., Fontaneto, D., Fontanier, C., Gooday, A. J., Kristensen, R. M., Ivanenko, V. N., Sørensen, M. V., Vanreusel, A., Thébault, J., Mea, M., Allio, N., Andro, T., Arvigo, A., Castrec, J., Danielo, M., Foulon, V., … Fernandes, D. (2015). Is the meiofauna a good indicator for climate change and anthropogenic impacts? *Marine Biodiversity: A Journal of the Senckenberg Research Institute/Senckenberg Forschungsinstitut Und Naturmudeum*, *45*(3), 505–535.

Zhou, X., Robinson, J. L., Geraci, C. J., Parker, C. R., Flint, O. S., Jr., Etnier, D. A., Ruiter, D., DeWalt, R. E., Jacobus, L. M., & Hebert, P. D. N. (2011). Accelerated construction of a regional DNA-barcode reference library: Caddisflies (Trichoptera) in the Great Smoky Mountains National Park. *Journal of the North American Benthological Society*, *30*(1), 131–162.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Macher, J.-N., Martínez, A., Çakir, S., Cholley, P.-E., Christoforou, E., Curini Galletti, M., van Galen, L., García-Cobo, M., Jondelius, U., de Jong, D., Leasi, F., Lemke, M., Rubio Lopez, I., Sánchez, N., Sørensen, M. V., Todaro, M. A., 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*, *24*, e13997. https://doi.org/10.1111/1755-0998.13997