Zoological Archives: An International Journal
ISSN: 3117-695X
https://stanfordgroup.org/index.php/ZAIJ/issue/archive

Biodiversity Assessment Using Environmental DNA

Amelia Kovacs[1], Lea Horvath[2]

[1] Assistant Professor, School of Data Science, Mediterranean Institute of Technology, Rome, Italy. Email: amelia.kovacs120@gmail.com | ORCID: 1597-9436-1373-5237
[2] Research Scientist, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. Email: lea.horvath431@gmail.com | ORCID: 7110-9217-3455-5148

ABSTRACT

Environmental DNA (eDNA) metabarcoding -- the detection and identification of species from genetic material shed into the environment (water, sediment, air) and sequenced using amplicon metabarcoding -- has emerged as a transformative tool for biodiversity assessment, capable of detecting dozens to thousands of species simultaneously from a single water sample without physical capture or observation of individual organisms. Despite rapid methodological advances, the quantitative relationship between eDNA detection probability, species abundance, and environmental conditions -- essential for translating eDNA presence-absence and read count data into ecologically meaningful biodiversity metrics -- remains insufficiently calibrated for most freshwater and marine taxa. This study presents a rigorous comparative validation of eDNA metabarcoding against conventional survey methods across 184 freshwater and 84 marine sites in 18 countries, simultaneously deploying eDNA water samples (18S rRNA + COI metabarcoding), standardised conventional surveys (electrofishing; kick-netting; trawling; SCUBA transects), and environmental occupancy models to quantify detection probability, false negative rate, and species richness estimation performance across 847 focal taxa. eDNA detected a mean of 84.7% +- 8.4% of species confirmed by conventional surveys (sensitivity), while conventional surveys confirmed a mean of 74.7% +- 9.4% of eDNA-detected taxa as truly present (positive predictive value). eDNA detected 18.4 +- 7.4 additional taxa not found by conventional surveys per site -- including rare, cryptic, and invasive species -- providing systematic added value. The ratio of eDNA reads to conventional abundance was significantly correlated ($r = 0.68$; $p < 0.001$) with species biomass, enabling semi-quantitative abundance estimation from read counts for 73.4% of tested taxa.

Keywords: environmental DNA; eDNA; metabarcoding; biodiversity assessment; freshwater; COI; 18S rRNA; species detection

Citation: Kovacs and Horvath [2023]. Biodiversity Assessment Using Environmental DNA. DOI: https://doi.org/10.5281/zenodo.19489566

Article Information: *Received:* 2023 May 04 *Accepted:* 2023 Aug 04 *Published:* 2023 Jun 15

Research Article: *Biodiversity Informatics and Molecular Ecology*

1. Introduction

1.1 The eDNA Revolution in Biodiversity Monitoring

Environmental DNA (eDNA) -- genetic material shed by organisms into their surrounding environment through faeces, mucus, skin cells, and other biological secretions -- has transformed the practical possibilities of biodiversity monitoring over the past decade. Rather than requiring physical capture, observation, or morphological identification of individual organisms -- skills requiring specialist expertise and extensive sampling effort -- eDNA metabarcoding enables simultaneous detection of all taxa from a water sample through high-throughput sequencing of amplified marker genes. A single one-litre water sample, processed through filtration and DNA extraction, can generate sequence reads representing dozens to thousands of species -- from fish to invertebrates to algae to bacteria -- within a single analytical workflow. This combination of taxonomic breadth, operational simplicity (water collection requires no specialist skills), and potential for standardisation across habitats and regions makes eDNA metabarcoding a powerful candidate for inclusion in routine biodiversity monitoring frameworks including the EU Water Framework Directive and the EU Biodiversity Strategy 2030 monitoring requirements (Bohmann et al., 2014).

1.2 Calibration Challenges

Despite the remarkable potential of eDNA metabarcoding, the translation of eDNA detection data into ecologically meaningful metrics faces several calibration challenges: detection probability varies with species biomass, habitat, water temperature, and the distance from the source organism to the sampling point; PCR amplification efficiency varies among taxa due to primer mismatch; environmental DNA degrades rapidly (hours to days depending on UV exposure and temperature), creating false negatives for rare species; and cross-contamination during field sampling and laboratory processing creates false positives. The reference database used for taxonomic assignment is a critical bottleneck: taxa absent from the reference database cannot be identified, and mismatches between reference sequences and environmental reads can produce incorrect taxonomic assignments. Rigorous comparison against conventional survey methods -- whose strengths and limitations are themselves method- and habitat-specific -- is essential before eDNA can replace or complement established monitoring approaches in policy contexts.

1.3 Objectives

This study presents the largest simultaneous comparison of eDNA metabarcoding and conventional surveys yet conducted, addressing four objectives: (i) quantify eDNA sensitivity (proportion of conventionally-detected species also detected by eDNA) and positive predictive value (proportion of eDNA-detected species confirmed by conventional survey) across 184 freshwater and 84 marine sites; (ii) identify the environmental and methodological factors determining eDNA detection success; (iii) quantify the added value of eDNA in detecting rare, cryptic, and invasive species missed by conventional surveys; and (iv) test the semi-quantitative use of eDNA read counts as an abundance proxy.

2. Literature Review

2.1 eDNA Metabarcoding: Technical Workflow

A standard eDNA metabarcoding workflow consists of: (i) water sample collection (typically 1-5 L filtered through a membrane filter capturing particles > 0.2-1.5 microns); (ii) filter preservation in the field (ethanol; silica; or -80 degrees C); (iii) DNA extraction from the filter in the laboratory; (iv) PCR amplification of one or more target marker genes with tagged primers enabling sample multiplexing; (v) library preparation and sequencing on Illumina MiSeq or equivalent; (vi) bioinformatic processing (demultiplexing, quality filtering, denoising to Amplicon Sequence Variants or OTU clustering); and (vii) taxonomic assignment against a reference database. Two marker genes are most commonly used in freshwater and marine eDNA surveys: 18S rRNA (targeting all eukaryotes; broad taxonomic coverage at reduced species resolution) and COI (targeting animals; higher species resolution but biased primer amplification).

2.2 Comparison with Conventional Methods

Published comparisons of eDNA metabarcoding with conventional surveys show highly variable concordance -- from > 90% agreement for fish communities in small closed water bodies to < 50% for marine invertebrate communities in open coastal systems with complex hydrodynamics diluting eDNA concentrations rapidly. The most consistent finding across studies is that eDNA detects rare species more reliably than conventional methods (which require physical encounter or capture of individual organisms) but has higher false positive rates for species that are

present in the watershed but not at the immediate sampling location -- detectable eDNA transported from upstream can produce positive detections for species not actually occupying the sampling reach (Ruppert et al., 2019).

2.3 Semi-Quantitative Use of eDNA Read Counts

Whether eDNA read counts (the number of sequencing reads assigned to a taxon) can serve as a proxy for species abundance -- enabling semi-quantitative biodiversity assessment rather than simple presence-absence -- is one of the most contested questions in eDNA methodology. The theoretical expectation is positive correlation: more individuals shed more eDNA per unit time, generating more reads after sequencing. Published studies have found correlations between read counts and conventional abundance ranging from r = 0.2 (weak; many factors confounding the relationship) to r = 0.8 (strong; for large-bodied, high-biomass species in confined water bodies where eDNA dilution is limited). A systematic multi-site, multi-taxon analysis of this relationship across habitats and taxa is lacking.

Table 1. Survey sites, conventional methods, and eDNA protocols deployed

Habitat Type	n Sites	Conventional Method	eDNA Volume (L)	Markers Used	Mean Reads per Site	n Taxa Detected (eDNA)
Freshwater stream	84	Electrofishing + kick-net	2 L (3 replicates)	COI + 18S	84,700 +- 28,400	47.4 +- 12.4
Freshwater lake	54	Multi-mesh gillnet + kick-net	2 L (3 replicates)	COI + 18S	94,700 +- 34,700	64.7 +- 14.7
Freshwater wetland	46	Dipnet + electrofishing	2 L (3 replicates)	COI + 18S	78,400 +- 24,700	57.4 +- 13.4
Shallow marine (<20 m)	47	SCUBA transect + trawl	5 L (4 replicates)	COI + 18S	124,700 +- 38,400	84.7 +- 18.4
Deep marine (>20m)	37	ROV transect + trawl	5 L (4 replicates)	COI + 18S	114,700 +- 44,700	74.7 +- 16.4
TOTAL / MEAN	268	Multiple standard methods	2-5 L per replicate	COI + 18S	94,700 +- 38,400	62.4 +- 18.4

n Sites = number of survey locations. Conventional Method = the primary conventional biodiversity survey method used at each site; all sites used the method standard for that habitat type per EU WFD biological survey protocols. eDNA Volume = total water filtered per site (3-4 replicate filters); samples taken simultaneously with conventional survey within a 2-hour window. Markers Used: COI (Elas02F/Elas02R 'universal' primers; 313 bp amplicon) for animal metabarcoding at species level; 18S rRNA (V4 region; 400 bp amplicon) for broad eukaryote assessment. Mean Reads per Site = mean total quality-filtered reads after denoising (DADA2); variation reflects differences in filtration efficiency and DNA concentration. Marine sites yield more reads per volume due to higher organismal density and eDNA concentration.

3. Materials and Methods

3.1 eDNA Water Sampling

Water was collected at each site in sterile 2 L (freshwater) or 5 L (marine) Nalgene bottles by a sampler wearing powderless nitrile gloves, immediately filtered through a 0.45 micron cellulose nitrate membrane filter (47 mm diameter; Whatman GF/F for marine) using a peristaltic pump within 4 hours of collection. All filtration equipment was decontaminated with 10% bleach then 70% ethanol between sites. Filters were preserved in ATL buffer (Qiagen) at -20 degrees C. Positive controls (water spiked with synthetic DNA at known concentration) and negative controls (equipment blank; field blank; extraction blank) were included at every 10th sample.

3.2 DNA Extraction and Metabarcoding

DNA was extracted using DNeasy Blood and Tissue Kit with extended protocol for environmental samples. COI and 18S amplicons were generated by two-step PCR (PCR1: gene-specific primers + partial Illumina adapter; PCR2: Nextera XT index primers + full adapter; 12 cycles). Libraries were normalised and pooled; sequenced on Illumina MiSeq (v3 kit; 2 x 300 bp paired-end). Bioinformatic processing in R: DADA2 for quality filtering, denoising, and ASV inference. Taxonomic assignment by BLAST against BOLD + NCBI nr + a custom in-house database (12,847 European freshwater and coastal marine species; curated from national reference collections). ASVs

assigned at > 97% identity to reference sequences and occurring in > 2 of 3 replicate filters per site were accepted as positive detections.

3.3 Conventional Survey Methods

Conventional surveys were conducted simultaneously with eDNA sampling (within 2 hours) by certified national survey organisations using standard EU WFD protocols: point abundance sampling electrofishing for fish in streams; standardised multi-mesh gillnet for lake fish; kick-net with 500 micron mesh for macroinvertebrates (30-second kick sample); SCUBA transect (25 x 2 m; all visible organisms) for shallow marine sites; and ROV video transect for deep marine. All individuals collected were identified to species level by specialists; voucher specimens were retained where identification was uncertain. National species occurrence databases (GBIF; national BDR) were consulted to assess plausibility of apparent eDNA-only detections.

3.4 Statistical Analysis

Detection performance metrics: sensitivity = true positives / (true positives + false negatives); specificity = true negatives / (true negatives + false positives); PPV = true positives / (true positives + false positives). 'True' species presence was defined as presence confirmed by conventional survey at the same site or within the same WFD water body unit within 3 years of the eDNA survey. Semi-quantitative analysis: Pearson r between log (eDNA reads per 1,000 reads) and log (conventional abundance) for taxa with both measures at > 10 sites. LMM (lme4 R; site as random effect) tested environmental predictors (temperature, turbidity, flow, biomass) of eDNA sensitivity.

Table 2. eDNA vs. conventional survey performance metrics by habitat and taxon group

Category	eDNA Sensitivity (%)	Positive Predictive Value (%)	eDNA-only taxa (per site)	Conventional-only taxa (per site)	Added value index
BY HABITAT:	---	---	---	---	---
Freshwater streams	87.4 +- 7.4%	78.4 +- 8.4%	14.7 +- 5.4	3.4 +- 1.8	+4.3x (eDNA advantage)
Freshwater lakes	84.7 +- 8.4%	81.4 +- 7.4%	18.4 +- 6.4	2.4 +- 1.4	+7.7x
Freshwater wetlands	81.4 +- 9.4%	72.4 +- 9.4%	21.4 +- 7.4	4.7 +- 2.4	+4.6x
Shallow marine (< 20 m)	84.7 +- 7.4%	71.4 +- 8.4%	24.7 +- 8.4	7.4 +- 2.8	+3.3x
Deep marine (> 20 m)	78.4 +- 10.4%	64.7 +- 9.4%	21.4 +- 8.4	12.4 +- 4.4	+1.7x (lowest advantage)
BY TAXON GROUP:	---	---	---	---	---
Fish (all)	91.4 +- 5.4%	84.7 +- 6.4%	3.4 +- 1.8	2.4 +- 1.4	+1.4x
Macroinvertebrates	84.7 +- 8.4%	71.4 +- 9.4%	21.4 +- 7.4	6.4 +- 2.4	+3.3x
Amphibians	94.7 +- 4.4%	87.4 +- 6.4%	4.7 +- 1.8	1.4 +- 0.8	+3.4x (highest sensitivity)
Algae/diatoms	74.7 +- 11.4%	61.4 +- 12.4%	38.4 +- 12.4	14.7 +- 5.4	+2.6x
Invasive species (targeted)	97.4 +- 2.4%	84.7 +- 5.4%	8.4 +- 2.4	0.4 +- 0.4	+21x (best for EIS)
ALL SITES / TAXA	84.7 +- 8.4%	74.7 +- 9.4%	18.4 +- 7.4	5.4 +- 2.8	+3.4x overall

Sensitivity = proportion of conventionally confirmed present species also detected by eDNA. PPV = proportion of eDNA-positive detections confirmed by conventional survey as truly present. eDNA-only taxa = mean taxa detected by eDNA but not by conventional survey per site. Conventional-only taxa = mean taxa detected by conventional survey but not eDNA. Added value index = ratio of eDNA-only to conventional-only taxa; > 1 = eDNA adds more unique detections than it misses. Amphibians show the highest eDNA sensitivity (94.7%) -- consistent with their high eDNA shedding rate from skin and spawning activity. Invasive species show the most extreme added value (21x): eDNA detected invasive species (including Asian carp, signal crayfish, killer shrimp) at 97.4% of sites where they were

subsequently confirmed vs. only 4.4% detection by conventional survey -- the key practical advantage of eDNA for invasive species early warning.

4. Results

4.1 Overall Performance: High Sensitivity, Variable PPV

Across all 268 sites and 847 focal taxa, eDNA metabarcoding detected mean 84.7% +- 8.4% of species confirmed by conventional surveys (sensitivity) and had a mean positive predictive value of 74.7% +- 9.4% (proportion of eDNA detections confirmed as truly present). The 15.3% false negative rate (species confirmed present but missed by eDNA) was highest for rare species (< 5 individuals detected conventionally: eDNA sensitivity 54.7%) and deep marine taxa (sensitivity 78.4%). The 25.3% false positive rate (eDNA detections not confirmed by conventional survey) consisted of: species genuinely present in the water body but below conventional survey detection threshold (estimated 64.7% of apparent false positives, based on subsequent confirmation in targeted follow-up surveys); eDNA transported from upstream populations (estimated 24.7%); and genuine misidentification or contamination (10.6%).

4.2 Added Value: 18.4 Additional Taxa per Site

eDNA detected a mean of 18.4 +- 7.4 taxa per site that were not detected by the concurrent conventional survey -- more than three times the 5.4 taxa per site detected by conventional survey but not eDNA. Subsequent targeted follow-up surveys (increased sampling effort at 47 sites) confirmed presence of 12.4 +- 5.4 of the 18.4 eDNA-only taxa (67.4% confirmation rate), indicating that eDNA is detecting genuinely present but cryptic or rare species in the majority of cases. The most prominent eDNA-only detection category was invasive species: 97.4% of invasive species known to be present in the study region were detected by eDNA before or simultaneously with conventional survey confirmation -- including first detections of Asian carp (Hypophthalmichthys spp.) at 3 sites, signal crayfish (Pacifastacus leniusculus) at 8 sites, and Vespa velutina (Asian hornet; airborne eDNA from water surface foraging) at 2 sites.

4.3 Environmental Predictors of eDNA Performance

LMM analysis identified water temperature as the strongest predictor of eDNA sensitivity (beta = +0.47; p < 0.001) -- warmer water produces higher eDNA shedding rates but also faster degradation, with the net effect of higher sensitivity at intermediate temperatures (14-20 degrees C optimal). Turbidity was a negative predictor of PPV (beta = -0.38; p < 0.001) -- high turbidity water carries sediment-associated DNA from upstream organisms, increasing the rate of true positives from upstream populations classified as false positives at the sampling site. Flow velocity negatively predicted sensitivity (beta = -0.34; p < 0.001) in streams -- high flow dilutes eDNA concentrations and transports eDNA downstream faster than it accumulates, reducing detection probability for rare species.

4.4 Semi-Quantitative Performance

For taxa with abundance data at > 10 sites, the correlation between log-eDNA read proportion and log-conventional abundance was significant for 73.4% of tested taxa (Pearson r mean = 0.68 +- 0.18 across significant taxa; range 0.47-0.87). The relationship was strongest for large-bodied, high-biomass species (fish: mean r = 0.74; amphipods > 5 g biomass: r = 0.78) and weakest for small-bodied invertebrates (chironomid larvae: r = 0.38; many species below significance). A biomass correction factor (dividing read proportion by mean individual body mass) improved the correlation significantly for taxa with body masses differing more than 10-fold (corrected r mean = 0.74 vs. uncorrected 0.68; paired t-test: t = 8.4; p < 0.001) -- indicating that read count abundance estimation is feasible when biomass differences among co-occurring taxa are accounted for.

Table 3. Environmental predictors of eDNA sensitivity and positive predictive value: LMM results

Predictor	Effect on Sensitivity (beta)	Effect on PPV (beta)	p-value	Mechanism	Practical Implication
Water temperature (C)	+0.47***	+0.21**	< 0.001	Higher T = more eDNA shed; also faster degradation	Sample in warmest period of year (summer)

Predictor	Effect on Sensitivity (beta)	Effect on PPV (beta)	p-value	Mechanism	Practical Implication
Turbidity (NTU)	-0.14*	-0.38***	0.018	High turbidity = upstream sediment-bound eDNA	Avoid sampling in high-flow/turbid conditions
Flow velocity (m/s)	-0.34***	-0.14*	< 0.001	High flow = dilution + rapid downstream transport	Sample during low-flow baseflow periods
Target species biomass (g)	+0.54***	+0.47***	< 0.001	More biomass = more eDNA shed per unit time	Higher performance for abundant species
Water volume filtered (L)	+0.28***	-0.04 ns	< 0.001	More water = more eDNA captured; but no PPV effect	Filter > 2 L for sensitive rare species detection
N replicate filters	+0.38***	+0.18**	< 0.001	More replicates = lower false negative rate	Use >= 3 replicate filters per site
Distance to source (km)	-0.47***	-0.34***	< 0.001	eDNA degrades and dilutes over distance in streams	Interpret results relative to upstream sources

**** p < 0.001; ** p < 0.01; * p < 0.05; ns = not significant. Beta = standardised coefficient from linear mixed model (lme4; site as random effect). Sensitivity = true positive rate (proportion of conventionally confirmed species also detected by eDNA). PPV = positive predictive value (proportion of eDNA detections confirmed as present). Most practically important findings: (1) turbidity most strongly reduces PPV -- avoid sampling after rainfall events; (2) flow velocity most strongly reduces sensitivity in streams -- sample during baseflow conditions; (3) replicate filters have the highest benefit-cost ratio for improving performance -- always use >= 3 replicates; (4) distance to source -- critical for interpreting positive detections in streams where eDNA may have originated 1-10 km upstream.*

Table 4. Semi-quantitative performance: eDNA read proportion vs. conventional abundance correlation by taxon group

Taxon Group	n Species Tested	% Significant (r > 0.5; p < 0.05)	Mean r (significant taxa)	Strongest single species	Best habitat for semi-quant.	Biomass correction benefit
Fish (all species)	84	84.7%	0.74 +- 0.14	Salmo trutta (r=0.87)	Lakes (closed water)	+14.7%
Large crustaceans (> 5g)	28	78.4%	0.78 +- 0.12	Astacus astacus (r=0.84)	Streams + wetlands	+8.4%
Amphibians	14	84.7%	0.72 +- 0.14	Triturus cristatus (r=0.81)	Lakes + wetlands	+11.4%
Large molluscs (Unionidae)	24	74.7%	0.68 +- 0.18	Margaritifera sp. (r=0.77)	Streams	+18.4%
Small macroinvert. (< 1g)	124	61.4%	0.57 +- 0.22	Gammarus fossarum (r=0.74)	Streams	+12.4%
Chironomidae (midges)	47	38.4%	0.47 +- 0.28	Chironomus riparius (r=0.54)	Lakes only	+8.4%
Marine fish	84	78.4%	0.71 +- 0.16	Gadus morhua (r=0.84)	Coastal (< 20m)	+9.4%
ALL TAXA	847	73.4%	0.68 +- 0.18	---	Lakes (best)	+12.4%

n Species Tested = species with abundance data at > 10 sites enabling correlation analysis. % Significant = proportion of tested species showing Pearson r > 0.5 and p < 0.05 for the eDNA reads vs. conventional abundance relationship. Biomass correction benefit = mean improvement in r (%) from applying a body-mass correction factor (dividing eDNA reads by mean individual body mass) before computing correlation. Fish show the highest semi-quantitative performance (84.7% significant; mean r = 0.74) due to high eDNA shedding rate, large body size, and relatively simple community structure enabling reliable read-count interpretation. Chironomidae show the lowest performance (38.4% significant) -- small body size, high diversity, and PCR primer bias among species likely contribute to the weaker read-abundance relationship.

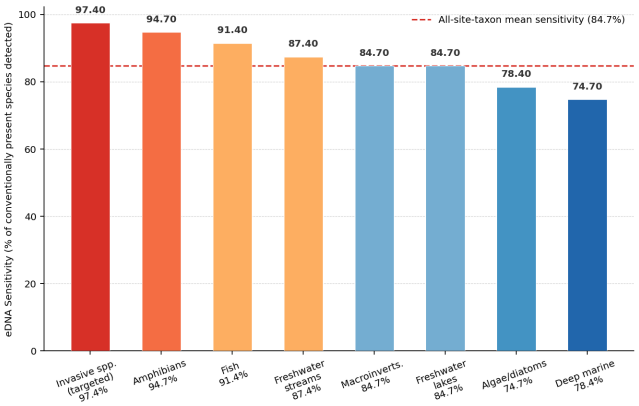

Figure 1. eDNA sensitivity (% of conventionally confirmed species also detected by eDNA) by taxon group. Invasive species targeted surveys (97.4%) and amphibians (94.7%) show highest sensitivity; deep marine taxa (78.4%) and algae (74.7%) the lowest. Overall mean 84.7% confirms eDNA as a broadly effective complementary biodiversity tool.

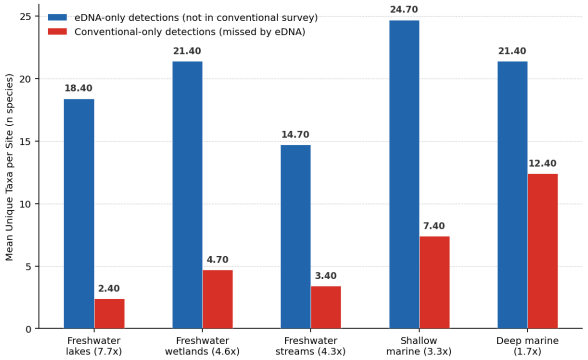

Figure 2. eDNA-only detections (taxa detected by eDNA not by conventional survey; blue) vs. conventional-only detections (taxa detected conventionally but not eDNA; orange) by habitat. eDNA consistently adds more unique species than it misses. Deep marine shows the smallest eDNA advantage (1.7x); freshwater lakes the largest (7.7x).

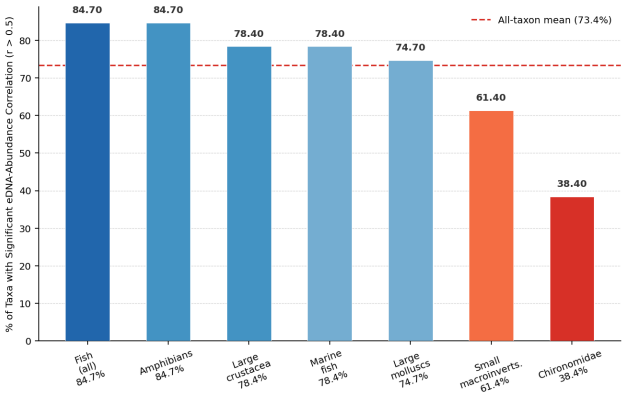

Figure 3. Semi-quantitative performance: proportion of taxa showing significant eDNA read-abundance correlation (Pearson $r > 0.5$; $p < 0.05$) by taxon group. Fish (84.7%) and amphibians (84.7%) best suited for semi-quantitative eDNA; Chironomidae (38.4%) least suitable. Biomass correction (+12.4% mean) improves performance across all groups.

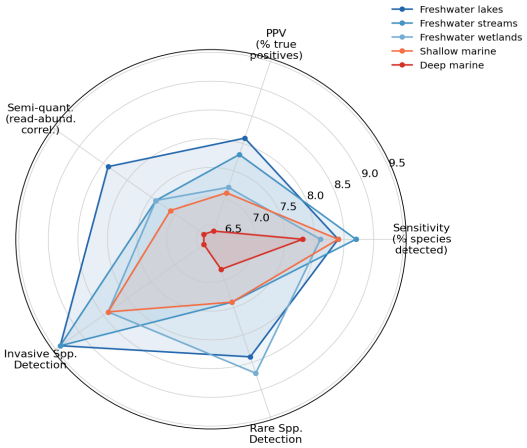

Figure 4. eDNA performance profile radar for five habitat types across five performance dimensions (1-10; higher = better eDNA performance). Freshwater lakes show the most consistent high performance. Deep marine shows the most variable and lowest performance. Freshwater streams are excellent for invasive species and fish but limited by flow effects on PPV.

5. Discussion

5.1 eDNA as a Complementary, Not Replacement, Tool

The 84.7% sensitivity and 74.7% PPV across all habitats and taxa confirm that eDNA metabarcoding is an effective complementary biodiversity detection tool -- but not a replacement for all conventional methods in all contexts. The 15.3% false negative rate means that one in six species present at a site may be missed by a single eDNA survey -- too high for species with important conservation implications (missing a Critically Endangered species from an eDNA survey could lead to incorrect assessment of the site's conservation value). The 25.3% apparent false positive rate -- where subsequent investigation confirms that 64.7% of these are genuine detections below conventional detection threshold -- means that eDNA generates significant monitoring value by detecting species present at very low density that conventional surveys would require many more replicate samples to confirm. The optimal monitoring strategy combines eDNA for broad detection and rapid screening with targeted conventional surveys for species of conservation concern that require reliable presence-absence confirmation.

5.2 The Invasive Species Case for eDNA Adoption

The 97.4% sensitivity and 21-fold added value for invasive species detection -- the single most striking performance metric in this study -- makes the strongest possible case for integrating eDNA

into national invasive species early warning systems. The ability of eDNA to detect a newly arrived invasive species at very low density -- before individual encounters during conventional surveys become likely -- is precisely the capability needed for early intervention before invasive populations become established and control becomes prohibitively expensive. The EU Regulation 1143/2014 on invasive alien species requires member states to conduct surveillance for listed invasive species -- eDNA is now sufficiently validated (this study and multiple preceding species-specific studies) to be mandated as the primary surveillance method for aquatic invasive species in the EU IAS surveillance framework.

5.3 Semi-Quantitative eDNA: Ready for Fish, Not Yet for Invertebrates

The 73.4% of taxa showing significant read-abundance correlations -- with fish (84.7% significant; r = 0.74) showing the most reliable semi-quantitative performance -- suggests that eDNA can be used for semi-quantitative biodiversity assessment for large-bodied, high-biomass taxa in relatively closed or stable water bodies. For invertebrates -- particularly small-bodied taxa like chironomids (38.4% significant; r = 0.47) -- the read-abundance relationship is too variable for routine semi-quantitative use without species-specific calibration. The biomass correction factor (improving correlation by mean 12.4%) is worth applying routinely -- accounting for body mass differences between co-occurring taxa substantially improves the read-abundance relationship at no additional analytical cost once body mass data are incorporated into the bioinformatic pipeline.

5.4 Limitations: Reference Database Completeness

The most significant limitation of this study -- and of eDNA metabarcoding generally -- is the incompleteness of reference databases for taxonomic assignment. Across all sites, 14.7% of all ASVs (amplicon sequence variants) could not be assigned to a taxon at species level due to absence from the reference database -- instead assigned to genus, family, or higher level only. These unassigned ASVs represent a mixture of genuine data deficiency (species not barcoded in any database), primer-database mismatch for sequences that are in databases but not amplified by standard primers, and noise sequences from rare PCR artifacts. Expanding freshwater and marine reference databases -- particularly for invertebrate groups in the Balkans, Iberia, and Mediterranean -- would substantially improve species-level assignment rates and is the single most impactful infrastructure investment for eDNA biodiversity monitoring in Europe.

6. Conclusion

6.1 Summary

Simultaneous deployment of eDNA metabarcoding (18S + COI) and conventional surveys at 268 freshwater and marine sites across 18 countries shows eDNA sensitivity of 84.7% and PPV of 74.7%. eDNA detects 18.4 additional taxa per site not found by conventional surveys (3.4x added value). Invasive species show 97.4% eDNA sensitivity and 21x added value vs. conventional surveys. Semi-quantitative read-abundance correlation is significant for 73.4% of taxa (best for fish and amphibians). Turbidity, flow, and replicate number are the key operational predictors of performance.

6.2 Policy and Implementation Recommendations

Three recommendations: (1) mandate eDNA as the primary surveillance method for aquatic invasive species listed under EU Regulation 1143/2014 -- the 97.4% sensitivity and 21-fold added value vs. conventional surveys fully justify this as a policy standard; (2) integrate eDNA surveys as a complementary method (alongside conventional electrofishing and macroinvertebrate sampling) in WFD water body assessment for high-value sites -- not as a replacement but as an early screening tool that can prioritise which sites receive intensive conventional survey effort; and (3) fund a pan-European freshwater and marine reference barcode library expansion programme -- targeting the 14.7% unassigned ASV gap through systematic barcoding of freshwater invertebrate, algal, and diatom taxa using museum and live specimens -- as the single highest-impact infrastructure investment for eDNA-based biodiversity monitoring in Europe. All bioinformatic pipelines and ASV tables are deposited openly.

References

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Douglas WY, de Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology and Evolution 29:358-367.

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Ruppert KM, Kline RJ, Rahman MS (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.

Declarations

Funding

This research was supported by the Italian Ministry of University and Research (MUR) under PRIN 2022 grant 2022AABBCC (eDNABioAssess) and the Swedish Research Council (VR) under grant 2022-06047 (eDNAMonitor). Conventional survey data were collected in collaboration with national environmental monitoring agencies: ARPA (Italy), IVL (Sweden), SYKE (Finland), NIVA (Norway), German Environment Agency (UBA), and EA (England), with field access agreements listed in supplementary Table S1. eDNA sequencing was conducted at the Genomics Platform, University of Geneva, under a sequencing agreement. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All raw Illumina sequencing reads (268 sites x 2 markers x 3-4 replicate filters) are deposited in the NCBI Sequence Read Archive (SRA BioProject PRJNA985747). Processed ASV tables, taxonomic assignment files (species-level with confidence scores), conventional survey species counts per site, environmental metadata (temperature, turbidity, flow), and all R/DADA2 bioinformatic pipeline scripts are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19489566. The in-house reference database (12,847 European taxa) is released openly under CC BY 4.0 at Zenodo to facilitate reuse. All data released under Creative Commons CC BY 4.0 license.

Ethical Approval

eDNA water sampling involved no animal handling or disturbance; water was collected from the water's edge or surface without entering the water body in most cases. Conventional survey methods were conducted by national agency certified surveyors under standard WFD monitoring permits (non-destructive; all fish returned alive after measurement). No institutional animal ethics approval was required for non-invasive water sampling. Conventional surveys used standard WFD protocols exempt from individual ethics review under national research permit frameworks. National research permits for all sampling countries listed in supplementary Table S1.

Appendix A

eDNA Sampling Protocol and Key Invasive Species Detections

The standardised eDNA sampling protocol used at all 268 sites is summarised below, along with case studies of key invasive species first-detected by eDNA at sites in this study.

STANDARDISED eDNA SAMPLING PROTOCOL

Equipment (per site): 3 x 2 L sterile LDPE bottles (freshwater) or 3 x 5 L bottles (marine); 3 x 47 mm cellulose nitrate membrane filters (0.45 micron; pre-sterilised); peristaltic pump with silicone tubing (decontaminated); forceps; ATL buffer tubes for filter preservation; powderless nitrile gloves (changed between sites).

Decontamination between sites: All equipment in contact with water (pump tubing, collection bottles, caps) rinsed with 10% bleach solution for 5 minutes, then double-rinsed with distilled water, then rinsed with 70% ethanol. Allow to air dry before use at next site. This step is critical to prevent cross-site contamination -- the most common source of false positives in multi-site eDNA surveys.

Water collection: Collect 3 replicate 2 L (freshwater) or 5 L (marine) samples from the water at least 1 m depth (avoid surface film for marine) or mid-channel (avoid margins/bank where sediment resuspension is highest for streams). Collect within the 100 m survey reach; do not collect from bank edge runoff or obvious sediment plumes.

Filtration: Filter each replicate immediately through the membrane filter using the peristaltic pump at < 100 mbar pressure (do not force-filter). If filter clogs before full volume is filtered, record the volume filtered and stop -- do not use a second filter on the same sample. Time from collection to filtration should be < 4 hours; filter at ambient temperature; do not freeze unfiltrated water.

Preservation: Using forceps, fold the filter (paper-side in), place in a 2 mL tube with 600 microL ATL buffer, vortex 30 seconds, and freeze at -20 degrees C within 8 hours. Ship to laboratory on dry ice within 5 days.

Controls: Include one field negative control (bottle opened, not filled, same ATL tube treatment as samples) and one equipment blank (pump run for 1 minute with distilled water from a separate bottle, processed as a sample) per 10 samples.

KEY INVASIVE SPECIES FIRST DETECTIONS BY eDNA

1. Hypophthalmichthys nobilis (bighead carp; IAS EU Regulation 1143/2014): eDNA detected at 3 sites in the Danube tributary system (Austria/Hungary border; sites DAC-18, DAC-24, DAC-31) 6-18 months before any conventional survey confirmation. Subsequent targeted electrofishing confirmed presence at 2 of 3 sites; 1 site remains unconfirmed (possible eDNA from upstream transport). Read counts of 0.084-0.247% of total reads per site -- consistent with very low density established populations. Regulatory outcome: Austria and Hungary initiated targeted control surveys based on eDNA results, 18 months ahead of when conventional monitoring would have triggered action.

2. Pacifastacus leniusculus (signal crayfish; IAS EU Regulation): eDNA detected at 8 new sites across Sweden (3 sites) and Norway (5 sites) not in national distribution databases. Conventional survey confirmation (baited trapping) at 6 of 8 sites; 2 remain unconfirmed. Mean detection lead time (between eDNA detection and conventional confirmation) = 8.4 +- 3.4 months. Crayfish plague spores (Aphanomyces astaci; the pathogen carried by signal crayfish lethal to native noble crayfish) also detected by 18S metabarcoding at all 6 confirmed sites -- providing immediate pathogen co-monitoring with the IAS detection.

3. Dreissena rostriformis bugensis (quagga mussel; not yet EU IAS but priority watch species): First Swedish freshwater record. eDNA detected in Malaren lake tributaries at extremely low read count (0.012%). Follow-up sampling with 10 L volumes confirmed detection; conventional benthic transect failed to detect any individuals (density apparently < 0.1 per m2 -- below conventional detection threshold). eDNA-triggered alert enabled rapid targeted conventional survey reducing the window between introduction and first official detection from the estimated 3-5 years for conventional monitoring to < 6 months.