Building a local reference library for metabarcoding survey of lake macrobenthos: oligochaetes and chironomids from Lake Maggiore

Silvia Zaupa[1,*], Diego Fontaneto[1,2], Raffaella Sabatino[1], Angela Boggero[1]

[1]National Research Council - Water Research Institute (CNR-IRSA), Verbania, Italy; [2]National Biodiversity Future Center (NBFC), Palermo, Italy

ABSTRACT

This study represents a first reference database of genetic diversity of macroinvertebrates for a barcoding marker for Lake Maggiore, focusing on the two dominant groups of the littoral benthic fauna (chironomids and oligochaetes), commonly used for biological monitoring of freshwater lakes. Sediment samples were sorted at the stereomicroscope and single animals were cut in two pieces, one piece to be used for morphological identification and one piece for DNA extraction. This study allowed us to collect and identify 427 organisms: 309 oligochaetes belonging to 27 identifiable taxa and 118 chironomid larvae belonging to 26 identifiable taxa. Four families of oligochaetes: Naididae, Lumbricidae, Lumbriculidae, and Enchytraeidae and five subfamilies of Chironomidae: Chironominae, Tanypodinae, Orthocladiinae, Diamesinae, and Prodiamesinae were found. The extraction and amplification of the DNA covered a total of 10 oligochaete taxa. For 7 of them (*Ophidonais serpentina, Uncinais uncinata, Vejdovskyella intermedia, Psammoryctides barbatus, Limnodrilus hoffmeisteri, Tubifex tubifex,* and *Bothrioneurum vejdovskyanum*), we found other sequences in GenBank to compare genetic similarities with available data. For the other taxa (Lumbriculidae, and Enchytraeidae, and *Nais* sp.) no best hits were found in GenBank. The extraction and amplification of the DNA covered a total of 21 chironomid taxa. For ten species (*Cladotanytarsus mancus, Cladotanytarsus atridorsum, Polypedilum scalaenum, Polypedilum nubeculosum, Benthalia carbonaria, Phaenopsectra flavipes, Clinotanypus nervosus, Paracladopelma laminatum, Cryptochironomus rostratus* and *Parakiefferiella finnmarkica*) sequences were available in GenBank to compare genetic similarities. For the other taxa (*Cryptochironomus* sp., *Demicryptochironomus vulneratus*, *Chironomus* sp., *Stictochironomus* sp., *Orthocladius* sp., *Cricotopus* sp., *Eukiefferiella* sp., *Procladius* sp., *Diamesa* sp., *Potthastia* sp., and *Monodiamesa bathyphila*) no best hits were found in GenBank. For chironomids, DNA taxonomy revealed the existence of several species complexes. Covering more populations and more genetic markers for those taxa within a rationale of integrative taxonomy could solve the taxonomic problems and provide a reliable description of diversity.

INTRODUCTION

The current era is identified by increasing anthropogenic impacts on the environment with cascading negative effects on biological diversity: from individuals to populations, species, communities, and whole ecosystems (Hooper *et al*., 2012; Jaureguiberry *et al*., 2022). These impacts affect ecosystem services and functioning, and eventually human well-being (Díaz *et al*., 2006; Cardinale *et al*., 2012). Biodiversity surveys are a pivotal step in increasing our knowledge of the world around us and what is happening to it (Purvis and Hector, 2000; Mendenhall *et al*., 2012).

Biodiversity surveys usually produce species lists, which will remain invaluable data for future reference (Chavan and Penev, 2011; Costello *et al*., 2013). Such species lists concur in creating our understanding of biological diversity at different spatial scales and, if performed through time, become the basis for understanding changes in relation to the increase in human presence and related anthropogenic impacts (Loh *et al*., 2005). Faunistic and floristic studies started to provide species lists long ago, but nowadays they can be improved by providing additional information that allows future studies to take advantage of objective descriptors of the species mentioned in a species list. Such an objective descriptor is represented by what has been dubbed DNA barcode (Hebert *et al*., 2003): an unambiguous DNA sequence that can be used to compare genetic diversity within and between species across time and space (Joly *et al*., 2014).

The aim of the present report is to build a reference database of genetic diversity in a barcoding marker (Weigand *et al*., 2019), focusing on the two dominant groups of the littoral benthic fauna commonly used for biological monitoring of freshwater lakes, oligochaete annelids and larvae of chironomid midges (Boggero *et al*. 2020; Kornijów *et al*., 2021), in Lake Maggiore, one of the best studied large and deep Italian subalpine lakes (de Bernardi *et al*., 1988; Arfè *et al*., 2019).

MATERIALS AND METHODS

Sampling

The study focused on Lake Maggiore (bounding box, WGS84 coordinates: latitude N 45.722039 - 46.179841, longitude E 8.81792 - 8.860820), a deep subalpine oligomictic lake, shared between Italy (Lombardy and Piedmont regions) and Switzerland (Canton Ticino). The lake is part of the Italian Long-Term Ecological Research Network (https://deims.org/f30007c4-8a6e-4f11-ab87-

569db54638fe). The lake shores host several protected areas, listed in Natura 2000 for Italy and in Emerald networks for Switzerland, together with stretches of highly modified areas, represented by towns, harbors, and other infrastructures. Samples for the study (Supplementary Table S1) were collected in the littoral areas, from the shores, and covered both natural and human-modified habitats, mostly sandy sediments in natural habitats and pebbles/rocks for human-modified habitats, to cover different habitats along the whole perimeter of the lake (Supplementary Figure S1).

For each sampling station (Supplementary Table S1), sediment samples were collected using a spatula, as abundance of species was not required. Aliquots of sediment were collected in plastic jars and kept at controlled temperatures between 4 and 7 °C to preserve live animals.

Laboratory methods

Sediment samples were sorted at the stereomicroscope (Leica M125, up to 80x magnification) to isolate oligochaetes and chironomid larvae. Single animals were isolated and preserved for further analysis. For animals of sufficiently large size to allow us to handle them with cutters, each animal was cut in two pieces, one piece to be used for morphological identification and one piece for DNA extraction. The piece used for morphological identification contained diagnostic characters and included the front end, including genital segments for oligochaetes and the head capsule for chironomid larvae. Species identification was performed by preparing Faure slides mounts of front end with chaetae distribution and genital apparatus for oligochaetes (Timm, 2009) and head for chironomid larvae. Identification was performed to species level whenever possible, using relevant taxonomic keys (*e.g.* Timm, 2009; Andersen *et al*., 2013). Photographs of taxonomically meaningful features for all analyzed individuals were obtained and used as a graphical representation in the phylogenetic trees.

The piece used for DNA extraction was preserved in ethanol (96%) and stored at -20°C until processing. DNA extraction was performed only for a selection of animals of each species for which we had specimens in good condition. After ethanol evaporation, for oligochaetes DNA was extracted using 40 µl of Chelex (BioRad, Segrate (MI), Italy) + 1 µl Proteinase K (PanReac AppliChem, Monza (MB), Italy, at 20 mg/mL in distilled water), incubating for 1 h at 56°C and 10 minutes at 100°C. For chironomid larvae, DNA was extracted with PureLink® Genomic DNAKit (Invitrogen by Life Technologies, Thermo Fisher Scientific, Monza (MB), Italy), following manufacturer's instructions DNA was amplified with PCR to obtain the barcoding Folmer fragment of the cytochrome c oxidase subunit I (COI), with primers HCO2198 and LCO1490 (Folmer *et al*., 1994; Hebert *et al*., 2003). PCR protocol included an initial step at 95°C per 5 min, 42 cycles at 95°C for 15 s, 40°C for 20 s, and 72°C for 1 min, and a final elongation at 72°C for 5 min. Negative controls were also included and all PCR products were visualised on agarose gel. Amplicons were, then, sequenced (Sanger sequencing) at Macrogen Europe (Amsterdam (BA), The Netherlands; https://www.macrogen-europe.com/); chromatograms were checked in FinchTV 1.5.0 (https://digitalworldbiology.com/FinchTV), with forward and reverse sequences merged using Mesquite 3.6 (https://www.mesquiteproject.org/). Sequences were aligned with MAFFT 7 (Katoh *et al*., 2013) using default settings, and then checked for absence of stop codons, indels, and other deviations that could be interpreted as evidence of numts.

Alignments were then used to perform DNA taxonomy. First, all sequences were checked in GenBank through Blast searches to check their identity at the high taxonomic level and confirm that they indeed were from chironomids and oligochaetes. Then, we reported the best hits in GenBank, saving their accession numbers, species names, and geographic origin. We then followed the accepted threshold of 3% as a cutoff value for DNA barcoding in COI in invertebrates (Hebert *et al*., 2003) to confirm species identity. Then, we downloaded from GenBank all available sequences of the same species; for taxa identified at the genus level, we downloaded all species of the same genus. With such data we prepared phylogenetic reconstructions to check that sequences we obtained for each species formed monophyletic clades with sequences of the same species downloaded from GenBank. Uncorrected raw genetic distances were also calculated within species and between species of the same genus using the R 4.1.3 (R Core Team, 2022) package ape v 5.6.2 (Paradis & Schliep, 2019).

A visual inspection of phylogenetic relationships within each species was obtained by phylogenetic reconstructions with Maximum Likelihood approaches in PHYML v3.0 (Guindon *et al*., 2010) with a GTR+invgamma evolutionary model. For each species, we used an outgroup selected from one of the closest taxa available in GenBank.

RESULTS

We collected and identified 427 organisms: 309 oligochaetes belonging to 27 identifiable taxa and 118 chironomid larvae belonging to 26 identifiable taxa (Supplementary Figure S2). Four families of oligochaetes: Naididae, Lumbricidae, Lumbriculidae, and Enchytraeidae and five subfamilies of Chironomidae: Chironominae, Tanypodinae, Orthocladiinae, Diamesinae, and Prodiamesinae, were found. The specimens were deposited in the CNR-IRSA collection.

Oligochaeta

The extraction and amplification of the DNA was performed successfully for a total of 36 oligochaetes (GenBank Accession numbers: OP933791 - OP933826). These covered a total of 10 taxa. For 7 of them (*Ophidonais serpentina, Uncinais uncinata, Vejdovskyella intermedia, Psammoryctides barbatus, Limnodrilus hoffmeisteri, Tubifex tubifex,* and *Bothrioneurum vejdovskyanum)* we found other sequences in GenBank to compare genetic similarities with available data. For other taxa (Lumbriculidae, Enchytraeidae and *Nais* sp.), no best hits were found in GenBank with sequences below the 3% distance threshold (Table 1).

Table 1. Uncorrected raw genetic distances between sequences of oligochaetes and chironomids from Lake Maggiore and GenBank. Percent similarity (%) is reported for the corresponding best hit with accession number and country of origin. In addition, for ambiguous hits, the secondary best hit is also reported.

Query	%	Best hit	Accession number	Country	Second %	Second best hit	Accession number	Country
Lumbriculidae								
1585_l004_ Lumbriculidae	96.50	*Lumbriculus variegatus*	FJ639300	Sweden				
Naididae								
1585_os02_Ophidonais_ serpentina	99.56	*Ophidonais serpentina*	LT903820	Switzerland				
1585_os04_Ophidonais_ serpentina	100.00	*Ophidonais serpentina*	OM033378		100	*Ophidonais serpentina*	AF534846	USA
1585_os08_Ophidonais_ serpentina	100.00	*Ophidonais serpentina*	LN810257	Switzerland				
1585_os10_Ophidonais_ serpentina	100.00	*Ophidonais serpentina*	LN810257	Switzerland				
1685_n012_Nais_sp	100.00	*Nais stolci*	MT186471	Canada	100.00	*Nais* sp.	LT903787	Switzerland
1685_n014_Nais_sp	100.00	*Nais* sp.	LT903787	Switzerland	100.00	*Nais stolci*	JQ519894	
1685_nv01_Nais_sp	99.8	*Nais* sp.	LT903787	Switzerland	99.85	*Nais stolci*	JQ519894	
1685_n021_Nais_sp	98.72	*Nais* sp.	LT903787	Switzerland	98.72	*Nais stolci*	JQ519894	
9349_n004_Nais_sp	99.38	*Nais stolci*	MT186471	Canada	99.38	*Nais* sp.	LT903787	Switzerland
1685_n009_Nais_sp	98.34	*Nais communis*	LT903795	Switzerland				
1685_uu04_Uncinais_ uncinata	98.96	*Uncinais uncinata*	LT903783	Switzerland				
1685_uu06_Uncinais_ uncinata	99.85	*Uncinais uncinata*	LT903783	Switzerland				
1685_uu09_Uncinais_ uncinata	99.68	*Uncinais uncinata*	LT903783	Switzerland				
1685_uu05_Uncinais_ uncinata	99.17	*Uncinais uncinata*	KY633410	Sweden				
9349_uu01_Uncinais_ uncinata	97.57	*Uncinais uncinata*	KY633410	Sweden				
1685_vi01_Vejdovskyella_ intermedia	98.46	*Vejdovskyella intermedia*	LT905363	Switzerland				
1685_pb02_ Psammoryctides_barbatus	98.92	*Psammoryctides barbatus*	LT598625	Switzerland				
1685_pb12_ Psammoryctides_barbatus	98.94	*Psammoryctides barbatus*	LT598625	Switzerland				
1685_pb08_ Psammoryctides_barbatus	99.09	*Psammoryctides barbatus*	LT598625	Switzerland				
1685_pb09_ Psammoryctides_barbatus	98.94	*Psammoryctides barbatus*	LT598625	Switzerland				
1685_pb11_ Psammoryctides_barbatus	98.78	*Psammoryctides barbatus*	LT598625	Switzerland				
1685_pb07_ Psammoryctides_barbatus	98.69	*Psammoryctides barbatus*	LT598625	Switzerland				
9450_Limnodrilus_ hoffmeisteri	99.35	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_a004_ Limnodrilus_ hoffmeisteri	100.00	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				

To be continued on next page

Tab. 1. Continued from previous page.

Query	%	Best hit	Accession number	Country	Second %	Second best hit	Accession number	Country
9450_pl05_Limnodrilus_hoffmeisteri	100.00	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_pl01_Limnodrilus_hoffmeisteri	99.79	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_pl04_Limnodrilus_hoffmeisteri	99.79	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_a003_Limnodrilus_hoffmeisteri	99.47	*Limnodrilus hoffmeisteri*	LT899877	Switzerland				
9450_l005_Limnodrilus_hoffmeisteri	99.81	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_lh01_Limnodrilus_hoffmeisteri	99.83	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_pl02_Limnodrilus_hoffmeisteri	100.00	*Limnodrilus hoffmeisteri*	LR025079	Switzerland				
9450_tt01_Tubifex_tubifex	98.67	*Tubifex tubifex*	LN810418	Switzerland	98.67	*Tubifex* sp.	MW703511	Canada
9450_a002_Bothrioneurum_vejdovskyanum	99.17	*Bothrioneurum vejdovskyanum*	LN999166	Switzerland				
9450_lh02_Bothrioneurum_vejdovskyanum	99.82	*Bothrioneurum vejdovskyanum*	LN999111	Switzerland				
Enchytraeidae								
7258_e001_Enchytraeidae	91.70	*Globulidrilus riparius*	MF801981	Norway				
Chironominae								
A115_Cladotanytarsus_mancus	99.09	*Cladotanytarsus mancus*	KC250762	Sweden				
A116_Cladotanytarsus_atridorsum	99.24	*Cladotanytarsus atridorsum*	KM901073	Canada				
A93_Polypedilum_scalaenum	98.67	*Polypedilum scalaenum*	MT535132	Montenegro				
A109_Polypedilum_nubeculosum	98.24	*Polypedilum nubeculosum*	MT534950	Montenegro				
A112_Polypedilum_nubeculosum	98.38	*Polypedilum nubeculosum*	MT535387	Montenegro				
A107_Polypedilum_nubeculosum	96.91	*Polypedilum nubeculosum*	MT534950	Montenegro				
A108_Polypedilum_nubeculosum	96.15	*Polypedilum nubeculosum*	MT534950	Montenegro				
A90_Benthalia_carbonaria	99.19	*Benthalia carbonaria*	MZ656833	Finland				
A92_ Phaenopsectra_flavipes	98.61	*Phaenopsectra flavipes*	MT535315	Montenegro				
A58_Paracladopelma_laminatum	99.24	*Paracladopelma laminatum*	MZ659988	Finland				
A74_Cryptochironomus_sp	98.25	*Cryptochironomus rostratus*	MZ657228.	Finland				
A118_Cryptochironomus_sp	98.94	*Cryptochironomus albofasciatus*	MZ657673	Finland	98.94	*Cryptochironomus obreptans*	MZ657544	Finland
A75_Cryptochironomus_sp	97.50	*Cryptochironomus supplicans*	MT535135	Montenegro				
A70_Cryptochironomus_sp	97.50	*Cryptochironomus supplicans*	MT535135	Montenegro				
A114_Cryptochironomus_sp	97.49	*Cryptochironomus supplicans*	MT535135	Montenegro				
A2_Demicryptochironomus_vulneratus	97.15	*Demicryptochironomus vulneratus*	MZ631468	Finland				
A61_Demicryptochironomus_vulnerautus	95.75	*Demicryptochironomus vulneratus*	MZ631468	Finland				
A35_Stictochironomus_sp	92.55	*Stictochironomus sticticus*	MZ659009	Finland				

To be continued on next page

Tab. 1. Continued from previous page.

Query	%	Best hit	Accession number	Country	Second %	Second best hit	Accession number	Country
A29_Stictochironomus_sp	92.04	*Stictochironomus pictulus*	MT534725	Montenegro				
A52_Stictochironomus_sp	92.07	*Stictochironomus pictulus*	MT534725	Montenegro				
A42_Stictochironomus_sp	93.31	*Stictochironomus sticticus*	MZ659009	Finland				
A46_Stictochironomus_sp	92.55	*Stictochironomus sticticus*	MZ659009	Finland				
A49_Stictochironomus_sp	93.18	*Stictochironomus sticticus*	MZ659009	Finland				
A30_Stictochironomus_sp	93.31	*Stictochironomus sticticus*	MZ659009	Finland				
A53_Stictochironomus_sp	92.66	*Stictochironomus sticticus*	MZ659009	Finland				
A51_Stictochironomus_sp	93.10	*Stictochironomus sticticus*	MZ659009	Finland				
A41_Stictochironomus_sp	87.41	*Stictochironomus sticticus*	MZ659009	Finland				
A44_Stictochironomus_sp	90.38	*Stictochironomus pictulus*	MT534725	Montenegro				
A38_Stictochironomus_sp	83.79	*Stictochironomus pictulus*	MZ657768	Finland				
A87_Chironomus_sp	97.82	*Chironomus cingulatus*	AF192191					
A89_Chironomus_sp	94.09	*Chironomus nipponensis*	JF412096					
Orthocladinae								
A94_Orthocladius_sp	98.33	*Orthocladius* sp.	JN275432	Norway				
A95_Orthocladius_sp	100.00	*Orthocladius oblidens*	KC250814	Sweden	100.00	*Orthocladius decoratus*	MZ632329	Finland
A97_Orthocladius_sp	100.00	*Orthocladius oblidens*	KC250814	Sweden	100.00	*Orthocladius decoratus*	MZ632329	Finland
A113_Orthocladius_sp	100.00	*Orthocladius oblidens*	KC250814	Sweden	100.00	*Orthocladius decoratus*	MZ632329	Finland
A111_Orthocladius_sp	99.85	*Orthocladius oblidens*	KC250814	Sweden	99.85	*Orthocladius decoratus*	MZ632329	Finland
A98_Cricotopus_sp	96.35	*Cricotopus vierriensis*	MZ658113	Finland				
A102_Eukiefferiella_sp	94.68	*Eukiefferiella minor*	JF870931.1	Norway				
A104_Eukiefferiella_sp	92.81	*Eukiefferiella minor*	JF870931.1	Norway				
A96_Parakiefferiella_sp	99.85	*Parakiefferiella finnmarkica*	MZ658845	Finland				
Tanypodinae								
A85_Procladius_sp	99.85	*Procladius culiciformis*	MT535048	Montenegro	99.85	*Procladius pectinatus*	MZ660730	Finland
A82_Procladius_sp	95.62	*Procladius culiciformis*	LC462322	Japan				
A84_Clinotanypus_nervosus	99.39	*Clinotanypus nervosus*	MZ657158	Finland				
A83_Clinotanypus_nervosus	99.77	*Clinotanypus nervosus*	MZ660411	Finland				
Diamesinae								
A99_Diamesa_sp	99.71	*Diamesa tonsa*	NC_063859		99.85	*Diamesa cinerella*	LN897667	
A119_Diamesa_sp	99.56	*Diamesa tonsa*	NC_063859		99.54	*Diamesa cinerella/tonsa*	LN897583	
A110_Diamesa_sp	99.85	*Diamesa tonsa*	LN897648		99.54	*Diamesa cinerella*	MT362508	Russia
A60_Potthastia_sp	90.45	*Diptera* sp.	JN290907	Canada				
Prodiamesinae								
A3_Monodiamesa_bathyphila	91.73	*Monodiamesa* sp.	MW888706					

For the species *Ophidonais serpentina*, 16 COI sequences were found in Genbank. The analysis revealed that the sequences from Lake Maggiore and all those present in the database had very high similarities (Supplementary Figure S3), with a maximum uncorrected raw genetic distance of 0.8% (Table 1). The sequences of the same species deposited in GenBank came from Switzerland, Canada, Sweden and the USA.

For the species *Uncinais uncinata*, only two sequences were found in GenBank. The five sequences from Lake Maggiore clustered in two groups (Supplementary Figure S4): one group of three sequences (uu04, uu09, uu06) was more similar to a sequence of Swiss origin (with distance of 0.4%) and the other group of two sequences (uu05, uu01) was more similar to a Swedish sequence (with distance of 0.8%). The maximum distance between the two groups, within the same species, was 3.3%.

For the species *Vejdovskyella intermedia*, only one sequence was found in GenBank, from Switzerland, and it had a distance of 2.5% to the one found in Lake Maggiore (Table 1).

The six sequences of *Psammoryctides barbatus* were compared with 25 sequences found in GenBank of the same species, all from Switzerland. They all had high similarity (Supplementary Figure S5), with a maximum uncorrected raw genetic distance of 1.8%.

For *Bothrioneurum vejdovskyanum*, 15 sequences from GenBank were available, with a maximum genetic distance within species of 5.4% (Supplementary Figure S6). One of the sequences from Lake Maggiore (a002) appeared to be very similar to sequences found in Switzerland and Canada (maximum distance 0.3%), while the other sequence (lh02) was more similar to other Swiss sequences (maximum distance 0.2%).

The nine sequences of *Limnodrilus hoffmeisteri* from Lake Maggiore were compared with 450 sequences found in GenBank (Supplementary Figure S1). The maximum distance between all sequences was 21.8%, showing a high diversity within what was reported as belonging to the same species. The most similar sequences to the ones from Lake Maggiore, within a maximum distance of 1.9%, belonged to organisms sampled in Switzerland, Sweden, Italy, The Netherlands, Finland, and China.

For the single sequence of the species *Tubifex tubifex* from Lake Maggiore, the analysis with 243 sequences present in GenBank (Supplementary Figure S2) revealed that the diversity within the species was high, with a maximum distance of 22.8%. The closest sequences to the ones from Lake Maggiore, within a maximum distance of 1.6%, were all from Switzerland.

For unidentified individuals of the families Enchytraeidae and Lumbriculidae and the genus *Nais*, the analysis of the sequences through comparison with GenBank did not give clear results for the assignment of the taxa.

Chironomidae

The extraction and amplification of the DNA was performed successfully for a total of 49 chironomid larvae (GenBank Accession numbers: OP948306 - OP948354). These covered a total of 21 taxa. For ten species *(Cladotanytarsus mancus, Cladotanytarsus atridorsum, Polypedilum scalaenum, Polypedilum nubeculosum, Benthalia carbonaria, Phaenopsectra flavipes, Clinotanypus nervosus, Paracladopelma laminatum, Cryptochironomus rostratus* and *Parakiefferiella finnmarkica*), sequences were available in GenBank to compare genetic similarities. For the other taxa (*Cryptochironomus* sp., *Demicryptochironomus vulneratus*, *Chironomus* sp., *Stictochironomus* sp., *Orthocladius* sp., *Cricotopus* sp., *Eukiefferiella* sp., *Procladius* sp., *Diamesa* sp., *Potthastia* sp., and *Monodiamesa bathyphila*), no best hits were found in GenBank with sequences below the 3% distance threshold. A phylogeny-based analysis was attempted to obtain inference on taxonomic assignment.

For *Cladotanytarsus mancus*, the phylogenetic analysis using 37 sequences identified as *Cladotanytarsus mancus* in GenBank (Supplementary Figure S7) revealed three distinct groups. The single sequence obtained from Lake Maggiore fell into one of these groups, with maximum genetic distances of 4% within the group. The distances between the Lake Maggiore individual and those of the other two groups were greater than 14.8%. Individuals from GenBank that were present in the same group with the sequence from Lake Maggiore originated from Finland, Sweden, and Montenegro.

For *Cladotanytarsus atridorsum*, the analysis with the 123 sequences of other individuals present in GenBank showed a maximum genetic distance of 3.0%. The phylogenetic reconstruction (Supplementary Figure S8) confirmed the low genetic diversity within the species, regardless of the area of origin, even with samples from Canada with a distance of 0.8% from the individual from Lake Maggiore.

Sequence comparisons for the only sequence of *Polypedilum scalaenum* obtained from Lake Maggiore and the sequences found in GenBank for the same species confirmed the species identity and showed a genetic distance of up to 13.7%. The phylogenetic tree (Supplementary Figure S9) showed at least seven well-differentiated groups. Within the group to which the individual from Lake Maggiore fell, there are identical sequences, all from Montenegro, and other sequences up to 4% distance, also from Montenegro.

For the species *Polypedilum nubeculosum*, the comparison between the four sequences from Lake Maggiore and the ones from GenBank confirmed the identification of the species, but revealed maximum distances up to 16.5%. The phylogenetic tree (Supplementary Figure S10) confirmed the presence of three groups. The sequences from Lake

Maggiore fell into the same group, along with individuals from Finland, Germany, Montenegro, Sweden and China, with a maximum distance of 4.8%.

The identification of *Benthalia carbonaria* was confirmed by comparison with GenBank, with maximum distances with other individuals of this species of 2.2%. The closest sequences had a maximum distance of 0.9% and came from Albania, Montenegro, and Finland. The phylogenetic tree (Supplementary Figure S11) confirmed the relatively low genetic diversity of the species.

Analysis with GenBank sequences for the only sequence of *Phaenopsectra flavipes* confirmed its identity and showed a maximum genetic distance of 4.3% with the 18 GenBank sequences from Finland, Montenegro, and Sweden. The phylogenetic tree (Supplementary Figure S12) confirmed the genetic homogeneity of the data for this species.

For *Paracladopelma laminatum*, the taxonomic identification of the only sequence from Lake Maggiore was confirmed by comparison with the two GenBank sequences from Finnish samples, which showed a maximum distance of 0.9% (Supplementary Figure S13).

Five sequences from Lake Maggiore were obtained from unidentified species of the genus *Cryptochironomus*. Only for one of them a best hit was found in Genbank: sequence A74_Cryptochironomus had a distance of 1.7% to a sequence of *C. rostratus* from Finland. Three other sequences (A114_Cryptochironomus, A75_Cryptochironomus, and A70_Cryptochironomus) were similar to *C. supplicans* from Montenegro and Finland (distances between 1.8% and 2.9%). Sequence A118_Cryptochironomus appeared to be similar both to *C. albofasciatus* (genetic distance 1.1%) and to *C. obreptans* (genetic distance 1.1%), both from Finland.

The two sequences of the species *Demicryptochironomus vulneratus* were compared with the 11 available sequences of the genus in GenBank (Supplementary Figure S14). The two sequences clustered with other two of the same species, from Finland, with a genetic distance to them ranging from 1.7% to 2.8%.

Four sequences for *Parakiefferiella finnmarkica* were found in GenBank, to be compared with the only sequence obtained from Lake Maggiore. The maximum genetic distance within the species was 11.4%. Two groups were present in the tree (Supplementary Figure S15): one group formed by the sequence from Lake Maggiore and the sequence from Finland, with a genetic distance of 0.1% between them.

The taxonomic identification of two sequences of *Clinotanypus nervosus* from Lake Maggiore was confirmed by comparison with GenBank. The maximum genetic distance within the species was 14.9%. The phylogenetic tree (Supplementary Figure S16) supported the presence of two groups: the two sequences from Lake Maggiore fell into the same group, with individuals from Finland and Montenegro and with genetic distances below 0.9%.

We obtained only one sequence for the genus *Potthastia*. The sequence indeed clustered within the clade formed by the 12 sequences of the genus available in GenBank (Supplementary Figure S17), but no best hit was available to match it with any known species.

For the species *Monodiamesa bathyphila*, a low similarity was found for the only sequence from Lake Maggiore to the available sequences in GenBank for the same species, with a minimum distance of 8.3%. The phylogenetic reconstruction (Supplementary Figure S18) indeed showed that the species contained at least three clusters, which did not even form a monophyletic clade.

No further analyses were performed for the sequences of the genera *Stictochironomus*, *Chironomus*, *Orthocladius*, *Cricotopus*, *Eukiefferiella*, *Procladius* and *Diamesa* given that no species could be identified from Lake Maggiore and no best hit was found to any species in GenBank. Interestingly, a high level of taxonomic ambiguity was present: four sequences of *Orthocladius* had 0% difference to two different species in GenBank, three sequences of *Diamesa* had from 0.2% to 0.5% differences to two different species in GenBank, and a sequence of *Procladius* had 0.1% difference with two different species in GenBank (Table 1).

DISCUSSION

This study allowed the creation of the first database of DNA sequences for the COI gene for macroinvertebrates belonging to Chironomidae and Oligochaeta in Lake Maggiore. The results provide a reference library to be used for future surveys using metabarcoding tools (Baird and Hajibabaei, 2012) including 36 COI barcode DNA sequences for a total of 7 identified oligochaete species, and 49 COI barcode DNA sequences for a total of 10 identified chironomid species.

Taxonomic correspondence was found between the morphological identification and sequence analysis for Oligochaeta. All oligochaete taxa identified to species level were confirmed by sequences already present in GenBank. Six of the seven identified species revealed clusters corresponding to species and with genetic distances and tree topologies compatible with the existence of only one independent evolutionary entity for each species. For one morphological species, *Tubifex tubifex*, the potential occurrence of a complex of cryptic species was highlighted by high genetic divergence and by the presence of separate clusters in the phylogenetic reconstruction. Such a pattern of high genetic differentiation within the same species name is not unexpected, given previous knowledge on this taxon (Erséus and Gustafsson, 2009; Marotta *et al.*, 2014).

For chironomids, DNA taxonomy based on genetic distances only in few cases allowed arriving unambiguously to species level. For five (*Polypedilum scalaenum*, *Polypedilum nubeculosum*, *Benthalia carbonaria*, *Demicryptochironomus vulneratus*, *Clinotanypus nervosus*) of the six taxa identified at species level, GenBank unambiguously confirmed the taxonomic identification, with other sequences of the same species at short genetic distance. For six species, including three of the five just mentioned (*Cladotanytarsus mancus*, *Polypedilum scalaenum*, *Polypedilum nubeculosum*, *Parakiefferiella finnmarkica*, *Clinotanypus nervosus*, *Monodiamesa bathyphila*), genetic distances supposedly within the same species were higher than 10%, suggesting the possible existence of species complexes or taxonomic problems.

CONCLUSIONS

Several factors may influence taxonomic identification from genetic distances in the framework of barcoding and metabarcoding (Magoga *et al*., 2021). Apart from potential mistakes in the reference libraries (Beentjes *et al*., 2019), some taxa may indeed reflect evolutionary trajectories that do not fit with the expectation of clusters of closely related sequences. Some groups indeed may have high levels of cryptic species, which then is expressed in problems in DNA-based identifications. Detailed analyses showed that for animals the error rate of GenBank for genus-level identification is generally low (~0.7/3.5%) (Ficetola *et al*., 2021). Indeed, we confirm that for all taxa of oligochaetes and chironomids from Lake Maggiore, genus-level identification was fine. Yet, whereas for oligochaetes species identification was mostly concordant, for chironomids, evidence of potential taxonomic problems was found for several species.

In a reference database to be used for biological survey through metabarcoding (Baird and Hajibabaei, 2012), sequences not unambiguously associated with species levels may reduce the accuracy of the ecological inference based on them (Sigovini *et al*., 2016; Gadawski *et al*., 2022; Magoga *et al*., 2022).

More information is surely needed to understand the reasons behind the existence of potentially problematic taxa, especially for chironomids. For example, covering more populations and more genetic markers for those taxa within a rationale of integrative taxonomy (Padial *et al*., 2010) could solve the taxonomic problems and provide a reliable description of diversity. Making databases like ours publicly available, starting from morphological identification of taxa, comparison with already available sequences, and then increasing the amount of available information will surely improve the reliability of reference libraries for lake macroinvertebrates. The reliability of a reference barcode library like the one we built for Lake Maggiore may alleviate such a potential issue, given that preserved vouchers for morphological comparisons, associated to DNA sequences of the same individuals, could be used for detailed taxonomic investigations aimed at solving the complexes of cryptic species.

Corresponding author: Silvia Zaupa, National Research Council - Water Research Institute (CNR-IRSA), Largo Tonolli 50, 28922 Verbania, Italy.
E-mail: silvia.zaupa@gmail.com

Authors' contributions: All the authors made a substantive intellectual contribution. All the authors have read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: The authors declare no potential conflict of interest.

Funding: The present study was funded by the International Commission for the Protection of Italian-Swiss Waters (CIPAIS) project and by ARPA Lombardia (Regional Agency for Environmental Protection). Additional funding for molecular analyses were obtained within the projects Progetti@CNR SOS Acque funded by CNR, and 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. Lake Maggiore is part of the Long-Term Ecological Research Network (https://deims.org/f30007c4-8a6e-4f11-ab87-569db54638fe).

Availability of data and materials: The georeferenced database with all records used for the manuscript is available at GBIF: https://doi.org/10.15468/8wbqfc.
GenBank accession numbers of all COI sequences obtained for the study: OP933791 - OP933826 for Oligochaeta, OP948306 - OP948354 for Chironomidae.

Key words: DNA barcode, macroinvertebrates, lentic waters, midges, freshwater annelids.

Received: 20 November 2022.
Accepted: 18 December 2022.

Publisher's note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Licensee PAGEPress, Italy
Advances in Oceanography and Limnology, 2022; 13:11051
DOI: 10.4081/aiol.2022.11051

REFERENCES

Andersen T, Cranston PS, Epler JH (Eds.), 2013. Chironomidae of the Holarctic Region: Keys and diagnoses. Part 1 - Larvae. Entomological Society of Lund, Sweden. 573 pp.

Arfè A, Quatto P, Zambon A, MacIsaac HJ, Manca M, 2019. Long-Term Changes in the Zooplankton Community of Lake Maggiore in Response to Multiple Stressors: A Functional Principal Components Analysis. Water. 11:962.

Baird DJ, Hajibabaei M, 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next□generation DNA sequencing. Mol. Ecol. 21:2039-44.

Beentjes K, Speksnijder A, Schilthuizen M, Hoogeveen M, Pastoor R, van der Hoorn BB, 2019. Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. PLoS ONE. 14:e0226527.

Boggero A, Zaupa S, Bettinetti R, Ciampittiello M, Fontaneto D, 2020. The Benthic Quality Index to Assess Water Quality of Lakes may be Affected by Confounding Environmental Feature. Water. 12:2519.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, et al., 2012. Biodiversity loss and its impact on humanity. Nature. 486:59-67.

Chavan V, Penev L, 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinform. 12:1-12.

Costello MJ, Michener WK, Gahegan M, Zhang ZQ, Bourne PE, 2013. Biodiversity data should be published, cited, and peer reviewed. Trends Ecol. Evol. 28:454-61.

de Bernardi R, Giussani G, Manca M, Ruggiu D, 1988. Long-term dynamics of plankton communities in Lago Maggiore (N. Italy). Verh. Internat. Verein Limnol. 23:729-33.

Díaz S, Fargione J, Chapin III FS, Tilman D, 2006. Biodiversity loss threatens human well-being. PLoS Biol. 4:e277.

Erséus C, Gustafsson D, 2009. Cryptic speciation in clitellate model organisms. In: Annelids in Modern Biology, p.31-46. Shain DH, (Ed.). John Wiley & Sons, Inc.: Hoboken, NJ, USA.

Ficetola G, Boyer F, Valentini A, Bonin A, Meyer A, Dejean T, et al., 2021. Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. Mol. Ecol. 30:3189–202.

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. Mol. Mar. Biol. Biotechnol. 3:294-9.

Gadawski P, Montagna M, Rossaro B, Giłka W, Pešić V, Grabowski M, et al., 2022. DNA barcoding of chironomidae from the Lake Skadar region: reference library and a comparative analysis of the European fauna. Divers Distrib. 28:2838-57.

Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O, 2010. New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307-21.

Hebert PD, Cywinska A, Ball SL, DeWaard JR, 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biol. Sci. 270:313-21.

Hooper DU, Adair EC, Cardinale BJ, Byrnes JE, Hungate BA, Matulich KL, et al., 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486:105-8.

Jaureguiberry P, Titeux N, Wiemers M, Bowler DE, Coscieme L, Golden A, et al., 2022. The direct drivers of recent global anthropogenic biodiversity loss. Sci. Adv. 8:eabm9982.

Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, et al., 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Mol. Ecol. Resour. 14:221-32.

Katoh K, Standley DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772-80.

Kornijów R, Pawlikowski K, Błędzki LA, Drgas A, Piwosz K, Ameryk A, Całkiewicz J, 2021. Co-occurrence and potential resource partitioning between oligochaetes and chironomid larvae in a sediment depth gradient. Aquat. Sci. 83:51.

Loh J, Green RE, Ricketts T, Lamoreux J, Jenkins M, Kapos V, Randers J, 2005. The Living Planet Index: using species population time series to track trends in biodiversity. Philosophical Transactions of the Royal Society B: Biol. Sci. 360:289-95.

Magoga G, Fontaneto D, Montagna M, 2021. Factors affecting the efficiency of molecular species delimitation in a species-rich insect family. Mol. Ecol. Resour. 21:1475–89.

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 DNAmetabarcoding: COins. Database 2022:baac055.

Marotta R, Crottini A, Raimondi E, Fondello C, Ferraguti M, 2014. Alike but different: the evolution of the *Tubifex tubifex* species complex (Annelida, Clitellata) through polyploidization. BMC Evol. Biol. 14:1-14.

Mendenhall CD, Daily GC, Ehrlich PR, 2012. Improving estimates of biodiversity loss. Biol. Conserv. 151:32-34.

Padial JM, Miralles A, De la Riva I, Vences M, 2010. The integrative future of taxonomy. Front. Zool. 7:16.

Paradis E, Schliep K, 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-8.

Purvis A, Hector A, 2000. Getting the measure of biodiversity. Nature 405:212-9.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

Sigovini M, Keppel E, Tagliapietra D, 2016. Open nomenclature in the biodiversity era. Methods Ecol. and Evol. 7:1217-1225. doi: 10.1111/2041-210X.12594

Tarmo T, 2009. A guide to the freshwater Oligochaeta and Polychaeta of Northern and Central Europe. Lauterbornia. 66:235.

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, et al., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Sci Total Environ. 678:499-524.