repliSTREAM:

Replicating Ballini et al. (2024) for Comparative Analysis of Bioinformatics Pipelines in eDNA Metabarcoding of Italian Freshwater Fish Communities

(An Independent Learning Project)

Author:

Nicole Marie Anthony

Collaborator:

Daniel Stratis

September 11, 2025

Introduction

Environmental DNA (eDNA) refers to genetic material shed by organisms in their environment (Taberlet et al., 2012). eDNA metabarcoding enables simultaneous identification of multiple species from a variety of sample types, including soil, air, and water (Ruppert et al., 2019). In aquatic environments, water samples can be filtered and sequenced to identify species without the need to capture or observe organisms directly (Taberlet et al., 2018). eDNA has revolutionized biodiversity monitoring because it is a non-invasive, affordable and sensitive tool for detecting rare, endangered or invasive species and for assessing ecosystem health (Ruppert et al., 2019).

To interpret eDNA data, bioinformatics pipelines are essential. Although many different pipelines have been developed, they all share common steps, including demultiplexing, trimming primers and adapters, quality filtering and merging paired-end reads, chimera filtering, denoising or clustering, post-clustering curation, taxonomic assignment, and decontamination (Hakimzadeh et al., 2024). However, the way that each pipeline performs these steps can vary, which means that species detections can differ between workflows (Hakimzadeh et al., 2024). This is important to consider when choosing a bioinformatics pipeline because eDNA results are increasingly used to inform conservation decisions and policies (Pochon et al., 2025).

Recent research highlights these differences. For example, dos Santos & Blabolil (2025) showed that different bioinformatics pipelines processing the same 12S ribosomal RNA (rRNA) metabarcoding data often detect slightly different fish species. Denoising pipelines like DADA2 may detect more species but at the cost of some false positives (Nearing et al., 2018), whereas alignment-based pipelines like Barque (Normandeau, 2024) are more conservative to avoid over-assigning ambiguous reads but might miss rare species (dos Santos & Blabolil, 2025). Consequently, biodiversity estimates and the management decisions based on them can depend heavily on the choice of pipeline, highlighting the need for reproducibility in eDNA analyses.

While pipeline choice can significantly influence species detection results, usability also shapes adoption by researchers. Most eDNA workflows rely on command-line interface (CLI) tools, but graphical user interfaces (GUIs) are increasingly being developed to improve accessibility (Dufrense et al., 2019; Ratnasingham, 2019; Macher et al., 2021; Buchner et al., 2022; Wheeler et al., 2024). Although the number of eDNA bioinformatic tools continues to grow, direct comparisons of multiple pipelines on the same dataset remain rare (dos Santos & Blabolil, 2025). The primary motivation for this project was to gain practical experience with eDNA workflows while assessing reproducibility and usability across three open-source pipelines: Barque (Normandeau, 2024), APSCALE-CLI (Buchner et al., 2022), and GUI-based eDNA-Container App (Wheeler et al., 2024).

Barque is a CLI pipeline that denoises reads using UNOISE3 and annotates them via VSEARCH alignment, producing operational taxonomic units (OTUs) or amplicon sequence variants (ASVs). It is an alignment-based workflow designed to avoid over-assignment, though this can limit resolution at the species level (Normandeau, 2024). APSCALE, available in both CLI and GUI formats, combines denoising, clustering, and LULU for post-clustering curation (Frøslev, et al., 2017) before taxonomic assignment via BLAST+. The eDNA-Container App offers a GUI that runs QIIME2 within Docker, applying DADA2 for error correction and taxonomic assignment, and automatically generating standardized reports for users without coding experience (Wheeler et al., 2024). These three pipelines were selected to reflect different approaches in eDNA bioinformatics, enabling comparison of alignment-based, clustering-based, and GUI-driven workflows.

Methods

Ballini et al. (2024) Dataset and Workflow

We analysed Illumina paired-end 12S rRNA reads from freshwater fish communities in six rivers of northwestern Italy (Roia, Bevera, Argentina, Carpasina, Nervia, Tanaro), originally sequenced by Ballini et al. (2024) using two mitochondrial markers: Tele02 (Taberlet et al.,

2018) and Vert01 (Riaz et al., 2011). Their workflow combined Barque, LULU (Frøslev et al., 2017), and microDecon (McKnight et al., 2019) for decontamination.

repliSTREAM Workflow

Figure 1 depicts an overview of the repliSTREAM bioinformatics workflow. All instructions and scripts used in this project are available at the GitHub page: https://github.com/nanthony80/repliSTREAM.

Processing of Raw Reads

Raw sequencing data from Ballini et al. (2024) were downloaded directly from the National Centre for Biotechnology Information (NCBI) using accession IDs listed in `SRR_Acc_List.txt`. Downloads were performed with the SRA Toolkit prefetch utility via `01_fastq_download.sh` and resulting .sra files were converted to paired-end FASTQ format with fasterq-dump via `02_convert_fastq.sh`. Quality control was performed on all FASTQ files using FastQC and MultiQC via `03_qc_fastqc_multiqc.sh`. Finally, paired-end FASTQ files were compressed and renamed via `04_rename_fastq.sh`, which referenced `metadata.csv` to label each file by river and replicate for ease of identification in downstream analyses.

Reference Database Construction

Custom reference databases for Tele02 and Vert01 were created with CRABS (v0.1.7) via `05_CRABS_ref_db.sh`, following the approach of Ballini et. al. (2024) but with minor modifications. Briefly, mitochondrial 12S rRNA sequences were downloaded from NCBI and the Barcode of Life Database (BOLD), merged, and trimmed by in-silico PCR. Amplicons were filtered by length and ambiguous bases, aligned with pairwise global alignment, assigned taxonomy, dereplicated, and curated using custom scripts (`length-filtering.py`, `dbrefinement_strict.py`, `idt2barque.py`) from the STREAM repository (Staffoni, n.d.). Unlike Ballini et al. (2024) who used CRABS v0.1.8, we implemented v0.1.7 as v0.1.8 was not available for download at the time of this project. Additional NCBI queries for Amphibia, Archelosauria, and Mammalia included the mitochondrion[filter] flag

to reduce download size and prevent repeated errors during retrieval on a personal computer.

To control for reference database content, the CRABS-curated Tele02 and Vert01 databases were produced in Barque format and reformatted for APSCALE and the eDNA-Container App (Figure 1). Each database contained sequences for a single marker: 28,254 for Tele02 and 41,792 for Vert01, closely matching the totals reported by Ballini et al. (2024) (26,674 and 40,556, respectively).

Replication of Barque Pipeline

We replicated the workflow of Ballini et al. (2024) using Barque (v1.8.5), LULU (v0.1.0) and microDecon (v1.2.0) following steps in the manuscript and the STREAM repository (Staffoni, n.d.). Briefly, paired-end reads were merged with FLASH and trimmed with Trimmomatic, allowing up to four primer mismatches. Chimeras were identified and removed with VSEARCH. Taxonomic assignment was carried out using the Barque VSEARCH alignment module against the CRABS-derived databases, with thresholds of ≥98% identity for species, 95–98% for genus, and <95% for family. Low-abundance sequences were filtered with a minimum of 10 reads per sample and across the dataset. As in the original workflow, we performed two sequential Barque runs. The first run produced denoised OTUs and initial taxonomic assignments. The second run used these OTUs as the reference database to generate read counts per sample. Unlike Ballini et al. (2024), we did not apply `remove_species.py` to exclude "unwanted species" from the reference database prior to the second run. Instead, all filtering of non-target taxa was done post-assignment (see multihits_and_nontarget_species.md).

Following Barque, error reduction was applied with LULU (minimum match similarity = 87%) via `07_lulu_barque.Rmd`. For Vert01, OTUs with multiple hits were manually consolidated into the assignments made by LULU and documented in `multihits_and_nontarget_species.md`. Finally, decontamination was conducted with *microDecon* using negative controls via `08_microdecon_barque.Rmd`.

APSCALE-CLI Pipeline

We analysed the same data using APSCALE-CLI (v4.1.3). The pipeline was configured via a Settings.xlsx file. For Tele02, quality filtering was set to a maximum expected error of 1 and read lengths restricted to 109–229 bp and Vert01 reads were filtered between 36–152 bp. Denoising used UNOISE3 defaults, and clustering was applied with a sequence group threshold of 97% to approximate Barque's direct alignment strategy. Replicate merging and negative control removal were disabled (set to `FALSE`) to maintain comparability with the Barque workflow. CRABS-derived BLAST databases was created via `09_CRABS_to_APSCALE.py` and taxonomic assignment was conducted with APSCALE-BLAST (v1.2.7) via `11_apscale_blast.sh` against the database using identity thresholds identical to Barque (≥98% species, 95–98% genus, <95% family). The read tables and taxonomy outputs were merged in TaxonTableTools2 (Macher et al., 2021) for downstream analyses. Since APSCALE integrates LULU, only decontamination was applied with microDecon using negative controls via `13_microdecon_apscale.Rmd`.

eDNA-Container Pipeline

The final workflow implemented was the eDNA-Container App (v2.0.6), which wraps QIIME2 (v2023.2) in a containerised, GUI. Input FASTQ files were renamed to meet eDNA-Container App conventions (Wheeler, 2025), and each marker was processed separately. Primer trimming was conducted with cutadapt, and denoising and ASV inference with DADA2 using default settings (trunc-len = 0; maxEE forward = 2, reverse = 4; truncQ = 2; chimera removal = consensus). Taxonomic assignment used QIIME2's feature-classifier classify-sklearn, retrained on the CRABS Tele02 and Vert01 reference databases converted to QIIME2 classifiers via `15_CRABS_to_QIIME2.sh`. The resulting ASV tables were subsequently filtered with LULU via `17_lulu_edna_container_app.Rmd` and decontaminated using microDecon via`18_microdecon_edna_container_app.Rmd` in the same manner as for Barque and APSCALE outputs.

Comparative Analyses

After decontamination with microDecon, the outputs from all three pipelines consisted of standardized tables at two levels: (i) per-marker, per-river summaries (reads summed

across samples) and (ii) per-marker tables disaggregated by river and sample. Non-target taxa were manually removed and documented in`multihits_and_nontarget_species.md`. Like Ballini et al. (2024), species were considered present in a river if more than 10 reads were detected. Marker-specific analyses included tables of species detected and read counts for each pipeline, stratified by marker and river via `19_analyses_richness_river_all_pipelines.Rmd` and `20_analyses_reads_all_pipelines.Rmd`. Species overlap and uniqueness among pipelines were visualized with chord plots via `21_analyses_circlize_all_pipelines.Rmd`, illustrating shared and unique detections. Community structure was assessed using Principal Coordinates Analysis (PCoA) based on Bray–Curtis dissimilarity, with ordinations generated for individual river samples for each pipeline and marker via `22_analyses_PCoA_all_pieplines.Rmd`. Statistical differences in community composition were evaluated using PERMANOVA, and differences in dispersion among groups were tested using PERMDISP.

Results

Overview of Pipeline Outputs

All three pipelines successfully processed the six river eDNA datasets with the curated CRABS reference databases. Each pipeline recovered broadly comparable fish communities, but they differed in how finely they could identify taxa (species vs. genus), how many taxa they reported at each site, and which taxa were uniquely detected by one pipeline but not the others. Since the focus of this study was on replicating fish taxa, only fish results are reported here, and detections of birds, amphibians, and mammals were excluded from downstream analyses.

Fish Community Composition Across Pipelines

Anguillidae

Relative to Ballini et al. (2024), *Anguilla Anguilla* was consistently detected in the same four rivers (Argentina, Bevera, Nervia, Roia) with similar read counts by the eDNA-Container App with Tele02 (Table 2). APSCALE detected this species only in Bevera and at very low depth for both markers (Table 2-3). Barque-repliSTREAM produced read counts for both markers closely matching Ballini et al. (2024) but only at the genus level (*Anguilla sp.*) (Table 2-3).

Blenniidae

Salariopsis fluviatilis was consistently detected in Roia across all pipelines (Table 1, Figure 2 A-B), with near-identical read counts for Tele02 (Table 2). For Vert01, detections by Barque-repliSTREAM and APSCALE matched Ballini et al. (2024), but the eDNA-Container App did not detect this species (Table 3, Figure 2A-B).

Cottidae

Detection of *Cottus gobio* varied across workflows (Table 1, Figure 2A-B). Compared to Ballini et al. (2024), APSCALE expanded detections beyond Tanaro to all six rivers with high read counts for Vert01 (Table 3). Barque-repliSTREAM resolved only *Cottus* sp. with Vert01, and the eDNA-Container App did not detect this taxon with either marker (Tables 2–3, Figure 2A-B).

Cyprinidae

Barbus caninus was consistently detected in Argentina by APSCALE, Barque-repliSTREAM, and the eDNA-Container App with both markers, but was absent from Ballini et al. (2024) (Tables 2-3, Figure 2A-B).

Barbus meridionalis was detected in Bevera and Roia across all four pipelines with Tele02 (Figure 2A) with similar read counts among workflows (Table 2). With Vert01, this species was not detected by Ballini et al. (2024), but was detected in Bevera and Roia by APSCALE, Barque-repliSTREAM, and the eDNA-Container App (Table 3, Figure 2B), and additionally in Nervia by APSCALE and the eDNA-Container App (Table 3). *Barbus plebejus* was detected in Argentina, Bevera, and Roia by all four pipelines with both markers (Table 2-3, Figure 2A-B). In addition, APSCALE, Barque-repliSTREAM, and the eDNA-Container App extended

detections of *Barbus plebejus* to Nervia with Vert01, whereas Ballini et al. (2024) did not report the species at this site (Tables 1–3).

Gobiidae

Neogobius nigricans was consistently detected in Carpasina across all four pipelines (Table 1) for both markers (Table 2-3, Figure 2A-B), demonstrating strong reproducibility for this endangered species.

Leuciscidae

The genus *Phoxinus* showed clear differences among pipelines. In Ballini et al. (2024), *Phoxinus* sp. was reported in both Bevera and Roia only at the genus level for both markers, a finding reproduced by Barque-repliSTREAM (Tables 2–3, Figure 2A–B). In contrast, APSCALE and the eDNA-Container App resolved multiple species: *Phoxinus csikii* (APSCALE, Bevera and Roia, both markers; Table 2-3), *Phoxinus phoxinus* (APSCALE, Roia, Tele02 only, Table 2), *Phoxinus septimaniae* (eDNA-Container App, Bevera and Roia, Tele02, Table 2), and *Phoxinus bigerri* (APSCALE, Barque-repliSTREAM, and eDNA-Container App in Bevera, and eDNA-Container App in Roia ,Vert01, Table 3).

Barque-repliSTREAM was the only pipeline to detect *Rutilus pigus*, which appeared in Carpasina with Vert01 (Table3).

Squalius squalus was detected in Bevera and Roia across all workflows for both markers, with the exception of Bevera with Vert01 for the eDNA-Container App (Table 2-3). Barque-repliSTREAM additionally detected this species in Carpasina with Vert01, a detection absent from other pipelines (Table 3).

Telestes muticellus was consistently detected in all rivers except Tanaro, with near-identical read counts across workflows for Tele02 (Table2). With Vert01, only the eDNA-Container App detected this species (Table 3). In contrast, *Telestes souffia* was uniquely detected by APSCALE with Vert01 in all rivers except Tanaro. In the same rivers, Barque-Ballini et al. (2024) and Barque-repliSTREAM reported only *Telestes sp.* with Vert01 (Table 1,3, Figure 2B).

Salmonidae

Pipeline differences were also apparent in salmonid detections. With Tele02, Ballini et al. (2024) and Barque-repliSTREAM detected *Salmo* sp. in Bevera, Carpasina, and Tanaro (Table 2). In contrast, APSCALE uniquely resolved these detections with Tele02 at the species level to *Salmo carpio* and *Salmo marmoratus*, while the eDNA-Container App did not detect salmonids with any marker (Table 1, Figure 2A-B).

With Vert01, Ballini et al. (2024) and Barque-repliSTREAM again reported *Salmo* sp. in the same rivers and additionally in Argentina, but no salmonids were detected by APSCALE or the eDNA-Container App (Table 3).

Additionally, Barque-repliSTREAM uniquely identified *Oncorhynchus* sp. in Carpasina with Vert01, a result not produced by other pipelines (Table3).

Community Structure Across Rivers (PCoA)

PCoA showed that fish communities differed significantly among rivers for all pipelines and both markers (PERMANOVA, $p = 0.001$; Figure 3). These differences were not caused by uneven variability within river samples (PERMDISP, $p > 0.05$) but reflected real differences among rivers. Overall, APSCALE and Barque gave the most balanced separation of river communities with both markers. The eDNA-Container App also distinguished rivers, but most of the variation was compressed into one axis and it missed detections in Tanaro with Tele02, reducing resolution.

Discussion

This study shows that while all three pipelines recovered broadly similar fish communities from the same eDNA dataset, important differences emerged in taxonomic resolution and marker-specific detections. Relative to Ballini et al. (2024), APSCALE and the eDNA-Container App often resolved taxa to species level where Barque-repliSTREAM reported only genus-level assignments (e.g., *Anguilla* and *Phoxinus*). APSCALE also produced

unique detections, such as *Telestes souffia* with Vert01, while the eDNA-Container App did not recover some key taxa, including all salmonids.

The genus *Phoxinus* illustrates these contrasts. Barque consistently returned *Phoxinus* sp., while APSCALE and the eDNA-Container App resolved multiple species (*Phoxinus septimaniae, Phoxinus phoxinus, Phoxinus bigerri*). APSCALE also detected *Phoxinus csikii* in Bevera and Roia. This Danube-native minnow is now considered invasive in the Po basin (De Santis et al., 2021). Its appearance in these additional rivers may reflect further spread or misassignment within this cryptic complex (Palandačić et al., 2020).

Similar patterns were observed for other taxa. APSCALE and/or the eDNA-Container App detected *Phoxinus spetimaniae, Telestes souffia*, *Salmo carpio*, and *Salmo marmoratus*, whereas Barque workflows identified only *Telestes muticellus* and *Salmo* sp. Ballini et al. (2024) documented that *Phoxinus spetimaniae* and *Telestes souffia* are native to the six river basins based on historical and monitoring data. Their recovery here therefore supports the idea that more resolving pipelines may capture expected biodiversity overlooked by conservative workflows. In contrast, the exclusive detections of *Salmo carpio* and *Salmo marmoratus* with Tele02 by APSCALE should be interpreted with caution. *Salmo carpio* is endemic to Lake Garda (Randazzo et al., 2022), so detections outside this system likely reflect misassignment or DNA transport. *Salmo marmoratus*, while native to the Po basin and Adriatic drainages (Pengal et al., 2023), is also known to hybridize extensively with *Salmo trutta*, making species-level assignments particularly uncertain. These issues align with findings from Morey et al. (2024), who documented frequent resolution blind spots in Salmonidae due to limited marker discriminatory power and gaps in reference databases.

Pipeline usability and run time varied. All pipelines were well-documented and had similar set up requirements. APSCALE-CLI had the fastest run time (~2 minutes) followed by Barque (~5 minutes per run) and the eDNA-Container App (~10 minutes). APSCALE and Barque require command-line familiarity but offer flexibility and reproducibility. Although

APSCALE also has a GUI version, it was not tested here. The eDNA-Container App provided a beginner-friendly GUI, improving accessibility.

These findings highlight both strengths and trade-offs among different eDNA metabarcoding bioinformatic workflows. Alignment-based pipelines like Barque provide conservative but reproducible results, possibly at the cost of fine-scale resolution. APSCALE offers higher resolution and captured additional taxa, though some detections were at low read depth. The eDNA-Container App provides accessibility through a GUI but showed gaps in detection that may limit its use for certain applications. Despite these differences, all three workflows produced broadly consistent representations of northwestern Italian river fish communities, indicating that pipeline choice primarily affects taxonomic resolution rather than overall community patterns. For conservation and monitoring, Barque and APSCALE may offer stronger reliability, while the eDNA-Container App can expand accessibility but may need refinement to match the performance of CLI workflows.

Acknowledgements

We thank Ballini et al. (2024) for making their data and workflow openly available, which enabled this project to be reproduced as a learning exercise. Generative AI tools, including ChatGPT, DALL·E (OpenAI), and GitHub Copilot (GitHub, Inc.), were used to support coding, fish icon creation, editing, and documentation. Their use formed part of the learning process, helping to build technical skills and support clarity and reproducibility of this project.

Author Contributions

DS selected Ballini et al. (2024) as a suitable study to replicate as a learning project. NMA and DS collaborated to prepare the CRABS curated reference databases and to run the Barque-repliSTREAM bioinformatics analyses. NMA ran the APSCALE-CLI and eDNA-

Container App analyses. NMA and DS collaborated to run the post-bioinformatics analyses. NMA wrote the report.

Data Availability

A detailed workflow and all scripts used in this project are available at the GitHub page: https://github.com/nanthony80/repliSTREAM.

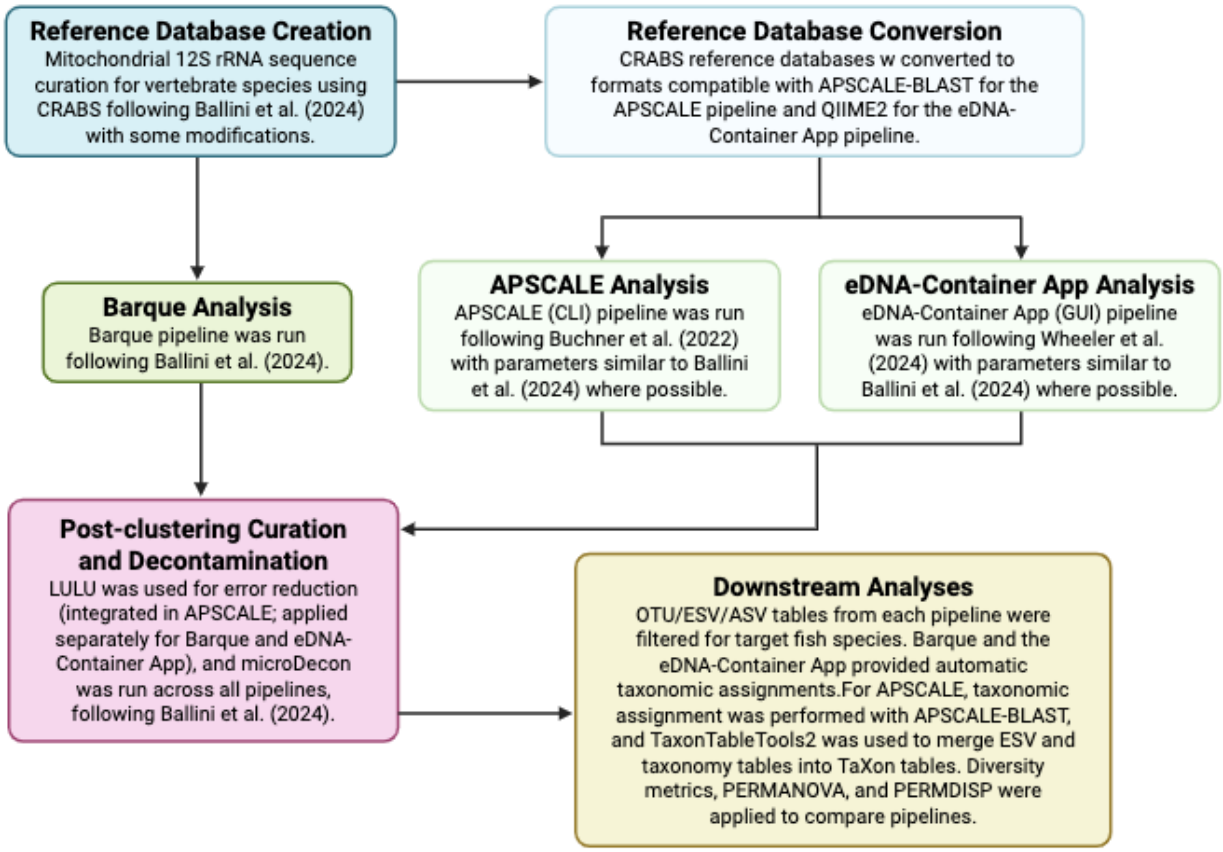
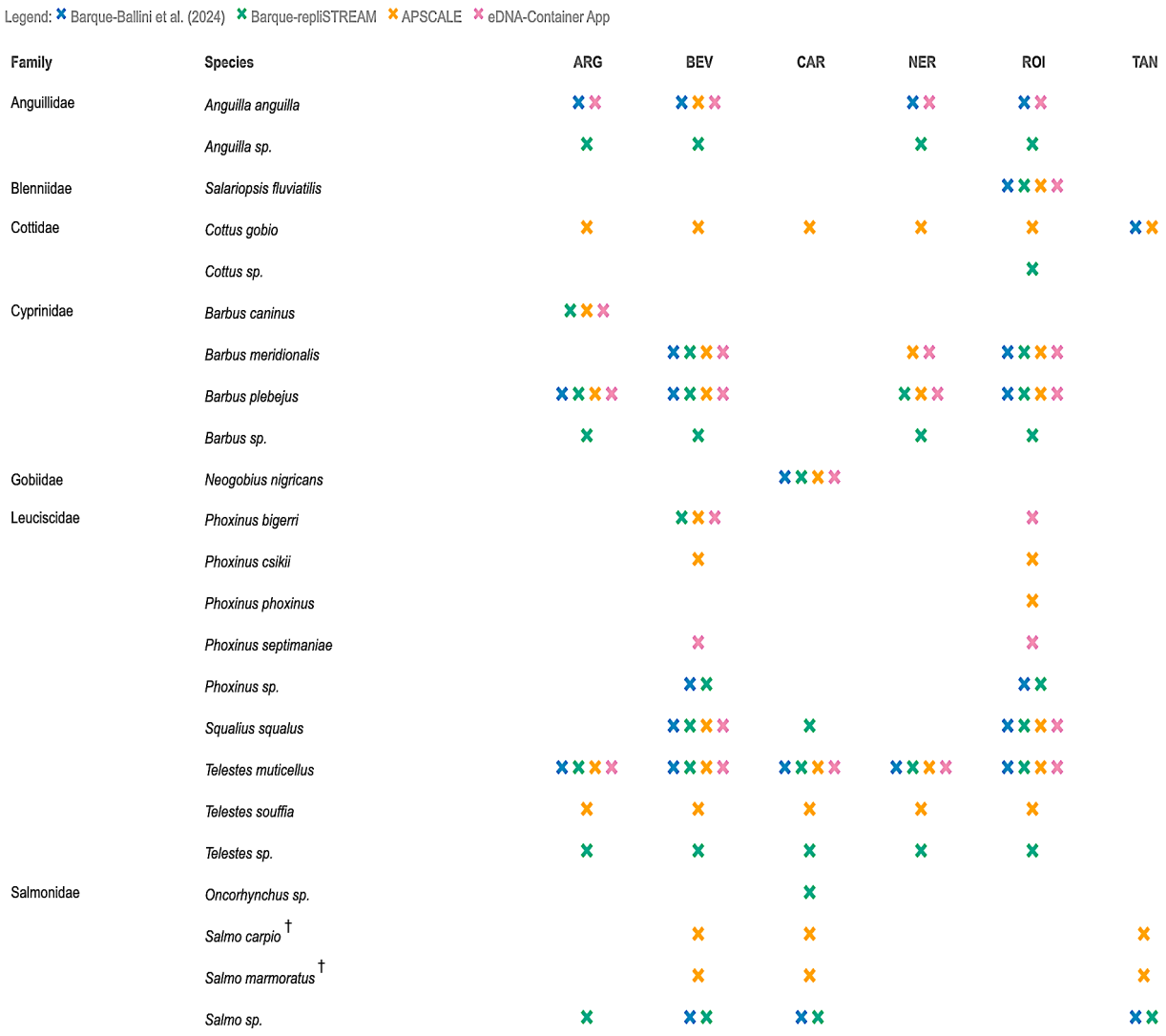


Figure 1: *Overview of the repliSTREAM bioinformatics workflow comparing Barque, APSCALE, and eDNA-Container App pipelines* (created with BioRender.com).

Table 1: Fish communities by river for each pipeline*



*Presence–absence data reflect combined detections from the mitochondrial 12S rRNA markers Tele02 and Vert01.

†Possible false-positive given literature consensus on species distribution (see Discussion).

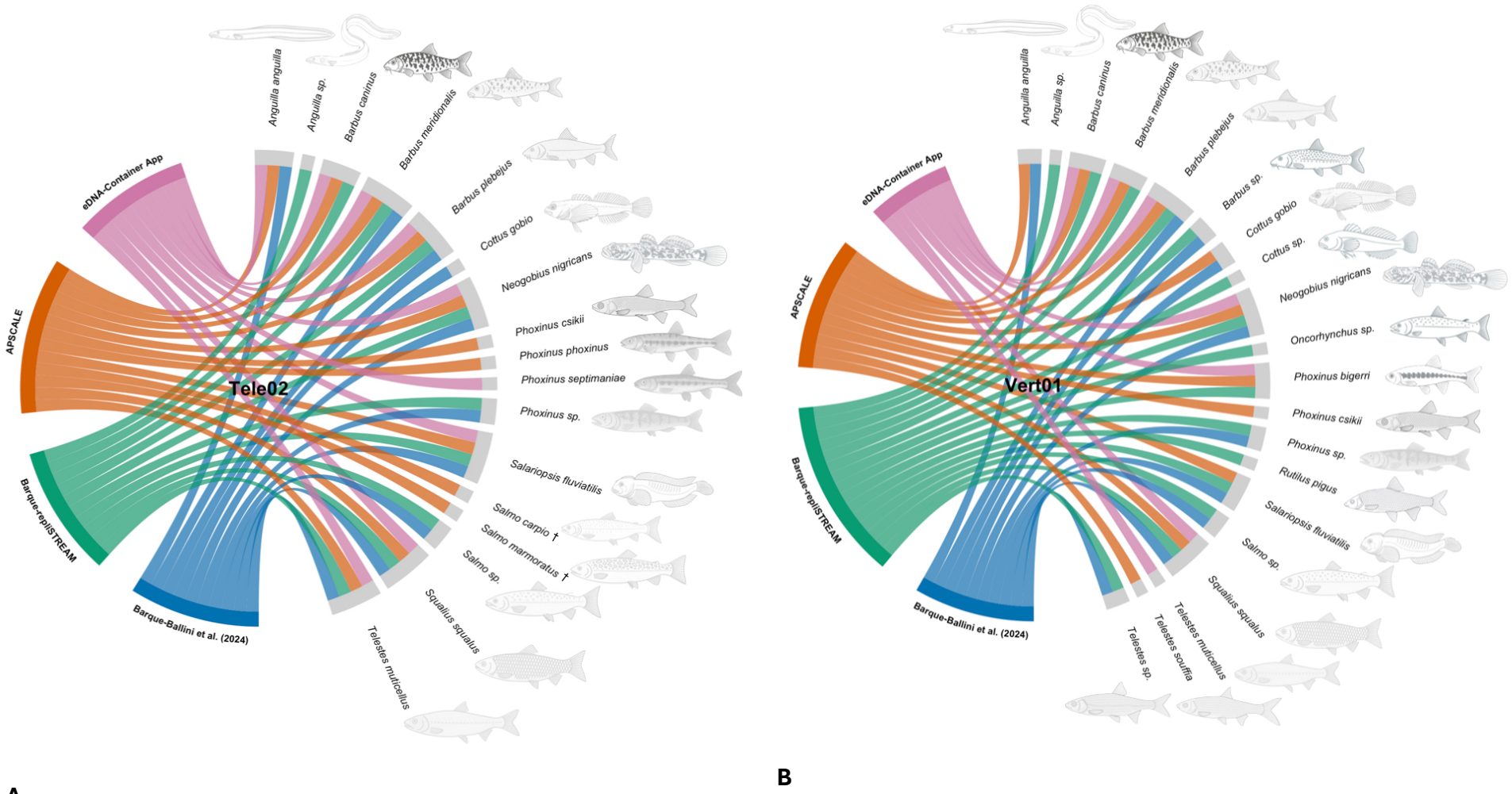
Table 2: Fish species and read counts per river by pipeline using the Tele02 primer pair.

Family	Species	Pipeline	ARG	BEV	CAR	NER	ROI	TAN
Anguillidae	*Anguilla anguilla*	APSCALE	0	16	0	0	0	0
Anguillidae	*Anguilla anguilla*	Barque-Ballini et al. (2024)	277	1457	0	12187	1838	0
Anguillidae	*Anguilla anguilla*	eDNA-Container App	290	1420	0	12626	1870	0
Anguillidae	*Anguilla sp.*	Barque-repliSTREAM	279	1465	0	12284	1850	0
Blenniidae	*Salariopsis fluviatilis*	APSCALE	0	0	0	0	306	0
Blenniidae	*Salariopsis fluviatilis*	Barque-Ballini et al. (2024)	0	0	0	0	391	0
Blenniidae	*Salariopsis fluviatilis*	Barque-repliSTREAM	0	0	0	0	392	0
Blenniidae	*Salariopsis fluviatilis*	eDNA-Container App	0	0	0	0	392	0
Cottidae	*Cottus gobio*	Barque-Ballini et al. (2024)	0	0	0	0	0	5327
Cyprinidae	*Barbus caninus*	APSCALE	525	0	0	0	0	0
Cyprinidae	*Barbus caninus*	Barque-repliSTREAM	727	0	0	0	0	0
Cyprinidae	*Barbus caninus*	eDNA-Container App	738	0	0	0	0	0
Cyprinidae	*Barbus meridionalis*	APSCALE	0	47408	0	0	1274	0
Cyprinidae	*Barbus meridionalis*	Barque-Ballini et al. (2024)	0	50040	0	0	1646	0
Cyprinidae	*Barbus meridionalis*	Barque-repliSTREAM	0	50505	0	0	1676	0
Cyprinidae	*Barbus meridionalis*	eDNA-Container App	0	51797	0	0	1702	0
Cyprinidae	*Barbus plebejus*	APSCALE	10029	8648	0	0	14876	0
Cyprinidae	*Barbus plebejus*	Barque-Ballini et al. (2024)	11670	9743	0	0	17045	0
Cyprinidae	*Barbus plebejus*	Barque-repliSTREAM	11790	10080	0	0	17160	0
Cyprinidae	*Barbus plebejus*	eDNA-Container App	11755	10237	0	0	17618	0
Gobiidae	*Neogobius nigricans*	APSCALE	0	0	33	0	0	0
Gobiidae	*Neogobius nigricans*	Barque-Ballini et al. (2024)	0	0	48	0	0	0
Gobiidae	*Neogobius nigricans*	Barque-repliSTREAM	0	0	48	0	0	0
Gobiidae	*Neogobius nigricans*	eDNA-Container App	0	0	48	0	0	0
Leuciscidae	*Phoxinus csikii*	APSCALE	0	31301	0	0	21013	0
Leuciscidae	*Phoxinus phoxinus*	APSCALE	0	0	0	0	87	0
Leuciscidae	*Phoxinus septimaniae*	eDNA-Container App	0	35834	0	0	24640	0
Leuciscidae	*Phoxinus sp.*	Barque-Ballini et al. (2024)	0	35330	0	0	24353	0
Leuciscidae	*Phoxinus sp.*	Barque-repliSTREAM	0	35725	0	0	24537	0
Leuciscidae	*Squalius squalus*	APSCALE	0	7965	0	0	20022	0
Leuciscidae	*Squalius squalus*	Barque-Ballini et al. (2024)	0	9722	0	0	23070	0
Leuciscidae	*Squalius squalus*	Barque-repliSTREAM	0	9776	0	0	23361	0
Leuciscidae	*Squalius squalus*	eDNA-Container App	0	9895	0	0	23673	0
Leuciscidae	*Telestes muticellus*	APSCALE	30808	715	27251	220197	448	0
Leuciscidae	*Telestes muticellus*	Barque-Ballini et al. (2024)	35343	940	29938	225808	585	0
Leuciscidae	*Telestes muticellus*	Barque-repliSTREAM	35207	919	30100	228580	585	0
Leuciscidae	*Telestes muticellus*	eDNA-Container App	35996	938	30457	235024	583	0
Salmonidae	*Salmo carpio*†	APSCALE	0	110	34457	0	0	6622
Salmonidae	*Salmo marmoratus*†	APSCALE	0	323	1347	0	0	1589
Salmonidae	*Salmo sp.*	Barque-Ballini et al. (2024)	0	571	39315	0	0	9678
Salmonidae	*Salmo sp.*	Barque-repliSTREAM	0	575	39605	0	0	9711

†Possible false-positive given literature consensus on species distribution (see Discussion).

Table 3: Fish species and read counts per river by pipeline using the Vert01 primer pair.

Family	Species	Pipeline	ARG	BEV	CAR	NER	ROI	TAN
Anguillidae	*Anguilla anguilla*	APSCALE	0	71	0	0	0	0
Anguillidae	*Anguilla anguilla*	Barque-Ballini et al. (2024)	0	4138	0	0	11237	0
Anguillidae	*Anguilla sp.*	Barque-repliSTREAM	0	4144	0	0	11277	0
Blenniidae	*Salariopsis fluviatilis*	APSCALE	0	0	0	0	105	0
Blenniidae	*Salariopsis fluviatilis*	Barque-Ballini et al. (2024)	0	0	0	0	115	0
Blenniidae	*Salariopsis fluviatilis*	Barque-repliSTREAM	0	0	0	0	115	0
Cottidae	*Cottus gobio*	APSCALE	6713	2301	4962	11394	6026	28305
Cottidae	*Cottus gobio*	Barque-Ballini et al. (2024)	0	0	0	0	0	16423
Cottidae	*Cottus sp.*	Barque-repliSTREAM	0	0	0	0	4351	0
Cyprinidae	*Barbus caninus*	APSCALE	3443	0	0	0	0	0
Cyprinidae	*Barbus caninus*	Barque-repliSTREAM	3554	0	0	0	0	0
Cyprinidae	*Barbus caninus*	eDNA-Container App	3528	0	0	0	0	0
Cyprinidae	*Barbus meridionalis*	APSCALE	0	135453	0	814	13318	0
Cyprinidae	*Barbus meridionalis*	Barque-repliSTREAM	0	1765	0	0	564	0
Cyprinidae	*Barbus meridionalis*	eDNA-Container App	0	80800	0	917	14174	0
Cyprinidae	*Barbus plebejus*	APSCALE	61981	26869	0	852	92556	0
Cyprinidae	*Barbus plebejus*	Barque-Ballini et al. (2024)	57489	137504	0	0	100284	0
Cyprinidae	*Barbus plebejus*	Barque-repliSTREAM	57881	139134	0	1612	100859	0
Cyprinidae	*Barbus plebejus*	eDNA-Container App	64242	28224	0	980	95301	0
Cyprinidae	*Barbus sp.*	Barque-Ballini et al. (2024)	0	1286	0	0	149	0
Cyprinidae	*Barbus sp.*	Barque-repliSTREAM	19	420	0	10	168	0
Gobiidae	*Neogobius nigricans*	APSCALE	0	0	296	0	0	0
Gobiidae	*Neogobius nigricans*	Barque-Ballini et al. (2024)	0	0	317	0	0	0
Gobiidae	*Neogobius nigricans*	Barque-repliSTREAM	0	0	318	0	0	0
Gobiidae	*Neogobius nigricans*	eDNA-Container App	0	0	342	0	0	0
Leuciscidae	*Phoxinus bigerri*	APSCALE	0	84	0	0	0	0
Leuciscidae	*Phoxinus bigerri*	Barque-repliSTREAM	0	85	0	0	0	0
Leuciscidae	*Phoxinus bigerri*	eDNA-Container App	0	118	0	0	453	0
Leuciscidae	*Phoxinus csikii*	APSCALE	0	101473	0	0	136051	0
Leuciscidae	*Phoxinus sp.*	Barque-Ballini et al. (2024)	0	98133	0	0	131641	0
Leuciscidae	*Phoxinus sp.*	Barque-repliSTREAM	0	99054	0	0	133031	0
Leuciscidae	*Rutilus pigus*	Barque-repliSTREAM	0	0	1090	0	0	0
Leuciscidae	*Squalius squalus*	APSCALE	0	22619	0	0	132267	0
Leuciscidae	*Squalius squalus*	Barque-Ballini et al. (2024)	0	22384	0	0	128453	0
Leuciscidae	*Squalius squalus*	Barque-repliSTREAM	0	22567	12	0	129831	0
Leuciscidae	*Squalius squalus*	eDNA-Container App	0	0	0	0	136222	0
Leuciscidae	*Telestes muticellus*	eDNA-Container App	149216	4353	58786	153284	0	0
Leuciscidae	*Telestes souffia*	APSCALE	146352	3807	56824	149069	6203	0
Leuciscidae	*Telestes sp.*	Barque-Ballini et al. (2024)	139038	4138	54509	142545	6480	0
Leuciscidae	*Telestes sp.*	Barque-repliSTREAM	136557	4108	54707	143268	6540	0
Salmonidae	*Oncorhynchus sp.*	Barque-repliSTREAM	0	0	651	0	0	0
Salmonidae	*Salmo sp.*	Barque-Ballini et al. (2024)	320	911	120284	0	0	15271
Salmonidae	*Salmo sp.*	Barque-repliSTREAM	320	911	120367	0	0	15348



A B

Figure 2: Fish taxa detected from eDNA by pipeline for the Tele02 (A) and Vert01 (B) 12S markers. Chord plots illustrate species detected by each pipeline and their assignment to the respective marker. Chord diagrams were generated in R using the package *circlize* (Gu et al., 2014). Fish icons were created using ChatGPT's *DALL-E* generative image tool (OpenAI, 2025).

†Possible false-positive given literature consensus on species distribution (see Discussion).

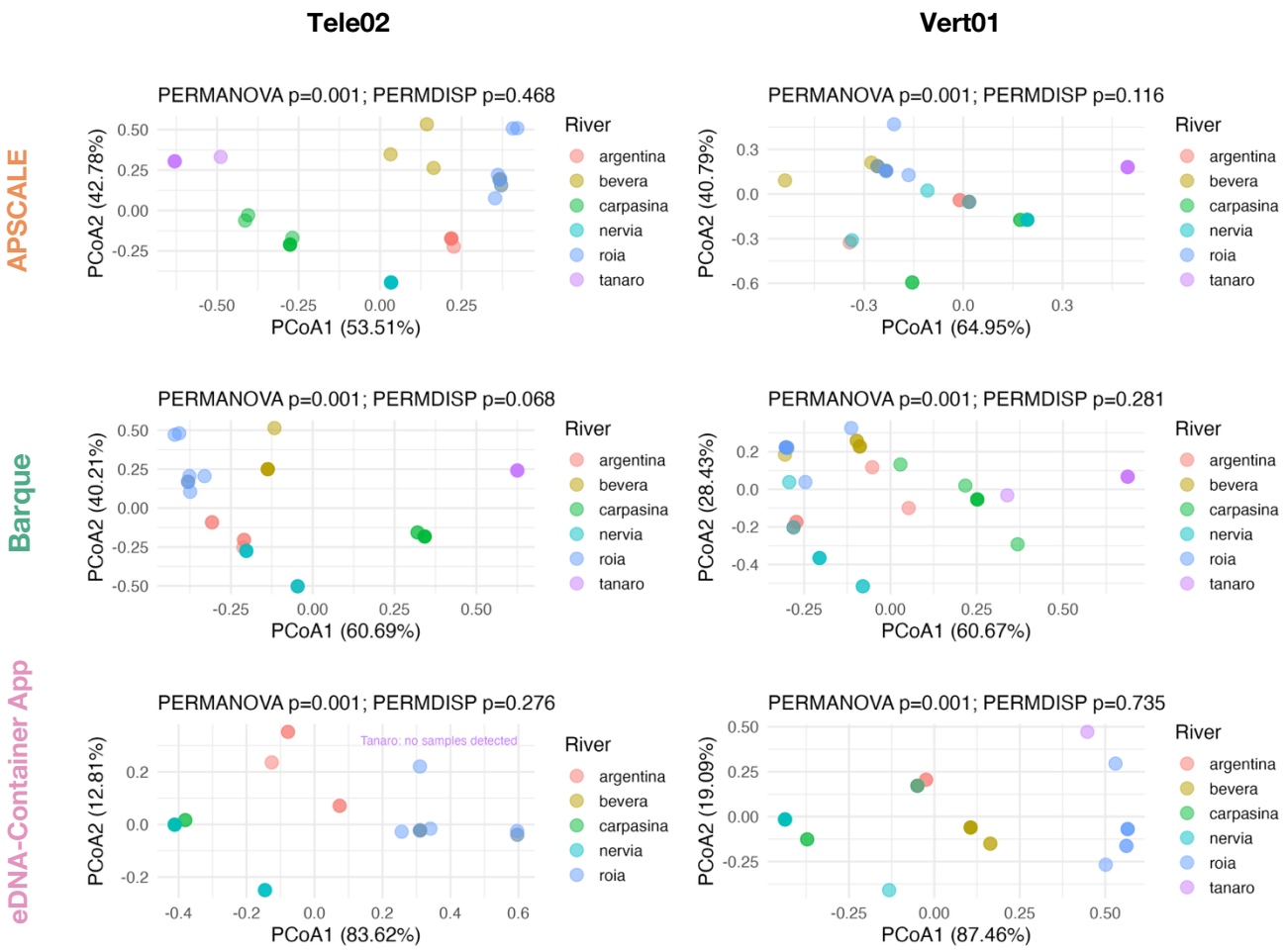


Figure 3: PCoA with ordinations for individual river samples for each pipeline (horizontal) and marker (vertical). Statistical differences in community composition were evaluated using PERMANOVA, and differences in dispersion among groups were tested using PERMDISP.

References

Ballini, Lorenzo & Staffoni, Giorgia & Nespoli, Davide & Dario, Ottonello & Candiotto, Alessandro & Forte, Simone & Vezza, Paolo & Iannucci, Alessio & Fratini, Sara. (2024). Environmental DNA metabarcoding as an efficient tool to monitor freshwater systems in northwestern Italy. Hydrobiologia. 852. 791-803. 10.1007/s10750-024-05723-y.

Buchner, D., Macher, T. H., & Leese, F. (2022). APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. *Bioinformatics (Oxford, England)*, *38*(20), 4817–4819. https://doi.org/10.1093/bioinformatics/btac588

De Santis, V., Delmastro, G.B., Vanetti, I. *et al.* (2021). Species composition of introduced and natural minnow populations of the *Phoxinus* cryptic complex in the westernmost part of the Po River Basin (north Italy). *Biol Invasions* 23, 657–668 https://doi.org/10.1007/s10530-020-02406-2

dos Santos RA, Blabolil P. Comparison of Bioinformatic Pipelines for eDNA Metabarcoding Data Analysis of Fish Populations. *Fishes*. 2025; 10(5):214. https://doi.org/10.3390/fishes10050214

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L.A. *et al.* SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinformatics* **20**, 88 (2019). https://doi.org/10.1186/s12859-019-2663-2

GitHub, Inc. (2025). *GitHub Copilot* [AI code generation tool]. https://github.com/features/copilot

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics (Oxford, England)*, *30*(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature communications*, *8*(1), 1188. https://doi.org/10.1038/s41467-017-01312-x

Hakimzadeh, A., Abdala Asbun, A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J. G., Curd, E., Djemiel, C., Brandström Durling, M., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandeau, E., Özkurt, E., M Palmer, J., Pascal, G., ... Anslan, S. (2024). A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. *Molecular ecology resources*, *24*(5), e13847. https://doi.org/10.1111/1755-0998.13847

Macher, T. H., Beermann, A. J., & Leese, F. (2021). TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. *Molecular ecology resources*, *21*(5), 1705–1714. https://doi.org/10.1111/1755-0998.13358

McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA*. 2019; 1: 14–25. https://doi.org/10.1002/edn3.11

Morey, K.C., Myler, E., Hanner, R. and Tetreault, G. (2024), Taxonomic Blind Spots: A Limitation of Environmental DNA Metabarcoding-Based Detection for Canadian Freshwater Fishes. Environmental DNA, 6: e70054. https://doi.org/10.1002/edn3.70054

Nearing, J. T., Douglas, G. M., Comeau, A. M., & Langille, M. G. I. (2018). Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ*, *6*, e5364. https://doi.org/10.7717/peerj.5364

Normandeau, E. (2024). *Barque v1.8.5: Environmental DNA metabarcoding analysis* [Computer software]. GitHub. https://github.com/enormandeau/barque

OpenAI. (2025). *ChatGPT* (March 2025 version) [Large language model]. https://chat.openai.com/

OpenAI. (2025). *DALL·E* [Generative image model]. https://openai.com/dall-e

Palandačić, A., Kruckenhauser, L., Ahnelt, H., & Mikschi, E. (2020). European minnows through time: museum collections aid genetic assessment of species introductions in freshwater fishes (Cyprinidae: Phoxinus species complex). *Heredity*, *124*(3), 410–422. https://doi.org/10.1038/s41437-019-0292-1

Pengal, P., Cokan, B., Økland, F., Höjesjö, J., Tambets, M., & Thorstad, E. B. (2023). Comparing dominance relationships and movement of native marble trout (Salmo marmoratus) and introduced rainbow trout (Oncorhynchus mykiss). *Journal of fish biology*, *102*(4), 883–892. https://doi.org/10.1111/jfb.15333

Pochon, X., Bowers, H. A., Zaiko, A., & Wood, S. A. (2025). Advancing the environmental DNA and RNA toolkit for aquatic ecosystem monitoring and management. *PeerJ*, *13*, e19119. https://doi.org/10.7717/peerj.19119

Randazzo, B., Zarantoniello, M., Secci, G., Faccenda, F., Fava, F., Marzorati, G., Belloni, A., Maradonna, F., Orazi, V., Cerri, R., Povinelli, M., Parisi, G., Giorgini, E., & Olivotto, I. (2022). Towards the Identification of a Suitable Commercial Diet for Carpione (*Salmo carpio*,

Linnaeus 1758): A Multidisciplinary Study on Fish Performances, Animal Welfare and Quality Traits. *Animals*, *12*(15), 1918. https://doi.org/10.3390/ani12151918

Ratnasingham S (2019) mBRAVE: The Multiplex Barcode Research And Visualization Environment. Biodiversity Information Science and Standards 3: e37986. https://doi.org/10.3897/biss.3.37986

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic acids research*, *39*(21), e145. https://doi.org/10.1093/nar/gkr732

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, 17, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Staffoni, G. (accessed 31, August 2025). STREAM. GitHub. https://github.com/giorgiastaffoni/STREAM

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, Pierre & Bonin, Aurélie & Zinger, Lucie & Coissac, Éric. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, Oxford, U.K.

Wheeler, D., Brancalion, L., Kawasaki, A., & Rourke, M. L. (2024). The eDNA-Container App: A Simple-to-Use Cross-Platform Package for the Reproducible Analysis of eDNA Sequencing Data. *Applied Sciences*, *14*(6), 2641. https://doi.org/10.3390/app14062641

Wheeler, D. (accessed 31, August 2025). The eDNA-container app. GitHub. https://github.com/dwheelerau/edna-container