NATURE METRICS
DNA-BASED MONITORING

Phase 2 Main Report - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements

NatureMetrics
June 2023

Project Facilitation

Authors

Judith Bakker[1]
Hayley Craig[1]
Lynsey R. Harper[1]
Edward Wort[1]
Douglas W. Yu[1,2]
Owen Middleton[1]
Christian Devenish[1]
Hannah Flintham[1]
Nathan R. Geraldi[1]
Evie Gardner[1]
Mike Morris[1]
James Whiting[1]
Bastian Egeter[1]

[1] NatureMetrics, 1 Occam Court, Surrey Research Park, Guildford GU2 7HJ, United Kingdom
[2] University of East Anglia, Earlham Rd., Norwich, Norfolk NR18 9NS, United Kingdom

NatureMetrics Project Management:
Bastian Egeter (2021-2023)
Cath Tayleur (2021-2022)
Sam Lacey (2022)
Judith Bakker (2022-2023)

SG & SEPA Project Management:
Willie Duncan (2021-2022)
Pauline Lang (2022-2023)

Please reference this report as follows: Bakker J., Craig H., Harper L. R., Wort E., Yu D.W., Middleton O., Devenish C., Flintham H., Geraldi N.R., Gardner E., Morris M., Whiting J., Egeter B. (2023) Phase 2 Main Report - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements. NMP/001/20. NatureMetrics.

Dissemination status: Unrestricted

Acknowledgements: Funding for this work is provided by SG-RESAS, the Scottish Government's Rural and Environment Science and Analytical Services. The authors would like to thank the Phase 2 Project Management Steering Group (MSG): Pauline Lang (Contract Manager, Lead Project Partner), Alistair Duguid (Scottish Environment Protection Agency, SEPA), Colin Bean (NatureScot), Iveta Matejusova (The Marine Directorate of Scottish Government (formerly Marine Scotland Science, MSS)), and Helen Jones (the Scottish Government's Rural and Environment Science and Analytical Services division, SG-RESAS) for their substantial support and advice throughout the project. We also thank Colin Bean (NatureScot), Iveta Matejusova (The Marine Directorate (formerly Marine Scotland Science, MSS)), Pauline Lang, and Alistair Duguid (Scottish Environment Protection Agency, SEPA), as well as advisory board members Douglas Yu[§] (NatureMetrics), Pippa Howard[§] (NatureMetrics), Nadia Barsoum[§] (Forest Research) for providing project guidance, technical advice and peer reviewing this work.

We thank the Technical Reviewing Group (TRG) for providing timely and constructive feedback to inform Phase 2 project development and delivery: We also thank a broad range of project stakeholders including key representatives from the Phase 2 Project Technical Reviewing Group (TRG); the Project Advisory Board; the Scottish DNA Hub; Loch Lomond and the Trossachs National Park (LLTNP), Cairngorms National Park; and other organisations for engaging their expertise and support during project development and production of this deliverable output. We specifically thank Nadia Barsoum[§] (Forest Research); Kenny Kortland (Forestry and Land Scotland); Alan Bell, Dom Hall (LLTNP); Tom Butterworth[§] (WSP); Simon Franks (LLTNP Trees and Woodland advisor); Scot Mathieson (SEPA); Colin Adams (Scottish Centre for Ecology and the Natural Environment, SCENE, University of Glasgow), Douglas Yu[§] (University of East Anglia, NatureMetrics), Laia Rovira-Craven (SEPA), Nick Schurch (BioSS), Paul Woodcock (JNCC), Nigel Willby (University of Stirling), Bernd Haenfling (University of Highlands and Islands, UHI), Lauren Cook (Natural History Museum; CEFAS), Pete Hollingsworth (RBGE), Philip Boulcoutt (The Marine Directorate, formerly Marine Scotland Science, MSS), David Cooke (James Hutton Institute), Rob Ogden (University of Edinburgh) for providing their specialist expertise, and feeding ideas or suggestions into project development. We also thank SEPA lead and SG-CAMERAS Board partner Peter Singleton for project support.

We are extremely grateful to The Marine Directorate (formerly Marine Scotland Science, MSS), SEPA, and NatureScot for provisioning extensive in-kind support to both the Phase 1 pilot study and the Phase 2 main sampling campaign, which included facilitating staff time, their expertise, and operational resources needed to enable site access and field monitoring and/or laboratory analysis for the project to be mobilised across the four different habitat types, especially the following people and organisations: Iveta Matejusova (Marine habitat) from The Marine Directorate (formerly Marine Scotland Science, MSS); Peter Pollard, Kirsten Davidson, Lorraine Quinn, Ian Lorimer, Annette Ross, Ian Milne, and Pauline Lang (Marine, Freshwater, Peatland, and Woodland habitats) from SEPA; Colin Bean (Freshwater and Woodland habitats) from NatureScot. We are extremely grateful to all SEPA Marine Ecology team members who were involved in undertaking the sample analyses for marine benthic invertebrates and particle size analysis (PSA) from the marine habitat sampled during Phase 2 project work, with particular thanks to: Myles O'Reilly, Stephen Nowacki, Ryan Eustace, Emma Priestley, Calum Clark, Nick Woods, and Will Townshend. This in-kind support was needed to produce laboratory results for comparing outcomes of using conventional and eDNA methods in the marine environment. For the freshwater fieldwork, we thank Oliver Taylor and Lewis Campbell (NatureMetrics) for assistance with sample collection, Hannele Honkanen (SCENE, University of Glasgow) and David Scott-Park (Portnellan Farm) for permitting sampling at Loch Lomond shoreline sites. For the peatland fieldwork we thank Marco Fioratti (NatureMetrics) for assistance with sample collection, Richard Cooper (LLTNP Peatland ACTION) for help with identifying all the peatland sites and meeting us on site in 2021; Hamish Thomson (Woodlands Trust Glen Finglas), site contact and for providing transport for sampling in 2021, and Emily Gray for sampling permission; Emma Paterson, Nicola Colquhoun, and Ian Dingwall (Auchlyne Estate) for permitting sampling at the site and facilitating site access; Royal Scottish Forestry Society (owners of Cashel site) and Peter Phillips (Cashel site contact); Jane Lindsay and Claire Campbell (SEPA) for providing transport and expertise which enabled site access and the peatland condition assessments to be undertaken at Auchlyne and Glen Finglas in 2022; Alan Bell, Richard Cooper, Guy Cole, and Natasha Craven (LLTNP) providing transport and expertise which enabled site access and the peatland condition assessments to be undertaken at Cashel in 2022. For the woodland fieldwork we thank Marco Fioratti (NatureMetrics for assistance with sample collection, site contact Fraser Lamont (RSPB Inversnaid) for meeting us on site to show the experimental plots for woodland sampling in 2021; Colin Leslie, Juli Titherington, Emyr Algieri, Brian Duff (Forestry and Land Scotland) for assisting with site

information, site permissions, and/or identifying potential sampling locations; Alan McDonnell, and Marian Bruce (Trees for Life) for providing site information for the 2022 forest sites in Glen Affric and Glen Moriston; Piers Voysey and Peter Ferguson (Rothiemurchus Estate); Richard Lewis (Collie Coire Chulic site contact/owner); Alex Caraffi (Glen Falloch site contact); Emily Warner for providing advice and site info on Glen Affric and Glen Moriston sites; Kirsty North, Stan Phillips (Nature Scot) for advice on SSSI consent requirements; and Brodie Thomas (NatureScot) for field assistance during the woodland sampling in 2022. We are extremely grateful to Forest Research for providing in-kind support for planning and carrying out a significant part of the woodland work, including sharing data and metadata from a parallel Forest Research project, facilitated by Nadia Barsoum. We are grateful to Cathy Benett, Alison Bell, Annette Ross, and Andy Gowans for their expertise and help in facilitating SEPA data requests. We are also thankful to Tim Foster (SEPA) for their guidance and support with data management. We thank MSG and TSG members for peer review and constructive feedback to help inform and improve the final version of the Phase 2 deliverable outputs.

The authors are also very grateful to all participants of the project knowledge-exchange (KE) events, representing a broad range of individuals and organisations, in Scotland and beyond, for their interest in the project outcomes and using DNA-based approaches for biodiversity monitoring/reporting purposes, as well feedback provided and expert contributions to discussion on 3rd May 2023, 17th May 2023, and 7th June 2023. We specifically want to thank the 7 June 2023 KE event hosts at LLTNP, including Dom Hall, Simon Jones, and Gordon Watson together with their colleagues managing event coordination especially Jane Cook, Cara Thom, Lauren McInnes, and Rachael McLauchlan. Thanks to the KE event sponsor Pete Hollingsworth of RBGE on behalf of the Scottish DNA Hub. We also thank KE event coordinator Pauline Lang, co-chairs Alistair Duguid, Laia Rovira-Craven, Colin Bean, Iveta Matejusova, and convener Scot Mathieson for creating the conditions for stakeholder engagement and facilitating discussions to help the project outcomes. We also thank representatives from the following organisations for participating in the May and June 2023 KE events and for helping to improve project deliverables: Scottish Environment Protection Agency (SEPA), The Marine Directorate (formerly Marine Scotland Science, MSS), NatureScot, Scottish Government, Scottish Government's Rural and Environmental Science and Analytical Services (SG-RESAS) division, Loch Lomond and the Trossachs National Part (LLTNP) Authority, Cairngorms National Park (CNP) Authority, Royal Botanic Garden Edinburgh (RBGE), Moredun Research Institute, Science and Advice for Scottish Agriculture (SASA), Forestry and Land Scotland, Scottish Forestry, Forest Research, James Hutton Institute (JHI), Biomathematics and Statistics Scotland (BioSS), Joint Nature Conservation Committee (JNCC), Natural England, University of Stirling, University of the Highlands and Islands (UHI), Natural History Museum (NHM), Scottish Association for Marine Science (SAMS), Trees For Life, Scottish Badgers, Royal Zoological Society of Scotland (RZSS), Historic Environment Scotland (HES), University of Glasgow, Natural Resources Wales (NRW), University of Edinburgh, Loch Lomond Fisheries Trust (LLFT), Scotland's Rural College (SRUC), National Museums Scotland (NMS), Centre for Environment, Fisheries & Aquaculture Science (CEFAS), Department for Environment, Food & Rural Affairs (DEFRA), Shetland Oil Terminal Environmental Advisory Group (SOTEAG).

Finally, we thank the Scottish Government project funder (SG-RESAS Contract Research Fund) and the Co-ordinated Agenda for Marine, Environment and Rural Affairs Science (SG-CAMERAS) Board Partnership, in collaboration with the Scottish DNA Hub, for commissioning this project and their ongoing support.

[§]Project Advisory Board Member

Contents

Executive Summary

Biodiversity loss is widely recognised as one of the most urgent global challenges to be addressed in the next decade. The Scottish Biodiversity Strategy sets out a clear ambition to be Nature Positive by 2030, and to have restored and regenerated biodiversity across the country by 2045 (Scottish Government 2022). To protect, restore, and regenerate biodiversity, it is necessary to be able to accurately describe and quantify ecological change. Biodiversity monitoring through environmental DNA (eDNA) analysis is increasingly being used for tracking species diversity and composition in ecosystems as it is a scalable and high-resolution method. The overall goal of this project was to investigate and test the applicability of eDNA-based monitoring approaches for biodiversity assessment and reporting purposes across a broad range of habitat types in Scotland.

Samples were collected across four habitat types: marine lochs, freshwater lochs, woodland, and peatland. The survey sites were mostly situated in and around the focal study area of Loch Lomond and the Trossachs National Park (LLTNP) but eDNA sampling included other parts of Scotland such as the Cairngorms National Park. This was a Proof-of-Concept study across small numbers of sites and gradients of condition across Scotland. This work was undertaken to help establish scientific evidence, blended with practical learning-by-doing experience, and provide key recommendations, including future perspectives, to inform the development and implementation of eDNA-based habitat monitoring programmes for Scotland going forward. Throughout this document we use the term 'eDNA-based' to encompass all DNA collected from environmental substrates, which includes both extracellular DNA and whole organisms such as soil fauna samples and microeukaryotes in marine sediments (Pawlowski et al. 2020).

Across the four surveyed habitat types we found that eDNA-based data can detect compositional shifts in species communities that are associated with ecosystem state or habitat classification (freshwater: loch Water Framework Directive Overall Status, marine: biotope, woodland: restoration/regeneration class, peatland: restoration class). Using Random Forest algorithms, the eDNA-based data can be used to classify sites according to ecosystem state or restoration gradient class. These findings were most evident for the freshwater and woodland habitats. While the data for the marine and peatland habitats were not sufficient for classification.

Numerous species with important biodiversity monitoring designations can also be detected including SSSI[1]-listed species, IUCN[2] threatened species, PMF[3] and invasive species (species whose introduction or spread threatens biological diversity).

In some cases, eDNA-based data can likely be fed directly into existing community-based indicator metrics. For example, marine sediment health scoring categories[4] were comparable to those calculated from morphological surveys and freshwater loch chironomid scoring produced similar values to those produced using best-matching conventional (CPET) data[5]. However, this approach underuses much of the data and alignment of eDNA-based data into existing models can produce differing results and

[1] Sites of Special Scientific Interest
[2] International Union for Conservation of Nature
[3] Priority Marine Feature
[4] Using the AZTI Marine Biotic Index (AMBI)
[5] Using the Water Framework Directive Chironomid Pupal Exuvial Technique (CPET)

might consequently not be accepted. eDNA-based data should primarily be viewed as a 'new' tool, with new models, not necessarily as a tool to shoehorn into existing indices (with exceptions).

Developing national eDNA-based datasets to operationalise these findings will require well-considered site choices along well-defined gradients that are of highest priority for meeting monitoring and reporting needs. The breadth of potential applications is large. Successful future development and implementation will depend on posing targeted biomonitoring questions for specific objectives within national and international reporting frameworks.

The specific Key Recommendations from this project are:

- For freshwater lochs, build a national ecosystem-state prediction tool based on the methods presented in the project. This would be a scalable and efficient method for tracking loch quality state and change. It will require multiple lochs across a wide geographic range.
- As part of the ecosystem-state prediction tool, conduct a validation study for chironomid scoring by conducting side-by-side studies with conventional methods (CPET).
- For marine monitoring of vertebrate PMF species, develop standard monitoring guidance using aquatic eDNA sampling.
- For marine sediment health scoring, validate further at sites with greater pollution gradients.
- For marine biotope classification, conduct further research into optimal eDNA assays for maximised indicator species detection.
- For woodland, and other terrestrial habitats which use fungi as part of SSSI selection, further validate eDNA-based approaches for detection of SSSI-listed fungal species.
- For woodland restoration/regeneration monitoring, we initially recommend using eDNA-based data at the site level to monitor programme progress. In the longer term, a national eDNA-based survey across multiple woodland types in Scotland could be used as input to a systematic conservation planning exercise to rank woodlands by conservation value, and the higher-value woodlands can then be used as restoration targets.
- For peatland, there was a clear difference between degraded and restored peatlands, but the classification model was unable to predict status, due to the small size of this dataset. Classification of restoration status from eDNA-based data will require a large training dataset with a suitable sampling design and clear status definitions.

The Key Knowledge Gaps & Barriers are:

- Using eDNA-based data for biomonitoring at a national level in a regulatory context requires ecological frameworks based on national baselines, such as Ecological Quality Ratio (EQR) models. There are currently very few such frameworks based on or incorporating eDNA data (the Lake Fish Classification Index being the exception). Developing such frameworks requires large scale studies with focussed objectives. Biomonitoring at local scales is already possible through careful study design.
- The number of samples required for biomonitoring at the national level using eDNA-based data remains largely unanswered. This is partially due to fact that the breadth of potential applications is large, spanning numerous taxonomic groups, habitats, and biomonitoring objectives. Identifying the number of samples required for each specific biomonitoring objective is required.

- There are numerous eDNA-based projects being carried out in Scotland at various scales and in various contexts of biomonitoring, yet the data is not being captured in a systematic and unified way. Standardised guidance for formatting and storing eDNA-based data in publicly available databases would allow research in this area to progress faster. Systems such as the European Nucleotide Archive provide platforms for that could be used to store and access eDNA-based data for biomonitoring.
- There remains opportunity to develop minimum standards and validation scales to ensure consistency across projects and providers.

In most cases, eDNA-based approaches can be used to classify sites along ecological gradients. Until larger ecological biomonitoring frameworks for eDNA-based data are developed, eDNA-based approaches for national level reporting will likely remain underutilised. In the meantime, practitioners are using them for efficient surveying of key taxonomic groups. Local and regional projects are already using eDNA-based approaches to monitor negative and positive impacts of land management and restoration. The true power of eDNA-based data lies in the ability to generate huge datasets that can be used build national-level models of biodiversity and characterise ecological conditions for robust and consistent monitoring and reporting purposes.

Overall, eDNA-based approaches can provide the necessary scaling up of biodiversity monitoring for a national monitoring strategy across multiple habitat types, increasing the number of samples that can realistically be collected and analysed, and improve the reporting efficiency through standardised field and laboratory methodologies and data formats.

1 Introduction

1.1 Background

Biodiversity loss is widely recognised as one of the most urgent global challenges to be addressed in the next decade. The Scottish Biodiversity Strategy sets out a clear ambition to be Nature Positive by 2030, and to have restored and regenerated biodiversity across the country by 2045 (Scottish Government 2022). To protect, restore, and regenerate biodiversity, it is necessary to be able to accurately describe and quantify ecological change. Biodiversity monitoring through environmental DNA (eDNA) analysis is increasingly being used for tracking species diversity and composition in ecosystems as it is a scalable and high-resolution method.

The vast complexity of nature makes it almost impossible to capture the full scope of biodiversity, and despite decades of monitoring efforts, large gaps remain in our knowledge of biodiversity and how it responds to different pressures. Critical to solving the global biodiversity crisis is improving the pace, scope, and scale of data collection to better monitor progress relative to actions and interventions to better inform future responses. eDNA-based monitoring has the potential to be truly transformative in this regard, enabling biodiversity to be measured and monitored at large geographic scales. Additionally, there is a need for straightforward information that is easy to communicate. eDNA-based monitoring generates sufficient data for the application of ecological statistics, but also holds the potential for the development of biotic indicators and/or metrics for a wide range of habitats and geographies.

The goal of the overall project was to identify the potential to implement eDNA-based methods for efficient biodiversity assessment, by or for the Scottish Government, as part of the Scottish Biodiversity Strategy. The first output of this project (Phase 1; Cruickshanks et al. 2022) was a literature review of the state-of-art DNA-based monitoring technology, including the identification of areas where such approaches can be applied to fulfil biodiversity reporting requirements in Scotland. The second output (Phase 1: Egeter et al. 2023) followed the completion of small-scale pilot studies across four different habitats (marine, freshwater, woodland, and peatland). It included a habitat-scale sampling plan for a Proof-of-Concept eDNA-based survey for Phase 2 of the project that could potentially be applied, upscaled, and further developed by the Scottish Government and its collective organisations to inform future habitat monitoring programmes. This Phase 2 Main Report is the third output of the project, containing core work, where we undertook eDNA-based sampling and analyses across ecological categories of interest to the Scottish Government in each of the four habitats and tested whether eDNA-based data could detect trends across gradients/categories. The project results have been used to develop a set of recommendations for future development. This Phase 2 Main Report is also underpinned by the Phase 2 Technical Appendices output (Bakker et al. 2023c) and its key messaging has been succinctly captured elsewhere (see Summary Brief; Bakker et al. 2023b) to help inform a wide range of end users and key decisions going forward.

Stakeholder workshops have been held at key points throughout the project, to help facilitate knowledge-exchange (KE) and capacity-building opportunities between the contractors, researchers, key project partners, and the wider end-user community, spanning many people and organisations. These initial building blocks are essential for end-users to further develop the potential utility of eDNA-based monitoring approaches, which apply metrics for protecting, restoring, and regenerating nature in innovative and cost-effective ways. It is anticipated that operationalising future habitat monitoring

programmes, by deploying eDNA-based monitoring methods at the habitat-scale, will need to be informed by an overarching national biodiversity monitoring plan. Therefore, the development of a strategic biodiversity monitoring plan was considered out of scope for this project.

By completion, this project envisages bringing Scottish Government conservation and environmental regulatory agencies closer to implementing eDNA-based approaches and applying eDNA-based metrics, as a replacement, addition, or improvement to the suite of conventional monitoring methods that are currently being used, for enabling post-2020 biodiversity monitoring and for reporting requirements to be met at national and international scales.

For the full justification of the sampling plans, details on the sites chosen in this report, initial pilot study results, and laboratory and bioinformatics methods for each habitat, please refer to Egeter et al. (2023). All bioinformatics analyses can also be found in the Phase 2 Technical Appendix output (Bakker et al. 2023c).

1.2 Aim & Key Objectives

The overall project aim was to investigate and test the applicability of eDNA-based monitoring approaches for biodiversity assessment and reporting purposes across a broad range of habitat types in Scotland, including marine, freshwater, woodland, and peatland habitat types (Cruickshanks et al. 2022; Egeter et al. 2023). Overarching research questions were set out to be addressed and were as follows:

Can eDNA-based metabarcoding community data be used in habitat-scale monitoring programmes to:

- Apply eDNA-based units/metrics that can act as indicators of biodiversity and be used or potentially aggregated to (*a*) distinguish (classify or sort) habitat condition and (*b*) detect ecological pressures or restoration stage of marine, freshwater, woodland, and peatland habitats in Scotland?
    - If we find any differences, are these data sufficiently clear and predictable to be used to assess and enhance our understanding of how pressures (e.g., land use, pollution/water quality, invasive non-native species (INNS) and/or climate change) or restoration are affecting ecosystem health (condition, function, and resilience) in different aquatic and terrestrial habitats in Scotland?
- Explore, wherever feasible and sufficient data exists, if there are any differences or similarities in habitat condition or biodiversity and community composition from any re-surveyed locations between timescales (e.g., seasons or years) and protected status (e.g., designated vs non-designated sites) for each of the four habitat types sampled in Scotland?
- Extend DNA-derived metrics from taxonomic descriptions to assess possible functional changes across a broad range of a range of habitats in Scotland?

How does the utility of eDNA metabarcoding data compare, where best available matching data exists, with using conventional morphology-based or scoring-based methods for:

- Assessing the ecological status and, where relevant, restoration stage of marine, freshwater, woodland, and peatland habitats in Scotland?
- Assessing whether eDNA metabarcoding community data can provide similar and reliable outcomes as conventional methods across a broad range of a range of habitat types in Scotland? What are the benefits and disbenefits that arise from eDNA-based vs conventional methods?

- Assessing site condition (e.g., presence-absence of selected or agreed conservation features; species-assemblage recorded) of high conservation value sites in Scotland's designated site network (e.g., Marine Protected Areas, MPAs; Special Protection Areas, SPAs; Special Areas of Conservation, SACs; Sites of Special Scientific Interest, SSSIs), where sampling design reasonably coincided with protected areas across the marine, freshwater, woodland, and/or peatland habitats selected for this study?

What are the key recommendations and future perspectives to develop and operationalise standardised eDNA-based methods at national scale for monitoring biodiversity and the impacts of pressures or restoration across a broad range of habitat types in Scotland?

- How should we sample?
- What should we sample?
- Where should we sample?
- When should we sample?
- Where do the knowledge gaps and barriers exist, and why is that? What practical solutions and technology innovations would help overcome these types of challenges going forward?
- What have we (*a*) learned to do the same or differently and (*b*) anticipated are key opportunities to develop, upscale, and implement DNA-based methods for each habitat type in Scotland?
- What other habitat types across Scotland do we need to consider sampling for eDNA-based biodiversity monitoring and reporting purposes in the future?

It was recognised that the breadth of biodiversity and ecosystem health assessment across each of the four habitat types would be constrained on the project by the available resources and the contractor's current portfolio of analytical products and services, and as such there was potential for ecological gaps and uncertainties. However, the decisions about eDNA-based analyses used were reasonably balanced to deliver the project scope against the maximum budget and timeline available.

This was a Proof-of-Concept study across small numbers of sites and ecological gradients or categories. Each of the habitats monitored are subject to their own range of human-induced pressures, from pollution in freshwater lochs to the draining of peatlands, which have caused widespread declines in ecosystem integrity. We therefore focussed our efforts on identifying eDNA-derived indicators along restoration and pressure gradients that are specific to each habitat.

1.3 Focal Study Area

The project study area was mostly focussed on the Loch Lomond and Trossachs National Park (LLTNP) (https://www.lochlomond-trossachs.org/) and its adjacent areas. The LLTNP was chosen because it contains a mosaic of different habitat types occurring across Scotland and includes areas of conservation importance and protected status, as a part of Scotland's designated site network (e.g., Marine Protected Areas, MPAs; Special Protection Areas, SPAs; Special Areas of Conservation, SACs; Sites of Special Scientific Interest, SSSIs). However, as the project developed, the geographic radius for undertaking eDNA-based sampling and analysis was relaxed and extended beyond LLTNP boundaries, including the Cairngorms National Park. The purpose of this was to enable adequate study design and establish more habitat condition replicate sites where needed.

1.4 Research Approach

To address the objectives of the project, sites were selected that reflect a category or gradient of ecological condition for each habitat. For each site, eDNA surveys were designed to effectively capture

ecological variation across the site and either water (marine and freshwater), sediment (marine) or soil (woodland and peatland) samples were collected. These samples were sent to NatureMetrics laboratories, where eDNA was extracted and metabarcoding performed to generate community datasets. Information on survey design, sample collection, and analyses can be found in the methods section of this report and in Egeter et al. (2023). We then used a combination of sample-level biodiversity metrics and community analyses to compare the eDNA data across the conditions and categories. Where feasible, we also calculated conventional indicator values that are typically generated using morphological-based identification and compared these values with best available matching data obtained by conventional methods. We focussed on taxonomic groups that are either currently monitored or have key ecological roles in ecosystems.

1.5 Report Structure

This report provides the overall background, justification, and aims of the project, as well as the results of the Phase 2 study and future recommendations. The report is framed around specific research questions for each habitat in the context of the overarching research questions. In this main report, we have also synthesised and summarised the key results and findings. Further detailed results are captured in the Phase 2 Technical Appendices output (Bakker et al. 2023c) and referred to where appropriate. In section 2, we cover the applied methods, in section 3 the key findings per habitat, in section 4 a summary of key findings, knowledge gaps, barriers, recommendations and future perspectives, in section 5 a glossary, and in section 6 a bibliography. Section 7 contains several appendices providing additional background information.

2 Methods

2.1 Site Selection & Sampling Approach

The geographic scope of sites sampled in this project was limited, so we prioritised selection of sites that have clear restoration or pressure gradients that are relevant across similar habitats in Scotland. Therefore, while the environmental context of other sites across Scotland may create local-level differences in biodiversity responses, the eDNA-derived metrics and analyses for different taxonomic groups tested in this project will be applicable to sites across similar restoration or pressure gradients. A brief overview of the sampling plan is given separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c). For the full justification of the sampling plans, details on the sites chosen, initial pilot study results, and laboratory and bioinformatics methods for each habitat, please refer to Egeter et al. (2023).

2.1.1 Marine

Within the overarching questions for the marine habitat we asked whether eDNA-based metabarcoding community data can be used to identify specific biotopes, and identify Priority Marine Features (PMF).

A total of twenty sampling locations were selected in the Loch Long region (Loch Long, Loch Striven, Kyles of Bute, Gare Loch, Holy Loch, Loch Goil; Figure 1), representing four Level 5 biotopes that were not geographically clustered, split between two Level 4 biotopes according to 2013/14 survey reports (C. R. Allen et al. 2013). These were:

- Seapens and burrowing megafauna in circalittoral fine mud (SS.SMu.CFiMu.SpnMeg), which is a PMF biotope (Level 4 SS.SMu.CFiMu)

- Burrowing megafauna *Maxmuelleria lankesteri* in circalittoral mud (SS.SMu.CFiMu.MegMax) also a PMF biotope (Level 4 SS.SMu.CFiMu)
- *Cerianthus lloydii* and other burrowing anemones in circalittoral muddy mixed sediment (SS.SMx.CMx.ClloMx; Level 4 SS.SMx.CMx)
- *Ophiothrix fragilis* and/or *Ophiocomina nigra* brittlestar beds on sublittoral mixed sediment (SS.SMx.CMx.OphMx; Level 4 SS.SMx.CMx

We aimed to collect triplicate water and sediment eDNA samples from each station to increase the likelihood of detecting PMF species, enable statistical comparison across stations and biotopes and, where possible, advise on the minimum number of samples required to detect PMF species. For one grab from each station, macrofaunal morphological analysis and PSD (Particle Size Distribution) analysis were undertaken, to provide accurate and contemporary conventional biotope information, and to compare species detected with eDNA-based data with morphological methods. Sampling was conducted on 30-31 August 2022. Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

Water sample collection

Forty-six out of a targeted 60 water samples were collected from a vessel using a vertical 7.5-litre Niskin bottle (+CTD) from a depth of 25m (where possible), as well as three field blanks. Although triplicate samples were planned, only two samples could be collected at 14 stations due to intensive resource efforts (e.g., boat access, expertise, and staff time for field sampling), and distances between the individual sampling sites, resulting in time restraints during sampling. Five litres of sea water per sample was collected, with one exception of 2.5 L due to a leaking bottle. Each sample was passed through an encapsulated 0.8 µm PES filter (including a 5 µm glass fiber prefilter) supplied in a NatureMetrics 'Pump Aquatic eDNA Kit'. The volume of water passed through each filter was recorded. Filters were preserved with Longmire's solution and kept at ambient temperature for two days, and subsequently stored at -20°C until return to the NatureMetrics laboratory where they were stored at -20°C until analysis. NatureMetrics have found no difference between communities recovered from samples that have been stored at ambient temperatures in Longmire's solution for up to 2 weeks and from samples that have been immediately frozen after collection.

Sediment grab samples for DNA analysis

Fifty-six out of a targeted 60 marine sediment samples were collected using a 0.1 m^2 Day grab. One replicate was not obtained for station LL54, and grabs came up empty at LL27. Four syringe core samples were collected from each grab and mixed in a plastic bag to form a composite sample. The samples were not processed further on-site. Samples were preserved in cold storage for two days and then stored at -20°C until shipment to the NatureMetrics laboratory, where they were again stored at -20°C prior to further processing.

Sediment grab samples for morphological analysis

One grab per station, totaling 20 samples, was analysed using morphoanalysis. The contents of each grab were sieved on board the vessel using a 1 mm sieve mesh and retained material was preserved with formaldehyde, and subsequently transported to SEPA laboratories for analysis following the NMBAQC methods (Worsfold and Hall 2010).

Sediment grab samples for particle size distribution (PSD) analysis

One grab per station, totaling 20 samples, was selected for PSD analysis. 100 ml of sediment was collected using a syringe corer and mixed in a plastic bag. Samples were preserved in cold storage for

two days and then stored at -20°C until transport to SEPA laboratories for analysis following the NMBAQC methods (Mason 2016). Grabs came up empty at LL27. Both morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

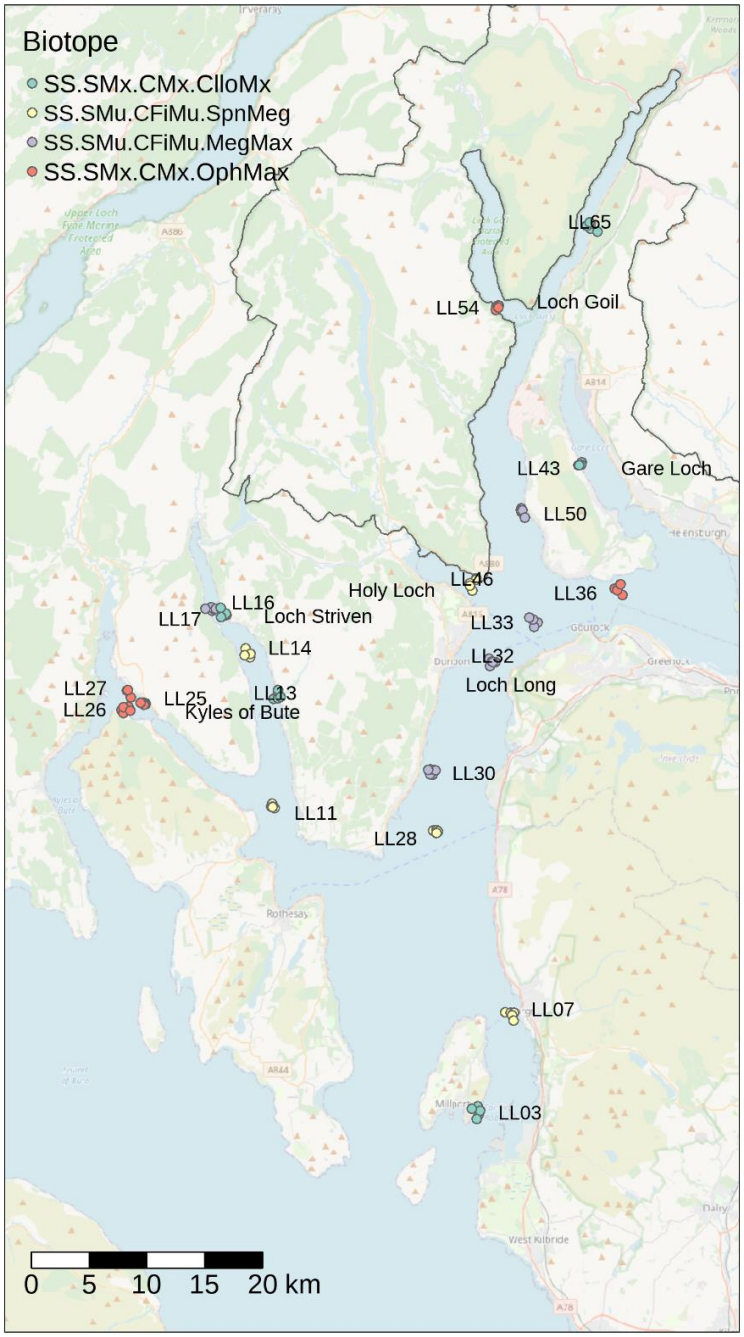


Figure 1: Map of the marine sampling locations. The outline within the map represents the boundary of LLTNP. Biotope classifications based on previous surveys are colour coded. Basemap: OpenStreetMap.

eDNA assay selection

Assay selection focused on detecting as many PMF species as possible and on assessing whether eDNA-based community data could show discernable differences between biotopes. Many of the marine PMF species are benthic invertebrates. Accordingly, using at least one invertebrate assay for the sediment samples was essential. The 18S invertebrate assay used (Capra et al. 2016) did not perform well in either PMF species detection or overall species resolution during the pilot study (Egeter et al. 2023) and alternative assays were selected: the first targeting a sequence within the 18S invertebrate gene (Amaral-Zettler et al. 2009, and the second targeting a sequence within the COI eukaryote gene(cytochrome c subunit I; Leray et al. 2013). The 16S bacteria assay used in the pilot study (Egeter et al. 2023) was also selected as it showed high levels of heterogeneity across sampling stations and this taxonomic group has the potential to be used assessing pressures (Borja 2018; Lejzerowicz et al. 2021). Based on the pilot results of the water sample analyses, where the fish assay (Miya et al. 2015) detected ten PMF fish species, and two PMF marine mammal species, the same assay was selected. The 16S bacteria assay was not selected for the water samples because results from the pilot study indicated a very high level of heterogeneity in bacterial communities, even at a very small scale. Increased sampling effort would therefore be required to capture a more complete representation of the bacterial community. Instead, it was decided to reallocate effort to the sediment samples to capture as much invertebrate diversity as possible.

Biotoping

Marine biotoping for Scotland is based on the EUNIS habitat classification, with methods outlined by JNCC (Parry 2019). Initial habitat classification (Levels 1 and 2) is made based on the physical characteristics of depth, salinity, and proximity to shore. The habitat is then further defined based on the substrate type (Levels 3 and 4). In the previous reports (for our selected stations) this was based on benthic imagery data (C. R. Allen et al. 2013; C. G. Moore 2013), although particle size distribution (PSD) analysis of sieved sediment, followed by the application of the Folk classification method for sediment classification is the preferred method. For this study, PSD analysis was therefore conducted on one sample per station to provide details on the benthic substrate, using the Folk classification method. Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

For assigning the final biotope level (Level 5), the biological communities were compared with the characterising species lists available from JNCC (https://mhc.jncc.gov.uk/). Where this conflicted with the PSD observed, the biotope was assigned based on the biological community. Level 5 biotoping was also attempted using the sediment eDNA-based data.

2.1.2 Freshwater

The overarching question we aimed to answer for the freshwater habitat, is whether eDNA metabarcoding community data can be used to assess habitat condition of Scottish freshwater lochs. Rivers, streams, and ponds were not targeted. Rivers were sampled for the pilot project, but given the indications from the pilot study results that 5-10 samples were sufficient for lochs, while multiple replicate samples per sampling point are possibly required for rivers, focusing on lochs enabled the inclusion of more sites, better representing a range of habitat conditions, rather than a very limited number of rivers.

The Water Framework Directive (WFD) uses a system of five ecological status quality classes for waterbodies in Scotland. The classification encompasses many categories such as ecology, fish status, water chemistry, hydrology, and morphology. It is a hierarchical ‘one-out-all-out’ system, whereby the

overall class is determined by the worst class of the individual elements. The WFD classification system was used to help identify and select the range of freshwater lochs, mostly situated within LLTNP focal study area.

We specifically aimed to test whether lochs experiencing similar hydrological pressures (so much as is possible to control) contain different biological communities that can indicate overall condition. The principal habitat condition gradient investigated was the WFD 'Overall Status'. We also investigated land use, alkalinity, chlorophyll-a, nitrate, nitrite, dissolved oxygen, and total phosphorus. Land use was extracted for each catchment area and expressed as percentage cover for the four principal land use types (Moor, Arable, Woodland, and Urban). For nutrient and chemical data, annual mean values were used for each loch.

A total of 15 freshwater lochs (Figure 2) were selected based on their location within, or their proximity to LLTNP, accessibility by road, and where shoreline sampling would be sufficient (to minimise resource constraints, also keeping in mind potential future monitoring programmes). All 15 lochs had previously been classified by SEPA using the WFD 'Overall Status' designation and we chose them against the criteria specified that could be met for High, Good, Moderate, and Poor classification status. However, no representative loch examples were available for Bad classification status that met all the criteria we had specified and were not significantly impacted by hydrology and/or morphology pressures. See the pilot study findings & phase 2 sampling plan for further details (Egeter et al. 2023). Six sampling locations (i.e. six samples) were collected for each loch. It should be noted that this is lower than the minimum sampling effort of 10 samples required to detect ≥85% of fish species present in UK lakes (Li et al. 2019), or the minimum of 10 required for the Lake Fish Classification Index (Willby et al. 2020). However, as the goal was not to specifically detect all fish species or validate the Lake Fish Classification Index, and we needed as many separate lochs as possible to apply the planned statistical approaches, this fixed number was considered the reasonable balance between maximising the number of lochs and the available resources (budgetary constraints) to deliver the project work. Sampling locations were set at equidistant intervals where possible around the perimeter of the loch , as is best practice for lake fish eDNA sampling (Hänfling et al. 2016; Lawson Handley et al. 2019; Li, Lawson Handley, et al. 2019; Zhang et al. 2020).

Water sample collection

Sampling was conducted between the 23rd and 27th August 2022. At each sampling location, 6 L of water was collected and split across two sterile sampling bags. Each 3 L sample was comprised of 12 x 250 mL subsamples, with subsamples collected at 10 m intervals along 120 m of shoreline using a sterile dipper ladle. This broadly follows existing eDNA sampling protocols for lochs, albeit the subsample volumes and overall sample volume differ (Bedwell and Goldberg 2020; Hänfling et al. 2016; Lawson Handley et al. 2019; Li, Lawson Handley, et al. 2019). A total of 90 samples and five field negative controls (one per day of sampling) were collected. Each sample was passed through an encapsulated 0.8 μm PES filter (including a 5 μm glass fiber prefilter) supplied in a NatureMetrics 'Aquatic eDNA Kit'. The volume of water passed through each filter was recorded. Filters were preserved with Longmire's solution and transported to the NatureMetrics laboratory at ambient temperature and stored at -20°C until further processing.

eDNA assay selection

Based on the pilot project results, all freshwater samples were analysed using vertebrates (Riaz et al. 2011; Kelly et al. 2014), freshwater invertebrates (Leese et al. 2021), and bacteria (Caporaso et al. 2011) assays. For the vertebrates and invertebrates assays the primary reason was because these assays

detect numerous species that are currently monitored. The bacteria assay showed potential for being able to classify sites according to habitat condition.

Figure 2: Map of the freshwater sampling locations. The outline within the map represents the boundary of LLTNP. Sites are coloured by the Water Framework Directive Overall Status. Basemap: OpenStreetMap. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

2.1.3 Woodland

Woodland regeneration is central to Scottish Government ecosystem restoration goals and efficient indicators are needed to understand progress of forest restoration schemes and to assess woodland condition. Note that we use the term "condition" for woodland in a general sense of the restoration gradient that was targeted. We do not imply that this is equivalent to the use of the term in the context of Site Condition Monitoring as defined by NatureScot guidance (https://www.nature.scot/professional-advice/protected-areas-and-species/protected-areas/site-condition-monitoring). The primary research goal for the woodland habitat was to assess whether eDNA data can be used to monitor woodland restoration, from unforested, to recently planted/regenerating, to mature status, using the different stages of restoration as a proxy for

monitoring over time. We focused on Caledonian Scots pine as it is a priority habitat and, given the number of samples that could feasibly be collected within this project, multiple woodland types would have risked an overly variable and noisy dataset that could have hindered interpretation.

Although classifying Caledonian Scots pine woodland as mature or otherwise could be done using conventional methods, eDNA approaches have the potential to 1) monitor below ground biota with relatively simple field collection methods and 2) monitor the environment without the requirement of extensive in-field (e.g. plant identification) or in-lab (e.g. invertebrate identification) taxonomic expertise. eDNA metabarcoding has the potential to be used in a standardised way to produce numeric values/indicators to track progress along a restoration gradient.

Sample collection was undertaken in nine sites within four geographic areas (LLTNP, Cairngorms, Glen Affric, and Glen Moriston; Figure 3). Sampling locations within sites were chosen to represent a restoration gradient of Unforested, Recently Planted/Regenerating, and Mature Scot's Pine habitats. It should be noted that not every site possessed all three conditions. The samples collected from Rothiemurchus in the Cairngorms National Park were part of a separate concurrent project undertaken by NatureMetrics for Forest Research. Permission was obtained from Forest Research to include those samples in this project as they provided an opportunity to increase the dataset size in the Recently Planted/Regenerating and mature gradient categories. "Unforested" sites were characterised primarily by an absence of trees but comprised a range of habitats dominated by grasses, sedges, heathers, and rushes. No vegetation surveys were undertaken as part of this project. It was not assumed that these sites were inherently of low condition or value due to their classification as "Unforested". "Recently Planted" sites were also variable in their composition and represented a range of intermediate conditions where trees had either been planted or were regenerating. Site photographs are provided separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c).

A total of 107 samples were collected, each consisting of nine subsample cores collected across a ~10m x 10m plot and mixed into one composite soil sample. A metal soil augur (inner core diameter 14 mm) was used to collect subsample cores to a depth of ~10 cm. Any living moss at the top of the core was discarded. The subsamples were thoroughly mixed in a plastic grip seal bag by shaking and massaging. Note that further manual mixing is carried out in the lab prior to DNA extraction. Samples are ~ 40 g and 10 g is used for DNA extraction. Internally, NatureMetrics has compared processing samples in triplicate and found very little difference in communities, indicating that this homogenisation process is effective. Samples were kept on ice in a cool bag in the field and while in transit and stored in a freezer prior to transport to the laboratory where they were stored at -20°C until further processing.

eDNA assay selection

The woodland soil samples were analysed using the bacteria, fungi, and soil invertebrate assays. These were chosen as they are the most common taxonomic groups targeted in soil metabarcoding studies and all have potential for habitat characterisation and classification.

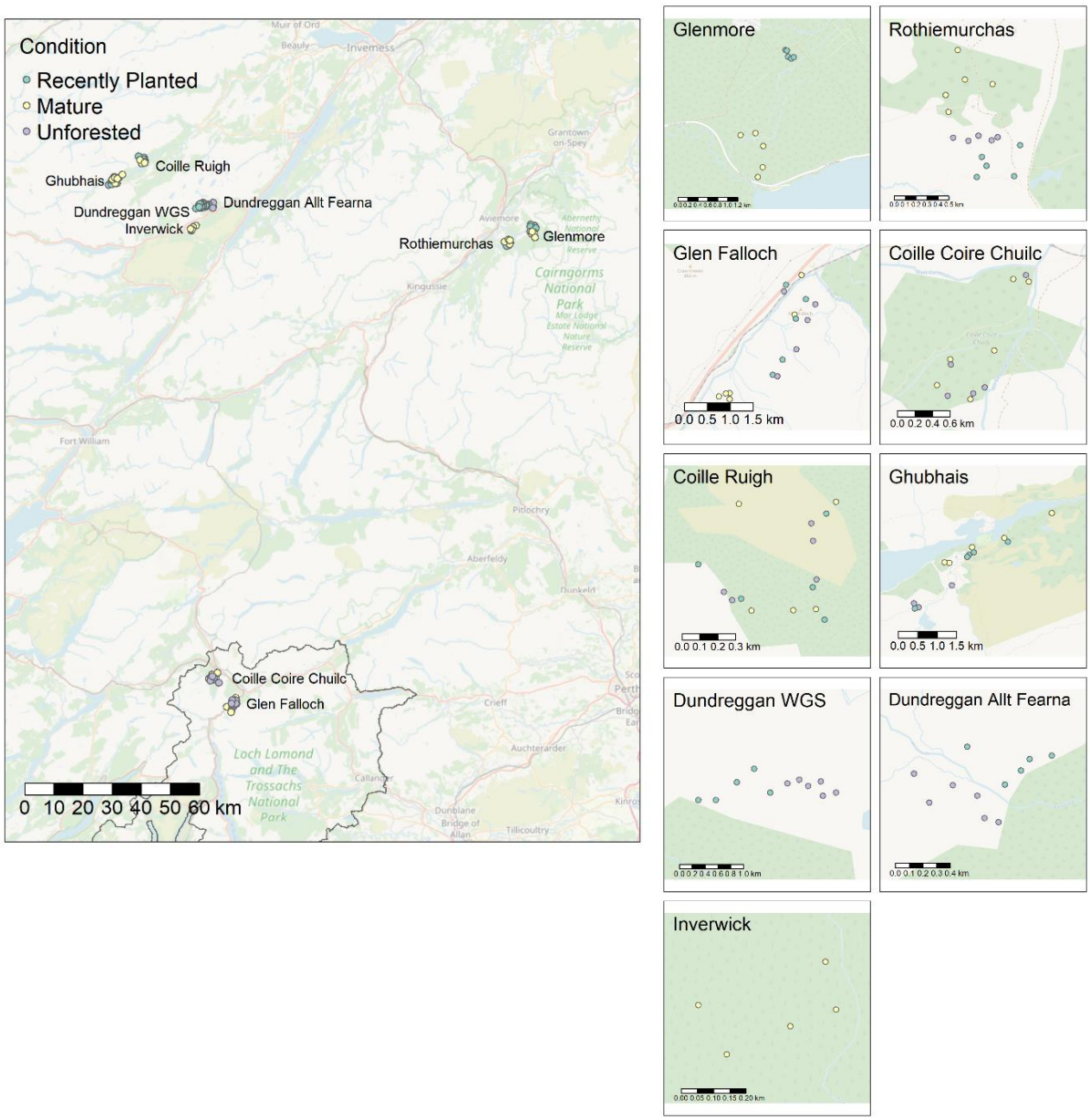


Figure 3: Map of the woodland sampling locations. The outline within the map represents the boundary of LLTNP. The term Recently Planted refers to sites that had either planted or young regenerating Scots pine. Basemap: OpenStreetMap.

2.1.4 Peatland

Many Scottish peatland areas are degraded, due to high intensity grazing and the installation of drainage channels and require suitable management and restoration action as a result. The primary goal was to assess whether eDNA data can be used to monitor peatland restoration, from degraded to restored. Although classifying peatland as degraded or restored can be done using conventional methods, eDNA approaches have the potential to 1) monitor below ground biota with relatively simple field collection methods and 2) monitor the environment without the requirement of extensive in-field (e.g. plant identification) or in-lab (e.g. invertebrate identification) taxonomic expertise. eDNA community data have the potential to be used in a standardised way to produce numeric values/indicators to track progress along a restoration gradient.

Site selection criteria required sites with two peat conditions; degraded and restored. Originally a third category (unimpacted) was proposed. However, 70% of Scotland's blanket bog and 90% of Scotland's raised bog peatland is degraded (Artz et al. 2014) and no unimpacted peatland sites could be identified to be included in the study. All three peatland sampling locations, Glen Finglas, Auchlyne, and Cashel are situated within the LLTNP (Figure 4). Glen Finglas and Auchlyne contain drained and restored (through grip blocking) peatland. The Cashel site covers a large area on the south-east side of LLTNP but did not contain any areas that were not drained. Site photographs are provided in the Technical Appendix.

Peatland soil sample collection

A total of 50 samples were collected from ~10 x 10 m plots of homogenous habitat at each location, adjacent to blocked/unblocked drains. Nine subsample cores were collected per plot to align with the woodland sampling methodology and the collection process was identical.

eDNA assay selection

Based on the key outcomes from the pilot study, the soil samples were analysed using the bacteria, fungi, and invertebrate assays. These were chosen as they are the most common taxonomic groups targeted in soil metabarcoding studies and all have potential for habitat characterisation and classification.

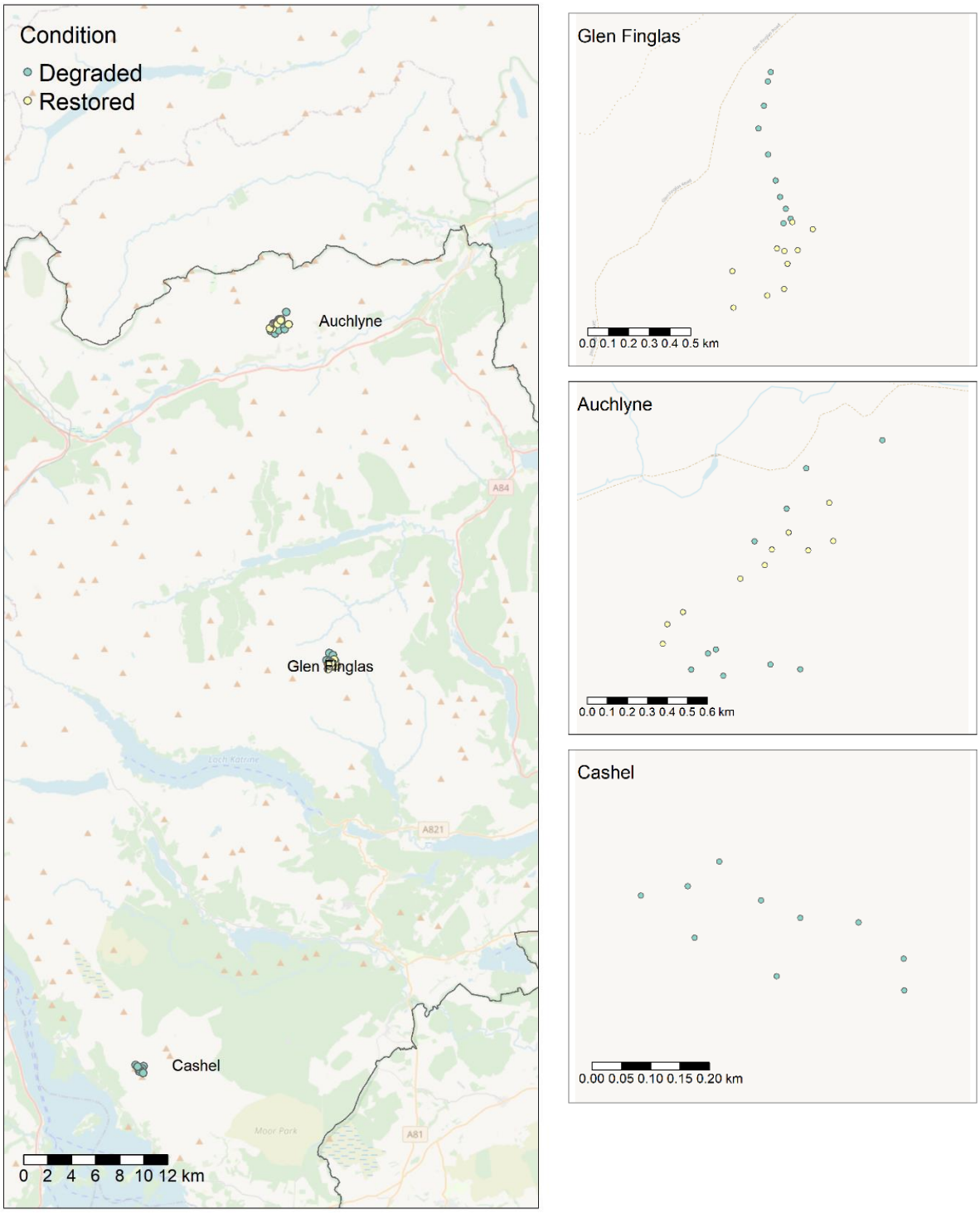


Figure 4: Map of the peatland sampling locations; Glen Finglas, Auchlyne, and Cashel. The outline within the map represents the boundary of LLTNP. Basemap: OpenStreetMap.

2.2 Data analysis

For all data analyses in this project, only target Operational Taxonomic Units (OTUs) were utilised. These are OTUs belonging to taxa that are targeted by the selected assay. For example, for the fish assay, only OTUs identified as fish were used (with the exception of reporting marine mammal PMF species). See the Phase 2 Technical Appendices output (Bakker et al. 2023c) for a breakdown of what is considered target for each assay. OTUs that could not be assigned to at least Kingdom level were excluded. For the freshwater vertebrate dataset, we limited analyses to the fish species detected because incidental detections of other more sporadically detected vertebrates, especially bird species, led to ecologically irrelevant outcomes. For example, gulls (Laridae) were playing a significant part in the model, when gull eDNA detection is likely to be detected sporadically and have little ecological relevance to water quality or health monitoring.

Many metabarcoding studies use an OTU approach as standard as it limits the effect of PCR and sequencing errors on detected species diversity by clustering highly similar sequences. In contrast, Amplicon Sequence Variants (ASVs) keep each unique DNA sequence separate but attempt to fix potential PCR and sequencing errors based on error models. We report OTUs because they are easier to interpret as they are closely related to the species concept. However, our taxonomic assignments (Egeter et al. 2023) are made via sequence similarity searches of the ASV sequences, and only after taxonomic assignment are the DNA sequences clustered to produce OTUs. This is more robust than the traditional approach of clustering of OTUs prior to taxonomic assignment.

2.2.1 Species of Note

For each of the habitats studied, we highlighted relevant notable species. These included:

- Invasive species: Species identified were checked against the Global Register of Introduced and Invasive Species (GRIIS), which is an IUCN Invasive Species Specialist Group initiative (following methods in Pagad et al. 2022; 2018).
- IUCN listed species: species identified were checked against the IUCN Red List (https://www.iucnredlist.org/).
- Fungal Sites of Special Scientific Interest (SSSI)-listed species: fungal species were checked against the JNCC Guidelines for the Selection of Biological SSSIs Chapter 14 Non-lichenised fungi (Bosanquet et al. 2018).
- Priority Marine Feature (PMF) species: species were checked against the NatureScot Priority Marine Features in Scotland's seas (Tyler-Walters et al. 2016).
- During the review process, expert opinion was also sought from project team members and report draft reviewers on detections of species, with the specific request to note observations that were novel, unusual, or potentially suspect (potential false positives).
- Note that none of the above registers/lists are relevant for bacteria and as such there are no bacteria listed as species of note in this report.

2.2.2 Sample-level Indicators from eDNA

To assess whether simple sample-level indicators derived from eDNA could provide additional insights or metrics that are easier to interpret as predictors of habitat condition or restoration stage, we calculated a range of metrics and compared them across the treatments. All sample-level metrics were compared using linear models with the metric value as the response and the treatment as predictor.

Site (or station for the marine habitat) was included as a random factor to account for geographical variation. Covariates collected during sampling were included in the models (pH, soil moisture, temperature, depth, etc). Models were run in R version 4.2.3 (R Core Team 2021) using the lme4 package version 1.1.33 (Bates et al. 2015) in the form of:

$$metric\_value \sim treatment + (1 | Site) + covariate.1 + covariate.2 + covariate.3$$

Where significant differences were observed across treatments, post hoc comparisons were carried out using the emmeans package version 1.8.5 (Lenth 2023). As all response variables were continuous, we used Gaussian link functions for all metrics. Residuals were checked by plotting residual versus fitted values. All residuals plots were reasonable, indicating assumptions of normality were not violated. Model outputs are provided separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c). We focus in this document on the sample-level indicators that were significantly different across habitat conditions or restoration stage. The sample-level indicators that were tested, include:

All habitats and assays:
Species Richness: A count of the number of OTUs of the target taxa detected in a sample or site.

Evolutionary Diversity: The evolutionary breadth of species that are present in a sample. This is also known as phylogenetic diversity. It was calculated following Luo et al. (2020) and Lund et al. (2022). Briefly, sequences were aligned using the MUSCLE algorithm (R. C. Edgar 2004). Based on this alignment, a matrix of pairwise dissimilarities was built and used to generate a phylogenetic tree based on unsupervised clustering (UPGMA). Faith's PD (Faith 1992) was calculated for each sample in turn, using the total length of sample tree branches.

Bacteria and fungi assays only:

- Bacteria Functional Diversity: The range of ecological functions that a bacterial community has the capacity to perform. This was calculated by processing the OTU tables through the Picrust2 pipeline. This takes the sequence associated to each OTU, places it into an extensive reference phylogeny based on a maximum-likelihood similarity algorithm (Douglas et al. 2020) and predicts the abundance of all gene families in the EC database (Enzyme commission; Bairoch 2000) for each OTU. The resulting OTU/gene family table is used to build a clustering tree of OTUs based on their similarity in terms of functional profile. The branching length of the tree was calculated for each sample (Petchey and Gaston 2006).
- Fungal Functional Diversity: The range of ecological functions that a fungal community has the capacity to perform, specifically the range of feeding methods and resources used by the community. This was calculated by comparing the OTUs within each sample to a database of fungal ecological traits, FUNGUILD, and assigning OTUs that had a match to one or more ecological categories (Nguyen et al. 2016). The resulting guild profile was used to categorise OTUs based on their ecological role and build a clustering tree with the resulting dissimilarity matrix. The branching length of the tree was calculated for each sample (Petchey and Gaston 2006).

Marine and freshwater habitats only:
Fish species were compared to a global fish database, FishBase, and three metrics calculated. FishBase has detailed information on how these are calculated (https://www.fishbase.se/manual/english/key%20facts.htm). All fish metrics were treated as continuous response variables rather than ordinal. We chose this approach because FishBase calculations (including mean of prey trophic level) provide a number with 2 decimal places. We then

used the mean of this per sample to calculate a z-score based on all samples in project (data is near normal distribution). Previous peer-reviewed publications analyse FishBase trophic level as a continuous variable (Mancinelli et al. 2013; Romanuk, Hayward, and Hutchings 2011; Sethi, Branch, and Watson 2010; Ghilardi et al. 2021; Pauly et al. 1998; Pauly and Watson 2005).

- Fish Trophic Level Index: this is the mix of trophic levels of the species identified in a sample. Trophic level refers to the food chain position, with plants being 1 and top predators being 3-5. Higher trophic level fish species are usually larger species that are often the first to be removed by fishing (fishing down the food chain). Lower trophic level fish are associated with overfishing and nutrient pollution (Caddy 1993; Pauly and Watson 2005; Soler et al. 2015), while higher trophic levels are related to healthier ecosystems.
- Fish Vulnerability Index: Species with low reproductive and slow maturation rates are more susceptible to change. Often larger species are the first to be affected by environmental disturbances (Duggan-Edwards et al. 2020; G. J. Edgar et al. 2014). The vulnerability per sample was calculated by taking the mean vulnerability of the species detected in the sample. A greater mean vulnerability indicates a healthier community, while a lower vulnerability of indicates a more degraded community.
- Fish Economic Value Index: Based on the market price/demand for each detected species, the price category (low, medium, high, and very high) for species detected was assigned a number (1-4) and the mean per sample was calculated.

2.2.3 Validating existing indices with eDNA data

For marine and freshwater habitats, we assessed the capacity for eDNA data to produce reliable values for use with established taxonomy-based indices. For AMBI and CPET (see below), we calculated the index value using morphological datasets and compared these to the index values generated from eDNA data from the same locations.

- AMBI (AZTI's Marine Biotic Index): This tool was developed by AZTI (http://ambi.azti.es) to assign mean Marine Biotic Index values for each sample. It is an established metric to determine the ecological status of coasts and estuaries using a database of >10,000 categorised benthic species. More recently, genomic AMBI (gAMBI) was developed (Aylagas et al. 2018) and showed a high correlation with AMBI (0.66 or 0.88 r^2). We trialled the application of AMBI and gAMBI to assess how similar the outputs are in the context of the Scottish marine environment. We used the species identified by sediment morphological analysis and by eDNA analysis as inputs to AMBI v6.0 (Borja, Mader, and Muxika 2012). Based on Aylagas et al. (2018), both inputs were binary; we did not include count data, as the authors found this was the most comparable approach.
- Chironomid pupal exuviae technique (CPET): Aquatic benthic invertebrates, of which chironomids are the largest family, are good indicators of nutrient enrichment and can be used to assess lake water quality (Kranzfelder et al. 2015; Poikane et al. 2016; Saulino et al. 2021). Typically, sampling involves the collection of passively drifting pupal skins that are easily collected, identified, and counted. The metric is based on the composition of chironomid species or groups of species in the sampled lake. Following the standard methodology (UKTAG 2008), we matched the nutrient sensitivity scores of chironomid taxa detected and divided the observed score by the predicted score to calculate the Ecological Quality Ratio (EQR), which is used to assign a water quality status from Bad to High. EQR ranges from one, indicating

invertebrate communities close to the natural state, to zero indicating a high level of anthropogenic impact. Matching morphological data was available from four lochs (Lake of Menteith, Loch Doilet, Loch Lomond (North), and Loch Lubnaig). We generated EQR scores using previously collected morphological data and compared this to EQR scores generated from eDNA-based community data.

- Lake Fish Classification Index: This tool was developed by Willby et al. (2020) to describe the impact of nutrient pressures on fish populations, by calculating EQRs for lakes, based on sample occupancy from eDNA sampling, that is, the proportion of samples which had a positive result for the presence of different species or combinations of species. The approach uses five species groups (brown trout, percidae, roach, salmon & charr & coregonids, and carp & bream), which were selected based on their ability to differentiate between sites with high and low nutrient pressures. It uses a clear classification system that categorises lakes into four classes: oligotrophic, mesotrophic, eutrophic, and hypertrophic. This was extensively discussed as part of the project (including with the developers of the tool). As the tool requires at least 10 samples per lake, any results of applying the tool to the dataset in this project (which had 6 samples per lake) could be misleading, so we did not incorporate this into our final report. Detections of fish via eDNA-based data from the development of the tool were provided (data not included in this report) and brief qualitative comparison of the results is included.

2.2.4 Classifying Habitat Conditions through Community Analyses

To assess the potential for eDNA-based data to distinguish habitat condition or restoration stage, we analysed community datasets for each assay in each habitat using two general approaches. First, we performed *supervised classification* using the 'randomForest' package v4.7-1.1 (Liaw and Wiener 2002). Second, we used the 'gllvm' package v1.4.1 (Niku et al. 2019; van der Veen et al. 2022) for *model-based community ordination* with site as a random factor and using (i) no predictor variables (henceforth '*unconstrained ordination*'), and (ii) including potentially influential covariates (henceforth '*constrained ordination*'), both of which are forms of *joint species distribution modelling (JSDM).* Supervised classification and JSDM are separate but complementary methods for analysing multivariate datasets.

We used supervised classification to quantify the ability of eDNA datasets to *predict* predefined habitat conditions, such as the WFD Overall Status. We used JSDM to infer (some of) the reasons why different conditions have different species compositions.

Supervised classification

We carried out site-level cross-validation tests by successively training a Random Forest model on all sites but one and testing how well the trained model predicted the condition of the samples from the hold-out site. We summarised predictive performance for each condition by calculating the percentage of correctly classified samples. For instance, if there are three condition classes (e.g. unforested, recently planted, and mature woodland), a model performing no better than random will classify samples correctly 33% of the time (assuming equal numbers of samples per class). For two condition classes, the random-level performance is 50%.

The general Random Forest call was:

$$\text{randomForest::randomForest}(x = x,\ y = y)$$

where x is the OTU table and y is a vector of condition or habitat classes.

Given the small datasets, we did not attempt to tune the Random Forest model. We also did not pool samples per site (e.g. per loch, woodland, peatland, or marine station), but we note that doing so via

an occupancy model might improve predictive performance, since sample-to-sample noise would be reduced by such a procedure. However, this would require further testing.

Model-based ordination

In unconstrained ordination, variation in community compositions is summarised in one or more 'latent variables' (i.e. unexplained environmental variables) and visualised in a scatterplot. The positions of the sites in the ordination suggest unmeasured environmental covariates that could explain species compositions if they were included in the model. A recent extension to ordination is *concurrent* ordination (van der Veen et al. 2022), which is a type of constrained ordination in which one's measured environmental covariates are included in the model while acknowledging that there are still one or more *unmeasured* environmental covariates contributing to community compositions. Concurrent ordination estimates the total contribution that the measured covariates make toward explaining variation in community composition and estimates the relative contributions that each measured covariate makes toward the ordering along each of the latent variables.

The general gllvm call for unconstrained ordination was:

```
gllvm::gllvm(y = y, family = binomial(), num.lv = 2, studyDesign = Site, row.eff = ~(1 | Site),
control.start = list(n.init = 10))
```

where *y* was the OTU table and Site was included as a random factor to account for multiple samples per site.

The general gllvm call for constrained ordination was:

```
gllvm(y = y, X = X, family = binomial(), num.lv.c = 2, lv.formula = ~ Condition + Area + Moisture,
studyDesign = Site, row.eff = ~(1 | Site), control.start = list(n.init = 10))
```

where *y* was the OTU table, *X* was the table of predictors, nothing was modelled at full-rank (no formula term), and lv.formula indicated the environmental predictors used to model the latent variables (e.g. Condition, Area, and Moisture for woodlands). Site was included as a random factor to account for multiple samples per site. We chose to extract either one or two (concurrent) latent variables to make plotting interpretable. Optimising a gllvm model for latent variables would have required many days of runtime, which we did not have the budget for. Note that plotting three or more latent variables requires multiple 2- or 3-dimensional plots (depending on the number of latent variables), and in our experience, are very hard to interpret.

The OTU table supplied to the calls was converted to binary detections – that is, read counts were not used in the models, only the presence/absence of OTUs. We confirmed for all models that residual patterns were satisfactory. In our opinion, it is incorrect to use OTU read counts as abundance information, due to differences in species detectabilities. This is discussed in Diana et al. (2022).

3 Results & Discussion

For each habitat, we first present a Sequencing Data Summary showing the number of OTUs detected and the percentage of OTUs assigned at each taxonomic level. We generally targeted 100k reads per sample, although the final number of reads per sample is marker and sequencing run dependent. Read depth of final target reads ranged from 33k to 83k per sample. The reason for not being able to assign an OTU to species level (or other taxonomic levels) is a mix of two primary factors: marker resolution and database completeness. Marker resolution refers to the fact that multiple species may have

identical, or very similar, DNA sequences for the region being amplified by PCR. In these cases, it is not possible to distinguish between those species, even when they are all present in reference databases. This is the primary reason, for example, that the percentage of vertebrate sequences in the freshwater habitat are identifiable to species level for 60% of OTUs. Most UK vertebrates (especially fish) have 12S reference sequences available but not all species can be distinguished. Bird species in particular are poorly distinguished by this marker region. Database incompleteness, the absence of reference sequences for given species, is another challenge and has a greater impact on the lesser studied taxa e.g. invertebrates, bacteria, and protist eukaryotes. When there are no high identity matches for a given OTU, it can only be assigned at higher taxonomic levels.

Following a summary of metabarcoding results for each sample type, we present:

- eDNA-based metrics showing a significant difference between habitat conditions.
- If applicable to the habitat, established indices values derived from best matching available data compared to eDNA-based data.
- An assessment of whether community composition data was able to:
- distinguish between habitat conditions/categories.
- predict the status of unknown samples with adequate confidence.
- identify other potentially influential environmental variables.
- Future sampling recommendations.

Within the Phase 2 Technical Appendices output (Bakker et al. 2023c) we provide:

- Taxonomic heat trees showing an overview of taxa detected.
- Sample-level metrics results (boxplots and model outputs). Results that showed clear and meaningful trends are discussed in the main report.
- Random Forest Classification tables.
- Further information on the assays used.
- Other outputs relevant to particular habitats.

3.1 Marine

The marine habitat specific research questions (MRQs) posed and the key findings (MKFs), are summarised in this section.

3.1.1 Marine Sequencing Data Summary – Water samples

Fish sequence data were obtained from 44 of the 46 eDNA water samples. The final dataset contained a total of 62 OTUs. The average taxon richness was 12 and ranged from 4 to 22. The taxa detected belong to 13 orders, 28 families, and 49 genera (Table 1), from a total of 2,674,886 sequence reads (Table 3). The most abundant species detected by read count were European sprat (*Sprattus sprattus*), Atlantic mackerel (*Scomber scombrus*), Atlantic herring (*Clupea harengus*), and haddock (*Melanogrammus aeglefinus*). None of the negative field controls produced any target reads.

Table 1: Number of OTUs detected and the percentage of OTUs identified at each taxonomic level for fish.

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Fish	62	100%	100%	100%	100%	96.8%	87.1%

3.1.2 Marine Sequencing Data Summary - Sediment samples

Three different DNA metabarcoding assays were applied to DNA extracted from each sediment sample: bacteria (16S), eukaryotes (18S), and invertebrates (COI). A total of 1,488 taxa were detected across the 56 samples (from a total of 4,468,988 sequence reads); 753 bacteria, 575 eukaryotes, and 160 invertebrates (Table 2). More 18S eukaryote and COI invertebrate OTUs were identified at the species level compared to bacteria.

In the bacteria dataset, OTUs were detected across 20 different phyla. The average bacteria taxon richness per sample was 260 and ranged from 197 to 314. In the eukaryote dataset, OTUs were detected across 37 different phyla within the kingdoms Animalia, Chromista, Fungi, Plantae, and Protozoa. The average eukaryote taxon richness per sample was 119 and ranged from 47 to 155. In the invertebrate dataset, OTUs were detected across 15 different phyla within the kingdom Animalia. The average invertebrate taxon richness per sample was 12 and ranged from 1 to 26.

Table 2: Number of OTUs detected and the percentage of OTUs identified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	753	74.6%	58.2%	34.5%	26.2%	6.5%	1.2%
Eukaryotes	575	96.7%	73.9%	67.3.2%	65.7%	30.4%	11.7%
Invertebrates	160	99.4%	91.2%	86.2%	76.9%	60%	41.9%

Table 3: Marine sequencing data summary for water and sediment samples

Assay	N reads obtained	N samples	N samples that did not produce data	Samples that did not produce data
Aquatic Fish	2,674,886	46	2	LL11.1 (Loch Long, SS.SMu.CFiMu.SpnMeg), LL32.2 (Loch Long SS.SMu.CFiMu.MegMax)
Sediment Bacteria	1,847,184	56	0	
Sediment Eukaryotes	1,473,894	56	0	
Sediment Invertebrates	1,147,910	56	0	

3.1.3 Marine Conventional Sampling Data Summary

The PSD results generated Level 4 EUNIS habitat that was consistent with those previously reported (C. R. Allen et al. 2013) in all except two cases:

- Station LL13 was defined as SS.SMx.CMx where previously it had been defined as either SS.SMu.CFiMu or SS.SMu.CSaMu
- Station LL65 was defined as SS.SMu.CFiMu where previously it had been defined as SS.SMx.CMx
- Station LL27 had no PSD data but was classified as SS.SMx.CMx due to the description of stones and shells in the macrofaunal sample

Using morphoanalysis, a total of 289 macrofauna taxa were identified (Table 4), composed of a total of 5114 individuals. The average macrofauna taxon richness per sample was 40 and ranged from 7 to 108

taxa. Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

Table 4: Summary of the number of taxa detected and the percentage of taxa successfully identified at each taxonomic level for macrofauna.

Number of taxa	Phylum	Class	Order	Family	Genus	Species
289	100%	95.2%	85.5%	93.8%	90.0%	73.4%

Detailed assignment information for each station is given separately in the Technical Appendices output (Bakker et al. 2023c), but in most cases the community had shifted slightly since the 2013 survey (C.R. Allen et al. 2013) to similar, but different Level 5 biotopes (Table 5).

Table 5: Marine conventional habitat classification summary.

Station	Loch name	Allen (2013) Level 5 (Imagery)	New Level 4 (PSD)	Allen (2013) Level 5 (Grab)	New eDNA-based Level 5 (Biotope)
LL03	Loch Long	SS.SMx.CMx.ClloMx	SS.SMx.CMx	-	SS.SMx.CMx.ClloMx
LL07	Loch Long	SS.SMu.CFiMu.SpnMeg	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL11	Loch Long	SS.SMu.CFiMu.SpnMeg	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL13	Loch Long	SS.SMx.CMx.ClloMx	SS.SMx.CMx	-	SS.SMx.CMx.KurThyMx
LL14	Loch Long	SS.SMu.CFiMu.SpnMeg	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL16	Loch Striven	SS.SMu.CFiMu.MegMax	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL17	Loch Striven	SS.SMx.CMx.ClloMx	SS.SMx.CMx	-	SS.SSa.CMuSa.AalbNuc
LL25	Kyles of Bute	SS.SMx.CMx.OphMax	SS.SMu.CSaMu	-	SS.SMx.CMx.ClloModHo
LL26	Kyles of Bute	SS.SMx.CMx.OphMax	SS.SMx.CMx	-	SS.SMx.CMx.KurThyMx
LL27	Kyles of Bute	SS.SMx.CMx.ClloMx	SS.SMx.CMx	-	SS.SMx.CMx.OphMx
LL28	Loch Long	SS.SMu.CFiMu.SpnMeg	SS.SMu.CSaMu	SS.SMu.CSaMu.AfilMysAnit	SS.SMu.CSaMu.AfilKurAnit

Station	Loch name	Allen (2013) Level 5 (Imagery)	New Level 4 (PSD)	Allen (2013) Level 5 (Grab)	New eDNA-based Level 5 (Biotope)
LL30	Loch Long	SS.SMu.CFiMu.MegMax	SS.SMu.CFiMu	SS.SMu.CSaMu.AfilMysAnit	SS.SSa.CMuSa.AalbNuc
LL32	Loch Long	SS.SMu.CFiMu.MegMax	SS.SMx.CMx	SS.SMu.CSaMu.LkorPpel	SS.SMx.CMx.KurThyMx
LL33	Loch Long	SS.SMu.CFiMu.MegMax	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL36	Loch Long	SS.SMx.CMx.OphMax	SS.SMx.CMx	-	SS.SMx.CMx.KurThyMx
LL43	Gare Loch	SS.SMx.CMx.ClloMx	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL46	Holy Loch	SS.SMu.CFiMu.SpnMeg	SS.SMu.CFiMu	SS.SMu.CSaMu.AfilMysAnit	SS.SSa.CMuSa.AalbNuc
LL50	Loch Long	SS.SMu.CFiMu.MegMax	SS.SMu.CFiMu	-	SS.SSa.CMuSa.AalbNuc
LL54	Loch Goil	SS.SMx.CMx.OphMax	SS.SMx.CMx	-	SS.SMx.CMx.KurThyMx
LL65	Loch Long	SS.SMx.CMx.ClloMx	SS.SMu.CFiMu	-	-

The new Level 5 biotopes were:

- *Abra alba* and *Nucula nitidosa* in circalittoral muddy sand or slightly mixed sediment (SS.SSa.CMuSa.AalbNuc)
- *Kurtiella bidentata* and *Thyasira* spp. in circalittoral muddy mixed sediment (SS.SMx.CMx.KurThyMx)
- Sparse *Modiolus modiolus*, dense *Cerianthus lloydii* and burrowing holothurians on sheltered circalittoral stones and mixed sediment (SS.SMx.CMx.ClloModHo)
- *Ophiothrix fragilis* and/or *Ophiocomina nigra* brittlestar beds on sublittoral mixed sediment (SS.SMx.CMx.OphMx)
- *Amphiura filiformis*, *Kurtiella bidentata* and *Abra nitida* in circalittoral sandy mud (SS.SMu.CSaMu.AfilKurAnit)
- *Cerianthus lloydii* and other burrowing anemones in circalittoral muddy mixed sediment (SS.SMx.CMx.ClloMx)

3.1.4 Marine Research Questions & Key Findings

3.1.4.1 Summary of Marine Habitat Key Findings

Biodiversity monitoring through eDNA sampling is a useful tool for detecting PMF species across multiple taxonomic groups (including fish, marine mammals, and invertebrates), and can classify

between biotopes, although limitations exist. The number of species-level identifications for sediment invertebrates was lower than that of the morphological analyses. Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

We found that although broadly different biotopes (i.e. different substrate types) can be classified by metabarcoding community data, higher-level biotopes were not well-defined by ecological community eDNA-based datasets in this project. This is likely due to different communities being sampled by the original morphological analysis and videography methods (mostly surface and larger invertebrates) that are used to originally describe the biotopes, versus the ones sampled by eDNA-based methods (smaller species within the sediment).

We also show that AMBI scores derived from eDNA-based datasets are comparable to those calculated from morphological surveys, providing a platform to scale-up pollution monitoring of marine habitats.

3.1.4.2 Research Questions & Key Findings

MRQ1: Are there any DNA-based units/metrics that can act as indicators of biodiversity and be used to (a) distinguish habitat condition (using biotope as a proxy for marine habitat condition) and (b) detect ecological pressures or restoration stage in Scottish marine loch habitats?

To address this question, we investigated whether there were clear trends evident across biotopes for any of the following:

- species richness (fish, invertebrates, eukaryotes, bacteria)
- evolutionary diversity (fish, invertebrates, eukaryotes, bacteria), and
- functional diversity (bacteria).

Community-based statistics, species of interest results, and AMBI results are detailed under the MRQs that follow below.

MKF1a: Species richness and evolutionary diversity are significantly different across biotopes for bacteria and invertebrates from the marine sediment samples. There may be future potential to use these metrics to monitor ecological changes. However, as there was no clear pressure gradient across the locations available in the study (see later MRQs), and there is currently no framework for placing these values in terms of what a biotope should possess, this requires further research.

MKF1a.1: Species richness of bacteria ($p < 0.001$) and invertebrates ($p = 0.04$) from sediment samples were significantly different between the six new morphological biotopes. When considering Level 4 biotopes, using the PSD analysis, bacteria ($p = 0.003$) and invertebrate ($p = 0.007$) species richness were also significantly different.

Evolutionary diversity of bacteria and invertebrates were significantly different among the six new morphological biotopes ($p < 0.001$) and the level-four PSD biotopes ($p = 0.01$).

The sampling design did not allow for the testing for ecological pressures or restoration stages in Scottish marine loch habitats, as the marine project focus was centred around biotoping and PMF detection. However, the results indicate that as biotopes change, these values could be used to track such shifts. As there is limited interpretation that can be provided on these results, all figures and model results are provided separately in the Technical Appendices output (Bakker et al. 2023c).

MRQ2: Can eDNA metabarcoding community data from sediment be used to distinguish (classify or sort) between different biotopes or other units/metrics used to classify Scottish marine loch habitats, using Loch Goil/Long as an example?

MKF2: *Both supervised classification and unconstrained ordination found that eDNA metabarcoding data for bacteria and eukaryotes distinguishes between muddy versus sandy/gravelly biotopes but not among sandy/gravelly biotopes.*

MKF2.1: We found that it was possible to classify Level 4 biotopes successfully using supervised classification and cross-validation tests with all four community datasets. We tested if eDNA data could predict two new habitat categorisations: the Level 5 biotope (new.morpho.biotope) based on the combined morphological analysis and PSD from this survey campaign and the Level 4 habitat based on only the PSD data (Level.4.PSD) Morphological and PSD data is not included in this report but can be requested from SEPA for different purposes. Testing against both of these, classifications allowed us to consider the more accurate contemporary sediment classification compared to the 2013/2014 surveys. Furthermore the Level.4.PSD was not always consistent with the new.morpho.biotope. We constructed these two new biotope response variables because pre-analysis found that eDNA community structures did not predict the 2013/14 biotope categories. It is not clear whether this mismatch is because eDNA data are unable to detect biotope differences, or because the morphological and DNA-based methods select for different combinations/communities of invertebrate species. It may also be that the station biotopes themselves have changed since 2013/14.

There are two new.morpho.biotope classes represented by more than one station; this means that random classification would achieve a predictive accuracy of ~50%. The Random Forest cross-validation test shows that across all four assays, the SS.SSa.CMuSa.AalbNuc biotope can be identified with >80% predictive accuracy, whereas predictive accuracy to the SS.SMx.CMx.KurThyMx biotope is much worse than random (see Phase 2 Technical Appendices output; Bakker et al. 2023c).

There are three Level.4.PSD biotopes represented by more than one station, so random classification would achieve a predictive accuracy of ~33%. The Random Forest cross-validation test shows that across all four assays, the SS.SMu.CFiMu biotope can be identified with 73-90% accuracy whereas predictive accuracies to the other two biotopes are worse than random. The SS.SSa.CMuSa.AalbNuc (new.morpho.biotope) and SS.SMu.CFiMu (Level.4.PSD) biotopes both indicate a muddy substrate, so supervised classification with sediment or fish eDNA can separate muddy from gravelly/sandy substrates but cannot differentiate amongst gravel/sandy substrates. This may be due to finer sediments having more sorted sediments, hence potentially a greater homogeneity of habitat conditions for biological communities than coarser and more mixed sediments.

MKF2.2: Ecological community compositions form clusters within biotopes, particularly for bacteria and eukaryote datasets, as is shown through unconstrained ordination (Figure 5 and Figure 6). For the concurrent ordination, we included water depth and temperature as potentially influential covariates, alongside condition. With sediment bacteria and eukaryotes, water depth and temperature had no significant effects on composition, but both kinds of biotope (new.morpho.biotope and Level.4.PSD) had strongly significant effects on composition. For fish and sediment invertebrates, there was an effect of water depth (see Phase 2 Technical Appendices output; Bakker et al. 2023c).

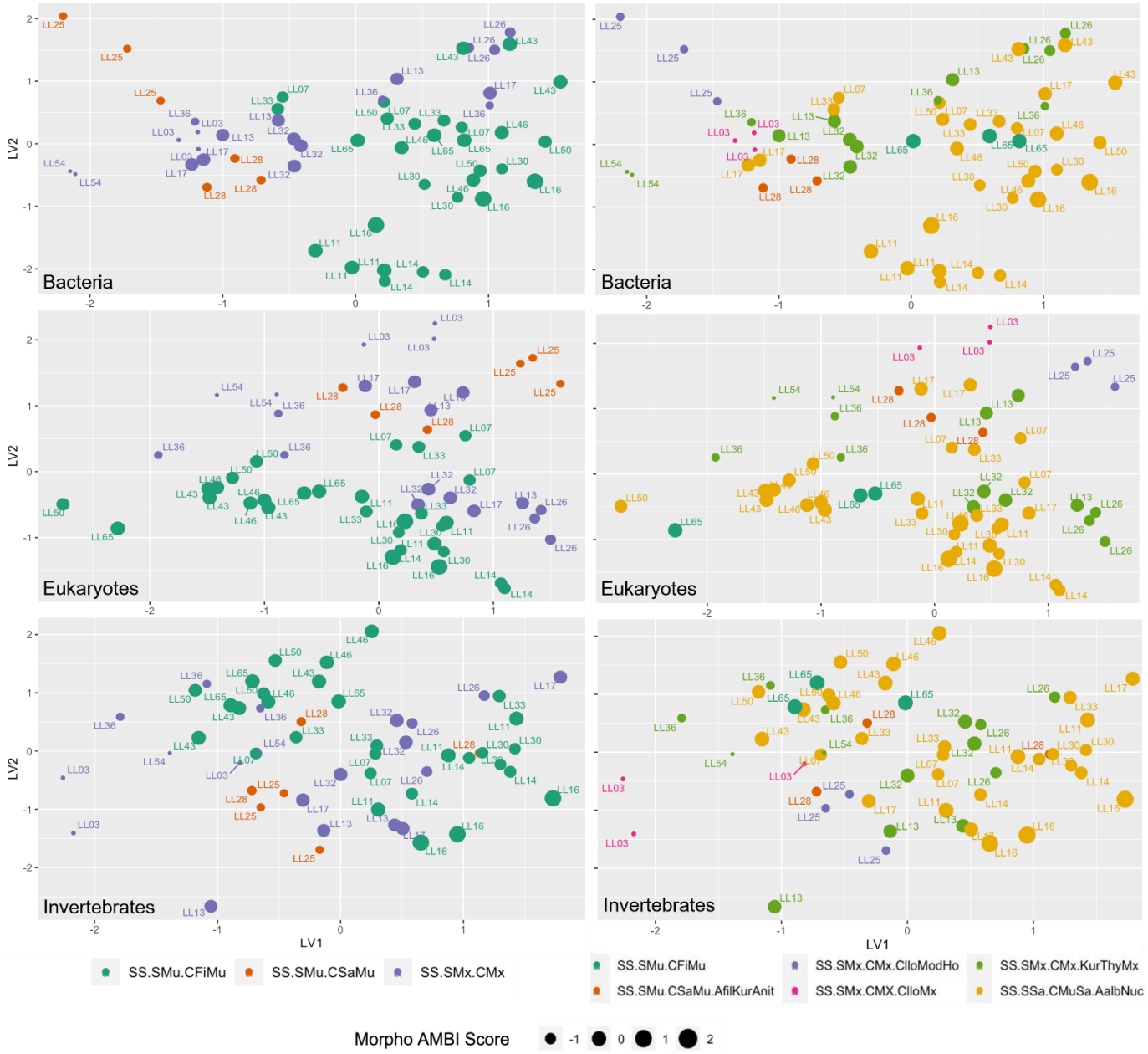


Figure 5: Unconstrained ordination for marine sediment bacteria, eukaryotes and invertebrates. Each point is a sample, and samples that are closer together are more similar. Plots on the left are coloured by Level 4 Biotopes based on PSD and plots on the right by Level 5 Biotopes based on PSDA and morphoanalysis. Community compositions, particularly for the bacteria and eukaryotes datasets, were differentiated between mud biotopes and other biotopes, but not between those other biotopes.

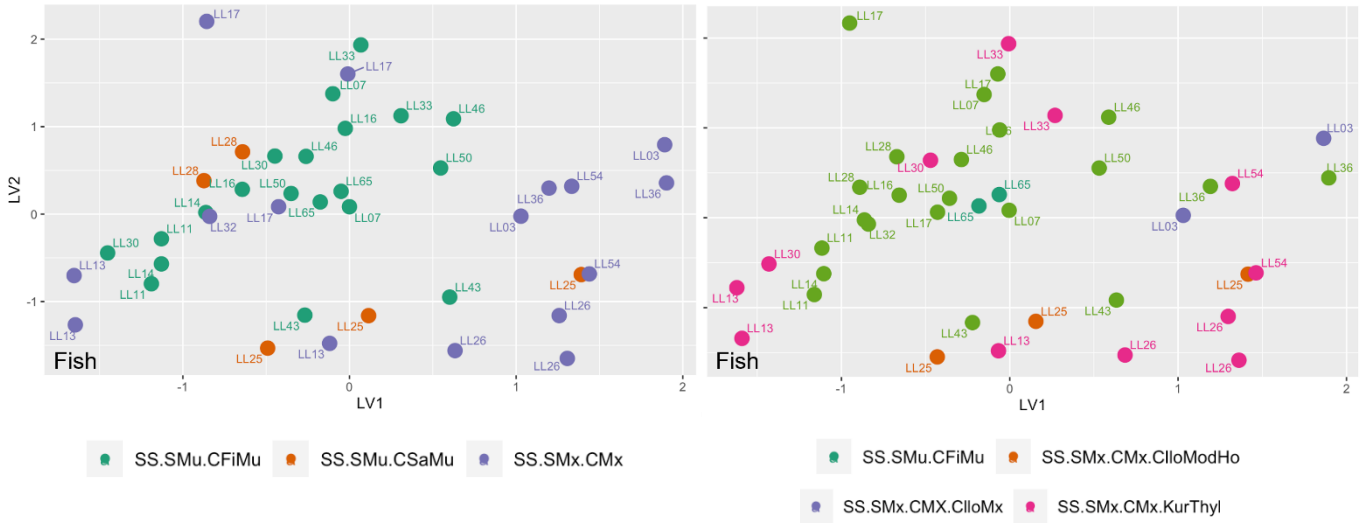


Figure 6: Unconstrained ordination for marine fish. Each point is a sample, and samples that are closer together are more similar. The plot on the left is coloured by Level 4 Biotopes and the plot on the right by Level 5 Biotopes.

MRQ3: Can eDNA metabarcoding community data provide similar and reliable outcomes as conventional invertebrate morphology-based or scoring-based methods (e.g., AMBI and/or infaunal quality index IQI) to assess the ecological status of Scottish marine loch habitats, using best available matching data from Loch Goil/Long as an example?

MKF3: *The AMBI categories from the morphological and eDNA datasets from the same samples agreed in 78% of the samples (14 out of 18 samples).*

The AMBI scores of all samples based on morphological data were categorised as good, which indicated no gradient in pollution for these locations (see Phase 2 Technical Appendices output; Bakker et al. 2023c) for scoring results; Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose). The AMBI and gAMBI categories agreed in 78% of the samples (14 of 18 samples), as one paired grab for eDNA was empty (LL54.3). Two of the four samples that did not agree were assigned adjacent categories. The other two samples that did not agree had gAMBI categorised as bad or poor and the AMBI categorised as good (LL14.3 and LL32.3). LL32.3 differed because no species were matched with the gAMBI database for gAMBI (only 7 invertebrate OTUs identified) compared to 30 of 31 species assigned an AMBI score for the morphological dataset. The other sample, LL14.3, had 4 of 11 invertebrate OTUs assigned to an gAMBI score compared to 9 species identified and assigned AMBI scores for the morphological dataset which differed in the species and their respective assignments. The pattern of morphological data matching more species to AMBI scores compared to gAMBI was evident overall. The AMBI matched a mean of 38.0 taxa per sample with 3.5% of taxa unassigned, while gAMBI matched a mean of 12.4 taxa per sample with 47% of invertebrate taxa unassigned. It is promising that eDNA was still comparable considering the AMBI database was built on morphologically identified species and eDNA matches could be improved as taxa identified by eDNA are added. Combining multiple assays may also lead to increased matching of taxa with gAMBI. Here we have only applied the gAMBI method to the results of the invertebrate assay, and not those of the eukaryote assay, because of the analysis pipeline available to us at the time of data analyses.

In general, there was good agreement among the three replicates for the gAMBI categorisation. For 17 stations with AMBI scores for all 3 replicates; 8 stations had perfect agreement among replicates and 6 sites had 2 replicates that were the same category. The agreement between AMBI and gAMBI was not improved if the mean gAMBI score of the three eDNA samples from replicate grabs was used in the comparison, although the mean did reduce differences. For example, taking the mean gAMBI score of three eDNA replicates compared to the single replicate changed 2 samples from poor or bad status to moderate when the morphological data categorised them as moderate and the mean eDNA replicate changed 1 sample from high to good when the morphological data categorised it as good. However, there were also examples that using the mean of eDNA based score made existing differences larger. For example, the mean replicate eDNA-based AMBI score changed 2 samples to poor or moderate that were good based on the single replicate when the morphological data categorised them as good. This suggests replicates may improve data robustness. It is also recommended that replicate cores are collected for morphological AMBI given the variability among replicate cores (Aylagas et al. 2018).

The majority of AMBI scores suggested "unpolluted" (0 to 1.2) and "slightly polluted habitat" (1.3 to 3.3), with the previously designated SS.SMu.CFiMu.MegMax appearing significantly more polluted ("slightly polluted" and "moderately polluted" (3.4 to 5) (Figure 7). The finer sediment is indicative of a lower energy environment, which may therefore be more prone to the retention of pollutants.

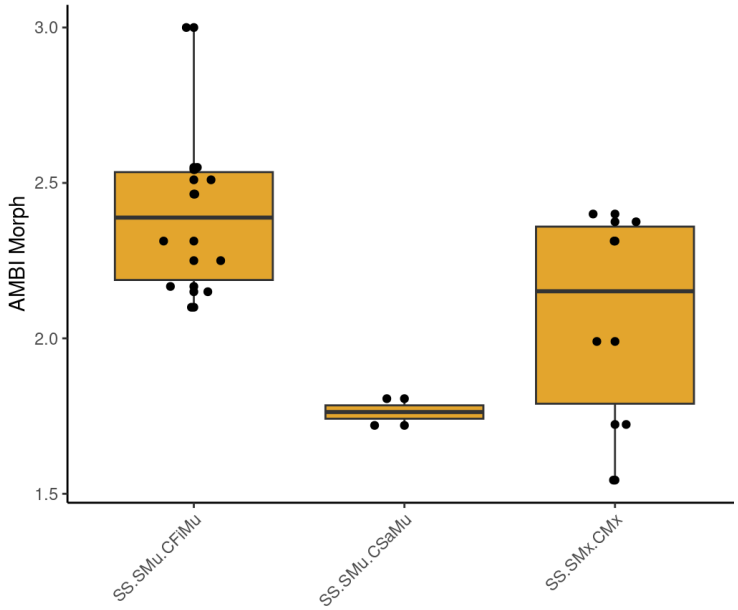


Figure 7: There is a significant effect (p = 0.02) of Level 4 PSD biotope on the morphological invertebrate AMBI score. Boxplots show the distribution of data for each biotope, including the median (mid-line in box) and the lower and upper quartiles (limits of orange box) representing the point at which 75% and 25% of the data falls below. The whiskers represent the limits of data points that fall outside of the interquartile range (the values between the upper and lower quartiles.

Note that we also applied a marine benthic bacteria index, designed specifically for aquaculture monitoring (Frühe et al. 2021). As we did not have an aquaculture gradient in this project these results should be interpreted with caution, and are provided in the Appendices.

MRQ4: Can the DNA-derived metrics extend from taxonomic descriptions to assess possible pressure or restoration-related changes in Scottish marine loch habitats (e.g., using gAMBI and/or bacteria-derived metrics)?

MKF4: *The AMBI and gAMBI categories, which give an indication of pollution levels, from the morphological and eDNA-based datasets from the same samples, agreed in 78%* (MKF3) *of the samples.*

Accordingly, the application of gAMBI should be investigated further, at sites with a greater pollution gradient as a metric, to assess possible pressure or restoration-related changes in Scottish marine loch habitats, based on eDNA-based data. The sampling design did not allow for further testing of any other ecological pressures or restoration stages in Scottish marine loch habitats.

MRQ5: Are there any differences or similarities in biotopes or biodiversity and community composition from any re-surveyed locations between timescales (e.g., seasons or years) and protected status (e.g., designated vs non-designated sites) for the Scottish marine loch habitats sampled?

MKF5: *a) There were some differences in re-surveyed locations between years. b) There were no differences in designated vs non-designated status between sampling stations.*

In MKF2 we discuss the reclassification of biotopes, and the identification of new biotopes compared to the 2013/2014 biotope assignments. The changes in biotope classification based on the contemporary dataset may be due to the different methodological approach applied in this survey campaign (macrofauna morphological analysis combined with PSD) compared to the 2013/2014 survey campaign (benthic imagery, with macrofauna data and PSD at four of our selected stations). Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

MRQ6: Can the eDNA metabarcoding primers used deliver detailed and species-specific data for consistent detection of Priority Marine Features (PMF) (e.g., invertebrates and fish) across multiple sites?

MKF6: *The primers used can identify the presence of invertebrate and fish species indicators including some PMF species, although the taxonomic resolution with morphological analysis of marine invertebrates was greater overall.*

MKF6.1: The two PMF invertebrate species detected through eDNA analysis were *Arctica islandica* and *Modiolus modiolus.* Both were also detected in the morphological dataset (Table 6). Only one additional invertebrate PMF species (*Maera loveni*) was detected in the morphological dataset and not in the eDNA-based dataset. Several reasons, or a combination of these, may explain why only very few invertebrate PMF species were detected:

- No PMF species were present in most of the sampling stations
- No PMF species were captured in the sediment samples
- PMF species were present in the samples, and their DNA was present but below detection levels
- PMF species were present in the sample, and were amplified, but due low marker resolution and/or to an incomplete reference database (see section 3), could not be assigned to species level. For further details, please see the Gapfinder output in the Phase 2 Technical Appendices output (Bakker et al. 2023c).

Table 6: Invertebrate PMF species detected in the sediment samples, for each station, by metabarcoding assays for invertebrates and eukaryotes, and morphological macrofauna analysis.

Station		LL03	LL13	LL17	LL25	LL27	LL28	LL30	LL32	LL36	LL46	LL54
Analysis	Species											
Invertebrates (eDNA, COI)	*Arctica islandica*	■	■	■	■				■	■		
	Modiolus modiolus			■	■							
Eukaryote (eDNA, 18S)	*Modiolus modiolus*			■	■					■		■
Macrofauna (Morphological)	*Arctica islandica*	■			■	■	■	■				
	Modiolus modiolus				■	■				■		
	Maera loveni										■	

MKF6.2: Fish and marine mammals that are designated PMF species and/or IUCN listed species were also detected (Table 7). Some of these are endangered or threatened, and/or commercially important and some of them are actively surveyed to monitor their populations and assess their conservation status. The commercially important species [6] included Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), and Atlantic horse mackerel (*Trachurus trachurus*). Conservation priority species included the critically endangered European eel (*Anguilla anguilla*), and the iconic threatened[7] harbour porpoise (*Phocoena phocoena*) which populations are being monitored (L. D. Williamson et al. 2022).

[6] https://www.gov.scot/publications/scotlands-marine-atlas-information-national-marine-plan/pages/32/

[7] https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/biodiversity-status/marine-mammals/harbour-porpoise-bycatch/#:~:text=Harbour%20porpoise%20(Phocoena%20phocoena)%20is,and%20drowning%20in%20fishing%20nets

Table 7: All marine species of note detected in the DNA-based data. Indicated is whether a species has PMF designation and/or IUCN status, and for some detected species specialist input is provided.

Species	Common name	PMF	IUCN status	Specialist input
Anguilla anguilla	European eel	Yes	CR	Conservation significance due to huge decline across Europe resulting in its IUCN CR status
Gadus morhua	Atlantic cod	Yes	VU	
Melanogrammus aeglefinus	Haddock		VU	
Trachururs trachurus	Atlantic horse mackerel	Yes	VU	
Clupea harengus	Atlantic herring	Yes		
Merlangius merlangus	Whiting	Yes		
Pollachius virens	Saithe	Yes		
Trisopterus esmarkii	Norway pout	Yes		
Pomatoschistus minutus	Sand goby	Yes		
Scomber scombrus	Atlantic mackerel	Yes		
Salmo salar	Atlantic salmon	Yes		
Salmo trutta	Trout	Yes		
Atherina boyeri	Big-scale sand smelt			Perhaps surprising to find as distribution recorded as much more southerly than Scotland, as far north as south England and The Netherlands. Detected at only one station and in only one sample, possible false positive.
Phoxinus phoxinus	Minnow			Freshwater species. Surprising to see in marine environments no known association even with brackish conditions.
Rutilus rutilus	Common roach			Detected in one sample, 511 reads. Freshwater/brackish water species – not native to most of Scotland. Complex history of natural and artificial introductions. Debatable native status in southwest Scotland (Maitland), introduced elsewhere? Surprising in marine environment, possibility of eDNA being transported from freshwater sources. Present in Loch Lomond South, River Leven and River Clyde (NBN Atlas). Sampling location 1.5 km from small streams connected to Lindowan Reservoir (although presence in those areas not known).
Hippoglossoides platessoides	American plaice		EN	Native. IUCN classifies it as Endangered due to over-fishing in the west Atlantic
Phocoena phocoena	Harbour porpoise	Yes		
Arctica islandica	Ocean quahog	Yes		
Modiolus modiolus	Horse mussel	Yes		

MKF6.3: When comparing marine sediment invertebrate taxa detection between morphological and eDNA-based data from the same samples, the taxonomic resolution from morphological analysis outperformed the eDNA-based data on all taxonomic levels (Table 8), and particularly on the species level (Figure 8). This is not surprising as with the eukaryote assay, only 11.7% out of 575 OTUs have been identified down to species level. For the invertebrate assay, this was 41.9% out of 160 OTUs (Table 8),

whereas the morphological method detected 289 taxa, of which 72% were identified to species level. Even when both the invertebrate and the eukaryote assays are combined, the overlap with morphology only increases with 2% on species level (Figure 9 and Figure 10). Additional figures showing overlap between the phylum level can be found in the Phase 2 Technical Appendix output (Bakker et al. 2023c). Metabarcoding of marine sediment invertebrate data clearly works very well, however, the incompleteness of the reference databases still causes the method to be inefficient for species level detection compared to morphological analysis. Morphological and PSD data is not included in this report but can be requested from SEPA for a different purpose.

Table 8: Marine sediment invertebrate taxa detected; agreements between morphological and eDNA analyses on taxonomic levels from species to phylum. % Agreement = number of species detected by both methods in either the overall dataset or at a given station divided by the number of species detected in the dataset/station overall times 100.

Taxonomic Level	% Agreement (overall dataset)	% Agreement (by station)
Species	8.10	3.32
Genus	12.50	4.63
Family	20.70	9.58
Order	29.90	20.44
Class	48.60	23.21
Phylum	55.00	37.50

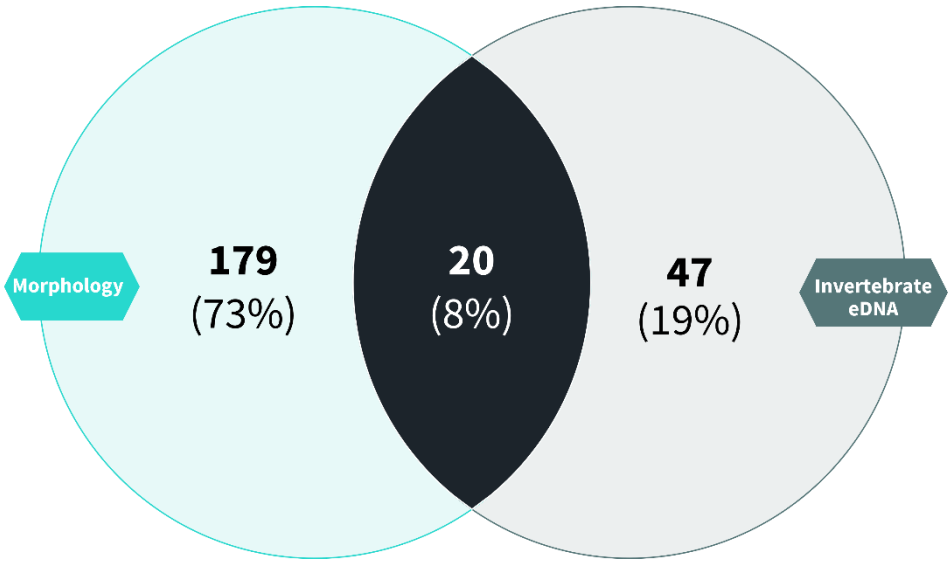


Figure 8: Venn diagram showing the level of overlap in the overall dataset between species level detection of marine sediment invertebrate species, between morphology and the invertebrate eDNA assay.

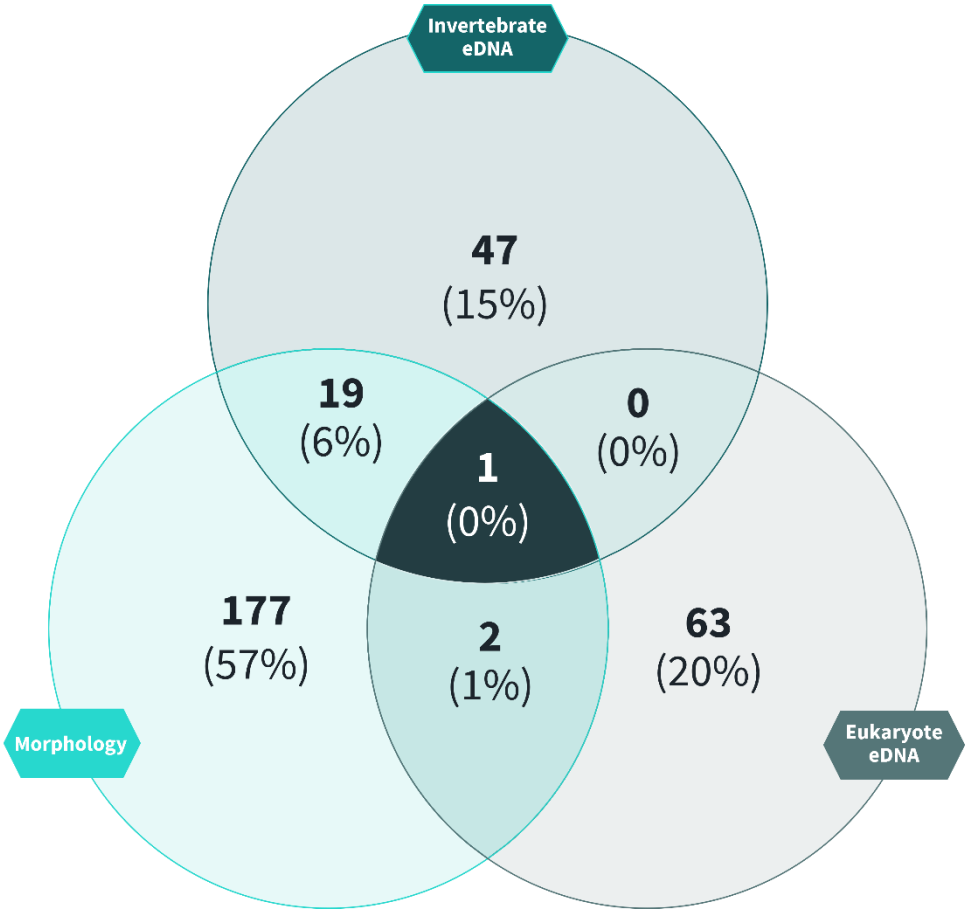


Figure 9: Venn diagram showing the level of overlap in the overall dataset between species level detection of marine sediment invertebrate species, between morphology and the invertebrate and eukaryote eDNA assays.

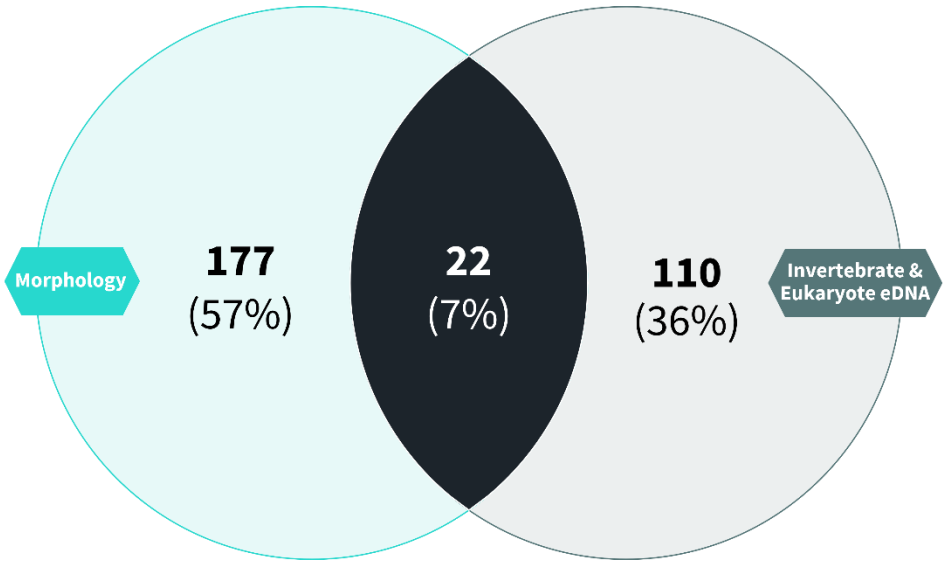


Figure 10: Venn diagram showing the level of overlap in the overall dataset between species level detection of marine sediment invertebrate species, between morphology and the invertebrate and eukaryote eDNA assays combined.

MRQ7: What is the variability between samples collected at the same station for both water and sediment samples and can we statistically predict the potential impact of primer bias and subsampling of sediment samples on the overall outcome?

MKF7.1: Sample replicates are required to provide statistical power for models that perform predictive mapping, interpolate and extrapolate species richness, and estimate species detection probability or likelihood of false positives (i.e. species detected that is not actually present) and false negatives (i.e. species not detected that is actually present). We are working on an analysis (Occupancy+) to estimate the number of sample replicates needed to detect a given percentage of all the species in a sample. Some initial work on sample replicates using occupancy modelling has already been performed (Figure 11; also see Phase 2 Technical Appendices output (Bakker et al. 2023c)).

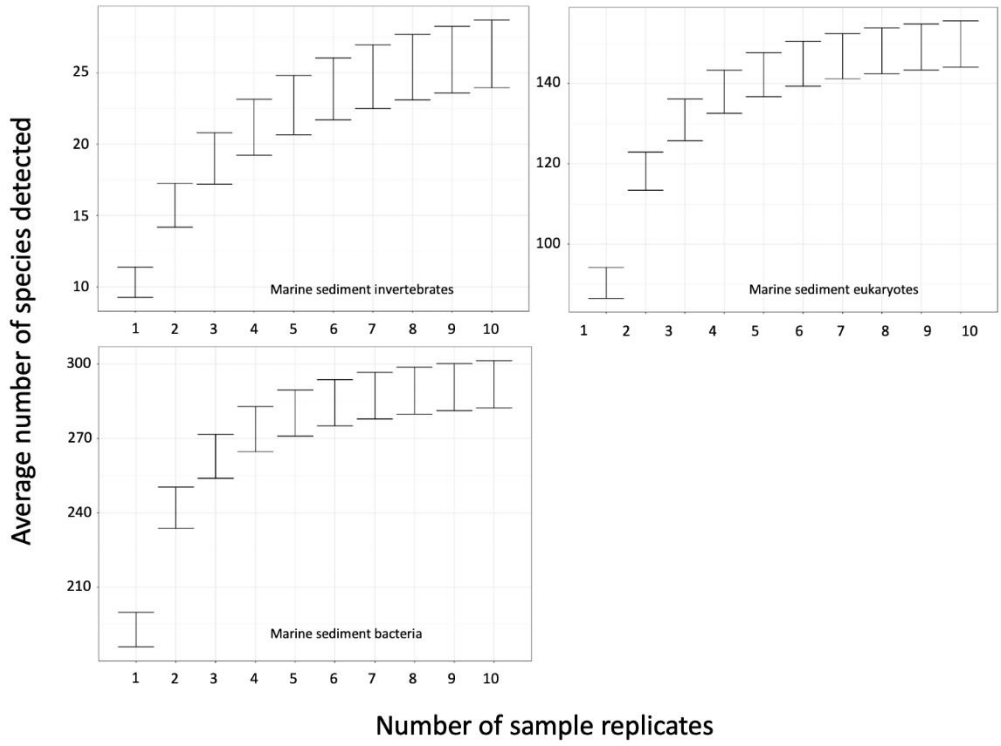


Figure 11: Occupancy modelling to estimate probability distributions of detection for marine sediment invertebrates, eukaryotes, and bacteria for 1 to 10 sample replicates. Indicating that increasing the number of biological replicates, increases the number of species detected.

MKF7.2: Analysis of primer bias requires complex laboratory and statistical analyses outlined in Williamson et al. (2021), which combine individual qPCRs with metabarcoding and a spike-in. We did not have capacity to carry out this analysis.

MRQ8: What are the key recommendations and future perspectives to develop and operationalise standardised eDNA-based methods at national scale for monitoring biodiversity and the impacts of pressures or restoration in Scottish marine loch habitats? Each question must be responded to in turn:

MKF8.1: How should we sample?

Sampling in the marine environment is most often conducted from a vessel, but samples may also be collected from the shoreline. Marine water sampling is most effectively done using a Kemmerer/Niskin

type sampler, which can be combined with a CTD. These samplers can also be integrated into a mechanical sampling system on larger research vessels such as a rosette. When using the Kemmerer/Niskin sampler to collect water near the bottom, it is important to not disturb the sediment as sediment present in water samples can clog filters and may contain 'old' eDNA that can skew inferences of species' presence (Goldberg et al. 2016). Thorough flushing of the bottle sampler with seawater at the (new) sampling site (by moving it through the water column) is often sufficient to reduce carryover between stations/sites. To further decrease carryover risk, samplers can be beach cleaned between sampling stations, but this also requires additional measures to avoid environmental contamination. Field negative controls (blanks) should always be included to track potential contamination.

It is possible to use a manual syringe for water filtration, but peristaltic or vacuum pumps are recommended and more commonly used for larger volume samples, which are required in the marine environment (Cowart, Murphy, and Cheng 2022).

MKF8.2: What should we sample?
Environmental DNA in marine systems is generally much more dilute compared to that in freshwater systems (Bruce et al. 2021; Harrison, Sunday, and Rogers 2019). This is in part dependent on the target group or species but is particularly pertinent for larger vertebrates (extra-organismal vertebrate eDNA) (K. J. Harper et al. 2020; Suarez-Bregua et al. 2022). Planktonic or microbial taxa usually require smaller volumes (Bruce et al. 2021). Therefore, sample volume should be maximised to be representative of the environment and the taxa that are targeted (Stauffer et al. 2021). Each sample should be at least 2 L volume (Bessey et al. 2020), with 5 L being typical although larger volumes may yield more data (Bruce et al. 2021; Valsecchi et al. 2021). Turbidity is usually less of a problem in marine water compared to freshwater, although inshore areas (e.g. mangrove forests, marinas, areas with high population density) can become turbid due to coastal run-off and wave action disturbing the sea floor (Bruce et al. 2021; Hallam et al. 2021).

Sample number will depend on the spatial scale of the study or monitoring project (Bruce et al. 2021). In order to characterise a community or compare sites, a minimum of 20 samples, even for relatively small areas, is strongly advised (Stauffer et al. 2021). This usually involves collecting independent samples (rather than subsamples from one and the same sample) spread out across the sampling area (Bruce et al. 2021).

MKF8.3: Where should we sample?
In marine waterbodies, we have limited understanding of how hydrological conditions affect eDNA transportation and distribution. However, it has been shown that communities obtained from marine eDNA metabarcoding are highly representative of the immediate local habitat where the sample was collected, both on horizontal and vertical planes (Ely et al. 2021; Jeunen et al. 2020; Larson et al. 2022; Monuki, Barber, and Gold 2021; Yamamoto et al. 2017; Guri et al. 2023), and across short timescales (Ely et al. 2021; Jensen et al. 2022; Murakami et al. 2019). This means that when samples are collected in a transect going from shore to offshore, different communities will be detected, sometimes even within the range of tens of metres (Allan et al. 2021; Jeunen et al. 2020; Monuki, Barber, and Gold 2021; Port et al. 2016). As in lakes, vertical stratification of the water column (as a result of thermoclines) restricts mixing of eDNA, meaning that water samples should be collected from each depth zone of interest to fully characterise the marine communities at the sampling location (Allan et al. 2021; Bruce et al. 2021; Canals et al. 2021; Jeunen et al. 2020; Monuki, Barber, and Gold 2021).

MKF8.4: When should we sample?

Sampling design will be dependent on the size of the sampling area, and also on the type of ecosystem (reef, pelagic, etc.) that is targeted. Season (and even time of day) may need to be taken into consideration. Many fish species move to (in)shore areas for mating and spawning and move to deeper or warmer waters in winter, while other species may prefer cold, deep water during summer (Sigsgaard et al. 2017). Furthermore, some fish and other marine vertebrate and invertebrate species exhibit diel migration where they move to a different position in the water column during the day or night which is reflected by eDNA analysis (Canals et al. 2021; Easson et al. 2020; Jensen et al. 2022). Thus, it is important to consider migration patterns as well as mating, spawning, and pupping sites.

MKF8.5: Where do the knowledge gaps and barriers exist, and why is that? What practical solutions and technology innovations would help overcome these types of challenges going forward?
The designation of PMF species in Scotland is important for their conservation and management. It provides legal protection and helps to guide marine spatial planning. Consequently, their detection and monitoring are essential, and the application of eDNA-based methods instead of the labour-intensive method of visual morphological analysis, would be a significant step forward in the upscaling and simplification of the processes involved, regarding consistency, efforts, cost, and time, especially when aiming to monitor species across the tree of life. This is particularly pertinent for the marine sediment invertebrate PMF species. However, when comparing marine sediment invertebrate taxa detection between morphological and eDNA-based data from the same samples, morphological analysis outperforms the eDNA-based data in taxonomic resolution. Thus, improving marine invertebrate metabarcoding assays and the barcoding of Scottish marine invertebrate (PMF) species and subsequent deposition of their sequences into publicly available databases is imperative.

Variable species detectabilities across marine sediment types (i.e. differing probabilities of detection when the species is truly present) might result in artefactual spatial patterns. Traditional biotopes might have reflected the variables that drive species detectabilities, not the variables that drive species distributions. With marine sediments, we could imagine that sediment particle-size distribution could affect species detectabilities, for both eDNA-based and conventional methods. For example, soft-bodied organisms on hard substrates may be damaged during sampling, leading to them not being included in morphological analysis. Testing for such effects will require a multi-species occupancy model analysis paired with a suitable study design and candidate detection covariates.

Even if detectabilities do not vary across species, traditional biotope classifications might not capture the true drivers of sediment community structure. Important environmental covariates might remain unknown and unmeasured, and/or a combination of dispersal limitation, environmental nonstationary ('constantly changing environments'), and priority effects (including competitive exclusion) might prevent sediment biota from settling into predictable communities. This raises the possibility that the eDNA surveys are revealing community structure that has remained unseen until now, resulting in different community clustering patterns (not clustering into the traditional biotopes). The statistical toolkit to disentangle these explanations is rapidly improving but tends to be data hungry. Since it is probably not the Scottish Government's priority to carry out an expensive basic-science study on marine sediment metacommunity structure, we suggest the following priorities:

Reanalyse the dataset after occupancy model correction. There are 2 or 3 sample replicates per station, which can be used to estimate detectability of at least some species.

Focus future sampling effort on marine stations that describe independently known gradients of anthropogenic impact, with direct measures of impact, such as pollutant concentrations. These gradients can then be used to define new indicator communities.

eDNA analysis detected differences between muddy versus sandy/gravelly marine sediment biotopes but did not differentiate amongst different sandy/gravelly biotopes. One possibility is that conventional, morphology-based sandy/gravelly biotope definitions are themselves based on inadequate sampling, given the high cost of ship time. Another is that eDNA is more easily dispersed across sandy/gravelly substrates, obscuring the pattern. Thus, for marine sediment monitoring, our recommendation is to focus effort on optimising survey design, by determining the number of sample replicates per station needed to produce a robust biodiversity measurement, comparing conventional with eDNA-based results. One potential advantage of eDNA-based monitoring is that multiple sample replicates could be achieved from a single grab, as opposed to separate individual grabs, while conventional morphology does require multiple grabs, which greatly increases survey costs. We have started this process by estimating species detectabilities for the three eDNA assays (bacteria, invertebrates, and eukaryotes) and showed that the rate of species accumulation begins to asymptote after three grab sample replicates. This is consistent with a similar project regarding replication within and between grab samples around oil and gas infrastructure in Norway (Hestetun, Lanzén, and Dahlgren 2021; Hestetun et al. 2020). See also MKF7.1 and the Phase 2 Technical Appendices output (Bakker et al. 2023c).

MKF8.6: What have we (a) learned to do the same or differently and (b) anticipated are key opportunities to develop, upscale, and implement DNA-based methods for marine loch habitats in Scotland?

MKF8.6a:
The analysis of eDNA from water samples can provide useful information regarding the distribution of vertebrate PMF species, which can be used to inform management and targeted monitoring. The detection of PMF biotopes and species is less well implemented through the eDNA metabarcoding of invertebrates from sediment samples. Clustering of sediment invertebrate communities detected by eDNA metabarcoding appears consistent with PSD data, which was not the case for the morphological macrofaunal analysis. Marine sediment bacteria and eukaryotes show the clearest grouping based on the macrofaunal and PSD data and should be considered for characterising new biotopes.

MKF8.6b:
eDNA-based methods for assessing pollution were consistent with morphological analysis, although this was not tested over a broad pollution gradient as that was not within the project scope. To expand the scope, stations could be sampled along a pollution gradient over relatively homogenous habitat for further testing and expanding of the gAMBI method for tracking pollution in Scottish marine lochs and beyond to coastal and pelagic habitats.

MKF8.7: What other marine habitat types across Scotland do we need to consider sampling for eDNA-based biodiversity monitoring and reporting purposes in the future?
Other PMF habitat types can be considered in the future. However, as discussed above, the differences between conventional and eDNA-based methods have led to low concordance between the datasets. This may be partly due to the high habitat homogeneity in the Loch Long region. We would therefore recommend sampling PMF habitat types more widely and evenly distributed in a given area for clearer methods comparisons and calibration.

3.2 Freshwater

The freshwater habitat specific research questions (FRQs) posed and the key findings (FKFs), are summarised in this section.

3.2.1 Freshwater Sequencing Data Summary

Three different eDNA metabarcoding assays were applied to eDNA extracted from each water filter sample; vertebrates, freshwater invertebrates, and bacteria. A total of 1033 taxa were detected across the 90 samples; 60 vertebrates, 1111 invertebrates, and 862 bacteria from a total of 15,816,361 sequence reads (Table 9, Table 10).

In the vertebrate dataset, taxa detected belong to 20 orders, 38 families, and 44 genera (from a total of 4,254,276 vertebrate sequence reads. The average taxon richness was 9 and ranged from 3 to 23. Eurasian minnow (*Phoxinus phoxinus*), which accounted for 43.3% of the total target sequence reads, was among the most abundant in terms of sequences. Among the most commonly detected target taxa were European eel (*Anguilla anguilla*), Eurasian minnow (*Phoxinus phoxinus*) and brown trout (*Salmo trutta*) which were detected in 75, 73, and 70 samples, respectively. High-quality vertebrate sequence data were obtained for 89 of the 90 eDNA samples (Table 10). Three field negative control samples produced target reads below our reporting threshold. The remaining two field negative control samples contained only common contaminant (non-target) DNA. All laboratory controls behaved as expected.

In the invertebrate dataset, a total of 1111 taxa were detected (from a total of 7,522,286 invertebrate sequence reads). The average taxon richness was 119 and ranged from 12 to 247. The most abundant sequence was a fly in the order Diptera (*Chaoborus flavicans*) which accounted for 7.3% of the total target sequence reads. Among the most commonly detected target taxa were three Chironomid midges (*Ablabesmyia monilis*, *Pseudorthocladius filiformis*, and *Stempellinella brevis*), which were detected in 74, 72, and 72 samples, respectively. Commonly used taxonomic groups for freshwater aquatic indices were also found: 15 Ephemeroptora (mayflies) OTUs (14 to species level); 21 Plecoptera (stoneflies) OTUs (15 to species level) and; 19 Trichoptera (caddisflies) OTUs (19 to species level). No Odonata (dragonflies) were found. High-quality invertebrate sequence data were obtained for 89 of the 90 eDNA samples (Table 10). The negative control samples did not amplify any DNA. All laboratory controls behaved as expected.

In the bacteria dataset, a total of 826 taxa were detected. OTUs were detected across 20 different phyla within the kingdom Bacteria (from a total of 4,039,799 bacterial sequence reads). The average taxon richness was 120 and ranged from 29 to 206. The phylum with the highest proportion of OTUs was Proteobacteria and the most abundant sequence was from the family Enterobacteriaceae, which was detected in all samples. High-quality bacterial sequence data were obtained for 89 of the 90 eDNA samples (Table 10). One out of five field negative control samples contained bacterial target reads, the other four did not amplify any DNA. All laboratory controls behaved as expected. Note that this assay does detect some cyanobacteria. However, it is not a cyanobacteria-specific assay.

Table 9: Number of OTUs detected and the percentage of OTUs identified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Vertebrates	60	100%	100%	95%	83.3%	81.7%	60%
Invertebrates	1111	99.8%	98.9%	95.8%	84.7%	64%	43.1%
Bacteria	862	80.9%	72.3%	58.6%	45.9%	17.3%	4.5%

Table 10: Freshwater sequencing data summary

Assay	N reads obtained	N samples	N samples that did not produce data	Samples that did not produce data
Aquatic Vertebrates	4,254,276	90	1	VOIL-05 (Loch Voil – Good)
Aquatic Invertebrates	7,522,286	90	1	LLSS-01 (Loch Lomond South – Moderate)
Aquatic Bacteria	4,039,799	90	1	LLSS–04 (Loch Lomond South – Moderate)

3.2.2 Freshwater Research Questions & Key Findings

3.2.2.1 Summary of Freshwater Habitat Key Findings

Freshwater eDNA-based data can differentiate freshwater lochs between High-Good and Moderate-Poor WFD overall status categories. Freshwater eDNA-based data could not distinguish freshwater lochs between High and Good statuses. Although the Poor status loch, Castle Semple Loch, was clearly distinguished from other lochs, it was the only Poor status loch in the project, so we cannot make further generalisations. Freshwater invertebrates appear to be the most powerful for classifying lochs by Overall Status. Fish communities also showed a relatively high predictive accuracy. This provides a very strong indication that a national model can be developed to support WFD-type loch classification based on eDNA-based data.

CPET results using eDNA-based data produce similar values to those produced using best-matching morphological data. This provides support that freshwater invertebrate eDNA-based data from freshwater lochs could be integrated into the existing CPET classification system.

3.2.2.2 Research Questions & Key Findings

FRQ1: Are there any eDNA-based units/metrics that can act as indicators of biodiversity and be used to (a) distinguish habitat condition and (b) detect ecological pressures or restoration stage in Scottish freshwater loch habitats?

To address this question, we investigated whether there were clear trends evident across WFD Overall Status classification for any of the following:

- species richness (vertebrates, invertebrates, bacteria)
- evolutionary diversity (vertebrates, invertebrates, bacteria), and
- functional diversity (bacteria).

Community-based statistics, species of interest results, and CPET results are detailed under the FRQs that follow below.

FKF1: There were no clear trends with respect to the metrics investigated. Figures and model outputs are provided separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c). This means

that although sample-level metrics such as species richness can form useful descriptive statistics, they are not a reliable indicator of loch habitat condition in Scotland. Neither pH nor conductivity (a measure for salinity) showed a significant relationship for any of the metrics.

FRQ2: Can eDNA metabarcoding community data be used to assess ecosystem health (condition, function, and resilience) of Scottish freshwater loch habitats in response to pressures such as land use, pollution/water quality, INNS, and/or climate change?

To address this question, we investigated whether there were clear trends evident across mean annual values of Alkalinity, Chlorophyll-a, Nitrate, Nitrite, Dissolved Oxygen, Total Phosphorus, and land use for any of the following: species richness (vertebrates, invertebrates, bacteria), evolutionary diversity (vertebrates, invertebrates, bacteria), and functional diversity (bacteria). We also performed ordinations incorporating these pressures to assess whether they might be influencing the communities detected. Species of interest results, and CPET results are detailed under the FRQs that follow below.

FKF2.1: *There were no clear trends with respect to most metrics investigated. This means that although sample-level metrics such as species richness can form useful descriptive statistics, they are not a reliable indicator of loch habitat condition in Scotland*.

The only sample-level metric that had a significant p-value and reasonable R^2 value was the Fish Trophic Level Index. There was a negative relationship between mean annual nitrate values and the Fish Trophic Level Index ($p = 0.03$, marginal $R^2 = 0.42$, conditional $R^2 = 0.69$, Figure 12). This suggests that the prevalence of predatory fish is reduced in higher nitrate lochs. However, we recommend that this metric is calculated only if five or more species are detected in a sample at the species level. Only 29 samples met this condition so this is based on a very limited dataset and the result should be treated cautiously. Another note of caution is that the fish species present in a loch may also partly be an artifact of stocking and different native ranges (such as the natural lack of roach in the highlands).Nonetheless, this could warrant further investigation in future studies.

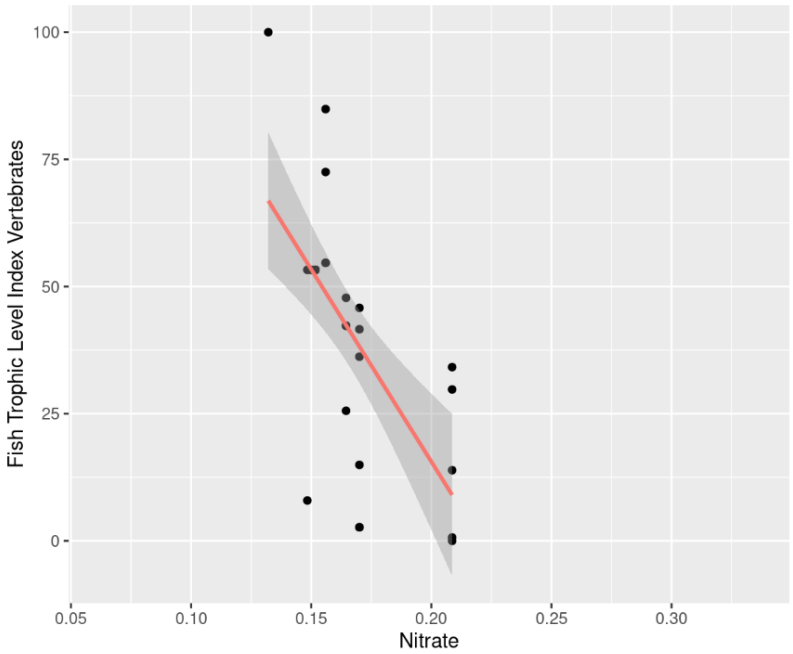


Figure 12: Relationship between Nitrate and Fish Trophic Level Index. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

FKF2.2: *There were no clear and consistent trends in eDNA-based community data across mean annual values of any of the predictor variables.*

Given the relatively small size of the freshwater dataset, running multiple tests for continuous predictor variables was not statistically appropriate. Thus, we only visualised the correlation between the variables and the ordination scores. We point out where there appears to be some clustering along the LV1 axis, and these patterns can be used to generate hypotheses for future investigation. It should also be noted that project constraints only allowed the comparison of community data against mean annual values. It may be the case that trends could be more evident using values corresponding to the month of sampling. For mean annual alkalinity (Figure 13) and Chlorophyll-a (Figure 14), a weak trend was evident. There was no trend evident for the other variables. In all cases, the site scores (LV1) are from the original gllvm unconstrained ordination.

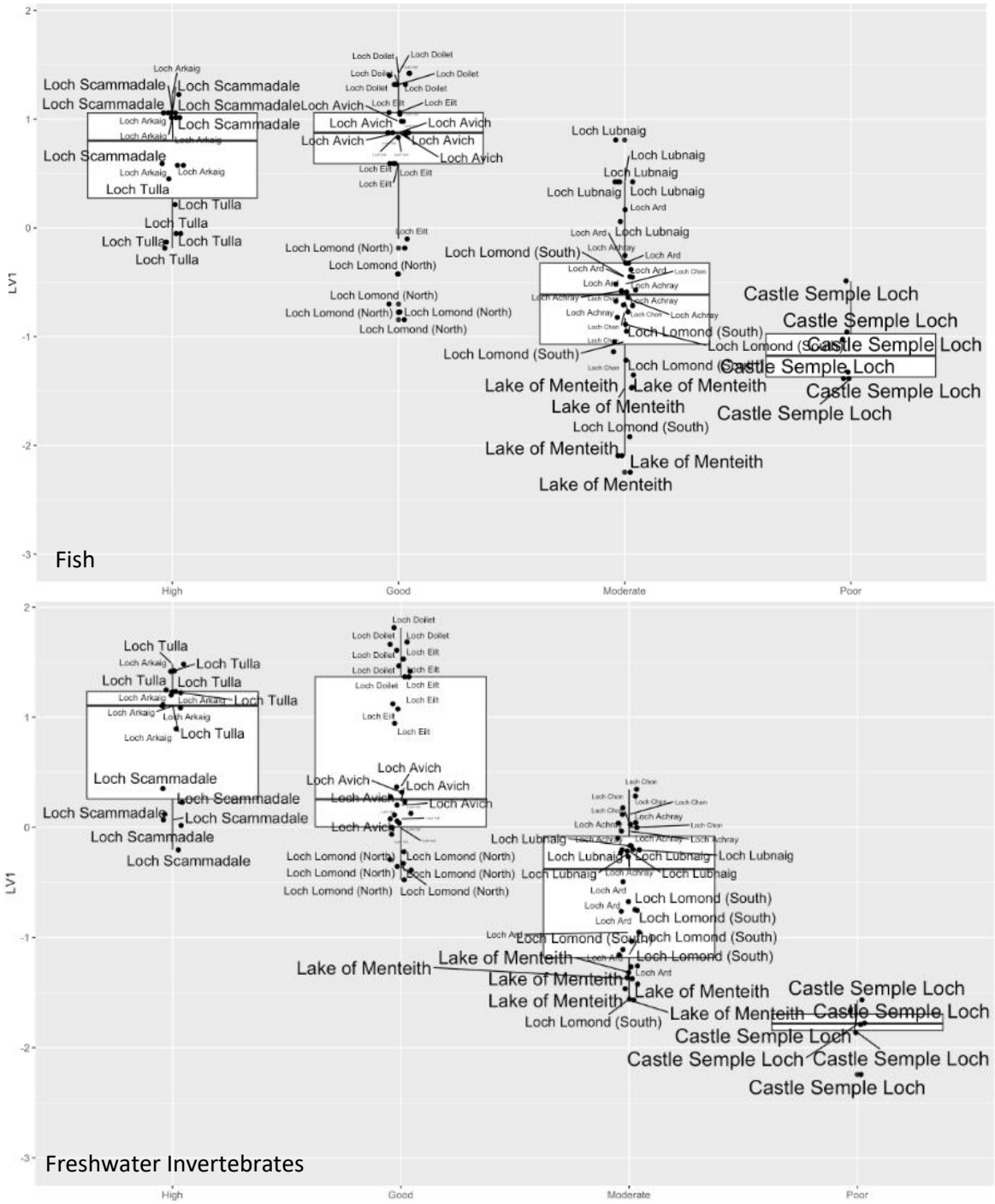


Figure 13: Relationship between loch alkalinity (mean annual values) and either fish or freshwater invertebrate eDNA-based data. Larger text indicates higher alkalinity values. The existing WFD Overall Status is used for grouping lochs in four boxplots to provide further context. A clear trend would be an increase in the size of the text along the y-axis. While the trend is not consistent across all samples, the lochs with the highest alkalinity values (Lake of Menteith and Castle Semple Loch) do appear to be grouped together, indicating that there may be an effect of alkalinity that warrants further investigation. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

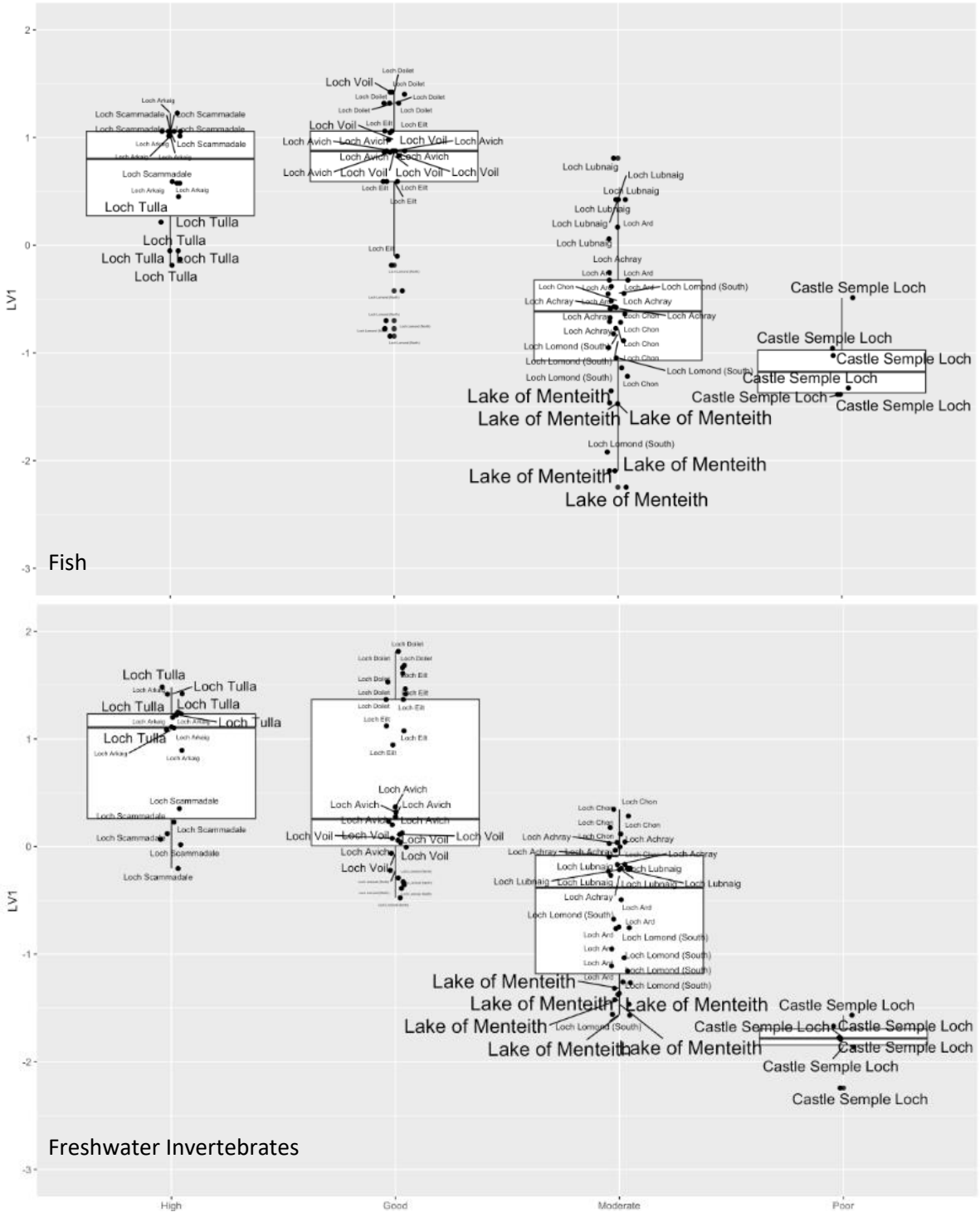


Figure 14: Relationship between Chlorophyll-a (mean annual values) and either fish or freshwater invertebrate eDNA-based data. The size of the text indicates higher Chlorophyll-a values. The existing WFD Overall Status is used for grouping lochs to provide further context. A clear trend would be an increase in the size of the text along the y-axis. The relationship between Chlorophyll-a values and site (LV1) scores is, at best, weak, with Lake of Menteith and Castle Semple Loch having the highest Chlorophyll-a values, but these overlap with lochs which have low Chlorophyll-a values. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

FRQ3: Can the DNA-derived metrics extend from taxonomic descriptions to assess possible pressure or restoration-related changes in Scottish freshwater loch habitats?

FKF3*: From other FRQs we have shown that eDNA metabarcoding data can be used to distinguish and classify between lochs of different WFD Overall Statuses with reasonable accuracy, and that the effects of possible pressures can be investigated. These methods do not rely on taxonomy.*

We did not observe clear trends for evolutionary diversity (vertebrates, invertebrates, bacteria), or functional diversity (bacteria) across either WFD Overall Status or possible pressures. Similarly, we did not observe clear trends for community data across possible pressures (mean annual values), although we make note of limitations for these statistical analyses.

We also investigated loch catchment land cover as a potential pressure that may influence species composition in the lochs. Our analyses showed that land cover was not a significant contributor to species composition for freshwater fish and invertebrates. Partial R^2 values for condition (Overall_Status_Grouped) were 0.826 and 0.472 for fish and invertebrates, respectively, whereas the landcover variable (Principal Component 1), which differentiates Moor versus Urban/Woodland, had partial R^2 values of 0.029 and 0.079, respectively. These partial R^2 values are the proportion of generalised variance explained by all predictors and latent variables and provide a way to compare variable importances.

FRQ4: Can eDNA metabarcoding community data provide similar and reliable outcomes as conventional morphology-based or scoring-based methods for invertebrates (e.g., CPET taxonomic input (species composition and abundance) and WFD metric output data), bacteria (e.g., PLUTO cyanobacteria taxonomic input (species composition and abundance) and WFD metric output data), and fish (e.g., taxonomic input (species composition and abundance) and WFD metric output data) to assess the ecological status of Scottish freshwater loch habitats, using best available matching data examples?

To address this question, we explored using the freshwater invertebrate eDNA-based data as input for the CPET index method. We also compared the detections of fish with known historical records. We did not have capacity in this project to explore the PLUTO metric.

FKF4.1: The eDNA-based CPET status from this project agreed with the 2020 conventional WFD Macroinvertebrates CPET scoring in 70% of cases (7/10 lochs; Table 11). Where there were disagreements, they were all between High and Good (i.e. a single status difference). Applying eDNA-based data to the CPET metric shows promise for method development. It must be noted that only five lochs have in-loch conventional CPET data, the remaining lochs have CPET classes predicted based on lochs they are grouped with for WFD reporting purposes.

Table 11: CPET status for each loch for the conventional approach (Macroinvertebrates CPET, based on 2020 data from the SEPA Classification Hub) and the eDNA-based approach in this project (eDNA CPET). eDNA CPET was calculated based on the average EQR of the six samples collected . NA = Not Available. It must be noted that only five lochs have in-loch conventional CPET data, the remaining lochs have CPET classes predicted based on lochs they are grouped with for WFD reporting purposes. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

Loch	Macroinvertebrates CPET	eDNA CPET	Grouped data
Loch Arkaig	NA	High	Y
Loch Tulla	High	High	Y
Loch Scammadale	High	Good	Y
Loch Doilet	High	High	N
Loch Voil	High	High	Y
Loch Eilt	High	High	Y
Loch Avich	High	High	Y
Loch Lomond (North)	High	High	N
Loch Lomond (South)	High	High	N
Loch Achray	NA	High	Y
Loch Lubnaig	Good	High	N
Loch Ard	NA	High	Y
Lake of Menteith	High	Good	N
Loch Chon	NA	Good	Y
Castle Semple Loch	NA	Good	Y

Comparison of matching available data using raw EQR values

CPET EQR scores from two lochs (Loch Doilet and Loch Lubnaig) were equivalent using either the conventional Macroinvertebrates CPET approach or the eDNA CPET approach (Figure 15). Two of the lochs (Lake of Menteith and Loch Lubnaig) had more variable results and, on average, assigned CPET status higher with eDNA-based than morphological data . It should be noted that the data for the conventional Macroinvertebrates CPET approach was compiled from various years from 2007 to 2018 and had substantial variation. There was no significant difference between methods when pooling lochs (ANOVA: $F_{1, 49} = 0.09$, $p > 0.05$) or for any individual lochs (ANOVA: $F_{3, 49} = 2.4$, $p > 0.05$).

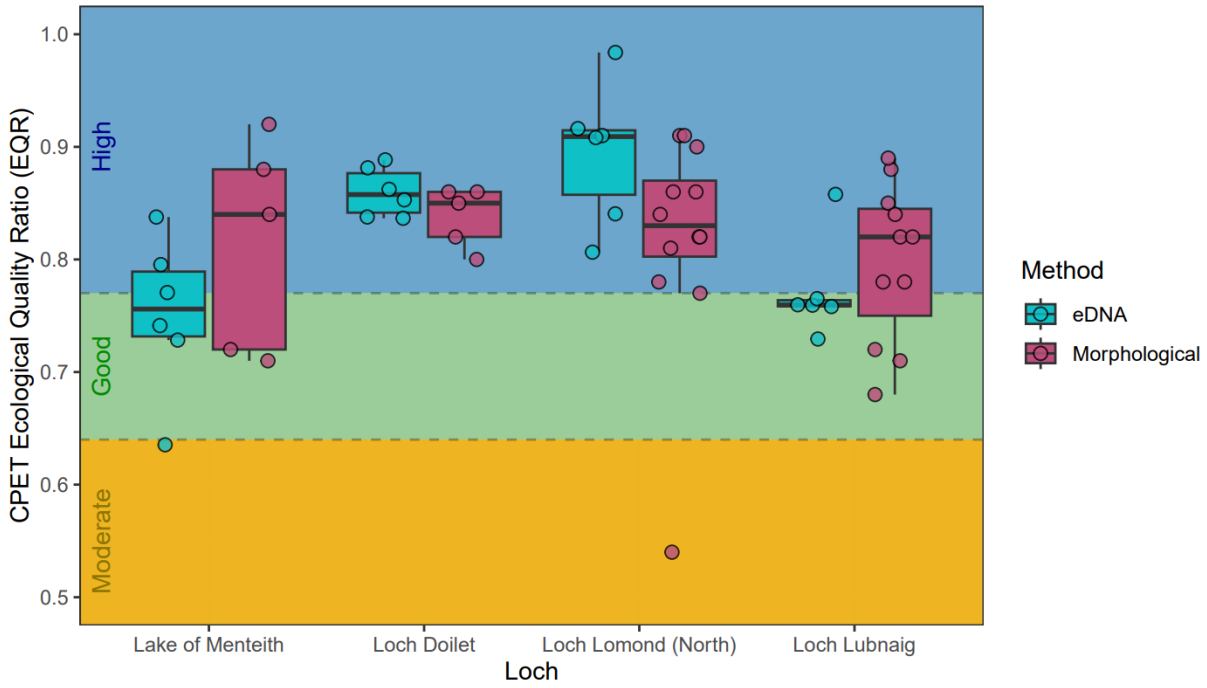


Figure 15: Comparison of CPET EQRs generated for four lakes (Lake of Menteith, Loch Doilet, Loch Lomond (North), and Loch Lubnaig) using the conventional Macroinvertebrates CPET approach or the eDNA CPET approach. Conventional Macroinvertebrates CPET data are historical so are not directly comparable. Coloured background indicates the thresholds of a water body having moderate (yellow), good (green), or high (blue) CPET status boundaries, as defined by the WFD (UKTAG 2008). Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

We conducted an initial investigation into the CPET taxa that are known to be detectable using the approach taken in this project. At least 47% of taxa are detectable using the eDNA CPET approach. However, it is likely that in reality this proportion is far higher, as there are likely to be many cases where the taxonomy used by the Macroinvertebrates CPET approach is different to that used by NatureMetrics (gbif backbone). That said, there are likely to be taxa that are currently not detected by the eDNA CPET method and this warrants further investigation. It should also be noted that the eDNA CPET approach detects many chironomid taxa that are not currently used for Macroinvertebrates CPET scoring. The results of the initial investigation are provided separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c).

Fish eDNA Metabarcoding Results

We have shown under previous FRQs that fish eDNA community data can be used for distinguishing habitat condition (WFD Overall Status) and has potential to be used to track functional changes in response to pressures through trophic structure changes. As there is already a Lake Fish Classification tool being used for loch monitoring, we present here the results from our fish eDNA-based data from the Vertebrates assay in the context of historical known occurrences of fish in these lochs.

The results from this project concur with previous work that has shown eDNA metabarcoding is a useful approach to surveying for fish in lochs. We identified freshwater fish species of note, each of which can inform environmental managers of particular health characteristics of the water body (Table 12). These include endangered species, such as the European eel (*Anguilla anguilla*) and commercially important species, such as the Atlantic salmon (*Salmo salar*) which is highly valued for sport fishing. Non-native fish species like the rainbow trout (*Oncorhynchus mykiss*) were also detected. *Oncorhynchus mykiss* has

been introduced to Scotland's rivers and lochs for sport fishing and commercial aquaculture. While our search for historical records was not exhaustive, we compiled records from 15 sources (see Phase 2 Technical Appendices output; Bakker et al. 2023c). Based on these records we have added 21 records to the known distribution of fishes in these lochs (Figure 16). However, it is also evident that the approach taken in this project does not detect all fish species in a loch. It should be noted that a number of the species records from previous surveys may be truly historical and that some species may no longer be present in some of the lochs. For example, the Balloch barrage was built on the Loch Lomond outflow in the early 70s, which may now be inhibiting the presence of species such as mullet in Loch Lomond. Numerous species were also detected in only one sample (Figure 17). Previous work has shown that a minimum of 10 samples are needed to detect 85% of fish species in UK lakes (Li et al. 2019). It has also been shown that summer sampling may be less effective for detecting certain species, such as arctic charr and coregonids (Table 12) (Sellers et al. 2023), which the results of this study suggest as well. Hence sampling outside of the winter season, may require more intense sampling.

The detection of Ruffe (*Gymnocephalus cernua*) in Loch Achray, Loch Chon, and Lake of Menteith was surprising. This species is at a high abundance in Loch Lomond, being first discovered there in 1982 (Adams and Maitland 1998). It is not known from these other lochs and this warrants further investigation. There is potential that these are not true positives as the signal for these was extremely low (< 35 reads) and they were only detected in either one or two samples from each loch (Figure 17). Potential sources of such a result include “environmental contamination” (for example Ruffe are known to be moved as live bait, or Ruffe eDNA could be introduced to a loch by predatory bird defecation) and “technical contamination” which can occur during laboratory or sequencing steps. The latter is carefully monitored through the use of negative controls, but the former is harder to assess. We recommend following this up with further eDNA surveys in the suspect lochs followed by conventional surveys if more detections are found.

Table 12: Freshwater fish species of note detected, including IUCN designation and input from specialist members of the project Management Steering Group (MSG) and the Technical Reviewing Group (TRG).

Species	Common name	IUCN status	Specialist input
Anguilla anguilla	European eel	CR	Expected to be present in all locations with free access to the sea. Interesting rarity in Loch Arkaig, suggests effect of major waterfalls downstream.
Salmo salar	Atlantic salmon		Expected to be present in all locations with free access to the sea, where water quality is sufficient.
Salmo trutta	Trout		Very common species, where water quality is sufficient. Rarity in Lake of Menteith interesting.
Abramis brama	Common bream		Introduced species in Scotland. Known records from Loch Lomond
Phoxinus phoxinus	Minnow		Very common species, complex history of natural and artificial introductions. Absence from Lake of Menteith is interesting.
Leuciscus leuciscus	Dace		Introduced species in Scotland. Known records from Loch Lomond (C. Adams). Lake of Menteith record is interesting.
Rutilus rutilus	Roach		Complex history of natural and artificial introductions. Debatable native status in southwest Scotland (Maitland), introduced elsewhere?
Tinca tinca	Tench		Introduced species in Scotland. Known records from Loch Lomond. Lake of Menteith record is interesting.
Barbatula barbatula	Stone loach		Common species in southern Scotland. Predominantly running waters.
Esox lucius	Northern pike		Common species, complex history of natural and artificial introductions.
Gasterosteus aculeatus	Three-spined stickleback		Expected to be present in all locations with free access to the sea. Surprisingly low number of records in this dataset.
Gymnocephalus cernua	Ruffe		Introduced species in Scotland, considered high risk invasive species. Presence in Loch Achray, Loch Chon and Lake of Menteith may be surprising
Perca fluviatilis/Sander lucioperca	Perch/Zander		No records of zander in Scotland. Perch widespread, with a complex history of natural and artificial introductions
Oncorhynchus mykiss	Rainbow trout		*O. mykiss* commonly stocked for sport angling. Lake of Menteith a major stocked fishery, so expected there. Record in Loch Achray unexpected.
Salvelinus alpinus/Salvelinus fontinalis	Arctic or Brook charr		Arctic charr native and of conservation interest. Would expect more records, e.g. Lochs Lubnaig, Tulla, Avich. (Likely due to low number of samples and also the summer sampling season from shoreline only). Brook charr very rare, and introduced, in Scotland.
Lampetra fluviatilis/Lampetra planeri	River lamprey/brook lamprey		Brook lamprey are widely distributed throughout much of Scotland, particularly south of the Great Glen. River lamprey show a similar distribution.

Figure 16: Contrast between fish species detected in this project versus compiled historical occurrences records for each loch. Where necessary, taxonomic names have been aligned to those used in the rest of this project. Since zander have never been recorded in Scotland, assignments flagged in this project as *Perca fluviatilis*/*Sander lucioperca* were modified to *P. fluviatilis*. Assignments to Salvelinus are either *Salvelinus alpinus* (Arctic charr) or *Salvelinus fontinalis* (brook charr). Assignments to *Lampetra* are either *Lampetra fluviatilis* (river lamprey) or *Lampetra planeri* (brook lamprey).

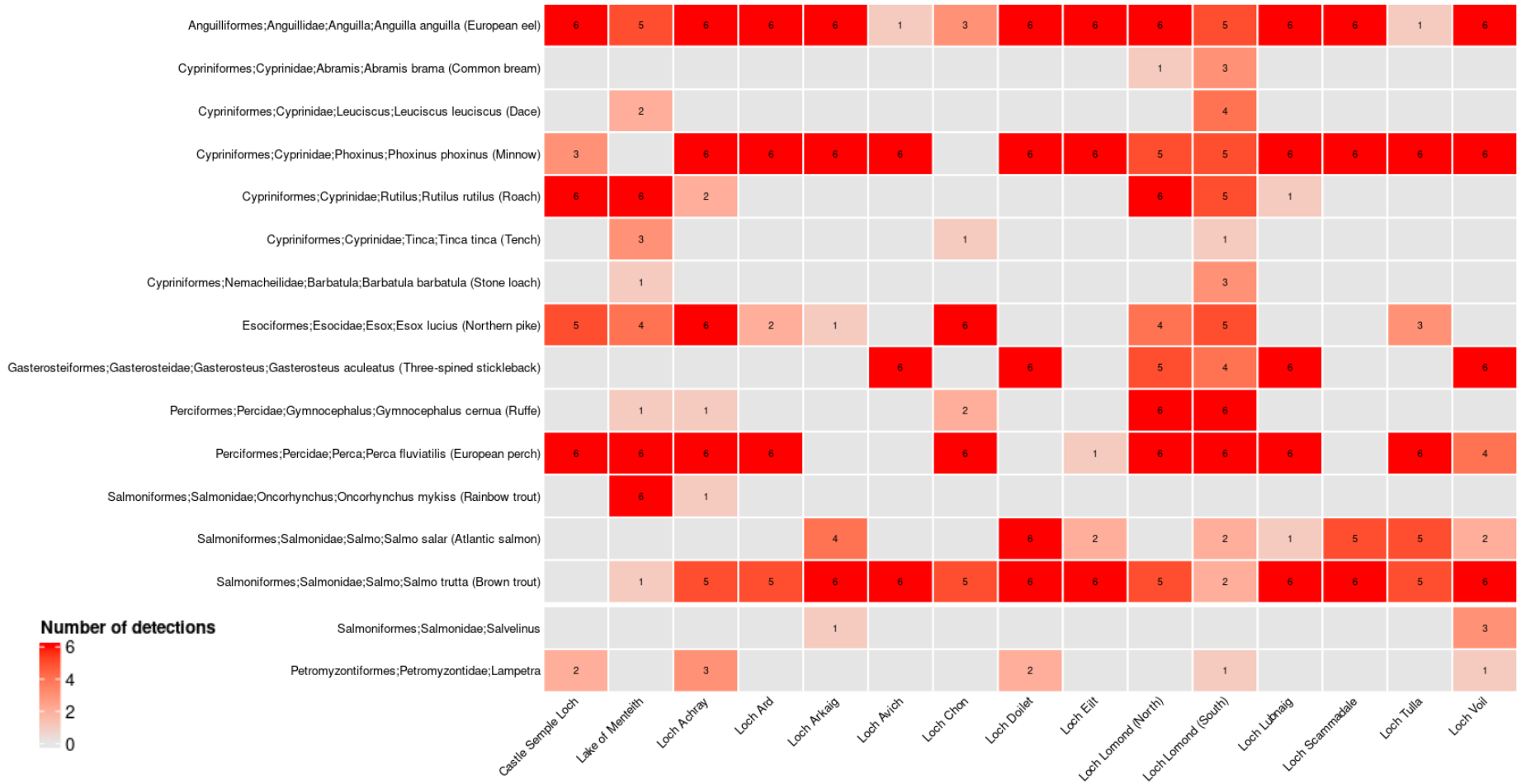


Figure 17: Fish species detected in this project for each loch. Numbers indicate the number of samples each species was detected in. Since zander have never been recorded in Scotland, assignments flagged in this project as *Perca fluviatilis*/*Sander lucioperca* were modified to *P. fluviatilis*. Assignments to Salvelinus are either *Salvelinus alpinus* (Arctic charr) or *Salvelinus fontinalis* (brook charr). Assignments to *Lampetra* are either *Lampetra fluviatilis* (river lamprey) or *Lampetra planeri* (brook lamprey).

FRQ5: Are eDNA metabarcoding data from water samples broadly similar or different in biodiversity and community composition amongst Scottish freshwater loch habitats of ranging ecological status (WFD high, good, moderate, and poor classification categories)? If we find any differences, are these data sufficiently clear and predictable to be used to distinguish (classify or sort) between water samples from Scottish freshwater loch habitats of ranging ecological status (WFD high, good, moderate, and poor classification categories)?

FKF5: *Fish and invertebrate eDNA metabarcoding community datasets were broadly similar between lochs with High-Good classifications and those with Moderate-Poor classifications, this indicates that classification of water samples to an ecological status based on eDNA-based data is possible. We discuss the latter point in more detail in FKF6, which specifically focuses on classifying water samples with an unknown status classification.*

In the following concurrent ordination models, we included mean depth (m), and/or conductivity, and different combinations of two land cover covariates, alongside WFD water quality status. The land cover covariates are the first two principal components extracted from the four raw land cover percentages (urban, woodland, arable, and moorland), whereby PC1 represents a transition from moorland (low values) to urban/woodland (high PC1 values), and PC2 represents a transition from arable land (low PC2 values) to woodland/moorland (high PC2 values). The WFD scoring categories, including loch catchment land cover, can be found separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c).

FKF5.1 (Fish): In both unconstrained and concurrent (constrained) ordinations for freshwater fish (Figure 18), there is a separation into two community clusters that correspond to two groups of conditions: High/Good and Moderate/Poor. In the unconstrained ordination, Loch Lomond North, which is classified as WFD Good, clusters with Loch Lomond (South), which is classified as Moderate. Given that the two halves of Loch Lomond are contiguous, we expect the fish (and/or their eDNA) to disperse at least up to a certain point between the two sides, even though the habitat is very different between the two parts of the loch. In the concurrent ordination, Overall Status explains 82.6% of the variation explained by the two latent variables, which is ~18 times greater than the variation explained by landcover (2.9%) (see Phase 2 Technical Appendices output; Bakker et al. 2023c).

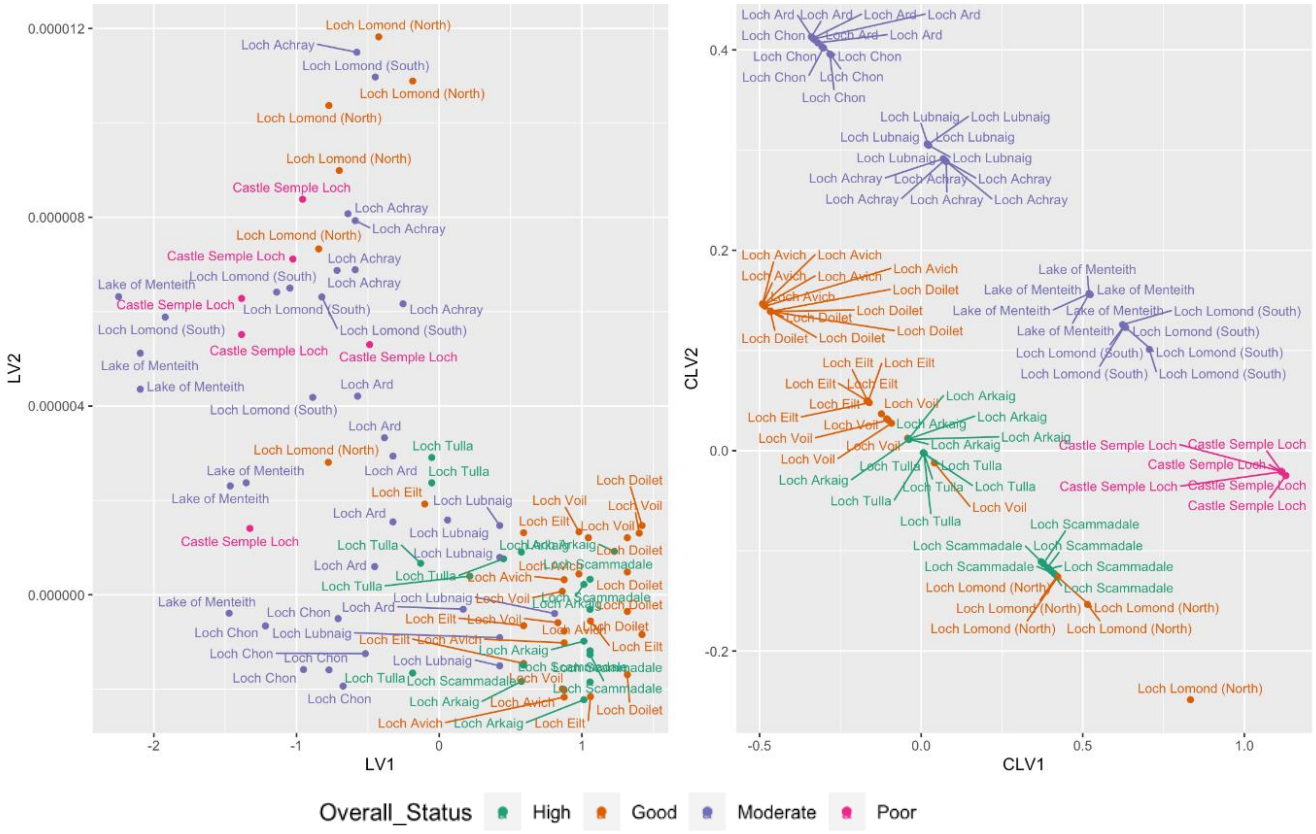


Figure 18: Unconstrained (left) and constrained/concurrent (right) ordinations for freshwater fish. Each point is a sample and samples that are closer together are more similar. In both models, there is a separation by Overall Status: High/Good and Moderate/Poor. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

Eight freshwater fish species were associated with a higher number of incidences in Moderate-Poor conditions and three species with more incidences in High-Good conditions (Table 13).

Table 13: Species with higher numbers of incidences in lochs with Moderate-Poor and High-Good overall WDF statuses. With the assay that was used it is not possible to distinguish between *S. alpinus* and *S. fontinalis,* but based on local records and specialist knowledge, the char species are almost certainly arctic char, *S. alpinus*.

Species Common Name	Scientific Name	Higher Incidence in…
European perch / Zander	*Perca fluviatilis / Sander lucioperca*	Moderate-Poor
Stone loach	*Barbatula barbatula*	Moderate-Poor
Tench	*Tinca tinca*	Moderate-Poor
Rainbow trout	*Oncorhynchus mykiss*	Moderate-Poor
Common dace	*Leuciscus leuciscus*	Moderate-Poor
Perch species (unidentified)	*Percidae*	Moderate-Poor
Northern pike	*Esox lucius*	Moderate-Poor
Common roach	*Rutilus rutilus*	Moderate-Poor
Common minnow	*Phoxinus phoxinus*	High-Good
Brown trout	*Salmo trutta*	High-Good
Arctic char / Brook trout	*Salvelinus alpinus / Salvelinus fontinalis*	High-Good

FKF5.2 (Invertebrates): In both unconstrained and concurrent (constrained) ordinations for invertebrates (Figure 19), there is a separation into three community clusters that correspond to three groups of conditions: High/Good, Moderate, and Poor. Notably, even Loch Lomond (North) clusters

with the High-Good lochs, unlike with the fish dataset, which suggests that the invertebrate communities are more spatially structured, allowing for the higher resolution detection of local changes in water quality given the current sampling strategy. Also, since there is only one loch with Poor status, it is not possible to know whether Castle Semple Loch is different because of its condition or some other loch-specific characteristic.

Land cover, condition, and conductivity are all significant contributors to the constrained ordination. The total variance explained by the latent variables is substantial (R^2 = 46%), and condition is ~10 times more important than land cover or conductivity in contributing to this explanatory power. Samples toward the bottom in Figure 19 are High-Good, and samples toward the right are surrounded by more moorland. See Phase 2 Technical Appendices (Bakker et al. 2023c) for statistical output from the ordination plots for freshwater invertebrate communities.

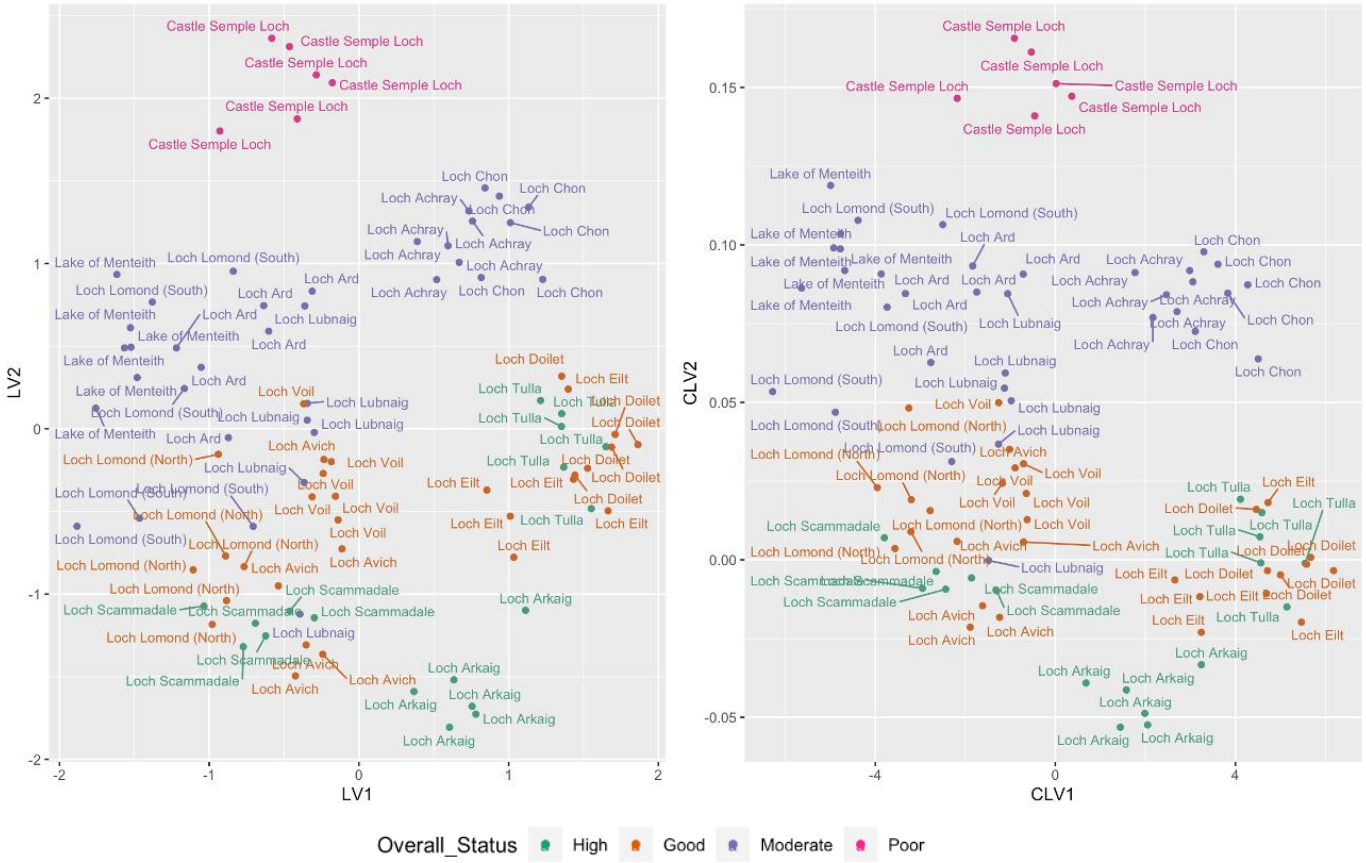


Figure 19: The unconstrained and constrained ordinations for freshwater invertebrates. Each point is a sample and samples that are closer together are more similar. In both models, there is a separation by Overall Status: High/Good and Moderate/Poor. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

FKF5.3 (Bacteria): Bacteria community compositions are clustered by loch (i.e. they show a strong geographic structure) but not by condition (Figure 20). See Phase 2 Technical Appendices (Bakker et al. 2023c) for statistical output from the ordination plots for freshwater bacteria communities.

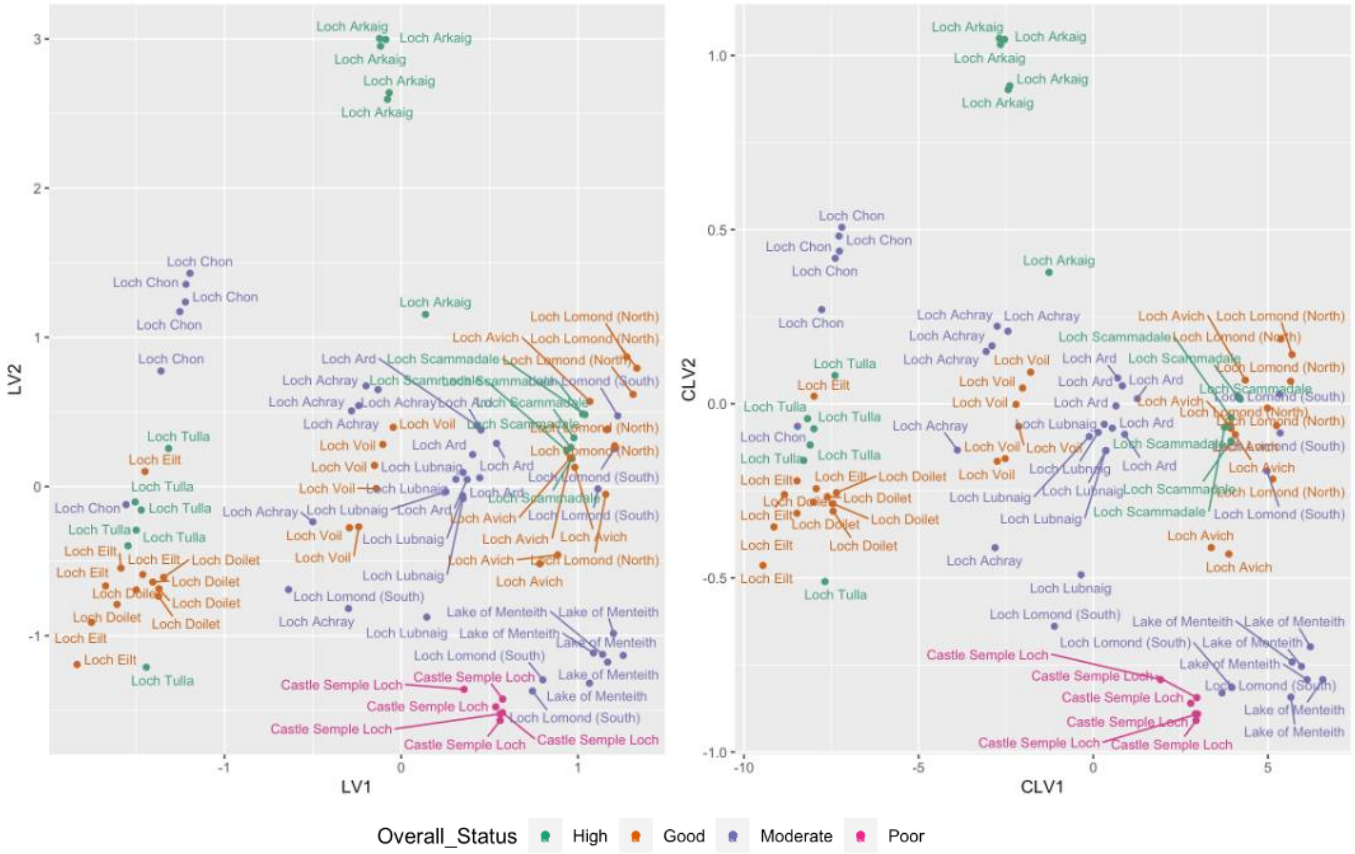


Figure 20: The unconstrained and constrained ordinations for bacteria. In both cases, samples cluster primarily by loch and not condition. In this dataset, the primary drivers of composition are land cover (PC1 and PC2 accounting for nearly all the variation along CLV1 and CLV2), followed distantly by conductivity. Contains SEPA data © Scottish Environment Protection Agency and database right (2023). All rights reserved.

FRQ6: Can we make similar and reliable DNA-based ecological predictions for unmonitored (grouped) Scottish freshwater loch habitats when compared with examples of similar-expected health status (condition category) and lake typology but where no matching contemporary data exists?

FKF6.1*: From the lochs that were sampled, fish and invertebrate community datasets can be used to classify Scottish lochs to High-Good and Moderate-Poor health statuses, with good to high accuracy. CPET results from previous FRQs showed that there was a high agreement between existing grouped lochs and the CPET values derived in this project.*

We pooled the four WFD categories into two: High-Good and Moderate-Poor. One reason is that there is only one Poor class loch included (Castle Semple Loch), which made it impossible to train a model to classify to Poor status. Another reason is that High and Good categories were indistinguishable in all of the datasets, so by using only two categories, the predictive error rate focuses on whether eDNA-based data can distinguish the most important transition, the one from Moderate-Poor to High-Good.

With these two condition classes, random classification would achieve a predictive accuracy of 50% for unmonitored Scottish freshwater lochs. The Random Forest cross-validation test (Phase 2 Technical Appendices output; Bakker et al. 2023c) shows that:

(1) Freshwater bacteria can classify samples to High-Good with ~62% accuracy and to Moderate-Poor with ~54% accuracy (random).

(2) Freshwater fish can classify samples to High-Good with ~70% accuracy and to Moderate-Poor with ~70% accuracy.

(3) Freshwater invertebrates can classify samples to High-Good with ~81% accuracy and to Moderate-Poor with ~71% accuracy.

Freshwater invertebrates show the highest predictive accuracy. Part of the reason for predictive error is that Loch Lomond received two classifications (North = Good, South = Moderate), but eDNA (particularly for fish) is likely dispersed between the two parts of the basin.

FRQ7: Are there any differences or similarities in habitat condition or biodiversity and community composition from any re-surveyed locations between timescales (e.g., seasons or years) and protected status (e.g., designated vs non-designated sites) for the Scottish freshwater loch habitats sampled?

FKF7.1: Of the 33 fish species detected previously in lochs (through a combination of fishing methods), 16 to 19 species were identified from the eDNA samples. The percentage of known species detected ranged from 33% (Loch Lomond) to 300% (Loch Doilet). However, for larger lochs, several freshwater species were not identified, namely the ninespine stickleback (*Pungitius pungitius*), European chub (*Squalius cephalus*), and the introduced crucian carp (*Carassius carassius*). Interestingly, both *Salmo salar* and *Salvelinus alpinus* were not detected in six lochs where they had been previously recorded.

Compared to previous eDNA surveys carried out during winter for the development of the Fish Lake Classification tool, there were very few additional species occurrences. These results are not directly comparable as the sampling intensity in the current project was lower, the sampling season differed, and the bioinformatics used for the development of the tool are somewhat more conservative than the approach used in this project (for example, the pipeline of the Fish Lake Classification tool uses quite a conservative filtration threshold of 0.1% of total reads per sample, whereas in this project we used 0.02%). However, the overall results are encouraging, with an average agreement rate of 0.73 (the number of fish species detected by both methods in a given loch divided by the number of species detected in the loch overall). This is despite the differences in methodologies and seasons, showing that eDNA-based surveying for fish in lochs in reasonably robust.

Figure 21: Contrast between fish species detected in this project versus previous eDNA-based surveys for each loch. Where necessary, taxonomic names have been aligned to those used in the rest of this project. Since zander have never been recorded in Scotland, assignments flagged in this project as *Perca fluviatilis/Sander lucioperca* were modified to *P. fluviatilis*. For the same reason, assignments flagged as Percidae in previous eDNA projects were modified to *P. fluviatilis*. As *Coregonus lavaretus* is the only valid taxonomic unit of the genus ever recorded in Loch Lomond (based on historic data compiled for this report), detections of genus *Coregonus* have been elevated to *C. lavaretus*. Assignments to Salvelinus are either *Salvelinus alpinus* (Arctic charr) or *Salvelinus fontinalis* (brook charr). Assignments to *Lampetra* are either *Lampetra fluviatilis* (river lamprey) or *Lampetra planeri* (brook lamprey).

FRQ8: What are the key recommendations and future perspectives to develop and operationalise standardised eDNA-based methods at national scale for monitoring biodiversity and the impacts of pressures or restoration in Scottish freshwater loch habitats?

FKF8.1: How should we sample?

Even in large lakes such as Loch Lomond, eDNA can still be localised, so the collection of multiple water samples remains key to capturing most of the eDNA present (Bruce et al. 2021; Di Muri et al. 2022; Hänfling et al. 2016; Hervé et al. 2022; Lawson Handley et al. 2019; Zhang et al. 2020). For using the Fish Lake Classification tool, there is already guidance that 10-20 samples are required, and we do not make any recommendations to the contrary.

Where resources are limited and the purpose of the surveys is to increase the knowledge of distribution of fish species in, for example, previously unsurveyed lochs, then reducing the sampling effort to six samples may suffice to provide a snapshot of all but the least abundant species.

We recommend using filtration techniques and this is the most common approach in eDNA sampling, having superseded ethanol precipitation methods.

FKF8.2: What should we sample?

This will depend on the purpose of the sampling event. From this project we can recommend sampling for fish and freshwater invertebrates. We cannot currently recommend the same for bacteria as we did not observe a strong trend across the gradient being investigated. Further work is required to assess the application of sampling for bacteria in lochs.

FKF8.3: Where should we sample?

While sampling from boats and sampling at depth has often been carried out by studies using eDNA sampling in large lakes (e.g. Egeter et al. 2022), this can be expensive and logistically challenging. In this project, the shoreline approach was similar (though not identical) to the approach recommended by Willby et al. (2020) and yielded data that could be used to classify lochs by WFD status. It also yielded chironomid data that produced similar scores to conventional methods. On balance, we recommend shore sampling.

We recommend collecting independent samples, each comprised of subsamples (e.g. 2 L comprised of 5 x 400 mL subsamples), collected from multiple locations around the lake perimeter (Bedwell and Goldberg 2020; Li, Lawson Handley, et al. 2019).

FKF8.4: When should we sample?

This will depend on the focus of the survey/monitoring event and the taxonomic group being targeted.

There is already current advice to sample lochs during the winter season when using the eDNA-based Fish Lake Classification tool (Willby et al. 2020). Further work is being conducted which verifies this (Griffiths, Hänfling, and Bean 2023). Collecting samples when a lake is not thermally stratified is ideal as more mixing of the water will occur. This means that there is a higher chance of collecting eDNA from both shallow and deep-water aquatic species (Bruce et al. 2021; Hervé et al. 2022; Lawson Handley et al. 2019; Littlefair et al. 2020). However, the detection of invertebrates and some other taxonomic groups is generally lower in colder temperatures (Schmidt et al. 2021; Reinholdt Jensen et al. 2021; Troth et al. 2021; K. Harper et al. 2018) and this might reduce detection rates.

Based on the results of this study, for classification of lochs according to the WFD Overall Status, August is a suitable time to conduct sample collection for eDNA-based detection of both fish and invertebrates, as the classification rate was reasonably high. However, as the fish communities detected in the current project are similar to those detected during the development of the Fish Lake Classification tool,

sampling in winter would also be likely to achieve the desired results (for fish). Given the results of this study, we can recommend late summer for using freshwater invertebrate data, but sampling at other times of year may also be possible and could potentially yield higher accuracy results. However, summer sampling may be less effective for the detection of certain fish species, such as arctic charr and coregonids, unless sampling is carried out offshore and/or at depth (Table 12) (Sellers et al. 2023). The frequency of surveying should align with existing monitoring programmes.

FKF8.5: Where do the knowledge gaps and barriers exist, and why is that? What practical solutions and technology innovations would help overcome these types of challenges going forward?
The lack of a national framework for assessing freshwater lochs using eDNA-based methods is the primary current barrier. Putting a national framework in place to monitor lochs using eDNA would mean that government, NGOs, and commercial entities could monitor loch status in Scotland.

For freshwater invertebrates, it is not clear whether all species currently used in the CPET classification index can be detected using the approach taken in this project. It is likely that not all species are currently reliably detected. To start using eDNA-based data as a CPET-like tool, further work is needed to address this challenge.

We undertook an analysis of potential pressures (nutrient/chemistry mean annual values) for which eDNA-based community data may be used to monitor. This was a limited analysis and there remain a lot of unknowns in this area.

FKF8.6: What have we (a) learned to do the same or differently and (b) anticipated are key opportunities to develop, upscale, and implement DNA-based methods for freshwater loch habitats in Scotland?
Sampling from multiple shoreline locations spread around the perimeter of a loch yields data from fish and freshwater invertebrate communities that can be used to classify lochs according to WFD statuses and has the potential to be incorporated into other metrics. The general sampling strategies and primers used in this project are adequate for monitoring at scale, although further work is required for bacteria.

An additional set of data analyses, that was not performed in this project, would be to look further into the bacterial traits that could be assigned to the eDNA-based data and to assess whether there are any further insights that can be gleaned from it. For example, are there certain bacterial functions associated with freshwater statuses or potential pressures, and can these be used to improve classification of sites? This would be a key opportunity to gain a deeper understanding of the functional ecology, redundancy, and resilience in the bacterial communities across the loch gradients. This is possible using the picrust2 pipeline (Douglas et al. 2020). This additional analysis could be carried out using the data produced in this project.

We analysed the data from each marker separately for this project. Combining the data from multiple markers could potentially yield higher classification accuracies. This would be valuable to explore. This additional analysis could be carried out using the data produced in this project.

FKF8.7: What other freshwater habitat types across Scotland do we need to consider sampling for eDNA-based biodiversity monitoring and reporting purposes in the future?
The primary other freshwater habitats to target are rivers, streams, and ponds. Observations and experiments show that eDNA can be transported from tens to thousands of metres downstream in lotic systems. eDNA appears to behave similarly to fine particulate organic matter and the distance it travels is affected by flow rate, volume, and local environmental conditions. If the location of taxa within a lotic system is important for a given monitoring program (rather than simply having observations that a

species occurs within a watercourse or catchment), then models to estimate where the eDNA originated from are essential. Only recently have attempts been made to overcome this challenge. For instance, (Carraro et al. 2020) developed and applied the eDITH (eDNA Integrating Transport and Hydrology) model, which uses hydrological first principles, to sequence read counts for 50 Ephemeroptera, Plecoptera, and Trichoptera (EPT) genera, and revealed the role of environmental covariates in driving the spatial distribution of single taxa and EPT diversity across the study catchment. The results are very promising, as the model-predicted low-diversity reaches corresponded to polluted sites. There is an opportunity to trial these models and to determine sampling strategies required to effectively apply such water bodies.

Initially this would be carried out intensively on a single catchment (multiple sampling locations at approximately 2.5 km intervals and multiple replicates per location) but it would be expected that the baseline data would inform a less intensive monitoring program. Any taxa can be targeted provided there is conventional data that could be used for verification. For example, if insects were targeted and the eDNA-based data generated was largely made up of chironomids, then follow-up conventional surveys would be carried out at numerous sites to assess the predictions made by the model. The outcome would be a sampling strategy for the taxonomic group(s) studied that would predict biodiversity (i.e., the location of taxa detected) throughout an entire lotic system, which could be repeated on an annual basis. By collecting physicochemical and hydrological data, the sampling strategy and model could also be assessed in terms of its power to detect particular impacts (such as point pollution). It is anticipated that a baseline study would need to be conducted on every new catchment being monitored.

3.3 Woodland

The woodland habitat specific research questions (WRQs) posed and the key findings (WKFs), are summarised in this section.

3.3.1 Woodland Sequencing Data Summary

Three different metabarcoding assays were applied to eDNA extracted from each soil sample: soil bacteria, soil fungi, and soil invertebrates. A total of 3,542 taxa were detected, from a total of 14,997,454 sequence reads, across the 107 samples; 1,800 bacteria, 1,572 fungi, and 170 soil invertebrates (Table 14). More fungal and invertebrate OTUs were identified at the species level compared to bacteria.

High quality bacterial sequence data were obtained for 100 of the 107 soil samples (4,466,425 sequence reads). Amplification was not successful for 7 samples despite troubleshooting steps (Table 15). In the bacterial dataset, OTUs were detected across 24 different phyla within the kingdom Bacteria. The average bacterial species richness was 281 and ranged from 164 to 445. The phylum with the highest richness of OTUs was Proteobacteria. The bacterial OTU with the most reads was from the phylum Acidobacteriota. This OTU was detected in all samples.

High quality fungal sequence data were obtained for 99 of the 107 soil samples (5,424,495 sequence reads). Amplification was not successful for 8 samples despite troubleshooting steps (Table 15). In the fungal dataset, OTUs were detected across 5 different phyla within the kingdom Fungi. The average fungal taxon richness per sample of rarefied data was 103 and ranged from 30 to 200. The phylum with the highest richness of OTUs was Ascomycota. The fungal OTU with the most reads was the family Hyaloscyphaceae.

High quality invertebrate sequence data were obtained for 97 of the 107 soil samples (5,106,534 sequence reads). Amplification was not successful for 10 samples despite troubleshooting steps (Table

15). In the invertebrate dataset, OTUs were detected across 8 different phyla within the kingdom Animalia. The average invertebrate taxon richness per sample of rarefied data was 11 and ranged from 3 to 41. The phylum with the highest richness of OTUs was Arthropoda. The OTU with the highest proportion of reads was from the family Enchytraeidae (a group of annelid worms). None of the extraction and PCR negative controls produced any target reads.

Table 14: Number of OTUs detected and the percentage of OTUs identified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	1,800	75.8%	63.5%	43.6%	31.4%	13.2%	2.4%
Fungi	1,572	99.4%	90.4%	76.8%	53.8%	32.2%	15.6%
Invertebrates	170	97.6%	86.5%	81.8%	70%	37.1%	15.3%

Table 15: Woodland sequencing data summary

Assay	N reads obtained	N samples	N samples that did not produce data	Samples that did not produce data
Soil Bacteria	4,466,425	107	7	AF22-G-04 (Glen Moriston – Unforested), CR22-G-04 (Glen Affric – Unforested), CR22-G-05 (Glen Affric – Unforested), GH22-G-01 (Glen Affric – Unforested), RE-G-03 (Cairngorms – Unforested), RE-G-04 (Cairngorms – Unforested), RE-R-03 (Cairngorms – Recently planted)
Soil Fungi	5,424,495	107	8	AF22-G-04 (Glen Moriston – Unforested), CR22-G-04 (Glen Affric – Unforested), CR22-G-05 (Glen Affric – Unforested), CR22-M-04 (Glen Affric – Mature), GH22-G-01 (Glen Affric – Unforested), GH22-R-01 (Glen Affric – Recently planted), MO-M-04 (Cairngorms – Mature), RE-G-05 (Cairngorms – Unforested)
Soil Invertebrates	5,106,534	107	10	CR22-G-04 (Glen Affric – Unforested), CR22-G-05 (Glen Affric – Unforested), CR22-M-04 Glen Affric – Mature), CR22-R-02 (Glen Affric – Recently planted), DW22-R-05 (Glen Moriston – Recently planted), GH22-G-01(Glen Affric – Unforested), IW-M-05 (Glen Moriston – Mature), RE-G-05 (Cairngorms – Unforested), RE-M-03 (Cairngorms – Mature), RE-R-03 (Cairngorms – Recently planted)

3.3.2 Woodland Research Questions & Key Findings

3.3.2.1 Summary of Woodland Habitat Key Findings

Soil sampling and subsequent eDNA metabarcoding to produce community datasets can be effectively used to monitor the recovery of soil biological communities along a gradient of woodland restoration and to also detect species of interest. This would allow restoration tracking to go beyond tree monitoring and to ensure the recovery of key taxonomic groups in soils that are associated with good condition habitats. In particular, fungal communities had the most consistent results as indicators of restoration progress, including consistent eDNA-based biodiversity metrics for fungal communities. Bacteria also had a high accuracy for assigning samples to each of the restoration classes.

Classification of samples into different woodland restoration stages based on community composition alone was found to be highly accurate for unforested and mature woodland. This was less so for recently planted, the transitional habitat, because community compositions overlapped between unforested and mature woodland which themselves were highly distinct. It is likely that recently planted sites will increase in overlap with mature woodland as succession occurs, which can be tracked to monitor woodland restoration progress. As eDNA-based data allows for the monitoring of biological

communities in soils, these findings demonstrate there is high potential to track the recovery of biological conditions of soils, which goes beyond the monitoring restoration progress based on tree growth alone.

The results of this study support that eDNA-based methods can be used to effectively establish a national baseline for woodland restoration progress across Scotland. We suggest monitoring a subset of the country's woodland restoration sites alongside comparable sites of unforested and mature woodlands to track progress in the recovery of soil biological conditions, although extension to other woodland types would require further validation. We present recommendations on rules of thumb for soil sampling to achieve this.

3.3.2.2 Research Questions & Key Findings

WRQ1: Are there any DNA-based units/metrics that can act as indicators of biodiversity and be used to (a) distinguish habitat condition and (b) detect ecological pressures or restoration stage in Scottish woodland habitats?

To address this question, we investigated whether there were clear trends evident across restoration stage for any of the following: species richness and evolutionary diversity (soil bacteria, soil fungi, soil invertebrates), bacterial and fungal functional diversity. Restoration stage was used as an equivalent to habitat condition in this habitat. We included soil pH and soil moisture content as covariates and reference to these is made where relevant. Community-based statistics and species of interest results are detailed under the WRQs that follow below.

On average, unforested sites were characterised by higher pH (mean = 4.9, SE = 0.06, 95% CI 4.76-5.0) and soil moisture content (mean = 78.3, SE = 4.77, 95% CI 65.7-91) than mature sites (pH: mean = 4.2, SE 0.06, 95% CI 4.11-4.35; soil moisture content: mean = 66.9, SE = 4.73, 95% CI: 54.2-79.6), and recently planted sites had intermediate values (pH: mean = 4.6, SE 0.06, 95% CI 4.44-4.68; soil moisture content: mean = 74.7, SE = 4.76, 95% CI: 62.0-87.4). The purpose of recording these values was only to assess whether they played a substantial role in the differences between the communities detected, rather than to assess their utility as indicators. The ranges of the pH and soil moisture values were reasonably similar across the restoration classes.

WKF1: *Fungal biodiversity metrics derived from eDNA-based data are a promising indicator of woodland restoration.*

WKF1.1: Neither soil bacteria nor soil invertebrate eDNA-based data showed clear trends across restoration stages with respect to the metrics investigated. This is unlikely to be due to a lack of taxonomic resolution, as none of the metrics used relied on taxonomy being assigned. Soil pH was correlated with bacterial species richness (R^2 = 0.26, full model R^2 = 0.3) and evolutionary diversity (R^2 = 0.27, full model R^2 = 0.31) indicating that this covariate, which is not of primary interest, plays a role in bacterial community composition. However, the range of pH was relatively narrow across all the sites, and this is not likely to have impacted the key findings. Although bacterial functional diversity was significantly different across the restoration gradient ($p = 0.01$), there was no clear trend and very little variation was explained by it (R^2 = 0.09). Soil moisture was correlated with invertebrate species richness (R^2 = 0.32, full model R^2 = 0.38) and evolutionary diversity (R^2 = 0.33, full model R^2 = 0.37), however, the range was relatively narrow across all the sites, and this is not likely to have impacted the key findings.

WKF1.2: Across the restoration gradient, there was a significant increase in fungal species richness ($p < 0.001$), evolutionary diversity ($p < 0.001$), and functional diversity ($p = 0.006$; Figure 22). Mature sites had higher values than unforested or recently planted sites, however, the differences between unforested

and recently planted sites were not significant. pH did not have a significant relationship for any of the fungal biodiversity metrics, and the range of pH was relatively narrow across all the sites and is not likely to have impacted the key findings. Fungal species richness ($R^2 = 0.08$, $p = 0.01$) and evolutionary diversity ($R^2 = 0.1$, $p = 0.008$) were negatively correlated with soil moisture content, but only weakly so, indicating that the restoration gradient is the primary factor influencing fungal communities.

WRQ2: Can eDNA metabarcoding community data from soil samples be used to assess ecosystem health (condition, function, and resilience) of Scottish woodland habitats in response to pressures such as land use, INNS, and/or climate change?

WKF2: *Based on key findings WKF1 and WKF5, fungal eDNA-based data is a promising indicator of woodland ecosystem health, and all three assays could classify site habitat condition with reasonable accuracy.*

An increase in species richness, evolutionary diversity, and functional diversity, such as observed for fungal eDNA-based data in this study, is generally equated to an increase in woodland maturity. The primary pressures related to the restoration gradient targeted in this study are land use (tree removal for timber or for agricultural use of land), grazing by deer, and the spread of non-native trees (Edwards 2006). However, we did not explicitly quantify grazing pressure or non-native vegetation in this study. Furthermore, all three assays could classify site habitat condition with reasonable accuracy (see WRQ5), with bacteria and fungi outperforming the invertebrate assay.

WRQ3: Can the DNA-derived metrics extend from taxonomic descriptions to assess possible pressure or restoration-related changes in Scottish woodland habitats?

WKF3: *Fungal evolutionary diversity and functional diversity both increased along the restoration gradient investigated in this project (unforested, recently planted/regenerating, and mature) (Figure 22).*

Furthermore, all three assays could classify site woodland maturity class with reasonable accuracy (WRQ5), with bacteria and fungi outperforming the invertebrate assay.

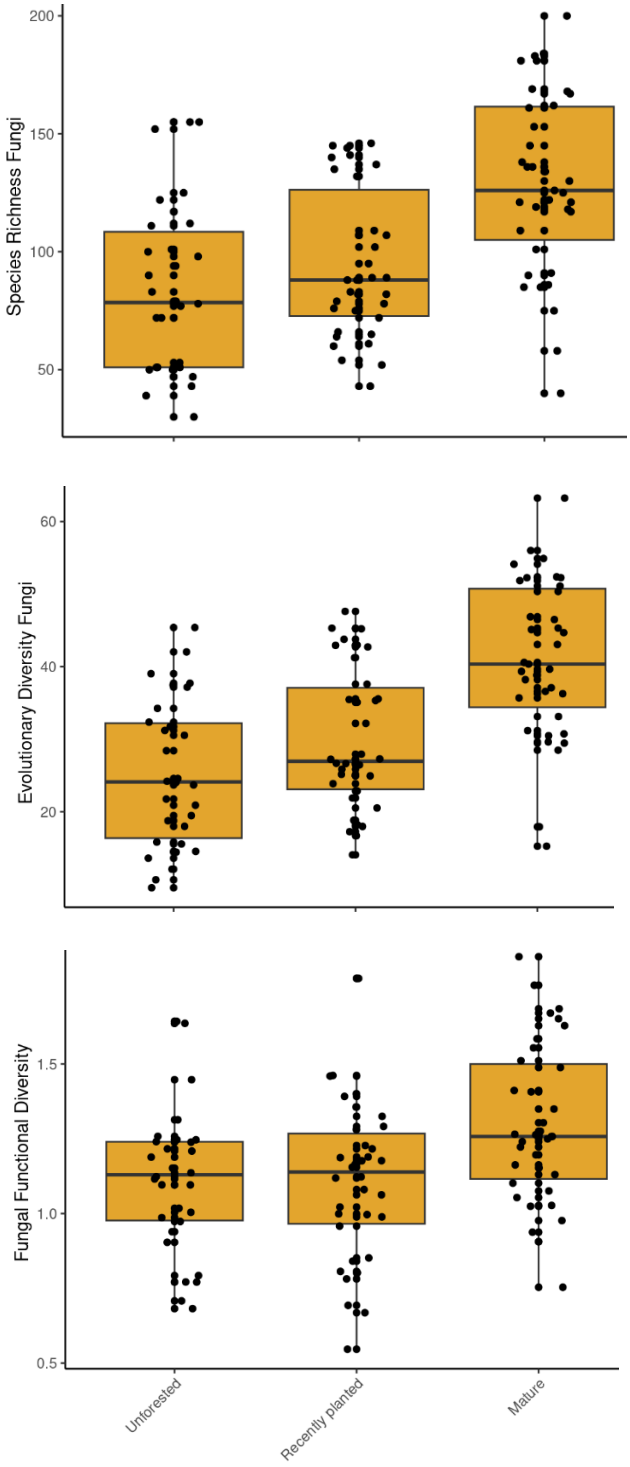


Figure 22: There is a significant positive effect of restoration gradient on woodland soil fungi species richness ($p < 0.001$), evolutionary diversity ($p < 0.001$), and functional diversity ($p = 0.006$)

These metrics do not necessarily rely on taxonomy to be assigned. It should be noted that the method we employed to generate fungal functional diversity measures did require taxonomy to be assigned first, however, alternative methods can be performed that are taxonomy-free, such as using the ITS gene function matching database FungalTraits (Põlme et al. 2020).

WRQ4: Can eDNA metabarcoding community data provide similar and reliable outcomes as conventional morphology-based or scoring-based methods for soils to assess the health status of Scottish woodland habitats, using best available matching data examples?

WKF4: *Currently we cannot answer this question as we did not have available matching conventional morphology or scoring based-based methods for the woodland habitat. However, the fungal assay did detect numerous SSSI species that are targeted in conventional surveys.*

We detected numerous fungal species that are of key interest for conventional survey methodologies. This includes two species listed as Endangered, and four species listed as Vulnerable on the IUCN Red List (Table 16). We also detected 27 SSSI-listed (Sites of Special Scientific Interest) fungal species (Table 17). Fungal species identified were checked against the JNCC Guidelines for the Selection of Biological SSSIs Chapter 14 'Non-lichenised fungi' (Bosanquet et al. 2018). The presence of such rare, threatened, or otherwise important fungal species in an area indicates the presence of a unique or valuable ecosystem which can contribute to its designation as an SSSI. Fungi are normally surveyed for by skilled ecologists/mycologists in autumn during fruiting body walkover surveys.

Table 16: Fungal species detected which have IUCN designated statuses.

Species	Common name	IUCN status
Trichoglossum walteri	Short-spored earthtongue	VU
Entoloma prunuloides	Mealy pinkgill mushroom	VU
Cuphophyllus lacmus	Grey waxcap	VU
Gloioxanthomyces vitellinus	Glistening waxcap	EN
Pseudotricholoma metapodium	Mealy meadowcap	EN
Cuphophyllus flavipes	Yellow foot waxcap	VU

Table 17: Fungal species detected that may contribute to the selection of SSSIs.

Restoration Status	Site	SSSI listed species detected	N of SSSI species
Unforested	Rothiemurchas	*Clavaria flavipes* *, *Clavaria zollingeri* *, *Clavulinopsis helvola* *, *Clavulinopsis laeticolor* *, *Cuphophyllus flavipes*^, *Gliophorus irrigatus*[§], *Hygrocybe coccinea*[§], *Hygrocybe insipida*[§], *Hygrocybe reidii*[§]	9
	Dundreggan WGS	*Russula paludosa*[±]	1
	Coille Coire Chuilc	*Hygrocybe miniata*[§]	1
	Dundreggan Allt Fearna	*Hygrocybe cantharellus*[§]	1
	Coille Ruigh		0
	Glen Falloch		0
	Ghubhais		0
Recently Planted	Glen Falloch	*Trichoglossum walteri*[Δ], *Clavulinopsis helvola* *, *Gloioxanthomyces vitellinus*[§], *Hygrocybe cantharellus*[§], *Hygrocybe reidii*[§]	5
	Ghubhais	*Russula paludosa*[±]	1
	Dundreggan Allt Fearna	*Suillus flavidus*[¥]	1
	Rothiemurchas	*Cortinarius gentilis*[±], *Cuphophyllus lacmus* ^	2
	Glenmore	*Clavaria zollingeri* *, *Hygrocybe coccinea*[§]	2
	Coille Ruigh		0
	Dundreggan WGS		0
Mature	Glen Falloch Mature	*Trichoglossum walteri*[Δ], *Clavulinopsis helvola* *, *Cortinarius subtortus*[±]	3
	Coille Coire Chuilc	*Cortinarius caperatus*[¥], *Russula decolorans*[¥], *Russula paludosa*[±]	4
	Coille Ruigh	*Cortinarius caperatus*[¥], *Russula decolorans*[¥], *Russula paludosa*[±]	3
	Ghubhais	*Cortinarius limonius*[±], *Russula decolorans*[¥], *Russula paludosa*[±]	3
	Inverwick	*Cortinarius gentilis*[±], *Russula integra*[±], *Russula vinosa*[¥]	3
	Rothiemurchas	*Tricholoma portentosum*[±], *Russula paludosa*[±]	2
	Glenmore	*Cortinarius gentilis*[±], *Cortinarius scaurus*[±], *Suillus flavidus*[¥], *Russula paludosa*[±]	4

*Clavarioid fungi for scoring (grassland), ^Grassland waxcap high diversity indicator, [§]Grassland waxcap, [±]Caledonian pinewoods additional species of interest, [Δ]Earthtongue for scoring (grassland), [¥]Caledonian pinewood scoring species.

WRQ5: Are eDNA metabarcoding data from soil samples broadly similar or different in biodiversity and community composition amongst Scottish woodland habitats of ranging ecological status and restoration stage (unforested, recently planted/reforested, and mature condition categories)? If we find any differences, are these data sufficiently clear and predictable to be used to distinguish (classify or sort) between soil samples from Scottish woodland habitats of ranging ecological status and restoration stage (unforested, recently planted/regenerating, and mature condition categories)?

WKF5: *Community compositions across all assays were found to be broadly different among woodland restoration stage, and samples could be classified according to restoration stage with reasonable accuracy.*

WKF5: In general, the patterns of community composition were as would be expected for all three assays, with communities transitioning from Unforested to Recently Planted to Mature (Figure 23 and Figure 24). This is encouraging, because if such a pattern was not evident it would mean that using eDNA-based community data to classify woodland habitat condition would not be feasible. This is despite samples being collected from relatively distant locations. For example, bacterial communities from unforested sites in Dundreggan are very similar to unforested communities in Glen Falloch, although the sites are situated c. 90 km apart. Similarly, communities from mature woodland in Rothiemurchus are very similar to mature communities in Coille Ruigh although the sites are situated c. 70 km apart. However, it is important to note that after sampling sites are included in the analyses, the trends across the restoration gradient become more evident, and geographic area does play a substantial role in community composition. This becomes evident in the difference between the unconstrained and constrained ordinations.

Bacteria showed the clearest separation of the restoration gradient with 72% of the samples being correctly classified, followed by fungi (69%) and soil invertebrates (55%) (see Phase 2 Technical Appendices (Bakker et al. 2023c) for further statistical output). There are three condition classes, so random classification would achieve a predictive accuracy of ~33%. The hardest samples to classify are those from Recently Planted sites, as the communities overlap with the other restoration classes. There was no clear trend for samples from Recently Planted sites being classified more often as Unforested or Mature. This is most likely due to the higher variation in the sites in the Recently Planted class. See Phase 2 Technical Appendices output (Bakker et al. 2023c) for site photos. Considering only mature and unforested sites, samples could be classified correctly by bacteria with 83% accuracy, fungi with 74% accuracy, and soil invertebrates with 72% accuracy.

This means that for these study sites, if new samples were collected and, for example, analysed using the bacteria assay, they would have an 83% chance of being classified into the correct restoration class.

This demonstrates that if a baseline is established across Scotland's Caledonian pinewoods, using known highly semi-natural (mature) sites as the target reference condition, then samples can be collected and classified with reasonable accuracy to assess and track progress. This is further discussed in WRQ7.

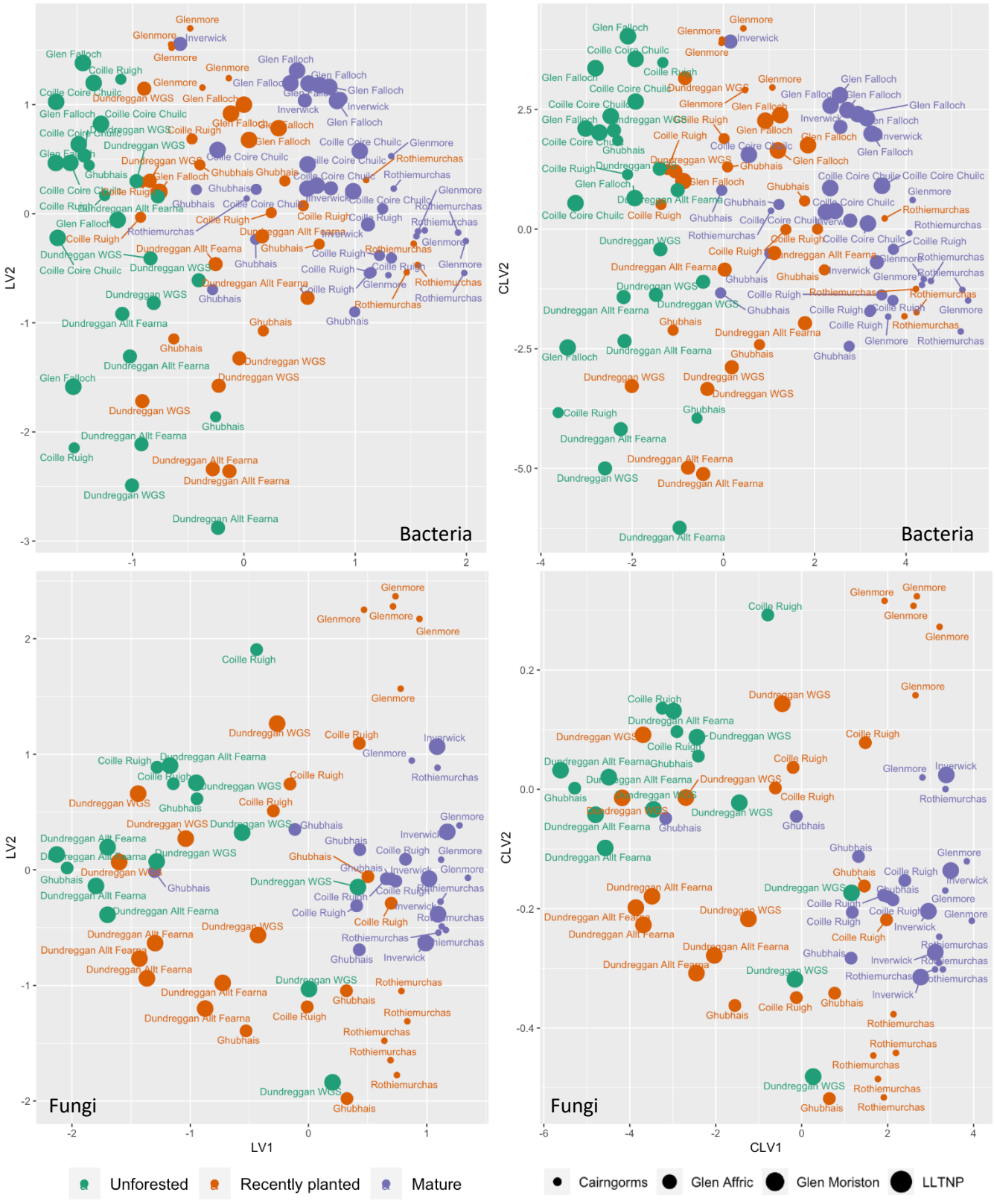


Figure 23: Unconstrained (left) and constrained/concurrent (right) ordinations for soil bacteria and fungi. Each point is a sample and samples that are closer together are more similar. In both models, there is a separation between unforested and mature conditions, with recently planted forming and intermediate stage.

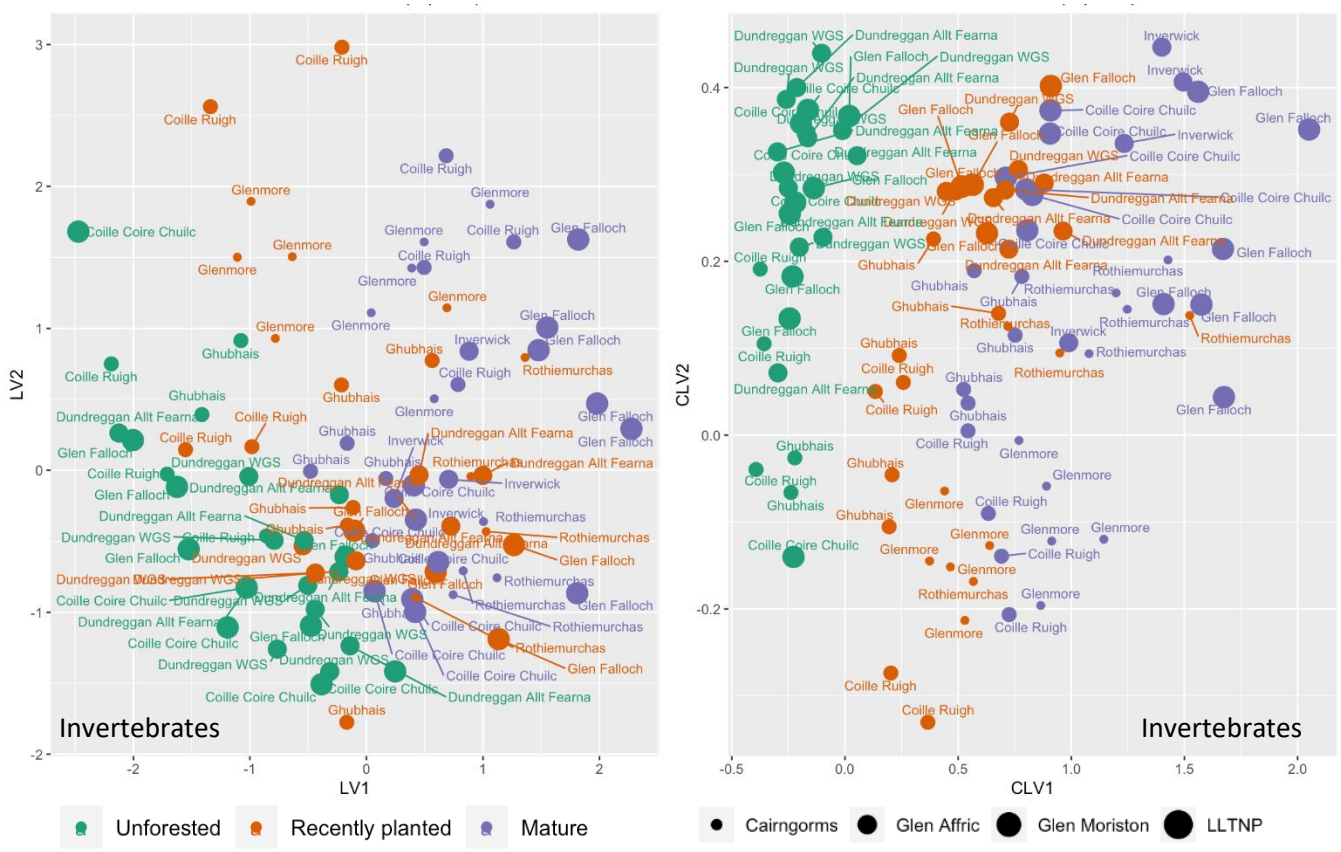


Figure 24: Unconstrained (left) and constrained/concurrent (right) ordinations for soil invertebrates. Each point is a sample and samples that are closer together are more similar. In both models, there is a separation between unforested and mature conditions, with recently planted forming and intermediate stage.

Geography (LLTNP, Glen Affric, Glen Moriston, Cairngorms) was included as a full-rank predictor, and the concurrent ordinations are thus based on the residuals after this effect of geography was removed. Concurrent latent variable 1 (CLV1) is highly significantly correlated with condition (unforested, recently planted, mature, $p<0.001$) for bacteria and invertebrates, and condition explains 5.2% (bacteria) and 58.3% (invertebrates) of the variation explained by the two latent variables (partial R^2 analysis). Note that although there is evident sorting by condition in the soil fungi dataset, some of the unforested samples are on the right, and one of the mature samples is on the left.

It should be noted that constrained ordination revealed that geography (sites located in four areas: LLTNP, Glen Moriston, Glen Affric, Cairngorms) nonetheless had a significant effect on community composition. The potential implication of a geographic effect is that distinct areas (like the Cairngorms) could require training area-specific classification models.

WRQ6: Are there any differences or similarities in habitat condition or biodiversity and community composition from any re-surveyed locations between timescales (e.g., seasons or years) and protected status (e.g., designated vs non-designated sites) for the Scottish woodland habitats sampled?

WKF6: *This project did not involve resurveying the same woodland locations across seasons or years and did not have a balanced design to compare designated vs non-designated sites. This is because the pilot study results guided a change in research strategy and the choice of a balanced design with respect to restoration classes was a priority, as opposed to including as many sites as possible that already had assigned designations.*

All the woodland areas in this study have SSSI status, apart from Glen Moriston. This precluded comparison of SSSI vs non-SSSI status sites as any observed effect would be impossible to disentangle from geographic effects. If detection of a signal caused by SSSI status is of interest for future work, this would require a comparison of SSSI and non-SSSI sites that are geographically close together.

This project did not involve resurveying the same locations across seasons or years. The Phase 1 pilot study was carried out in different locations (in October 2021) with the initial focus on comparing deer exclosure vs deer-grazed woodland plots, but the results indicated that the eDNA-based data signals would not be strong enough to detect significant differences of this effect.

Sampling was conducted in late summer/autumn for both the main sampling campaign and the pilot study. Both sampling events yielded high numbers of taxa being detected across all three assays. Only two SSSI-listed fungal species were detected in the pilot study (*Hygrocybe miniata* and *Hygrocybe cantharellus*). It is not possible to determine whether this was an effect of year or season, given that the sampling locations we not the same.

WRQ7: What are the key recommendations and future perspectives to develop and operationalise standardised eDNA-based methods at national scale for monitoring biodiversity and the impacts of pressures or restoration in Scottish woodland habitats?

WKF7.1: How should we sample?

Soil augers are a common approach in soil sample collection for eDNA-based surveys, and this project supports the fact that the approach is suitable for detecting high numbers of species. Using plots (such as in this project) rather than transects allows for the easy integration with satellite GIS data layers. There is variation in the literature in choosing plot sizes. We used 10 m x 10 m plots whereas for example Griffith et al. (2022), opted for 30 m x 30 m to match existing plots used for fungal fruiting body surveys. The primary benefit of smaller (10 m x 10 m) plots is that they are easier to measure out and easier to place within difficult terrain or patchy habitat. A potential downside of smaller plots is that repeat monitoring will need to ensure that plots, and particularly exact sampling locations, are not unduly impacted during surveys. Care should be taken to minimise trampling of vegetation, and sampling locations within plots should be close to previous locations, but not in exactly the same places (e.g. 2-3 m away).

Collecting multiple cores within a plot to ~10 cm depth, combining the subsamples in a plastic bag, keeping them chilled in a cool box and using refrigerated transport to the lab, generally aligns with the literature (e.g. Griffith, Detheridge, and Bye 2022; M. C. Allen et al. 2023). Freezing the soil samples as soon as possible after collection and processing the samples in the laboratory within one month is highly recommended (Kirse et al. 2021a). One limitation of sampling to a standardised depth is that different layers/horizons might be sampled. For example, organic horizons only develop in more mature woodlands and are likely not present in grassland. This can have a significant impact on the species to be found. Ideally, horizons present and their depths should be recorded, so these can be considered when interpreting the results.

WKF7.2: What should we sample?

This study focussed on soil samples. While there is potential to survey water bodies within woodlands, these are not always present and linking the results from aquatic surveys to particular land parcels is difficult. For surveying invertebrates, other sample types such as pitfall traps, malaise traps (Kirse et al. 2021b) and even leaf samples (Krehenwinkel et al. 2022) can be collected and processed using DNA metabarcoding. These generally require additional resources such as multiple visits and additional equipment. However, when surveying for above ground and flying invertebrates, these methods are

more appropriate. The collection and processing of airborne eDNA is a recent development in the molecular analyses of terrestrial biodiversity, including plant (pollen) (Korpelainen and Pietiläinen 2017; Johnson, Cox, and Barnes 2019), fungi (Banchi et al. 2020; Rosa et al. 2020), insect (Roger et al. 2022), and vertebrate (Clare et al. 2022; Klepke et al. 2022) communities. However, alongside practical methodological advancements, gaps in technological knowledge need to be addressed before airborne eDNA can be integrated into routine biodiversity monitoring. An advantage of using soil samples is being able to survey for multiple taxonomic groups such as bacteria, fungi, and invertebrates simultaneously. Soil samples can also be used to survey for terrestrial mammals (Mena et al. 2021). Although we did not trial this in this study, NatureMetrics have used this approach to detect mice, shrews, voles, deer, and badgers (unpublished data).

For general habitat condition assessment and for classifying sites to a maturity status along woodland restoration gradients (for Caledonian pine woodlands), fungi and bacteria should be targeted. . Soil invertebrates did show a community shift, but the data were not as robust for classifying Scots pine maturity status. This may differ for other woodland habitats.

WKF7.3: Where should we sample?
Soil sampling is subject to high local variability in species and communities detected, so it is essential that long term monitoring plots are established for repeated surveys in order to detect habitat condition status and temporal changes. Monitoring plots for eDNA-based methods should align with existing monitoring plots where possible, such as the National Forestry Inventory monitoring plots.

Site-based monitoring

To monitor temporal changes and condition improvements at a particular site, a similar strategy to the approach used in this study can be employed, whereby multiple plots are established in each treatment. In this context, treatments should ideally include a control (i.e. unforested area), a target (i.e. an area of mature woodland), and the restoration site of interest (i.e. a planted or grazing-exclosed area), and the plots should be representative of the treatment. Using such a before-after-control-impact (BACI) design provides the most robust method to track progress at a given site, particularly to account for other general factors, such as weather conditions or climate change. There is little guidance on how many samples or plots should be used for eDNA-based soil monitoring within each treatment. In this study, 15 samples per site (5 samples per treatment) was sufficient to attain reasonable accuracy in habitat condition classification.

For example, the Glenfalloch site was approximately 120 ha and classification had 94% accuracy using the bacteria assay (1 out of 16 samples was incorrectly classified). The Coille Ruigh site was approximately 21 ha and had 77% accuracy. The Ghubhais site was approximately 80 ha and only had an accuracy of 58%. accuracy. Therefore, on balance, in the context of habitat condition monitoring at the site level for Caledonian pine woodlands, we would not suggest going below 5 samples per treatment for sites up to ~150 ha, and increasing the number of samples per treatment would likely improve accuracy. Site variability, in particular for the Recently Planted/Regenerating categories, is likely to play an important role in classification accuracy. Where there is a high variability within sites, sites should be subdivided into reasonably similar habitat types.

Based on the Native Woodland Survey of Scotland (Forestry Commission Scotland 2014), there are an estimated ~88,000 ha of native Scots pine in Scotland of which 22% was estimated to be highly semi-natural. Sampling at the site-based intensity would require ~2,900 samples to cover the whole area and ~640 samples to cover the highly semi-natural woodland, costing ~£1.2 million or ~£255,000 respectively for eDNA-based analysis.

Establishing a national baseline

Another approach would be to establish a national baseline comprising a percentage of the area of native Scots pine in Scotland as well as adjacent or nearby unforested areas, and then using that baseline to compare site-based restoration projects against. For example, if a baseline was constructed by sampling 10% of the native Scots pine area and an equal area unforested land, which would cost ~£240,000, this could form a national monitoring framework for Scots pine restoration projects. Projects could then sample at the site-based intensity and their sites could be classified as unforested, mature, or intermediate with reasonable accuracy. An increase in the percentage of samples being classified as mature would be a reasonable indicator that restoration was progressing. By including highly semi natural sites in the baseline, this could also form a highly desirable category within the mature restoration class. The baseline would need to be repeated every ~5 years (see below) to account for any shifts in communities that could be caused by climate change. It should be noted that with such a framework in place, site-based monitoring would not require a full BACI design, as control and target sites would no longer be needed, reducing the cost of monitoring these sites.

An example of a conceptually similar baseline model was carried out by Wilhelm et al. (2022). They used similar classification methods to predict soil health from 16S bacteria metabarcoding data, using a final dataset of 598 soil samples collected from farmlands across the USA and Canada. The authors note that regional scale or cropping-specific models could be valuable for improving their prediction accuracy.

Of course, these are indicative sample numbers only and a more comprehensive desk study that accounts for the national targets for Scots pine woodland restoration would need to be undertaken to fully conceptualise the sampling strategy for a national baseline. What we have shown in this project is that this would be a scientifically feasible strategy.

WKF7.4: When should we sample?
Regarding SSSI designated fungi, conventional surveys are limited to the autumn months when fruiting bodies are present. However, as the JNCC guidelines point out; 'DNA-based detection of fungi in environmental samples, such as roots, soil, and water, highlights that fruitbody recording only provides a partial picture of fungal distribution and that some fungi rarely ever fruit' (Bosanquet et al. 2018). As eDNA-based methods do not rely on fruiting bodies for fungal detection, there is the potential to survey at other times of the year. Based on the results of this study, August through September is a suitable time to conduct soil sample collection for eDNA-based detection of both fungi and bacteria, as the classification rate was reasonably high. However, as bacteria and fungi communities do vary seasonally (Ma et al. 2021), it is essential that monitoring is carried out in the same season each year. Given the results of this study, we can recommend late summer and autumn, but sampling at other times of year may also be possible and could potentially yield higher accuracy results. Griffith et al. (2022) also used eDNA metabarcoding to survey for CHEGD fungi in the Peak District National Park and identified 137 CHEGD species and 11 IUCN threatened species during late summer/autumn.

The frequency of surveying should align with existing monitoring programmes. While annual sampling is perhaps the ideal, woodlands generally do not change dramatically on an annual basis (unless affected by extreme events such as flooding or burning). However, to monitor restoration progress the intervals between sampling events should not be so long that adaptive management cannot be undertaken in a timely manner. Conducting eDNA sampling for the monitoring of restoration sites every 3-6 years would likely be appropriate.

WKF7.5: Where do the knowledge gaps and barriers exist, and why is that? What practical solutions and technology innovations would help overcome these types of challenges going forward?

The lack of a national framework for assessing woodlands using eDNA-based methods is the primary current barrier. Putting a national framework in place to monitor woodlands (as outlined above) would mean that government, NGOs, and commercial entities could monitor their restoration progress of woodlands in Scotland using eDNA-based methods.

For fungal surveys, it is not clear whether all SSSI species can be amplified using the primers used in this study and whether the current sampling strategy would detect all SSSI species known to be present in a site. It is also not documented whether eDNA sampling outside of the conventional survey season can detect these species. As Griffith et al. (2022) pointed out, even if eDNA-based data does not become accepted as a method to designate SSSI sites, eDNA-based methods are likely to be transformative for fungal conservation since it permits "eDNA-guided surveying", informing follow up surveys. A suitable follow up study would be to conduct concurrent co-located fruiting body and eDNA-based surveys using multiple fungal primers, and to conduct out-of-season surveys at the same sites. The study should include an analysis of available fungal barcodes in public databases. Potential primer candidates would be those used by Tedersoo et al. (2014). It should be noted that eDNA-based methods have the potential to detect more species than conventional survey methods, which, when implemented as a standard monitoring method, would require a revision of the thresholds currently used to assign a site as SSSI (Bosanquet et al. 2018).

For invertebrates, other studies have noted that COI (cytochrome c oxidase subunit I) markers are more suitable for detecting community shifts. These were not trialled in this habitat. We would suggest that any remaining DNA from this project be subjected to these primers to assess the differences as this may increase the power of using invertebrate eDNA metabarcoding for community-based approaches and potentially generate higher number of species level assignments (Kirse et al. 2021a; 2021b).

WKF7.6: What have we (a) learned to do the same or differently and (b) anticipated are key opportunities to develop, upscale, and implement DNA-based methods for woodland habitats in Scotland?
The general sampling strategies and assays used in this project are adequate for monitoring at scale. As noted above, there are likely improvements to be explored around assays to be used for invertebrates, and potentially enhancements to fungal assays to detect the full suite of species of interest.

An additional set of data analyses, that was not performed in this project, would be to look further into the bacterial and fungal traits that could be assigned to the eDNA-based data and to assess whether there are any further insights that can be gleaned from it. For example, are there certain bacterial functions or fungal feeding guilds associated with woodland habitat conditions and can these be used to improve classification of sites? This would be a key opportunity to gain a deeper understanding of the functional ecology and redundancy and resilience in the communities across the restoration gradient. This is possible using databases such as FunGuild (Nguyen et al. 2016), FungalTraits (Põlme et al. 2020) and the picrust2 pipeline (Douglas et al. 2020). This additional analysis could be carried out using the data produced in this project.

WKF7.7: What other woodland habitat types across Scotland do we need to consider sampling for eDNA-based biodiversity monitoring and reporting purposes in the future?
The Native Woodland Survey of Scotland lists upland birchwoods, native pinewoods, wet woodland, lowland mixed deciduous woodland, upland oakwoods, upland mixed ashwoods and native woodland scrub as the native woodland types in Scotland. In order to consider sampling these habitats, a similar study to the current project would need to be carried out in each woodland type to assess the efficacy of classification rates and species of interest detected.

3.4 Peatland

The peatland habitat specific research questions (PRQs) posed and the key findings (PKFs), are summarised in this section.

3.4.1 Peatland Sequencing Data Summary

Three different eDNA metabarcoding assays were applied to DNA extracted from each soil sample: bacteria, fungi, and invertebrates. A total of 1,850 taxa were detected, from a total of 5,512,482 sequence reads, across the 50 samples; 934 bacteria, 801 fungi, and 115 invertebrates (Table 18). More fungal and invertebrate OTUs were identified at the species level compared to bacteria.

High quality bacterial sequence data were obtained for 49 of the 50 soil samples (2,100,161 sequence reads). Amplification was not successful for 1 sample despite troubleshooting steps (Table 19). In the bacterial dataset, OTUs were detected across 20 different phyla within the kingdom Bacteria. The average bacterial taxon richness per sample was 224 and ranged from 152 to 357. The phylum with the highest richness of OTUs was Proteobacteria. The bacterial OTU with the most reads was from the phylum Acidobacteriota. This OTU was detected in all 49 successfully sequenced samples.

High quality fungal sequence data were obtained for all soil samples (1,621,976 sequence reads) (Table 19). In the fungal dataset, OTUs were detected across 5 different phyla within the kingdom Fungi. The average fungal taxon richness per sample was 86 and ranged from 25 to 238. The phylum with the highest richness of OTUs was Ascomycota. The fungal OTU with the most reads was from the phylum Ascomycota.

High quality invertebrate sequence data were obtained for all soil samples (1,790,345 sequence reads) (Table 19). In the invertebrate dataset, OTUs were detected across 8 different phyla within the kingdom Animalia. The average invertebrate taxon richness per sample was 14 and ranged from 2 to 45. The phylum with the highest richness of OTUs was Arthropoda. The OTU with the highest proportion of reads was from an Annelid worm species (*Cernosvitoviella atrata*).

None of the extraction and PCR negative controls produced any target reads.

Table 18: Number of OTUs detected and the percentage of OTUs identified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	934	75.1%	62.6%	41.8%	27.8%	10.2%	2%
Fungi	801	99%	89.5%	76.7%	47.4%	26.2%	12%
Invertebrates	115	97.4%	77.4%	77.4%	73.9%	47%	22.6%

Table 19: Peatland sequencing data summary.

Assay	N reads obtained	N samples	N samples that did not produce data	Samples that did not produce data
Soil Bacteria	2,100,161	50	1	GF22-D-01 (Glen Finglas – Degraded)
Soil Fungi	1,621,976	50	0	
Soil Invertebrates	1,790,345	50	0	

3.4.2 Peatland Research Questions & Key Findings

3.4.2.1. Summary of Peatland Habitat Key Findings

The use of soil sampling and eDNA metabarcoding would allow restoration tracking to go beyond vegetation monitoring and to ensure the recovery of key taxonomic groups in soils that are associated with good condition habitats. Soil fungi alpha diversity measures (fungal species richness, evolutionary diversity, and functional diversity) showed a difference between degraded and restored samples; they were all lower in restored peatland. Fungi are important decomposers, playing a key role in breaking down dead organic matter and recycling nutrients back into the soil. As such, they are sensitive to changes in soil quality and moisture levels, and are influenced by factors such as soil compaction, nutrient availability, and changes in temperature and precipitation patterns. We do note that lower alpha diversity does not mean 'worse' in any way, since one of the primary goals of rewetting peatland is to reduce carbon emissions, not to support higher biodiversity *per se*.

In contrast, community analyses could not reliably classify samples to degraded and restored classes. However, the peatland dataset is small with only two restored locations. Therefore, we trained the classification model to recognise restored peatland with just one location's samples, which meant that the model had no way to distinguish species that indicated restoration status *per se* from species that were indicators of that location *per se*.

Before a national baseline could be suggested (such as for woodland), further work is needed to identify peatland sites that are true target conditions that Scotland can aim towards. In this project, despite multiple stakeholder consultations, identifying such sites within the LLTNP and surrounding area was not possible. Future work should focus on identifying sites outside of this area that have the conditions required as being as near pristine as possible so that a national baseline can be set.

3.4.2.2. Research Questions & Key Findings

PRQ1: Are there any DNA-based units/metrics that can act as indicators of biodiversity and be used to (a) distinguish habitat condition and (b) detect ecological pressures or restoration stage in Scottish peatland habitats?

To address this question, we investigated whether there were clear trends evident across restoration stage for any of the following: species richness and evolutionary diversity (soil bacteria, soil fungi, soil invertebrates), and bacterial and fungal functional diversity. Restoration stage was used as an equivalent to habitat condition in this habitat. We included soil pH and soil moisture content as covariates and reference to these is made where relevant. Community-based statistics and species of interest results are detailed under the PRQs that follow below.

Degraded and restored sites did not have significantly different pH levels (degraded: mean = 4.9, SE = 0.1, 95% CI 4.56-5.19; restored: mean = 4.8, SE = 0.13, 95% CI 4.41-5.1; p = 0.4). Although soil moisture was, on average, higher in restored sites, this was not significant (degraded: mean = 88.9, SE = 1.03, 95% CI 85.5-92.3; restored: mean = 91.5, SE = 1.29, 95% CI 87.9-95.0; p = 0.054)

PKF1: *Fungal biodiversity metrics derived from eDNA data are a promising indicator of peatland restoration.*

PKF1.1: Fungi species richness ($p < 0.001$), evolutionary diversity ($p < 0.001$), and functional diversity ($p < 0.001$) were significantly different between the two peatland conditions (Figure 25) and may function as an indicator for peatland soil health. However, all fungal biodiversity metrics were found to be lower in restored habitats, and the reason for this is unclear. Neither soil pH nor soil moisture content showed

a significant relationship with any of the fungal metrics, indicating that these covariates were of lesser influence than the primary factor of interest, habitat condition. This is not surprising in this dataset given the lack of a significant difference for these covariates in the sites sampled. However, general relationships between pH, soil moisture, and fungal metrics may exist and we do not attempt to generalise this statement to other sites. The figures for bacteria and invertebrates results are shown separately in the Phase 2 Technical Appendices output (Bakker et al. 2023c).

PKF1.2: Bacterial evolutionary diversity was significantly different between the two peatland conditions ($p = 0.03$) (Figure 26). However, the effect size was relatively small (an average difference of 1.56 units (range of all samples = 10.29 to 21.45). Bacterial species richness and bacterial functional diversity showed no such trends. Similar to woodland, of the parameters included in the model, soil pH explained most of the variation in bacterial species richness ($R2 = 0.39$, full model $R2 = 0.46$) and evolutionary diversity ($R2 = 0.41$, full model $R2 = 0.53$), indicating that pH, which is not of primary interest, plays a substantial role in bacterial community composition. Note that the model for functional diversity failed and was not interrogated further.

Soil invertebrates showed no significant trends for species richness or evolutionary diversity across habitat condition, pH, or soil moisture content.

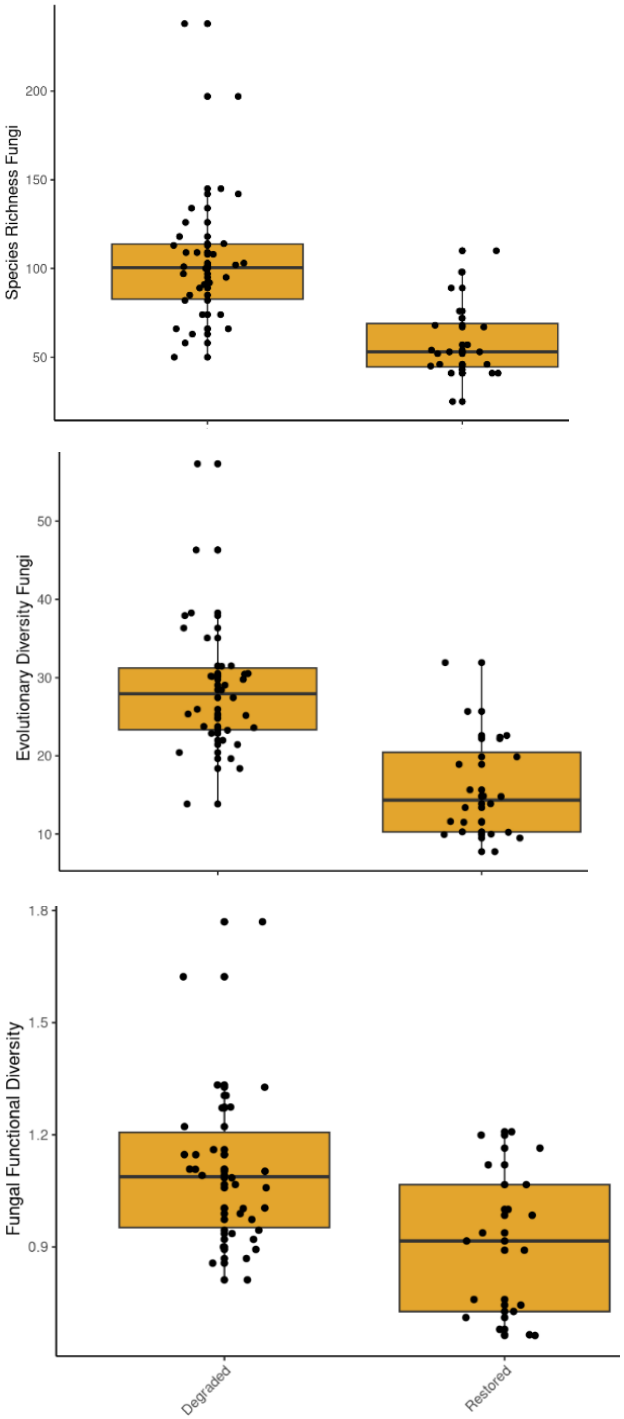


Figure 25: There is a significant effect of peatland condition on soil fungi species richness, evolutionary diversity, and functional diversity. All three metrics show a decrease from degraded to restored condition.

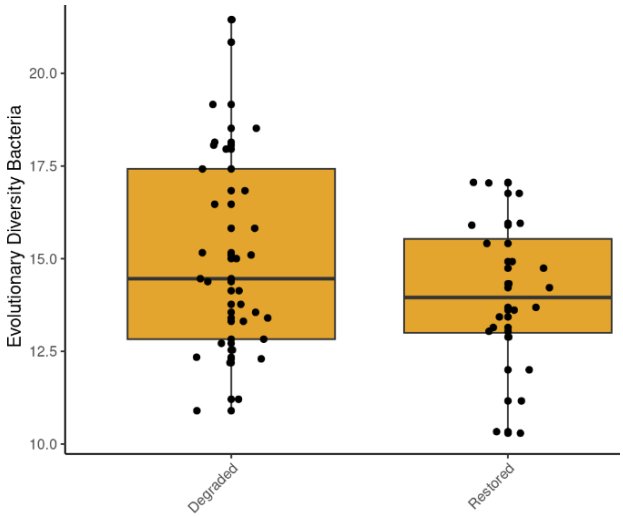


Figure 26: There is a significant, but minor, effect of peatland condition on soil bacterial evolutionary diversity. The metric decreases from degraded to restored condition.

PRQ2: Can eDNA metabarcoding community data from soil samples be used to assess ecosystem health (condition, function, and resilience) of Scottish peatland habitats in response to pressures such as land use, INNS, and/or climate change?

PKF2: *Fungal eDNA-based data is a promising indicator for peatland restoration tracking (degraded and restored sites, based on rewetting) but the direction of the trend is not what might be expected.*

An increase in species richness, evolutionary diversity, and functional diversity, such as that observed for fungal eDNA-based data in this study, is generally equated to an increase in ecosystem condition. However, the results of this project indicate that while differences might exist for fungal eDNA-based data between drained and restored sites, we do not yet have a full picture of what the expected trends should be. The primary pressures related to the restoration gradient targeted in this study is land use (drainage). It is possible that the time since drain blocking, and the level of disturbance associated with restoration activities, could have impacted the results and should be accounted for in future studies. Furthermore, surface soil moisture (rather than soil moisture calculated from composite core samples) and/or water table depth can have strong confounding effects on community composition, and we would recommend including this measurement in future studies.

However, as we were not able to classify site habitat condition accurately across all the sites (see PRQ5), further work is needed to be able to answer this question.

PRQ3: Can the DNA-derived metrics extend from taxonomic descriptions to assess possible pressure or restoration-related changes in Scottish peatland habitats?

PKF3: *Fungal evolutionary diversity and functional diversity both respond to the land-use pressure of peatland draining (and by extension the restoration effect of rewetting) and have the potential to be used to assess changes resulting from these impacts.*

These metrics do not necessarily rely on taxonomy to be assigned. It should be noted that the method we employed to generate fungal functional diversity measures did require taxonomy to be assigned first, however, alternative methods can be applied that are taxonomy-free, such as using the ITS gene function matching database FungalTraits (Põlme et al. 2020).

PRQ4: Can eDNA metabarcoding community data provide similar and reliable outcomes as conventional peatland morphology-based or scoring-based methods for soils to assess the health status and restoration stage of Scottish peatland habitats, using best available matching data examples?

PKF4: *Currently we cannot answer this question as we did not have available matching conventional morphology data for the peatland habitat.*

Peatland condition assessment was carried out at Auchlyne and Glen Finglas. While most of the assessments aligned with our classification as being restored or degraded, it was noted that in some cases the vegetation indicated that some "restored" locations are still classified as moderately degraded. This was the case for one location in Glen Finglas and three locations in Auchlyne. In fact, only two locations were classified as near natural. We suggest that future work conducts concurrent co-located vegetation assessments using this method and eDNA-based sampling to obtain a more balanced study design.

However, we did detect a few fungal species that are of key interest. This includes one species listed as Endangered on the IUCN Red List (Table 20). We also detected multiple Sites of Special Scientific Interest (SSSI) listed fungal species (Table 20). The fungal species identified were checked against the JNCC Guidelines for the Selection of Biological SSSIs Chapter 14 'Non-lichenised fungi' (Bosanquet et al. 2018). The presence of rare, threatened, or otherwise important fungal species in an area indicates the presence of a unique or valuable ecosystem and can contribute to its designation as an SSSI. However, we note that peatland is not a habitat covered by the JNCC Guidelines (Bosanquet et al. 2018) and the species detected are all grassland indicators.

Table 20: An overview of detected fungal species, including one IUCN endangered species.

Restoration Status	Site	SSSI listed species detected	N of SSSI species
Degraded	Glen Finglas	*Geoglossum barlae*[ᴬ], *Geoglossum fallax*[ᴬ], *Hygrocybe cantharellus*[§]	3
	Auchlyne	*Gloioxanthomyces vitellinus*[§∞], *Hygrocybe cantharellus*[§]	2
	Cashel	*Hygrocybe cantharellus*[§]	1
Restored	Auchlyne	*Hygrocybe cantharellus*[§]	1
	Glen Finglas	*Hygrocybe cantharellus*[§]	1

[§]Grassland waxcap, [ᴬ] Earthtongue for scoring grassland, ∞Listed as Endangered on the IUCN Red List

PRQ5: Are eDNA metabarcoding data from Scottish soil samples broadly similar or different in biodiversity and community composition amongst peatland habitats of ranging ecological status and restoration stage (degraded, restored, and good condition categories)? If we find any differences, are these data sufficiently clear and predictable to be used to distinguish (classify or sort) between soil samples from Scottish peatland habitats of ranging ecological status and restoration stage (degraded, restored, and good condition categories)?

PKF5: *Community compositions for fungi and invertebrates were found to be broadly different between the degraded and restored sites. Bacteria composition did not exhibit such a clear trend (Figure 27). The classification approach did not perform as well as for woodland. However, these results were heavily limited by the fact that only two sites were identified in the project that had both restored and degraded conditions, making modelling less robust. We cannot currently recommend whether these results are representative across the rest of Scotland.*

PKF5.1: In general, after accounting for the effects of sampling sites and areas, the patterns of community composition were as would be expected for fungi and invertebrates. Communities showed a clear transition from degraded to restored (Figure 28 and Figure 29). This is encouraging, because if such a pattern would not be evident it would mean that using eDNA-based community data to classify peatland condition (degraded/restored) would not be feasible. However, it is important to note that when sampling sites are taken into account in the analyses, the trends across the restoration gradient become more evident, and geographic area does play a substantial role in community composition (Figure 28 and Figure 29). This is evident in the difference between the unconstrained and constrained ordinations. There were limitations in this habitat, and further work is needed to develop models for peatlands.

PKF5.2: Using bacteria eDNA-based data, samples had an overall classification rate of 51%. There are two condition classes, so random classification would achieve a predictive accuracy of ~50%.

This low accuracy was almost entirely caused by misclassification for the Glen Finglas site, which had an accuracy of 37%. For Auchlyne, the accuracy was 77%. The Cashel site did not have a restored area as none could be identified during the field sampling. Removing the Cashel site from the training model did not generally improve predictive performance. Samples could be classified to the degraded condition with 60% accuracy vs 25% accuracy for restored.

Using fungal eDNA-based data, samples had an overall classification rate of 62%. For Glen Finglas, the accuracy was 45%. For Auchlyne, the accuracy was 80%. Samples could be classified to the degraded condition with 75% accuracy vs 45% accuracy for restored.

Using soil invertebrate eDNA-based data, samples had an overall classification rate of 50%. For Glen Finglas, the accuracy was 40%. For Auchlyne, the accuracy was 50%. Samples could be classified to the degraded condition with 73% accuracy vs 15% accuracy for restored.

Overall, the Glen Finglas “restored” sampling locations were very often classified as degraded. It may be that restoration at this site is not working effectively, or is only in the early stages of restoration, and that this is confounding the results.

More importantly, the dataset is almost certainly too small to train a robust classification model. Since there were only two restored-class sites, the model is being asked to learn from just one restored site (from one geographic region) to predict a restored site (from a different geographic region). It is likely that the inclusion of more sites with both habitat conditions would improve accuracy. We also suggest that future work conducts side by side vegetation assessments and eDNA-based sampling to obtain a more balanced study design.

We also ran constrained ordinations for each community dataset, including condition, moisture, and pH into the model. The unconstrained and constrained ordinations for peatland bacteria and fungi were equivalent. For peatland invertebrates, there was a sorting of communities along a moisture and pH gradient, which were not primary factors of interest, but condition was still not a significant contributor to the new ordination.

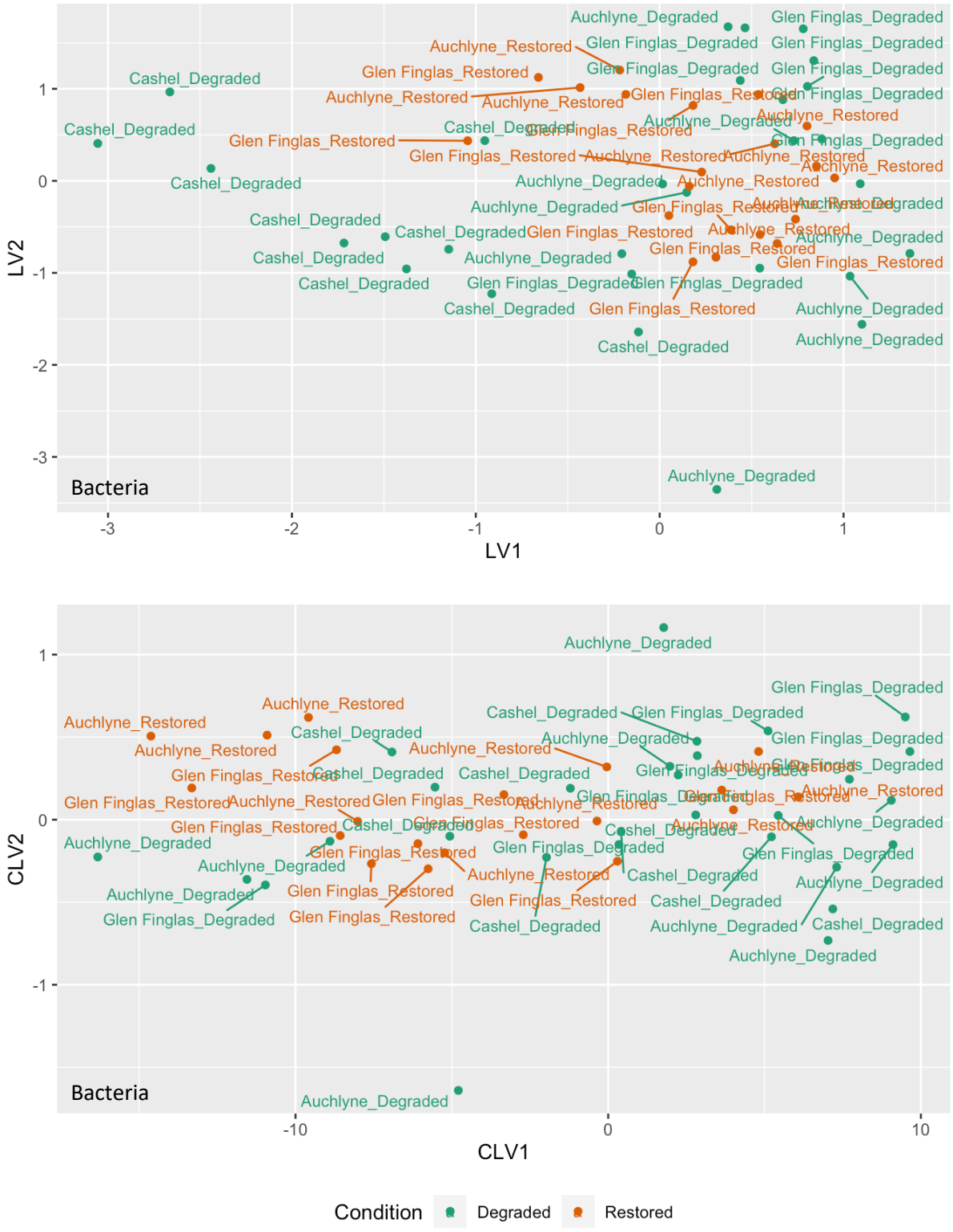


Figure 27: Unconstrained (top) and constrained/concurrent (bottom) ordinations for soil bacteria. Each point is a sample and samples that are closer together are more similar. In neither model was there a clear separation between degraded and restored conditions. Geography (Glen Finglas, Auchlyne, Cashel) was included as a full-rank predictor, and the ordination is thus based on the residuals after this effect of geography was removed.

Figure 28: Unconstrained (top) and constrained/concurrent (bottom) ordinations for soil fungi. Each point is a sample and samples that are closer together are more similar. Constrained ordination showed a clearer separation between degraded and restored conditions. Geography (Glen Finglas, Auchlyne, Cashel) was included as a full-rank predictor, and the ordination is thus based on the residuals after this effect of geography was removed.

Figure 29: Unconstrained (top) and constrained/concurrent (bottom) ordinations for soil invertebrates. Each point is a sample and samples that are closer together are more similar. Constrained ordination showed a clearer separation between degraded and restored conditions. Geography (Glen Finglas, Auchlyne, Cashel) was included as a full-rank predictor, and the ordination is thus based on the residuals after this effect of geography was removed.

PRQ6: Are there any differences or similarities in habitat condition or biodiversity and community composition from any re-surveyed locations between timescales (e.g., seasons or years) and protected status (e.g., designated vs non-designated sites) for the Scottish peatland habitats sampled?

Auchlyne and Glen Finglas were also targeted in the pilot study (6 samples per site, 3 in each condition) in August 2021. The pilot study included a relatively small number of samples and only two locations at Auchlyne and two at Glen Finglas were situated close to locations sampled during this main sampling campaign. This did not enable a statistical comparison of the sampling events.

Regarding fungi, it is of note that one of the SSSI-listed species detected in this main sampling campaign, *Hygrocybe cantharellus*, was also detected in the pilot study and an additional two species (*Hygrocybe miniata* and *Clavulinopsis helvola*) were also detected in the pilot study. The smaller sample sizes in the pilot study preclude a thorough analysis of whether sampling year or season would have caused the differences in SSSI species detected.

PRQ7: What are the key recommendations and future perspectives to develop and operationalise standardised eDNA-based methods at national scale for monitoring biodiversity and the impacts of pressures[1] or restoration in Scottish peatland habitats?

PKF7.1: How should we sample?

Soil augers are a common approach in soil sample collection for eDNA-based surveys, and this project supports the fact that the approach is suitable for detecting high numbers of species. The approach we took of sampling multiple cores within a plot to ~10 cm depth, combining the subsamples in a plastic bag, keeping cool in a cool box and using refrigerated transport to the lab, generally aligns with the literature for eDNA-based soil sampling (e.g. Griffith, Detheridge, and Bye 2022; M. C. Allen et al. 2023). However, there are not many examples of soil sampling for eDNA in peatland habitats and further work is required to make robust recommendations on soil sampling depth. Generally, for eDNA-based soil sampling, vegetation (including moss) is pushed aside prior to collecting a core. It can be difficult to ensure consistency for peatland in this respect as the divide between decomposing matter and vegetation can be difficult to ascertain. The primary benefit of smaller (10 m x 10 m) plots is that they are easier to measure out and potentially easier to place within difficult terrain or patchy habitat – this is particularly important in peatland habitats where drier and wetter areas can be very close together. Freezing the soil samples as soon as possible after collection and processing the samples in the laboratory within one month is highly recommended (Kirse et al. 2021a).

PKF7.2: What should we sample?

This study focussed on soil samples. While there is potential to survey water bodies within peatlands, these are not always present and linking the results from aquatic surveys to particular parcels of land is difficult. For surveying invertebrates, other sample types such as pitfall traps, malaise traps (Kirse et al. 2021b) and even leaf samples (Krehenwinkel et al. 2022) can be collected and processed using metabarcoding. These generally require additional sources such as multiple visits and additional equipment. However, where surveying for above ground and flying invertebrates, these methods are more appropriate. The use of airborne eDNA is a recent development in the molecular analyses of terrestrial biodiversity, including plant (pollen) (Korpelainen and Pietiläinen 2017; Johnson, Cox, and Barnes 2019), fungi (Banchi et al. 2020; Rosa et al. 2020), insect (Roger et al. 2022), and vertebrate (Clare et al. 2022; Klepke et al. 2022) communities. However, alongside practical methodological advancements, gaps in technological knowledge, need to be addressed before airborne eDNA can be integrated into routine biodiversity monitoring. An advantage of using soil samples is being able to survey for multiple taxonomic groups such as bacteria, fungi, and invertebrates simultaneously. Soil

samples can also be used to survey for terrestrial mammals (Mena et al. 2021), although we did not trial this is this study.

For general habitat condition assessment, fungi show the most promise, given the differences in species richness, evolutionary diversity, and functional diversity. However, future work targeting a higher number of sites with both restored and degraded conditions could improve the accuracy of classification models and bacteria and soil invertebrates may still prove useful for this purpose. As discussed in other sections, the effects of disturbance due to restoration works, and conducting concurrent vegetation surveys to choose sampling locations should also be incorporated in future work. Furthermore, surface soil moisture (rather than soil moisture calculated from composite core samples) and/or water table depth can have strong confounding effects on community composition , and we recommend including this measurement in future studies.

PKF7.3: Where should we sample?
Soil sampling is subject to high local variability in species and communities detected, so it is essential that long term monitoring plots are established for repeated surveys to detect habitat condition status and temporal changes.

To monitor temporal changes and condition improvements at a particular site, a similar strategy to the approach used in this study can be employed, whereby multiple plots are established in each treatment. In this context, treatments should ideally include a degraded control site, a target (i.e. a pristine area of peatland, or area considered to be as near pristine as possible) and the restoration site of interest (i.e. a site undergoing rewetting), and the plots should be representative of the treatment. Using such a BACI design provides the most robust method to track progress at a given site. There is little guidance on how many samples or plots should be used for eDNA-based soil monitoring within each treatment.

In this study, 20 samples per site (10 samples per treatment) in Auchlyne (~30 ha area) was sufficient to attain reasonable accuracy in habitat condition classification.. However, without more supporting evidence that classification can be robust across multiple sites, it is not possible to suggest further sampling strategies in this context. For both Auchlyne and Glen Finglas, the sampling intensity in this project was sufficient to detect changes in fungal species richness, evolutionary diversity, and functional diversity.

Before a national baseline could be suggested (such as for woodland), further work is needed to identify peatland sites that are true target conditions that Scotland can aim towards. In this project, despite multiple stakeholder consultations, identifying such sites within the LLTNP and surrounding area was not possible. Future work should focus on identifying sites outside of this area that have the conditions required as being as near pristine as possible so that a national baseline can be set. We suggest that future work conducts concurrent co-located vegetation assessments and eDNA sampling to obtain a more balanced study design. Once such a baseline is set, the basic concepts would be very similar as for the woodland habitat.

PKF7.4: When should we sample?
The pilot study included a relatively small number of samples and only two locations at Auchlyne and two at Glen Finglas were situated close to locations sampled during this main sampling campaign. This did not enable a statistical comparison of the sampling events. Based on the varied results of this study, we cannot currently recommend a surveying season. We sampled in August 2022 and community trends were observed between degraded and restored sites. However, there were limitations with this habitat, we recommend that follow up studies are carried out to investigate this further.

PKF7.5: Where do the knowledge gaps and barriers exist, and why is that? What practical solutions and technology innovations would help overcome these types of challenges going forward?
Peatlands are one of the least targeted habitats for eDNA-based studies to date and there are many unknowns. The lack of a national framework for assessing peatlands using eDNA-based monitoring is the primary current barrier. Putting a national framework in place to monitor peatlands using eDNA-based methods would mean that government, NGOs, and commercial entities could monitor the restoration progress of peatlands in Scotland. However, to ensure that building such a framework is feasible, further work is required to sample near pristine peatlands to ensure that clear ecological and condition signals can be found (as outlined above).

In this project we identified some SSSI-listed fungal species but note that the current guidelines do not cater for peatlands specifically. If using fungi as an indicator of peatland importance and condition is a priority, then similar barriers exist for this purpose as for woodland (see woodland section).

For invertebrates, other studies have noted that COI markers are more suitable for detecting community shifts. These were not trialled in this habitat. We would suggest that any remaining DNA from this project be subjected to these primers to assess the differences as this may increase the power community–based approaches based on invertebrate eDNA metabarcoding data and potentially generate a higher number of species level assignments (Kirse et al. 2021a; 2021b).

PKF7.6: What have we (a) learned to do the same or differently and (b) anticipated are key opportunities to develop, upscale, and implement DNA-based methods for peatland habitats in Scotland?
As noted previously, there were confounding factors in selecting sites for this habitat. We recommend that a similar study should be conducted in near pristine peatland sites to use eDNA-based data for large scale monitoring in this habitat.

An additional set of data analyses that was not performed in this project, would be to look further into the bacterial and fungal traits that could be assigned to the eDNA-based data and to assess whether there are any further insights that can be gleaned from it. For example, are there certain bacterial functions or fungal feeding guilds associated with peatland habitat conditions and can these be used to improve classification of sites? This would be a key opportunity to gain a deeper understanding of the functional ecology, redundancy, and resilience in the communities across the restoration gradient. This is possible using databases such as FunGuild (Nguyen et al. 2016), FungalTraits (Põlme et al. 2020) and the picrust2 pipeline (Douglas et al. 2020).

PKF7.7: What other peatland habitat types across Scotland do we need to consider sampling for eDNA-based biodiversity monitoring and reporting purposes in the future?
There are four main natural peatland habitat types in Scotland: blanket bog, raised bog, fen, and bog woodland. To consider sampling these habitats, a similar study to the current project (along with recommendations for future work) would need to be carried out in each peatland type to assess the efficacy in classification rates and species of interest detected.

4 Summary

Nature is complex and monitoring its recovery is a considerable challenge. Yet, we demonstrate that eDNA-based methods can simplify this complexity and would be highly effective at measuring progress of the Scottish Biodiversity Strategy using metrics and community analyses. In this project report, we have demonstrated the key findings, highlighted advantages, alongside the disadvantages, scoped key knowledge gaps and barriers, reflected on key lessons learned, and provided a set of key recommendations for eDNA-based biodiversity monitoring and reporting purposes going forward.

4.1 Key Findings

- Across all ecological gradients investigated in this project, eDNA metabarcoding community data showed clear compositional differences.
- Statistical approaches were presented showing how the data can be used to classify samples according to ecological gradients.
- These approaches can be used to build national frameworks and baselines to be used in regulatory biomonitoring at large scales.

Marine lochs

- eDNA-based methods can detect PMF (Priority Marine Feature) species across multiple taxonomic groups, including fish, marine mammals, and aquatic invertebrates. eDNA-based approaches have the potential to be implemented in a standardised way to survey for these indicator species.
- AMBI categories calculated from eDNA metabarcoding datasets are comparable to those calculated from morphological survey datasets, providing an efficient solution to scale-up the monitoring of water pollution in marine habitats.
- Marine eDNA-based data can differentiate broadly different biotopes (i.e., different substrate types), but higher-level biotopes were not well-defined by eDNA metabarcoding community datasets. Using eDNA-based data for monitoring biotopes requires further research.
- eDNA-based taxa lists were not suitable (by themselves) for conventional biotope classification in this project. This is likely due to different communities being used for conventional biotope classification (mostly surface and larger invertebrates), compared to the communities detected by eDNA-based methods, which tend to be dominated by meiofauna. This is consistent with broader research on the subject showing a low overlap between the two methods (Steyaert et al. 2020; Wangensteen et al. 2018).

Freshwater lochs

- Freshwater eDNA-based data can differentiate freshwater lochs between High-Good and Moderate-Poor WFD Overall Status categories, with invertebrates and fish eDNA-based data being the most powerful for classification. This provides a very strong indication that a national model can be developed to support WFD-type loch classification, based on eDNA data.
- CPET results using eDNA-based data produce similar values to those produced using best-matching morphological data. This provides support that eDNA-based invertebrate data from freshwater lochs can be used in the existing CPET classification system.

Woodland

- eDNA-based community datasets can differentiate between unforested, recently planted/regenerating, and mature Caledonian pine forest. Samples from recently planted/regenerating woodlands lie in the ordination transition zone between the unforested and mature samples, suggesting that soil eDNA-based data can provide a general, quantitative forest-condition metric.
- Fungi communities had the most consistent results as indicators of restoration progress, including consistent eDNA-derived biodiversity metrics.
- Soil fungal eDNA-based data detected many species of fungi that are typically used for SSSI designation for woodlands, and the eDNA-based approach has the potential to be used for standardised surveys of these important indicator species.

Peatland

- Model-based ordination of peatland eDNA-based invertebrate data found a significant and clear difference between degraded and restored peatlands, but the Random Forest supervised classification model was unable to predict restoration/degradation status. This is likely due to the small size of this dataset (N=3 degraded + N=2 restored locations) combined with a substantial effect of geography and possibly an internally inconsistent classification of peatland into restored and degraded classes. Thus, classification of restoration status from eDNA-based data will likely require a large training dataset with a suitable sampling design and clearer status definitions.

4.2 Advantages and Disadvantages of eDNA-based vs Conventional Methods

eDNA-based methods have many advantages, including the ability to provide broader taxonomic coverage and a wider view of biodiversity, being life-stage independent (many species cannot be morphologically identified in their juvenile forms) and able to detect species that may be difficult to observe due to being shy, nocturnal, small in body size, or otherwise elusive.

Nonetheless, as with any survey method, there are important aspects to consider when using eDNA-based monitoring methods. These are discussed in more detail in the Phase 1 report (Cruickshanks et al. 2022) of this project. We limit this section to discussion of the key advantages and disadvantages that were observed during Phase 2 of this project.

Key Advantages

- eDNA-based sampling creates large community datasets quickly and efficiently. Establishing community datasets of comparable size using conventional methods requires extensive expertise in methods and taxonomic identification.
- While there were limitations to this project (such as lack of impact gradients and limited numbers of samples), we have shown that eDNA-based metabarcoding data is effective for classifying the condition/restoration state of sites across a broad range of habitats.
- Insights into microbiomes, i.e. generating lists of microbial species and using these to classify conditions, through metabarcoding is unprecedented.

- eDNA-based community datasets are in some cases comparable to morphological-based datasets for conventional metrics. This provides the opportunity to perform monitoring using already well-established and approved metrics to track the state of nature.
- eDNA-based methods can identify many species of scientific interest (e.g., IUCN, PMF, SSSI, etc.) across many taxonomic groups.
- eDNA-based sample collection requires minimal training and expertise. However, minimising cross-contamination is a key learning principle to instil for eDNA practitioners.

Key Disadvantages

- In the specific context of using eDNA-based metabarcoding to identify taxa to the species level, there is a reliance on reference sequence databases. Some taxonomic groups have an underrepresentation in DNA reference databases which can impact detection rates and taxonomic resolution. This is less of an issue for vertebrates but is a larger issue for invertebrates and microbes. For marine invertebrates, for example, we found that morphology-based methods outcompeted the eDNA-based methods for species level identification.
- Numerous existing indicators use abundance data (counts of individuals most commonly). Further research is needed to address whether eDNA sequence reads can be used to reflect abundances in a standardised way across taxa and habitats.
- While eDNA sample collection requires relatively little training and expertise, sample processing and bioinformatic methods do require a lot of laboratory and computer-based analyses training and expertise. This highlights a capacity building need for upskilling people and organisations interested in applying eDNA-based approaches going forward.
- The key disadvantage regarding assays/primers is the often incomplete coverage of the target taxa (or a bias towards certain species within the target taxa), leading to incomplete or biased representation of (target taxon) biodiversity in the sampled habitat. The challenge is selecting the most appropriate assay that can effectively amplify a wide range of species within the target taxonomic group. This can be mitigated by using multiple assays/primer sets to increase the range of species coverage (which is more costly). Increasing species coverage in reference databases combined with continued assay improvement and new assay development enhance coverage and reduce biases.

4.3 Key Knowledge Gaps & Barriers

The results in this report identified some key biodiversity knowledge gaps as well as key knowledge gaps and barriers for eDNA-based sampling and monitoring approaches.

- There are numerous eDNA-based projects being carried out in Scotland at various scales and in various contexts of biomonitoring, yet the data is not being captured in a systematic and unified way. Standardised guidance for formatting and storing eDNA-based data in publicly available databases would allow research in this area to progress faster. Systems such as the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser) provide platforms for such applications and could be used to store and access eDNA-based data for biomonitoring purposes.
- Assigning taxonomy, particularly at the species level, remains a challenge for some taxonomic groups. Developing maintained and curated reference sequence databases would allow for more comparable data to be generated across biomonitoring programmes. Although it is noted that this is not required for all applications of biomonitoring. Over time these databases

will improve and while it is not likely that there will ever exist a database that contains DNA references for the entirety of all taxa across the globe, barcoding and genome assembly initiatives are being conducted at the regional, national, and global levels (e.g. https://ibol.org/programs/bioscan/; https://www.darwintreeoflife.org/) that will greatly improve the power of metabarcoding surveys.

- The number of samples required for biomonitoring at the national level using eDNA-based methods remains largely unanswered. This is partially due to fact that the breadth of potential applications is large, spanning numerous taxonomic groups, habitats and biomonitoring objectives. Identifying the number of samples required for each specific biomonitoring objective will require large scale studies with focussed objectives. Biomonitoring at local scales is already possible through careful study design.
- Using eDNA-based methods for biomonitoring at a national level in a regulatory context requires ecological frameworks based on national baselines, such as Ecological Quality Ratio (EQR) models. There are currently very few such frameworks based on, or incorporating, eDNA-based data (the Lake Fish Classification Index (Willby et al. 2020) being the exception). Developing such frameworks requires large scale studies with focussed objectives. Biomonitoring at local scales is already possible through careful study design.
- Calibration with existing indicators and frameworks can be challenging. If for example a new eDNA-based EQR framework produces different outcomes to conventional methods, it can delay the adoption of new methods, or require more extensive calibration studies to be undertaken.
- Many existing indicators use the abundance of individuals as the underlying data. Although we did not investigate this in this project, it was raised as a potential concern. Individual organisms contribute varying amounts of eDNA due to differences in size, behaviour, body composition, life stage, etc., making it difficult to estimate numbers of individuals without controlling for these factors, which is not always possible in natural settings. Differences in shedding and decay rates have also been noted between taxa associated with different temperature regimes (Andruszkiewicz Allan et al. 2021). Although with targeted approaches, such as qPCR, a good relationship of abundance with eDNA concentration has been found in laboratory conditions, this relationship is weaker in natural settings (Yates, Fraser, and Derry 2019). In the case of metabarcoding workflows, this is further compounded by biases introduced during the PCR amplification process, where some species' DNA will amplify more efficiently than others. Nonetheless, many fish eDNA metabarcoding studies have shown a strong correlation between sequence read counts and known relative abundance (Li, Hatton-Ellis, et al. 2019; Di Muri et al. 2020), while assessment of occupancy within a landscape can give a strong indication of how common a species is. Further research is needed in this area before abundance estimates can be widely adopted for biomonitoring.

4.4 Key Lessons Learned

- Until larger ecological biomonitoring frameworks for eDNA-based methods are developed, eDNA-based approaches for national level reporting will likely remain underutilised. Although in the meantime practitioners will be able to use them for efficient surveying of key taxonomic groups, the true power of eDNA-based methods lies in their ability to generate huge datasets that can be used to build national-level models of biodiversity and to characterise ecological

conditions. Local and regional projects are already using eDNA-based approaches to monitor negative and positive impacts of land management and restoration.

- The gradients used for the marine and peatland habitats should be better defined for future projects to progress eDNA-based biomonitoring at the regional or national level.
- For soil sampling, capturing additional factors such as soil horizon layers and soil surface moisture should be factored into future research.
- To align with results of previous work, further work on progressing eDNA sampling methods targeting fish in lochs should be carried out in winter months in the UK.

4.5 Key Recommendations

Using eDNA-based monitoring is not the only emerging technology for routine and investigative national monitoring purposes. It should be combined with existing monitoring tools, as well as considering other emerging technologies such as earth observation (satellite images), lidar (light scanning technology that can create 3D habitat models) and bioacoustics (recording and analysing sounds to monitor species). From an eDNA development perspective we make the following recommendations:

Freshwater lochs

Create eDNA-based tools (sampling and data analysis methods) to predict the condition of freshwater lochs at a regional or national level. These tools will help us understand impacts of pressures on ecosystem health and monitor the quality of the lochs. We can use existing classifications of loch health to develop these tools. To make them more accurate, we need to collect samples from multiple lochs across a wider geographic area and capture eDNA-based signal from a wide range of organisms which may be present. This approach could be expanded to assess the status of other freshwater habitats such as rivers, streams, and ponds. The data provided in this project could already be used as a pilot for this at a regional scale, namely LLTNP area. This could be done by building a tool, for example an online form, that accepts eDNA-based sample data (data generated using the same sampling and processing methods as in this project) and returns loch WFD Overall Status based on modelling of the existing dataset.

Marine lochs

Develop guidelines for how to collect samples of marine species in a consistent way, enabling eDNA-based monitoring programmes in marine lochs. For vertebrates, this means choosing a standard approach for sampling and processing samples, as detection of this group has been shown to be highly effective. For sediment sampling this requires further tests on the number of samples required for specific monitoring objectives. Further validate a method of scoring the health of marine sediment by using sites with different impacts of pollution. Additionally, research the best lab techniques for identifying PMF invertebrate species using eDNA-based methods, which can help us better characterise marine habitats at a regional and national level.

Woodland

To monitor the progress of woodland restoration projects, we recommend focusing on individual sites from the beginning and tracking ecological response through time. We see value in including mixed deciduous forest and, more broadly, sites that have been deemed of high conservation value. In the longer term, it would be possible to conduct a national survey using eDNA-based data analysis to rank woodlands based on their conservation value. This will allow prioritisation of restoration efforts for the most valuable woodlands. We also recommend validating the use of eDNA-based monitoring to detect specific fungal species that are listed as important for conservation and designation of SSSI sites.

Peatland

We found that degraded and restored peatlands have different communities. However, our current model was unable to accurately predict the status of peatlands due to the limited data available. To develop a reliable model, we need a large eDNA-based dataset with clear definitions of peatland condition.

4.6 Future Perspectives

- Adopting eDNA-based approaches holds incredible promise for enhancing existing frameworks, surveying priority species, and assessing sites along ecological gradients. To fully unlock their potential, we need to build extensive datasets and train robust models. Furthermore, the ongoing advancements in eDNA-based assays and cost reduction in sample processing are pivotal for widespread adoption in national reporting frameworks.
- While we await the development of larger eDNA-based biomonitoring frameworks, it is important to recognise that using eDNA-based methods can already make a positive impact on biodiversity monitoring at local scales. Through thoughtful study design, we can gather valuable insights and contribute to the understanding of our local ecosystems.
- There is a potential strategic opportunity for coordinating governance of DNA-based priorities and strengthening partnerships to enhance synergies and underpin capacity building needs for Scotland going forward. This may involve complementing use of existing monitoring tools with emerging technology including, though not constrained to, expanding eDNA methods development, implementation, and training plans or developing a national monitoring plan for biodiversity assessment purposes and operationalising eDNA sampling efforts across the country. Collaborative networks (such as the SG-CAMERAS Board Partnership and Scottish DNA Hub) have key roles in enabling discussion amongst different stakeholders and identifying next steps for utilising innovative approaches that could be applied more broadly and upscaled in future habitat monitoring programmes.
- We emphasise the impacts and legacy value of the project as a springboard to inform and unlock the potential for Scotland to implement eDNA methods for biodiversity monitoring and reporting purposes in the future. For example, from the new eDNA insights gained and by tackling some previous unknowns, through to practical end use in method development and decision-making spaces, as well as having produced extensive datasets and many remaining physical samples which will be made available for future reuses.
- The true power of eDNA-based approaches lies in their ability to generate vast datasets using standardised sampling kits by those appropriately trained. These datasets serve as tools for characterising ecological conditions, ensuring consistent monitoring, and facilitating accurate reporting. By harnessing them, we can gain deeper insights into our environment for working towards its preservation and restoration in the future.

5 Glossary

Benthic	Anything associated with or occurring on the bottom of a body of water. The animals and plants that live on or in the bottom are known as the benthos.
bioinformatics	Refers to a data processing pipeline that takes the raw sequence data from high-throughput sequencing (often 20 million sequences or more) and transforms it into usable ecological data. Key steps for metabarcoding pipelines include quality filtering, trimming, merging paired ends, removal of sequencing errors such as chimeras, clustering of similar sequences into molecular operational taxonomic units (OTUs; each of which approximately represents a species) and matching one sequence from each cluster against a reference database. The output is a species-by-sample table showing how many sequences from each sample were identified as each species.
biotope	Community of species (biotic elements) associated with particular environmental conditions (abiotic elements), giving a particular habitat type.
extraction blank	A DNA extraction with no sample added to assess potential contamination during the DNA extraction process.
gel electrophoresis	The process in which DNA is separated according to size and electrical charge via an electric current, while in a gel. The process is used to confirm the successful amplification of a specific size fragment of DNA.
high-throughput sequencing	Technology developed in the 2000s that produces millions of sequences in parallel. Enables thousands of different organisms from a mixture of species to be sequenced at once, so community DNA can be sequenced. Various different technologies exist to do this, but the most commonly used platform is Illumina's MiSeq. Also known as Next-Generation Sequencing (NGS) or parallel sequencing.
Jaccard similarity index	This index is a calculation that compares two samples to see which taxa are shared and which are distinct. The higher the value, the more similar two samples are in their community composition.
metabarcoding	Refers to identification of species assemblages from community DNA using barcode genes. PCR is carried out with non-specific primers, followed by high-throughput sequencing and bioinformatics processing. Can identify hundreds of species in each sample, and 100+ different samples can be processed in parallel to reduce sequencing cost.
negative control	A PCR reaction with no sample added to determine if PCR reactions are contaminated.

NMDS	Non-metric multidimensional scaling (NMDS) is a method that allows you to visualise the similarity of each sample to one another. The dissimilarity between each sample is calculated, taking into account shared taxa (Jaccard similarity index), and then configured into a 2D ordinal space that allows you to see the relationship of each sample to one another. Samples that are closer together are more similar to one another in terms of community composition, while samples that are further apart are less similar. This type of clustering analysis allows you to see if certain types of samples, for example, those from a particular habitat type, are more clustered together and therefore more similar to one another compared to other groups.
Nekton	The actively swimming aquatic organisms in a body of water, that are able to move independently of currents.
OTU	Short for Operational Taxonomic Unit. Similar sequences are clustered into OTUs at a defined similarity threshold. OTUs are approximately equivalent to species and are treated as such in our analyses. Species-level taxonomic assignments may or may not be possible, depending on the availability of reference sequences and the similarity between closely related species in the amplified marker. It may be possible to refine the taxonomic assignment for an OTU later as more sequences are added to reference databases.
PCR	Short for Polymerase chain reaction. A process by which millions of copies of a particular DNA segment are produced through a series of heating and cooling steps. Known as an 'amplification' process. One of the most common processes in molecular biology and a precursor to most sequencing-based analyses.
positive control	Used to determine whether the PCR is working correctly.
primers	Short sections of synthesised DNA that bind to either end of the DNA fragment to be amplified by PCR. Can be designed to be totally specific to a particular species (so that only that species' DNA will be amplified from a community DNA sample), or to be very general so that a wide range of species' DNA will be amplified. Good design of primers is one of the critical factors in DNA-based monitoring.
rarefy	A normalisation technique which transforms the data to remove biases associated with uneven sampling depth (number of reads) across samples. The sampling depth of each sample is standardised to a specified number of reads (usually that of the sample with the lowest depth) by random resampling.
reference databases	Over time, the DNA sequences of many species have been compiled into publicly accessible databases by scientists from around the world. These databases serve as a reference against which unknown sequences can be queried to obtain a species identification. The most commonly accessed database is NCBI (National Center for Biotechnology Information), which is

	maintained by the US National Institute of Health. Anyone can search for DNA sequences at https://www.ncbi.nlm.nih.gov
richness	The total number of taxa within a sample.
sample coverage	A measure of how complete the sample is in detecting all taxa of an assemblage.
sequence	A DNA sequence is made up of four nucleotide bases represented by the letters A, T, C & G. The precise order of these letters is used to compare genetic similarity among individuals or **species** and to identify species using **reference databases**. In **high-throughput sequencing** analyses (e.g. **metabarcoding**), many identical copies of the same sequence are obtained for each species in the sample. The number of copies obtained per species is known as the number of sequence reads, and this can be often - although not always - related to the relative abundance of the species.
taxon (s.)/**taxa** (pl.)	Strictly, a taxonomic group. Here we use the term to describe groups of DNA sequences that are equivalent to **species**. We do not use the term species because we are unable to assign complete identifications to all of the groups at this time due to gaps in the available reference databases.
taxonomy species (s./pl.)	A group of genetically similar organisms that show a high degree of overall similarity in many independent characteristics. Related species are grouped together into progressively larger taxonomic units, from genus to kingdom. *Homo sapiens* (human) is an example of a species. **genus** (s.) / **genera** (pl.) - A group of closely related species. Each genus can include one or more species. Homo is an example of a genus. **family** (s.) / **families** (pl.) - A group of closely related genera. Homo sapiens is in the Family Hominidae (great apes). **order** (s.) / **orders** (pl.) - A group of closely related families. Homo sapiens is in the Order Primates. **class** (s.) / **classes** (pl.) - A group of closely related orders. Homo sapiens is in the Class Mammalia.
phylum (s.) / **phyla** (pl)	A group of closely related classes. Homo sapiens is in the Phylum Chordata.

6 Bibliography

Adams, Colin E., and Peter S. Maitland. 1998. “The Ruffe Population of Loch Lomond, Scotland: Its Introduction, Population Expansion, and Interaction with Native Species.” *Journal of Great Lakes Research* 24 (2): 249–62. https://doi.org/10.1016/S0380-1330(98)70817-2.

Allan, Elizabeth Andruszkiewicz, Michelle H. DiBenedetto, Andone C. Lavery, Annette F. Govindarajan, and Weifeng G. Zhang. 2021. “Modeling Characterization of the Vertical and Temporal Variability of Environmental DNA in the Mesopelagic Ocean.” *Scientific Reports* 11 (1): 21273. https://doi.org/10.1038/s41598-021-00288-5.

Allen, Craig R., M Axelsson, S Dewey, and Laura Clark. 2013. “Marine Biological Survey to Establish the Distribution of Priority Marine Features within the Clyde Sea Area.” 437. Scottish Natural Heritage Commissioned Report.

Allen, Michael C., Robert Kwait, Anthony Vastano, Alex Kisurin, Isabelle Zoccolo, Benjamin D. Jaffe, Jordan C. Angle, Brooke Maslo, and Julie L. Lockwood. 2023. “Sampling Environmental DNA from Trees and Soil to Detect Cryptic Arboreal Mammals.” *Scientific Reports* 13 (1): 180. https://doi.org/10.1038/s41598-023-27512-8.

Amaral-Zettler, Linda A., Elizabeth A. McCliment, Hugh W. Ducklow, and Susan M. Huse. 2009. “A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes.” Edited by Gordon Langsley. *PLoS ONE* 4 (7): e6372. https://doi.org/10.1371/journal.pone.0006372.

Andruszkiewicz Allan, Elizabeth, Weifeng Gordon Zhang, Andone C. Lavery, and Annette F. Govindarajan. 2021. “Environmental DNA Shedding and Decay Rates from Diverse Animal Forms and Thermal Regimes.” *Environmental DNA* 3 (2): 492–514. https://doi.org/10.1002/edn3.141.

Artz, Rebekka R. E., D Donnelly, Roxane Andersen, Ruth Mitchell, Stephen Chapman, Jo Smith, Pete Smith, Roger Cummins, Bedru Balana, and Andrew Cuthbert. 2014. “Managing and Restoring Blanket Bog to Benefit Biodiversity and Carbon Balance – a Scoping Study.” 562. Scottish Natural Heritage Commissioned Report.

Aylagas, Eva, Ángel Borja, Iñigo Muxika, and Naiara Rodríguez-Ezpeleta. 2018. “Adapting Metabarcoding-Based Benthic Biomonitoring into Routine Marine Ecological Status Assessment Networks.” *Ecological Indicators* 95 (December): 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044.

Bairoch, A. 2000. “The ENZYME Database in 2000.” *Nucleic Acids Research* 28 (1): 304–5. https://doi.org/10.1093/nar/28.1.304.

Bakker, Judith, Hayley Craig, Lynsey R Harper, Douglas W Yu, Christian Devenish, Hannah Flintham, Nathan R. Geraldi, et al. 2023a. “Phase 2 Main Report - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements.” NMP/001/20.

———. 2023b. “Phase 2 Summary Brief - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements.” NMP/001/20.

———. 2023c. “Phase 2 Technical Appendices - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements.” NMP/001/20.

Banchi, Elisa, Claudio G. Ametrano, Enrico Tordoni, David Stanković, Silvia Ongaro, Mauro Tretiach, Alberto Pallavicini, et al. 2020. “Environmental DNA Assessment of Airborne Plant and Fungal Seasonal Diversity.” *Science of The Total Environment* 738 (October): 140249. https://doi.org/10.1016/j.scitotenv.2020.140249.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Bedwell, Mallory E., and Caren S. Goldberg. 2020. “Spatial and Temporal Patterns of Environmental DNA Detection to Inform Sampling Protocols in Lentic and Lotic Systems.” *Ecology and Evolution* 10 (3): 1602–12. https://doi.org/10.1002/ece3.6014.

Bessey, Cindy, Simon N. Jarman, Oliver Berry, Ylva S. Olsen, Michael Bunce, Tiffany Simpson, Matthew Power, James McLaughlin, Graham J. Edgar, and John Keesing. 2020. “Maximizing Fish Detection with EDNA Metabarcoding.” *Environmental DNA* 2 (4): 493–504. https://doi.org/10.1002/edn3.74.

Borja, Angel. 2018. “Testing the Efficiency of a Bacterial Community-Based Index (MicrogAMBI) to Assess Distinct Impact Sources in Six Locations around the World.” *Ecological Indicators* 85 (February): 594–602. https://doi.org/10.1016/j.ecolind.2017.11.018.

Borja, Angel, Julien Mader, and Iñigo Muxika. 2012. “Instructions for the Use of the AMBI Index Software (Version 5.0).” *Revista de Investigación Marina, AZTI-Tecnalia* 19 (3).

Bosanquet, Sam, Martyn Ainsworth, Sean Cooch, David Genney, and Tim Wilkins. 2018. “Chapter 14 Non-Lichenised Fungi.” In *Guidelines for the Selection of Biological SSSIs. Part 2: Detailed Guidelines for Habitats and Species Group.* Peterborough: Joint Nature Conservation Committee.

Bruce, Kat, Rosetta Blackman, Sarah J. Bourlat, Ann Micaela Hellström, Judith Bakker, Iliana Bista, Kristine Bohmann, et al. 2021. *A Practical Guide to DNA-Based Methods for Biodiversity Assessment*. Pensoft Publishers. https://doi.org/10.3897/ab.e68634.

Caddy, J. F. 1993. “Toward a Comparative Evaluation of Human Impacts on Fishery Ecosystems of Enclosed and Semi-enclosed Seas.” *Reviews in Fisheries Science* 1 (1): 57–95. https://doi.org/10.1080/10641269309388535.

Canals, Oriol, Iñaki Mendibil, Maria Santos, Xabier Irigoien, and Naiara Rodriguez-Ezpeleta. 2021. “Vertical Stratification of Environmental DNA in the Open Ocean Captures Ecological Patterns and Behavior of Deep-Sea Fishes.” *Limnology and Oceanography* 6 (6): 14. https://doi.org/10.1002/lol2.10213.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. “Global Patterns of 16S RRNA Diversity at a Depth of Millions of Sequences per Sample.” *Proceedings of the National Academy of Sciences* 108 (Supplement_1): 4516–22. https://doi.org/10.1073/pnas.1000080107.

Capra, E., R. Giannico, M. Montagna, F. Turri, P. Cremonesi, F. Strozzi, P. Leone, G. Gandini, and F. Pizzi. 2016. "A New Primer Set for DNA Metabarcoding of Soil Metazoa." *European Journal of Soil Biology* 77 (November): 53–59. https://doi.org/10.1016/j.ejsobi.2016.10.005.

Carraro, Luca, Elvira Mächler, Remo Wüthrich, and Florian Altermatt. 2020. "Environmental DNA Allows Upscaling Spatial Patterns of Biodiversity in Freshwater Ecosystems." *Nature Communications* 11 (1): 1–12. https://doi.org/10.1038/s41467-020-17337-8.

Clare, Elizabeth L., Chloe K. Economou, Frances J. Bennett, Caitlin E. Dyer, Katherine Adams, Benjamin McRobie, Rosie Drinkwater, and Joanne E. Littlefair. 2022. "Measuring Biodiversity from DNA in the Air." *Current Biology* 32 (3): 693-700.e5. https://doi.org/10.1016/j.cub.2021.11.064.

Cowart, Dominique A., Katherine R Murphy, and C.-H Christina Cheng. 2022. "Environmental DNA from Marine Waters and Substrates: Protocols for Sampling and EDNA Extraction." In *Methods in Molecular Biology*. Vol. 2498. Marine Genomics. https://doi.org/10.1007/978-1-0716-2313-8_11.

Di Muri, Cristina, Lori Lawson Handley, Colin W. Bean, Marco Benucci, Lynsey R. Harper, Ben James, Jianlong Li, Ian J. Winfield, and Bernd Hänfling. 2022. "Spatio-temporal Monitoring of Lake Fish Spawning Activity Using Environmental DNA Metabarcoding." *Environmental DNA*, July, edn3.343. https://doi.org/10.1002/edn3.343.

Diana, Alex, Eleni Matechou, Jim Griffin, Douglas Yu, Mingjie Luo, Marie I Tosa, Alex Bush, and Richard Griffiths. 2022. "EDNAPlus: A Unifying Modelling Framework for DNA-Based Biodiversity Monitoring." https://arxiv.org/abs/2211.12213.

Douglas, Gavin M., Vincent J. Maffei, Jesse R. Zaneveld, Svetlana N. Yurgel, James R. Brown, Christopher M. Taylor, Curtis Huttenhower, and Morgan G. I. Langille. 2020. "PICRUSt2 for Prediction of Metagenome Functions." *Nature Biotechnology* 38 (6): 685–88. https://doi.org/10.1038/s41587-020-0548-6.

Duggan-Edwards, Mollie F., Jordi F. Pagès, Stuart R. Jenkins, Tjeerd J. Bouma, and Martin W. Skov. 2020. "External Conditions Drive Optimal Planting Configurations for Salt Marsh Restoration." Edited by Joslin Moore. *Journal of Applied Ecology* 57 (3): 619–29. https://doi.org/10.1111/1365-2664.13550.

Easson, Cole G., Kevin M. Boswell, Nicholas Tucker, Joseph D. Warren, and Jose V. Lopez. 2020. "Combined EDNA and Acoustic Analysis Reflects Diel Vertical Migration of Mixed Consortia in the Gulf of Mexico." *Frontiers in Marine Science* 7 (July): 552. https://doi.org/10.3389/fmars.2020.00552.

Edgar, Graham J., Rick D. Stuart-Smith, Trevor J. Willis, Stuart Kininmonth, Susan C. Baker, Stuart Banks, Neville S. Barrett, et al. 2014. "Global Conservation Outcomes Depend on Marine Protected Areas with Five Key Features." *Nature* 506 (7487): 216–20. https://doi.org/10.1038/nature13022.

Edgar, R.C. 2004. "MUSCLE: Multiple Sequence Alignment with Improved Accuracy and Speed." In *Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004.*, 689–90. Stanford, CA, USA: IEEE. https://doi.org/10.1109/CSB.2004.1332560.

Edwards, C. 2006. "Stand Structure and Dynamics of Four Native Scots Pine (Pinus Sylvestris L.) Woodlands in Northern Scotland." *Forestry* 79 (3): 261–77. https://doi.org/10.1093/forestry/cpl014.

Egeter, Bastian, Hayley Craig, Lynsey R Harper, Edward Wort, and Judith Bakker. 2023. "Phase 1 Pilot Study Findings & Phase 2 Sampling Plan - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements." NMP/001/20. NatureMetrics.

Egeter, Bastian, Joana Veríssimo, Manuel Lopes-Lima, Cátia Chaves, Joana Pinto, Nicoletta Riccardi, Pedro Beja, and Nuno A. Fonseca. 2022. "Speeding up the Detection of Invasive Bivalve Species Using Environmental DNA: A Nanopore and Illumina Sequencing Comparison." *Molecular Ecology Resources*, April, 1755-0998.13610. https://doi.org/10.1111/1755-0998.13610.

Ely, Taylor, Paul H. Barber, Lauren Man, and Zachary Gold. 2021. "Short-Lived Detection of an Introduced Vertebrate EDNA Signal in a Nearshore Rocky Reef Environment." Edited by Hideyuki Doi. *PLOS ONE* 16 (6): e0245314. https://doi.org/10.1371/journal.pone.0245314.

Faith, Daniel P. 1992. "Conservation Evaluation and Phylogenetic Diversity." *Biological Conservation* 61 (1): 1–10. https://doi.org/10.1016/0006-3207(92)91201-3.

Forestry Commission Scotland. 2014. "Scotland's Native Woodlands - Results from the Native Woodland Survey of Scotland." Natural Scotland, Scottish Government.

Ghilardi, Mattia, Nina M. D. Schiettekatte, Jordan M. Casey, Simon J. Brandl, Samuel Degregori, Alexandre Mercière, Fabien Morat, Yves Letourneur, Sonia Bejarano, and Valeriano Parravicini. 2021. "Phylogeny, Body Morphology, and Trophic Level Shape Intestinal Traits in Coral Reef Fishes." *Ecology and Evolution* 11 (19): 13218–31. https://doi.org/10.1002/ece3.8045.

Goldberg, Caren S., Cameron R. Turner, Kristy Deiner, Katy E. Klymus, Philip Francis Thomsen, Melanie A. Murphy, Stephen F. Spear, et al. 2016. "Critical Considerations for the Application of Environmental DNA Methods to Detect Aquatic Species." Edited by M. Gilbert. *Methods in Ecology and Evolution* 7 (11): 1299–1307. https://doi.org/10.1111/2041-210X.12595.

Griffith, Gareth W, Andrew P Detheridge, and Ruby Bye. 2022. "Use of EDNA Analysis of Soil Samples to Evaluate the Fungal Conservation Value of Grassland Areas in the South West Peak; Lessons for Woodland Creation Proposals."

Griffiths, Nathan P., Bernd Hänfling, and Colin W Bean. 2023. "Environmental DNA Based Approaches to Site Condition Monitoring of Fish in Standing Waters (EDNA-SCM)." NatureScot Research Report.

Gunn, I.D.M., L. Carvalho, C.E. Davies, F.K. Edwards, M.T. Furse, P.S. Maitland, C. Raper, G.M. Siriwardena, and I.J. Winfield. 2018. "UK Checklist of Freshwater Species." NERC Environmental Information Data Centre. https://doi.org/10.5285/57653719-434b-4b11-9f0d-3bd76054d8bd.

Guri, Gledis, Jon-Ivar Westgaard, Nigel Yoccoz, Owen S. Wangensteen, Kim Præbel, Jessica Louise Ray, Ryan P. Kelly, Andrew Olaf Shelton, Tanja Hanebrekke, and Torild Johansen. 2023. "Maximizing Sampling Efficiency to Detect Differences in Fish Community Composition Using ENVIRONMENTAL DNA Metabarcoding in Subarctic Fjords." *Environmental DNA*, April, edn3.409. https://doi.org/10.1002/edn3.409.

Hallam, Jane, Elizabeth L. Clare, John Iwan Jones, and Julia J. Day. 2021. "Biodiversity Assessment across a Dynamic Riverine System: A Comparison of EDNA Metabarcoding versus Traditional Fish Surveying Methods." *Environmental DNA* 3 (6): 1247–66. https://doi.org/10.1002/edn3.241.

Hänfling, Bernd, Lori Lawson Handley, Daniel S. Read, Christoph Hahn, Jianlong Li, Paul Nichols, Rosetta C. Blackman, Anna Oliver, and Ian J. Winfield. 2016. "Environmental DNA

Metabarcoding of Lake Fish Communities Reflects Long-Term Data from Established Survey Methods." *Molecular Ecology* 25 (13): 3101–19. https://doi.org/10.1111/mec.13660.

Harper, Kirsten, Patrick Anucha, James Turnbull, Colin Bean, and Michael Leaver. 2018. "Searching for a Signal: Environmental DNA (EDNA) for the Detection of Invasive Signal Crayfish, Pacifastacus Leniusculus (Dana, 1852)." *Management of Biological Invasions* 9 (2): 137–48. https://doi.org/10.3391/mbi.2018.9.2.07.

Harper, Kirsten J., Kelly D. Goodwin, Lynsey R. Harper, Erin L. LaCasella, Amy Frey, and Peter H. Dutton. 2020. "Finding Crush: Environmental DNA Analysis as a Tool for Tracking the Green Sea Turtle Chelonia Mydas in a Marine Estuary." *Frontiers in Marine Science* 6 (January): 810. https://doi.org/10.3389/fmars.2019.00810.

Harrison, Jori B., Jennifer M. Sunday, and Sean M. Rogers. 2019. "Predicting the Fate of EDNA in the Environment and Implications for Studying Biodiversity." *Proceedings of the Royal Society B: Biological Sciences* 286 (1915): 20191409. https://doi.org/10.1098/rspb.2019.1409.

Hervé, Alix, Isabelle Domaizon, Jean-Marc Baudoin, Tony Dejean, Pierre Gibert, Pauline Jean, Tiphaine Peroux, et al. 2022. "Spatio-Temporal Variability of EDNA Signal and Its Implication for Fish Monitoring in Lakes." Edited by Hideyuki Doi. *PLOS ONE* 17 (8): e0272660. https://doi.org/10.1371/journal.pone.0272660.

Hestetun, Jon Thomassen, Anders Lanzén, and Thomas G. Dahlgren. 2021. "Grab What You Can—an Evaluation of Spatial Replication to Decrease Heterogeneity in Sediment EDNA Metabarcoding." *PeerJ* 9 (June): e11619. https://doi.org/10.7717/peerj.11619.

Hestetun, Jon Thomassen, Anders Lanzén, Katrine Skaar, and Thomas Dahlgren. 2020. "The Impact of DNA Extract Homogenization and Replication on Marine Sediment Metabarcoding Diversity and Heterogeneity." Preprint. Preprints. https://doi.org/10.22541/au.160486389.98416006/v1.

Jensen, Mads Reinholdt, Eva Egelyng Sigsgaard, Marcelo de Paula Ávila, Sune Agersnap, William Brenner-Larsen, Mita Eva Sengupta, Yingchun Xing, et al. 2022. "Short-term Temporal Variation of Coastal Marine EDNA." *Environmental DNA* 4 (4): 747–62. https://doi.org/10.1002/edn3.285.

Jeunen, Gert-Jan, Miles D. Lamare, Michael Knapp, Hamish G. Spencer, Helen R. Taylor, Michael Stat, Michael Bunce, and Neil J. Gemmell. 2020. "Water Stratification in the Marine Biome Restricts Vertical Environmental DNA (EDNA) Signal Dispersal." *Environmental DNA* 2 (1): 99–111. https://doi.org/10.1002/edn3.49.

Johnson, Mark D., Robert D. Cox, and Matthew A. Barnes. 2019. "Analyzing Airborne Environmental DNA: A Comparison of Extraction Methods, Primer Type, and Trap Type on the Ability to Detect Airborne EDNA from Terrestrial Plant Communities." *Environmental DNA* 1 (2): 176–85. https://doi.org/10.1002/edn3.19.

Kelly, Ryan P., Jesse A. Port, Kevan M. Yamahara, and Larry B. Crowder. 2014. "Using Environmental DNA to Census Marine Fishes in a Large Mesocosm." Edited by Gretchen E. Hofmann. *PLoS ONE* 9 (1): e86175. https://doi.org/10.1371/journal.pone.0086175.

Kirse, Ameli, Sarah J. Bourlat, Kathrin Langen, and Vera G. Fonseca. 2021a. "Unearthing the Potential of Soil EDNA Metabarcoding—Towards Best Practice Advice for Invertebrate Biodiversity Assessment." *Frontiers in Ecology and Evolution* 9 (May): 630560. https://doi.org/10.3389/fevo.2021.630560.

———. 2021b. “Metabarcoding Malaise Traps and Soil EDNA Reveals Seasonal and Local Arthropod Diversity Shifts.” *Scientific Reports* 11 (1): 10498. https://doi.org/10.1038/s41598-021-89950-6.

Klepke, Martin Johannesen, Eva Egelyng Sigsgaard, Mads Reinholdt Jensen, Kent Olsen, and Philip Francis Thomsen. 2022. “Accumulation and Diversity of Airborne, Eukaryotic Environmental DNA.” *Environmental DNA*, July, edn3.340. https://doi.org/10.1002/edn3.340.

Korpelainen, Helena, and Maria Pietiläinen. 2017. “Biodiversity of Pollen in Indoor Air Samples as Revealed by DNA Metabarcoding.” *Nordic Journal of Botany* 35 (5): 602–8. https://doi.org/10.1111/njb.01623.

Kranzfelder, Petra, Alyssa M. Anderson, Alexander T. Egan, Jane E. Mazack, Jr. Bouchard R. William, Moriya M. Rufer, and Jr. Ferrington Leonard C. 2015. “Use of Chironomidae (Diptera) Surface-Floating Pupal Exuviae as a Rapid Bioassessment Protocol for Water Bodies.” *JoVE*, no. 101 (July): e52558. https://doi.org/10.3791/52558.

Krehenwinkel, Henrik, Sven Weber, Rieke Broekmann, Anja Melcher, Julian Hans, Rüdiger Wolf, Axel Hochkirch, et al. 2022. “Environmental DNA from Archived Leaves Reveals Widespread Temporal Turnover and Biotic Homogenization in Forest Arthropod Communities.” *ELife* 11 (November): e78521. https://doi.org/10.7554/eLife.78521.

Larson, Wesley A., Patrick Barry, Willie Dokai, Jacek Maselko, John Olson, and Diana Baetscher. 2022. “Leveraging EDNA Metabarcoding to Characterize Nearshore Fish Communities in Southeast Alaska: Do Habitat and Tide Matter?” *Environmental DNA* 4 (4): 868–80. https://doi.org/10.1002/edn3.297.

Lawson Handley, Lori, Daniel S. Read, Ian J. Winfield, Helen Kimbell, Harriet Johnson, Jianlong Li, Christoph Hahn, et al. 2019. “Temporal and Spatial Variation in Distribution of Fish Environmental DNA in England’s Largest Lake.” *Environmental DNA* 1 (1): 26–39. https://doi.org/10.1002/edn3.5.

Leese, Florian, Mandy Sander, Dominik Buchner, Vasco Elbrecht, Peter Haase, and Vera M. A. Zizka. 2021. “Improved Freshwater Macroinvertebrate Detection from Environmental DNA through Minimized Nontarget Amplification.” *Environmental DNA* 3 (1): 261–76. https://doi.org/10.1002/edn3.177.

Lejzerowicz, Franck, Andrew John Gooday, Inés Barrenechea Angeles, Tristan Cordier, Raphaël Morard, Laure Apothéloz-Perret-Gentil, Lidia Lins, et al. 2021. “Eukaryotic Biodiversity and Spatial Patterns in the Clarion-Clipperton Zone and Other Abyssal Regions: Insights From Sediment DNA and RNA Metabarcoding.” *Frontiers in Marine Science* 8 (May): 671033. https://doi.org/10.3389/fmars.2021.671033.

Lenth, Russell. 2023. “_emmeans: Estimated Marginal Means, Aka Least-Squares Means_. R Package Version 1.8.5,.” https://github.com/rvlenth/emmeans.

Leray, Matthieu, Joy Y Yang, Christopher P Meyer, Suzanne C Mills, Natalia Agudelo, Vincent Ranwez, Joel T Boehm, and Ryuji J Machida. 2013. “A New Versatile Primer Set Targeting a Short Fragment of the Mitochondrial COI Region for Metabarcoding Metazoan Diversity: Application for Characterizing Coral Reef Fish Gut Contents.” *Frontiers in Zoology* 10 (1): 34. https://doi.org/10.1186/1742-9994-10-34.

Li, Jianlong, Tristan W. Hatton-Ellis, Lori Jayne Lawson Handley, Helen S. Kimbell, Marco Benucci, Graeme Peirson, and Bernd Hänfling. 2019. “Ground-Truthing of a Fish-Based Environmental

DNA Metabarcoding Method for Assessing the Quality of Lakes." *Journal of Applied Ecology* 56 (5): 1232–44. https://doi.org/10.1111/1365-2664.13352.

Li, Jianlong, Lori J. Lawson Handley, Lynsey R. Harper, Rein Brys, Hayley V. Watson, Cristina Di Muri, Xiang Zhang, and Bernd Hänfling. 2019. "Limited Dispersion and Quick Degradation of Environmental DNA in Fish Ponds Inferred by Metabarcoding." *Environmental DNA* 1 (3): 238–50. https://doi.org/10.1002/edn3.24.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest" 2.

Littlefair, Joanne E., Lee E. Hrenchuk, Paul J. Blanchfield, Michael D. Rennie, and Melania E. Cristescu. 2020. "Thermal Stratification and Fish Thermal Preference Explain Vertical EDNA Distributions in Lakes." *Molecular Ecology* 30 (13): 3083–96. https://doi.org/10.1111/mec.15623.

Lund, Mads, Jacob Agerbo Rasmussen, Jazmín Ramos-Madrigal, Ruairidh Sawers, M. Thomas P. Gilbert, and Christopher James Barnes. 2022. "Rhizosphere Bacterial Communities Differ among Traditional Maize Landraces." *Environmental DNA* 4 (6): 1241–49. https://doi.org/10.1002/edn3.333.

Luo, Ye, Xu Wei, Shuai Yang, Yuan-Hao Gao, and Zhu-Hua Luo. 2020. "Fungal Diversity in Deep-Sea Sediments from the Magellan Seamounts as Revealed by a Metabarcoding Approach Targeting the ITS2 Regions." *Mycology* 11 (3): 214–29. https://doi.org/10.1080/21501203.2020.1799878.

Ma, Wenxu, Zhen Yang, Lisong Liang, Qinghua Ma, Guixi Wang, and Tiantian Zhao. 2021. "Seasonal Changes in Soil Microbial Community and Co-Occurrence Network of Species of the Genus Corylus." *Microorganisms* 9 (11): 2228. https://doi.org/10.3390/microorganisms9112228.

Mancinelli, Giorgio, Salvatrice Vizzini, Antonio Mazzola, Stefano Maci, and Alberto Basset. 2013. "Cross-Validation of **Δ**15N and FishBase Estimates of Fish Trophic Position in a Mediterranean Lagoon: The Importance of the Isotopic Baseline." *Estuarine, Coastal and Shelf Science* 135 (December): 77–85. https://doi.org/10.1016/j.ecss.2013.04.004.

Mason, Claire. 2016. "Particle Size Analysis (PSA) for Supporting Biological Analysis." NMBAQC's Best Practice Guidance.

Mena, José Luis, Hiromi Yagui, Vania Tejeda, Emilio Bonifaz, Eva Bellemain, Alice Valentini, Mathias W. Tobler, Pamela Sánchez-Vendizú, and Arnaud Lyet. 2021. "Environmental DNA Metabarcoding as a Useful Tool for Evaluating Terrestrial Mammal Diversity in Tropical Forests." *Ecological Applications* 31 (5): e02335. https://doi.org/10.1002/eap.2335.

Miya, M., Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, et al. 2015. "MiFish, a Set of Universal PCR Primers for Metabarcoding Environmental DNA from Fishes: Detection of More than 230 Subtropical Marine Species." *Royal Society Open Science* 2 (7): 150088. https://doi.org/10.1098/rsos.150088.

Monuki, Keira, Paul H. Barber, and Zachary Gold. 2021. "EDNA Captures Depth Partitioning in a Kelp Forest Ecosystem." Edited by Hideyuki Doi. *PLOS ONE* 16 (11): e0253104. https://doi.org/10.1371/journal.pone.0253104.

Moore, C G. 2013. "Biological Analyses of Underwater Video from Research Cruises in the Clyde Sea (Loch Goil and the South of Arran) and in Orkney (Rousay Sound and Stronsay Firth)." 631. Scottish Natural Heritage Commissioned Report.

Murakami, Hiroaki, Seokjin Yoon, Akihide Kasai, Toshifumi Minamoto, Satoshi Yamamoto, Masayuki K. Sakata, Tomoya Horiuchi, et al. 2019. "Dispersion and Degradation of Environmental DNA from Caged Fish in a Marine Environment." *Fisheries Science* 85 (2): 327–37. https://doi.org/10.1007/s12562-018-1282-6.

Muri, Cristina Di, Lori Lawson Handley, Colin W. Bean, Jianlong Li, Graeme Peirson, Graham S. Sellers, Kerry Walsh, Hayley V. Watson, Ian J. Winfield, and Bernd Hänfling. 2020. "Read Counts from Environmental DNA (EDNA) Metabarcoding Reflect Fish Abundance and Biomass in Drained Ponds." *Metabarcoding and Metagenomics* 4: 97–112. https://doi.org/10.3897/MBMG.4.56959.

Nguyen, Nhu H., Zewei Song, Scott T. Bates, Sara Branco, Leho Tedersoo, Jon Menke, Jonathan S. Schilling, and Peter G. Kennedy. 2016. "FUNGuild: An Open Annotation Tool for Parsing Fungal Community Datasets by Ecological Guild." *Fungal Ecology* 20 (April): 241–48. https://doi.org/10.1016/j.funeco.2015.06.006.

Niku, Jenni, Francis K. C. Hui, Sara Taskinen, and David I. Warton. 2019. "Gllvm: Fast Analysis of Multivariate Abundance Data with Generalized Linear Latent Variable Models in R." Edited by Sarah Goslee. *Methods in Ecology and Evolution* 10 (12): 2173–82. https://doi.org/10.1111/2041-210X.13303.

Osburn, Ernest D., Steven G. McBride, Frank O. Aylward, Brian D. Badgley, Brian D. Strahm, Jennifer D. Knoepp, and J. E. Barrett. 2019. "Soil Bacterial and Fungal Communities Exhibit Distinct Long-Term Responses to Disturbance in Temperate Forests." *Frontiers in Microbiology* 10 (December): 2872. https://doi.org/10.3389/fmicb.2019.02872.

Pagad, Shyama, Stewart Bisset, Piero Genovesi, Quentin Groom, Tim Hirsch, Walter Jetz, Ajay Ranipeta, Dmitry Schigel, Yanina V. Sica, and Melodie A. McGeoch. 2022. "Country Compendium of the Global Register of Introduced and Invasive Species." *Scientific Data* 9 (1): 391. https://doi.org/10.1038/s41597-022-01514-z.

Pagad, Shyama, Piero Genovesi, Lucilla Carnevali, Dmitry Schigel, and Melodie A. McGeoch. 2018. "Introducing the Global Register of Introduced and Invasive Species." *Scientific Data* 5 (1): 170202. https://doi.org/10.1038/sdata.2017.202.

Parry, M E V. 2019. "Guidance on Assigning Benthic Biotopes Using EUNIS or the Marine Habitat Classification of Britain and Ireland (Revised 2019)."

Pauly, Daniel, Villy Christensen, Johanne Dalsgaard, Rainer Froese, and Francisco Torres. 1998. "Fishing Down Marine Food Webs." *Science* 279 (5352): 860–63. https://doi.org/10.1126/science.279.5352.860.

Pauly, Daniel, and Reg Watson. 2005. "Background and Interpretation of the 'Marine Trophic Index' as a Measure of Biodiversity." *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1454): 415–23. https://doi.org/10.1098/rstb.2004.1597.

Pawlowski, Jan, Laure Apothéloz-Perret-Gentil, Elvira Mächler, and Florian Altermatt. 2020. "Environmental DNA Applications for Biomonitoring and Bioassessment in Aquatic Ecosystems." https://doi.org/10.5167/UZH-187800.

Petchey, Owen L., and Kevin J. Gaston. 2006. "Functional Diversity: Back to Basics and Looking Forward." *Ecology Letters* 9 (6): 741–58. https://doi.org/10.1111/j.1461-0248.2006.00924.x.

Poikane, Sandra, Richard K. Johnson, Leonard Sandin, Ann Kristin Schartau, Angelo G. Solimini, Gorazd Urbanič, Kęstutis Arbačiauskas, et al. 2016. "Benthic Macroinvertebrates in Lake Ecological Assessment: A Review of Methods, Intercalibration and Practical Recommendations." *Science of The Total Environment* 543 (February): 123–34. https://doi.org/10.1016/j.scitotenv.2015.11.021.

Põlme, Sergei, Kessy Abarenkov, R. Henrik Nilsson, Björn D. Lindahl, Karina Engelbrecht Clemmensen, Havard Kauserud, Nhu Nguyen, et al. 2020. "FungalTraits: A User-Friendly Traits Database of Fungi and Fungus-like Stramenopiles." *Fungal Diversity* 105 (1): 1–16. https://doi.org/10.1007/s13225-020-00466-2.

Port, Jesse A., James L. O'Donnell, Ofelia C. Romero-Maraccini, Paul R. Leary, Steven Y. Litvin, Kerry J. Nickols, Kevan M. Yamahara, and Ryan P. Kelly. 2016. "Assessing Vertebrate Biodiversity in a Kelp Forest Ecosystem Using Environmental DNA." *Molecular Ecology* 25 (2): 527–41. https://doi.org/10.1111/mec.13481.

R Core Team. 2021. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing." Vienna, Austria. https://www.R-project.org/.

Reinholdt Jensen, Mads, Eva Egelyng Sigsgaard, Sune Agersnap, Jes Jessen Rasmussen, Annette Baattrup-Pedersen, Peter Wiberg-Larsen, and Philip Francis Thomsen. 2021. "Seasonal Turnover in Community Composition of Stream-associated Macroinvertebrates Inferred from Freshwater Environmental DNA Metabarcoding." *Environmental DNA* 3 (4): 861–76. https://doi.org/10.1002/edn3.193.

Riaz, Tiayyba, Wasim Shehzad, Alain Viari, François Pompanon, Pierre Taberlet, and Eric Coissac. 2011. "EcoPrimers: Inference of New DNA Barcode Markers from Whole Genome Sequence Analysis." *Nucleic Acids Research* 39 (21): e145–e145. https://doi.org/10.1093/nar/gkr732.

Roger, Fabian, Hamid Ghanavi, Natalie Danielsson, Niklas Wahlberg, Jakob Löndahl, Lars B. Pettersson, Georg K.S. Andersson, Niklas Boke Olén, and Yann Clough. 2022. "Airborne Environmental DNA Metabarcoding for the Monitoring of Terrestrial Insects - a Proof of Concept," Environmental DNA, 4 (4): 790–807. https://doi.org/10.1101/2021.07.26.453860.

Romanuk, Tamara N., April Hayward, and Jeffrey A. Hutchings. 2011. "Trophic Level Scales Positively with Body Size in Fishes." *Global Ecology and Biogeography* 20 (2): 231–40. https://doi.org/10.1111/j.1466-8238.2010.00579.x.

Rosa, Luiz Henrique, Thamar Holanda da Silva, Mayara Baptistucci Ogaki, Otávio Henrique Bezerra Pinto, Michael Stech, Peter Convey, Micheline Carvalho-Silva, Carlos Augusto Rosa, and Paulo E. A. S. Câmara. 2020. "DNA Metabarcoding Uncovers Fungal Diversity in Soils of Protected and Non-Protected Areas on Deception Island, Antarctica." *Scientific Reports* 10 (1): 21986. https://doi.org/10.1038/s41598-020-78934-7.

Saulino, Hugo H.L., Miguel Cañedo-Argüelles, Susana Trivinho-Strixino, Guilherme Rossi Gorni, and Juliano José Corbi. 2021. "Chironomid Pupal Exuviae Communities Support the 'Field of Dreams' Hypothesis after the Riparian Vegetation Recovery in Headwater Urban Streams." *Ecological Indicators* 127 (August): 107766. https://doi.org/10.1016/j.ecolind.2021.107766.

Schmidt, Kristie J., Daniel A. Soluk, Sarah E. Mays Maestas, and Hugh B. Britten. 2021. "Persistence and Accumulation of Environmental DNA from an Endangered Dragonfly." *Scientific Reports* 11 (1): 18987. https://doi.org/10.1038/s41598-021-98099-1.

Scottish Government. 2022. "Biodiversity Strategy to 2045: Tackling the Nature Emergency." https://www.gov.scot/publications/scottish-biodiversity-strategy-2045-tackling-nature-emergency-scotland/pages/1/.

Sellers, Graham S., Christopher L Jerde, Lynsey R Harper, Marco Benucci, Cristina Di Muri, Jianlong Li, Graeme Peirson, et al. 2023. "Optimising Species Detection Probability and Sampling Effort in Lake Fish EDNA Surveys." Preprint. Metabarcoding and Metagenomics. https://doi.org/10.3897/arphapreprints.e105785.

Sethi, Suresh A., Trevor A. Branch, and Reg Watson. 2010. "Global Fishery Development Patterns Are Driven by Profit but Not Trophic Level." *Proceedings of the National Academy of Sciences* 107 (27): 12163–67. https://doi.org/10.1073/pnas.1003236107.

Sigsgaard, Eva Egelyng, Ida Broman Nielsen, Henrik Carl, Marcus Anders Krag, Steen Wilhelm Knudsen, Yingchun Xing, Tore Hejl Holm-Hansen, Peter Rask Møller, and Philip Francis Thomsen. 2017. "Seawater Environmental DNA Reflects Seasonality of a Coastal Fish Community." *Marine Biology* 164 (6): 128. https://doi.org/10.1007/s00227-017-3147-4.

Soler, German A., Graham J. Edgar, Russell J. Thomson, Stuart Kininmonth, Stuart J. Campbell, Terence P. Dawson, Neville S. Barrett, et al. 2015. "Reef Fishes at All Trophic Levels Respond Positively to Effective Marine Protected Areas." Edited by Dennis M. Higgs. *PLOS ONE* 10 (10): e0140270. https://doi.org/10.1371/journal.pone.0140270.

Stauffer, Salomé, Meret Jucker, Thomas Keggin, Virginie Marques, Marco Andrello, Sandra Bessudo, Marie-Charlotte Cheutin, et al. 2021. "How Many Replicates to Accurately Estimate Fish Biodiversity Using Environmental DNA on Coral Reefs?" *BioRxiv Preprint*, 30. https://doi.org/10.1101/2021.05.26.445742.

Steyaert, Margaux, Victoria Priestley, Owen Osborne, Alba Herraiz, Richard Arnold, and Vincent Savolainen. 2020. "Advances in Metabarcoding Techniques Bring Us Closer to Reliable Monitoring of the Marine Benthos." Edited by Vitor Paiva. *Journal of Applied Ecology* 57 (11): 2234–45. https://doi.org/10.1111/1365-2664.13729.

Suarez-Bregua, Paula, Miguel Álvarez-González, Kim M. Parsons, Josep Rotllant, Graham J. Pierce, and Camilo Saavedra. 2022. "Environmental DNA (EDNA) for Monitoring Marine Mammals: Challenges and Opportunities." *Frontiers in Marine Science* 9 (September): 987774. https://doi.org/10.3389/fmars.2022.987774.

Tedersoo, Leho, Mohammad Bahram, Sergei Põlme, Urmas Kõljalg, Nourou S. Yorou, Ravi Wijesundera, Luis Villarreal Ruiz, et al. 2014. "Fungal Biogeography. Global Diversity and Geography of Soil Fungi." *Science (New York, N.Y.)* 346 (6213): 1256688. https://doi.org/10.1126/science.1256688.

Troth, Christopher R., Michael J. Sweet, Jen Nightingale, and Alfred Burian. 2021. "Seasonality, DNA Degradation and Spatial Heterogeneity as Drivers of EDNA Detection Dynamics." *Science of The Total Environment* 768 (May): 144466. https://doi.org/10.1016/j.scitotenv.2020.144466.

Tyler-Walters, H, Ben James, Morven Carruthers, C Wilding, O Durkin, C Lacey, E Philpott, et al. 2016. "Descriptions of Scottish Priority Marine Features (PMFs)." 406. Scottish Natural Heritage Commissioned Report.

UKTAG. 2008. "UKTAG LAKE ASSESSMENT METHODS BENTHIC INVERTEBRATE FAUNA CHIRONOMID PUPAL EXUVIAE TECHNIQUE (CPET)." Water Framework Directive - United Kingdom Technical Advisory Group (WFD-UKTAG). http://www.wfduk.org/resources%20/lake-cpet.

Valsecchi, Elena, Antonella Arcangeli, Roberto Lombardi, Elizabeth Boyse, Ian M. Carr, Paolo Galli, and Simon J. Goodman. 2021. "Ferries and Environmental DNA: Underway Sampling From Commercial Vessels Provides New Opportunities for Systematic Genetic Surveys of Marine Biodiversity." *Frontiers in Marine Science* 8 (August): 704786. https://doi.org/10.3389/fmars.2021.704786.

Veen, Bert van der, Francis K. C. Hui, Knut A. Hovstad, and Robert B. O'Hara. 2022. "Concurrent Ordination: Simultaneous Unconstrained and Constrained Latent Variable Modelling." *Methods in Ecology and Evolution* 14: 683–95. https://doi.org/10.1111/2041-210X.14035.

Wangensteen, Owen S., Creu Palacín, Magdalena Guardiola, and Xavier Turon. 2018. "DNA Metabarcoding of Littoral Hard-Bottom Communities: High Diversity and Database Gaps Revealed by Two Molecular Markers." *PeerJ* 6 (May): e4705. https://doi.org/10.7717/peerj.4705.

Wilhelm, Roland C., Harold M. van Es, and Daniel H. Buckley. 2022. "Predicting Measures of Soil Health Using the Microbiome and Supervised Machine Learning." *Soil Biology and Biochemistry* 164 (January): 108472. https://doi.org/10.1016/j.soilbio.2021.108472.

Willby, Nigel, Alan Law, Colin Bull, Bernd Hänfling, Lori Lawson Handley, and Ian Winfield. 2020. "A Tool for Classifying the Ecological Status of Lake Fish in Britain Based on EDNA Metabarcoding. Report to the Scottish Environment Protection Agency (SEPA)."

Williamson, Brian D., James P. Hughes, and Amy D. Willis. 2021. "A Multiview Model for Relative and Absolute Microbial Abundances." *Biometrics* doi:10.1111/biom.13503 (June). https://doi.org/10.1111/biom.13503.

Williamson, Laura D., Beth E. Scott, Megan Laxton, Janine B. Illian, Victoria L.G. Todd, Peter I. Miller, and Kate L. Brookes. 2022. "Comparing Distribution of Harbour Porpoise Using Generalized Additive Models and Hierarchical Bayesian Models with Integrated Nested Laplace Approximation." *Ecological Modelling* 470 (August): 110011. https://doi.org/10.1016/j.ecolmodel.2022.110011.

Worsfold, Tim, and David Hall. 2010. "Guidelines for Processing Marine Macrobenthic Invertebrate Samples: A Processing Requirements Protocol."

Yamamoto, Satoshi, Reiji Masuda, Yukuto Sato, Tetsuya Sado, Hitoshi Araki, Michio Kondoh, Toshifumi Minamoto, and Masaki Miya. 2017. "Environmental DNA Metabarcoding Reveals Local Fish Communities in a Species-Rich Coastal Sea." *Scientific Reports* 7 (1): 40368. https://doi.org/10.1038/srep40368.

Yates, Matthew C., Dylan J. Fraser, and Alison M. Derry. 2019. "Meta-Analysis Supports Further Refinement of EDNA for Monitoring Aquatic Species-Specific Abundance in Nature." *Environmental DNA* 1 (1): 5–13. https://doi.org/10.1002/edn3.7.

Zhang, Shan, Qi Lu, Yiyan Wang, Xiaomei Wang, Jindong Zhao, and Meng Yao. 2020. "Assessment of Fish Communities Using Environmental DNA: Effect of Spatial Sampling Design in Lentic Systems of Different Sizes." *Molecular Ecology Resources* 20 (1): 242–55. https://doi.org/10.1111/1755-0998.13105.

7 Appendices

7.1 Mutually Agreed IP Statement & NM Standard IP Policy

[NMP/001/20] DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS

Purpose of this project IP document, its situational need and intended use

The purpose of this document was to establish clarity, shared understanding, and a record of the mutually agreed Intellectual Property (IP) position, applied in a specific context for the project [NMP/001/20 DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS] which was funded by the Scottish Government's Rural and Environment Science and Analytical Services (SG-RESAS) Contract Research Fund and commissioned by the Co-ordinated Agenda for Marine, Environment and Rural Affairs Science (SG-CAMERAS) Board Partnership, in collaboration with the Scottish DNA Hub.

There was a situational need for (1) establishing clear and reasonable project boundaries about what exactly was and was not considered the contractor's proprietary information regarding specific information related to the 'wet'[8] and 'dry'[9] aspects of DNA-based materials and methods used for project sampling, analysis, data interpretation, and reporting by the contractors (e.g., sample collection, laboratory processing and analysis protocols, bioinformatics pipelines, statistical data analysis, and DNA-based products); and (2) aligning clear and reasonable expectations amongst the key project actors (i.e., the lead requesters[10]; the contractors[11]; the key stakeholders[12]) regarding what

[8] Wet aspects include sampling kits & components, laboratory analysis (everything from DNA extraction to loading on a DNA sequencer), remaining DNA.

[9] Dry aspects include sampling strategy, bioinformatics pipelines, statistical data analysis, data interpretation, reports.

[10] The project [NMP/001/20 DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS] lead requesters, aka the core Management Steering Group (MSG) leadership team, represent the Scottish Government (SG) and its collective partner organisations (including SEPA, The Marine Directorate of Scottish Government (formerly Marine Scotland Science, MSS) and NatureScot and other key project stakeholders represented by the Scottish DNA Hub, Technical Steering/Reviewing Group (TSG/TRG), SG-CAMERAS Board Partnership, and Scotland's National Parks), together with the SG-RESAS project funder, and SG lead project manager/partner agency.

[11] The project [NMP/001/20 DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS] contractors, NatureMetrics, are expected to work closely and collaboratively with the SG lead project manager, the MSG, and above-mentioned key stakeholders as needed during the project lifespan. Additionally, NatureMetrics are expected to consult with the Advisory Board as outlined in the main contract.

[12] The project [NMP/001/20 DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS] involves a broad collective 'hub' of the key stakeholders orbiting the project (including the wider Technical Steering/Reviewing Group (TSG/TRG), the Scottish DNA Hub, SG-CAMERAS Board Partnership, and Scotland's National Parks) who need to be

specific information related to the 'wet' and 'dry' aspects of DNA-based materials and methods used for project sampling, analysis, data interpretation, and reporting by the contractors (e.g., sample collection, laboratory processing and analysis protocols, bioinformatics pipelines, statistical data analysis, and DNA-based products) was needed and was to be made available to project users at a satisfactory level of detail (i.e., by sitting outside of the contractor's proprietary information or within a realm of the contractor's adaptive capacity to reasonably meet precise project requests). There was a need for providing information that enabled (*a*) successful project delivery, use, and impact; (*b*) project methods to be followed and reasonably repeated by others in the future; and (*c*) where feasible open access information from the project that can be shared and potentially used to advance the development and implementation of DNA-based expertise, approaches, and technology innovations across local and regional to national and international levels.

Furthermore, we intended for this document to be used in a relevant and timely manner for enabling clear and transparent communications regarding the project's mutually agreed IP position for considering, informing, and supporting the:

- needs of many different people and wide range of organisations connecting with, interested in, and contributing to the project, as well as its practical useability and impacts going forward (e.g., baseline monitoring end-uses) – especially ahead of knowledge-exchange activities and engagement opportunities with key stakeholders during the project's remaining lifespan; and
- potential needs of subsequent users across the wider research:policy community in Scotland, the UK, and globally for enabling key knowledge-exchanges (e.g., Scottish DNA Hub collaboration) to thrive beyond the project's lifespan and include future uses of the project's related deliverable outputs – for example by making its DNA-based sequencing outputs publicly available following project completion.

This document was a 'living' document during the project duration to help facilitate joint discussion, until a clear and mutually agreed IP position had been reached by all parties in a timely manner which aligned with the project lifespan and the need to share draft work for a wider peer review. To ensure transparency in communication and shared understanding with key project stakeholders, a final cleaned version of the project's clear and mutually agreed IP position statement was produced and appended to the contractor's standard IP policy document. It is included here as an Appendix to the Phase 2 Main Report deliverable output, being made available during peer review and embedded in the published finalised (i.e., signed-off) version of the Phase 2 Main Report at project completion.

reasonably consulted, included, and considered as part of a two-way knowledge-exchange (KE) process. This was to ensure their specific needs as well as collective roles and contributions were reasonably engaged, voices heard in a safe space, and diverse perspectives respected for the purpose of enhancing mutual collaborative benefits, adding value to discussions, and aiding development of draft project work. This approach will help to bring others along on a learning-by-doing journey of developing, upscaling and implementing DNA-based methods for biodiversity assessment and reporting purposes. We will do this by co-creating purposeful, constructive, and respectful conditions for enabling quality and positive KE engagement opportunities at the most relevant points during the project's lifespan.

For additional relevant background information, please refer to the appended document - Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf and we signpost readers to an indication of the level of detail presented in the Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (Egeter et al. 2023).

To help establish a clear and mutually agreed IP position for information and use, applied in the specific context of the project, key IP themes identified as needs relating to materials and methods needs for "sample collection", "laboratory processing and analysis protocols", "bioinformatics pipelines", "statistical data analysis" and "DNA-based products" have been responded to under each header and structured into the relevant sections as non-sharable or sharable information, with "other key IP considerations" noted separately.

Non-sharable Information

Information that cannot be shared by the project contractors for proprietary (or commercial) reasons:

"Sample Collection"

- Supplier details of kits or kit components (commercial reasons).
- Itemised costs of individual kit components (commercial reasons).
- Anything listed as "N" in the "Can be shared (Y/N)" column of Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
    - Exceptions:
        - NM will share the buffer used for aquatic kits in this project.

"Laboratory Processing and Analysis Protocols"

- Any laboratory process not documented in Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
- Anything listed as "N" in the "Can be shared (Y/N)" column of Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
    - Exceptions:
        - NM will share primer sequences (excluding indexes or unique molecular identifiers) for all metabarcoding primers used in the project to date. This does not extend to other metabarcoding primers that were not discussed for inclusion. If it is decided that any further metabarcoding primers are to be included as part of this project, specific details of what can be shared will be taken into consideration and discussed with NatureMetrics and the MSG prior to this decision.
- See Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (Egeter et al. 2023) for the precise level of detail that can be shared for each item.
    - Primers or lab methodology for fungal:bacterial (F:B) ratio.

Notes:

- DNA extraction – In this project NM used off-the-shelf brand items so this can be shared.

"Bioinformatics Pipelines"

- Any process not documented in Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
- Anything listed as "N" in the "Can be shared (Y/N)" column of Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
- Note that the Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (Egeter et al. 2023) provides the level of detail that can be shared for each item.
- NM reserves the right not to provide a greater level of detail than this.
- Any bioinformatics code/scripts.
- Sequencing run quality or quantity statistics.
    - Exceptions:
        - For this project NM will provide DNA quality scores.
- Other sequencing run output information.
- Details of any other samples or projects that were included on a sequencing run.
- Any methods used to curate or error-check public databases for the purposes of taxonomic assignment or lookups of taxonomic attributes.
- DNA barcode gap analysis software and any improvements.

"Statistical Data Analysis"

- Any statistical data analysis code/scripts.
- Exceptions to "NatureMetrics does not share any methodology pertaining to statistical analyses or other analyses post generation of the OTU table" (Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf)
    - Description of statistical data analyses performed. See below.
        - NM reserves the right not to provide a greater level of detail than this.

"DNA-based products"

- Note for Fastq files – all adapters and primer sequences will be removed prior to providing Fastq files. Fastq files will include all samples that proceeded to sequencing as part of the project. Sample metadata will also be provided, and it will be clear which samples each Fastq file is related to.

Sharable Information

Information the project contractors expect to generate as part of the funded project, as part of their obligations under the contract and is not deemed as proprietary information:

"Sample Collection"

- The methodology used to collect the samples, including sampling strategy, sampling protocols and sample storage conditions.
- Itemised list of kit components. See Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (*Egeter et al. 2023).*

"Laboratory Processing and Analysis Protocols"

- Anything listed as "Y" in the "Can be shared (Y/N)" column of Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
- Note that the Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (*Egeter et al. 2023)* provide the level of detail that can be shared for each item.
- NM reserves the right not to provide a greater level of detail than this.

"Bioinformatics Pipelines"

- Anything listed as "Y" in the "Can be shared (Y/N)" column of Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf
- Note that the Phase 1 Pilot Study Findings and Phase 2 Sampling Plan main report deliverable output (*Egeter et al. 2023)* provide the level of detail that can be shared for each item.
    - NM reserves the right not to provide a greater level of detail than this.
- An overview of bioinformatics processing steps.

"Statistical Data Analysis"

- Overview of basic statistical data analyses performed, including basic descriptions of each metric, relevant citations for key steps, and basic descriptions of statistical comparison of metrics between habitat conditions/categories. For added clarity, examples of the level of detail that will be shared is provided below. Core statistical packages and core model calls will be shared.
- Note that the Phase 2 Main Report deliverable output (Bakker et al. 2023a) will provide the level of detail that can be shared for each item.

Examples of basic descriptions of statistics and metrics. Note that these may not be the final exact methods used in the project and are provided here only to provide the SG with the expected level of detail that will be provided by the contractors in the final deliverable outputs:

- Data were analysed using the *R* v4.1.0 statistical environment in the RStudio IDE. *Tidyverse* v1.3.1 packages were used for data manipulation and formatting. The total read counts per sample were used to calculate the proportional read counts for each taxon. Bubble plots showing positive detections in samples were produced using the package *ggplot2* v3.3.5. The

mean of each metric between sampling groups was compared using the lm or lmer packages in R.

- We used the randomForest package (Liaw & Wiener 2002) for supervised classification and the gllvm package (Niku et al. 2019, van Veen et al. 2022) for generalised linear latent variable modelling and concurrent ordination. For the woodland habitat we used the following gllvm calls (this will be included for all habitats).
    - gllvm(y = y, family = binomial(), num.lv = 2, studyDesign = Site, row.eff = ~(1 | Site), control.start = list(n.init = 10))
    - gllvm(y = y, X = X, formula = ~Condition, family = binomial(), num.lv.c = 2, lv.formula = ~Area + Moisture, studyDesign = Site, row.eff = ~(1 | Site), control.start = list(n.init = 10))
    - Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. R News. 2, 18-22.
    - Niku, J., Hui, F. K. C., Taskinen, S., and Warton, D. I. (2019). gllvm - Fast analysis of multivariate abundance data with generalized linear latent variable models in R. Methods in Ecology and Evolution, 10, 2173-2182.
- Species Richness: the number of unique taxonomic units (OTUs) detected in a sample. Only target OTUs were considered.
- Evolutionary Diversity: This was calculated following Luo (2020) and Lund (2022). Briefly, sequences were aligned using the MUSCLE algorithm (Edgar, 2004). Based on this alignment, a matrix of pairwise dissimilarities was built and used to generate a phylogenetic tree based on unsupervised clustering. Faith's PD (Faith, 1992) was calculated for each sample in turn using the total length of sample tree branches.
    - D. P. Faith (1992). Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10.
    - Luo, Y., Wei, X., Yang, S., Gao, Y. H., & Luo, Z. H. (2020). Fungal diversity in deep-sea sediments from the Magellan seamounts as revealed by a metabarcoding approach targeting the ITS2 regions. Mycology, 11(3), 214-229.
    - Lund, M., Agerbo Rasmussen, J., Ramos-Madrigal, J., Sawers, R., Gilbert, M. T. P., & Barnes, C. J. (2022). Rhizosphere bacterial communities differ among traditional maize landraces. Environmental DNA.
    - Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
- Bacterial Functional Diversity: This was calculated by processing the OTU table through the Picrust2 pipeline. This takes the sequence associated to each OTU, places it into an extensive reference phylogeny based on a maximum-likelihood similarity algorithm (Douglas et al. 2020), predicts the abundance of all gene families in the EC (Enzyme commission; Bairoch, 2000) database for each OTU. The resulting OTU/gene family table is used to build a clustering tree of OTUs based on their similarity in terms of functional profile. The branching length of the tree made with the OTUs detected in each sample provides a measure of the breadth of functions performed at the community level (Petchey and Gaston, 2006).

- Bairoch, A. (2000). The ENZYME database in 2000. Nucleic acids research, 28(1), 304-305.
- Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., ... & Langille, M. G. (2020). PICRUSt2 for prediction of metagenome functions. Nature biotechnology, 38(6), 685-688.
- Petchey, O. L., & Gaston, K. J. (2006). Functional diversity: back to basics and looking forward. Ecology letters, 9(6), 741-758.

- Fungal Functional Diversity: This was calculated by comparing the OTUs within each sample to a database of fungal ecological traits. OTUs that had a match are assigned to one or more ecological categories (Nguyen et al. 2016), based on splitting the output of the guild category in the FunGuild database in its individual components (e.g., Plant Pathogen, Ectomycorrhizal, Wood Saprotroph, Lichenized). The resulting guild profile is used to categorise OTUs based on their ecological role and build a clustering tree with the resulting dissimilarity matrix. The branching length of the tree made with the OTUs detected in each sample provides a measure of the breadth of functions performed at the community level (Petchey and Gaston, 2006).

  - Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., ... & Kennedy, P. G. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology, 20, 241-248.
  - Petchey, O. L., & Gaston, K. J. (2006). Functional diversity: back to basics and looking forward. Ecology letters, 9(6), 741-758.

- Fish metrics
  - The fish species detected were compared to a global fish database, FishBase, and each species was assigned a trophic level score. The trophic level for each sample is the mean of all the scores obtained. A mean trophic level of 2 or less indicates a degraded community while greater than 3 indicates a better community condition. FishBase has detailed information on how they calculate metrics for species - https://www.fishbase.se/manual/english/key%20facts.htm. A similar method, using FishBase, was used to assign the price category and vulnerability index.
- Marine Sediment Pollution Index: We used the AMBI pollution index (http://ambi.azti.es) to assign a mean AMBI value for each sample. AMBI is an established metric to determine the ecological status of coasts and estuaries using a database of >10,000 categorised benthic species (http://ambi.azti.es). More recently, genomic AMBI (gAMBI) was developed given the many benefits of using eDNA (Aylagas et al. 2014). gAMBI has a high correlation with AMBI (0.66 or 0.88 *r2*; Aylagas et al. 2018). Following these studies, we use gAMBI to categorise the pollution level based on invertebrate communities detected by eDNA.

  - Aylagas, Borja, Muxika, Rodríguez-Ezpeleta (2018) Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. Ecological Indicators 95:1.

- Soil Fungal:Bacterial (F:B) Ratio
  - This is a metric that we can calculate using a quantitative PCR (qPCR) assay that measures the number of fungal and bacterial gene copies to calculate the ratio of fungal to bacterial (F:B) gene copies in a soil DNA sample.

- Invasive species
    - Species identified were checked against the Global Register of Introduced and Invasive Species (GRIIS), which is an IUCN Invasive Species Specialist Group initiative. Species listed as Invasive are reported (following Pagad et al. 2018 and Pagad et al. 2020)
    - Pagad, S., Genovesi, P., Carnevali, L. et al. (2018) Introducing the Global Register of Introduced and Invasive Species. Sci Data 5, 170202.
    - Pagad, S., Bisset, S., Genovesi, P. et al. (2022) Country Compendium of the Global Register of Introduced and Invasive Species. Sci Data 9, 391.
- IUCN species
    - The number of named species in an OTU table that have Red List status of Vulnerable, Endangered, Critically Endangered.

"DNA-based products"

- Fastq sequencing files, as defined in - Data_and_Methodology_Sharing_NatureMetrics_public_v1.0.pdf.
    - "..demultiplexed, paired end Fastq files, including base quality scores. To enable NatureMetrics to provide this service whilst protecting its Intellectual Property, all adapters and primer sequences will be removed prior to providing. Henceforth this is referred to as "DNA sequence data"".
    - "Provision of DNA sequence data from control samples used internally by NatureMetrics will not necessarily be provided. This is because control samples with DNA or PCR product below quantifiable levels are not sequenced."
    - Fastq files will be provided by a customer specific, time-limited, download link. These will be provided in a format that has the internal NatureMetrics identifier and marker gene in the filename. A separate table will provide a link between the internal NatureMetrics sample ID and the client sample name. DNA Sequence data will be removed from the download link, 3 months after the data is released. It is the responsibility of the customer (the SG in this case) to ensure they have downloaded and securely backed up the DNA sequence data within this time frame. NatureMetrics will not retain responsibility for providing DNA sequence data after this window.
    - Fastq data and OTU tables will be owned by SG and will be made publicly available by SG following delivery by the NatureMetrics at successful project completion for future use by anyone, which may include for purposes such as bioinformatics pipeline developments and improvements; reanalysing samples and applying new technology; developing product innovations; independent data validation and testing bioinformatics pipeline reproducibility.
- OTU tables (OTU by sample matrix):
    - Excel files with all OTU tables.

- Scottish Government will be the owner of the Fastq sequence data, OTU tables, and remaining physical DNA sample material; as well as owning the specified PDF deliverable outputs from both Phase 1 and Phase 2 (i.e., reports, technical appendices, summary brief, etc.).
- Remaining DNA will be handed over to the Scottish Government at project completion. Note that not all samples will have DNA remaining. All remaining DNA samples will be housed as 'one project' at The Marine Directorate (formerly Marine Scotland Science, MSS) Marine Laboratory in Aberdeen. The MSG will need to retain shared governance of appropriately assessing any future R&D requests for use and re-use of the project's remaining physical DNA sample material (refer to the project specification document for more details), at least until there are no leftover DNA samples.

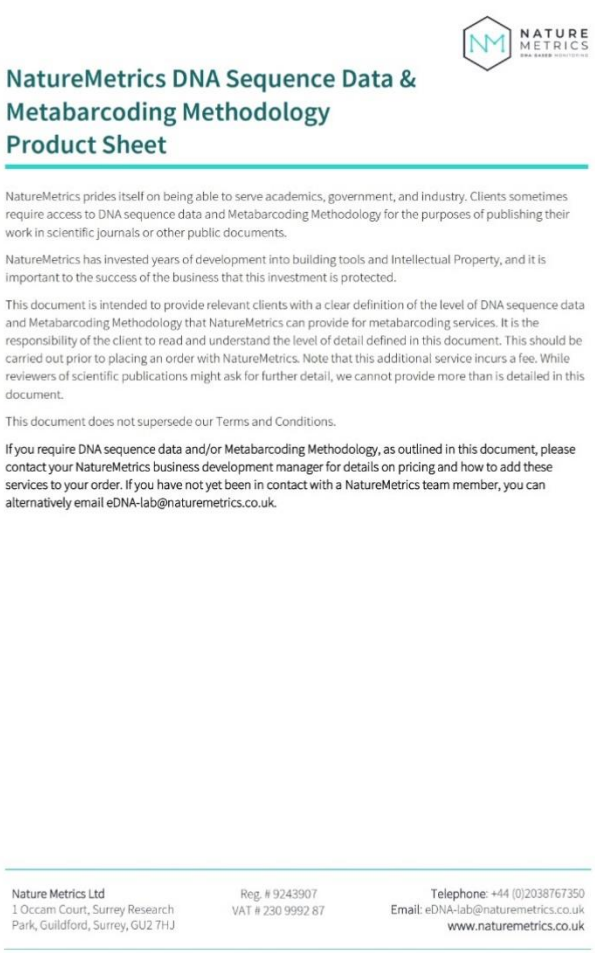
Other Key IP Considerations

- If additional DNA-based samples are collected in parallel with those funded by Scottish Government, for example if NatureMetrics collects additional samples, or is there is a need to collaborate with another project and/or previous SG research contractors for mutual benefit, then a data sharing agreement will need to be completed. However, the specific template used for data sharing purposes will need to be considered on a situational case-by-case basis. Such data sharing agreements will be between the contractor and the data owner/third party to ensure best available data can be used for the project outputs.
- The structure and format of all expected project deliverable outputs from both Phase 1 and Phase 2 produced by the contractor will align with the project specification document, as detailed by the MSG representing needs of the SG and its collective partner organisations. All specified project deliverable outputs including their respective file formats (e.g., PDF, Excel, Fastq, etc.) will be owned by the SG, made publicly available for future use and re-use purposes under Open Government Licence (OGL), stored and backed-up appropriately, and published via a SG project webpage (hosted by SEPA as the lead SG project partner e.g., a weblink with nested and/or signposted project deliverable outputs to be made publicly available by project completion).
- SEPA will be the Data Controller over any basic personal data (e.g., name, email address, employer organisation/affiliation) the project needs to capture and process from registered attendees for the project knowledge-exchange (KE) events and its related communications. The information submitted will be held and used by SEPA for organising and facilitating the Project Knowledge-Exchange (KE) Events and, if consent is provided, to send follow-up communications after the event and inform you about outputs of the project when these have been published. SEPA's purpose for collecting this information is so we can facilitate the event and provide you with an acceptable service. The lawful basis we rely on for processing your personal data is your consent under article 6(1)(a) of the UK GDPR. Personal data will be held securely and will only be used for the purposes noted above. It will be securely deleted within 6 months of completion of the project. By submitting a response, registered attendees are consenting to the use of personal information for the purpose described above. For more information on how SEPA handles personal information, please see also SEPA's general Privacy Policy (https://www.sepa.org.uk/help/privacy-policy/).

  - The contractors will be noting key comments and relevant stakeholder feedbacks (via chat function and as verbally communicated on the day) at each project KE event. This input will be considered and anonymised if it is to be used for informing the final project deliverable outputs emerging from Phase 2 work.

- Publication(s) of project findings in peer-reviewed academic journals are not specified project deliverable outputs from this project, nor is there a timeline/deadline expectation on when subsequent publications might be produced by. However, it is anticipated that the contractors may be interested in pursuing publication following successful project completion and when contractual obligations have been fulfilled. It is reasonably understood that key MSG and TSG/TRG members will be interested in contributing to relevant papers, for example where there is basis, motivation, and capacity to do so, and will need to be approached by NatureMetrics at the relevant time and dialogues held regarding who leading or coauthoring papers, on a case-by-case basis with NM and all the relevant parties involved. As such, academic papers led and coauthored by NM related to the project are expected to fully acknowledge the role of SG-RESAS as project funder as well as SG-CAMERAS and the Scottish DNA Hub as project commissioners, as well as extensive in-kind support provided by relevant people and organisations (specifying names and affiliations if needed), and the Advisory Board (specifying names and affiliations if needed). Similarly, if key MSG, TSG/TRG, or Advisory Board members wish to pursue academic publication, such academic papers led and coauthored by those members related to the project are expected to fully acknowledge NatureMetrics and will need to approach NatureMetrics at the relevant time and dialogues held regarding who leading or coauthoring papers, on a case-by-case basis with NM and all the relevant parties involved.
  - It is generally agreed that those involved in the project who have made a substantial contribution to the project and are interested in reasonably contributing to academic papers will be given the opportunity, by the relevant lead affiliate(s), to contribute as co-author(s) of project-derived publications after work is completed. Therefore, those parties will need stay in touch beyond the project lifespan to progress any publication plans, following a successful project completion in due course.
    - Example of author contributions e.g., X conceptualised the study. Y led the data analysis and wrote up the research supported by A, B, C, and D. Extensive reviewing and editing of the paper were undertaken by all authors. Z led the review process. We gratefully acknowledge […] for funding the project and […] for providing in-kind support/resources to the project.

7.2 NM Standard IP Policy



NATURE METRICS

NatureMetrics DNA Sequence Data & Metabarcoding Methodology Product Sheet

NatureMetrics prides itself on being able to serve academics, government, and industry. Clients sometimes require access to DNA sequence data and Metabarcoding Methodology for the purposes of publishing their work in scientific journals or other public documents.

NatureMetrics has invested years of development into building tools and Intellectual Property, and it is important to the success of the business that this investment is protected.

This document is intended to provide relevant clients with a clear definition of the level of DNA sequence data and Metabarcoding Methodology that NatureMetrics can provide for metabarcoding services. It is the responsibility of the client to read and understand the level of detail defined in this document. This should be carried out prior to placing an order with NatureMetrics. Note that this additional service incurs a fee. While reviewers of scientific publications might ask for further detail, we cannot provide more than is detailed in this document.

This document does not supersede our Terms and Conditions.

If you require DNA sequence data and/or Metabarcoding Methodology, as outlined in this document, please contact your NatureMetrics business development manager for details on pricing and how to add these services to your order. If you have not yet been in contact with a NatureMetrics team member, you can alternatively email eDNA-lab@naturemetrics.co.uk.

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

NATURE METRICS

Table 1. Services for which DNA sequence data and Metabarcoding Methodology can be shared. It should be noted that this list is subject to change. **NatureMetrics offers a range of other services – if they are not listed on this table they are not available for DNA sequence data or Metabarcoding Methodology sharing. It is important to also note that lab and bioinformatics methodology is only available for a subset of these services.**

Kit Family	Service	DNA sequence data	Metabarcoding Methodology
Aquatic[1]	Fish	✓	✓
Aquatic[1]	Vertebrates	✓	✓
Aquatic[1]	Mammals	✓	✓
Aquatic[1]	Aquatic Invertebrates (Freshwater eDNA)	✓	✓
Aquatic[1]	Aquatic Invertebrates (Marine eDNA)	✓	✓
Aquatic[1]	Unionid Mussels	✓	×
Aquatic[1]	Venerid Mussels	✓	×
Aquatic Bacteria	Aquatic Bacteria	✓	✓
Soil and Sediment[2]	Soil Bacteria	✓	✓
Soil and Sediment[2]	Sediment Bacteria	✓	✓
Soil and Sediment[2]	Soil Fauna	✓	✓
Soil and Sediment[2]	Sediment Fauna	✓	✓
Soil and Sediment[2]	Soil Fungi	✓	✓
Soil and Sediment[2]	Vertebrates from Sediment	✓	✓

[1] Aquatic Standard, Aquatic Maxi, Aquatic Pump, Aquatic Maxi Pump
[2] Soil and Sediment Cold Storage, Soil and Sediment Preservation Buffer

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

Metabarcoding DNA Sequence Data

For metabarcoding services listed in Table 1, NatureMetrics can provide demultiplexed, paired end fastq files, including base quality scores. To enable NatureMetrics to provide this service whilst protecting its Intellectual Property, all adapters and primer sequences will be removed prior to sharing. Henceforth this is referred to as "DNA sequence data". **For services not listed in Table 1 DNA sequence data cannot be shared.**

Sharing DNA sequence data is not part of the standard NatureMetrics service. It needs to be requested and incurs a fee. The fee is based on a single deliverable: delivery of a dataset and report. Where projects have multiple reporting deliverables (e.g. multi-season or multi-year projects), the fee will be applied to each deliverable.

For projects undertaken prior to the date of implementation of this service definition (September 2022), NatureMetrics does not guarantee the provision of DNA sequence data.

There is no guarantee that projects will be provided a dedicated Illumina run. Samples from multiple projects may be pooled on an Illumina run. It must be noted that sequences are occasionally known to be assigned to incorrect samples on Illumina sequencers at a minimal variable background rate. This is a known attribute of Illumina technology: Effects of Index Misassignment on Multiplexing and Downstream Analysis (illumina.com). In the NatureMetrics Bioinformatics and Reporting pipelines this is accounted for, and robust filters are applied. However, these filters are generated empirically and on a run-by-run basis in the context of the entire run dataset. It is not possible to apply these filtering thresholds to DNA sequence data from a partial Illumina run as the data is out of context of the entire run dataset. As a result, clients must be aware that sequences from other projects are likely to occur at a minimal variable background rate in DNA sequence data that would otherwise not appear in a client report.

Note also that sequences will be present in DNA sequence data that are not part of project reports or OTU tables. NatureMetrics Bioinformatics and Reporting pipelines remove known contaminants and non-target taxa (such as human) that are not deemed relevant to reporting outputs.

Provision of DNA sequence data from control samples used internally by NatureMetrics will not necessarily be provided. This is because control samples with DNA or PCR product below quantifiable levels are not sequenced.

FASTQ files will be provided by a customer specific, time-limited, download link. These will be provided in a format that has the internal NatureMetrics identifier and marker gene in the filename. A separate table will provide a link between the internal NatureMetrics sample ID and the client sample name. DNA Sequence data will be removed from the download link, 3 months after the data is released. It is the responsibility of the customer to ensure they have downloaded and securely backed up the DNA sequence data within this time frame. NatureMetrics will not retain responsibility for providing DNA sequence data after this window.

Sharing DNA Sequence Data Metadata

It must be noted that NatureMetrics cannot share metadata associated with DNA sequence data such as, but not limited to, the information listed below. This is because the information will pertain to samples from multiple projects and should not be used out of context.

- Sequencing run quality or quantity statistics
- Other sequencing run output information
- Details of any other samples or projects that were included on a sequencing run

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

Metabarcoding Methodology

It is expected that by requesting additional level of detail, at least one NatureMetrics scientific team member will be invited as a co-author of any resulting publications, where publications meet a standard deemed acceptable by NatureMetrics.

NatureMetrics has taken the decision to share detailed methodology for a selection of services only. This allows us to preserve our wider Intellectual Property while enabling clients to use NatureMetrics for public contracts or for industry clients that require more detail. Detailed methodology for any other services cannot be shared. This applies to current projects, future projects and to projects undertaken prior to the date of implementation of this service definition (September 2022). NatureMetrics reserves the right to change this service definition without notice.

This document is limited to our metabarcoding services, including data processing to generate OTU tables. It is limited to our Illumina-based approaches only. NatureMetrics does not share any methodology pertaining to statistical analyses or other analyses post generation of the OTU table.

Table 2. Methodological detail that NatureMetrics can share – specifically for the services indicated in Table 1. Detailed methodology for any other services cannot be shared. Please see the Appendix for a written example of the level of methodology that can be shared.

General	Can be shared (Y/N)
Model, make, manufacturer or supplier of lab equipment including pipettes, robotics, PCR machines	N
Any other laboratory or bioinformatics procedures not detailed in this document	N
Sequences of indexes or unique molecular identifiers	N
Methodology pertaining to statistical analyses or other analyses post generation of OTU tables	N
Primer sets	
Primer sequences	Y
Kits	
Non-liquid kit components	
Measurements, type of plastic/material and filter pore sizes for non-liquid kit components	Y
Name, brand, supplier	N
Liquid kit components	
Where liquid components are off-the-shelf brand items	
Name, brand, supplier	Y This applies to the Aquatic Bacteria Kit
Where liquid components are custom-ordered recipes	

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

Composition of liquid/buffer	N This will be referred to only as "preservation buffer"
Name, brand, supplier	N This applies to Aquatic Kits (not including the Aquatic Bacterial Kit) and the Soil and Sediment Preservation Buffer Kit
DNA extraction	
Duration and temperature of DNA extraction steps	Y
Where DNA extraction kits are off-the-shelf brand items	
Name, brand, supplier	Y
Where reagents are off-the-shelf brand items	
Name, brand, supplier	Y
Concentration/volume/duration of application, where different from manufacturer recommendations	Y
Where liquid components are custom-ordered recipes	
Name, brand, supplier	N
Concentration, volume applied to sample, duration of application	Y
PCR	
PCR total volumes	Y
Number of PCR replicates	Y
Name, brand, and concentrations of PCR reagents	Y
Temperatures, cycles, and step durations of PCR cycling conditions	Y
Library preparation	
An overview of the library preparation steps	Y Note that for the services listed in Table 1 NatureMetrics follows the Illumina 16S Metagenomic Sequencing Library Preparation protocol, with some modifications.
Purification steps	Y
Final loading concentration	Y
The Illumina sequencing reagent kit used	Y
The Illumina sequencer used	Y
Sequences of indexes or unique molecular identifiers	N
Bioinformatics	
An overview of bioinformatics processing steps	Y
Any bioinformatics code whatsoever	N

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

Names of software used to demultiplex, merge paired-end reads, trim primers/adaptors/Unique Molecular Identifiers/Indexes, perform quality filtering, dereplicate, denoise	Y
Specific programme settings	N
Overview of steps that used in-house bioinformatics software	Y
Specific programme settings for in-house software	N
Sequence clustering threshold (where relevant)	Y
Names of the public databases used for taxonomic assignment	Y
Indication that in-house databases were used	Y
Specific details on the content of in-house databases	N
For taxonomic assignment using BLAST, the e-score, query cover and minimum similarity thresholds	Y

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

Appendix – Example of level of detail that can be supplied

Note that this is intended to be an example only. Different services have differing workflows, and these can be subject to alterations.

Example: Freshwater Aquatic Vertebrates

Field Methods

Note that this section is given to provide context – field sampling approaches will be project specific and carried out by the client. A field sampling section will not be provided in reports.

Five 3.3 L water samples were collected from each lake at approximately equidistant locations where possible, taking shoreline access by foot into consideration. Each sample was comprised of 10 x 330 mL subsamples, with subsamples taken at 10 m intervals along 100 m of shoreline. Once subsamples were deposited in the sterile sampling bag, the bag was sealed and shaken for 20-30 seconds to homogenise the water sample. Water was filtered using a NatureMetrics Manual MAXI Aquatic eDNA Kit, where 1.5 L of water was passed through both 0.8 µm PES filters supplied with the kit where possible. The volume of water passed through each filter was recorded.

Water was sampled at three locations along the river. At each location, three 2 L water samples were collected from the left bank, mid-channel, and right bank. Subsequently, 1 L from each sample was pooled into a new sampling bag to create a 3 L composite sample, leaving three 1 L independent samples. One litre from each independent and composite sample was passed through a 0.8 µm PES filter supplied with the NatureMetrics Manual Aquatic eDNA Kit.

All samples were transported to the NatureMetrics laboratory at ambient temperature and stored at -20 C until analysis.

DNA extraction

Before and after each step, all benches were decontaminated with CHEMGENE HLD4L wipes (STARLAB). Each step of the process had its own designated space, equipment, reagents, and consumables. DNA was extracted from each filter using a DNeasy Blood and Tissue Kit (Qiagen) with the following modifications: initial lysis happens on the filter to minimise potential contamination risk and a higher lysate volume is taken through in subsequent steps to maximise DNA yield. An extraction blank was processed with each batch of extractions to assess potential contamination in the extraction process. DNA was purified to remove PCR inhibitors using a DNeasy PowerClean Pro Cleanup Kit (Qiagen). Purified DNA extracts were quantified using a Qubit dsDNA HS Assay Kit on a Qubit 3.0 fluorometer (Thermo Scientific).

PCR amplification

Vertebrate service: A region of the 12S ribosomal RNA gene was amplified via a two-step PCR process. For each water sample, 12 PCR replicates were performed using a 12S primer set (Riaz et al. 2011).
The vertebrate eDNA amplification mixture contained 1X DreamTaq Green PCR Master Mix (Thermo Scientific), 0.4 µM of each primer, 1.5 mM of $MgCl_2$ (Thermo Scientific), 0.8 mg/ml of Bovine Serum Albumin (BSA, Thermo Scientific), 3% of Dimethyl Sulfoxide (DMSO, Thermo Scientific), 0.9 µl of template DNA, and PCR grade water (Thermo Scientific). Total PCR volume was 8uL.

Vertebrate eDNA PCR conditions consisted of an initial denaturation at 95 C for 2 min; 10 cycles at 95 C for 20 s, a 30 s touchdown annealing step (-0.5 C per cycle) starting at 60 C, then 72 C for 40 s; 35 cycles of 95 C for 20 s, 55 C for 30 s, and 72 C for 40 s; and a final elongation step at 72 C for 5 min. PCR positive controls (i.e., a mock community with a known composition of non-native species) were included to verify sequence quality and PCR negative controls (i.e., PCR grade water) were included to detect potential cross-contamination. Amplification success was confirmed via gel electrophoresis.

Sequencing

All PCR replicates per sample were pooled and purified using Mag-Bind® TotalPure NGS (Omega Bio-tek) magnetic beads. A sequencing library was prepared from the purified amplicons using a combinational dual index approach, following Illumina's 16S Metagenomic Sequencing Library Preparation protocol but using 1X DreamTaq PCR Master Mix (Thermo Scientific). Indexed PCR products were again purified using Mag-Bind® TotalPure NGS (Omega Bio-tek) magnetic beads. The purified index products were quantified using a Qubit dsDNA BR Assay Kit, normalized and pooled. The pooled purified index PCRs were sized using a TapeStation D1000 ScreenTape System (Agilent). The libraries were sequenced on an Illumina MiSeq with a V3 MiSeq Reagent kit, the final library was loaded at 10 pM with a 20% PhiX control spike.

Bioinformatics

Sequences were demultiplexed with bcl2fastq and subsequently processed via a single analysis pathway with parameters optimised for each assay where appropriate. Paired-end FASTQ reads for each sample were merged with USEARCH (Edgar 2010). Forward and reverse primers were trimmed from the merged sequences using cutadapt (Martin 2011) and a length filter applied as appropriate for the assay. These sequences were quality filtered with USEARCH and dereplicated by sample, retaining singletons to obtain zero-radius OTUs (ZOTUs). Unique sequences from all samples were denoised in a single analysis with UNOISE (Edgar 2016).
Taxonomic assignments were made via sequence similarity searches of the ZOTU sequences against reference databases appropriate for the assay. The NCBI nucleotide (NCBI *nt*) database was queried for the vertebrate analysis. Searches against NCBI *nt* were made using blastn (Altschul et al. 1990; Camacho et al. 2009) and required hits to have a minimum e-score of 1e-20 and cover at least 90% of the query sequence. The taxonomic identification associated with all hits was converted to match the GBIF taxonomic backbone to allow results from different databases to be combined.

Assignments were made to the lowest possible taxonomic level where there was consistency in the matches, with minimum similarity thresholds of 99%, 97% and 95% for species, genus, and higher-level assignments respectively. Country-based sense-checking against GBIF occurrence records was used to manually improve identifications in cases where there were equally good reference matches (rgbif; Chamberlain et al., 2022).

ZOTUs were then clustered at 97% similarity with USEARCH to obtain OTUs. An OTU-by-sample table was generated by mapping all dereplicated reads for each sample to the OTU representative sequences with USEARCH at an identity threshold of 97%.

All OTUs with species-level identifications were queried against the IUCN Red List (rredlist; Chamberlin 2018) to obtain global threat status. Finally, low abundance detections were omitted, with filter thresholds set at a percentage of the total reads per sample (0.02%). Results are presented for OTUs identified to the target taxonomic group only (vertebrate: Chordata excluding human and domestic animals).

Date	12/09/2022
Version no.	1.0

Nature Metrics Ltd
1 Occam Court, Surrey Research Park, Guildford, Surrey, GU2 7HJ
Reg. # 9243907
VAT # 230 9992 87
Telephone: +44 (0)2038767350
Email: eDNA-lab@naturemetrics.co.uk
www.naturemetrics.co.uk

7.3 GapFinder Analyses

GapFinder is an in-house software developed by NatureMetrics. It checks whether reference sequences are available for taxa using a particular metabarcoding analysis. The input taxa list is initially checked against GBIF and the taxonomy recovered. Sequences for the relevant genes (e.g. 18S and COI) are downloaded from publicly available databases (e.g. NCBI) for the target taxa. These sequences are then aligned with the primers for the metabarcoding analysis and if there is under 20% dissimilarity, considered to possibly amplify and allow for identification of the taxon. This allows us to ascertain whether there is a possibility of identifying taxa and/or which assay is optimal for a list of target taxa (as demonstrated by NatureMetrics changing the marine invertebrate assay between the pilot and main sampling campaign).

Gapfinder analyses were run on the priority species for this project, such as PMF species and PMF-habitat characterising species (Tyler-Walters et al. 2016), UK vertebrates (taken from GBIF), invertebrates from the UK Checklist of freshwater species (Gunn et al. 2018), and non-lichenised fungi used for the designation of SSSIs with the relevant metabarcoding assays ((Bosanquet et al. 2018). The output is provided in a separate Excel spreadsheet.

Reference sequences were predicted to be available for:

- all vertebrate PMF species with the vertebrate assay except *Lampetra fluviatilis*
- all PMF fish species with the exception of *Centrophorus squamosus* and *Centroscymnus coelolepis* with the fish assay
- 75% of UK vertebrates with the vertebrate assay
- 8% of non-vertebrate PMF species with the eukaryotes assay and 15% with the invertebrates assay. This showed the benefits of running both assays
- 69% of freshwater invertebrate UK Checklist species

7.4 Soil Fungal:Bacterial Ratio Metric

Bacteria and fungi are the two major groups of decomposers of organic matter in soil and play an important role in nutrient cycling. Bacteria and fungi are physiologically distinct; fungi utilise high molecular weight compounds like cellulose and lignin and are more efficient at assimilating and storing nutrients than bacteria. Soils with a higher fungal:bacterial (F:B) ratio are typical of more advanced stages of the ecological succession towards a mature state and are indicative of an abundance of stable and recalcitrant carbon in the soil (Osburn et al. 2019). A higher F:B ratio indicates a more sustainable or less disturbed soil ecosystem (Osburn et al. 2019). A higher F:B ratio is also associated with higher levels of organic matter, which is beneficial for the production of healthy plants and the resilience of an ecosystem. The F:B ratio has traditionally been measured using phospholipid fatty acid (PLFA) analysis and microscopy and is a widely adopted microbial indicator of environmental change and its impact on ecosystem functioning. The molecular soil F:B ratio is a metric that can be calculated using quantitative PCR (qPCR).

Midway during this project, we decided to trial this analysis on the terrestrial soil samples. However, due to a technicality that analysis should occur within a week of DNA extraction, it was too late to apply the analysis. We did however perform an initial test on a batch of 15 soil samples, using nine samples

from Glen Falloch (woodland) and six samples from Glen Finglas (peatland). But unfortunately, the results were of insufficient quality to draw any conclusions from. The samples did not amplify at all, or the amplification was below the limit for the fungi qPCR assay. Therefore, the batch test deemed it was not viable to progress with the other soil samples. Results are therefore not included nor discussed further in this report. However, we do highly recommend adding soil F:B ratio analyses to future woodland monitoring programmes in Scotland, for example to track the efficacy of restoration efforts over time. However, a key lesson learned from undertaking the batch test analysis work on this project is that it should only be applied if samples can be analysed within a week of DNA extraction.

7.5 Marine Benthic Bacterial Index

The analysis of marine benthic community structure, based on eDNA metabarcoding, has emerged as a powerful alternative to microscopy-based taxonomic approaches to monitor disturbances in coastal environments. Frühe et al (2021) have identified universal bacterial core taxa that indicate high, moderate, and low impact, regardless of sampling season, sampled country, seafloor substrate type, or local farming and environmental conditions. The study was aimed at aquaculture disturbances in coastal environments (including Scotland). We trailed the same bacterial index on our eDNA-based marine sediment bacterial dataset. Three of the families that the study identified as core taxa were detected in our datasets. All three families are associated with the 'low impact' category (Frühe et al. 2021). One of those families, Turicibacteraceae, was detected in every sediment sample. Based on this observation, none of the sites are impacted by aquaculture. The other two families are Microbacteriaceae and Polyangiaceae, and both were detected in only one sample. When expanding the scope of eDNA-based monitoring in the Scottish marine environment, beyond the high-quality marine lochs sampled in this project, to encompass a pollution gradient, this method would be a good candidate for tracking pollution in coastal and pelagic habitats.

7.6 Source and description of data used in this project

Habitat	Data Description	Source	Reason for Inclusion	Location in Deliverables	Data licensing considerations
Marine	Morphological invertebrate analysis from sediment grabs - taxa list	SEPA morphological analysis undertaken in this project	Required for having a biotope "truth", to compare eDNA methods against conventional method.	Not included. If required, requests must be made to SEPA	Must follow SEPA's licencing conditions. General (sepa.org.uk)
Marine	Particle Size Analysis sediment results	SEPA analysis undertaken in this project	Required for having a biotope "truth", to compare eDNA methods against conventional method.	Not included. If required, requests must be made to SEPA	Must follow SEPA's licencing conditions. General (sepa.org.uk)
Marine	Biotope Classifications	Moore et al. Allen et al.	This was required to choose sampling points within the marine habitat	Excel metadata tables	None
Marine	CTD data, conductivity (salinity), temperature, depth	Collected by Marine Science Scotland during this project	Used in statistical models.	Excel metadata tables	None
Peatland/ Woodland	soil moisture and pH	NM-collected data	These environmental covariates can affect the detection of species. Used in statistical models.	Excel metadata tables	None
Peatland/ Woodland	Elevation	NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. ASTER Global Digital Elevation Model V003. 2019, distributed by NASA EOSDIS Land Processes	These environmental covariates can affect the detection of species. Used in statistical models.	Excel metadata tables	All LP DAAC current data and products acquired through the LP DAAC have no restrictions on reuse, sale, or redistribution.

		DAAC, https://doi.org/10.5067/ASTER/ASTGTM.003. Accessed 2023-01-14.			
Woodland	Data from aligned Forest Research project	Forest Research	This refers to the sites that were surveyed as part of the FR project. It was used as part of the overall study to get an equal number of woodland categories	Excel metadata tables. Excel OTU tables	This data was shared under a data sharing agreement between NatureMetrics and Forest Research and is subject to same conditions as the other OTU and metadata in the report.
Freshwater	CPET scores from previous years	SEPA	To compare CPET values with best-matching available data	Only included as summarised data in the main report	Must follow SEPA's licencing conditions. General (sepa.org.uk)
Freshwater	Land use in catchment %s	Catchment areas for each loch were obtained from SEPA (River and loch waterbody nested catchments: https://www.sepa.org.uk/environment/environmental-data/), and 20 m resolution land use raster from Space Intelligence (https://www.space-intelligence.com/scotland-landcover/).	To investigate the effects of land use	Excel metadata tables	Maps © Space Intelligence Ltd. Reuse and modification are permitted, as well as commercial and non-commercial exploitation, providing Space Intelligence is named, following Open Government 3.0 licence https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ Must follow SEPA's licencing conditions. General (sepa.org.uk)
Freshwater	WFD Loch Statuses (Overall Status)	SEPA https://www.sepa.org.uk/data-visualisation/water-classification-hub/	This was required to choose sampling points within the freshwater habitat in order to answer the pertinent research questions.	Not included. If required, requests must be made to SEPA	Must follow SEPA's licencing conditions. General (sepa.org.uk)

Freshwater	Fish data – list of species known from each loch	Various (see Technical Appendix)	Qualitative exploration of the number of species detected in this study compared to known species occurrences.	Only included as summarised data in the main report	None
Freshwater	Mean annual values of Alkalinity, Chlorophyll-a, Nitrite, Nitrate, Oxygen-dissolved, pH, Total Phosphorous	SEPA	Needed for water quality gradient of freshwater lochs	Not included. If required, requests must be made to SEPA	Must follow SEPA's licencing conditions. General (sepa.org.uk)
Freshwater	Mean alkalinity and mean depth (m)	UK Lakes Portal	Used for exploration of Fish Classification Index	Excel metadata tables	None
ALL	Raw sequence data	NatureMetrics-generated data	A project deliverable	Separate fastq files	
ALL	OTU tables	NatureMetrics -generated data	A project deliverable	Excel OTU tables	

edna-lab@naturemetrics.com

+44 2038 767350

naturemetrics.co.uk

1 Occam Court, Surrey Research Park,
Guildford, GU2 7HJ United Kingdom

1590 Hanlon Creek Boulevard, Unit 11,
Guelph, Ontario, N1C 0A1, Canada

NATURE METRICS

DNA-BASED MONITORING