https://doi.org/10.64142/jebs.2.1.20

Journal of Environment and Biological Science

Mini-Review

JEBS-26-20

From DNA Detection to Reliable Biodiversity Assessment: Methodological Priorities for Environmental DNA Metabarcoding in the African Great Lakes Region

Jean de Dieu Habiyakare*, Doris Niyonsaba, Valentin Munezero and Olivier I Mudahunga

Department of Biology, University of Rwanda, Kigali, Rwanda

***Corresponding author:** Jean de Dieu Habiyakare, Department of Biology, University of Rwanda, Kigali, Rwanda, E-mail: jeandedieuhabiyakare4@gmail.com

Received date: September 17, 2026; **Accepted date:** October 16, 2026; **Published date:** November 17, 2026

Citation: Habiyakare JD, Niyonsaba D, Munezero V, Mudahunga OI (2026) From DNA Detection to Reliable Biodiversity Assessment: Methodological Priorities for Environmental DNA Metabarcoding in the African Great Lakes Region. J Environ Biol Sci. Vol.2 No.1: 20.

Abstract

Environmental DNA (eDNA) metabarcoding can broaden freshwater biodiversity assessment by detecting genetic material released by organisms into water. Its value for conservation, however, depends on how confidently sequence detections can be linked to taxa and interpreted in space and time. This narrative mini-review examines methodological priorities for tropical African freshwater monitoring, using the Lake Albert study in the Albertine Rift as a geographically direct example and the Rufiji River catchment in Tanzania as a comparative East African river-system case. A targeted, non-systematic literature search was used to identify these focal studies and methodological literature on reference libraries, marker resolution, eDNA transport and persistence, replication, contamination control and taxonomic assignment. The review distinguishes two related but different identification problems: missing or unreliable reference sequences and markers that cannot separate closely related species even when reference sequences are available. It also explains why DNA transported downstream or persisting after release can indicate genetic material without proving that a living population occupies the precise sampling point. Practical recommendations are presented for interpreting negative controls, replicate agreement and assignment confidence without imposing universal thresholds that are not supported across assays and ecosystems. The main contribution is a focused inference framework for moving from sequence detection to defensible biodiversity claims in the African Great Lakes region. Priorities for Rwanda and neighbouring countries include regionally curated reference libraries, spatially and temporally appropriate sampling, transparent quality assurance, conservative taxonomic reporting and validation against complementary surveys.

Keywords: Environmental DNA; Metabarcoding; Freshwater biodiversity; African Great Lakes; Reference libraries; Taxonomic resolution; Quality assurance; Rwanda

Introduction

Freshwater ecosystems support biodiversity, fisheries, food security, water supply and livelihoods, yet their biological communities are often incompletely documented. In tropical Africa, large lake systems and river catchments can be difficult to survey comprehensively because sites are extensive or remote, some organisms are rare or elusive and identification may require specialist taxonomic expertise. These constraints limit the spatial and temporal coverage of conventional biodiversity surveys, although field observations, specimen collection and morphology-based identification remain essential sources of evidence.

Environmental DNA (eDNA) metabarcoding offers a complementary approach. Organisms release DNA into their surroundings through cells, mucus, waste, gametes and tissue fragments. Researchers can collect environmental samples, extract DNA, amplify one or more marker regions, sequence the resulting amplicons and compare sequences with reference data to infer which taxa may have contributed DNA to the sample. This approach can reveal community patterns across multiple taxa, but every stage from sampling and preservation to primer choice, sequencing, bioinformatics and taxonomic assignment can influence the resulting species list [1,2].

Copyright © 2026| https://biology.academicsquare-pub.com/

https://doi.org/10.64142/jebs.2.1.20
Journal of Environment and Biological Science
Mini-Review

This review uses the term "African Great Lakes region" operationally to refer to the Great Lakes of the East and Central African Rift system, particularly Lakes Albert, Edward, Kivu, Victoria, Tanganyika and Malawi/Nyasa, together with relevant adjoining freshwater habitats and tributaries. The exact list of lakes included in regional definitions can vary; therefore, the geographic scope should be stated explicitly in any study. Lake Albert is a direct example within this operational scope. By contrast, Tanzania's Rufiji River catchment drains eastwards to the Indian Ocean and is not part of the African Great Lakes drainage system. It is considered here as comparative evidence from a nearby tropical East African river catchment, not as direct evidence from the Great Lakes basin.

Two recent studies provide useful but distinct examples. Bálint et al. used eDNA barcoding to examine eukaryotic and metazoan communities in Lake Albert and reported spatial differences in community composition alongside limitations associated with incomplete reference sequences [3]. Shechonge et al. applied fish eDNA metabarcoding across 49 sites and 174 samples in the Rufiji catchment, detecting 66 fish species against an estimated 91 species considered confidently present in the system [3,4]. Their work demonstrates the value of spatially distributed sampling and locally developed reference sequences, while also illustrating that detection and identification performance depend on assay and database characteristics.

A systematic review by Ibrahim et al. synthesizes eDNA applications for biodiversity monitoring across sub-Saharan Africa, including the geographic distribution of studies, methods used, reported limitations and knowledge gaps [5]. The present mini-review does not attempt to repeat that broad regional inventory or claim to be a comprehensive review of all African eDNA research. Instead, it develops a narrower methodological contribution: An inference-focused framework for deciding what a metabarcoding result can support, especially when reference coverage is incomplete, markers have limited discriminatory power, DNA may have moved from its source and quality-control evidence varies. The aim is to connect these issues to practical decisions relevant to freshwater monitoring in the African Great Lakes region and to identify priorities for Rwanda and neighbouring countries.

Literature-Selection Approach and Scope

This article is a narrative mini-review based on a targeted, non-systematic search rather than a systematic review. Searches were conducted using combinations of the terms "environmental DNA", "eDNA metabarcoding", "freshwater biodiversity", "African Great Lakes", "Lake Albert", "Rufiji River", "reference library", "taxonomic assignment", "DNA transport", "DNA persistence", "negative controls" and "replication". Records were located through PubMed-indexed records, publisher platforms and searches for cited methodological literature; reference lists of focal papers were also checked for relevant methods studies.

Studies were prioritised when they (i) addressed freshwater biodiversity assessment using eDNA or metabarcoding; (ii) provided a geographically relevant example from Lake Albert or a comparative tropical East African river system or (iii) offered methodological evidence on reference databases, marker discrimination, transport and persistence, contamination control, replication or taxonomic assignment. The Lake Albert and Rufiji studies were selected as illustrative case studies, not as a representative sample of all regional research. No exhaustive screening, formal risk-of-bias assessment or PRISMA-based selection process was undertaken and no claim is made that all eligible studies were identified. Consequently, this synthesis is interpretive and its recommendations should be read alongside the primary studies and wider methodological literature.

Environmental DNA Metabarcoding in Freshwater Biodiversity Assessment

From environmental samples to biodiversity information

An eDNA metabarcoding workflow typically includes site selection, sample collection and preservation, DNA extraction, amplification of selected genetic markers, sequencing, bioinformatic filtering and taxonomic assignment. Decisions at each stage affect which taxa are detected and the confidence attached to those detections. Sampling may fail to capture DNA from a taxon that is present; primers may amplify some taxa more efficiently than others; inhibitors or degradation may reduce amplification and sequence-processing choices can alter the final set of retained reads [1,2].

Accordingly, non-detection is not equivalent to biological absence. A taxon may be present but undetected because of low DNA shedding, patchy distribution, insufficient sample volume, unsuitable timing, primer mismatch, inhibition or inadequate sequencing depth. Conversely, a sequence detection is not automatically proof that a living organism occupies the precise point where the sample was collected. Reliable interpretation requires an explicit link between the study question and the evidence generated by the sampling and analytical design.

Copyright © 2026| https://biology.academicsquare-pub.com/

https://doi.org/10.64142/jebs.2.1.20
Journal of Environment and Biological Science
Mini-Review

Evidence from Lake Albert and the Rufiji River catchment

Bálint et al. applied eDNA barcoding to Lake Albert, sampling surface water at three locations and investigating eukaryotic and metazoan communities. They reported spatial differences in community composition and noted that incomplete reference-sequence resources constrained taxonomic assignment [3]. This study is directly relevant to the Great Lakes region and illustrates both the potential of eDNA for describing broad community patterns and the continuing importance of regional taxonomic resources.

Shechonge et al. examined fish diversity across the Rufiji River catchment in Tanzania [4]. Their study analysed 174 samples from 49 sites and detected 66 fish species relative to an estimated 91 species considered confidently present in the system. The spatially distributed design supported investigation of community structure in relation to environmental gradients, while development of a regional fish reference library strengthened assignments. Because the Rufiji drains to the Indian Ocean rather than into the African Great Lakes, its results should be used comparatively: They demonstrate methodological lessons for tropical East African rivers but do not establish the same detection performance or species patterns in Great Lakes ecosystems.

The two studies should not be compared as if they were equivalent biodiversity inventories. They differ in ecosystem type, spatial coverage, target groups, markers, sampling design and analytical procedures. Their shared relevance lies in showing that ecological interpretation depends on sampling design and reference information, while the specific findings remain tied to their respective study systems.

Methodological Priorities for Reliable Inference

Reference-library completeness and marker resolution are different problems

Taxonomic assignment depends on both the availability and the discriminatory power of reference sequences. These are related but distinct limitations. Reference-library incompleteness occurs when a relevant species is absent from the database, represented by an incorrectly identified sequence or represented by a sequence that does not cover the marker region amplified in the environmental sample. In such cases, a true sequence may remain unassigned or may be matched to a related taxon simply because the correct reference is missing. This issue is especially important where regional freshwater taxa are poorly represented in public databases, as highlighted by the Lake Albert and Rufiji studies [3,4].

Insufficient marker resolution is different: Even when reference sequences for the relevant species are available, the chosen marker may not contain enough variation to distinguish closely related species. Congeneric species may share identical or near-identical sequences at a short marker region; hybridisation, introgression, incomplete lineage sorting or low evolutionary divergence can further complicate species-level assignment. A larger database cannot solve a marker that does not distinguish the taxa in question. Marker choice should therefore be informed by the target group, expected taxonomic diversity, in-silico comparisons and where possible validation against known specimens [2,4].

Taxonomic reporting should be conservative and hierarchical. A species-level name is justified only when the marker is demonstrably discriminative for the plausible regional taxa, the reference sequence is trustworthy and relevant and the assignment method supports that level of certainty. If multiple species share the best-matching barcode or competing assignments cannot be resolved, the result should remain at genus, family or another defensible higher rank. A high percentage sequence identity alone is not sufficient evidence of species identity if closely related taxa share that sequence or if the reference annotation is uncertain. Reports should describe the database version, assignment algorithm, thresholds or decision rules, treatment of ambiguous matches and evidence used to validate taxonomic names.

A regional reference-library programme should prioritise locally occurring freshwater taxa and use voucher specimens or other traceable evidence of identification. Sequences should include specimen provenance, collection location, identifier, marker region and links to museum or institutional records where available. Independent taxonomic review and transparent curation are important because adding misidentified reference sequences can increase, rather than reduce, false certainty.

Sampling design and spatial representation

Detection probability varies across sites, habitats, seasons and environmental conditions. Water flow, depth, turbidity, temperature, rainfall, organism activity and DNA degradation can affect the amount and distribution of detectable material. A single sample is unlikely to represent all habitats or taxa within a large lake or river. Spatial replication should reflect the ecological question, for example, shoreline versus offshore habitats, tributaries versus main channels or sites above and below a potential

Copyright © 2026| https://biology.academicsquare-pub.com/

disturbance. Temporal replication may be necessary where seasonal flow, breeding, migration or water conditions influence detectability.

Replication should be designed at the appropriate level. Independent field samples capture spatial or temporal heterogeneity; repeated PCRs from the same extract primarily help assess stochastic amplification and laboratory repeatability. Technical replicates cannot substitute for field replication. Pilot sampling can help estimate how detection changes with sample number, water volume, filter type and spatial separation. The Rufiji study illustrates the information gained from broad spatial sampling, while the Lake Albert study shows that community composition can differ among sampling locations [3,4].

To support comparison among studies, authors should report coordinates, dates, habitat and environmental conditions, water volume, filtration and preservation procedures, time between collection and preservation, field decontamination, sample storage, extraction methods, primer sequences, PCR conditions, sequencing platform and bioinformatics workflow. Protocols should be adapted to local field conditions but described sufficiently to permit replication.

DNA transport and persistence: Detection is not proof of local occupancy

Environmental DNA is transported and transformed after release. In flowing waters, DNA can be carried downstream, diluted, deposited in sediments, resuspended or degraded. Its transport distance and persistence depend on flow, turbulence, particle association, temperature, ultraviolet exposure, microbial activity, water chemistry and the physical characteristics of the DNA-containing material [6,7]. A positive water sample can therefore reflect an organism located upstream or DNA released earlier, rather than a living organism at the exact sampling point.

This limitation is particularly important when interpreting river samples, where the sampling point integrates signals from parts of the upstream catchment. In lakes, water movement, shoreline inputs, depth and mixing can also separate the location of DNA detection from the source organism.

Persistence does not make eDNA uninformative; it changes the spatial and temporal meaning of the evidence. An inference of local occupancy is stronger when detections are repeated across independent samples, spatial patterns are consistent with the species' ecology, hydrological context is considered and independent survey evidence supports the interpretation.

Authors should therefore distinguish among (i) detection of DNA assigned to a taxon; (ii) evidence consistent with the taxon occurring within the sampled waterbody or connected catchment and (iii) confirmed local occupancy by living organisms. The third claim generally requires additional evidence or a study design specifically validated for that inference. eDNA read counts should not be treated as direct estimates of organism abundance without calibration, because DNA shedding, transport, degradation, primer bias and amplification efficiency can all affect read numbers.

Quality assurance: Controls, replicates and assignment confidence

Quality assurance should be planned before sampling and reported in a way that allows readers to evaluate the reliability of detections. At minimum, studies should describe field and laboratory negative controls, positive controls or other checks that the assay worked, procedures to prevent cross-contamination, replicate structure, sequence filtering and taxonomic-assignment rules. Negative controls should pass through the relevant stages of collection, extraction, PCR and library preparation so that contamination can be detected and, where possible, traced to its source [1].

Negative-control results require context-dependent interpretation. If a target taxon is detected in a field blank, extraction blank or PCR negative control, detections of that taxon in associated samples should be treated as potentially contaminated and investigated. The response should consider which controls were positive, whether the same taxon or sequence variant occurred in samples, the read abundance and distribution across the run, the likely contamination pathway and whether independent evidence supports the detection. There is no universally appropriate read-count cut-off that can automatically distinguish contamination from true occurrence across all markers, sequencing runs and study designs. The decision rule and any exclusion criteria should be specified before interpreting ecological patterns and reported transparently.

Agreement among replicates adds evidence, but replicates are not interchangeable. A taxon recovered from multiple independent field samples and supported by consistent technical replicates, generally provides stronger evidence than a taxon found once in one PCR. Nevertheless, low-template DNA can be stochastic, so failure to reproduce a rare detection does not automatically prove it false. The strength of evidence should be evaluated in relation to sampling effort, expected detection probability, replicate independence, contamination evidence and the consequences of false-positive or false-negative conclusions [1,2]. Studies should report how many field samples and PCR

Copyright © 2026| https://biology.academicsquare-pub.com/

replicates were positive, rather than presenting only a pooled species list.

Taxonomic-assignment confidence should also inform whether a detection is accepted at species level. Researchers should consider marker discrimination among plausible local species, the quality and geographic relevance of reference sequences, similarity and alignment coverage, consistency across assignment methods or phylogenetic analyses where appropriate and the possibility of ambiguous or conflicting matches. If these criteria do not support species-level resolution, the detection should be reported at the lowest defensible taxonomic rank, with uncertainty clearly stated. Positive controls can demonstrate that amplification and sequencing worked, but they do not by themselves validate every sample-level detection or taxonomic assignment.

An operational reporting framework can classify detections as: (1) Supported at the reported rank, where controls are clean or any control issue has been resolved, the sequence passes stated filters and the assignment is taxonomically defensible; (2) provisional, where a sequence is plausible but replicate support, reference coverage or taxonomic resolution is limited or (3) not accepted for ecological inference, where contamination or analytical artefacts cannot be ruled out. These are proposed reporting categories, not universal regulatory thresholds. Their use should be adapted to assay validation, study objectives and the relative costs of false-positive and false-negative decisions.

Integrating eDNA with conventional biodiversity surveys

eDNA metabarcoding is most informative when combined with complementary methods where feasible. Fish surveys, visual observations, specimen-based identification, museum records and local ecological knowledge can help validate taxonomic assignments and interpret unexpected detections. Conversely, eDNA may reveal taxa missed by conventional surveys or support broader spatial screening. The methods have different biases and should not be expected to produce identical species lists. Validation should be designed around the question being asked. For example, if the objective is to assess species presence, independent observations or targeted sampling may help confirm high-priority detections. If the objective is to compare community composition, sampling effort and taxonomic coverage should be considered across both methods. Disagreement between methods is itself informative and should prompt investigation of detection probability, marker bias, sampling scale and taxonomic uncertainty rather than automatic acceptance of one method as the reference standard.

Priorities for Rwanda and regional collaboration

For Rwanda, a practical starting point is to define priority freshwater questions and pilot eDNA protocols in selected habitats, including Lake Kivu and relevant rivers or wetlands. The design should identify target taxa and markers, map available regional reference sequences, establish contamination-control procedures and determine how many sites, field replicates and temporal visits are needed for the intended inference. The recommendations in this review are proposed priorities, not evidence that a comprehensive eDNA monitoring programme has already been assessed or implemented in Rwanda.

Regional collaboration could reduce duplication and improve reliability through shared reference specimens, taxonomic expertise, harmonised metadata, training, laboratory quality procedures and reproducible bioinformatics workflows. Reference libraries and sequence datasets should be curated with transparent provenance and deposited in accessible repositories where permissions allow.

Partnerships among universities, biodiversity institutions, conservation organisations and public agencies could also support local capacity for sample collection, sequencing access, data management and long-term monitoring.

Contribution, Limitations and Future Directions

The contribution of this mini-review is not a new inventory of all eDNA applications in sub-Saharan Africa. Ibrahim et al. already provides a broader systematic synthesis of regional study distribution, methods, limitations and gaps [5].

The present article instead focuses on the chain of inference from a sequence read to a defensible biodiversity claim in tropical freshwater systems. It develops four linked distinctions with practical implications: missing reference data versus inadequate marker resolution; DNA detection versus local organism occupancy; technical replication versus independent field replication and a sequence match versus a taxonomically defensible assignment. It then connects these distinctions to decision-making around controls, uncertainty and complementary evidence. This article has several limitations.

Copyright © 2026| https://biology.academicsquare-pub.com/

The literature search was targeted and non-systematic and the two focal case studies were chosen for illustrative relevance rather than to represent all research in the region. The review does not quantify the prevalence of methodological problems, compare study performance statistically or establish universal thresholds for control results, replicate agreement or taxonomic assignment. The geographic scope is operational and should not be interpreted as a claim that all countries or lakes associated with the African Great Lakes have equal evidence coverage. A future systematic review could use predefined eligibility criteria and reproducible database searches to map study coverage, markers, sampling designs, reference-library availability and validation practices across the Great Lakes basin.

Conclusion

Environmental DNA metabarcoding can strengthen freshwater biodiversity assessment in tropical Africa, but its reliability depends on more than sequencing. Evidence from Lake Albert and the Rufiji River catchment illustrates the value of molecular approaches while highlighting the importance of reference information and spatially appropriate sampling. The Rufiji study provides comparative methodological evidence from a tropical East African river system, not direct evidence from the African Great Lakes drainage basin.

Reliable interpretation requires researchers to distinguish reference-library gaps from marker limitations, account for DNA transport and persistence, separate technical from field replication, evaluate negative controls in context and report taxa only at the level supported by the evidence. For Rwanda and neighbouring countries, priorities include regionally curated reference libraries, transparent and reproducible quality assurance, pilot-informed sampling designs and integration with conventional biodiversity surveys.

These steps can help turn molecular detections into more defensible biodiversity assessments without overstating what a positive sequence result alone can demonstrate.

References

1. Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, et al. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol 7: 1299-1307. [Crossref] [Google Scholar]
2. Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods Ecol Evol 9: 134-147. [Crossref] [Google Scholar]
3. Bálint M, Tumusiime J, Nakintu J, Baranski D, Schardt L, et al. (2024). Environmental DNA barcoding reveals general biodiversity patterns in the large tropical rift Lake Albert. Sci Total Environ 957: 177308. [Crossref] [Google Scholar] [Indexed]
4. Shechonge AH, Collins RA, Ward S, Saxon AD, Smith AM, et al. (2024) Environmental DNA metabarcoding details the spatial structure of a diverse tropical fish assemblage in a major East African river system. Environ DNA 6: e70008. [Crossref] [Google Scholar]
5. Ibrahim JJ, Mkindi A, Mbega E, Baino A, Ndakidemi B (2026) Environmental DNA (eDNA) applications for biodiversity monitoring across countries in the Sub-Saharan African region: A systematic review. Front Conserv Sci 7: 1941471. [Crossref] [Google Scholar]
6. Deiner K, Fronhofer EA, Mächler E, Walser JC, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nat Commun 7: 12544. [Crossref] [Google Scholar] [Indexed]
7. Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc Biol Sci 286: 20191409. [Crossref] [Google Scholar] [Indexed]

Copyright © 2026| https://biology.academicsquare-pub.com/