ID-GENE
ecodiagnostics

IDE CHE-287.540.570 TVA

109, ch. Du Pont-du-Centenaire
1228 Plan-les-Ouates
www.id-gene.com
info@id-gene.com
tél: 076 693 51 69

Date: 03.05.2023

Mandate:
Kristina Rehberger
Kanton Bern
Amt für Wasser und Abfall
Gewässer- und Bodenschutzlabor

Rapport n°: BE23_03052023_v2
Pages: 25
Annexes: Annex 1 herein
Annex 2-FilteringStats.xlsx
Annex 3-SpeciesLists.xlsx
Annex 4-OligochaetesConcatenated.xlsx
Annex5 -ChironomidsConcatenated.xlsx

Contact
kristina.rehberger@be.ch

eDNA Analyse im Rahmen des Biomonitorings

(eDNA analysis for biological monitoring)

Final report

Summary

Ninety-four water eDNA samples, as well as 40 chironomids and 36 oligochaetes bulk samples coming from 43 sites of watercourses in canton of Bern were analyzed using COI metabarcoding. In total, more than 23 millions of high-quality sequences were obtained for two mitochondrial markers COI Leray and COI Leese. In water samples, 33'168 sequence variants (ASV) were identified, belonging mostly to insects in the COI Leese dataset and to a wide range of invertebrates in the COI Leray dataset. Among the two datasets, **970 species** could be identified, representing all common aquatic invertebrate taxa. In both water datasets, Diptera, Ephemeroptera, Plecoptera and Trichoptera count together for more than 90% of insect sequences, with Diptera being dominant and represented mainly by Chironomidae and Simulidae. The contribution of bulk samples to the diversity of chironomids and oligochaetes was relatively limited. The majority of chironomid phylotypes were found in water eDNA using Leese COI marker and few of them were present only in bulk samples. Regarding the oligochaetes, their diversity in bulk samples was very low.

Project director
Kristina Cermakova

Scientific collaborator
Anne Mondino

Chief Scientific Officer
Jan Pawlowski

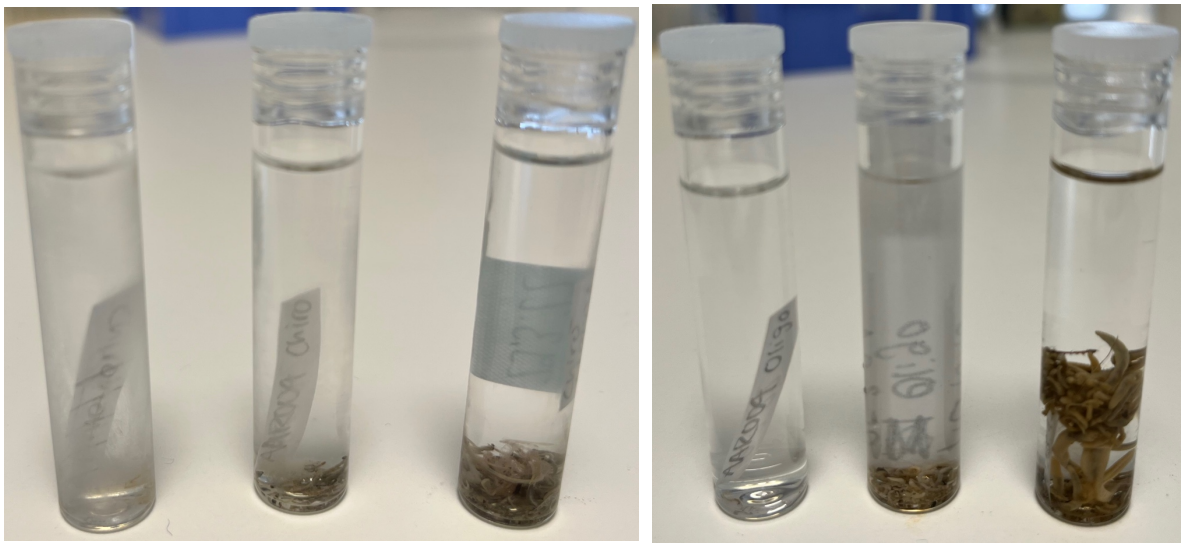
Table of contents

1. Methodology

Sampling. A total of 170 filters from water samples and 18 water negative controls, 40 chironomids bulk samples and 36 oligochaetes bulk samples in ethanol were received at ID-Gene laboratory for eDNA analysis. Detailed list of samples is provided in Annex 1 Table1. The samples were collected during four sampling campaigns: in March 22 and 26 2019, March 18 and 19 2020, March 24 and 25 2021 and March 28 and 29 2022. The sampling corresponds to the recommendations of the swiss guidelines (FOEN 2020). Water samples were collected and filtered on Sterivex filters (Millipore) in 43 sites with 4 samples per site, except in AAR04 and LUT001. where only 2 samples were taken. The samples from the left bank are indicated with L and from the right bank with R. The negative controls (filtration on site of distilled water) have been taken at the beginning and at the end of each campaign (18 in total). All Sterivex filters were stored at -20°C until shipment to the laboratory and DNA extraction. All equipment was sterilized, and gloves were used during filtration, in order to prevent contamination.

For bulk samples, the material come from kick-net samples collected at 43 sites and placed in ethanol. AquaPlus sorted morphologically oligochaetes and chironomids from the kick-net samples; chironomids were found in 40 sites and oligochaetes in 36 sites. The bulks were placed into 5ml tubes with new ethanol and kept at room temperature at AquaPlus, then at -20°C at ID-Gene upon reception in December 2022. The bulk samples contained very different biomass of oligochaetes or chironomids. In some samples, the specimens were invisible, or their biomass was very small. This was the case for 25% of chironomid samples and 39% of oligochaete samples. In other samples, the biomass of specimens was relatively large. Given the important differences in biomass of bulk samples, we introduced biomass classes: small, medium and big, as illustrated in Figure 1. The list of bulk samples with the sampling year and their approximate biomass is shown in Annex 1 Table 2.

Fig. 1. Chironomids bulks (left) and oligochaetes bulks (right) illustrating three approximate biomass classes : small, medium and big.



DNA extraction, PCR amplification, high-throughput sequencing (HTS). In the laboratory, eDNA from 188 Sterivex filters (170 samples and 18 negative controls) was extracted using the DNeasy PowerWater Sterivex kit (Qiagen), following manufacturer instructions. Ethanol from bulk samples was filtered on glass microfiber filters, Grade GF/F (Whatman), the filters were dried at room temperature and then incubated in the lysis buffer for 24h at 56°C. The eDNA was extracted using the Blood and Tissue Kit (Qiagen), following manufacturer instructions. All DNA extracts were stored at -20°C. For water samples, only 86 samples corresponding to 2 eDNA extracts per site and 8 negative controls were processed for further analysis (94 DNA extracts in total), together with 3 marine eDNA samples added as positive controls. These 97 samples and the 76 bulk samples were amplified using two mitochondrial markers. Both markers correspond to fragments of COI gene commonly used as barcode of animals (Leray et al. 2013, Leese et al. 2021). Primers sequences are shown in Table 1.

Tab. 1. List of primers for each marker.

Primer	Sequence 5' – 3'
COI Leese	
fwhF2	GGDACWGGWTGAACWGTWTAYCCHCC
EPTDr2n	CAAACAAATARDGGTATTCGDTY
COI Leray	
mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC
jgHCO2198	TANACYTCNGGRTGNCCRAARAAYCA

For each sample and marker, number of PCR replicates and quantity of DNA added to PCR reaction are shown in Table 2.

Tab. 2. Number of PCR replicates and quantity of DNA for each primer and sample

Marker	Samples	PCR replicates	DNA quantity (µl)
COI Leray	Bulk Chiro	4	2 replicates with 2ul, 1 replicate with 1ul, 1 replicate with 3ul
	Bulk Oligo	4	2 replicates with 2ul, 1 replicate with 1ul, 1 replicate with 3ul
	Water	4	3
COI Leese	Water	4	3
	Bulk Chiro	6	2

All the primers were bearing a tag of 8 or 9 nucleotides attached at each the 5'-extremity (Esling et al. 2015) to enable multiplexing of the PCR products in sequencing libraries for each marker. For each tagged primer, one PCR negative control was performed along with the sample amplification. The PCR replicates were pooled for each sample and then quantified with capillary electrophoresis using QIAxcel instrument (Qiagen). Equimolar concentrations of PCR products were pooled for each library and purified using High Pure PCR Product Purification kit (Roche Applied Science). The libraries preparation was performed using Illumina

TruSeq® DNA PCR-Free Library Preparation Kit (Illumina). The libraries were then quantified with qPCR using KAPA Library Quantification Kit (Roche Sequencing Store) and sequenced on a MiSeq instrument (Illumina) using paired-end sequencing for 300 and 500 cycles with Standard v2 kit for samples with Leese and Leray markers, respectively.

High-throughput sequencing (HTS) data analysis. Raw FASTQ reads were quality-filtered by removing any sequence with a mean quality score below 30, and also by removing all sequences with ambiguous bases or any mismatch in the tagged primer. These extremely stringent parameters ensure that we keep only high-quality reads. Then, paired-end reads assembly, chimera removing and formation of the amplicon sequence variants (ASVs) were performed using dada2 R package (Callahan et al., 2017). Taxonomic assignment was then performed with BOLD database (https://boldsystems.org/) and with MIDORI database (http://reference-midori.info/server.php) using VSEARCH (Rognes et al., 2016).

The ASVs were assigned to oligochaetes using VSEARCH with an updated version of a local database of oligochaetes phylotypes created by Vivien et al. 2017 and some other phylotypes (K1-10) were identified using phylogenetic analysis.

Unlike for the oligochaetes, there was no curated local genetic database for chironomids, so we had to add phylogenetic analyses in order to curate the ASV assignations. We grouped the ASVs that were assigned to chironomids with BOLD and MIDORI into phylotypes using phylogenetic analysis (BIONJ, K2P, 100 bootstrap replicates) as implemented in Seaview (Gouy et al. 2010). The phylotypes correspond to the phylogenetic clades supported by > 90% bootstrap. We also considered as a separate phylotype every single ASV that has been assigned to particular species. We created the local database with Genbank sequences of 293 chironomid species of which 275 were on a list of 363 chironomid species reported in Switzerland (Lods-Crozet, 1998, updated in 2005 on http://www.chironomidae.net/chklists/Swisslist.html) and 18 species were not in the Swiss list but were identified in our dataset using Midori and BOLD databases. We used this local database to build the phylogenetic trees and to verify the initial chironomid assignations.

Positive controls containing a marine community were used. The presence of marine species in samples collected in Switzerland allowed to choose a threshold (specific to dataset water Leray and water Leese), in terms of % of reads of an occurrence among the totality of reads of the given ASV, below which the occurrence might not be representative of the given sample, but might result from a sequencing artefact called tag jump. The occurrences below the threshold are marked in red in the ASVs tables of water datasets. They were taken into account for general analyses of sequence data and of high-rank taxonomic composition but removed in all the analyses going to species level.

2. Results

2.1. Sequence data

Overall sequence data statistics are shown in Table 3. In total, over 24 millions of sequences were obtained, of which about 92% were of high quality. The average number of sequences per water sample was 66'424 for COI Leray marker and 110'829 for COI Leese marker. For each chironomids bulk sample, an average number of sequences of 85'588 was obtained with COI Leray marker and 23'759 for COI Leese marker. The average number of sequences per oligochaetes bulk sample was 60'177. The number of ASVs shown in Table 3 correspond to the whole datasets, including unassigned reads.

Tab. 3. Sequence data statistics.

Marker	**COI Leese Water**	**COI Leray Water**	**COI Leese Bulk Chiro**	**COI Leray Bulk Chiro**	**COI Leray Bulk Oligo**
Raw reads (samples and controls)	10'144'039	7'018'279	1'417'992	3'578'364	2'310'522
High quality reads (samples and controls)	9'660'067	6'339'886	950'378	3'423'539	2'166'366
Average high quality reads number per sample (except controls)	110'829	66'424	23'759	85'588	60'177
Number of ASVs (samples and controls)	4'042	29'126	350	2'640	1'442

Sequence data statistics per sample are shown in Annex 2. There were high deviations among the samples - especially for the bulk samples. For oligochaetes bulks, we notice that 26 samples have less than 300 oligochaetes reads (highlighted in orange). For chironomids bulks, we observe 10 sites with less than 300 chironomids reads in both markers datasets and one more site in Leray dataset (all highlighted in orange). These samples failed to provide a sufficient number of reads of the taxon of interest either due to absence of the taxon in the sample or due to some technical problem (DNA too diluted or degraded).

Negative controls: For water negative controls, the average number of sequences was 734 and 18'101, for COI Leese and COI Leray markers respectively (Annex 2). The data in the tab "Negative controls" of Annex 2 show that no significant contamination occurred. Most occurrences have low number of reads and might thus be a result of technical biases enhanced in empty samples/controls. Besides human DNA contamination, two arthropod species were found with higher numbers of

reads in negative controls. A black fly *Prosimulium hirtipes* was found in high number of reads in both Leray and Leese datasets, in several samples and one control, indicating either a contamination of the filtered water with this species DNA or a cross-contamination in the laboratory. A mite *Nalepella brewerianae* was found in Leray dataset almost exclusively in controls. It is possible that this species was introduced to the filters or the control water or was present in the laboratory and it amplified mainly in the controls due to the absence of any other DNA to amplify.

2.2. Taxonomic composition

2.2.1. Water samples

2.2.1.1. General composition (high-rank level)

In total, 33'168 sequence variants (ASV) were identified in water samples. After a preliminary automatic assignation with MIDORI database, taxonomic composition of both COI Leray and COI Leese datasets obtained from water samples was analyzed. These "first step" analyses were conducted at phylum level for total datasets and at order level for insects. The results of these analyses are presented separately for the abundance (number of reads) and richness (number of ASVs) in Figures 2 to 4.

The analyses of COI Leray dataset show that it is dominated by the unassigned sequences, both at abundance (78%) and richness (87%) levels (Figure 2A). When the unassigned sequences are removed (Figure 3A), the COI Leray dataset is dominated by arthropods, followed by large diversity of eukaryotes (30 phyla), including fungi, oomycetes, diatoms, amoebae, and many others. Among metazoans, the Leray dataset include annelids, rotifers, mollusks, cnidarians, nematodes, gastrotrichs, porifera, and tardigrades.

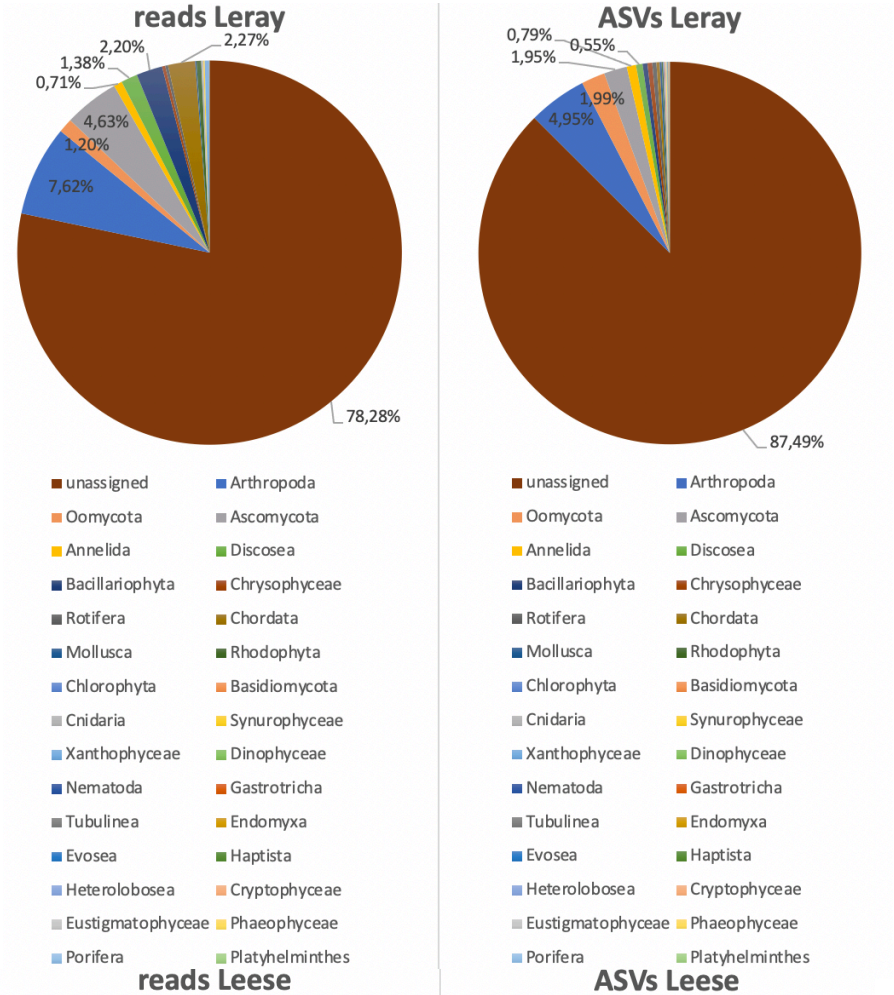
The COI Leese marker is much more specific. The unassigned sequences form only about 10% of reads and 26% of ASVs (Figure 2B). The dataset is largely dominated by arthropods (mainly insects) with 89% of reads and 71% of ASVs. The marker also recognizes 6 other phyla (Figure 3B), including annelids, mollusks and rotifers, as well as some amoebae, rhodophytes and synurophytes. However, the latter are uncommon both at reads and ASVs level.

The taxonomic analysis of insects (Figure 4) shows similar proportion of orders in COI Leray and COI Leese datasets. Both datasets are dominated by Diptera that ranges from 71% of reads and 60% of ASVs in COI Leray, to 83% of reads and 76% of ASVs in Leese dataset. The Diptera are followed by the EPT orders, with Ephemeroptera dominating in COI Leray dataset and Plecoptera dominating in COI Leese dataset. The order of Coleoptera followed in richness and abundance in both datasets. Interestingly, the orders Odonata and Dermaptera are present only in Leray dataset, while the orders Raphidioptera, Megaloptera, Siphonaptera and Neuroptera are found only in Leese datasets. In total, there are 12 and 13 insect orders recognized by COI Leray and COI Leese marker, respectively.

Highlight: eDNA from water samples analyzed with COI Leray marker enables detection of species of various animal groups, from invertebrates to vertebrates.

Fig. 2. Distribution of reads and ASVs assigned to different phyla in COI Leray and COI Leese datasets including the unassigned sequences. Values are shown for the groups representing at least 0,5% of the dataset.

A



B

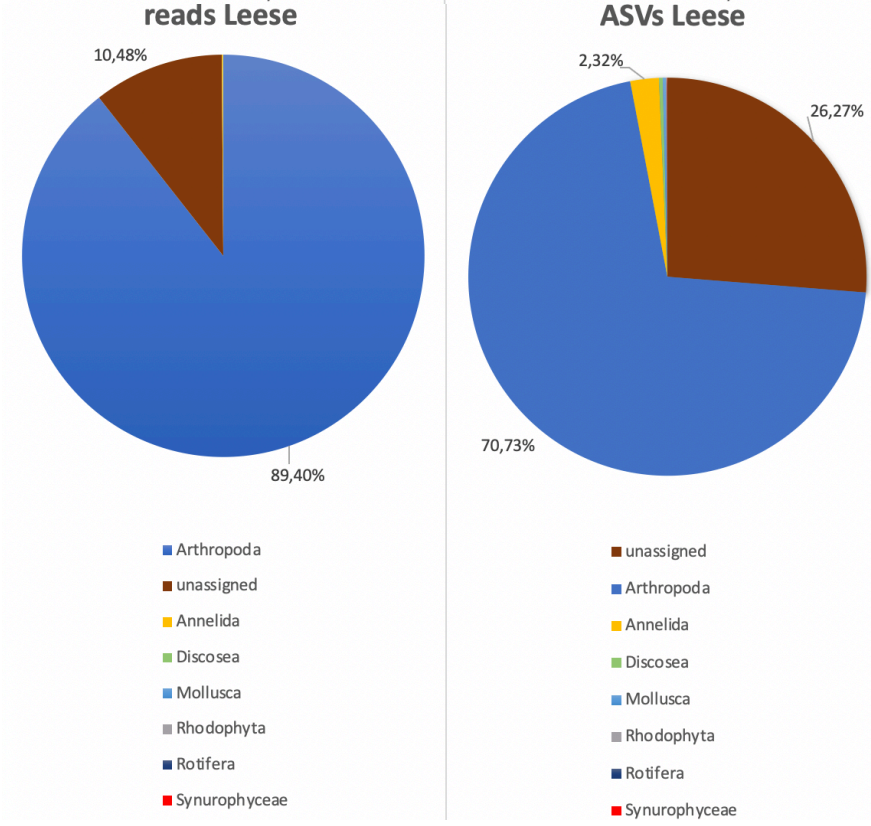
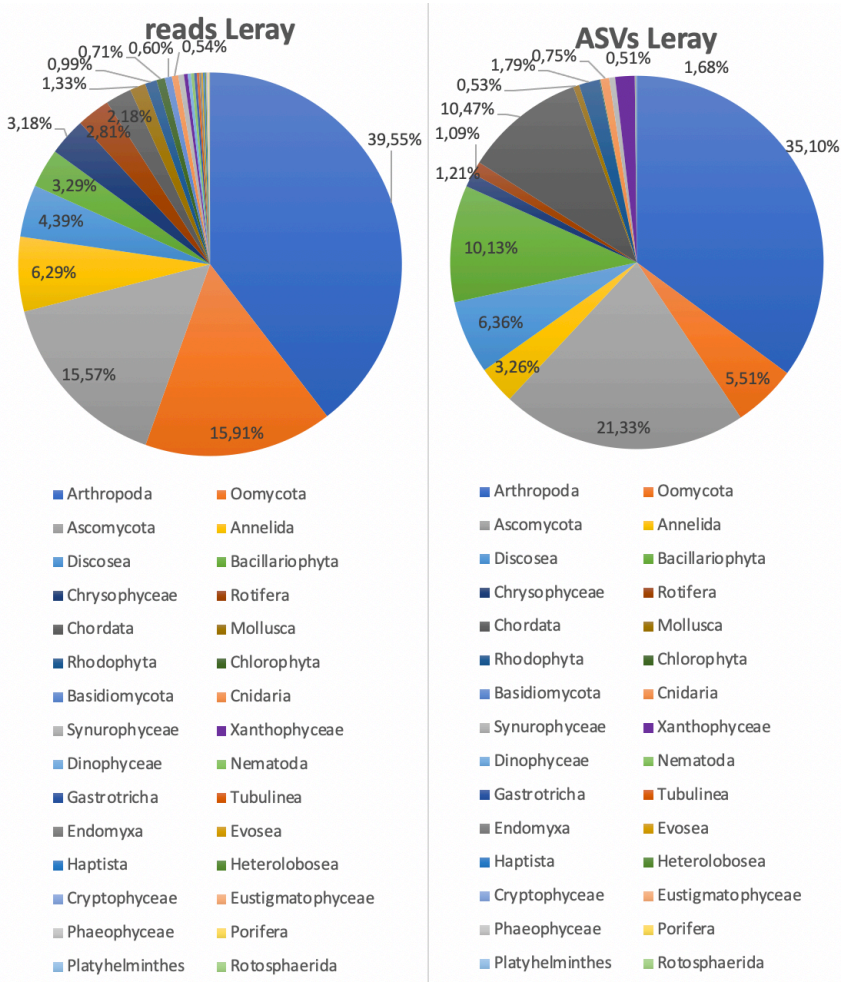


Fig. 3. Distribution of reads and ASVs assigned to different phyla in COI Leray and COI Leese datasets excluding the unassigned sequences. Values are shown for the groups representing at least 0,5% of the dataset.

A



B

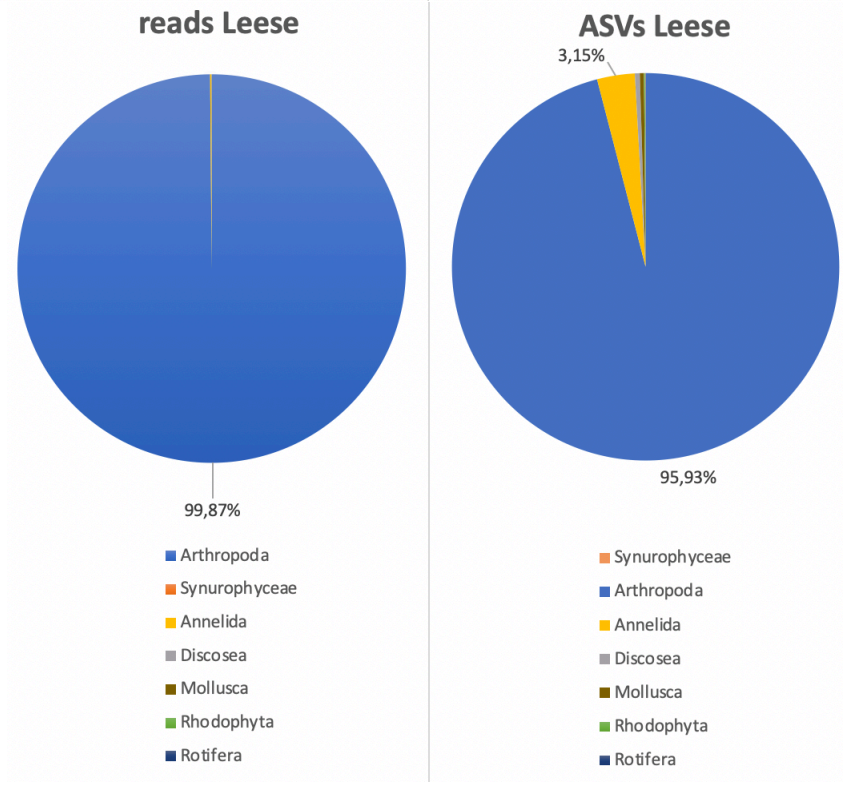
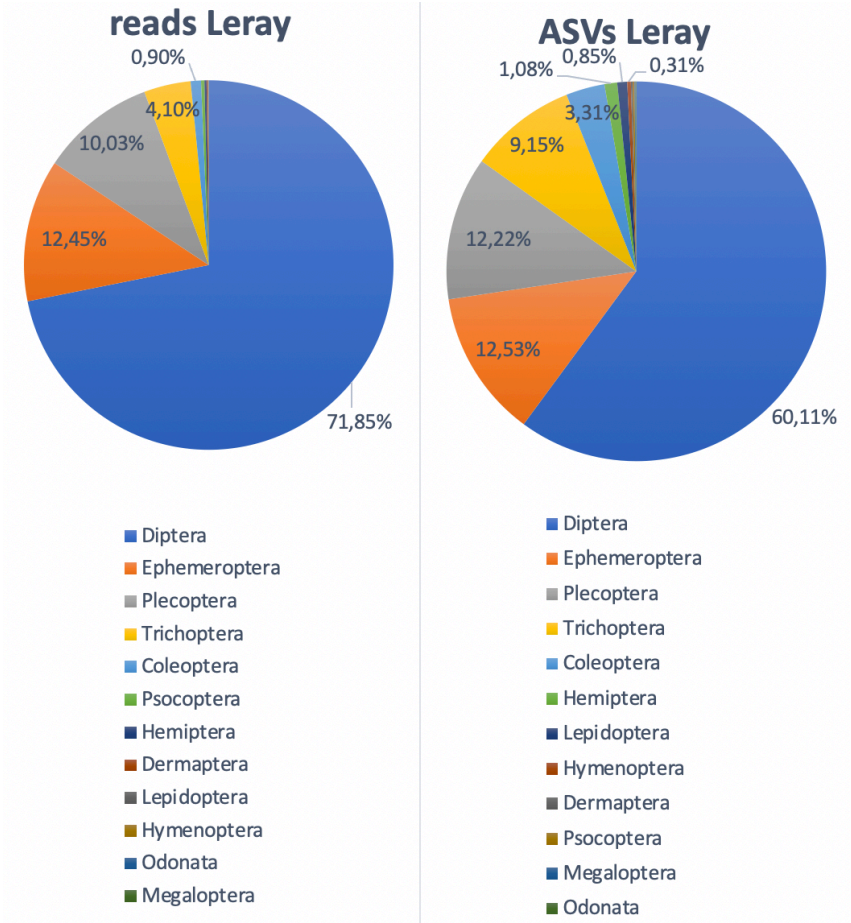
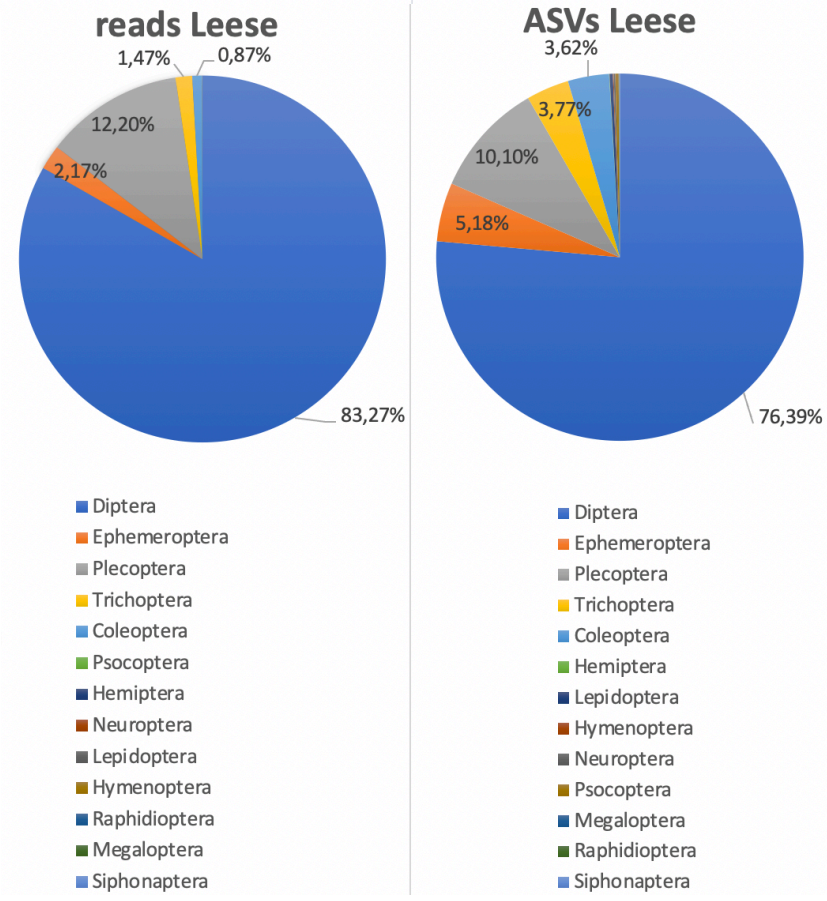


Fig. 4. Distribution of reads and ASVs assigned to insects' orders in COI Leray and COI Leese datasets. Values are shown for the groups representing at least 0,5% of the dataset.

A



B



2.2.1.2. Species lists

In the next step, in order to create species lists, taxonomic assignment was curated based on GenBank database and BOLD database (https://boldsystems.org/). ASVs assigned to chironomids and oligochaetes were further analyzed using local databases and phylogenetic trees in order to check the automated assignations and name different lineages and phylotypes. For water samples, we removed the occurrences below threshold.

970 metazoan species and phylotypes were identified in water samples (Annex 3), 278 species being found by both markers, 336 species found only in Leray dataset and 356 species found only in Leese dataset.

Among the 634 species found in water samples using Leese primer (Annex 3, tabs Leese marker), there were 54 oligochaetes, 159 chironomids, 390 other insects, 10 crustaceans and others. An average of 100 species was found per site, ranging from 54 to 166 species identified in different sites (Annex 3, tab Number of species per site).

Among the 614 species identified in water samples using Leray marker (Annex 3, tabs Leray marker), there were 85 oligochaetes, 109 chironomids, 259 other insects, 15 crustaceans and others. An average of 64 species was found per site, ranging from 13 to 170 species identified in different sites (Annex 3, tab Number of species per site).

We identified 172 species of Mayflies, Stoneflies and Caddisflies (Ephemeroptera, Plecoptera and Trichoptera) in water samples, of which 82 were shared in both markers' datasets, 41 were detected in Leese dataset only and 49 in Leray dataset only. Among them, there are 20 potentially endangered species, 3 endangered species, 8 vulnerable and 1 critically endangered species (Table 4). The latter, caddisfly *Brachycentrus maculatus* was detected with both markers in ALA010AlteAare. This site is in one of the two regions in Switzerland where the species has been observed after year 2000 according to InfoFauna data on website https://lepus.unine.ch. The most species of the list in Table 4 were found in AAR004Aare, CHI001Chirel, LUT001Luetschine (7 species) and in LUE001WeisseLuetschine (10 species). On the other hand, one of the species from the Table 4 were found in HKA001Hauptkanal, LAN002Langete, LYB002Lyssbach, TWB011Twannbach and WOR003Worble.

Among vulnerable amphibian species of Switzerland, we have detected the Common toad *Bufo bufo* in 3 sites and the Fire salamander *Salamandra salamandra* in ONZ007Oenz. The White-throated dipper *Cinclus cinclus*, a diving species of bird, was detected in 9 sites.

Among species considered as invasive in Switzerland we detect Signal crayfish *Pacifastacus leniusculus* in LYB002Lyssbach, *Oncorhynchus mykiss* in 5 sites, Brook trout *Salvelinus fontinalis* in CBA001Chraebsbach, bivalve *Corbicula fluminea* found in ALA010AlteAare, Brown rat *Rattus norvegicus* in 8 sites and Spanish slug *Arion vulgaris*, found in several sites.

Several mammals species (including also human and domestic animals) and another terrestrial animals (arthropods, molluscs,…) were detected with COI Leray marker. It is important to note that the detection of terrestrial species that do not live in the

water is incidental and no conclusions should be made about the absence of these species in other tested sites.

Tab. 4. Swiss Red list Ephemeroptera, Plecoptera and Trichoptera species detected in water samples at the indicated sites. Status: CR= critically endangered, EN=endangered, NT=potentially endangered, VU vulnerable.

Swiss status	Species	Sites
CR	Brachycentrus maculatus	ALA010AlteAare
EN	Isoperla oxylepis	19 sites
EN	Nemoura avicularis	GIB002Giesse
EN	Nemoura uncinata	EMM001Emme, SWA004Schwarzwasser
NT	Rhithrogena dorieri	6 sites
NT	Leuctra rauscheri	10 sites
NT	Nemoura cambrica	29 sites
NT	Nemoura minima	6 sites
NT	Nemoura sinuata	6 sites
NT	Perla marginata	HOB001Hornbach, CHA001RuisseaudeChaluet, SWA004Schwarzwasser
NT	Protonemura auberti	10 sites
NT	Rhabdiopteryx alpina	15 sites
NT	Ecclisopteryx guttulata	ONZ007Oenz,OEH001Oesch, ROT002Rot
NT	Metanoea flavipennis	HOB001Hornbach, SEN001ChaltiSense
NT	Rhithrogena nivata	LUT001Luetschine
NT	Capnia vidua	AAR004Aare, LUE001WeisseLuetschine
NT	Dictyogenus fontium/alpinu	LUE001WeisseLuetschine, SAA001Saane
NT	Leuctra pseudosignifera	HOB001Hornbach
NT	Nemoura obtusa	HOB001Hornbach
NT	Siphonoperla montana	LUE001WeisseLuetschine
NT	Cryptothrix nebulicola	SAA001Saane
NT	Lype reducta	10 sites
NT	Micrasema morosum	AAR004Aare, CHI001Chirel, ROA003Rotache
NT	S.galeatum/flavicorne	DSC001Dorfbach, ROT002Rot, TWB011Twannbach
VU	Protonemura algovia	4 sites
VU	Protonemura meyeri	SOR002LaSorne
VU	Baetis buceratus	URT004Urtenen
VU	Caenis beskidensis	DSC001Dorfbach
VU	Torleya major	CHA001RuisseaudeChaluet
VU	Leuctra niveola	AAR004Aare, LUT001Luetschine, LUE001WeisseLuetschine
VU	Lepidostoma basale	ONZ007Oenz
VU	Silo piceus	BIZ001LaBirse, ONZ007Oenz, DSC001Dorfbach

2.2.2. Bulk samples

2.2.2.1. General remarks

Numerous bulk samples provided poor quality amplifications, below the gel electrophoresis detection threshold, despite optimization (different quantities of DNA substrate per PCR reaction). For several bulk samples, the number of reads for the target taxa was very low (less than 300 target taxon reads highlighted in orange in Annex-Filtering stats). These problems might be either due to the degradation of DNA during long storage at room temperature or to the dilution of DNA during the processing of samples and changing the ethanol preservative. It is important to be cautious in the interpretation of the species lists from these low-quality bulk samples, because the absence of a species might actually be a false negative and a presence of a species very abundant in the overall dataset might be a result of technical biases enhanced in empty/too diluted/too degraded samples.

As different bulk samples contained very different volumes of specimens of oligochaetes or chironomids, we analyzed the correlation of their biomass with sequencing data. Table 5 shows the mean number of high-quality reads and the percentage of the target group reads in bulk samples in relation to the biomass classes introduced in Methodology section. The mean number of high-quality reads and mean percentage of chironomids or oligochaetes was higher in samples with larger biomass. This indicates that transferring the specimens into new ethanol diluted the DNA concentration and that only those samples that show high biomass provide good results.

We also analyzed the number of reads in relation to the year of sampling. As shown in the Table 5, the year of sampling does not seem to have a clear impact on number of reads for chironomids. In the case of oligochaetes, the mean percentage of oligochaetes reads per bulk was much higher in 2022 samples, compared to those collected in years 2019-21. This is not due to larger biomass of the samples that year, as shown by the mean biomass values calculated by giving a value of 1 to invisible biomass, 2 to small, 3 to medium and 4 to big biomass to each sample and calculating the mean value for the year. It suggests that the long storage of samples at room temperature reduce the preservation of specimens DNA.

Tab. 5. Mean number of high-quality reads and to the percentage of the target group reads in bulk samples in relation to the sampling year (left) and biomass (right).

Year	Mean nb high-quality reads Leray oligo	Mean % oligo reads	Mean biomasse
19	39918	4,5	3,0
20	46832	0,0	2,6
21	62378	5,3	3,3
22	104364	30,6	2,1

biomass	Mean nb high-quality reads Leray oligo	Mean % oligo reads
big	68960	8,4
medium	60619	11,1
small	61787	7,8
invisible	23230	0,1

Year	Mean nb high-quality reads Leray chiro	Mean % chiro reads	Mean biomasse
19	113578	40,9	3,3
20	42602	13,9	2,8
21	97933	22,5	3,4
22	87521	20,8	2,6

biomass	Mean nb high-quality reads Leray chiro	Mean % chiro reads
big	154438	25,7
medium	55765	18,3
small	39728	8,8
invisible	11610	0,0

Year	Mean nb high-quality reads Leese chiro	Mean % chiro reads	Mean biomasse
19	42259	79,7	3,3
20	2085	74,6	2,8
21	27936	83,4	3,4
22	20139	69,8	2,6

biomass	Mean nb high-quality reads Leese chiro	Mean % chiro reads
big	44174	86,5
medium	18959	77,6
small	3175	72,0
invisible	0	0

2.2.2.2. Oligochaetes

In this part of the study, we analysed the diversity of oligochaetes in water and bulk eDNA samples. Oligochaetes data were retrieved from water using two markers: COI Leese and COI Leray and from bulk samples using COI Leray primer. Bulks provide the highest mean number of oligochaetes reads (11'345) per sample, accounting on average for 8% of high-quality reads (Table 6).

Tab. 6. The mean number of sequences of target taxa and their percentage in relation to all high-quality sequences, for each type of sample and each marker.

Sample	Water		Oligochaetes bulks
Marker	Leray	Leese	Leray
mean number of oligochaetes reads	405	101	11'345
mean % among high-quality reads	0,64	0,06	8

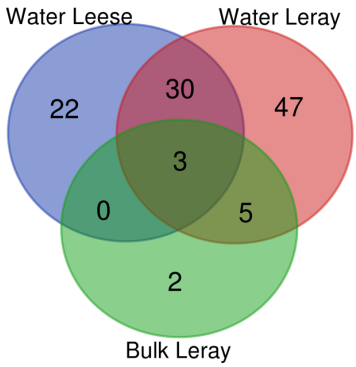
However, at the site level (Annex 2), we notice that 26 samples have less than 300 oligochaetes reads (highlighted in orange), with an average of 14 oligochaetes reads, accounting for 0,07% of high-quality reads. For the remaining 10 samples, the average number of oligochaetes reads was 40'807, accounting for 30% of high-quality reads.

For water samples, we removed the occurrences below threshold and then combined the reads of the two samples coming from each of the 36 sites. Concerning the pertinence of oligochaetes species lists derived from water datasets, it is important to note that they are based on relatively low number of sequences. The mean number of oligochaetes reads in Leese water dataset was 101 and 405 in Leray water dataset. Only 7 sites had above 300 oligochaetes reads in Leese water dataset and they were 27 in Leray water dataset.

Species list of oligochaetes present in bulk and water samples are presented in Annex 4. The species lists of the sites with low number of oligochaetes reads should be interpreted with caution because the absence of a species might actually be a false negative and a presence of a species very abundant in the overall dataset such as *Stylodrilus heringianus* genotype LL3 might be a result of technical biases enhanced in empty/too diluted/too degraded samples.

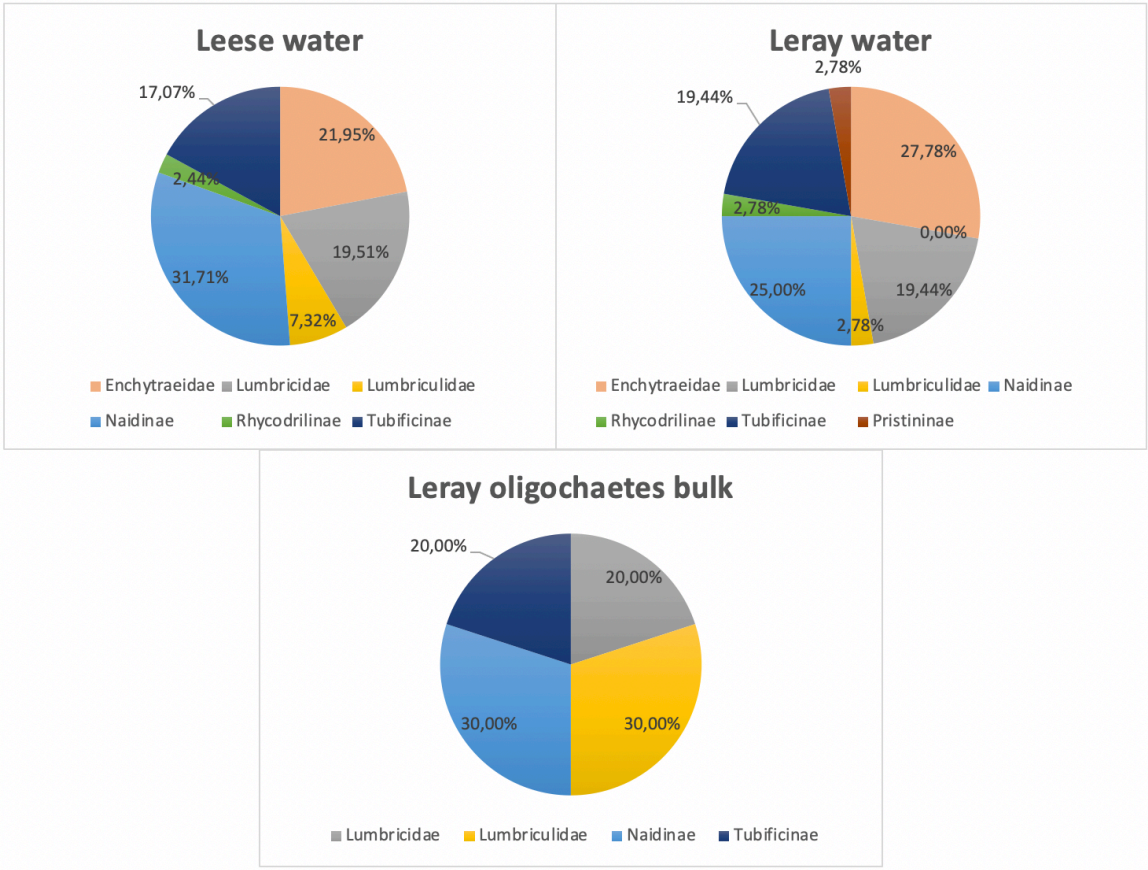
Overall, we could identify 87 phylotypes and 23 species of oligochaetes, of which only 10 were in the bulk samples (Figure 5).

Fig. 5. Venn diagram showing the species and phylotypes of oligochaetes shared between the three datasets.



The number of oligochaetes phylotypes classified in different families and subfamilies is similar between water dataset of Leray and Leese marker, while the 10 phylotypes of the bulk Leray dataset represent only 4 taxonomic groups, Enchytraeidae family being absent (Figure 6).

Fig. 6. Proportion of oligochaetes phylotypes representing different families and superfamilies, for both markers in water and bulk samples



Highlight: The highest richness of oligochaetes was obtained with COI Leray marker on water samples.

2.2.3. Chironomids

In this part of the study, we analysed the diversity of Chironomidae in water and bulk eDNA samples, using two markers: COI Leese and COI Leray. In total, we obtained 8'284'387 chironomid sequences. The largest mean proportion of target sequences was obtained with Leese marker; 78% of high-quality reads for bulk samples and 55,8% for water samples (Table 7).

Tab. 7. The mean number of sequences of target taxa and their percentage in relation to all high-quality sequences, for each type of sample and each marker.

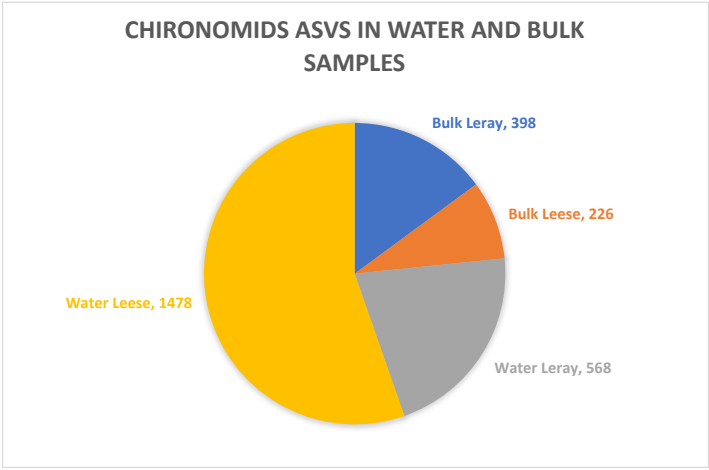
Sample	Water		Chironomids bulks	
Marker	Leray	Leese	Leray	Leese
mean number of chironomid reads	2'240	69'283	31'262	22'072
mean % among high-quality reads	3,40	55,8	25	78

Highlight: On average, more than a half of high-quality sequences obtained from bulks and water samples with COI Leese marker belong to chironomids.

If we look at the quantity of chironomid reads per site for chironomids bulks, we observe 10 sites with less than 300 chironomids reads in both markers' datasets and one more site in Leray dataset (highlighted in orange in Annex 2, bulks tab). In water samples 4 sites have less than 300 chironomids reads (Leray marker) and are also highlighted in orange (Annex 2). It is recommended to be cautious in the interpretation of species lists from these sites.

In total, 2670 ASVs assigned to Chironomidae were obtained. The majority of these ASVs (55%) were obtained from water samples with Leese marker. 21% were obtained from water samples using Leray marker, and 23% of ASVs were obtained from bulk samples (Figure 7).

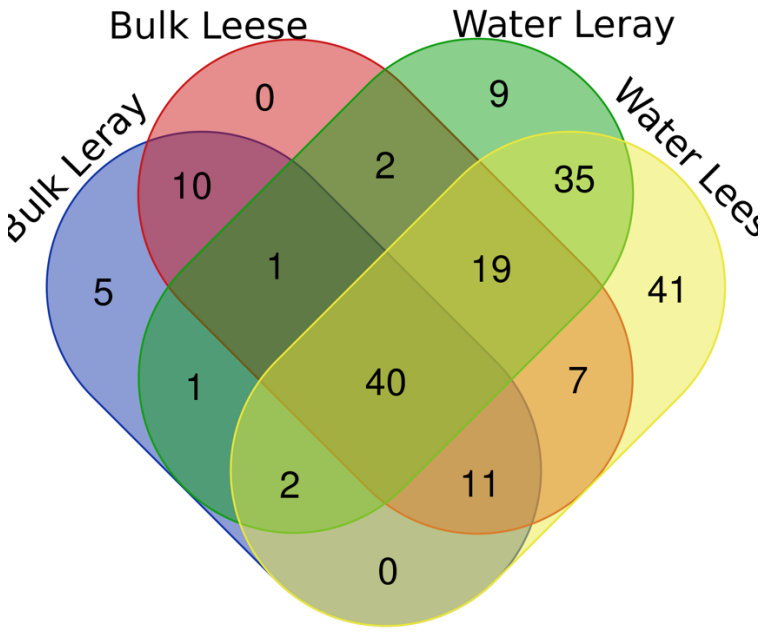
Fig. 7. Number of ASV present in four datasets: bulk and water, with both markers COI Leese and COI Leray.



Phylogenetic analysis allowed to group the majority (2478 ASVs and 8'129'042 reads) of ASVs assigned to chironomids into phylotypes. In total, 183 phylotypes were identified. The 192 ASVs assigned to chironomids that could not be grouped into phylotypes, represent less than 1% of all chironomid sequences.

The occurrence of 183 phylotypes in four datasets (water Leese, water Leray, bulk Leese, bulk Leray) is illustrated in Figure 8. 40 phylotypes were shared between all datasets. 85 phylotypes were found in water samples only (both markers), while only 15 phylotypes were present in bulk samples only (both markers). 48 phylotypes were detected with Leese marker only (both water and bulk), while only 15 phylotypes were detected with Leray marker only (both water and bulk). COI Leray marker detected 109 phylotypes in water, while COI Leese marker identified 155 phylotypes in water.

Fig. 8. Venn diagram showing the phylotypes shared between four datasets



Highlight: Water eDNA samples analyzed with COI Leese marker provide a dataset with the most chironomid phylotypes, compared to Leese marker on bulk samples and to COI Leray marker on water or bulk samples.

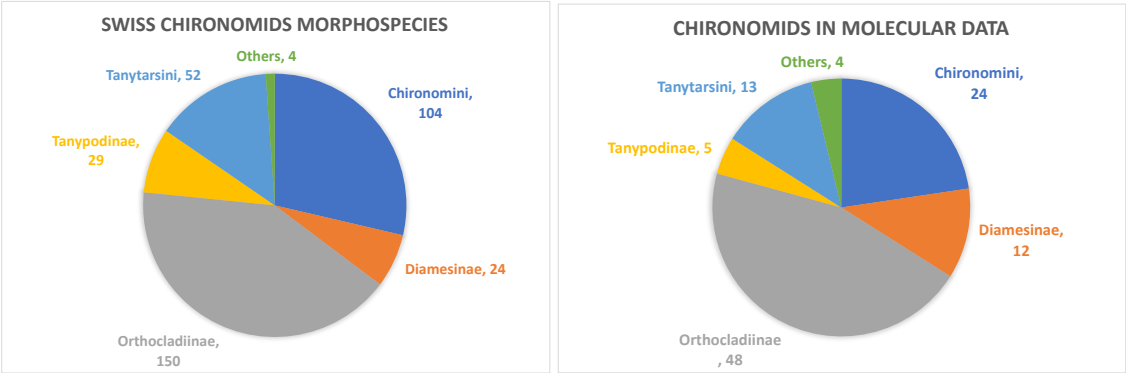
Among 183 phylotypes, 131 could be assigned to species or genus level. A few species comprise several phylotypes, that are either closely related suggesting high genetic variability of the given species or that are less related suggesting an error in the Genbank database. 39 phylotypes remained unassigned and have been given an informal names Chironomidae T9, etc. It is possible that several of these phylotypes are undetermined because Genbank reference database covers only 75% of Swiss species (Lods-Crozet 1998, updated 2005).

Despite the incompleteness of the genetic databases, we identified in this project 106 species (Annex 3, tab Chiro summary list), which corresponds to about 30% of the total number of chironomid species reported in Switzerland (Lods-Crozet 1998, updated 2005).

The distribution of species in different subfamilies or tribes is very similar for the Swiss morphospecies (Fig. 9 left) and for the phylotypes of this project (Fig. 9 right), confirming that there is no particular taxonomic bias in molecular data.

Highlight: About 30% of Swiss chironomids were identified in our data with no bias in representation of different subfamilies and tribes

Fig. 9. Distribution of Swiss morphospecies (left) and phylotypes of this project (right) in different subfamilies or tribes.



Species list of chironomids present in bulk and water samples are shown in Annex 5. For water samples, we removed the occurrences below threshold and then combined the reads of the two samples coming from each of the 40 sites.

Discussion and recommendations

The present study confirms the potential of eDNA metabarcoding as a tool to explore the diversity in aquatic ecosystems. Several studies applied this approach to monitor the biological quality of rivers in Switzerland, targeting macroinvertebrates (Mächler et al. 2014, Deiner et al. 2016, Brantschen et al. 2021), oligochaetes (Vivien et al. 2017, 2020), and diatoms (Visco et al. 2017, Apothéloz-Perret-Gentil 2021). Here, we used the COI metabarcoding to assess the diversity of macroinvertebrates in rivers of canton Bern, with focus on oligochaetes and chironomids.

We followed the study of Brantschen et al. (2021) and used two COI markers to cover wider range of taxa. Our results confirm the specificity of COI Leese marker, which was designed specifically to aquatic insects (Leese et al. 2021). The COI Leese datasets are largely dominated by Arthropods (mainly insects) with 89% of reads and 71% of ASVs assigned to these groups. Other macroinvertebrates are present but generally uncommon and their correct identification is not always possible, due to the short size of the marker (142 bp).

The COI Leray marker (Leray et al. 2013) gives longer amplicons (313 bp) and is more universal, allowing to detect wide range of invertebrates. Indeed, with this marker, we could detect practically all freshwater invertebrates taxa present in water eDNA samples, including molluscs, flatworms, sponges, cnidarians and gastrotrichs.

The Leray marker also allows to detect wide range of protist taxa, including oomycetes, diatoms, crysophytes and others. However, as the COI database of non-metazoan taxa is relatively limited, the majority of COI Leray data remained unassigned (78%). Nevertheless, the COI Leray marker is commonly used to identify oligochaetes (Vivien et al. 2017) and therefore this marker was used as the single marker in the case of oligochaetes bulk samples.

Besides using two markers, we also compared the diversity of oligochaetes and chironomids in water and bulk samples. As indicated in the Guidelines on Environmental DNA applications in biomonitoring and bioassessment of aquatic ecosystems (Pawlowski et al. 2020), different types of samples (water, sediment, biofilm and bulk) shall be preferentially used for detection of different taxonomic groups. For aquatic insects, water is qualified as 'good source' and bulks obtained by kick-net as the 'preferred source', while in the case of oligochaetes water is indicated as a 'moderate source', while the 'preferred source' is the bulk sample obtained by sorting specimens from sieved sediments (Vivien et al. 2019).

For aquatic insects, our data shown a higher abundance and richness of chironomids in water eDNA samples compared to bulk samples. This might be due to several factors. First, water eDNA has the advantage to have a potential to give a more exhaustive information, due to the presence of chironomid DNA traces from upstream, while the kick-net sampling gives a more localized information. Second, bulk samples might suffer from subsampling by kick-net (more patchiness compared to the more homogenous presence of traces in water) and from potential operator biases during sorting of specimens. Third, the sequencing depth influences the detection of species and the water samples (with two filters analyzed per site) had a bigger sequencing depth than the bulk samples (one bulk per site). Fourth, changing ethanol and non-optimal conditions of preservation might have decreased the availability of DNA in bulk samples. Fifth, since the publication of the Guidelines on eDNA applications (Pawlowski et al. 2020), a modified COI primers by Leese et al. 2021 were made available. While the traditional Leray COI marker provides better results with the bulk samples than with water samples in our chironomid data, the Leese primers' enhanced specificity for insects allowing for better chironomid detection in water eDNA samples compared to bulk samples, both in terms of abundance (reads number) and richness (number of ASVs and number of species).

In the case of oligochaetes, the detection in water eDNA samples was not very successful, compared to chironomids. Concerning the oligochaetes bulk samples, only a low the proportion of samples provided sufficient eDNA data. This is probably due to a dilution of DNA by changing the ethanol and to the degradation of DNA in ethanol samples stored for long time (up to few years) at room temperature. For comparison, the bulk ethanol samples stored in the freezer, were successfully used in other eDNA studies targeting oligochaetes (Vivien et al. 2016). It is also possible that sampling oligochaetes with a kick-net is not optimal. Indeed, according to sampling protocol for oligochaetes published by AFNOR (2016), the specimens should be sorted from sieved sediments, not from kick-net samples. Possibly, the problem could be solved by using sediments rather than kick-net to obtain oligochaetes bulk samples.

Our final remarks concern the DNA identification of targeted groups. As shown by this study, a comprehensive local reference database is essential for accurate species identification. When such local database exists as in the case of oligochaetes (Vivien et al. 2017), the assignment of sequences to particular morphospecies is relatively

straightforward. Albeit some new lineages of oligochaetes were discovered here, their number was relatively limited.

In the case of chironomids, the local barcoding database of Swiss species is not available. Although there is a large number of chironomid COI barcodes sequences available (over 200'000 sequences in the Genbank database), the proportion of chironomid metabarcodes that could not be assigned to a formally described species remain relatively high (85% of chironomid OTUs were reported as unassigned in Beermann et al. 2018). In our data, 120 out of 183 phylotypes could be assigned to morphospecies. Yet, the assignment was often based on a single reference DNA barcode, which makes it reliability uncertain. The number of metabarcoding studies targeting chironomids (Theissinger et al 2018, Uchida et al. 2020, Lin et al. 2020, Beermann et al. 2018, 2021) is rapidly increasing. It is therefore of great importance to develop a reference database of DNA barcodes for further metabarcoding studies.

Based on results of this study and the points discussed above, we propose some recommendations for the future projects:

1. **To use COI Leese marker to study aquatic insects (mainly Diptera, Ephemeroptera, Plecoptera and Trichoptera) in eDNA water samples.**
2. **To use the COI Leray marker to study various animal groups in eDNA water samples.**
3. **To use bulk samples obtained from sieved sediments rather than from kick-net to study the diversity of oligochaetes.**
4. **To preserve both filters used for water eDNA sampling and bulk ethanol samples in low temperature to avoid DNA degradation.**
5. **To filter rapidly the ethanol used for bulk samples and do not change it.**
6. **To preserve specimens for DNA barcoding reference database of Swiss chironomids and oligochaetes.**

References

AFNOR. 2016. Qualité de l'eau – échantillonnage, traitement et analyse des oligochètes dans les sédiments des eaux de surface continentales. Association française de normalisation (AFNOR), NF T 90-393. France: 14pp. + annexes.

Apothéloz-Perret-Gentil L, Bouchez A, Cordier T, Cordonier A, Guéguen J, Rimet F, Vasselon V, Pawlowski J. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Mol Ecol. 2021 Jul;30(13):2959-2968. doi: 10.1111/mec.15646.

Beermann AJ, Elbrecht V, Karnatz S, Ma L, Matthaei CD, Piggott JJ, Leese F. Multiple-stressor effects on stream macroinvertebrate communities: A mesocosm experiment manipulating salinity, fine sediment and flow velocity. Sci Total Environ. 2018 Jan 1;610-611:961-971. doi: 10.1016/j.scitotenv.2017.08.084.

Beermann, A.J., Zizka, V.M.A., Elbrecht, V. et al. DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. Environ Sci Eur 30, 26 (2018). https://doi.org/10.1186/s12302-018-0157-x

Brantschen J, Blackman RC, Walser JC, Altermatt F. Environmental DNA gives comparable results to morphology-based indices of macroinvertebrates in a large-scale ecological assessment. PLoS One. 2021 Sep 21;16(9):e0257510. doi: 10.1371/journal.pone.0257510. eCollection 2021.

Brantschen J, Pellissier L, Walser J-C, Altermatt, F. (2022). Evaluation of primer pairs for eDNA-based assessment of Ephemeroptera, Plecoptera, and Trichoptera across a biogeographically diverse region. Environmental DNA. 4. 10.1002/edn3.342.

Callahan B. J., McMurdie P. J., & Holmes S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. – The ISME Journal 11: 2639 – 2643 (2017).

Deiner K, Fronhofer EA, Mächler E, Walser JC, Altermatt F. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nat Commun. 2016 Aug 30;7:12544. doi: 10.1038/ncomms12544.

Esling P., Lejzerowicz F., & Pawlowski J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. – Nucleic Acids Res. 43: 2513 – 2524 (2015).

FOEN (ed.) 2020: Environmental DNA applications in biomonitoring and bioassessment of aquatic ecosystems. Federal Office for the Environment, Bern. Environmental studies no. 2010: 70 pp.

Hajibabaei M, Porter TM, Wright M, Rudar J. COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One. 2019 Sep 12;14(9):e0220953. doi: 10.1371/journal.pone.0220953.

Leese F. et al. Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environmental DNA 3, 261–276 (2021).

Leray M. et al. 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, 34 (2013).

Lin, Xiao-Long & Mo, Lidong & Bu, Wenjun & Wang, Xinhua. (2021). The first comprehensive DNA barcode reference library of Chinese Tanytarsus (Diptera: Chironomidae) for environmental DNA metabarcoding. Diversity and Distributions. 27. 10.1111/ddi.13209.

Lods-Crozet B. Chironomidae. 12. In : Merz, B., Bächli, G., Haenni, J.-P. & Gonseth, Y. (eds). Diptera - Check-list. *Fauna Helvetica* 1, 92-101 (1998), updated in 2005 on http://www.chironomidae.net/chklists/Swisslist.html

Pawlowski J., Apothéloz-Perret-Gentil L., Mächler E. & Altermatt F. 2020: Environmental DNA applications in biomonitoring and bio-assessment of aquatic ecosystems. Guidelines. Federal Office for the Environment, Bern. Environmental Studies. no. 2010: 71 pp.

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

Uchida N, Kubota K, Aita S, Kazama S. Aquatic insect community structure revealed by eDNA metabarcoding derives indices for environmental assessment. PeerJ. 2020 Jun 11;8:e9176. doi: 10.7717/peerj.9176.

Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J. (2015) Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environ Sci Technol. 49(13):7597-605.

Vivien R, Apothéloz-Perret-Gentil L, Pawlowski J, Werner I, Lafont M, Ferrari BJD. High-throughput DNA barcoding of oligochaetes for abundance-based indices to assess the biological quality of sediments in streams and lakes. Sci Rep. 2020 Feb 6;10(1):2041. doi: 10.1038/s41598-020-58703-2.

Vivien R, Holzmann M, Werner I, Pawlowski J, Lafont M, Ferrari BJD. Cytochrome c oxidase barcodes for aquatic oligochaete identification: development of a Swiss reference database. PeerJ. 2017 Dec 6;5:e4122. doi: 10.7717/peerj.4122.

Vivien R, Lejzerowicz F, Pawlowski J. (2016) Next-generation sequencing of aquatic oligochaetes: comparison of experimental communities. PLoS One. 2016 Feb 11;11(2):e0148644.