PHASE 1 PILOT STUDY FINDINGS & PHASE 2 SAMPLING PLAN - DEVELOPING HABITAT SCALE DNA MONITORING IN SUPPORT OF POST 2020 BIODIVERSITY REPORTING REQUIREMENTS

NatureMetrics
January 2023

PROJECT FACILITATION

Authors

Bastian Egeter[1]
Hayley Craig[1]
Lynsey R. Harper[1]
Edward Wort[1]
Judith Bakker[1]

[1] NatureMetrics, 1 Occam Court, Surrey Research Park, Guildford, GU2 7HJ, 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: Egeter B., Craig H., Harper L. R.., Wort E., Bakker, J. (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

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 1 Project Management Steering Group: Willie Duncan (Scottish Environment Protection Agency, SEPA), Colin Bean (NatureScot), Iveta Matejusova, and Philip Boulcott (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 (Marine Scotland Science, MSS), Willie Duncan, Pauline Lang, and Alistair Duguid (Scottish Environment Protection Agency, SEPA), as well as Nadia Barsoum[§] (Forest Research), Malcolm Starkey[§] (The Biodiversity Consultancy), and Tom Butterworth[§] (WSP) for providing project guidance, technical advice and peer reviewing this work. We also thank a broad range of project stakeholders including key representatives from the Phase 1 Project Technical Steering Group; the Project Advisory Board; the Scottish DNA Hub; Loch Lomond and the Trossachs National Park (LLTNP); and other organisations for 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 (LLTNP); Tom Butterworth[§] (WSP); Simon Franks (LLTNP Trees and Woodland advisor); Doug Yu[§] (University of East Anglia, NatureMetrics) for providing their specialist expertise, and feeding ideas or suggestions into project development. We are extremely grateful to Marine Scotland Science, 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 Marine Scotland Science; 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. For the freshwater fieldwork, we thank Hannele Honkanen (Scottish Centre for Ecology and the Natural Environment, University of Glasgow) and David Scott-Park (Portnellan Farm) for permitting sampling at Loch Lomond shoreline sites. For the peatland fieldwork we thank 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 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. 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 (CAMERAS) Board, in collaboration with the Scottish DNA Hub, for commissioning this project and their ongoing support.

[§]Project Advisory Board Member

Table of Contents

1 Introduction

1.1 The biodiversity monitoring challenge

Biodiversity loss is widely recognised as one of the most urgent global challenges to be addressed in the next decade. One element of the Global Biodiversity Framework is the 30x30 target, which aims to protect 30% of the planet for nature by 2030. Committed countries are expected to contribute to this global goal through domestic action to increase coverage of effectively managed protected areas. The Scottish Government 2020 Statement of Intent on Biodiversity outlined the commitment to the 30x30 target. Additionally, the 2021 Programme for Government committed to the deployment of Nature Networks. These two programmes are key components in increasing ecological connectivity and restoration of nature more widely, helping to deliver the Scottish Biodiversity Strategy. These two programmes are key components in increasing ecological connectivity and restoration of nature more widely, helping to deliver ambitions of the Scottish Biodiversity Strategy which was published in 2022 (https://www.gov.scot/publications/scottish-biodiversity-strategy-2045-tackling-nature-emergency-scotland/).

However, 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 speed, scope, and scale of data collection to better monitor progress relative to actions and interventions to better inform future responses. DNA-based monitoring uses high-throughput DNA sequencing to rapidly characterise the species diversity present in mixed-species and environmental samples and has the potential to be truly transformative in this regard, enabling biodiversity to be measured and monitored at large geographic scales. It generates sufficient data for the application of ecological statistics and holds the potential for development of new biotic indicators for a wide range of habitats and geographies. However, uncertainties remain around how to best utilise DNA-based monitoring approaches considering scientific, operational, and fiscal perspectives.

This report represents the second output of the project 'Developing habitat scale DNA monitoring in support of post 2020 biodiversity reporting requirements' funded by the Scottish Government (Reference: NMP/001/20). Its main purpose is to outline the (i) initial project findings, including key learning opportunities, from the Phase 1 pilot study sampling and analysis carried out during 2021, and (ii) sampling plan design and rationale for Phase 2 of this project.

1.2 Aim & Key Objectives

The overall project aim is to investigate and test the applicability of DNA-based monitoring approaches for biodiversity assessment and reporting purposes across a broad range of habitat types in Scotland, including a mix of aquatic and terrestrial ecosystems, specifically those situated in and around Loch Lomond and the Trossachs National Park (LLTNP). The project facilitates scientific evidence and practical experience to inform next conceivable actions in DNA-based method development and implementation that may be taken forward by the Scottish Government and its collective organisations. This will help end-users to further develop the potential utility of DNA-based monitoring to set goals for protecting and restoring nature. This is needed for tackling the twin climate and ecological crises in both innovative and cost-effective ways.

The key objectives of this project are:

Phase 1

1.1 To document the current scientific overview of using DNA-based monitoring technology to understand how these methods can contribute to biodiversity reporting, against global goals and targets, in the context of national and international biodiversity reporting frameworks.

1.2 To codesign and deliver a DNA-based sampling plan for use in Phase 2, in a form that outlines the sampling strategy and rationale for a Phase 2 eDNA survey, undertaken in and around LLTNP during summer 2022, to inform future habitat monitoring programmes.

Phase 2

2.1 To refine and deliver a Phase 2 eDNA survey, which aimed to test DNA-based metabarcoding approaches and statistical analysis across four key Scottish habitat types from aquatic (marine lochs, freshwater lochs) and terrestrial (woodland, peatland) sampling locations in and around LLTNP.

2.2 To create reasonable opportunities for enhancing knowledge-exchanges (KE) between the contractors, the key stakeholders, and lead requesters to (a) ensure that a range of end-users are engaged in learning-feedback loops and able to interpret and implement the outcomes of the Phase 2 eDNA survey and inform future habitat monitoring programmes going forward; and (b) inform the production and use of project outputs.

As part of Objective 1.2, an initial small-scale pilot sampling campaign was conducted between August – October 2021. The aim of the pilot studies was to generate preliminary eDNA field data to guide the methodological and analytical approaches to be employed in Phase 2 of the project.

This report provides the summary outcomes of the four pilot studies (with further details included as appendices), and, utilizing this information, as well as information gathered during

consultation with relevant stakeholders, practitioners and subject matter experts, the report also details the sampling strategy and rationale for the Phase 2 eDNA survey that was conducted during summer 2022 (Objective 1.2). The sampling strategy was designed to align as closely as possible, where practically feasible, to respond to the overall project ask and scope.

1.3 Scope for DNA-based Sampling & Analysis

There are numerous potential approaches that could have been taken to trial the use of DNA-based monitoring in the context of this project, each of which has merits. In consultation with key project stakeholders, and to the best of current knowledge from relevant published work, we identified and chose a set of sampling strategies that were, considered to be the most relevant with regards to enhancing existing biodiversity monitoring approaches, whilst also applying novel approaches for addressing the project's overarching aim and objectives.

The project had the resources to process 300-500 samples across four habitats and carry out ~900-1000 analyses (200 – 400 analyses per habitat type). It was not considered within the scope of the project to develop new products, bioinformatics pipelines, or analytical services. Nor was it within scope to modify existing bioinformatics pipelines. Additionally, as the project has wide-ranging goals across multiple taxonomic groups, the scope was restricted to multi-species metabarcoding only.

As the project focuses primarily on eDNA, the development of long-term monitoring approaches reliant on direct sampling of DNA from flora or fauna (e.g., biofilms, benthic diatoms, phytoplankton or aquatic plant communities, and bulk invertebrate sampling) was considered out of scope, except for conducting validation steps where needed such as taxonomic/morphological bulk invertebrate and particles size analyses (PSA) in the marine habitat, with extensive in-kind project sampling and analysis carried out by Marine Science Scotland and SEPA Marine Ecology respectively, required for producing a comparative biotope classification (as a proxy for marine habitat condition).

1.4 Focal Study Area

The project study area was mostly focussed on the Loch Lomond and Trossachs National Park (LLTNP) and its adjacent areas (https://www.lochlomond-trossachs.org/). The LLTNP was chosen specifically to investigate and test the applicability of DNA-based monitoring approaches for biodiversity assessment and reporting purposes because it contains a mosaic of different habitat types occurring in Scotland, including a mix of aquatic and terrestrial ecosystems. Furthermore, a broad range of 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), exist within the LLTNP which makes this area a desirable candidate location to test the DNA-based approach at habitat scale (https://www.nature.scot/professional-advice/protected-areas-and-species/protected-areas).

However, as the project developed during Phase 1, the geographic radius for undertaking DNA-based sampling and analysis was relaxed and expanded beyond LLTNP boundaries in Phase 2. The purpose of this was to enable adequate study design and establish more habitat condition replicate sites where needed for the freshwater, woodland, and peatland habitats. For example, additional woodland habitat sampling was undertaken at some locations within the Cairngorms National Park (https://cairngorms.co.uk/), capturing another area of conservation importance in Scotland. On the other hand, for the marine habitat, a relatively small geographic area of Loch Goil and Loch Long was selected to ensure research questions could be answered, taking on board both budgetary constraints, extent of in-kind SG resources needed to support the sampling and analysis (from MSS and SEPA respectively), as well drawing on relevant stakeholder knowledge to strategically avoid duplication of research efforts due to parallel R&D work elsewhere in Scotland.

1.5 Key Elements of the Phase 2 eDNA survey

To meet the project aim and key objectives, the fundamental elements of the Phase 2 eDNA survey were;

- To include sites that were geographically disparate (within the scope and operational confines of the project) but could be pre-categorised according to habitat condition and/or habitat classification.
- That sites falling within a single habitat condition and/or habitat classification category were not geographically clustered.
- To collect replicate samples for each site over and above what would be expected in future habitat monitoring programmes, in order to determine to minimal number of samples needed for future monitoring.

Designing the sampling strategy with these fundamental elements allowed the project to;

- Have expected hypotheses, e.g. ecological communities will be significantly different in composition (and other derived metrics such as species richness) among highly or mostly, intermediately, and least disturbed habitat conditions.
- Minimise the possibility that any observed statistical differences across habitat conditions/classifications were not purely an artefact of being geographically separated. In other words, if there are strong observed differences in composition or derived metrics among habitat conditions/classifications, this is not just because we chose highly or mostly disturbed condition sites in one location and least disturbed condition sites in a different location. In fact, if a strong difference or trend is observed across habitat conditions/classifications, then this would be *despite* geographic separation, so any such observations would provide strong evidence that the metrics are a reliable indicator of habitat condition across regional and likely broader scales and could be taken forward for future habitat monitoring programmes.

- Make recommendations on the minimum sampling effort required to take forward to future habitat monitoring programmes - as such programmes may be operationally constrained in terms of resources available to undertake field sampling per site and any subsequent laboratory analyses of samples collected.

It should be noted that the project applied numerous statistical analyses to the data to investigate which approaches reveal the most robust, and insightful, outputs. These analyses are not detailed in this document but include comparisons of both community composition and derived metrics such as functional diversity, phylogenetic diversity, and health indices. Such metrics must be robust but relatively simple to implement and understand if they are to be accepted broadly by relevant agencies.

2 Pilot study key outcomes

This section presents a brief summary of the key findings and learning opportunities from the pilot studies during Phase 1. For detailed methods and results for each habitat, see the appendices supplemented to this document and supplementary files (NM-DLY727 - ScotGov - Peatland pilot report appendices.xlsx, NM-MEA989 - ScotGov - Freshwater pilot report appendices.xlsx, NM-PPP825 - ScotGov - Marine pilot report appendices.xlsx, NM-RUZ072 - ScotGov - Woodland pilot report appendices.xlsx). The coordinates of each of the sampling sites for all four habitats have also been recorded in the supplementary file appendices. See also Table 1 in this section for an overview of the initial pilot study design, habitats sampled, and metabarcoding samples collected for analysis across the various aquatic (marine and freshwater) and terrestrial (woodland and peatland) ecosystems located within the LLTNP focal study area.

The pilot study was conducted to gather eDNA field data to guide the choices for methodological approaches to be employed in Phase 2. Cross-cutting questions being asked of the pilot study were:

- Do the eDNA primers initially being trialled provide useful taxonomic information in the context of long-term monitoring in each habitat?
- Do the resultant ecological communities provide an *indication* that they are likely to be different across habitat conditions/classifications?

2.1 Marine

Loch Long, including its arm of Loch Goil on the western side, is a sea loch situated within the LLTNP. The marine pilot study aimed to detect Priority Marine Feature (PMF) species using eDNA to enhance species inventories and detections, and to assess whether communities showed discernible biodiversity differences across different biotopes (as a proxy for habitat condition). Accordingly, we specifically wanted to verify whether the invertebrate and fish assays were appropriate for the detection of (as many as possible) PMF species. Additionally, to assess whether multiple samples were required within a station to 1) maximize the number of species detected and 2) compare communities across biotopes, we aimed to investigate the variability between and among stations.

Loch Goil has previously been surveyed for fish and macroinvertebrates using traditional monitoring approaches, by underwater camera footage and morphological identification of macroinvertebrates in benthic grabs samples (Allen, C., Axelsson, M., Dewey, S. & Clark, 2013; Moore, 2013). Samples were collected at three stations (two in Loch Goil and one in Loch Long), sites previously covered by these surveys. We focussed on benthic sampling as benthic and nektonic species are most commonly targeted for PMF surveys. Water sampling was carried

out as close to the substrate as possible (this was adapted during the main sampling campaign as only the fish assay was used on the water samples), without disturbing the substrate (three replicates per station) and samples were collected with Niskin bottle and processed according to NatureMetrics protocols. Sediment was collected from the same three stations, using a mini Van Veen grab (0.1 m^2). Two grabs were collected per station and samples were collected and processed according to NatureMetrics protocols. The remainder of the grab content was brought to shore, to be sieved through a 1 mm mesh and bulk macroinvertebrate samples were stored in a Ziplock bag fixed with 100% using with a 70% ethanol in 1L bottles. This was transported to the NatureMetrics laboratory, and subsequently sent to a specialist contractor for taxonomic/morphological analysis.

Bacteria (Caporaso et al., 2010), invertebrate (Capra et al., 2016), and fish (Miya et al., 2020) assays were used for the water samples. Bacteria and invertebrate analyses were used for the sediment samples.

Ten PMF fish species and two PMF marine mammal species were detected with the fish assay in the water samples. No PMF species were detected with the invertebrate assay in either the water or sediment samples. However the second statement had to take into consideration the fact that the invertebrate assay did not provide species-level resolution for most taxa. No species or genera were shared between the bulk sample, identified by morphology, and metabarcoding-identified taxa, while 44% of families were common to both methods.

Freshwater fish were detected in the sampling site closest to the head of Loch Goil (Station 3), while in Loch Long only marine fish species were detected (Station 1). Invertebrate and bacterial communities were discernible between stations, and this difference was more evident using sediment rather than water samples. In terms of variability within stations, although community patterns were similar among replicates, presence of rarer taxa varied between replicate water samples collected at the same stations. In particular, composition of bacteria in water samples showed a high degree of variation between replicates collected at the same station.

The pilot study relied on previous classifications of biotopes for sampling location selection and there were two complications with this: 1) Station 1 was classified differently according to two separate sources (Allen, C., Axelsson, M., Dewey, S. & Clark, 2013; Moore, 2013) and; 2) the actual sampling location for Station 2 was closer to a station originally classified as a biotope other than that intended.

Key outcomes

Do the eDNA primers initially being trialled provide useful taxonomic information in the context of long-term monitoring in each habitat?

The fish assay used is currently established as one of the best available, as it results in relatively low amounts of cross amplification of other vertebrates and captures high species diversity (Collins et al., 2019; Miya et al., 2020). Based on the results from the pilot study, it was determined to be sufficiently useful to monitor fish PMF species and it was decided to also use this assay for the Phase 2 eDNA survey. The invertebrate assay (Capra et al., 2016) that was trialled did not detect any of the PMF species, nor did it identify many sequences to species

level. Detection of PMF invertebrate species is one of important aspects of marine monitoring and designation of marine protected areas and it was decided that alternative primers would be used for the Phase 2 eDNA survey. Additionally, it was decided to only use the fish assay on the water samples and reallocate effort to the sediment samples to capture as much invertebrate diversity as possible from the sediment samples (as most invertebrate PMF species are benthic in the study area).

Do the resultant ecological communities provide an indication that they are likely to be different across habitat conditions/classifications?

Using the bacteria assay it was possible to discern sampling station, and potentially biotopes, from either water or sediment. However, even though only a small number of samples were collected, the results suggest that there is a high degree of variation between replicate water samples within a station for bacteria. Additionally, collecting water samples for bacterial analysis requires a specialized filter, with a smaller pore size, and therefore increases cost and sampling effort. It was decided that the bacterial assay would only be applied to the sediment samples in Phase 2.

Due to the complications arising from relying on previous classifications of biotopes, and the limited number of sites in the pilot study, it was not possible to answer definitively whether the ecological communities were different enough between biotopes and similar enough within biotopes to enable long-term monitoring. It was also not possible to determine the minimum number of replicates needed per station. Furthermore, the variability in detection of rare species was high within stations, and detection of such rare species is important for Marine Protected Area management. As a result, it was decided that 3 replicate samples would be used for the Phase 2 eDNA survey for both sediment and water samples to answer this question fully. Additionally, analysis of morphological invertebrate samples and Particle Size Analysis would be conducted on sediment samples to allow for more precise classification of biotopes.

Additionally, to remove the potentially confounding effects of freshwater influences on fish community composition (observed at Station 3), it was decided that sampling would be conducted predominantly in Loch Long to limit the impact of DNA originating from freshwater input into the loch. Moreover, as collecting water samples close to the seabed can potentially stir up the sediment that may contaminate the water samples and clog the water filters, and because only the fish assay would be applied to the water samples in Phase 2, it was decided to collect the water samples from ~25 meters depth using a vertical Niskin bottle. As within the project it was only possible to collect water samples from a single depth, we had to decide on a sampling depth that would not be complicated and time consuming to reach, and that would be achievable at every sampling site.

The full marine pilot study report can be found in Appendix 4.1

2.2 Freshwater

Loch Lomond is a major freshwater loch, including one of its tributaries the Endrick Water, situated in LLTNP, and it is also the largest loch in the UK by surface area. Both locations were selected as focus sites for the pilot study, as during the initial phase of the project there was an open question on whether to focus on lakes or rivers, and there were not the resources to include both. Loch Lomond has distinct ecological zones as well as contrasting geology, bathymetry, and land use between its north and south basins. The loch spans a habitat pressure gradient with disturbance from eutrophication (for example water quality impacts due to nutrient inputs from diffuse pollution), generally increasing from the rurally populated north towards the more densely populated south basin (Murphy et al., 1994). The Endrick Water is a river flowing into the eastern end of Loch Lomond and experiences water quality impacts from, for example, sewage treatment works and broad-scale land-use changes. eDNA sampling strategies for UK lakes are relatively well-established but vary widely in lotic systems in terms of number of sampling locations, number of replicate samples, volume of samples, and where samples are collected at each sampling location (Bruce et al., 2021; Hänfling et al., 2016). The freshwater eDNA pilot study aimed to compare the north and south basins of Loch Lomond (five samples from each basin) and from the upstream to downstream reaches of the Endrick Water. The pilot study also aimed to compare detection rates in individual replicate samples versus a single composite sample at each of three sampling locations along Endrick Water, to assess whether the more cost-effective route of composite sampling would yield similar results to replicate sampling. Samples were processed using vertebrate (Kelly et al., 2014; Riaz et al., 2011), freshwater invertebrate (Leese et al., 2021), and bacteria (Caporaso et al., 2010) assays.

Five vertebrate species of national interest (European eel [*Anguilla Anguilla*], Atlantic salmon [*Salmo salar*], common toad [*Bufo bufo*], Euraisian otter [*Lutra lutra*], red squirrel [*Sciurus vulgaris*]) and one vertebrate species of international interest (European eel) were detected. It was estimated that 5-10 shoreline samples should be sufficient to capture the majority of fish diversity present in Loch Lomond. This is generally in accordance with previous UK lake eDNA studies which indicate that 10 samples will detect c. 90% of fish species present in the majority of UK lakes (Hänfling et al., 2016). Over 400 invertebrate Operational Taxonomic Units (OTUs) were detected in the water samples collected from Loch Lomond. Many of these belonged to taxonomic groups that are targeted as part of ongoing monitoring programs, such as chironomids, caddisflies, stoneflies, and mayflies. The north basin of Loch Lomond showed higher invertebrate richness than the south basin. Community composition did not significantly differ between the north and south basins of Loch Lomond for vertebrates or invertebrates, but there were strong differences between the bacterial communities. Vertebrate composition differed between upstream and downstream sampling locations on the Endrick Water, but invertebrate composition did not. Composite samples contained fewer OTUs than the three independent replicate samples. In terms of community similarity, composite samples often differed from replicate samples from the same location.

Key outcomes

Given the indications that 5-10 samples were sufficient for lakes, while multiple replicate samples per sampling point are possibly required for rivers, it was decided to focus exclusively on lakes for Phase 2 of the project. This would allow for more sites, better representing a range of habitat conditions to be included, rather than focussing efforts on a very limited number of rivers with fewer habitat conditions.

Do the eDNA primers initially being trialled provide useful taxonomic information in the context of long-term monitoring in each habitat?

The results from the vertebrate analysis from both lochs and rivers were deemed relevant to supporting monitoring programs in Scotland and it was decided that this analysis should be kept for the Phase 2 eDNA survey. Due to the large number of invertebrate species detected, especially those from ecologically monitored groups, it was also decided to keep the invertebrate assay for the Phase 2 eDNA survey.

Do the resultant ecological communities provide an indication that they are likely to be different across habitat conditions/classifications?

As bacteria showed the most discernible communities between the north and south basins of Loch Lomond, this assay was also deemed appropriate for continuation in Phase 2, across multiple lakes, including potential comparison with conventional morphology-based methods for UK lake assessment (e.g., WFD-UKTAG, 2014). Although the fish and freshwater invertebrate communities did not differ significantly across the Loch Lomond basins, the species identified are commonly used in freshwater assessments, and fish eDNA results have been used in other studies to confidently assess lake eutrophication (Hänfling et al., 2016). Furthermore, the high number of chironomid species identified was viewed as a potential to test the use of metabarcoding data in the Chironomid Pupal Exuvial Technique (CPET) lake scoring system.

The full freshwater pilot study report can be found in Appendix 4.2.

2.3 Woodland

Overgrazing by herbivores can have a negative impact on woodland condition and is one of the main pressures on woodland habitats in the LLTNP. The aim of the woodland eDNA pilot study was to investigate the potential for assessing herbivore impacts on biodiversity and community composition of soil communities. The woodland pilot study was conducted at RSPB Inversnaid Nature Reserve along the eastern bank of Loch Lomond. Inversnaid is impacted by deer and goat grazing activity and the pilot project made use of existing experimental herbivore exclosure plots that have been in place for ~20 years. Sample collection consisted of three samples collected from each of eight 2m x 2m plots, four from within the fenced exclosures, and four in the immediately adjacent unfenced woodland, totalling 24 samples. Additionally, a composite sample was collected from each larger 10m x 10m plot (eight samples), to assess whether the more cost-effective route of composite sampling would

yield similar results to replicate sampling. Samples were processed using the bacteria (Caporaso et al., 2010), fungi (White et al., 1990), and soil fauna assays (Capra et al., 2016).

All three assays resulted in a high number of OTUs detected. There was no clear statistical evidence that communities were discernible between grazed and ungrazed plots. Overall, OTU richness captured by three replicate samples in a plot was higher than captured by the one sample composed of three replicates. In terms of community similarity, mixed samples generally coincided with replicates in visual representations, although with a substantial level of variation between plots and assays. For some applications, where capturing the entire biodiversity richness is not required, such as landscape or habitat model building, it appears to be reasonable to use a single composite sample comprised of multiple subsamples, allowing for more replicate plots to be included in studies and survey designs.

Key outcomes

Do the eDNA primers initially being trialled provide useful taxonomic information in the context of long-term monitoring in each habitat?

Invertebrate indicator species that are generally utilised in soils include, among others, earthworms, nematodes, mites, and springtails (Blair et al., 1996; Manu et al., 2021; Stork and Eggleton, 2009). The invertebrate assay used for soil in this pilot study does detect members of these taxonomic groups but does not identify many to species level. Similarly, the taxonomic resolution for the bacteria and fungal assays is generally at family or genus level. These assays are of limited use for identifying indicator species that are the focus of conventional methods.

Do the resultant ecological communities provide an indication that they are likely to be different across habitat conditions/classifications?

Based on the pilot results, it was considered unlikely that continuing to focus on grazing pressures would result in clearly different communities and it was decided not to pursue grazing pressure specifically as the key biodiversity pressure under investigation in the Phase 2 eDNA survey. Instead, a different gradient was sought – woodland restoration. Soil is the most popular substrate for eDNA sampling for ecosystem restoration studies, and soil microbial communities the most targeted taxa (van der Heyde et al., 2022). The peatland pilot study indicated that the soil fauna assay did result in community differences across the gradient being investigated (see following section). As a result of these considerations, and in the absence of pilot study data in a woodland restoration context, it was decided that all three assays be kept for the woodland restoration study.

Additionally, it was decided to use composite samples, with subsamples collected across a 10m x 10m plot, allowing more plots to be analysed across more sites.

The full woodland pilot study report can be found in Appendix 4.3

2.4 Peatland

Most of Scotland's peatland is considered degraded. This is mainly the result of artificial drainage channels, land use conversion, and overgrazing. The aim of the peatland eDNA pilot study was to investigate the potential for assessing degraded and restored peatland at two sites where restoration works have been undertaken to raise the water table: Auchlyne and Glen Finglas Estates. In total 14 samples were collected from two peatland sites: Auchlyne Estate (six samples) and Glen Finglas (eight samples). Sampling locations at each site were divided into 'degraded' and 'restored' in consultation with a Peatland ACTION representative and site managers. Samples were processed using the bacteria (Caporaso et al., 2010), fungi (White et al., 1990), and soil fauna assays (Capra et al., 2016).

Soil community composition significantly differed between degraded and restored areas for fauna and bacteria, but not for fungi. However, sample sizes were insufficient to include site effects or environmental characteristics in the analyses.

Key outcomes

Do the eDNA primers initially being trialled provide useful taxonomic information in the context of long-term monitoring in each habitat?

As noted for the woodland pilot study, the assays used in this pilot study are of limited use for identifying indicator species that are the focus of conventional methods.

Do the resultant ecological communities provide an indication that they are likely to be different across habitat conditions/classifications?

The pilot study data indicated a difference across the between degraded and restored areas. It was decided to extend this sampling strategy to further sites and locations for the Phase 2 eDNA survey, keeping the focus on degraded and restored locations within sites.

Although fungi communities did not appear to show such differences it was decided to keep all three assays to allow comparison with the woodland study and to fully assess the utility of the fungi assay in the broader Phase 2 eDNA survey.

To align with the woodland sampling strategy, it was decided to use composite samples, with subsamples collected across a 10m x 10m plot.

The full peatland pilot study report can be found in Appendix 4.4

Table 1: Summary of the Phase 1 pilot study conducted in 2021

Habitat	Survey period	Study area	Number of sampling locations	Samples / location	Taxonomic analyses	Number of analyses	Pilot outcomes
Marine	Oct	Loch Long and Loch Goil	3	Three aquatic samples, two sediment samples, and two bulk samples	Aquatic: invertebrates, fish, bacteria Sediment: invertebrates, bacteria Bulk: invertebrates	39	• Ten PMF teleost taxa were detected, but not always in all 3 station replicates • Sediment invertebrate and bacterial communities differed between stations and hence biotopes • Too few individuals and taxa were obtained from morphological analyses to observe any obvious patterns • No PMF invertebrate taxa were detected from either DNA metabarcoding or morphological analysis, and the assay used was determined not fit for purpose (detecting PMF species) • The variation in species richness of invertebrates from sediment was different between DNA analysis and morphological analysis • The station closest to the head of Loch Goil had freshwater taxa detected in a replicate. Accordingly, marine samples should be collected in less upstream/enclosed environments
Freshwater	Oct	Endrick Water	3 (and 1 blank)	Four aquatic samples (three separate and one composite) and one bulk sample	Aquatic: vertebrates, freshwater macroinvertebrates, bacteria Bulk: freshwater macroinvertebrates	42	Winter shoreline sampling detected most fish species present in Loch Lomond. • Vertebrate OTU richness and community composition did not significantly differ between the north and south basins of Loch Lomond. • Invertebrate OTU richness was higher in the north basin, but community composition did not differ between basins. • Extrapolation indicated 5-10 shoreline samples may be sufficient for the Lake Fish Classification Index for Loch Lomond. • Replicate samples produced higher detection rates and more similar communities than composite samples from Endrick Water sampling locations. • Vertebrate OTU richness and community composition changed from upstream to downstream, whereas invertebrate OTU richness and community composition did not.
		Loch Lomond	10	One aquatic sample	Vertebrates, freshwater macroinvertebrates, bacteria	33	

Table 1: Summary of the Phase 1 pilot study conducted in 2021

Habitat	Survey period	Study area	Number of sampling locations	Samples / location	Taxonomic analyses	Number of analyses	Pilot outcomes
Woodland	Oct	RSBP Inversnaid Nature Reserve	4	8 composite soil samples	Bacteria Fungi Fauna	96	• Soil pH and moisture showed some variability between plots with pH significantly higher in plot 4 and moisture differences between all plots except 1 and 2, and 3 and 4 • Observed OTU richness was highest in bacteria • Estimated sample coverage was highest for both fenced and unfenced treatments for bacteria • OTU richness did not show a consistent trend between fenced and unfenced treatments • Significant differences were observed in OTU community composition between locations for all taxonomic groups. Some differences in composition were observed between fenced/unfenced but these were not significant considering the limited sample size • Replicate samples produced higher detection rates and more similar communities than composite samples from all 3 replicate samples except one- soil fauna in the fenced area of plot 1 where the mixed sample showed higher richness • Mixed samples showed broadly similar species composition to the replicate samples
Peatland	Aug	Glen Finglas Estate	2	4 composite soil samples	Bacteria Fungi Fauna	24	• Moisture and pH data were broadly similar across sites, with moisture higher in restored peatland compared to degraded • OTU richness was highest for bacteria and fungi • Estimated sample coverage was highest for bacteria and lowest for fungi • OTU richness was generally higher for degraded transects compared to restored, but lower for soil fauna (based on estimated richness) • Community composition different between degraded and restored transects for soil fauna and bacteria but not fungi
		Auchlyne Estate	2	3 composite soil samples	Bacteria Fungi Fauna	18	

3 Sampling plan for Phase 2

To meet the project objectives, a sampling plan was developed and implemented for Phase 2, in consultation with key stakeholders. For each habitat, the plan provides the number of locations, samples per location (replicates), types of samples (water or sediment/soil), number and type of assays, and a justification for the decisions that were taken to arrive at this plan. An overview map of the sampling locations (from all four habitats) including their position in and around LLTNP is presented in Figure 1. The sample analysis costs for samples collected during both Phase 1 and 2 fieldworks are recorded for each habitat type in Appendix 4.5.

Reference is made to the pilot study results throughout this section but for full details please refer to Sections 2 and 4.

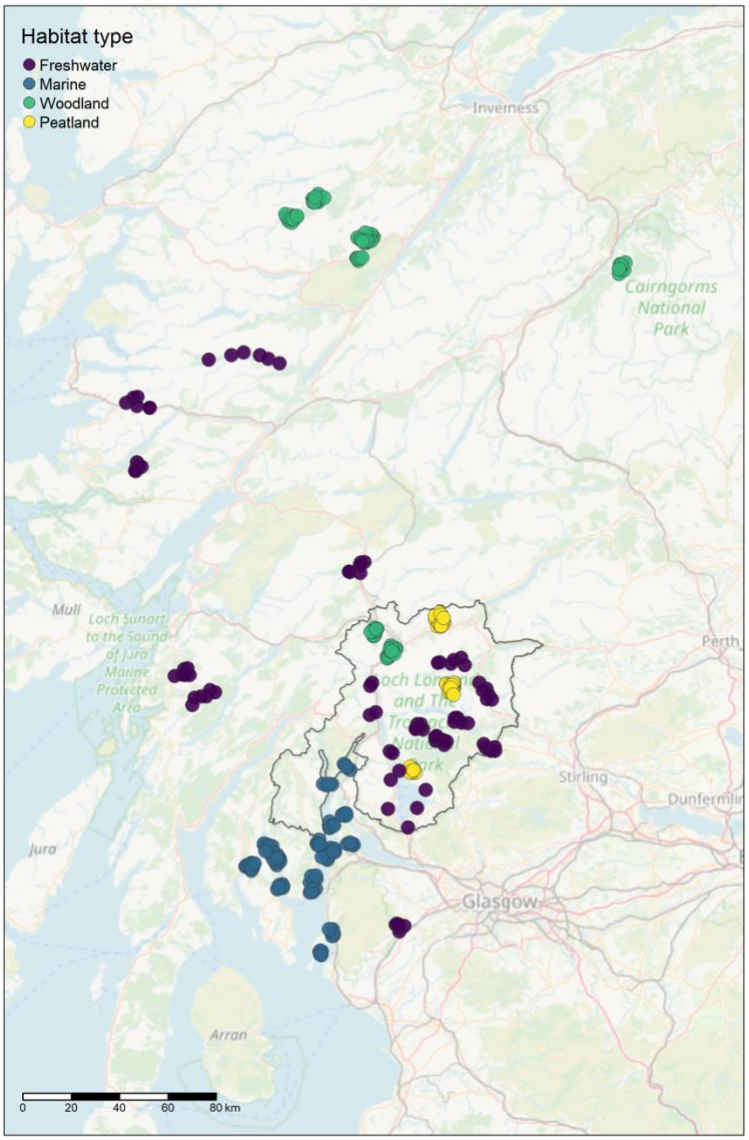

Figure 1: Overview map of the sampling locations from all four habitats. The outline within the map represents the boundary of LLTNP. Basemap: OpenStreetMap

3.1 Marine

Habitats and species that are considered conservation priorities in Scottish territorial waters are identified as Priority Marine Features (PMF). Traditionally these PMFs are monitored through methods such as benthic grab sampling, video and photography, diving surveys, acoustic monitoring, and fish trawls and plankton tows. Here we aimed to simplify species monitoring for habitat classification while simultaneously improving cost and resource effectiveness, by developing the ability to use DNA metabarcoding data to classify biotopes and detect DNA signatures of important protected features. Aided by comparing the species compositions obtained from the eDNA metabarcoding data (delivered by NatureMetrics) to habitat classification of the sampling locations obtained through PSA and morphoanalysis of simultaneously collected sediment samples (delivered by SEPA). Thus, species were classified to habitat instead of a pressure gradient such as is the case in the other three habitats. In addition, we also used sediment morphoanalysis data to generate ecosystem health scores using tools such as AZTI's Marine Biotic Index (AMBI) (Borja et al., 2000). If a gradient is identified using these scores, then the eDNA data can be compared against such a gradient.

3.1.1 Site selection

Based on the results of the pilot study, while Loch Goil is part of the Upper Loch Fyne and Loch Goil MPA (Table 4), water and sediment samples were collected by Marine Scotland Science predominantly from Loch Long to mitigate the effects of freshwater input. Four different sediment-based biotopes in Loch Long were selected for sample collection.

The following key sources of best available data were used to inform sampling design for the marine habitat:

- Moore 2013; NatureScot Commissioned Report 631: 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)
- Allen et al. 2013; SNH Commissioned Report 437: Marine biological survey to establish the distribution of Priority Marine Features within the Clyde Sea area
- Consultation with key stakeholders from Scottish Government agencies

3.1.2 Sampling location selection

To assess the ability of eDNA approaches to classify biotopes and compare ecological pressure indices using eDNA and conventional approaches, multiple locations per biotope were required, this limited biotopes within Loch Long with PMF species to two. Based on the above considerations, 20 sampling locations were chosen (Figure 2) in Loch Long including four different, known sediment-based biotopes (five stations each):

- PMF biotopes: SS.SMu.CFiMu.SpnMeg, SS.SMu.CFiMu.MegMax
- Non PMF biotopes: SS.SMx.CMx.ClloMx, SS.SMx.CMx.OphMax

- Sample locations were spread across Loch Long approximately from Ardgartan to Great Cumbrae, Loch Striven and Colintraeve (Figure 2)

3.1.3 Sampling approach

Samples were collected from four selected biotopes in Loch Long, 5 stations per biotope, and triplicate samples from each station for both water and sediment to increase the likelihood of detecting PMF species and enable statistical comparison across stations and biotopes and, where possible, to advise on the minimum number of samples required to detected PMFs. Biotopes had even numbers of samples, and sampling locations were not geographically clustered within Loch Long.

- Water samples
    - 49 water samples were collected from the boat using a vertical 7.5-litre Niskin bottle (+CTD) at a depth of 25m depth (where possible)
    - 5 litres per filter was collected, with one exception of 2.5 L due to a bottle leaking
    - Each sample was passed through 0.8 µm PES filters (and a 5 µm glass fibre prefilter) supplied in a NatureMetrics Pump Aquatic eDNA Kit
    - The volume of water passed through each filter was recorded
    - Filters were preserved with fixative supplied in a NatureMetrics Pump Aquatic eDNA Kit and kept at ambient temperature for two days and then stored at -20°C until return to the NatureMetrics laboratory where they were stored at -20°C until analysis
- Sediment grab samples for DNA analysis
    - 56 sediment samples were collected from the boat using a Day grab
    - four syringe core samples were collected from the grab, deposited and mixed in a plastic bag
    - The samples were not processed further on-site
    - Samples were preserved in cold storage for two days and then stored at -20°C until return to the NatureMetrics laboratory
- Sediment grab samples for morphoanalysis
    - Balancing budget with the need to better characterize biotopes and provide a comparison of traditional methods with eDNA methods, one grab per station was analysed using morphoanalysis, totaling 20 samples
    - For each grab the contents were sieved on board using a 1 mm sieve mesh and material was preserved with formaldehyde, and transported to SEPA
- Sediment grab samples for particle size analysis (PSA)
    - Similarly, one grab per station was selected for PSA analysis
    - For each grab, 100 ml sediment was collected from the Day grab, using a corer
    - The samples were deposited 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

3.1.4 Data collection

- Latitude and longitude coordinates
- Morphological and particle size analyses at each sampling location – to be used to generate biotope classification
- CTD: conductivity (salinity), temperature, and depth
- Where possible, previous survey data for PMF species will be sought from MSS, SEPA to allow qualitative assessment of species detected by the eDNA survey versus known species to be present at sampling locations.

3.1.5 eDNA assay selection

Assay selection focused on detecting as many PMF species as possible and to assess whether metabarcoding community data showed discernable differences between the different biotopes. Many of the marine PMF species are benthic invertebrates, thus using an invertebrate assay for sediment samples was decided. Based on the pilot results it was determined that the 18S invertebrate assay used (Capra et al., 2016) did not perform well in either PMF species detection or overall species resolution, hence an internal NatureMetrics testing (not associated with this project) was carried out testing two additional eukaryotic assays. As a result of this testing, a better performing assay based on the 18S gene (Amaral-Zettler et al., 2009), and a COI assay (Leray et al., 2013) were selected to provide better resolution.

The 16S bacteria assay used in the pilot studies was also selected as it showed high levels of heterogeneity across sampling stations and this taxonomic group has the potential to be used for classifying biotopes as well as being used for assessing pressures (Borja, 2018; Lejzerowicz et al., 2021).

Based on the pilot results of the water samples, where the fish assay detected ten PMF fish species, and two PMF marine mammal species, the same assay was selected.

3.1.6 Caveats

One of the overall caveats in the sampling design of all four habitats is that for both the pilot and the main study, the sampling was constrained to one sampling season only. It is essential that in future monitoring programs seasonality be taken into account. The Phase 2 eDNA survey marine sampling was focused on only one waterbody. Due to the different types of samples required to be collected from the marine habitat, intensive resource efforts (e.g., boat access, expertise, and staff time for field sampling and laboratory analyses), and distances between the individual sampling sites, instead of a total of 60 water filters, 49 filters were collected due to time restraints during sampling. Instead of a total of 60 sediment samples, 56 sediment samples were collected due to some grabs coming up empty.

Figure 2: 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

3.2 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, in response to pressures such as land use, water quality, pollution and climate change.

Here we aimed to test whether lochs experiencing similar hydrological and nutrient pressures (so much as is possible to control) result in different biological communities that can indicate overall condition using eDNA metabarcoding. This may help us in understanding the effects of the impacts of different pressures identified as driving the biodiversity crisis. The principal habitat condition gradient to be established for use in Phase 2 of this project was a pollution pressure gradient across ranging water quality (e.g. Water Framework Directive (WFD) statuses such as Overall Status, Overall Ecology Status, Biological Elements Status), and surrounding land use.

SEPA produces an annual Water Framework Directive (WFD) classification for all the identified waterbodies in Scotland (rivers, lochs, estuary, coastal and groundwater). Waterbodies are classified using a system of five ecological status quality classes: High, Good, Moderate, Poor, and Bad; The classification encompasses ecology, 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. According to SEPA's 2020 classification, many of these are of good or high status, with some at moderate status and fewer at poor or bad status. This information was used to help identify and select a range of freshwater lochs, mostly situated within LLTNP focal study area, for sampling during Phase 2 (see sections 3.2.1 and 3.2.2).

3.2.1 Site selection

Following the pilot study, freshwater lochs became the focus of the freshwater habitat Phase 2 eDNA survey. The total number of sites, sampling locations, and sample assays was initially decided based on balancing the project budget and resources available, with obtaining the range of sites and level of replication required to address whether eDNA metabarcoding can enable assessment of habitat condition of Scottish freshwater lochs. However, the decisions regarding exactly how many reasonably representative samples to collect and at which freshwater lochs sites to sample, beyond Loch Lomond, required extensive consideration and consultation with key project stakeholders with relevant technical expertise and practical experience of operationalising monitoring resources across Scotland.

The collection of 10 shoreline samples in winter was previously identified as the minimum sampling effort required to detect ≥85% of fish species present in UK lakes (Li et al., 2019). However, this level of sampling effort may or may not be achievable for freshwater lochs if eDNA shoreline monitoring approaches were upscaled in the future. For example, this might be due to sections of inaccessible shoreline or resource constraints to mobilise appropriately trained staff to collect eDNA samples from multiple shoreline locations (for subsequent laboratory processing) or to deploy boats if needed for in-lake open-water sampling purposes (where (part of) the shoreline is inaccessible by land), particularly for routine monitoring programmes (e.g., WFD networks) that need to be designed, implemented, and operationalised at a national scale. The pilot results in the present project also indicated that collection of 5-10 shoreline samples would be sufficient to capture most fish species present and to apply the Lake Fish Classification Index. While winter sampling maximises species

detection through better mixing of the water column, and subsequently eDNA, in the absence of thermal stratification (Hervé et al., 2022; Lawson Handley et al., 2019; Littlefair et al., 2021), this was impractical due to the time available for the Phase 2 eDNA survey, which needed to take place during summer 2022 and within the project delivery lifespan. One of the caveats in the sampling design is that for both the pilot and the main study, the sampling was constrained to one sampling season, which is likely to affect the communities detected when sampling from the surface only, without including the deeper stratified layers. The effect of the lack of sample collection during the winter season on the species composition, is further explored in the forthcoming phase 2 report. It is essential that in future DNA-based monitoring programmes seasonality be considered.

It was decided that six samples per freshwater loch were to be collected from the shoreline. This approach was standardised across all freshwater lochs sampled for the Phase 2 eDNA survey. This fixed sample number was considered the reasonable balance between the minimum required for DNA-based sampling, loch accessibility reasons (not all parts of the sampled lochs were accessible by land), and available contractor resources (budgetary constraints) to deliver the project work in 2022. By taking that key decision, it was possible to increase the total number of freshwater lochs that could be sampled for metabarcoding analysis from the shoreline, and in doing so expand breadth of the overall habitat pressure gradient assessed. The need for eDNA shoreline sampling was also based on future-proofing considerations such as potential constraints on SG resources (i.e. to support in-lake open-water sampling by boat) in future years for investigative and long-term biodiversity monitoring purposes, if DNA-based monitoring were rolled out for freshwater lochs at national scale. Following extensive consideration and consultation with key project stakeholders, including experts from SEPA and NatureScot, a total of 15 freshwater lochs (Figure 3) were selected based on their location within, or their proximity to LLTNP, accessibility by road, and where shoreline sampling would be sufficient (to minimize resource constraints, also keeping in mind potential future monitoring programmes). All 15 lochs had previously been classified by SEPA using the WFD ‘overall status’ designations 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. Within our selected criteria of ‘large’ and ‘low alkalinity’ (to control for any other potential confounding factors), three lochs with a ‘high’ status were available, five ‘good’, six ‘moderate’, and one ‘poor’ (Table 2). More than half (n = 8) of the freshwater lochs, including Loch Lomond north and south basins, selected for Phase 2 eDNA sampling were situated within the LLTNP, three of which also lie in the Teith Special Areas of Conservation (SAC) (Table 4). The other freshwater lochs selected were situated in the Renfrewshire (n = 1), Argyle and Bute (n = 3), and Highland (n = 3) council areas. Castle Semple is situated in the RSPB Lochwinnoch nature reserve. Table 4 gives an overview of the designations of conservation importance and protected statuses of each sampling location for all habitats. Those sites which did not fall within Scotland’s designated site network functioned to provide replicates for the habitat condition gradient established for freshwater lochs and will be used as a key learning

opportunity from Phase 2. Some of the freshwater lochs sampled are grouped (Table 2), as they are not routinely monitored by SEPA for WFD purposes and will also be used as a key learning opportunity from Phase 2.

It was important that the lochs were reasonably reflective of WFD overall status as high, good, moderate, and poor or bad ecological status, whilst also ensuring the overall hydrology status remained high so that impacts from major known confounding factors (such as hydrology pressure from impoundment or abstraction due to hydropower or water supplies) were reasonably minimized wherever feasible, especially if any Scottish lochs are designated as Heavily Modified Waterbodies (HMWBs) and Grouped Waterbodies. We selected lochs that were:

- Lowland situated (Altitude type <200 m according to WFD-UTKAG, 2004)
- Have a large surface area (size type ≥ 0.5 km^2 in surface area according to WFD-UTKAG, 2004)
- Situated within a reasonably similar geographic area and climatic envelope, with most lake sampling constrained to the LLTNP focal study area, with some acceptable distances up to a 100 km radius beyond LLTNP boundaries
- Reasonably representative of standing waterbodies located within the focal study area of LLTNP:
    - Mostly low alkalinity, with some acceptable and occasional deviation into moderate alkalinity (according to WFD-UKTAG, 2014)
    - A balanced mixture of deep and shallow waterbody depth types, with 'very shallow' being the occasional exception (according to WFD-UKTAG, 2014)
    - Mostly clear water colour types, with some acceptable and occasional deviation into humic, polyhumic, or unknown types (according to WFD-UKTAG, 2014)
- Reasonably representative of a range of land use categories including moorland, arable, woodland, and urban land cover in the surrounding catchments
- There is recent evidence that some freshwater lochs are impacted by climate change (May et al., 2022). It was found that Loch Achray and Loch Lubnaig situated in LLTNP to be amongst the most rapid warming standing waters in Scotland, with water temperatures having increased by between 1.0 and 1.3°C per year during 2015-2019)

The following key sources of best available data were used to inform sampling design for the freshwater habitat:

- Water Classification Hub (sepa.org.uk)
- UK Lakes Portal (ceh.ac.uk)
- wfd uktag | water framework directive e.g.,
    - WFD-UKTAG (2004) Guidance on Typology for Lakes for the UK | wfd uktag
    - WFD-UKTAG (2014) UKTAG Lake Assessment Methods (wfduk.org)
- Assessing climate change impacts on the water quality of Scottish standing waters | CREW | Scotland's Centre of Expertise for Waters
- https://www.space-intelligence.com/scotland-landcover/
- https://www.nature.scot/professional-advice/protected-areas-and-species/protected-areas

- Lochwinnoch Nature Reserve, Renfrewshire, Scotland - The RSPB
- Consultation with key stakeholders from Scottish Government organisations

A total of 15 freshwater lochs were selected based on the above criteria. These are shown in Table 2, with their key metadata categories and relevant supporting information.

Table 2: Sampled lochs with key metadata categories. Color codes represent the overall WFD status of the lochs based on SEPA 2020 classification results; blue = high, green = good, yellow = moderate, amber = poor (Water Classification Hub (sepa.org.uk)). Moorland, arable, woodland, and urban categories refer to the percentage of land use in the lake catchment (Space Intelligence, https://www.space-intelligence.com/scotland-landcover/). Heavily Modified Waterbodies – where achievement of the WFD target of good status would adversely impact a use of a waterbody (typically hydropower, water supply, or infrastructure) it can be designated an HMWB and has the lesser target of good ecological potential. Loch Lomond is an HMWB due to the morphological impact of the road. Grouped Waterbodies – because SEPA does not have resources to monitor all waterbodies, some of them are 'grouped' with a monitored waterbody on the basis of similarity of typology, geography and pressures. These grouped waterbodies are then assumed to be at the same status as the monitored waterbody, but SEPA have no actual data for them. The land use percentages for Loch Lomond N & S basins are representative of the whole loch.

Name	Size type	Depth type	Geology type	Humic type	Moorland	Arable	Woodland	Urban
Loch Arkaig	L	Deep	Low alkalinity	Clear	0.77	0.09	0.10	0.02
Loch Tulla	L	Shallow	Low alkalinity	Clear	0.83	0.09	0.06	0.00
Loch Scammadale	L	Deep	Low alkalinity	Unavailable	0.64	0.23	0.11	0.02
Loch Doilet	L	Shallow	Low alkalinity	Clear	0.52	0.03	0.43	0.01
Loch Voil	L	Shallow	Low alkalinity	Clear	0.73	0.07	0.19	0.00
Loch Eilt	L	Shallow	Low alkalinity	Clear	0.88	0.03	0.06	0.00
Loch Avich	L	Deep	Moderate alkalinity	Humic	0.27	0.08	0.61	0.05
Loch Lomond (N)	L	Deep	Low alkalinity	Clear	0.53	0.27	0.18	0.01
Loch Lomond (S)	L	Deep	Moderate alkalinity	Clear	0.53	0.27	0.18	0.01
Loch Achray	L	Shallow	Low alkalinity	Clear	0.74	0.07	0.18	0.01
Loch Lubnaig	L	Shallow	Low alkalinity	Clear	0.61	0.09	0.29	0.01
Loch Ard	L	Shallow	Low alkalinity	Clear	0.39	0.04	0.52	0.05
Lake of Menteith	L	Shallow	Moderate alkalinity	Clear	0.29	0.29	0.39	0.02
Loch Chon	L	Shallow	Low alkalinity	Humic	0.52	0.02	0.41	0.04
Castle Semple Loch	L	Very shallow	Moderate alkalinity	Humic	0.42	0.43	0.12	0.04

Figure 3: 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

3.2.2 Sampling location selection

Six sampling locations (i.e. six samples) were selected for each freshwater loch (noting that Loch Lomond comprises two waterbodies for WFD classification purposes i.e., its constituent north and south basins, with six samples collected from each basin). This number is also within the range identified by the pilot study as being suitable for capturing most fish species present within Scottish lochs.

Sampling locations were set at equidistant intervals where possible around the perimeter of the lake, as is best practice for lake fish eDNA sampling (Hänfling et al., 2016; Lawson Handley et al., 2019; Li et al., 2019; Zhang et al., 2020). Sampling locations were restricted to places where shoreline access was within reasonable walking distance from a parked vehicle, there were no health and safety risks, and sampling could be repeated in the field (without a need for mobilising boat work and related staffing resources) in the future.

3.2.3 Sampling approach

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 et al., 2019). Each sampling bag was sealed and shaken for 20-30 seconds to homogenise the water sample, then each sample was passed through one of two 0.8 µm PES filters with a Vampire Sampler. The volume of water passed through each filter was recorded.

A field negative control (500 mL of mineral water) was processed after the last lake was sampled on each day of sampling to monitor for potential contamination introduced in the field. Mineral water was poured into the 3.5 L sampling bag, the bag was sealed and shaken for 20-30 seconds, then the water was passed through the 0.8 µm PES filter (and a 5 µm glass fibre prefilter) using a Vampire Sampler (see pilot study for more details of the sampling equipment).

A total of 90 samples and five field negative controls were collected. These were transported to the NatureMetrics laboratory (Guildford, UK) at ambient temperature and stored at -20°C until analysis.

3.2.4 Data collection

- Latitude and longitude coordinates, water temperature (°C), pH, Total Dissolved Solids (ppt), conductivity (mS/cm), orthophosphate levels (ppm), and nitrite levels (ppm) were recorded for each lake. All instruments were properly calibrated prior to use. Google Maps was used to obtain coordinates, and water chemistry was measured using a Hanna Instruments HI-98130 Pocket EC/TDS and pH Tester (High Range), Hanna Instruments HI-713 Phosphate Low Range Handheld Colorimeter - Checker HC, and Hanna Instruments HI-707 Nitrite Low Range Handheld Colorimeter - Checker HC. The EC/TDS and pH Tester was calibrated on each day of use using Hanna Instruments pH 4.01 (HI 70004) and pH 7.01 (HI 70007) Buffer Solution Sachets, and Hanna Instruments 12880 µS/cm Sachets (HI 70030).
- Land use cover was collected using Welcome- Land Cover Maps (space-intelligence.com).
- WFD status information will be gathered from https://www.sepa.org.uk/data-visualisation/water-classification-hub/ and, where needed, from SEPA sources.

- We will be seeking additional fish and PLUTO (phytoplankton>cyanobacteria) data for comparative purposes and also Chironomid Pupal Exuvial Technique (CPET) data collected from some of these lochs as well as CPET models from SEPA.
- We may also seek the available suite of Water Chemistry parameters including phosphorous, alkalinity, etc, and surface water temperature from SEPA.
- Where possible, previous survey data for fish species will be sought from SEPA to allow qualitative assessment of species detected by the Phase 2 eDNA survey versus known species to be present at sampling locations.

3.2.5 eDNA assays

Based on the key outcomes from the pilot study, the samples were analysed using the vertebrates, freshwater invertebrates, and bacteria assays.

3.2.6 Caveats

Not all lochs selected have available matching water chemistry data or a similar suite of contemporary ecological monitoring data. Also, one or more lochs may be 'grouped' with routinely monitored lochs of similar typology, geography, and hydrology and morphology pressures. Algae and macrophytes were not monitored during this survey. The ecological gap or potential mismatch relating to a water quality pressure gradient assessed by diatoms, phytoplankton, and macrophytes is acknowledged.

Loch selection mostly involved a strategic desk-based assessment with stakeholder engagement within the brief timescale available. The environmental pressure status, i.e. ecological health condition may have changed at any of the lochs selected prior to, during, or following the project's Phase 2 eDNA survey in August 2022. Furthermore, the WFD approach was used to establish the principal environmental pressure/habitat condition gradient for this specific habitat type. However, there may be alternative ways to further explore the Phase 2 eDNA survey data (e.g. by using NatureScot's habitat condition assessment categories).

3.3 Woodland

Within the scope of the project objectives, we aimed to assess whether eDNA communities and derived metrics can indicate overall woodland condition across a restoration gradient, from unforested, recently planted/reforested, and mature Scots pine woodland habitats. We used eDNA metabarcoding data to track woodland restoration of monoculture stands of Scots pine at different stages of regeneration. This means that chronosequences (the different stages of regeneration) of restored woodland were used as a proxy for monitoring over time.

3.3.1 Site selection

Initially it was posed that the focus should be on deciduous forest. While deciduous forest is important to forest regeneration in Scotland, it was agreed that using this type of forest would be problematic due to the differences in species composition between the different stands.

Because species composition has a strong influence on soil community, this would function as a confounding factor that may drown out the differences in soil communities resulting from restoration status. Additionally, in most cases, the required level of detail of species mixes for the different woods within LLTNP is not available. Accordingly, it was decided to focus on restoration gradient in Caledonian pine forest, Scots pine (*Pinus sylvestris*). This increased our chances of obtaining clear, unequivocal results, which is a common aspiration across all key stakeholders. Moreover, using Scots pine has the benefit of tying into pre-existing and parallel work by Forest Research. Although these Forest Research experimental Scots pine sites are not within LLTNP, the setup of the sites warranted sufficient merit to include in this study. Because not all sites had all three categories, one of the Cairngorms Forest Research sites, Rothiemurchus, situated within the Cairngorms National Park, was chosen to be included in the main sampling campaign of this project. Moreover, Rothiemurchus had their own adjacent young and natural regeneration mature Scots pine, which made for a better comparison.

By using monoculture stands of Scots pine at different stages of regeneration, space was substituted for time by using chronosequences of restoration. Three chronosequence categories were chosen instead of four to obtain better replication per treatment; unforested, recently planted/reforested, and mature condition. All sites were required to contain all the chosen age categories. Within each site, the different categories were required to be the same forest type, i.e. Scots pine. To further exclude confounding factors, the different categories were also required to be in similar environments, e.g. we did not want to compare areas on a steep slope or high plateau with lochside areas. Ideally, the sites needed to have each of the categories in adjacent stands, or at least in close proximity to each other.

3.3.2 Sampling location selection

Categories were a chronosequence of forest age. Three categories were selected; unforested (which may range from grassland to moorland), recently planted/reforested, and mature condition. Unforested areas are representative of an area that would be forest if it wasn't grazed (such as grassland or moorland). Mature condition forest is the target, while recently planted/reforested is "regenerating" forest on its way to target status. The sampling locations (Figure 4) and their respective categories, that were chosen based on the above and on extensive consultation with all key stakeholders, are presented in Table 3. Two of the woodland sites selected for Phase 2 eDNA sampling were situated within the LLTNP and are both SSSI (Coille Coire Chuilc and Glen Falloch), while Rothiemurchus sits within the Cairngorms National Park. Coille Ruigh and Ghubhais are both SSSI and SAC areas (Tables 3 & 4). Those sites which did not fall within Scotland's designated site network functioned to provide replicates for the chronosequences established for the woodland sites and will be used as a key learning opportunity from Phase 2.

Table 3: Overview of the sampling sites, areas within the sampling sites and samples collected per chronosequence category

Site	Area	Unforested	Recently planted/reforested	Mature
Coille Ruigh	Glen Affric	5	5	5
Ghubhais	Glen Affric	3	5	5
Dundreggan WGS	Glen Moriston	6	5	
Dundreggan Allt Fearna	Glen Moriston	6	5	
Inverwick	Glen Moriston			5
Coille Coire Chuilc	LLTNP	5		6
Glen Falloch	LLTNP	5	5	6
Rothiemurchus	Cairngorms NP		5	5
	TOTAL	30	30	32

Figure 4: Map of the woodland sampling locations. The outline within the map represents the boundary of LLTNP. Basemap: OpenStreetMap

3.3.3 Sampling approach

- Samples were collected across 8 sites, but not every site has all three categories, so sites are further grouped into "area" (Glen Affric, Glen Moriston, LLTNP, and Cairngorms NP, see Tables 3 and 4).
- Based on the pilot study key outcomes, each sample consisted of nine subsample cores collected across a ~10m x 10m plot and mixed to give one composite soil sample. Plot size was selected to better enable potential linking of the results with satellite image data of the same resolution.
- Prior to coring, the surface litter layer was brushed aside to expose the fermentation layer
- 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 subsample cores were thoroughly mixed in a labelled plastic grip seal bag in the field to homogenise
- To prevent cross-contamination, after each plot the soil auger was wiped with Chemgene disinfectant wipes and the sampler's gloves were changed
- 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.

3.3.4 Data collection

- Latitude and longitude coordinates were recorded for each sampling site.
- Habitat condition (unforested, recently planted/reforested, and mature condition) was recorded for each sampling location.
- Soil moisture (% moisture on a wet weight basis), and soil pH were recorded for each sample. In the laboratory, the dry weight basis was also recorded and used for the analyses.
- Additional data for the Forest Research sites will be sought. It is not yet clear exactly which data will be pertinent or available. This will be explored as part of Phase 2.

3.3.5 eDNA assays

Based on the key outcomes from the pilot study, the soil samples were be analysed using the bacteria, fungi, and soil invertebrate assays.

3.3.6 Caveats

The study and subsequent interpretation of results were restricted to Scots pine forest habitat type. It was not possible to locate sites that contained all three habitat categories and as a result, the categories were unevenly split across sites. At some sites, patch sizes were relatively small, and trees were sparsely distributed, meaning that some edge effects may have been present in the study. This study cannot be considered a direct comparison of grazing vs no grazing as there was no herbivore exclusion fencing in place at all forested sites.

3.4 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. Within the scope of the project objectives we aimed to test whether peatland sites of differing condition categories (degraded or restored) have different biological communities that can indicate overall condition using eDNA metabarcoding.

3.4.1 Site selection

Site selection criteria required sites with varying peat condition - degraded and restored. Originally a third category (unimpacted) was proposed. However, because 70% of Scotland's blanket bog and 90% of Scotland's raised bog peatland is degraded (Artz et al., 2014), as such, the Peatland ACTION officer was unable to suggest any good/unimpacted condition peatland within LLTNP. Furthermore, despite searching while on site, no patches of good/unimpacted condition peatland were identified at any of the sites. Accordingly, it was not possible to find unimpacted areas to include in this study. Site selection was then based on two categories. All three peatland sampling locations are situated within the LLTNP (table 4).

1) Degraded peat: areas of peat in the vicinity of grips (drainage channels) where the water table is low because it is drained away from the site
2) Restored peat: Areas of peat that were formerly degraded but where restoration works have been undertaken to block the grips and raise the water table level back to more natural levels

Based on the criteria three sites were selected. Glen Finglas, Auchlyne, and Cashel (Figure 5). 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. However, restoration work is expected to start in 2023. When selecting damaged/drained areas this should be based on locations that are likely to go forward for restoration as this will allow future restoration time series assessments to be made.

3.4.2 Sampling location selection

Sampling locations were based on the following criteria:

- Approximate density of sampling points at 2 per km^2
- Within areas of known peat (e.g. using PEATMAP; (Xu et al., 2018); or Carbon and peatland 2016 map when available) and with varying condition between restored and degraded peat
- Within approximately 2 km of a road to allow accessibility
- Sample locations within Glen Finglas and Auchlyne were selected because these are upland blanket bog sites within LLTNP where restoration works have been undertaken as part of the Peatland ACTION project.
- Sampling locations were determined on site in consultation with a Peatland ACTION representative and site managers

Figure 5: Map of the peatland sampling locations. The outline within the map represents the boundary of LLTNP. Basemap: OpenStreetMap

Sampling approach

- Samples were collected from a 10 x 10 m plot of homogenous habitat at each location. Plot size was selected to better enable potential linking of the results with satellite image data of the same resolution.
- Cores were collected from a sampling depth of 0-10 cm. While peatland soil DNA studies in the scientific literature often go deeper than 10 cm, they usually separate soil from the cores into different depth categories. Sampling multiple depth categories would reduce the number of locations that we could sample from. The surface soil is where biological activity will be the most active due to higher levels of oxygen availability. It is also where fluctuations in water content will be at their most extreme and so will give a good indication of effectiveness of hydrological restoration works
- Samples were collected from plots situated adjacent to blocked/unblocked drains. Nine subsample cores were collected per plot to align with woodland sampling methodology
- Surface vegetation was pushed aside prior to coring
- A metal soil augur (inner core diameter 14 mm) was used to collect subsample peat cores to a depth of ~10 cm. Any living moss at the surface of the core was discarded
- The subsample cores were thoroughly mixed in a labelled plastic grip seal bag in the field to homogenise
- To prevent cross-contamination, after each plot the soil auger was wiped with Chemgene disinfectant wipes and the sampler's gloves were changed
- 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
- A total of 50 samples were collected from three sites (all within LLTNP)
    - 10 x drained and 10 x restored from Glen Finglas
    - 10 x drained and 10 x restored from Auchlyne
    - 10 x drained from Cashel (no restoration works have happened at this site)

3.4.3 Data collection

- Peatland Action Condition categories were recorded at some plots by an ecologist from SEPA who accompanied the sampling at Glen Finglas and Auchlyne. SEPA peatland categories mirror NM peatland categories
- Latitude and longitude coordinates were recorded for each sampling site
- Habitat condition, drained vs restored, was recorded for each sampling location
- Soil moisture (% moisture on a wet weight basis), and soil pH were recorded for each sample. In the laboratory, the dry weight basis was also recorded and used for the analyses.

3.4.4 eDNA assays

Based on the key outcomes from the pilot study, the soil samples were analysed using the bacteria, fungi, and invertebrates assays. Under further consideration was the application of a Fungal:Bacterial qPCR assay; the ratio of fungal to bacterial abundance. In general, higher fungal to bacteria ratios are associated with more fertile and sustainable soil.

3.4.5 Caveats

No unimpacted areas were included in this study. 90% of Scottish peatland is degraded, and as such, the Peatland ACTION officer was unable to suggest any good/unimpacted condition peatland within LLTNP. Furthermore, despite searching while on site, no patches of good/unimpacted condition peatland were identified at any of the sites. At some locations, the moisture/texture of the peat made it impossible to extract cores. We therefore continued along the channels until we encountered an area where cores could be extracted, and plots were situated accordingly.

Table 4: Overview of designations of conservation importance and protected status for each of the sampled locations of all four habitats. Green: type and name of designated area, Red: site has no special designation

Site Name	Sites of Special Scientific Interest (SSSI)	Special Areas of Conservation (SAC)	Special Protection Areas	National Parks	Marine Protected Areas (MPA)
Freshwater Sites					
Loch Arkaig	Loch Arkaig Pinewood (part of the catchment)				
Loch Tulla	Doire Darach (part of the catchment)				
Loch Scammadale					
Loch Doilet					
Loch Voil	Stronvar Marshes (small part of the loch)	River Teith		LLTNP	
Loch Eilt			Moidart and Ardgour, area around the loch		
Loch Avich					
Loch Lomond (N)	LLTNP, parts of the catchment	LLTNP, parts of the catchment	LLTNP, parts of the catchment	LLTNP	
Loch Lomond (S)	LLTNP, parts of the catchment	LLTNP, parts of the catchment	LLTNP, parts of the catchment	LLTNP	
Loch Achray		River Teith		LLTNP	
Loch Lubnaig	Loch Lubnaig Marshes, small part of the loch	River Teith		LLTNP	
Loch Ard				LLTNP	
Lake of Menteith	Lake of Menteith			LLTNP	
Loch Chon				LLTNP	
Castle Semple Loch	Castle Semple and Barr Lochs				
Marine Sites					
Loch Goil					Upper Loch Fyne and Loch Goil
Loch Long					
Woodland Sites					
Coille Ruigh	Glen Affric	Strathglass Complex			
Ghubhais	Glen Affric	Strathglass Complex			
Dundreggan WGS					
Dundreggan Allt Fearna					
Inverwick					
Coille Coire Chuilc	Coille Coire Chuilc			LLTNP	
Glen Falloch	Glen Falloch Pinewood			LLTNP	
Rothiemurchus	Cairngorms NP	Cairngorms NP	Cairngorms NP	Cairngorms NP	
Peatland Sites					
Glen Finglas				LLTNP	
Auchlyne				LLTNP	
Cashel				LLTNP	

4 Appendices

4.1 Pilot results marine

Introduction

The pilot sampling campaign took place in October 2021 and the main sampling campaign took place in August – September 2022. Within each of the habitats, the pilot study has investigated some key parameters of interest, such as the effects of subsampling or the starting sample type (e.g., water or sediment). The data generated in the pilot were interrogated to assess their usefulness for detecting key features or informing target indicators. The following report details the methods and results from the pilot phase of the project for the marine sampling. These results have fed into the decision making of both the sampling design and the analysis approach of the P2 Sampling Plan.

Methods

Study sites

Loch Goil was selected to conduct a pilot study to identify optimal sampling strategy, appropriate laboratory analysis, and project objectives to consider pursuing for the marine habitat in Phase 2. Loch Goil and Loch Long have been previously surveyed for fish and macroinvertebrates using traditional monitoring approaches such as underwater camera footage and morphological analysis of benthic fauna from grab samples (Allen, C., Axelsson, M., Dewey, S. & Clark, 2013; Moore, 2013). Loch Goil has distinct ecological zones as well as contrasting bathymetry and sediment composition. In 2014 Loch Goil was declared a Nature Conservation Marine Protected Area (NCMPA).

Samples were collected on the 7th of October 2021 from 3 stations with previously recorded contrasting biotopes, all of which are Priority Marine Features (PMFs; Table 5, Figure 6).

Table 5: Target sampling location coordinates and previously reported biotopes

Station	Target Lat	Target Long	Allen 2013 biotope	Moore 2013 biotope
Station 1	56.099518	-4.876821	SS.SMu.CfiMu.Meg Max	SS.SMu.CFiMu.SpnMeg
Station 2	56.123199	-4.898793	-	SS.SMu.CFIMu.SpnMeg
Station 3	56.14785	-4.896203	-	SS.SMu.IFiMu.Beg

Station 2 coordinates were closer to LFG18 rather than LFG17 from Moore et al. SS.SMu.IFiMu.Beg (56.13, 4.89), which was characterised by SS.SMu.IFiMu.Beg rather than SS.Smu.CFIMu.SpnMeg.

To capture the benthic community (benthic and nektonic PMF taxa) water sampling was carried out as close to the substrate as possible, without disturbing the substrate, at all three stations. Sampling depth was based on depth measurements recorded during previous

sampling campaigns (Moore, 2013). Three water samples of 4 L were collected from each station (Fig. 1). A Vampire Pump (Burlke, Germany) was used to filter 4 L of water through an enclosed casing containing a 0.8 µm PES filter and a 5 µm glass fibre prefilter (NatureMetrics, UK) and preserved with Longmire's preservation buffer (Longmire et al., 1997). One L of water was collected from Stations 2 and 3 and passed through a 0.2 µm PES filter and a 5 µm glass fibre prefilter (NatureMetrics, UK) and preserved in RNALater (NatureMetrics, UK). At Station 1, mineral water was filtered through a 0.2 µm PES filter as a field blank control.

Sediment was collected from the same three stations, using a mini Van Veen grab (0.1 m^2). Three grabs were planned per location, but poor sampling conditions reduced the number of grabs to two per station. Four subsamples were collected from each grab using a mini-corer (NatureMetrics, UK) and pooled following NatureMetrics (NatureMetrics, 2022). Pooled samples were placed in a Ziplock bag, sealed, and kept on ice until transported to the NatureMetrics laboratory and stored at -20°C until analysis. The remainder of the grab content was brought to shore, sieved through a 1 mm mesh and stored in 1L bottles fixed with 100% ethanol. This was transported to the NatureMetrics laboratory, and subsequently sent to be analysed by a subcontractor for morphological identification.

The pilot study relied on previous classifications of biotopes for sampling location selection and there were two complications with this: 1) Station 1 was classified differently according to two separate sources (Allen, C., Axelsson, M., Dewey, S. & Clark, 2013; Moore, 2013) and; 2) the actual sampling location for Station 2 was closer to a station originally classified as a biotope other than that intended.

Due to the complications arising from relying on previous classifications of biotopes, and the limited number of sites in the pilot study, it was not possible to answer definitively whether the ecological communities were different enough between biotopes and similar enough within biotopes to enable long-term monitoring. It was also not possible to determine the minimum number of replicates needed per station. Furthermore, the variability in detection of rare species was high within stations, and detection of such rare species is important for Marine Protected Area management. As a result, it was decided that 3 replicate samples would be used for the Phase 2 eDNA survey for both sediment and water samples to answer this question fully. Additionally, analysis of morphological invertebrate samples and Particle Size Analysis would be conducted on sediment samples to allow for more precise classification of biotopes.

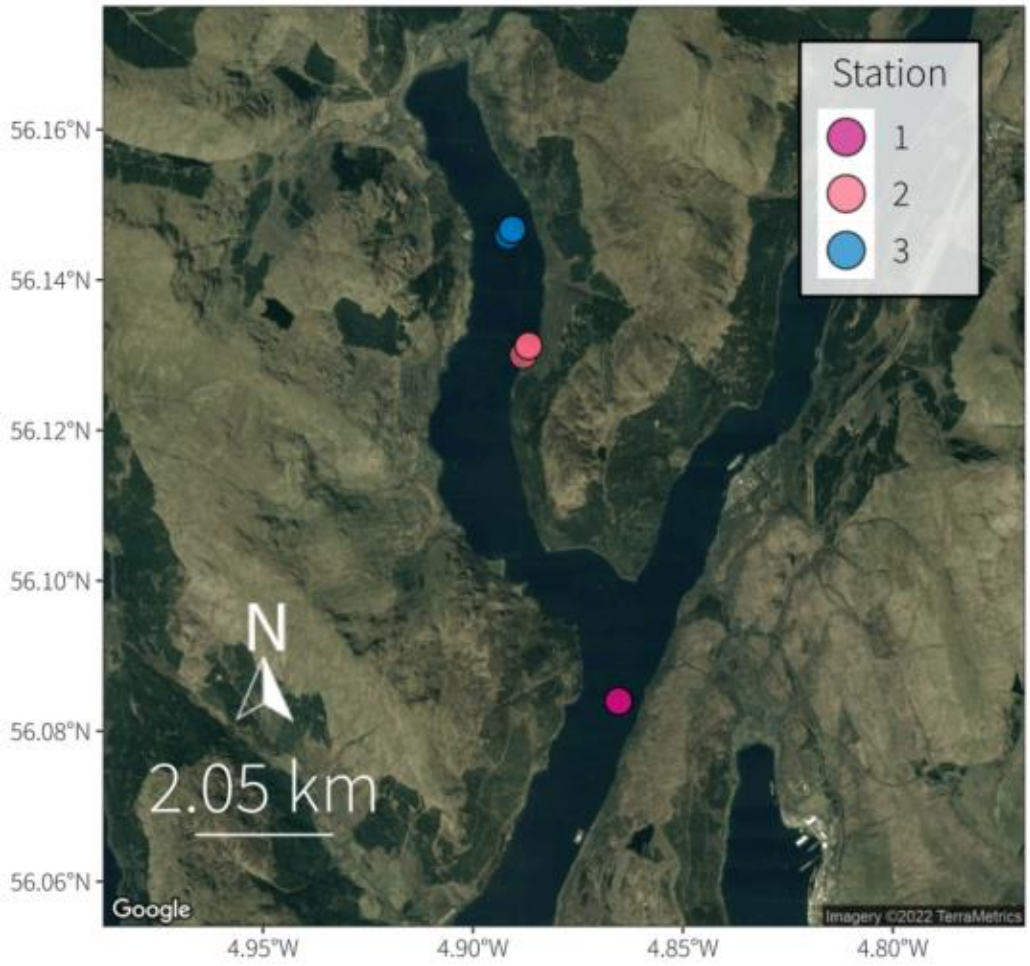

Figure 6: Map of marine sampling locations.

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. From water samples, DNA was extracted from each filter using a DNeasy Blood and Tissue Kit (Qiagen) following (Spens et al., 2017) method for disc filters in buffer with the following modifications: proteinase K was initially added directly to the filter housing; following incubation, 1 mL of lysate was carried forward for extraction and all DNeasy Blood and Tissue Kit (Qiagen) reagents were adjusted accordingly, with a final elution in 200 ul. The reasons for these modifications are to minimise potential contamination risks and 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). From sediment samples, DNA was extracted from approximately 10 grams of each sediment sample using a DNeasy PowerSoil Kit (Qiagen), following brief manual mixing in the bag. An extraction blank was also processed for each extraction batch. The DNA concentration was quantified using a Qubit DNA broad range kit according to the manufacturer's protocol.

PCR amplification

Regions of the 16S ribosomal RNA (bacteria, from the 0.2 µm PES filter samples) were amplified with modified 515F (Parada et al., 2015) and 806R primers (Apprill et al., 2015). The 18S ribosomal RNA (invertebrates) (Capra et al. 2016), and the 12S ribosomal RNA (teleost fish) (Miya et al. 2015) genes were amplified via a two-step PCR process. Tails were added at the 5′ end to be complementary with Illumina Nextera index primers. In the first step, 12 and 3 PCR replicates were performed on water and sediment samples, respectively. All PCRs were carried out in a total volume of 25 µL. For the first round PCR the fish amplification mixture contained 1X Phusion Green PCR Master Mix (Thermo Scientific), 0.4 µM of each primer, 1.5 mM of MgCl2 (Thermo Scientific), 0.6 mg/ml of BSA (Thermo Scientific), 3% DMSO, 0.9 µl of template DNA, and PCR grade water (Thermo Scientific). The invertebrate amplification mixture consisted of 1X Phusion Green PCR Master Mix (Thermo Scientific), 0.4 µM of each primer, 0.4 mg/ml of BSA (Thermo Scientific), 1 µl of template DNA, and PCR grade water (Thermo Scientific). The bacteria amplification mixture consisted of 1X DreamTaq PCR Master Mix (Thermo Scientific), 0.2 µM of each primer, 0.25 mg/ml of BSA (Thermo Scientific), 1 µl of template DNA, and PCR grade water (Thermo Scientific). Fish PCR conditions followed (Miya et al., 2015) but using a modified annealing temperature of 10 cycles touchdown PCR (- 0.5°C per cycle) starting at 69°C followed by 25 cycles of 72°C for 15 seconds. Invertebrate PCR conditions consisted of an initial denaturation at 98°C for 30 s; 35 cycles of 20 s at 98°C, 30 s at 65°C, and 30 s at 72°C; and a final elongation step at 72°C for 7 min. Bacteria PCR conditions followed (Caporaso et al., 2011) but using 30 cycles rather than 35. 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.

Another metazoan primer, LoboF1/LoboR1(Lobo et al., 2013), was also trialled using similar conditions as for the invertebrate PCRs, but amplification success was poor and was discontinued from further testing.

All first round PCR replicates per sample per marker 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 using the Nextera XT indexes (Illumina), but using 1X DreamTaq PCR Master Mix (Thermo Scientific). The second round 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 to 4 nM and pooled at equal volumes. The pooled 4 nM 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 to obtain Amplicon Sequence Variants (ASV).

For the fish and bacterial assays, 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 140-200 bp and 240-260 bp were applied for the fish and bacterial assays respectively. These sequences were quality filtered with USEARCH to retain only those with an expected error rate per base of 0.05 or below and dereplicated by sample, retaining singletons. Unique sequences from all samples were denoised in a single analysis with UNOISE, requiring retained sequences to have a minimum abundance of 8 in at least one sample (Edgar, 2016).

For the faunal assay the primer was removed from the forward reads using cutadapt (Martin, 2011) and further trimmed to 275 bp. Any reads with low quality bases (PHRED33 score <25) remaining at the 3' ends were discarded, as were all reverse reads. Because the reads for this marker do not overlap, only forward reads were used from this point. Sequences were quality filtered and denoised as above).

Taxonomic assignments were made via sequence similarity (blastn; (Altschul et al., 1990; Camacho et al., 2009) searches of the ASV sequences against two reference databases appropriate for the assay. The NCBI nucleotide database was queried for all three assays (NCBI *nt*; downloaded on 28-09-2021), with the bacterial and faunal datasets additionally queried against SILVA (Quast et al., 2012). Hits were required to have a minimum e-score of 1e-20 and cover at least 90% of the query sequence. The taxonomy associated with each hit was converted to the GBIF taxonomic backbone to allow blastn results from different databases to be comparable for the following step. Public databases are known to contain errors with DNA sequences frequently associated with incorrect species names. This poses a problem when assigning species names to metabarcoding DNA sequences. NatureMetrics uses proprietary methods to curate and filter out such errors through whitelisting, blacklisting, and human curation based on country lookups to ensure that there is reasonable support that the sequence has been correctly associated in the database. Accessions that are flagged as potentially erroneous are not used for taxonomic assignment of metabarcoding sequences. It should be noted that while it is possible to use uncurated public databases for taxonomic assignment, resulting outputs might differ (slightly), such as in species resolution, from those resulting from curated ones.

For the bacterial dataset, assignments were made to the lowest possible taxonomic level where there was consistency in the matches, with minimum similarity thresholds of 98%, 95% and 92% for species-, genus- and higher-level assignments respectively. The same process was applied to the faunal dataset, with an additional country-based sense-checking step subsequently applied to elevate identifications to the taxonomic level supported by GBIF occurrence records for the United Kingdom (rgbif; (Chamberlain et al., 2022).

A similar process was applied to the fish dataset, 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 for the United Kingdom was used to manually

improve vertebrate and invertebrate identifications in cases where there were equally good reference matches (rgbif; (Chamberlain et al., 2022).

ASVs 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; Chamberlain, https://cran.r-project.org/package=rredlist) to obtain global threat status. Finally, low abundance detections were omitted, with filter thresholds set at a percentage of the total reads per sample (bacterial: 0.025%; faunal: 0.025%; fish: 0.02). Results are presented for OTUs identified to the target taxonomic group only (bacterial: Bacteria; faunal: Animalia; fish: Actinopterygii and Cephalaspidomorphi).

Although none of the extraction or PCR blanks contained reads post filtering, for the invertebrate assay, we observed a number of OTUs in a positive control used to monitor the sequencing run. These are indicated on the OTU table as tentative detections.

Data analysis

Data were analysed using the statistical software R v4.1.0 with RStudio v1.4.1717. The package tidyverse v1.3.1 was 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. Using the package tidyverse v1.3.1, taxon richness was calculated for each sample and mean taxon richness between sampling locations and sample types for the was compared. Community similarity was visualised using Non-Metric Multi-Dimensional Scaling (NMDS) plots, created with the package vegan v2.5-7.

Sequencing data summary

The final dataset, across the nine water eDNA samples, contained a total of 385 OTUs: 278 bacteria, 69 invertebrates, and 38 fish (Table 6). The final sediment DNA dataset contained total of 808 OTUs: 694 bacteria, and 114 invertebrates. Taxon-by-sample tables of the samples are attached to this report (NM-PPP825 - ScotGov - Marine pilot report appendices.xlsx). More fish and invertebrate OTUs were identified at the species level compared to bacteria. This reflects differences in availability of reference sequences for different organisms within the reference databases and a higher proportion of assignment conflicts (100% matches to multiple species) in bacteria. Extraction and PCR blanks did not show evidence of amplification and were not sequenced.

Table 6. Summary of the number of OTUs detected and the percentage of OTUs successfully classified at each taxonomic level for each target in the nine water samples

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	278	76.6%	66.9%	48.9%	32%	10.4%	1.8%
Invertebrates	69	100%	95.7%	81.2%	75.4%	33.3%	13%
Fish	38	100%	100%	100%	100%	95.2%	85.7%

Table 7. Summary of the number of OTUs detected and the percentage of OTUs successfully classified at each taxonomic level for each target in the six sediment samples

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	694	68.2%	50.7%	32%	20.32%	3.75%	0.86%
Invertebrates	114	99%	90.4%	92.1%	73.7%	27.2%	4.4%

Results

Fish

In total, 38 fish taxa were detected across the nine water samples collected from Loch Goil (Fig. 8). Five species were unassigned taxa and removed from downstream analyses. Six mammal species (*Felis catus, Sus scrofa, Canis lupus, Bos taurus, Ovis aries*) were identified as contaminants, two of which (*Felis catus* and *Sus scrofa)* were only detected in the field control blank sample and removed from downstream analyses.

The most common taxa were European sprat (*Sprattus sprattus*), Atlantic cod (*Gadus morhua*), and Atlantic mackerel (*Scomber scombrus*), which were detected in all samples. Fifteen taxa were identified in fewer than two samples, including the European eel (*Anguilla anguilla*) and a lamprey species (*Lampetra* sp.) (Figure 8). Ten PMF teleost taxa were detected; European eel (*Anguilla anguilla*), Atlantic herring (*Clupea harengus*), Atlantic cod (*Gadus morhua*), lamprey species (*Lampetra* sp.), whiting (*Merlangius merlangus*), sand goby (*Pomatoschistus minutus*), Atlantic salmon (*Salmo salar*), brown trout (*Salmo trutta*), Atlantic mackerel (*Scomber scombrus*), and Norway pout (*Trisopterus esmarkii*).

Station 3 (closest to the head of Loch Goil) had one sample with fewer marine taxa (e.g., absence of *Clupea harengus* which was present in all other samples), and more freshwater taxa (e.g., was the only sample with the presence of minnow *Phoxinus phoxinus*). Freshwater input is likely to have influenced this sample, which could be due to proximity to shore and a river outlet, or sampling at a shallower depth where the water was less saline.

Three mammal taxa and one bird taxon were identified. Two of the mammal species, the common or harbour seal (*Phoca vitulina*) and the harbour porpoise (*Phocoena phocoena*), are also PMF taxa. The third mammal detected was the roe deer (*Capreolus capreolus*), and the bird DNA recovered was from a species of cormorant (Phalacrocoracidae).

Station 1 had the highest mean teleost fish richness of the three stations (mean ± s.d.: 15.33 ± 1.53),-followed by Station 2 (14.33 ± 1.15), and Station 3 (13.67 ± 0.58) (Figure 9). There is overlap in the NMDS plot between Station 2 and Station 3, whilst Station 1 (in Loch Long) forms a separate cluster (Figure 7).

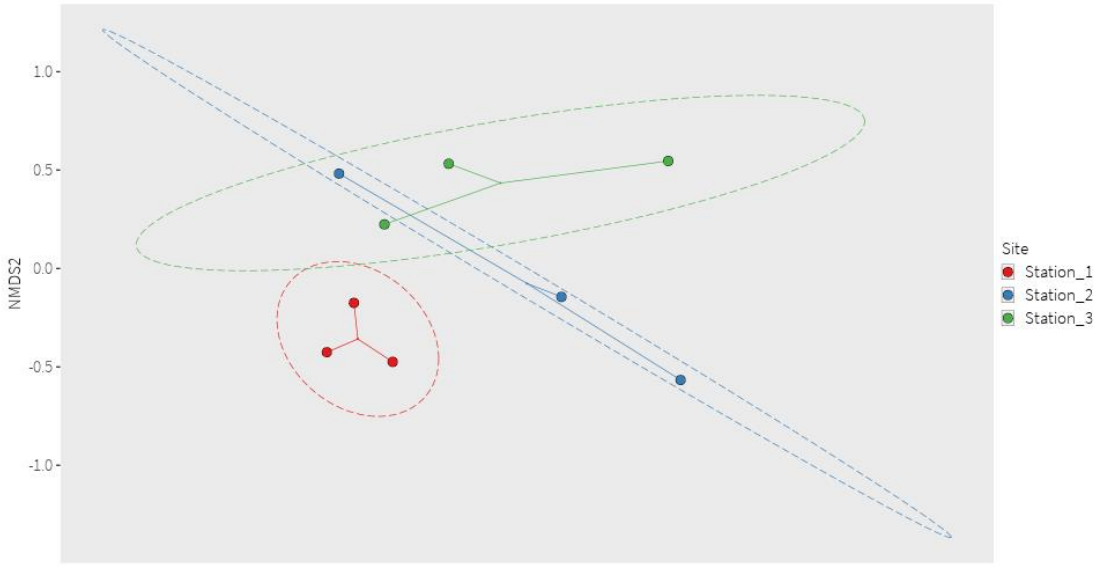

Figure 7: NMDS ordination plots based on Jaccard similarity index for vertebrate community taxonomic compositions. Points are coloured by Station.

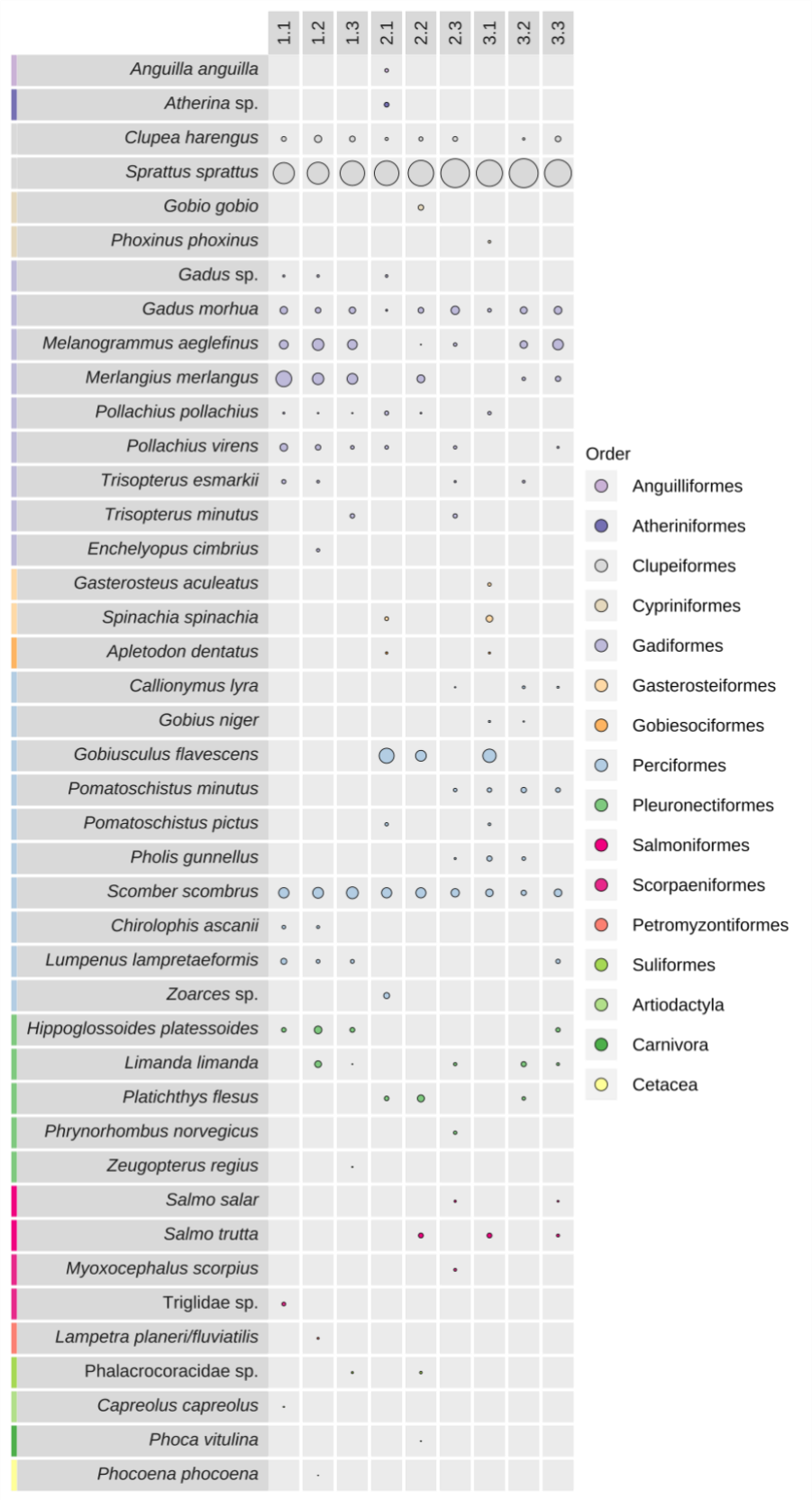

Figure 8: Bubble plot summarising vertebrate detection from water samples collected from Loch Goil with Station number (1 to 3) and Replicate (.1 to .3).

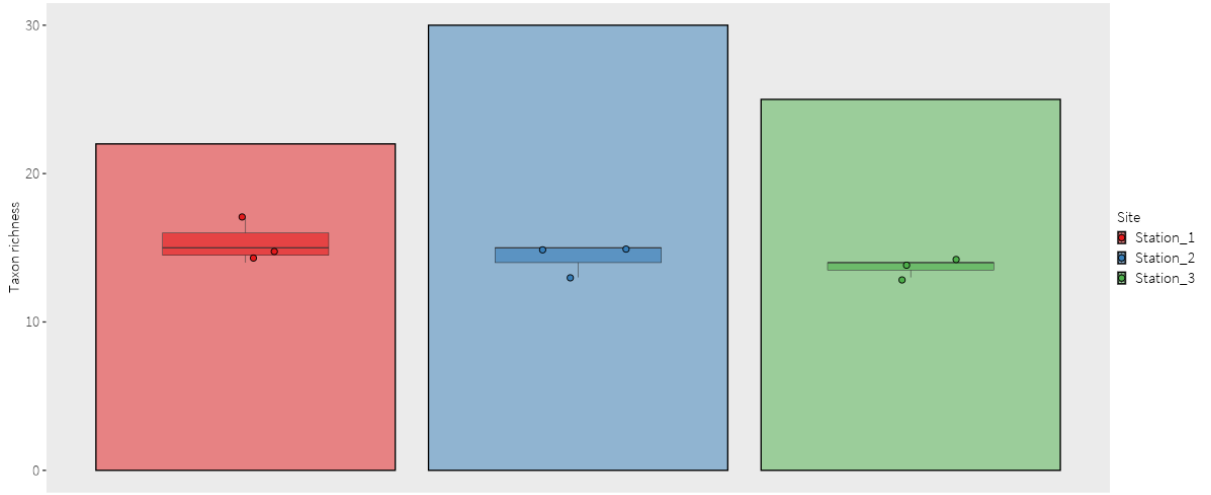

Figure 9: Vertebrate taxon richness from water samples based on eDNA data represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

Invertebrates

It is evident that different taxa have been recovered from the water compared to the sediment samples. The NMDS (Figure 10) illustrates the clear contrast between the different communities identified. No PMF invertebrate species were detected in any of the water or sediment samples, however this needs to take into account the fact that the invertebrate assay did not provide species-level resolution for most taxa. No species or genera were shared between the bulk sample, identified by morphology, and metabarcoding-identified taxa, while 44% of families were common to both methods.

The only PMF taxa without sequences in the reference database were *Alcyonium hibernicum* and *Parazoanthus anguicomus*.

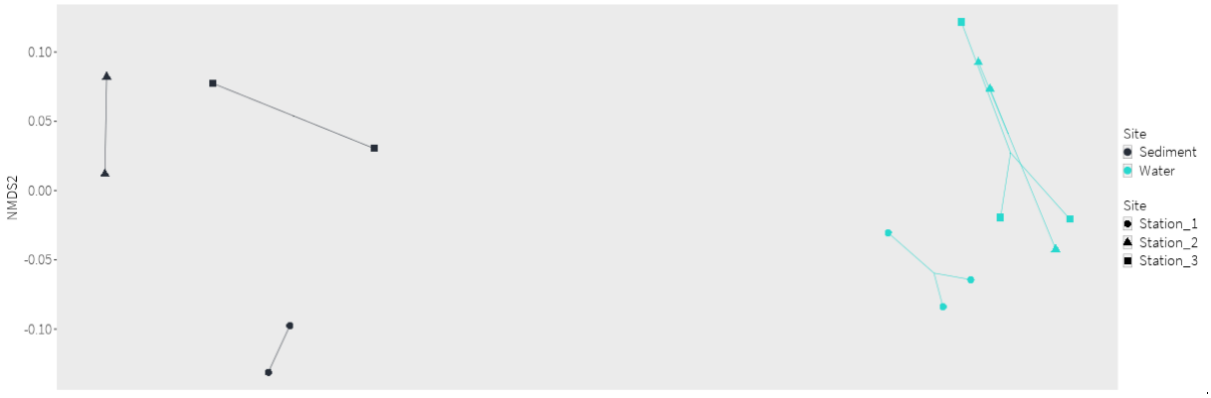

Figure 10: NMDS ordination plots based on Jaccard similarity index for water and sediment invertebrate community taxonomic compositions. Points are coloured by sample type.

Water

In total, 69 invertebrate taxa were detected across the nine water samples collected from Loch Goil. Six OTUs were unassigned taxa and removed from downstream analyses.

26 taxa were identifed fewer than two samples, including *Arenicola marina* (Annelida), *Evadne nordmanni* (Arthropoda), *Calanus finmarchius* (Arthropoda) and *Austrominius modestus* (Arthropoda).

Station 1 had the highest mean richness of the three stations (mean ± s.d.: 31 ± 3.61), followed by Station 3 (21.67 ± 2.65) and Station 2 (23 ± 3.21) (Figure 11). There is overlap in the NMDS plot between Station 2 and Station 3 (both in Loch Goil), whilst Station 1 (in Loch Long) forms a separate cluster (Figure 12), as was also the case for fish eDNA (Figure 7).

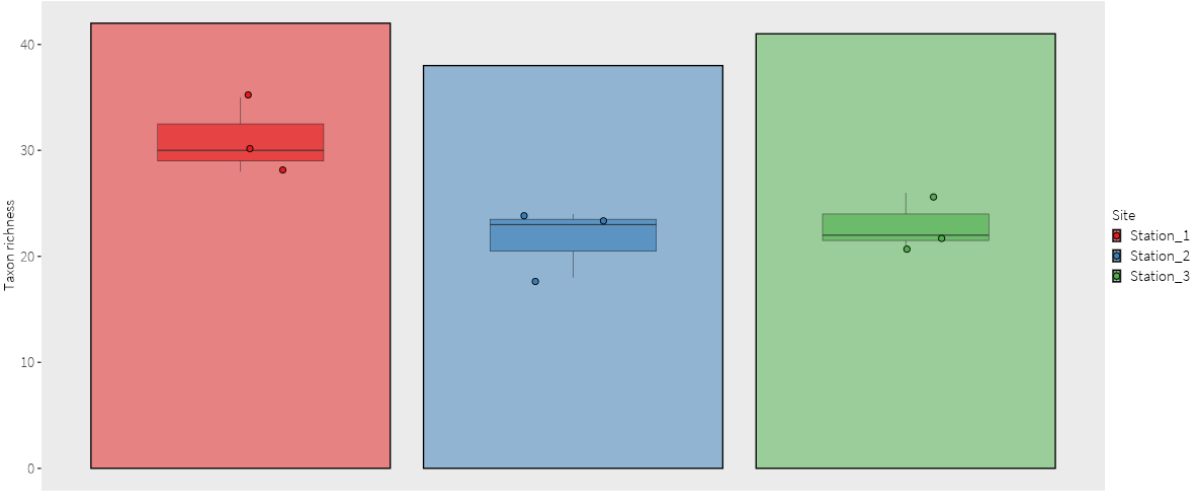

Figure 11: Invertebrate taxon richness from water samples based on eDNA data represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

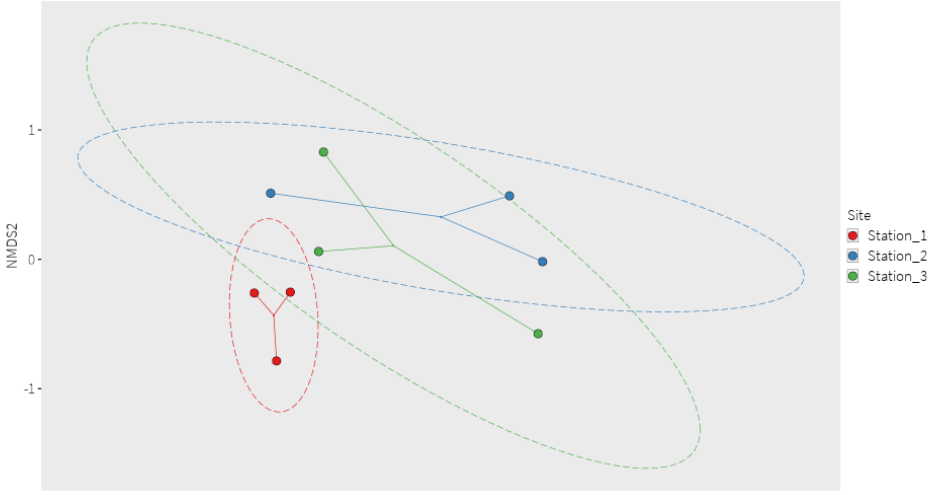

Figure 12: NMDS ordination plots based on Jaccard similarity index for water invertebrate community taxonomic compositions. Points are coloured by Station.

Sediment

In total, 114 invertebrate taxa were detected across the six sediment samples collected from Loch Goil. Of these, 99% were identified to phylum, 90.4% to class, 92.1% to order, 73.7% to family, 27.2% to genus, and 4.4% to species level. 28 OTUs were unassigned taxa (could not be assigned to phylum level) and were removed from downstream analyses.

The taxon with the most reads was a nematode Oncholaimidae sp., although it was not detected at Station 1. 50 taxa were identified in fewer than two samples.

Station 1 had the highest mean richness of the three stations (mean ± s.d.: 46 ± 17.0), followed by Station 2 (37 ± 1.41), and Station 3 (34.5 ± 2.12) (Figure 13). Stations are well grouped by duplicates, as is shown in the NMDS (Figure 14), suggesting low heterogeneity between duplicates within a station.

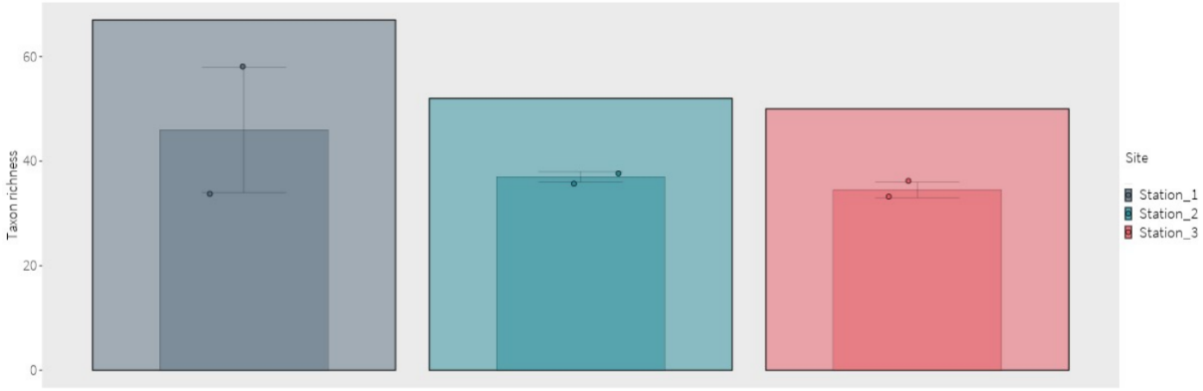

Figure 13: Invertebrate taxon richness from sediment samples based on eDNA data represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

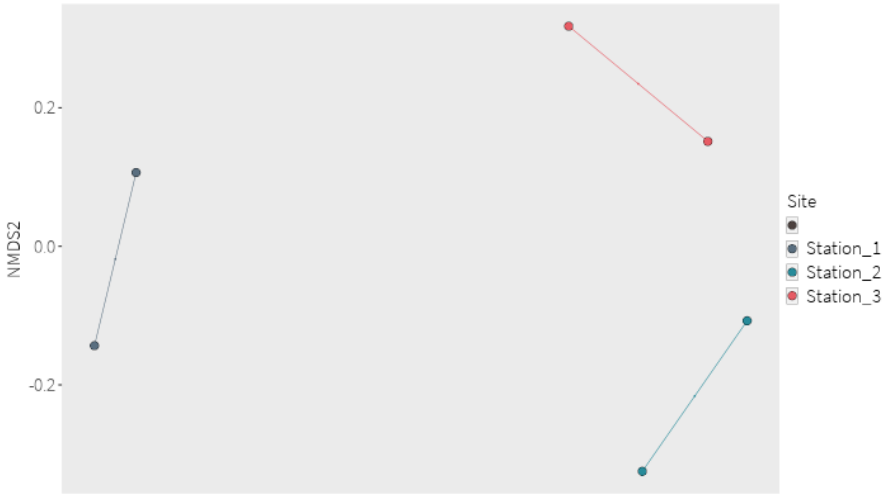

Figure 14: NMDS ordination plots based on Jaccard similarity index for sediment invertebrate community taxonomic compositions. Points are coloured by Station.

Morphological Invertebrate Analysis

In total, 26 invertebrate taxa were detected across the six sediment samples collected from Loch Goil. Of these, 100 % were identified to phylum, 92.3 % to class, 92.3 % to family, 80.7% to genus, and 69.2% to species level. Although no genera were shared between the morphologically identified and metabarcoded taxa, 44% of families identified in the morphological dataset were also found in the metabarcoded dataset. The primary reason for this difference is likely to be the low success rate of identifying species to genus or species level using the chosen primer set, but primer bias is likely to also play a role (van der Loos and Nijland, 2020).

It was noted that the samples were relatively small in terms of abundance and diversity. This may be caused by:

- Using a min-van Veen grab rather than a regular van Veen grab
- Using ethanol as a preservation buffer rather than formalin (as ethanol can reduce the diversity of taxa detected (de Souza and Barros, 2017)

The most abundant taxon was *Mediomastus fragilis* (Annelida) with a total of 51 individuals identified in both Station 3 samples, and in one of the Station 2 samples.

Station 3 had the highest mean richness of the three stations (mean ± s.d.: 11.5 ± 2.12), followed by Station 1 (8.5 ± 2.12), and Station 2 (7 ± 2.83) (Figure 15). Whilst Stations 1 and 2 have very similar communities identified between replicates, Station 3 shows a higher variation between replicate communities (Figure 16).

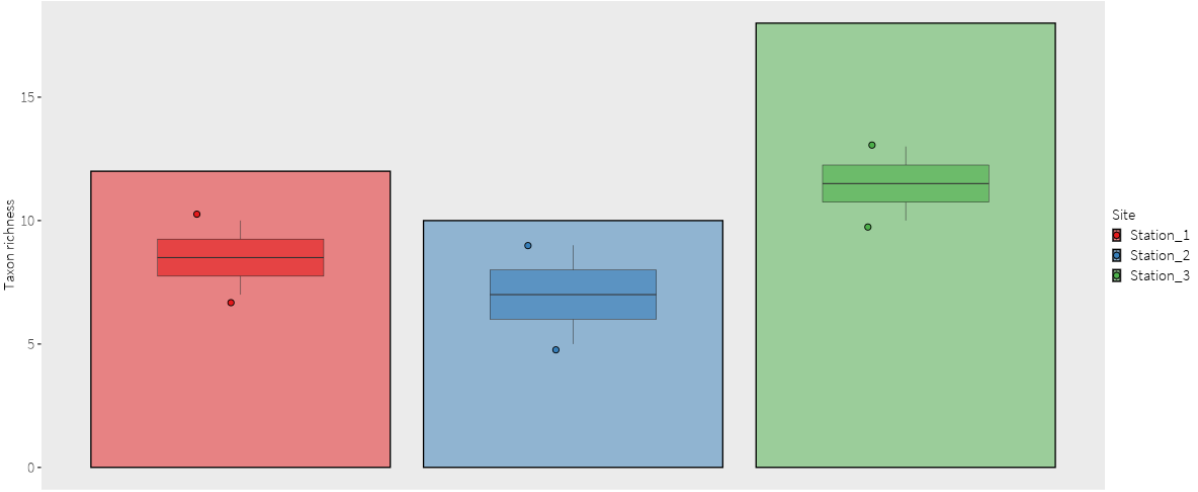

Figure 15: Morphologically identified invertebrate taxon richness from sediment samples represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

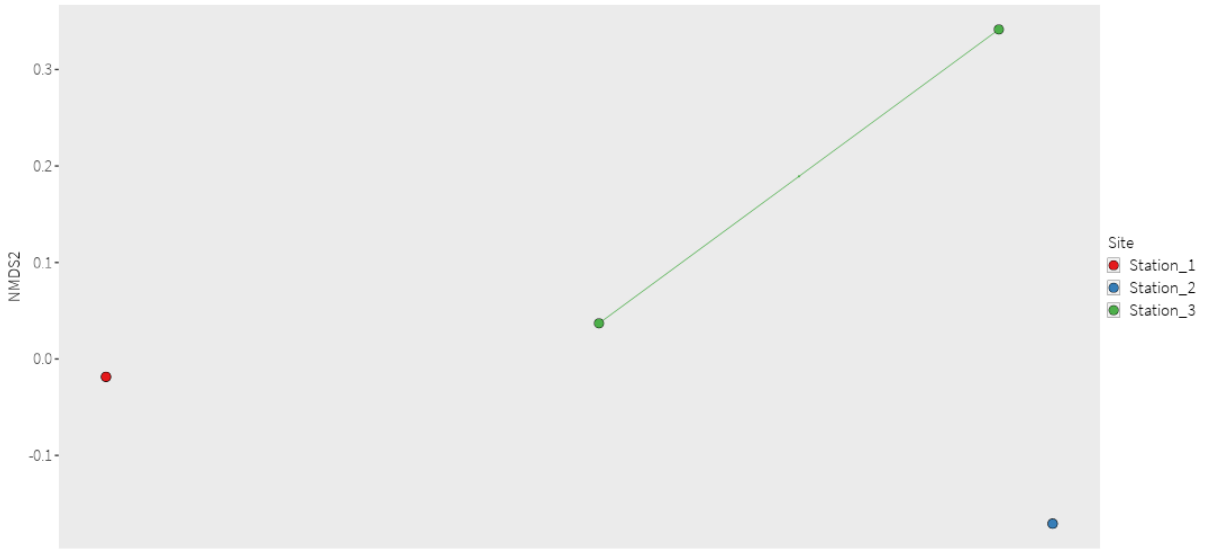

Figure 16: NMDS ordination plots based on Jaccard similarity index for sediment morphologically identified invertebrate community taxonomic compositions. Points are coloured by Station.

Bacteria

As with the invertebrates, different taxonomic groups were recovered from the water and sediment samples, and the data recovered from the two sample types should be treated separately.

Water

Unfortunately, one bacterial filter sample from each station failed to amplify during PCR. This may be related to preservation issues encountered in the field. In total, 278 bacteria taxa were detected across the five water samples collected from Loch Goil. 76.6% were identified to phylum, 66.9% to class, 48.9% to order, 32% to family, 10.4% to genus, and 1.8% to species level.

85 taxa were identified in fewer than two samples, including *Ilumatobacter nonamiensis* (Actinobacteriota), *Arcobacter aquimarinus* (Campylobacterota), *Salmonella enterica* (Proteobacteria), and the protozoan *Eutreptiella gymnastica* (Euglenozoa).

Station 2 had the highest mean richness of the three stations (mean ± s.d.: 151.5 ± 31.8), followed by Station 1 (132), and Station 3 (127.5 ± 7.78) (Figure 17). Even though only a small number of samples was collected, the results suggest that there is a high degree of variation between replicate samples within a station (Figure 18). This indicates that there may be a 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.

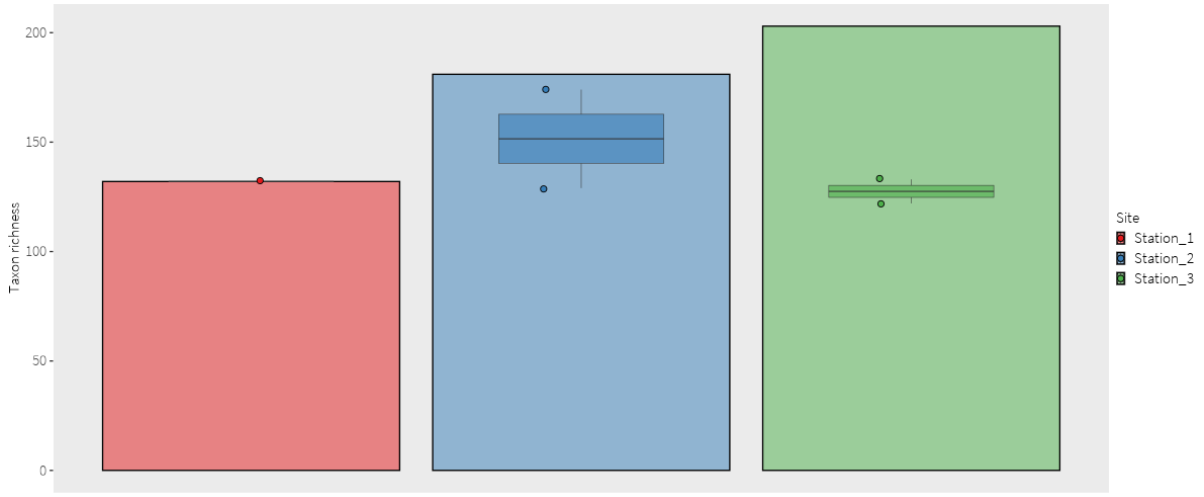

Figure 17: Bacteria taxon richness from water samples based on eDNA data represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

Figure 18: NMDS ordination plots based on Jaccard similarity index for water bacteria community taxonomic compositions. Points are coloured by Station.

Sediment

In total, 694 bacteria taxa were detected across the six sediment samples collected from Loch Goil. 20.32 % were identified to family, 3.75 % to genus, and 0.86 % to species level. 63 OTUs were unassigned taxa and removed from downstream analyses.

140 taxa were identified in fewer than two samples, including two Acidobacterium sp. (Acidobacteriota), *Colwellia* sp. (Proteobacteria), *Pelagibius* sp. (Proteobacteria), *Planctomyces* sp. (Planctomycetota), and 4 Archaea taxa.

Station 1 had the highest mean richness of the three stations (mean ± s.d.: 437.5 ± 6.37), followed by Station 3 (399 ± 2.82) and Station 2 (377.5 ± 9.19) (Figure 19). The variation in community structure between replicates was far less than the variation between stations (Figure 20), suggesting that bacteria sediment samples would be useful for distinguishing

between biotopes, if future research was done involving numerous biotopes to assess how the bacterial signature relates to each biotope.

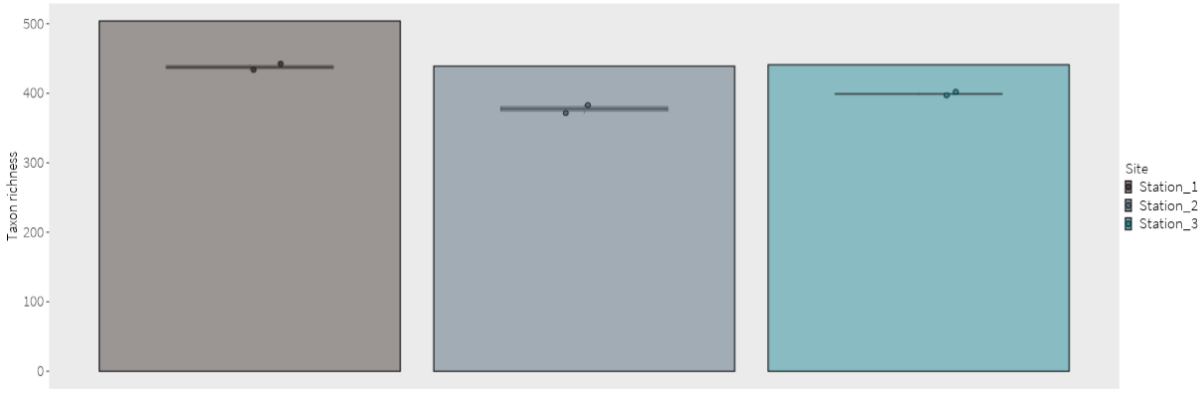

Figure 19: Taxonomic richness of bacteria from sediment samples collected represented per sample (smaller bar is the mean, individual samples are shown as points), and per station (larger bar).

Figure 20: NMDS ordination plots based on Jaccard similarity index for sediment bacteria community taxonomic compositions. Points are coloured by Station.

4.2 Pilot results freshwater

Introduction

The pilot sampling campaign took place in August – October 2021 and the main sampling campaign took place in August – September 2022. Within each habitat, the pilot study has investigated some key parameters of interest, such as the effects of subsampling or the starting sample type (e.g., water or sediment). The data generated in the pilot were interrogated to assess their usefulness for detecting key features or informing target indicators. This report details the methods and results from the pilot phase of the project for the freshwater sampling. These results have fed into the decision making to design the sampling and analysis approach of the P2 Sampling Plan.

Methods

Study sites

Loch Lomond and the Endrick Water were selected to conduct a pilot study to identify optimal sampling strategy, appropriate laboratory analysis, and project objectives to consider pursuing for the freshwater habitat in Phase 2. Loch Lomond has distinct ecological zones as well as contrasting bathymetry and land use between its north and south basins. The Endrick Water flows into Loch Lomond and has point impacts from sewage treatment works and major broad-scale impacts such as land-use changes (e.g. tourism, forestry, infrastructure).

Water was sampled from both basins in Loch Lomond for a preliminary comparison of biodiversity present in areas of high (South Basin) and low (North Basin) anthropogenic disturbance. The sampling strategy was based on protocols established by (Hänfling et al., 2016) with some minor modifications. Five 3.3 L water samples were collected from each basin at approximately equidistant locations where possible, taking shoreline access by foot into consideration (Figure 21; Appendix C). 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 3.5 L sterile sampling bag, the bag was sealed and shaken for 20-30 seconds to homogenise the water sample. To capture sufficient DNA for multiple downstream laboratory analyses, two filter units, each an enclosed casing containing a 0.8 µm PES filter and a 5 µm glass fibre prefilter (NatureMetrics, UK), were used per 3.3 L sample. Using a 100 mL syringe, up to 1.5 L of water was pushed through each filter unit and preserved with Longmire's preservation buffer (Longmire et al., 1997). The volume of water passed through each filter was recorded.

Water and macroinvertebrates were sampled at three locations along the Endrick Water for a preliminary comparison of biodiversity present along the length of the Endrick Water, sampling strategies and sample types. An established sampling strategy for rivers and streams, based on the collection of independent water samples across the width of the stream/river section (Bruce et al., 2021), was compared with an unconventional sampling strategy using a single composite water sample. At each location, three 2 L water samples were collected from the left

bank, mid-channel, and right bank, following which 1 L from each sample was pooled into a new sampling bag to create a 3 L composite sample, leaving three 1L independent samples (Figure 22). One litre from each independent and composite sample was passed through one of two 0.8 µm PES filters. The spare filters were used on the remaining water in the composite sample. A 3-minute kicknet survey was performed at each location by a SEPA ecologist to collect a macroinvertebrate sample. The sample was roughly sorted in the field to remove stones and vegetation, before specimens were preserved in 1 L of ≥99% analytical-grade ethanol.

All samples were transported to the NatureMetrics laboratory (Guildford, UK) at ambient temperature and stored at -20°C until analysis.

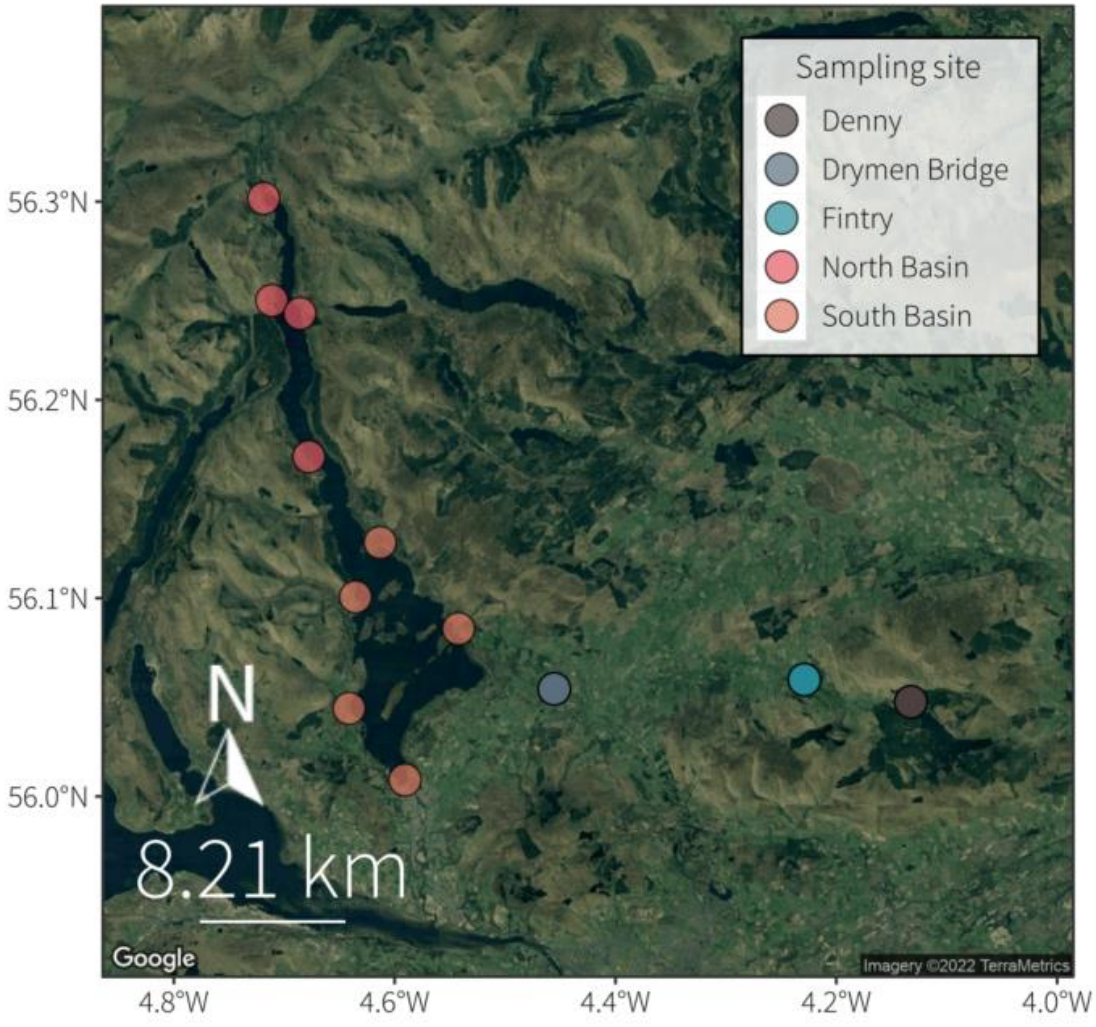

Figure 21. Map of sampling locations across Loch Lomond (North and South Basin) and the Endrick Water (Denny, Fintry, and Drymen Bridge).

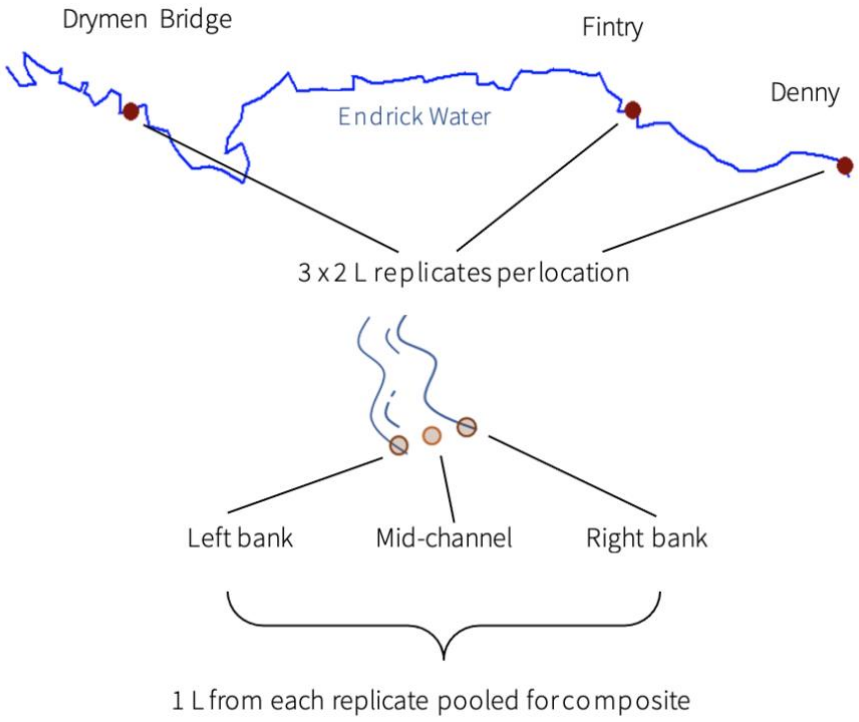

Figure 22: Sampling scheme for Endrick Water, showing three sampling **locations** (Drymen Bridge - downstream, Fintry – midsection, Denny - upstream), with three **replicate** samples and one **composite** sample per location, where the composite sample was made using 1 L from the three replicate samples.

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. Only one DNA extract resulting from one of two filters used for each sample collected from Loch Lomond and only DNA extracts resulting from the 1 L independent and composite samples from the Endrick Water were used for subsequent analyses. DNA was purified to remove PCR inhibitors using a DNeasy PowerClean Pro Cleanup Kit (Qiagen).

DNA was extracted from each bulk tissue sample using a homemade lysis buffer (Tris-HCl, NaCl, CaCL, EDTA, SDS, Proteinase K, DTT) with a DNeasy Blood and Tissue Kit (Qiagen). First, ethanol was poured off and samples were left to dry for 24 hours in an incubator at 30°C. Enough lysis buffer was added to cover each sample (200-400 ml), then samples were incubated at 56°C for 4 hours. After incubation, 14 ml of lysate from each sample was passed through a set of five spin columns (2.8 ml per spin column), followed by 500 µl of Buffer AW1, then 500 µl of Buffer AW2. Finally, 50 µl of Buffer AE was passed through each spin column and the five extracts per sample were pooled together.

An extraction blank was processed with each batch of extractions to assess potential contamination in the extraction process. DNA extracts were quantified using a Qubit dsDNA HS Assay Kit on a Qubit 3.0 fluorometer (Thermo Scientific).

PCR amplification

Regions of the 12S ribosomal RNA (12S; vertebrates), cytochrome c oxidase subunit I (COI; invertebrates), and 16S ribosomal RNA (16S; bacteria) genes were amplified via a two-step PCR process. Tails were added at the 5′ end to be complementary with Illumina Nextera index primers. For each water sample, 12 PCR replicates were performed using the 12S primers (Kelly et al., 2014; Riaz et al., 2011), and three replicates were performed using COI primers (Leese et al., 2021) and 16S primers (Apprill et al., 2015; Parada et al., 2016). For each bulk invertebrate sample, three replicates were performed using COI primers (Wangensteen et al., 2018). The COI primers for water samples (Leese et al., 2021) also used a nested PCR protocol, where 1 µl of the initial PCR product was amplified again under the same PCR conditions.

All PCRs were carried out in a total volume of 25 µL. 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). The invertebrate eDNA amplification mixture consisted of 1X DreamTaq Green PCR Master Mix, 0.3 µM of each primer, 0.8 mg/ml of BSA, 3% of DMSO, 1 µl of template DNA, and PCR grade water. The bacteria amplification mixture consisted of 1X DreamTaq Green PCR Master Mix, 0.2 µM of each primer, 0.25 mg/ml of BSA, 1 µl of template DNA, and PCR grade water. The bulk invertebrate amplification mixture consisted of 1X DreamTaq Green PCR Master Mix, 0.3 µM of each primer, 0.8 mg/ml of BSA, 0.9 µl of template DNA, and PCR grade water.

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. Invertebrate eDNA PCR conditions consisted of: an initial denaturation at 95°C for 5 min; 10 cycles of 95°C for 30 s, a 45 s touchdown annealing step (-0.5°C per cycle) starting at 53°C, then 72°C for 60 s; 30 cycles at 95°C for 30 s, 48°C for 45 s, and 72°C for 60 s; and a final elongation step at 72°C for 5 min. Bacteria PCR conditions followed (Caporaso et al., 2011) but using 30 cycles rather than 35 cycles. Bulk invertebrate PCR conditions consisted of: an initial denaturation at 95°C for 10 min; 35 cycles at 94°C for 60 s, 45°C for 60 s, and 72°C for 60 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.

Library preparation was identical to that in Section 4.1.

Bioinformatics

Bioinformatic processing followed the same pipeline as for marine assays detailed in Section 4.1 with specific differences noted below.

Length filters following merging were 80-120 bp and 120-160 bp for the vertebrate and invertebrate assays respectively.

The NCBI nucleotide (NCBI *nt*) database was queried for all four assays, with the invertebrate and bacterial datasets additionally queried against BOLD (Ratnasingham and Hebert, 2007) and SILVA (Quast et al., 2012; Yilmaz et al., 2014). Searches against BOLD were made using

Minimum similarity thresholds for taxonomic assignment for species-, genus- and higher-level assignments were 99%, 97% and 95% respectively for vertebrates, and 98%, 95% and 92% for invertebrates.

All OTUs with species-level identifications were queried against the IUCN Red List (rredlist; Chamberlain 2018, Hsieh et al., 2016) to obtain global threat status. Finally, low abundance detections were omitted, with filter thresholds set at a percentage of the total reads per sample (vertebrate eDNA: 0.02%; invertebrate eDNA: 0.015%; bacteria: 0.1%; bulk invertebrates: 0.01%). Results are presented for OTUs identified to the target taxonomic group only (vertebrate: Chordata excluding human and domestic animals; invertebrate: Animalia excluding Tetrapoda; bacterial: Bacteria).

Data analysis

Data were analysed using the statistical software R v4.1.0 with RStudio v1.4.1717. The package tidyverse v1.3.1 was used for data manipulation, formatting, and visualisation. The total read counts per sample were used to calculate the proportional read counts for each taxon. Bubble plots showing contaminants, NUMTs and true positive detections in samples were produced using the package ggplot2 v3.3.5. Using the package tidyverse v1.3.1, taxon richness was calculated for each sample and mean taxon richness between Loch Lomond's basins as well as sampling locations and sample types for the Endrick Water was compared. Community similarity was visualised using Non-Metric Multi-Dimensional Scaling (NMDS) plots created with the package vegan v2.5-7.

Species accumulation and sample coverage curves were calculated for incidence data (presence/absence) across samples in each group (North and South Basin of Loch Lomond) for the vertebrate and invertebrate markers using the iNEXT package v2.0.20 in R . For the Lake Classification example, accumulation curves were calculated only for OTUs within Actinopterygii and Cephalaspidomorphi. To evaluate whether OTU community composition differed amongst locations (North and South Basins of Loch Lomond; upstream, mid and downstream reaches of the Endrick Water), a model-based analysis of multivariate data based on binomial generalised linear models was used in R package mvabund v4.1.12 (Wang et al., 2012).

Results

Overview of detections by assay

In total, 63 vertebrate taxa were detected across 22 individual water samples (Endrick Water) collected from Loch Lomond (five per basin) and the Endrick Water (four per location). Twelve species were identified as contaminants and/or NUMTs and removed from downstream analyses (Figure 33). Eurasian minnow (*Phoxinus phoxinus*) and brown trout (*Salmo trutta*) were detected in all samples but also in the field blanks, and therefore eliminated from further analysis except the lake classification which brown trout was retained for. The most common species were European eel (*Anguilla Anguilla*) and three-spined stickleback (*Gasterosteus aculeatus*). Fourteen taxa were only detected in one sample, including the European bullhead (*Cottus gobio*), jack snipe (*Lymnocryptes minimus*), mandarin duck (*Aix galericulata*), Reeve's muntjac (*Muntiacus reevesi*), red squirrel (*Sciurus carolinensis*), and European otter (*Lutra lutra*) (Figure 34a). In Loch Lomond, perch or zander (*Perca fluviatilis / Sander lucioperca*) was the most common taxon and detected in eight samples, whereas European eel, three-spined stickleback and red deer (*Cervus elaphus*) were each detected in 12 samples. Twenty and nine OTUs were detected in just one sample from Loch Lomond and the Endrick Water, respectively (Appendix A). Five vertebrate species on the UKBAP (BRIG, 2007) list (*Anguilla anguilla*, *Salmo salar*, *Bufo bufo*, *Lutra lutra*, *Sciurus vulgaris*) were found at all but one of the sample locations across Loch Lomond (9) and Endrick Water (3).

For invertebrates, a total of 591 OTUs were detected across all samples. The most common taxa identified to species level were *Limnophyes edwardsi* (Chironomidae), *Baetis rhodani* (Baetidae) and *Leuctra fusca* (Leuctridae), detected in 17 samples each, followed by *Ectopsocus briggsi* (Ectopsocidae) and *Conchapelopia melanops* (Chironomidae) in 15 samples. In total, 196 OTUs were detected in just one sample. Three samples did not return any taxa for the invertebrate assay (two from Loch Lomond, and a mid-channel sample from the Endrick Water; Figure 35, Appendix B). One species, Iron Blue Mayfly (*Nigrobaetis niger*), showing a marked decline in UK was found from the UKBAP list (BRIG 2007) at Denny and Drymen Bridge (Endrick Water).

For bacteria, 1064 OTUs were detected across all samples. The most common genera detected in all 22 samples were *Flectobacillus*, *Rhodobacter*, *Novosphingobium*, *Polaromonas*, *Polynucleobacter*, *Methylobacter*, *Chthoniobacter*, *Opitutus*, and *Prosthecobacter*. The lease common genera present in only one sample were *Edaphobacter*, *Arthrobacter*, *Alloprevotella*, *Flavihumibacter*, *Solitalea*, *Pseudanabaena*, *Turicibacter*, *Roseburia*, *Azospirillum*, *Rubellimicrobium*, *Denitratisoma*, *Anaerobiospirillum*, and *Spirochaeta*.

Extraction and PCR blanks did not show evidence of amplification and were not sequenced.

Comparison of Loch Lomond and Endrick Water

For vertebrates, Loch Lomond and the Endrick Water were comparable in terms of observed overall richness (mean ± s.d.: Loch Lomond = 20.3 ± 6.2, Endrick Water = 21 ± 7.4), although individual samples varied (Figure 23a) and estimated richness was higher for Loch Lomond

(Table 8). Invertebrate observed OTU richness was also similar between Loch Lomond and Endrick Water (Loch Lomond: 99.6 ± 70; Endrick Water: 113 ± 85.5; Figure 25, Table 8). However, invertebrate OTU richness was more variable than vertebrate OTU richness (Table 8, Figure 23b), especially for each of the three replicates per sampling location along Endrick Water. Bacteria observed OTU richness was higher in Endrick Water than Loch Lomond (Loch Lomond: 262 ± 52.7; Endrick Water: 350 ± 71.6).

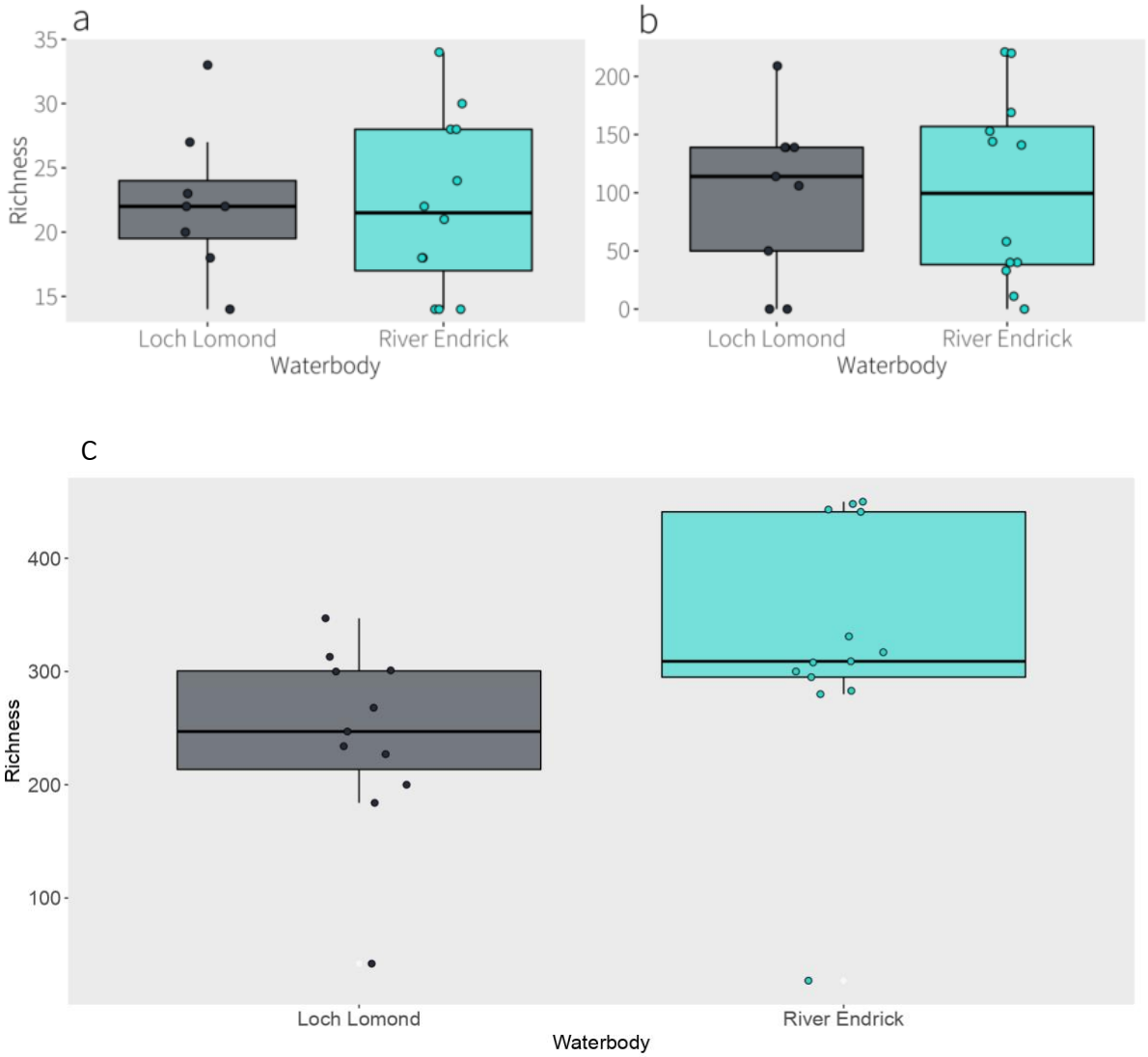

Figure 23. Comparison of observed OTU richness for (a) vertebrates, (b) invertebrates and (c) bacteria between samples collected from Loch Lomond and Endrick Water.

Comparison of Loch Lomond's North and South Basin

For vertebrates, Ardleish was the least rich location with 12 OTUs versus Luss where 33 OTUs were detected. For invertebrates, no OTUs were found at Inveruglas and Rowardennan (Scottish Centre for Ecology and the Natural Environment), whereas the highest richness of 209 OTUs was observed at Inversnaid. For bacteria, Luss was the most rich location with 347 OTUs versus the least rich location of Balloch with 184 OTUs.

Observed vertebrate OTU richness for the North Basin (18.4 ± 6.1) and South Basin (22.2 ± 6.3) was comparable, but richness of individual sampling locations varied (Figure 24). Estimated richness does not appear to show a significant difference between basins with the current number of samples (Figure 25a, b). However, for invertebrates, the North Basin had a higher observed and estimated richness (Figure 25c, d, Table 8), although both basins had high variability among samples (North Basin: 115.5 ± 86.9; South Basin: 86.8 ± 60.6). For bacteria, the North Basin had slightly higher observed richness than the South Basin (North Basin: 286 ± 27.4; South Basin: 238 ± 64).

In terms of community similarity, while ordination plots showed slightly less overlap for invertebrates than vertebrates between the North and South Basins (Figure 26), multivariate model-based comparisons between basins showed no significant differences (Table 10). However, bacterial communities displayed high dissimilarity between basins.

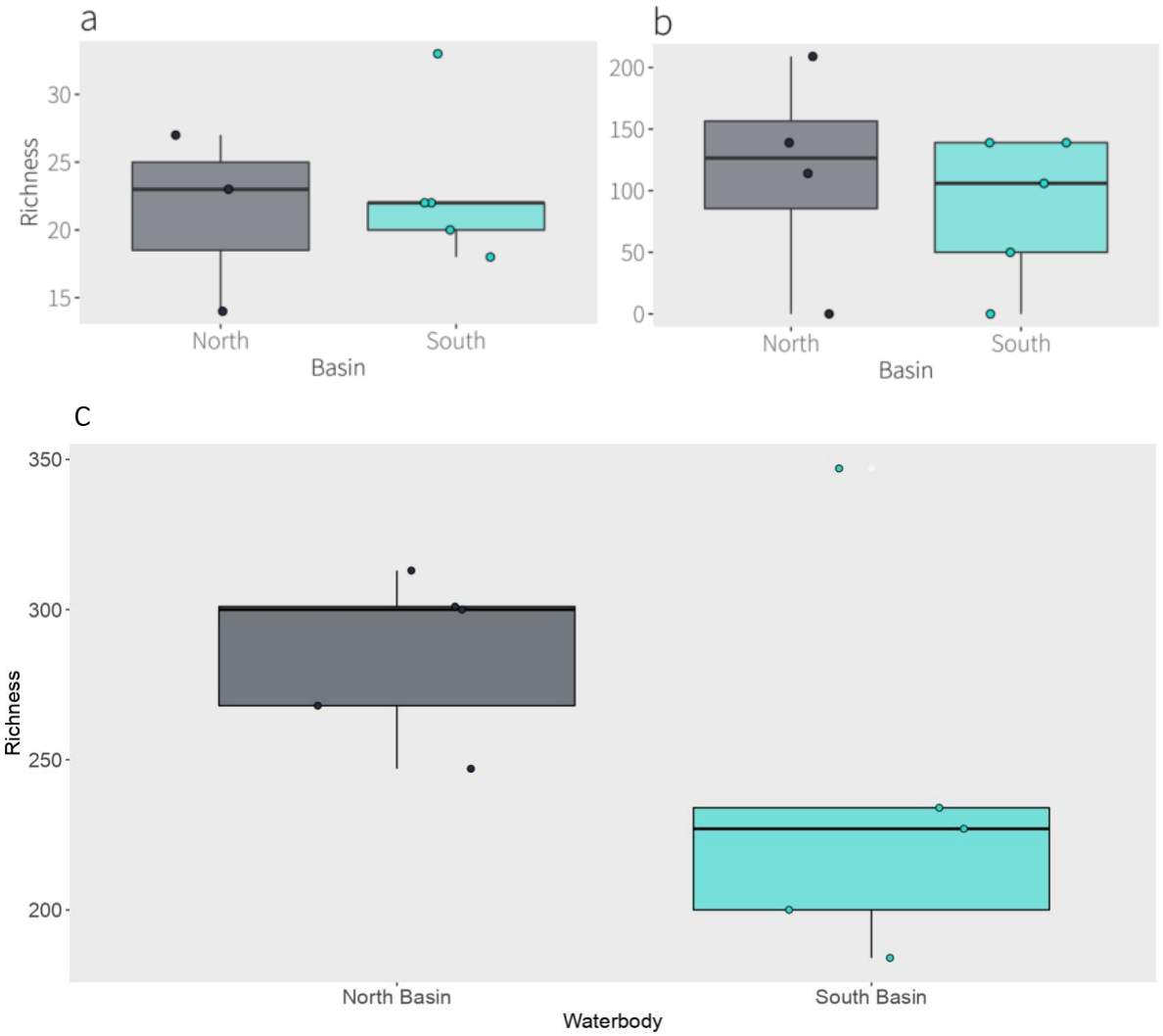

Figure 24. Box plot showing the observed OTU richness of (a) vertebrates, (b) invertebrates and (c) bacteria detected at each sampling location within the North and South Basins of Loch Lomond.

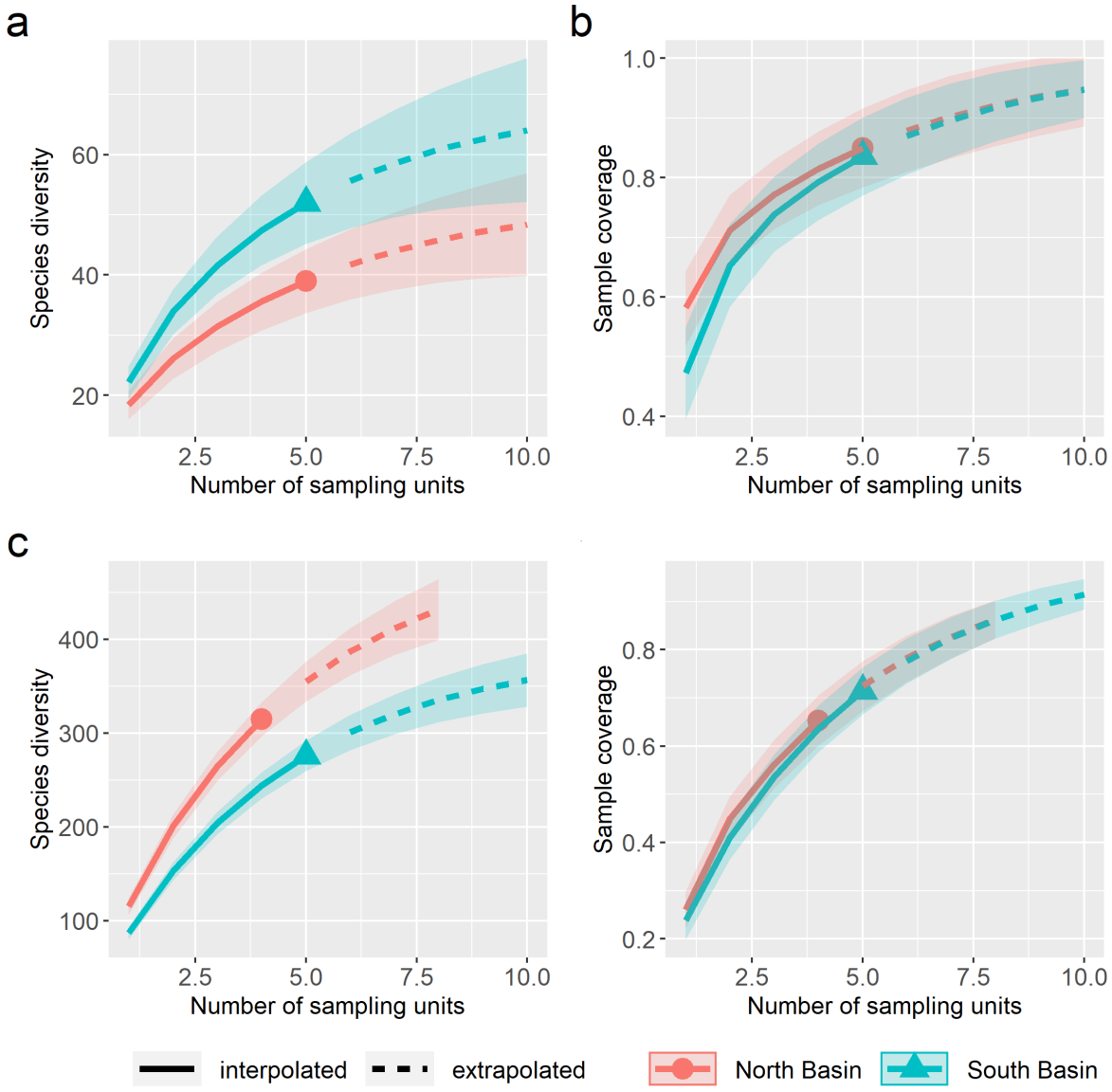

Figure 25. Accumulation curves for OTU richness (left column) and sample coverage (right) by number of sampling units for vertebrates **(a, b)** and invertebrates **(c, d)** in the North (n = 5) and South Basin (n = 5) of Loch Lomond.

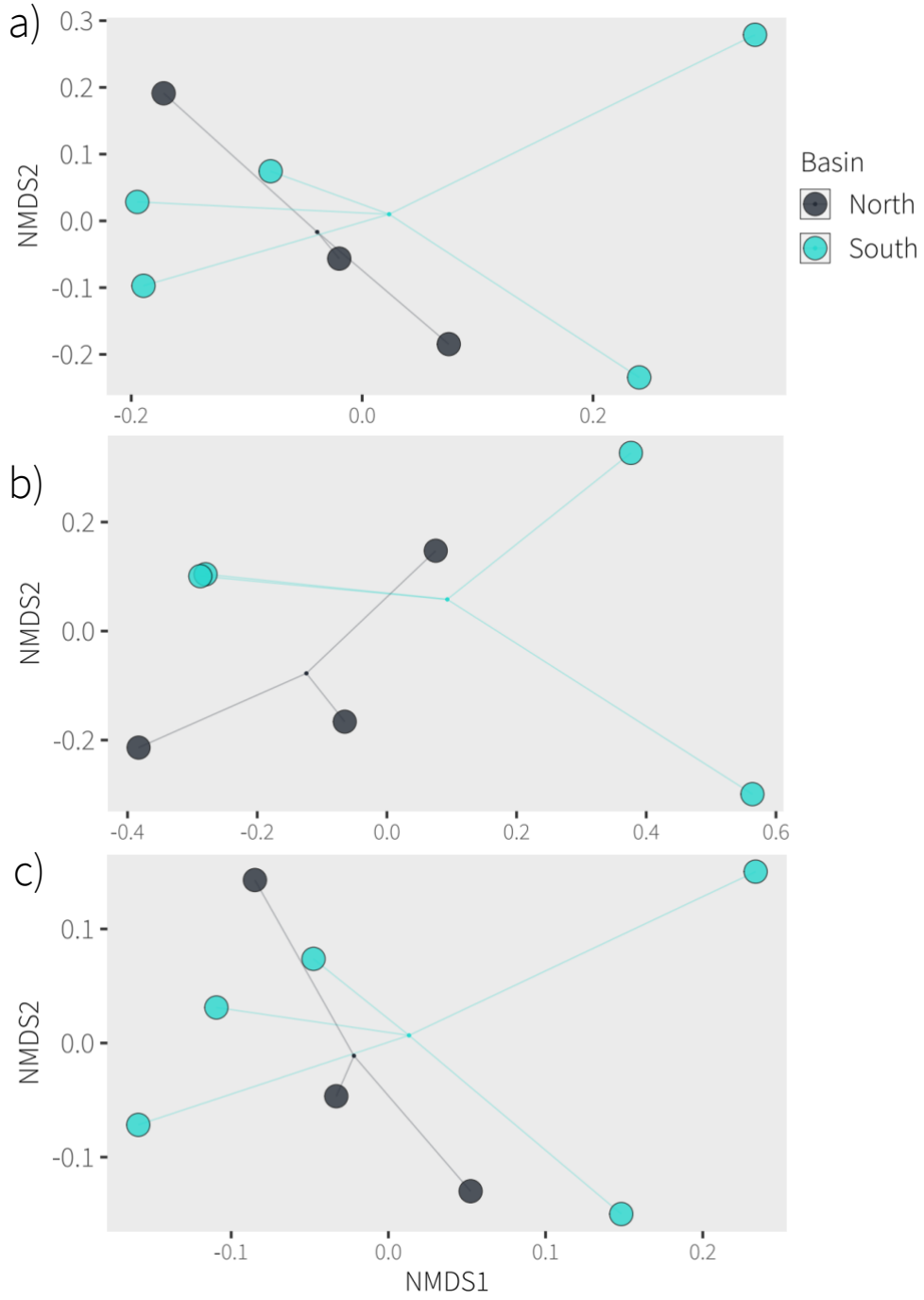

Figure 26. NMDS of (a) vertebrate, (b) invertebrate and (c) bacteria community dissimilarity between individual shoreline samples from the North and South Basins of Loch Lomond.

Comparison of downstream, midsection and upstream eDNA samples along the Endrick Water

Vertebrate richness increased from upstream (Denny) to downstream (Drymen Bridge) sampling locations (Denny = 13 ± 1, Fintry = 20.3 ± 2.5, Drymen Bridge = 29.7 ± 2.1) within the Endrick Water, possibly due to eDNA transport and accumulation (Figure 27). However, this pattern was not seen with invertebrates or bacteria, with Denny having the highest observed invertebrate (137.7 ± 91) and bacteria (444 ± 3.61) OTU richness compared to Fintry (invertebrates: 98.0 ± 106.6; bacteria: 313 ± 16.1) and Drymen Bridge (invertebrates: 103.3 ± 90.6;

bacteria: 296 ± 13), possibly due to eDNA dilution. Multivariate community comparison showed significant differences between upstream to downstream sampling locations for vertebrates and bacteria, but not for invertebrates (Table 9), as also visualised on ordination plots (Figure 27).

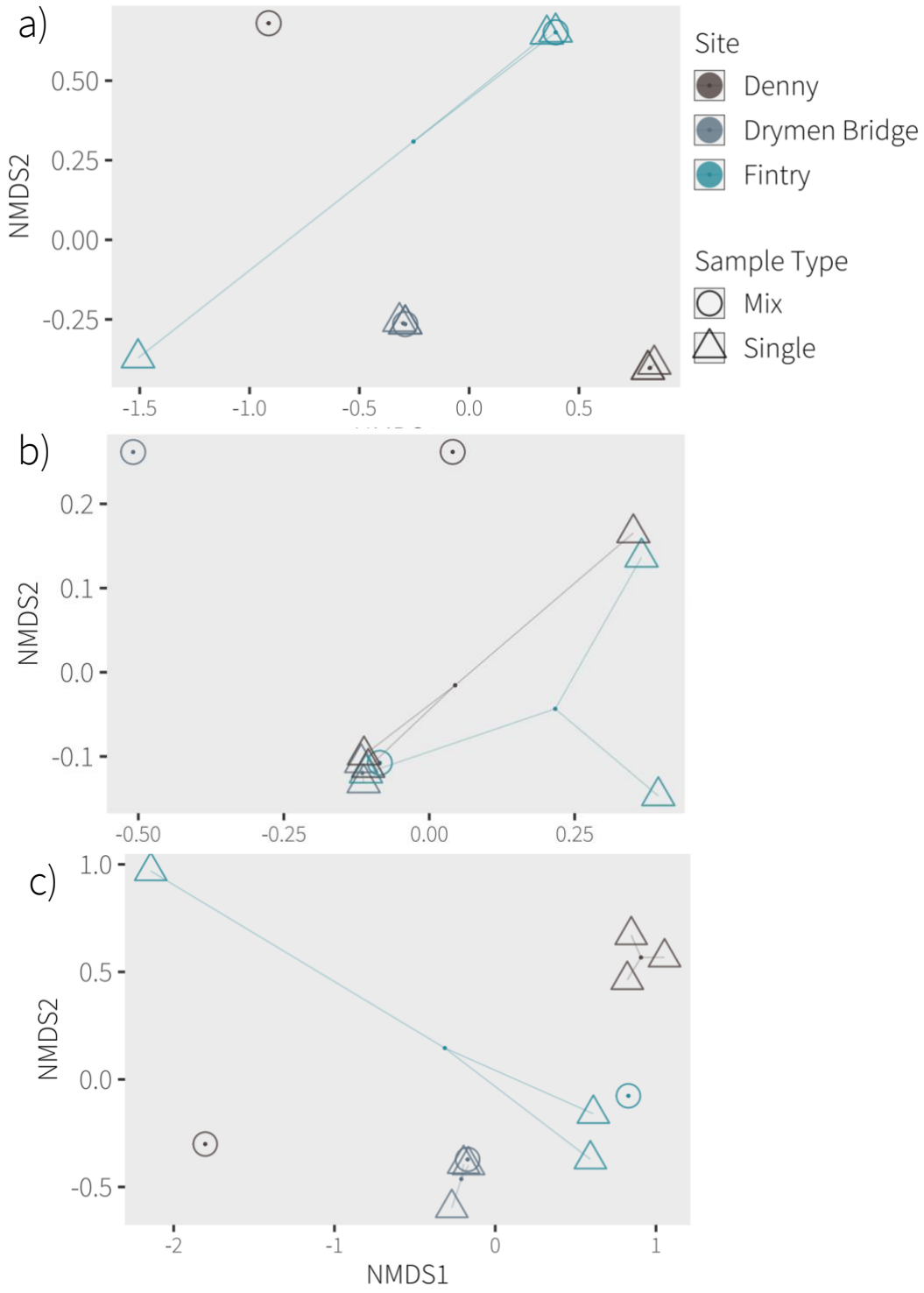

Figure 27. NMDS of (a) vertebrate, (b) invertebrate and (c) bacteria community dissimilarity for replicate (open points) and composite (outlined points) samples at three sampling locations along the Endrick Water: upstream (Denny), midsection (Fintry) and downstream (Drymen Bridge).

Comparison of composite versus replicate eDNA samples at each sampling location on Endrick Water

Overall vertebrate, invertebrate and bacteria OTU richness for all three sampling locations on Endrick Water was higher for replicate samples than for composite samples (Table 8). The relationship generally held at the level of sample locations except for vertebrate detections in the composite sample from Denny, which showed higher observed richness than the combined replicates across the river (left bank, mid-channel and right bank; Figure 27, Figure 28, Table 8).

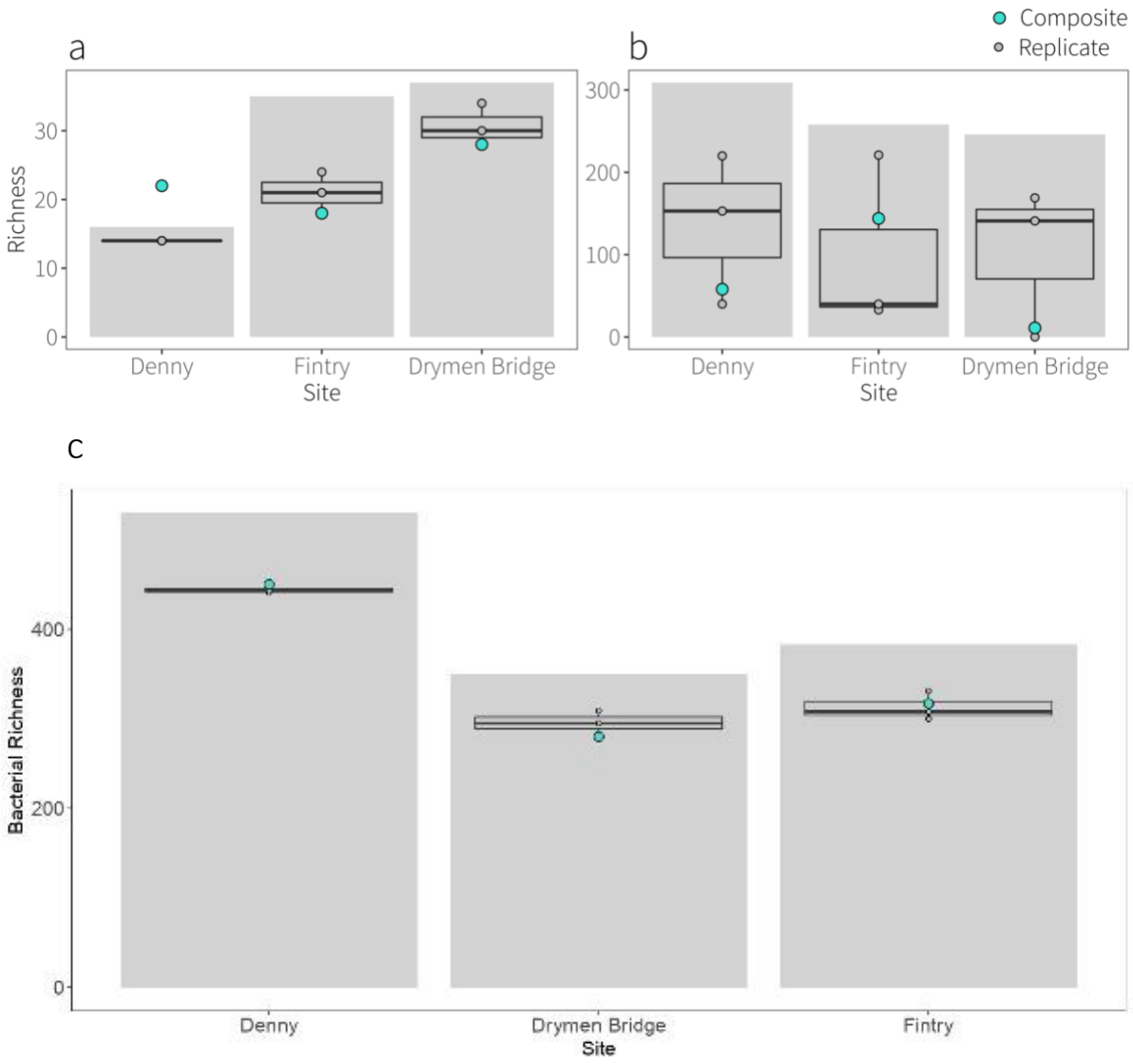

Figure 28. Summary of (a) vertebrate, (b) invertebrate and (c) bacteria OTU detections from replicate (boxplot) and composite (green points) water samples, as well as overall OTU richness (grey bars) from replicate water samples, collected at upstream (Denny), midsection (Fintry) and downstream (Drymen Bridge) sampling locations along the Endrick Water.

Comparison of eDNA and bulk invertebrate samples for Endrick Water

Taxon overlap for Endrick Water overall, and at each sampling location along Endrick Water, between eDNA and the bulk tissue DNA samples is summarised in Figure 29. The three replicate eDNA samples combined detected the most unique taxa overall (Figure 29a) and at each sampling location (Figure 29b-d). Bulk tissue DNA samples produced the most unique taxa after this (Figure 29a, b, d), except at Fintry where the composite eDNA sample detected more taxa than the bulk tissue DNA sample (Figure 29c). Overlap between sample types was low overall and at each location. A large number of taxa were identified by both replicate and composite eDNA samples (Figure 29a, c, d), except at Drymen Bridge where very few taxa were shared by the replicate samples and the composite eDNA sample (Figure 29b). Indeed, the composite eDNA sample from this location performed very poorly and detected only 11 taxa, potentially highlighting the additional habitat complexity at wider, downstream river locations and insufficient eDNA capture using composite sampling.

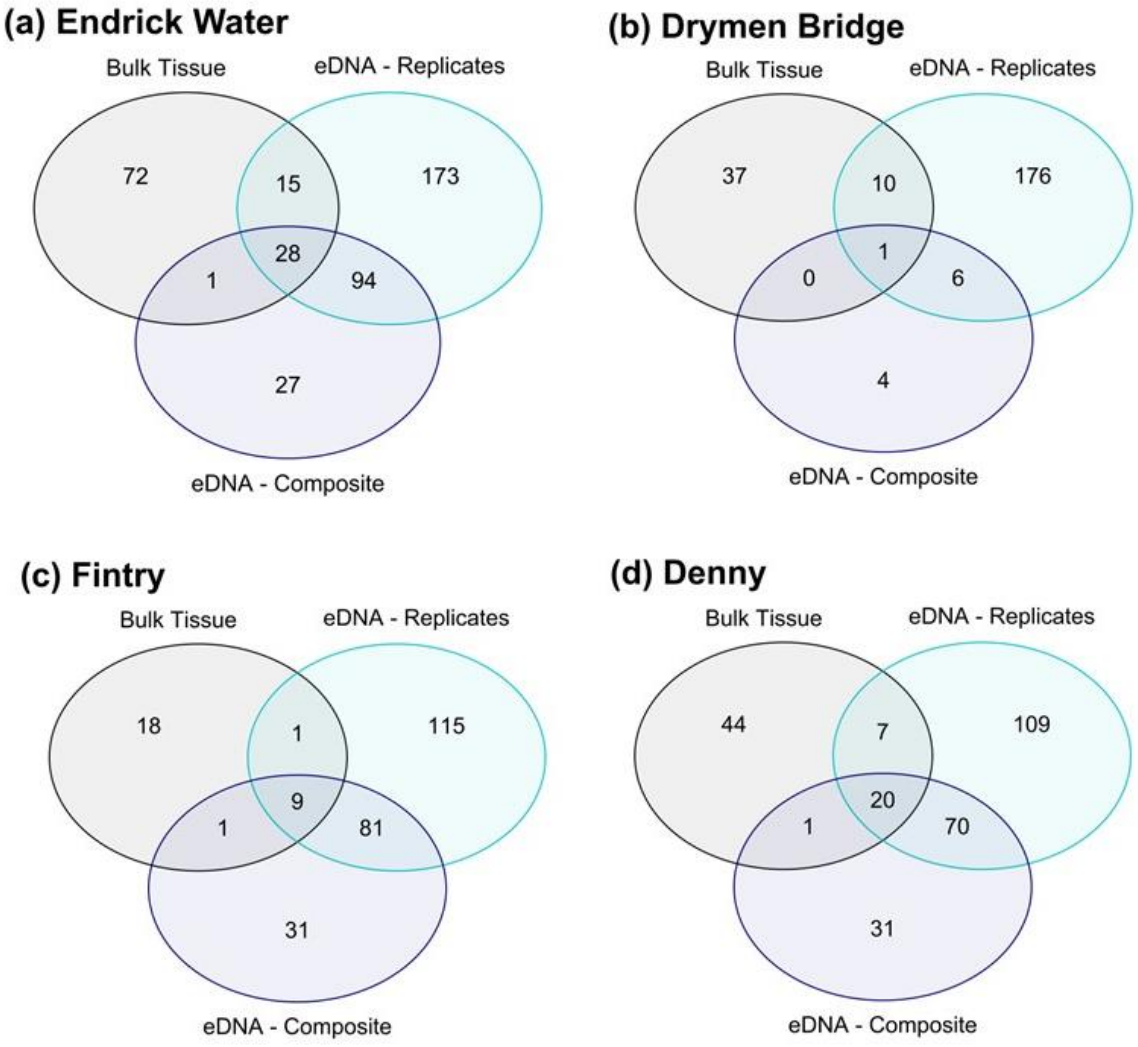

Figure 29. Venn diagrams summarising taxon overlap between bulk tissue DNA, replicate eDNA and composite eDNA samples from Endrick Water overall (a) and each sampling location along Endrick Water: (b) Drymen Bridge (downstream), (c) Fintry (mid-section), and (d) Denny (upstream).

Bulk tissue DNA samples detected less taxa (116 taxa) than replicate eDNA (310 taxa) and composite eDNA samples (150 taxa) overall. This also applied to individual sampling locations with the exception of Drymen Bridge where the bulk tissue DNA sample (48 taxa) outperformed the composite eDNA sample (11 taxa), but the replicate eDNA samples (193 taxa) combined vastly exceeded both other sample types. The proportion of sequencing output allocated to different families (Figure 30) and orders (Figure 31) for each sample from each sampling location is summarised below.

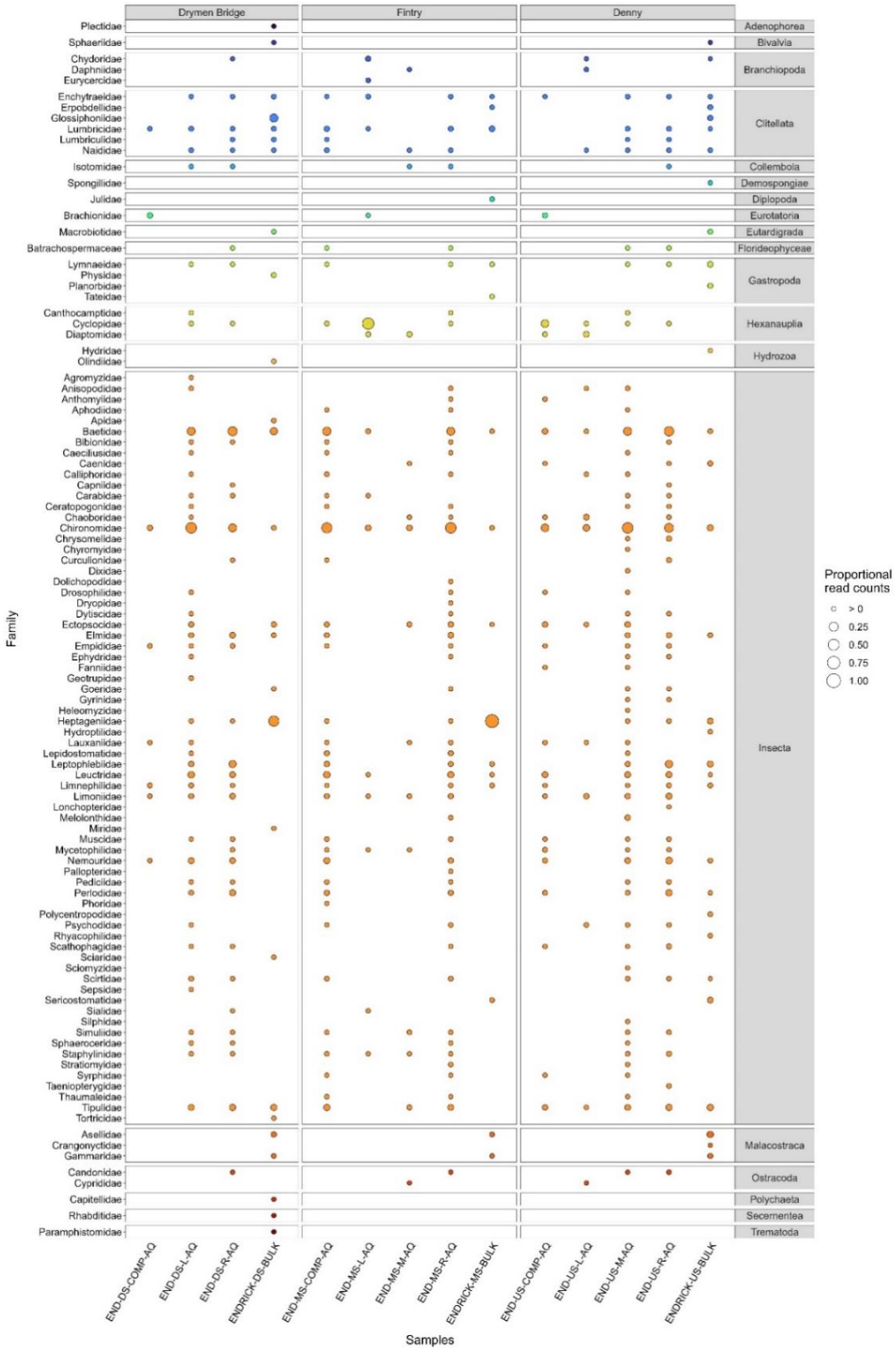

Figure 30. The proportion of the sequencing output allocated to the different families (rows) within each sample (columns). Each bubble per sample represents the proportion of DNA for each family for that sample. The size of the bubble is relative to the number of sequences from all families for that sample.

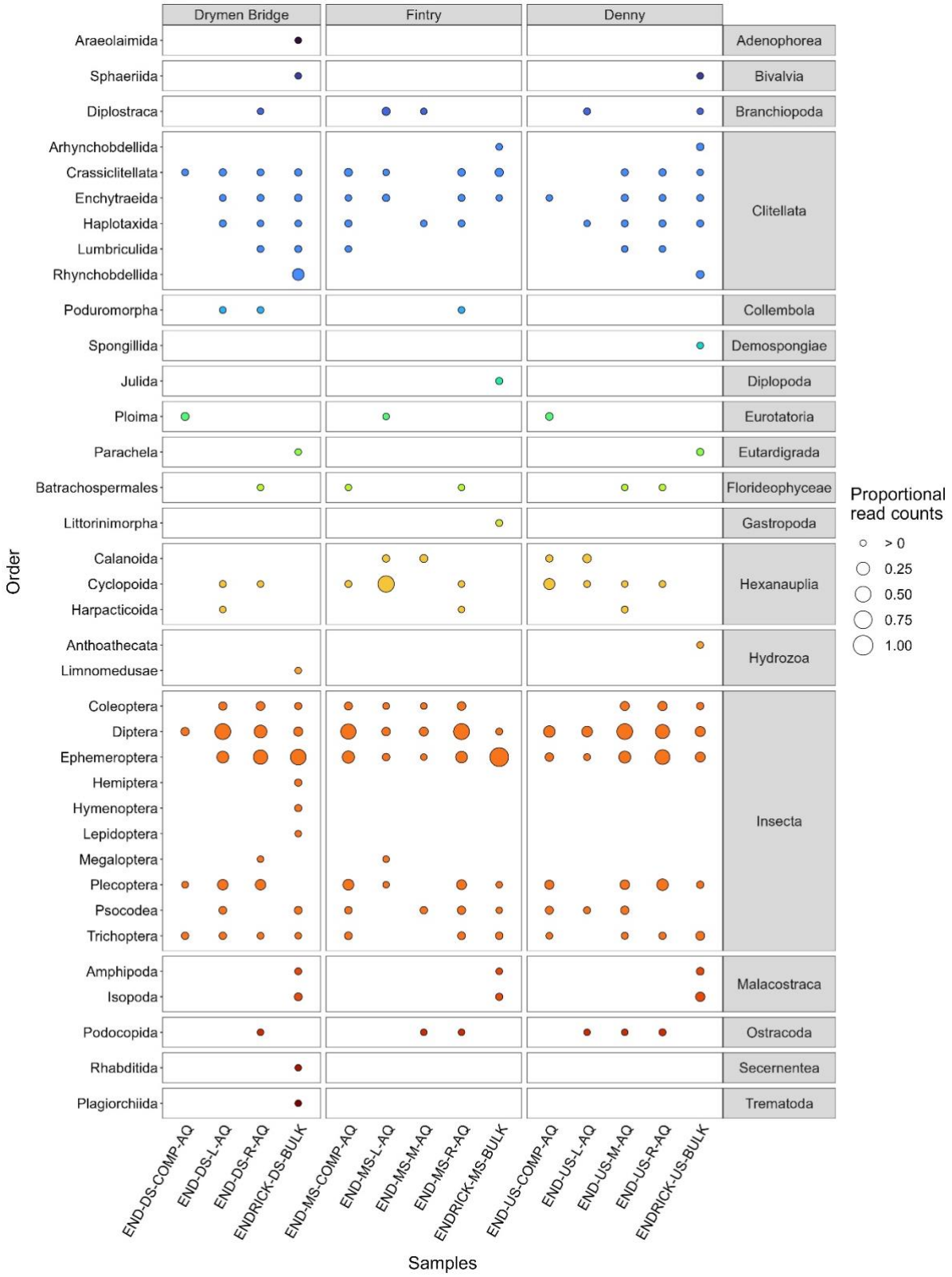

Figure 31. The proportion of the sequencing output allocated to the different orders (rows) within each sample (columns). Each bubble per sample represents the proportion of DNA for each order for that sample. The size of the bubble is relative to the number of sequences from all orders for that sample.

Box 1. Lake Classification – illustrative case study

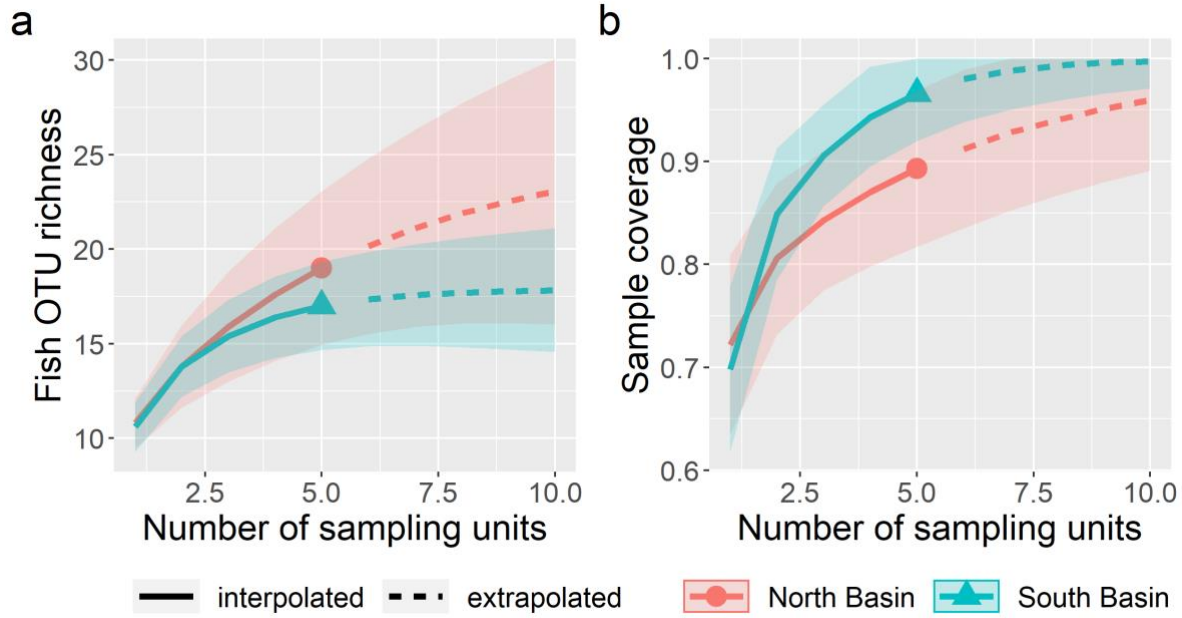

Figure 32. Accumulation of fish species (Actinopterygii, Cephalaspidomorphi) a) richness and b) sample coverage in the North and South Basins of Loch Lomond. Sample coverage is the estimated proportion of the whole community that belong to taxa in the sample.

Five indicator species used in the lake fish classification index were present in both basins, with preliminary mean EQR scores of 0.29 and 0.37 for the North and South Basins, respectively, both of which correspond to Moderate classification under the WFD. Asymptotic estimates of species richness for the above fish classes were 25.53 (s.e. 6.72) and 17.90 (s.e. 1.47) for the North and South Basins, respectively. Considering fish species only, both basins were close to the asymptotic estimates of species richness (Figure 32, Table 8). Conventional sampling methods over the last 50 years have documented a total of 19 species in Loch Lomond (Maitland, Adams, and Mitchell 2000).

Table 8. WFD scores, observed and estimated fish OTU richness for the North and South Basins of Loch Lomond.

Location	mean EQR (WPD score)	Observed richness	Estimated richness
North Basin	0.288 (Moderate)	19	25.53 ± 6.72
South Basin	0.367 (Moderate)	17	17.9 ± 1.47

Supplementary Information

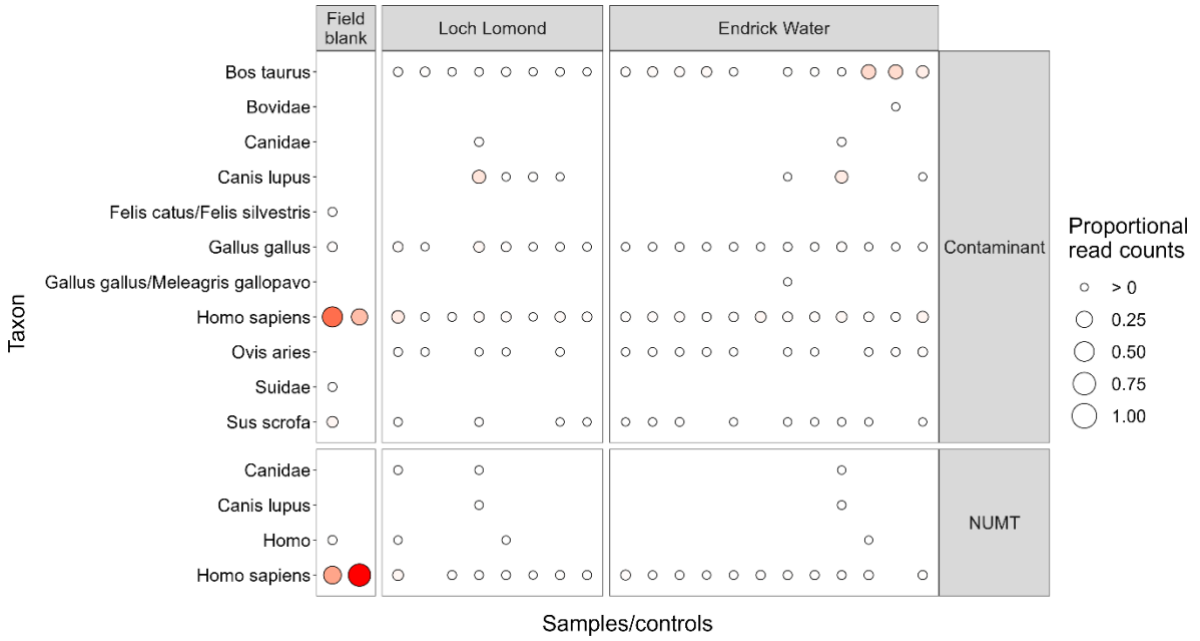

Figure 33. Bubble plot summarising vertebrate contaminants and NUMTs in water samples collected from Loch Lomond and Endrick Water.

Figure 34a. Bubble plot summarising vertebrate detection from water samples collected from North and South Basins of Loch Lomond.

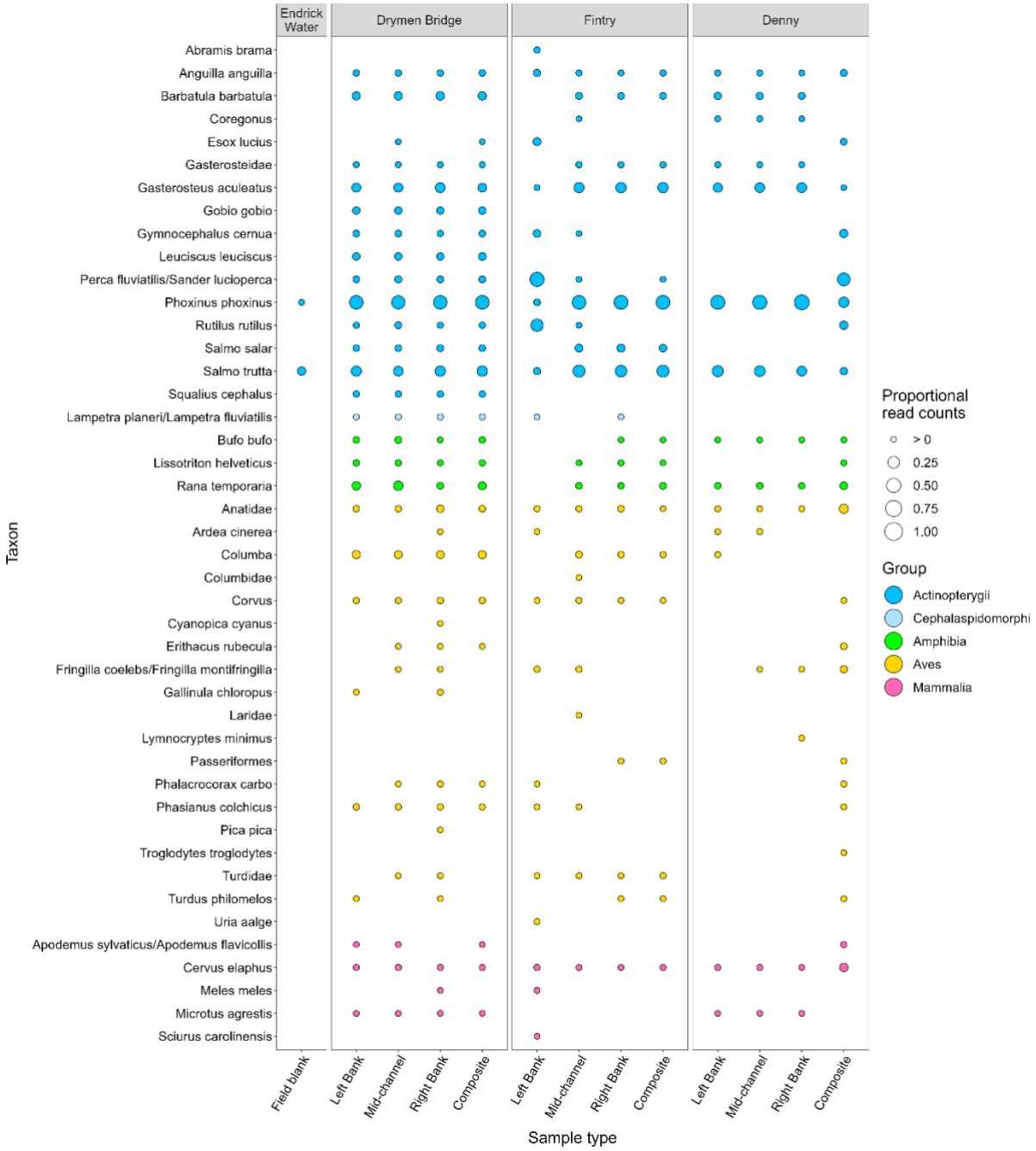

Figure 34b. Bubble plot summarising vertebrate detection from independent and composite water samples collected from upstream (Denny), mid-section (Fintry) and downstream (Drymen Bridge) locations along the Endrick Water.

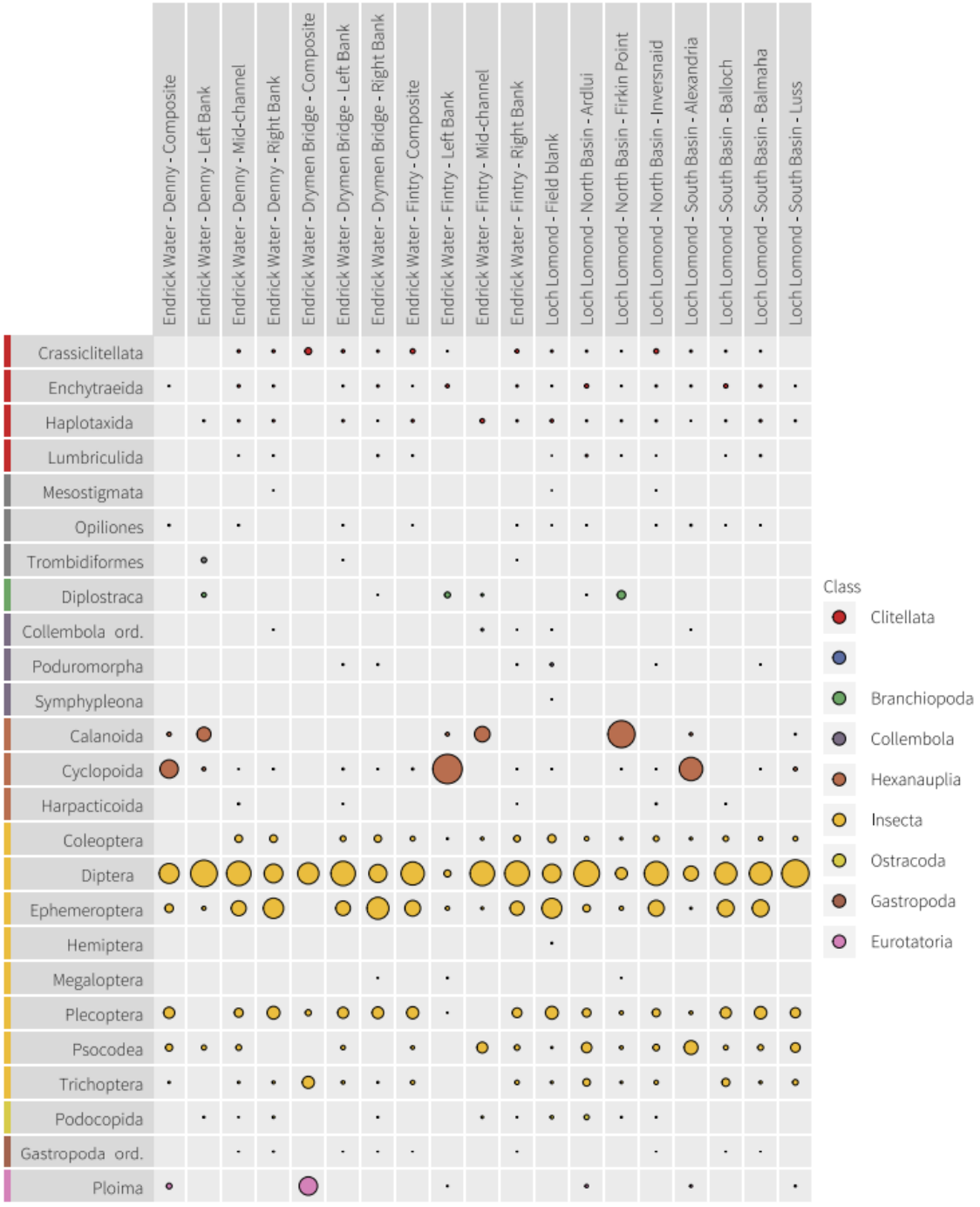

Figure 35. Bubble plot summarising invertebrate detection from water samples collected from the Endrick Water and Loch Lomond.

For Figures 34 and 35, the proportion of the sequencing output allocated to the different species (rows) within each sample (columns). Each bubble per sample represents the proportion of DNA for each species for that sample. The size of the bubble is relative to the number of sequences from all species for that sample.

Table 9. Observed, mean, overall and estimated OTU richness for vertebrates and invertebrates in the North and South Basins of Loch Lomond.

Assay	Sample location	Sample type (n)	OTU richness		
			mean (sd)	Observed overall	Estimated (LCL-UCL)
Vertebrates	North Basin	Multiple shoreline (5)	18.4 (6.1)	39	
	South Basin	Multiple shoreline (5)	22.2 (6.3)	52	
		Overall (10)	20.3 (6.2)	58	86.4 (67.5-143)
Invertebrates	North Basin	Multiple shoreline (4)	115.5 (86.9)	315	
	South Basin	Multiple shoreline (5)	86.8 (60.6)	276	
		Overall (9)	99.6 (70)	411	569.4 (519.3-642.7)

Table 10. Observed, mean and estimated OTU richness for vertebrates and invertebrates at three sampling locations along Endrick Water: Denny (upstream), Fintry (mid-section), Drymen Bridge (downstream).

Assay	Sample location	Sample type (n)	OTU richness		
			Mean (sd)	Observed	Estimated (LCL-UCL)
Vertebrates	Denny	Replicates (3)	13 (1)	17	
		Composite (1)		24	
	Fintry	Replicates (3)	20.3 (2.5)	35	
		Composite (1)		15	
	Drymen Bridge	Replicates (3)	29.7 (2.1)	36	
		Composite (1)		28	
		Overall - composite (3)	22.3 (6.7)	37	
		Overall - without composite (9)	21 (7.4)	46	51.6 (47.3-69.3)
Invertebrates	Denny	Replicates (3)	137.7 (91)	309	
		Composite (1)		58	
	Fintry	Replicates (3)	98 (106.6)	258	
		Composite (1)		144	
	Drymen Bridge	Replicates (3)	103.3 (90.6)	246	
		Composite (1)		11	
		Overall - composite (3)	71 (67.4)	180	
		Overall - without composite (9)	113 (85.5)	418	504.9 (474.8-550.9)

Table 11. Multivariate glm to show difference in OTU composition between the North and South Basins of Loch Lomond, and upstream, midsection and downstream locations along Endrick Water.

Taxonomic group	Waterbody	Coefficient	residual df	df	Deviance	p value
Vertebrates	Loch Lomond	Intercept	7			
		North/South Basin	6	1	47.86	0.588
	Endrick Water	Intercept	8			
		Upstream/Mid/Downstream	6	2	213.2	0.009
Invertebrates	Loch Lomond	Intercept	8			
		North/South Basin	7	1	532.9	0.341
	Endrick Water	Intercept	8			
		Upstream/Mid/Downstream	6	2	823.6	0.618
Bacteria	Loch Lomond	Intercept	9			
		North/South Basin	8	1	141674	0.001
	Endrick Water	Intercept	11			
		Upstream/Mid/Downstream	9	2	323795	0.001

4.3 Pilot results woodland

Introduction

The pilot sampling campaign took place in November 2021 and the main sampling campaign took place in August – September 2022. Within each habitat, the pilot study has investigated some key parameters of interest, such as the effects of subsampling or the starting sample type (e.g., water or sediment). The data generated in the pilot were interrogated to assess their usefulness for detecting key features or informing target indicators.

This report details the methods and results from the pilot phase of the project for the woodland sampling. These results have fed into the decision making to design the sampling and analysis approach of the P2 Sampling Plan.

Methods

Study sites

The RSPB Inversnaid Nature Reserve was chosen as the study site for the woodland pilot study. The key parameter of interest was to assess herbivore impacts on biodiversity and community composition of soil communities, by utilising existing fenced herbivore exclosure plots which have been in place within woodland areas at the RSPB Inversnaid Nature Reserve for ~20 years. Areas outside the exclosures are heavily browsed by deer and goat. The site was chosen through consultation with the Management Steering Group and Technical Steering Group.

Sampling was conducted at four locations within RSPB Inversnaid Nature Reserve (Figure 36) during 1-2 November 2021. Within the literature, there is no standardised soil sampling methodology relating to plot size, number of subsamples per plot, or sampling depth. We selected a plot size of 10 m x 10 m to align with the 10 m pixel resolution of SENTINEL-2 multispectral data. This gives the opportunity to compare soil DNA data with satellite derived indices in the next phase of this project.

To establish the level of variability within 10 m x 10 m plots, composite soil samples were collected in 2 m x 2 m subplots (three replicates within each 10 m x 10 m plot). A mixed sample was also prepared for each plot, combining equally sized subsamples from each of the homogenised subplot samples, to give an indication of how one composite sample consisting of soil from 15 cores across a 10 m x 10 m area compares with the three subplot replicates.

Most soil DNA studies sample to 5 cm depth, 10 cm depth, or keep the organic and mineral layers separated. As the depth of the organic layer is likely to differ within and among sites, a defined number of cores would result in different volumes of soil being collected for each composite sample which could affect the number of taxa contained within each sample. Keeping in mind that future soil sampling surveys would be deployed across different habitat types, to characterise the soil community at habitat scales we chose to collect composite samples consisting of a defined number of subsample cores per plot to 10 cm depth below the surface.

Sample collection

Two 10 m x 10 m plots were marked out at each location: one within the fenced exclosure, and one in the immediately adjacent unfenced woodland. Within each 10 m x 10 m plot composite samples were collected from three 2 m x 2 m subplots, totalling 24 composite samples. Each composite sample consisted of five soil cores (one core collected at the centre of the subplot and one core from each corner). Prior to coring, the surface litter layer was brushed aside to expose the fermentation layer. There were some subplots in wet patches with moss groundcover and at these locations the living moss was pushed aside, and the core was then collected from the underlying peaty soil. A sterile soil corer (a plastic syringe with the end cut off) was used to collect soil to 10 cm depth. The five cores were thoroughly mixed together in a labelled plastic snap-lock bag in the field to homogenise. To prevent cross contamination, fresh gloves and a new sterile corer were used for each composite sample. Samples were kept on ice in a cool bag in the field and while in transit, then transferred to a fridge upon arrival at the laboratory.

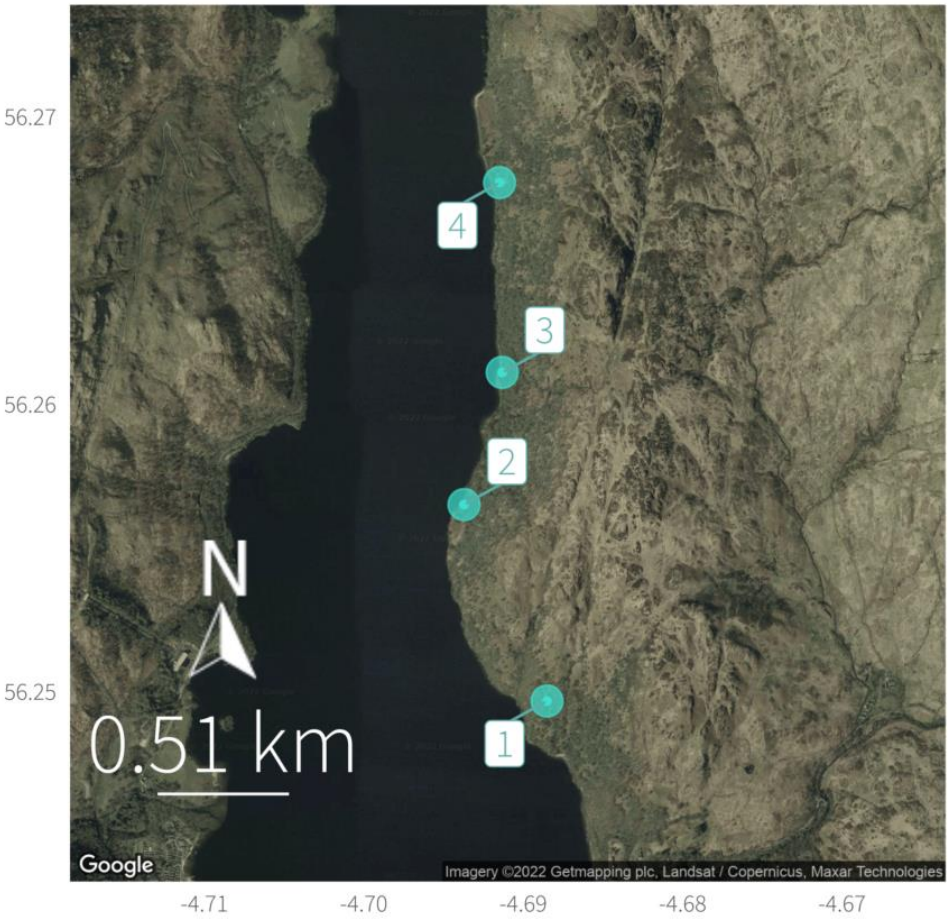

Figure 36. Locations of four paired plots (fenced and unfenced) within the RSPB Inversnaid Nature Reserve, Scotland.

Laboratory analysis

DNA extraction was identical to that for marine sediment.

Three replicate PCRs for each sample and extraction blank were amplified with modified 515F (Parada et al., 2016) and 806R primers (Apprill et al., 2015) targeting the V4 region of the 16S rRNA gene for bacteria, ITS3 and ITS4 primers targeting the internal transcribed spacer 2 (ITS2) region for fungi (White et al., 1990), and primers M620F and M1260R targeting the V4-V5 region of the 18S rRNA gene for fauna (Capra et al., 2016). Tails were added at the 5′ end to be complementary with Illumina Nextera index primers. All PCRs were performed in the presence of a negative control using only distilled water to detect cross contamination if present. A positive control samples (a sample known to amplify with those primers) was also run alongside each PCR. Success of the amplifications were confirmed by gel electrophoresis. No amplification was detected for negative controls and extraction blanks.

All PCRs were carried out in a total volume of 25 µL. The amplification mixture for 16S contained 1X DreamTaq Green PCR Master Mix (Thermo Scientific), 0.2 µM each primer, 0.25 mg/ml bovine serum albumin (Thermo Scientific) and 1µl template DNA, topped up with molecular grade water. The ITS2 amplification mixture was the same as the 16S mixture except 0.4 mg/ml of BSA was used. The 18S amplification mixture contained 1X Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Scientific), 0.4 µM of each primer, 0.4 mg/ml BSA and 1 µl DNA template, topped up with molecular grade water. The 16S PCR conditions followed (Caporaso et al., 2011) but using 30 cycles rather than 35°C. The ITS2 PCR conditions consisted of an initial denaturation at 95°C for 3 min followed by 35 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, and a final elongation step of 72°C for 10 min. The 18S PCR conditions consisted of an initial denaturation at 98°C for 30 sec followed by 35 cycles of 98°C for 20 sec, 65°C for 30 sec, 72°C for 30 sec, and a final elongation step of 72°C for 7 min.

Library preparation was identical to that in Section 4.1.

The pH (Fisherbrand™ accumet™ AB150 pH Benchtop Meter) and moisture content (%moisture:wer weight basis) of each soil sample was also measured. Soil %moisture:dry weight basis was later measured in the laboratory by drying the soil at 105°C for 72 hours in a Memmert Universal Oven UF110

Bioinformatics

Bioinformatic processing followed the same pipeline as for marine assays detailed in Section 4.1 with specific differences noted below.

For the fungal assay, dereplicated sequences were processed with ITSx (Bengtsson-Palme et al., 2013) to extract only fungal ITS2 sequences, removing the primers and any remaining ribosomal sequence.

The NCBI nucleotide (NCBI *nt*) database was queried for all three assays, with the bacterial and faunal datasets additionally queried against SILVA (Quast et al., 2012; Yilmaz et al., 2014) and the fungal dataset queried against UNITE (Nilsson et al., 2019).

For all soil datasets, assignments were made to the lowest possible taxonomic level where there was consistency in the matches, with minimum similarity thresholds of 98%, 95% and 92% for species-, genus- and higher-level assignments respectively.

All OTUs with species-level identifications were queried against the IUCN Red List (rredlist; Chamberlin 2018, https://cran.r-project.org/package=rredlist) to obtain global threat status. Finally, low abundance detections were omitted, with filter thresholds set at a percentage of the total reads per sample (bacterial: 0.05%; fungal: 0.025%; faunal: 0.05%). Results are presented for OTUs identified to target kingdom (bacterial: Bacteria; fungal: Fungi; faunal: Animalia) or below.

Data analysis

Data were analysed using the statistical software R v4.1.0 with RStudio v1.4.1717. The package tidyverse v1.3.1 was used for data manipulation and formatting, including taxon richness for each sample and mean taxon richness between sampling locations; ggplot2 v3.3.5 was used for graphics. Community similarity was visualised using Non-Metric Multi-Dimensional Scaling (NMDS) plots created with the package vegan v2.5-7.

Sequencing depths were considered sufficient for all samples to capture most of the taxon diversity. Species accumulation and sample coverage curves were calculated for incidence data (presence/absence) across samples in each group (fenced vs unfenced plots) for each marker using the iNEXT package in R (Hsieh et al., 2016). To evaluate whether soil community composition differed amongst groups (fenced or unfenced plots) for each marker, a model-based analysis of multivariate data based on binomial generalised linear models was used in R package mvabund (Wang et al., 2012).

Results

Study area environmental characteristics

Although pH was higher in unfenced treatments at two plots, differences were not significant (ANOVA: $p = 0.185$, $df = 1, 27$, $F = 1.85$), whereas it was significantly higher between plot 4 and the other three other plots (Anova: $p < 0.05$, $df = 3,27$, $F = 66.1$ and Tukey HSD; Figure 37). With regards to moisture, although there were no significant differences between treatments (ANOVA: $p = 0.127$, $df = 1,27$, $F = 2.48$), among plot differences showed a different pattern to pH with significant differences between all plots except plots 1 and 2, and 3 and 4 (ANOVA: $p < 0.05$, $df = 3,27$, $F = 27.66$ and Tukey HSD: Figure 37).

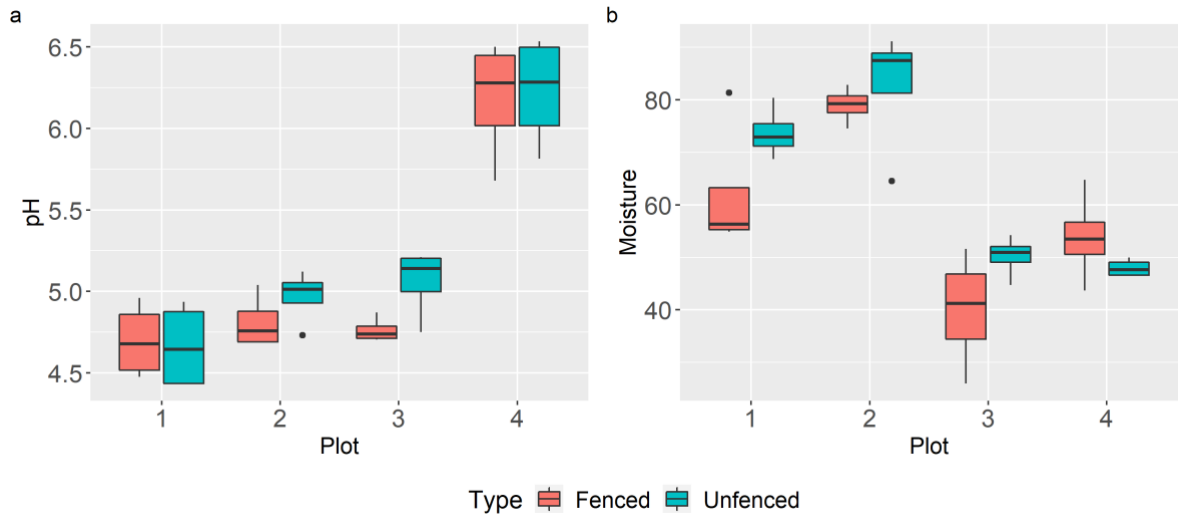

Figure 37: Soil pH and moisture levels (%moisture:dry weight basis) across woodland plots and fenced and unfenced treatments at Inversnaid RSPB reserve on Loch Lomond.

Sequencing data summary

The final dataset contained a total of 2,288 OTUs across the 32 soil DNA samples: 1,148 bacteria, 936 fungi, and 204 fauna (Table 12). Taxon-by-sample tables of the samples are attached to this report (Appendix A, Appendix B, and Appendix C) and sample details are provided in Appendix D. More fungal OTUs were identified at the species level compared to bacteria. This reflects differences in availability of reference sequences for different organisms within the reference databases and a higher proportion of assignment conflicts (100% matches to multiple species) in bacteria. Of the taxa that were identified to species level, the only species that were included on the IUCN Red List were classified as Least Concern. Extraction and PCR blanks did not show evidence of amplification and were not sequenced.

Table 12: Summary of the number of OTUs detected and the percentage of OTUs successfully classified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	1,148	78.1%	63.1%	44.8%	32.0%	13.7%	2.2%
Fungi	936	99.4%	91.0%	79.9%	58.5%	34.3%	15.9%
Fauna	204	98.0%	74.5%	86.3%	68.1%	31.9%	3.9%

In the bacterial dataset, OTUs were detected across 21 different phyla within the kingdom Bacteria. The average bacterial taxon richness per sample was 323.5 and ranged from 224 to 392. The bacterial phylum with the highest richness of OTUs was Proteobacteria. The bacterial OTU with the most reads was from the family Acidobacteriaceae. This OTU was detected in

32/32 samples. 25 bacterial OTUs were detected in every sample. There were 213 OTUs (18.6%) that were only detected in one sample each.

In the fungal dataset, OTUs were detected across 5 different phyla within the kingdom Fungi. The average fungal taxon richness per sample was 148.4 and ranged from 90 to 233. The fungal phylum with the highest richness of OTUs was Ascomycota. The fungal OTU with the most reads was the species *Mortierella pseudozygospora*. This OTU was detected in 31/32 of the samples. Nine fungal OTUs were detected in every sample. There were 267 OTUs (28.5%) that were only detected in one sample each.

In the faunal dataset, OTUs were detected across 10 different phyla within the kingdom Animalia. The average faunal taxon richness per sample was 31.7 and ranged from 3 to 67. The faunal phylum with the highest richness of OTUs was Nematoda. The OTU with the highest proportion of reads was from the family Enchytraeidae, a group of annelids. This OTU was detected in 32/32 of the samples and was the only faunal OTU that was detected in every sample. There were 63 OTUs (30.9%) that were only detected in one sample each.

Comparison between fenced and unfenced plots

Sample-level and plot-level taxon richness did not show a consistent trend between fenced and unfenced plots in any of the taxonomic groups (Figure 38). Plot-level faunal taxon richness was noticeably higher in fenced compared to unfenced at locations 2 and 3. When the cumulative richness was summed across all samples within each treatment group, bacterial and fungal treatment-level richness was higher in unfenced plots, but faunal richness was higher in fenced plots (Figure 39). The sample coverage curves indicate that most of the soil taxa were likely detected by this survey. Estimated sample coverage was highest in bacteria at 94% and 95% for fenced and unfenced treatments, respectively (Figure 37). This is also reflected in the smallest relative differences between observed and estimated richness in bacteria compared to other groups (Table 13). Soil fauna showed the largest differences between estimated and observed OTU richness (Table 13, Figure 39).

Multivariate community analysis showed significant differences in community composition among locations for all taxonomic groups but no significant differences between fenced and unfenced plots (Figure 40, Table 14). The ordination plots show that location 4 had the most soil distinct community composition.

Comparison between mixed core samples and individual replicates

For the three assays (fungi, bacteria and fauna) and across all four plots, the mixed samples showed consistently fewer observed OTU richness than the overall richness from the three replicate samples combined except for one case, fauna of the fenced sample in plot 1 where the mixed sample showed a higher observed richness (Figure 38). In several plots the mixed sample had higher richness than each of the replicate samples, but this was not consistent across all plots. The mixed samples were generally similar in species composition to the replicate samples (Figure 40).

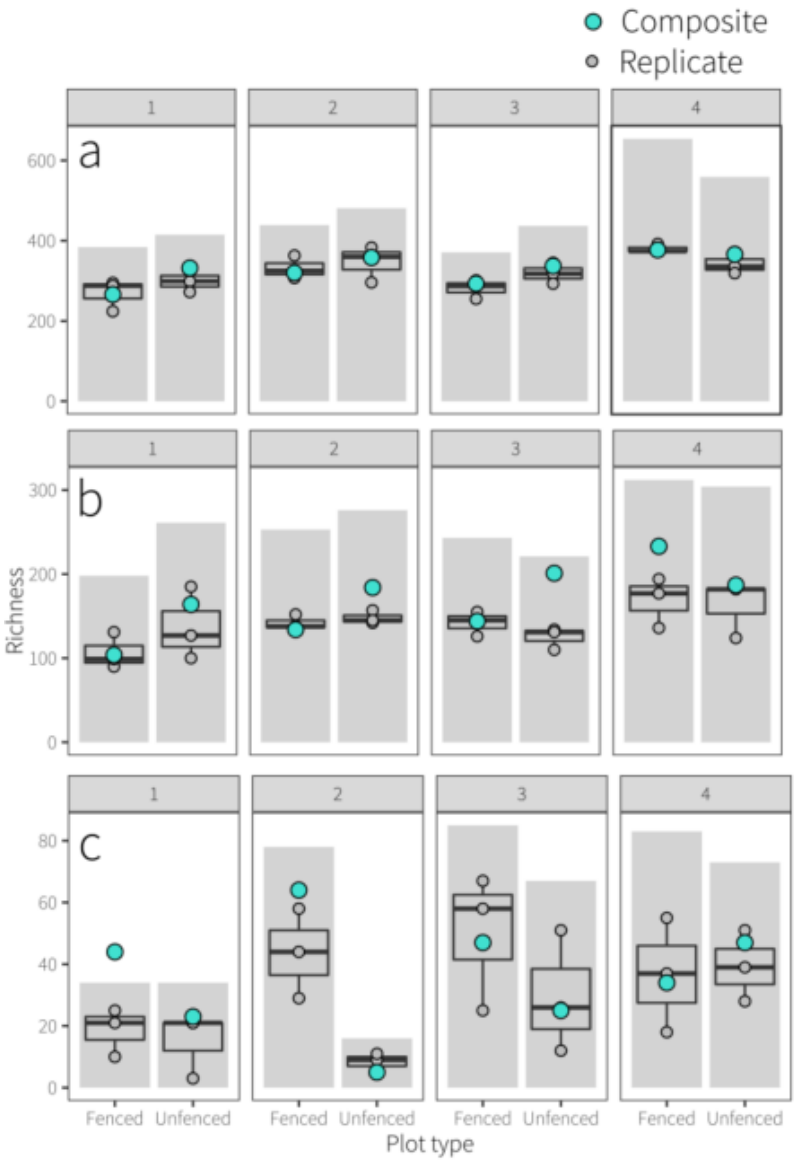

Figure 38: Taxon richness (number of OTUs) for soil (a) bacterial, (b) fungal, and (c) faunal communities at each location within fenced and unfenced plots. The bar shows plot-level taxon richness (cumulative richness for each plot). The boxplot shows sample-level richness, with the box depicting the median between the upper/lower quantiles, the whiskers indicating minimum and maximum values, and dots showing richness values of each sample. Any samples beyond the whiskers are considered as outliers which are 1.5x the interquartile range away from the upper or lower quartile. Mixed samples are denoted by green points.

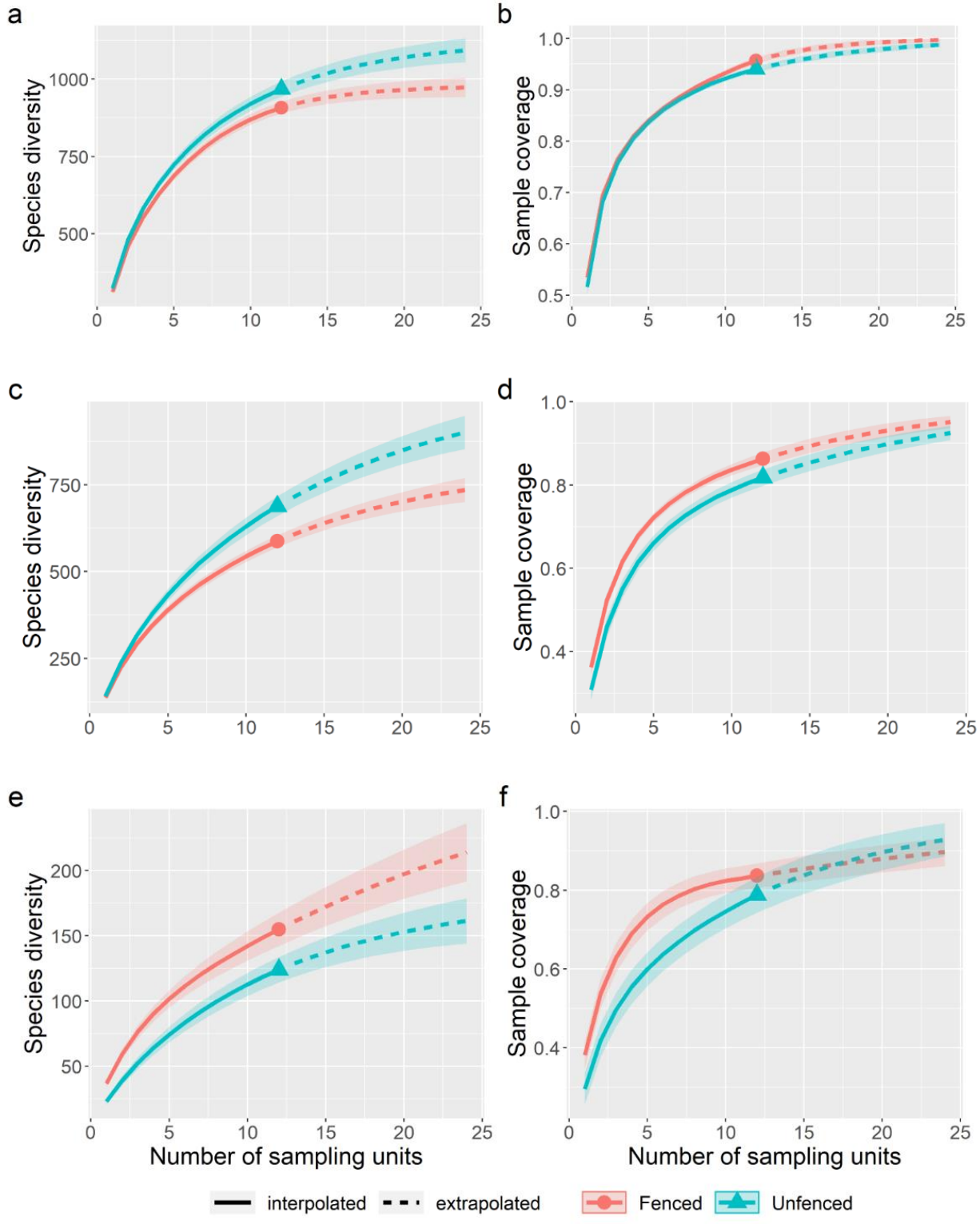

Figure 39: Accumulation curves for taxon richness (left column) and sample coverage (right) by number of sampling units for soil bacteria (a, b), fungi (c, d), and fauna (e, f) across all woodland samples taken in fenced and unfenced plots.

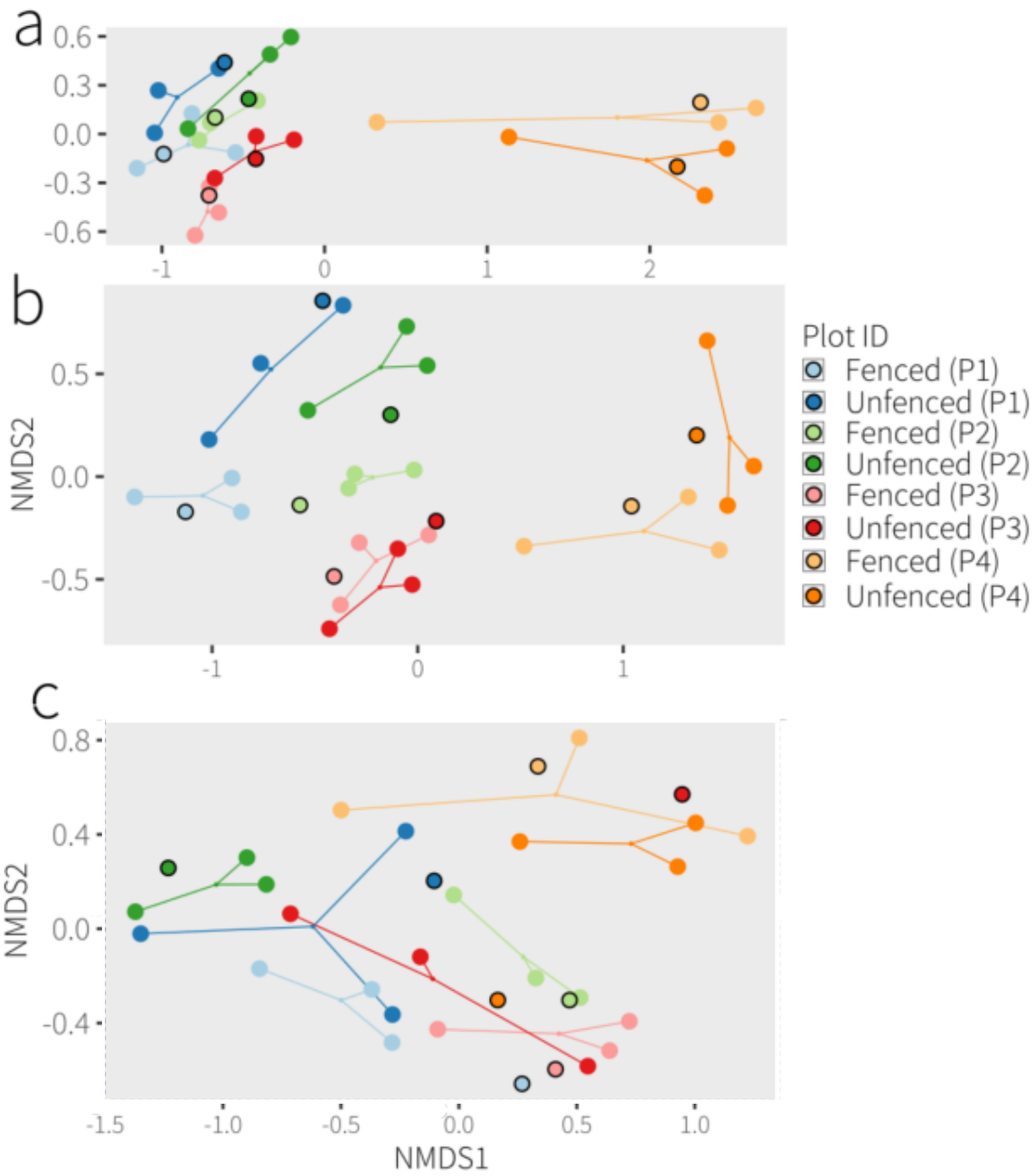

Figure 40: NMDS ordination plots based on Jaccard similarity index for soil (a) bacteria, (b) fungi, and (c) faunal community taxonomic compositions. Points are coloured by plot location with light colour denoting fenced plots and darker colour denoting unfenced plots. Replicate sub-plot samples are linked with solid lines and mixed samples are shown with black outlines

Limitations

Methodologies have been chosen based on the state of the art, but these choices inherently introduce specific limitations and biases. For each of the target groups we have chosen primers that in our experience capture their targets well. Each of these primer sets will inherently miss taxa and this will be a systematic error. Unfortunately, there is no one primer set that captures all the diversity, and the diversity present in soil makes it impossible to choose one primer set that balances specificity and resolution. This is only an issue if there is a particular target taxon (i.e. an indicator species) of interest.

Assigning taxonomic identities to the sequences is only possible through their comparison to reference databases, which are incomplete. This is not an issue if a taxonomy free approach is adopted – i.e. tracking changes over time by comparing datasets (as is advocated here), but it is a bigger concern if indicator species or functional groups (based on taxonomy) are required. It should be noted that multiple OTUs can be identified as belonging to the same species, which is most likely attributed to PCR or sequencing artefacts but potentially intraspecific genomic variation or cryptic diversity. Also, it is possible for closely related species to have identical sequences in the targeted gene region and if the species present at your site is not in the database it could be identified as a different closely related species.

What is happening among the communities (i.e. functioning) may be driven more so by the dominant taxa as opposed to the breadth of diversity. The abundance of taxa cannot be directly inferred from the number of sequence reads. While the number of sequence reads is a consequence of abundance, it is also impacted by biomass, body type, activity, surface area, condition, primer bias, and species-specific variation in the genome.

Table 13: Estimated, observed, and overall OTU richness across taxonomic groups and fenced/unfenced plots in Inversnaid RSPB reserve. Asymptotic estimated richness is given for each treatment by plots (n = 12 samples), with standard error (s.e.) lower (LCL) and upper (UCL) 95% confidence limits.

Taxonomi Group	Plot type	Number (type) of samples	Observed	Estimated	s.e.	LCL	UCL
Bacteria	Fenced	12 (single)	908	978.57	13.71	956.39	1010.92
		1 (mixed)	687				
		13 (total)	925				
	Unfenced	12 (single)	968	1126.55	25.73	1083.60	1185.47
		1 (mixed)	706				
		13 (total)	988				
		Overall	1128				
Fungi	Fenced	12 (single)	587	816.18	38.77	751.89	905.55
		1 (mixed)	399				
		13 (total)	616				
	Unfenced	12 (single)	688	1048.46	53.32	958.16	1168.92
		1 (mixed)	500				
		13 (total)	754				
		Overall	949				
Soil fauna	Fenced	12 (single)	155	316.13	56.29	237.86	468.35
		1 (mixed)	108				
		13 (total)	171				
	Unfenced	12 (single)	124	180.89	18.90	154.17	231.27
		1 (mixed)	72				
		13 (total)	134				
		Overall	201				

Table 14: Difference in community composition between fenced and unfenced treatments and across locations at for soil bacteria, fungi, and fauna in woodland at Inversnaid RSPB reserve.

Taxonomic group	Coefficient	residual df	df	Deviance	p value
Bacteria	Fenced/Unfenced	22	1	1160	0.48
	Location	19	3	10966	0.01
Fungi	Fenced/Unfenced	22	1	1402	0.07
	Location	19	3	6029	0.01
Fauna	Fenced/Unfenced	22	1	283.6	0.10
	Location	19	3	982.6	0.01

4.4 Pilot results peatland

Introduction

The pilot sampling campaign took place in August 2021 and the main sampling campaign took place in August – September 2022. Within each habitat, the pilot study has investigated some key parameters of interest, such as the effects of subsampling or the starting sample type (e.g., water or sediment). The data generated in the pilot were interrogated to assess their usefulness for detecting key features or informing target indicators.

This report details the methods and results from the pilot phase of the project for the peatland sampling. These results have fed into the decision making to design the sampling and analysis approach of the P2 Sampling Plan.

Methods

Study sites

Two upland blanket bog sites within Loch Lomond and The Trossachs National Park where restoration works have been undertaken as part of the Peatland ACTION project were chosen as the study sites for the peatland pilot study. The key parameter of interest was to assess degraded and restored areas in terms of biodiversity and community composition of soil communities. The sites were chosen through consultation with the Management Steering Group and Technical Steering Group. One site is located in the Glen Finglas Estate managed by the Woodland Trust, and the second site is located in the Auchlyne Estate which is privately owned (Figure 41). Sampling locations were determined on site in consultation with a Peatland ACTION representative and site managers.

The aim of the peatland DNA pilot project was to assess how soil biodiversity and community composition differ between:

- Degraded peat: areas of peat in the vicinity of grips (drainage channels) where the water table is low because it is drained away from the site
- Restored peat: Areas of peat that were formerly degraded but where restoration works have been undertaken to block the grips and raise the water table level back to more natural levels

Within the literature there is large variability in soil sampling designs for peatlands, particularly in regard to sampling depth. For this study we collected peat cores from 0-10 cm for analysis. While peatland soil DNA studies in the scientific literature often go deeper than 10 cm, they usually separate soil from the cores into different depth categories. Sampling multiple depth categories would reduce the number of locations that could be sampled. Surface soils will most likely be linked to vegetation indices that can be calculated from remote sensing data and the surface soil is where biological activity will be the most active due to higher levels of oxygen availability. It is also where fluctuations in water content will be at their most extreme and so will give a good indication of effectiveness of hydrological restoration works.

Samples were collected on 22-23 August 2021. Subsample cores were collected along 25 m transects running parallel to blocked and unblocked grips, approximately 30 cm from the edge of the grip. Two subsample cores were collected every 5 m and all 12 subsamples along a transect were combined to create one composite sample. The transects were laid out in this way as the peat immediately adjacent to the channel would be expected to see the strongest change in soil moisture once a grip is blocked and the water starts to flow over the edge of the channel.

Samples were collected on 22-23 August 2021 at two upland blanket bog sites within Loch Lomond and The Trossachs National Park where restoration works have been undertaken as part of the Peatland ACTION project. One site was located in the Glen Finglas Estate managed by the Woodland Trust, and the second site was located in the Auchlyne Estate which is privately owned (Figure 41). Sampling locations were determined on site in consultation with a Peatland ACTION representative and site managers.

The plan was to collect samples along three transects of each peat category (as defined above) at each site. Time permitted the collection of a fourth sample in each category at Glen Finglas. This resulted in a total of six samples being collected at Auchlyne Estate and eight samples collected at Glen Finglas. Subsample cores were collected into a large plastic bag. The cores were mixed together in the bag in the field and then a subsample of the homogenised soil (~50 g) was transferred to a smaller, labelled snaplock bag and the rest was discarded. Samples were kept on ice in a cool bag in the field and while in transit, then transferred to a fridge upon arrival at the laboratory.

On site, we found plastic syringe corers were not suitable for sampling soil that is highly water saturated. Therefore, a metal soil augur was used instead to collect the cores. To prevent cross-contamination, after each transect the soil auger was wiped with Chemgene disinfectant wipes and sampler's gloves were changed.

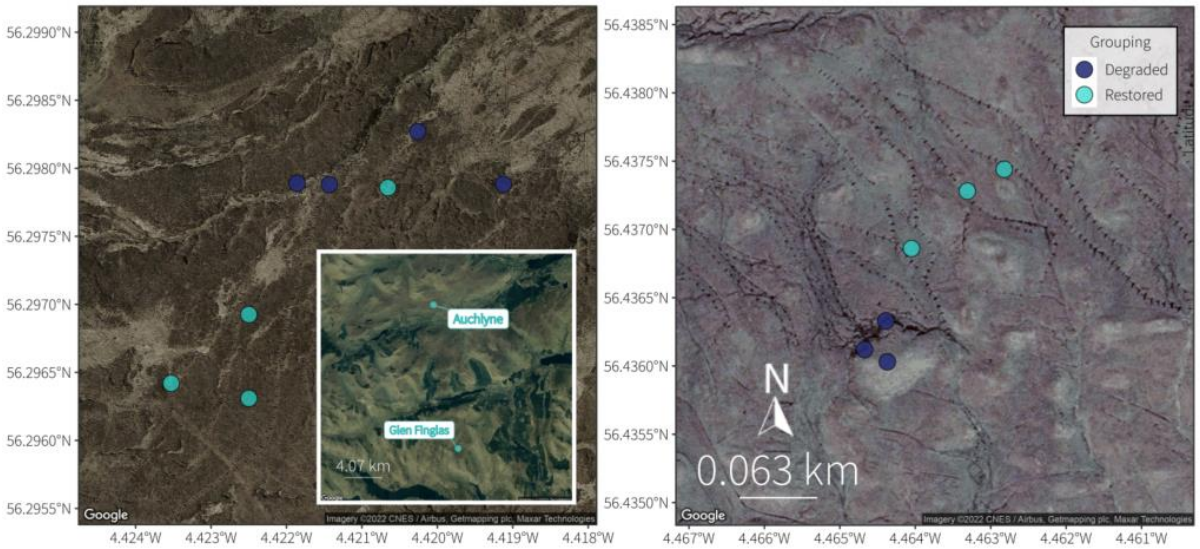

Figure 41: Soil sample transect locations in degraded and restored peatland areas across two sites: Glen Finglas (left) and Auchlyne Estate (right) within Loch Lomond and the Trossachs National Park, Scotland.

Laboratory analysis
All laboratory analysis was identical to that carried out for woodland.

Bioinformatics
All bioinformatic analysis was identical to that carried out for woodland.

Data analysis
Data were analysed using the statistical software R v4.1.0 with RStudio v1.4.1717. The package tidyverse v1.3.1 was used for data manipulation and formatting, including taxon richness for each sample and mean taxon richness between sampling locations; ggplot2 v3.3.5 was used for graphics. Community similarity was visualised using Non-Metric Multi-Dimensionsal Scaling (NMDS) plots created with the package vegan v2.5-7.

Sequencing depths were considered sufficient for all samples to capture most of the taxon diversity. Species accumulation and sample coverage curves were calculated for incidence data (presence/absence) across samples in each group (degraded vs restored peatland) for each marker using the iNEXT package in R (Hsieh et al., 2016). To evaluate whether soil community composition differed amongst groups, a model-based analysis of multivariate data based on binomial generalised linear models was used in R package mvabund (Wang et al., 2012).

Results
Study area environmental characteristics
Moisture and pH were broadly similar across sites and treatments although different patterns were observed (Figure 42). However, small sample size does not allow for detailed analysis. Moisture appears higher in restored plots compared to degraded.

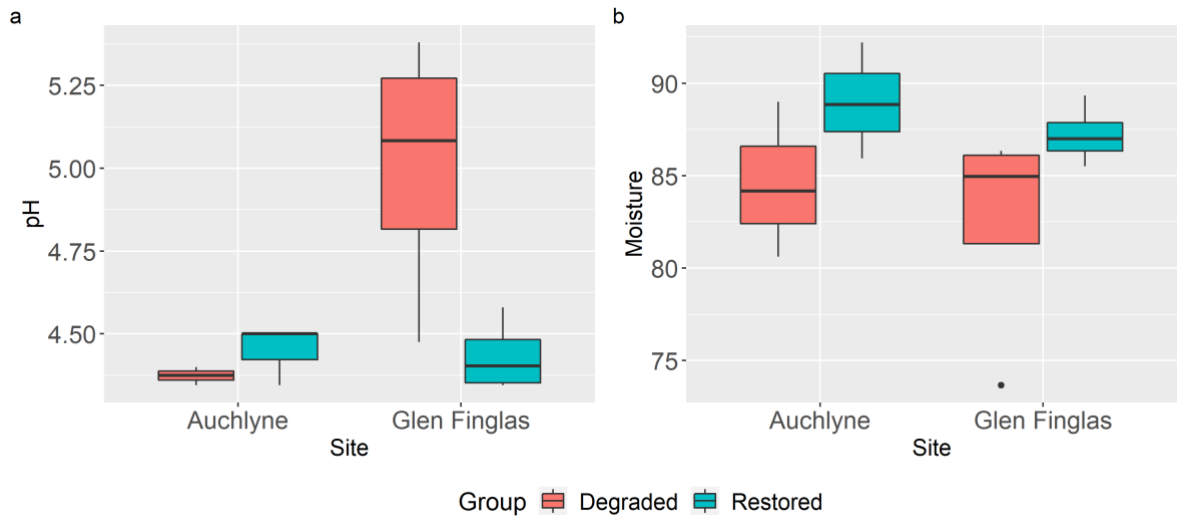

Figure 42: Soil pH and %moisture:dry wight levels across peatland transects in restored and degraded areas at Auchlyne and Glen Finglas Estates, Scotland

Sequencing data summary

The final dataset contained a total of 888 OTUs across the 14 soil DNA samples: 422 bacteria, 382 fungi, and 84 fauna (Table 15). Taxon-by-sample tables of the samples are attached to this report (Appendix A, Appendix B, and Appendix C) and sample details are provided in Appendix D. More fungal and faunal OTUs were identified at the species level compared to bacteria. This reflects differences in availability of reference sequences for different organisms within the reference databases and a higher proportion of assignment conflicts (100% matches to multiple species) in bacteria. None of the taxa that were identified to species level were listed on the IUCN Red List. Extraction and PCR blanks did not show evidence of amplification and were not sequenced.

Table 15: Summary of the number of OTUs detected and the percentage of OTUs successfully classified at each taxonomic level for each target

Target	Number of OTUs	Phylum	Class	Order	Family	Genus	Species
Bacteria	422	71.6%	59.2%	39.6%	28.9%	10.0%	1.7%
Fungi	382	99.5%	89.8%	77.5%	42.9%	22.3%	9.2%
Fauna	84	96.4%	75.0%	75.0%	72.6%	32.1%	1.2%

In the bacterial dataset, OTUs were detected across 19 different phyla within the kingdom Bacteria. The average bacterial taxon richness per sample was 190.4 and ranged from 131 to 247. The bacterial phylum with the highest richness of OTUs was Proteobacteria. The bacterial OTU with the most reads was only identified to Kingdom level. This OTU was detected in 14/14 samples. 58 bacterial OTUs were detected in every sample. There were 76 OTUs (18%) that were only detected in one sample each.

In the fungal dataset, OTUs were detected across five different phyla within the kingdom Fungi. The average fungal taxon richness per sample was 79.2 and ranged from 47 to 145. The fungal phylum with the highest richness of OTUs was Ascomycota. The fungal OTU with the most reads was only assigned to the Phylum level (Ascomycota). This OTU was detected in 13/14 of the samples. Two fungal OTUs were detected in every sample. There were 202 OTUs (52.9%) that were only detected in one sample each.

In the faunal dataset, OTUs were detected across seven different phyla within the kingdom Animalia. The average faunal taxon richness per sample was 20.9 and ranged from 8 to 45. The faunal phyla with the highest richness of OTUs were Arthropoda and Nematoda (18 OTUs each). The OTU with the highest proportion of reads was from the family Enchytraeidae, a group of annelids. This OTU was detected in 14/14 of the samples. Two faunal OTUs were

detected in every sample. There were 33 OTUs (39.3%) that were only detected in one sample each.

Comparison between degraded and restored plots

Sample-level fungal richness was higher in degraded compared to restored at both sites, but trends in bacterial and faunal richness between degraded and restored areas varied between sites (Figure 43). If we consider cumulative taxon richness across samples within each habitat condition, cumulative richness of bacteria and fungi were higher in the degraded areas compared to the restored, but soil faunal cumulative richness was higher in the restored areas (Figure 44). The sample coverage curves indicate that most of the soil taxa were likely detected by this survey. Estimated sample coverage was highest in bacteria at 96% for both degraded and restored areas, and lowest for fungi at 75% and 91% for degraded and restored areas, respectively (Figure 44).

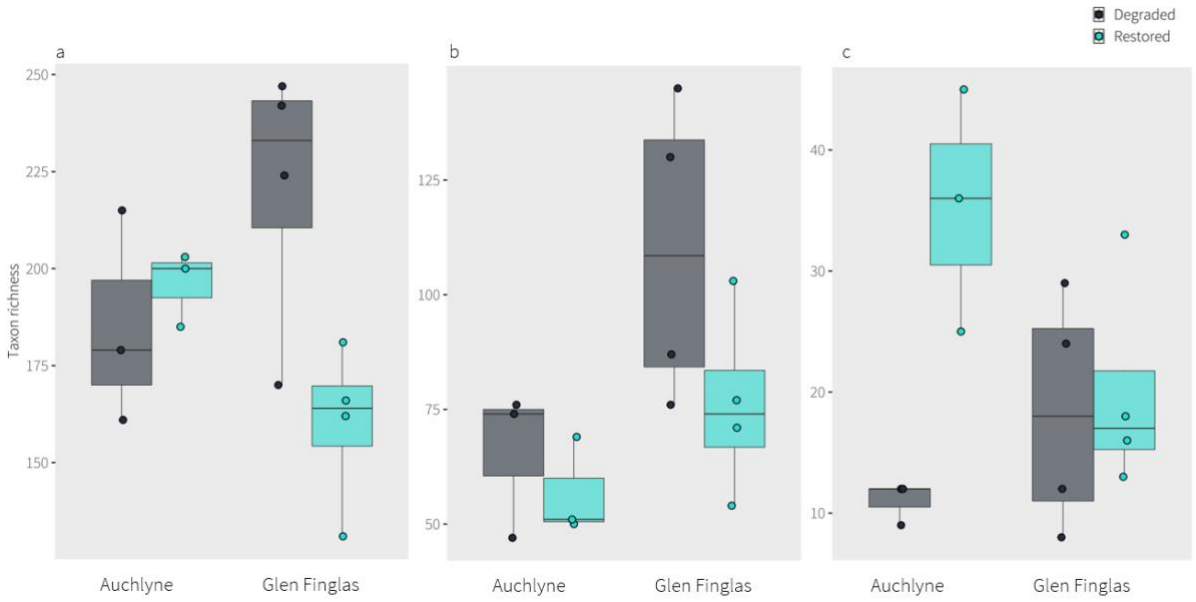

Figure 43: Taxon richness (number of OTUs) of soil a) bacteria, b) fungi, and c) fauna for degraded and restored areas at Auchlyne and Glen Finglas Estates. The boxplot shows sample-level richness, with the box depicting the median between the upper/lower quantiles, the whiskers indicating minimum and maximum values, and dots showing richness values of each sample. Any samples beyond the whiskers are considered as outliers which are 1.5x the interquartile-range away from the upper or lower quartile.

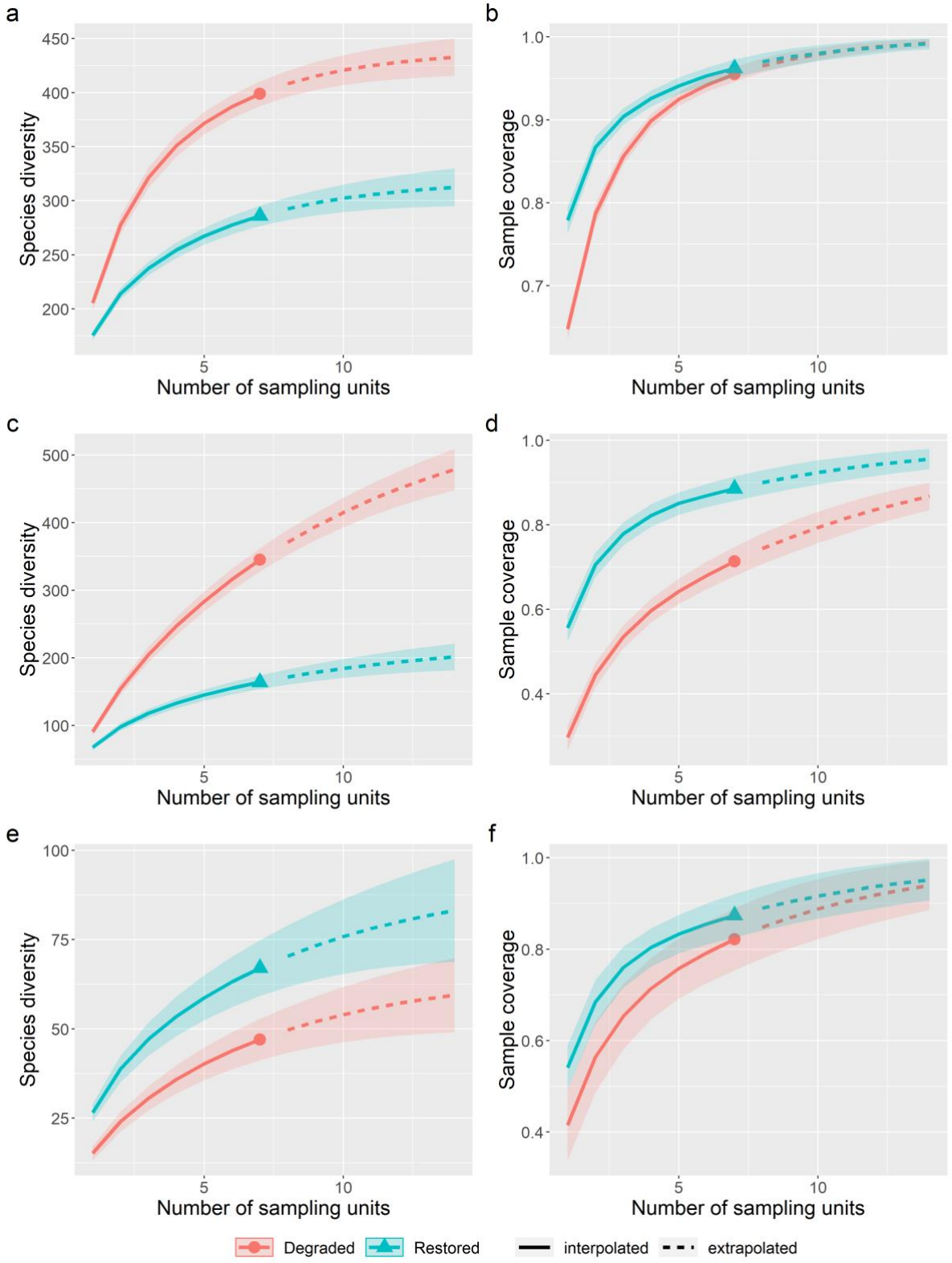

Figure 44: Accumulation curves for taxon richness (left column) and sample coverage (right) by number of sampling units for soil bacteria (a, b), fungi (c, d), and fauna (e, f) across all peatland plots (n = 7) for degraded and restored treatments Auchlyne and Glen Finglas estates.

Soil community composition significantly differed between degraded and restored areas for fauna (Deviance = 195.4, df = 1, p = 0.014) and bacteria (Deviance = 821.3, df = 1, p = 0.032, but

not for fungi (Deviance = 628, df = 1, p = 0.099) although sample sizes were insufficient to include site effects or environmental characteristics in the analysis. Moisture, but not pH, appears to influence community composition from preliminary analyses. The ordination plots indicate that site effects appear to be of importance, as the degraded and restored samples at each site cluster separately from each other, but when considered across both sites the restored peatland points form a cluster between degraded samples of Auchlyne and Glen Finglas (Figure 45).

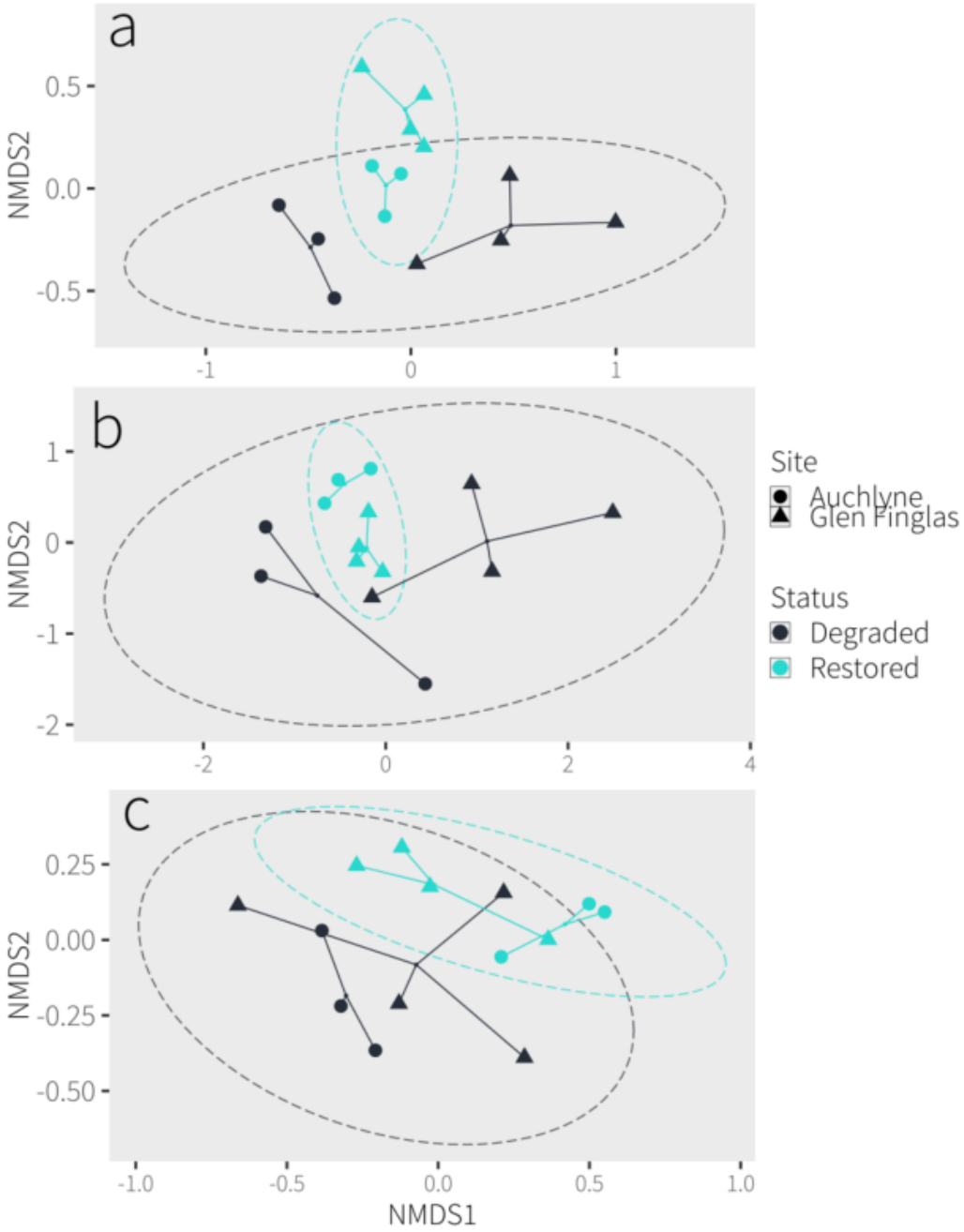

Figure 45: NMDS ordination plots based on Jaccard similarity index for soil (a) bacteria, (b) fungi, and (c) faunal community taxonomic compositions. Points are coloured by condition, shape indicates site. 95% confidence intervals for each condition are indicated by dashed ellipses.

Limitations

Methodologies have been chosen based on the state of the art, but these choices inherently introduce specific limitations and biases. For each of the target groups we have chosen primers that in our experience capture their targets well. Each of these primer sets will inherently miss taxa and this will be a systematic error. Unfortunately, there is no one primer set that captures all of the diversity, and the diversity present in soil makes it impossible to choose one primer set that balances specificity and resolution. This is only an issue if there is a particular target taxon (i.e. an indicator species) of interest.

Assigning taxonomic identities to the sequences is only possible through their comparison to reference databases, which are incomplete. This is not an issue if a taxonomy free approach is adopted – i.e. tracking changes over time by comparing datasets (as is advocated here), but it is a bigger concern if indicator species or functional groups (based on taxonomy) are required. It should be noted that multiple OTUs can be identified as belonging to the same species, which is most likely attributed to PCR or sequencing artefacts but potentially intraspecific genomic variation or cryptic diversity. Also, it is possible for closely related species to have identical sequences in the targeted gene region and if the species present at your site is not in the database it could be identified as a different closely related species.

What is happening among the communities (i.e. functioning) may be driven more so by the dominant taxa as opposed to the breadth of diversity. The abundance of taxa cannot be directly inferred from the number of sequence reads. While the number of sequence reads is a consequence of abundance, it is also impacted by biomass, body type, activity, surface area, condition, primer bias, and species-specific variation in the genome.

4.5 Sample analysis costs and kit components

Table 16: Overview of sample analysis costs for each habitat. Costs reflect standard NatureMetrics pricing in 2021/2022 and are provided only as an indication of costs for an end user. Note that in multiple assays were applied for the purposes of this scientific project and this is not required for targeted routine biomonitoring programmes.

		Pilot study			Main campaign			
	Sample type	n samples	n assays/ sample	Cost £	n samples	n assays/ sample	Cost £	Total Cost £
Marine	Water	9	3	9x475 = 4,275	49	1	49x275 = 14,700	19,650
	Sediment	6	2	6x375 = 2,250	56	3	56x475 = 26,600	33,350
Freshwater	Water	24	4	24x575 = 13,800	95	3	95x475 = 45,125	68,450
Woodland	Soil	32	3	32x475 = 15,200	77	3	77x475 = 36,575	59,950
Peatland	Soil	14	3	14x475 = 6,650	6464	3	64x475 = 30,400	42,900
Total costs £				42,175			152,175	194,350
Total n samples		85			341			426
Av. Cost/sample								456

Table 17: Kit components and costs per kit and per assay (per sample). Costs reflect standard NatureMetrics pricing in 2021/2022 and are provided only as an indication of costs for an end user. Note that in multiple assays were applied for the purposes of this scientific project and this is not required for targeted routine biomonitoring programmes.

Sample type	Kit (1 sample) Price £	Kit + 1 assay £	2nd assay £	3rd assay £	Sample x 2 assays £	Sample x 3 assays £	Sample x 4 assays £
Water filter	20	275	100	100	375	475	575
Sediment	NA	275	100	100	375	475	575

Table 18: Kit components for one aquatic sampling kit

	Kit Components Aquatic
n	Item
1 pair	Nitrile gloves
1	3.5 L sampling bag
1	60 ml Luer Lock plastic syringe
1	Enclosed filter (0.8 µm pore size, polyethersulfone (PES) and a 5 µm glass fibre prefilter)
1	Resealable bag
1	Small syringe filled with 1.5 ml Longmire's solution (sealed with reusable Luer lock cap)
1	Extra Luer lock cap
1	Specimen bag
1	Sampling data sheet

Table 19: Kit components for one soil/sediment sampling kit – cold storage

	Kit Components Soil – cold storage
n	Item
1 pair	Nitrile gloves
1	A4 grip seal bag
1	Soils & Sediments DNA Sampling Kit labels
1	Soil & Sediment cold storage datasheet
1	clear snap lock bag with write on labels
1	10ml syringe (corer)
1	sealable silver mailer bag (barcoded)

Table 20: Kit components for one soil/sediment sampling kit – Buffer storage

	Kit Components Soil – buffer storage
n	Item
1 pair	Nitrile gloves
1	A4 grip seal bag
1	Soils & Sediments DNA Sampling Kit labels
1	Soil & Sediment buffer storage datasheet
1	clear snap lock bag with write on labels
1	10ml syringe (corer)
2	125 mL sample pot
50ml	RNAlater buffer (Preservation buffer solution)

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.
extraction blank	A DNA extraction with no soil 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 percentage, 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	Used 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 segment 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 is 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

Allen, C., Axelsson, M., Dewey, S. & Clark, L., 2013. Marine biological survey to establish the distribution of Priority Marine Features within the Clyde Sea area. Scottish Natural Heritage Commissioned Report No.437. 1–240.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology 215, 403–410.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS ONE 4, e6372. https://doi.org/10.1371/journal.pone.0006372

Apprill, A., McNally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat. Microb. Ecol. 75, 129–137. https://doi.org/10.3354/ame01753

Artz, R.R.E., Donnelly, D., Andersen, R., Mitchell, R., Chapman, S.J., Smith, J., Smith, P., Cummins, R., Balana, B., Cuthbert, A., 2014. Managing and restoring blanket bog to benefit biodiversity and carbon balance – a scoping study.

Bedwell, M.E., Goldberg, C.S., 2020. Spatial and temporal patterns of environmental DNA detection to inform sampling protocols in lentic and lotic systems. Ecology and Evolution 10, 1602–1612. https://doi.org/10.1002/ece3.6014

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A.S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y.J.K., Sanli, K., Eriksson, K.M., Vik, U., Veldre, V., Nilsson, R.H., 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol n/a-n/a. https://doi.org/10.1111/2041-210X.12073

Blair, J.M., Bohlen, P.J., Freckman, D.W., 1996. Soil Invertebrates as Indicators of Soil Quality, in: Doran, J.W., Jones, A.J. (Eds.), SSSA Special Publications. Soil Science Society of America, Madison, WI, USA, pp. 273–291. https://doi.org/10.2136/sssaspecpub49.c16

Borja, A., 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, 594–602. https://doi.org/10.1016/j.ecolind.2017.11.018

Borja, A., Franco, J., Pérez, V., 2000. A marine Biotic Index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40, 1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8

Bruce, K., Blackman, R., Bourlat, S.J., Hellström, A.M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., Elbrecht, V., Fazi, S., Fonseca, V., Hänfling, B., Leese, F., Mächler, E., Mahon, A.R., Meissner, K., Panksep, K., Pawlowski, J., Schmidt Yáñez, P., Seymour, M., Thalinger, B., Valentini, A., Woodcock, P., Traugott, M., Vasselon, V., Deiner, K., 2021. A practical guide to DNA-based methods for biodiversity assessment. Pensoft Publishers. https://doi.org/10.3897/ab.e68634

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107

Capra, E., Giannico, R., Montagna, M., Turri, F., Cremonesi, P., Strozzi, F., Leone, P., Gandini, G., Pizzi, F., 2016. A new primer set for DNA metabarcoding of soil Metazoa. European Journal of Soil Biology 77, 53–59. https://doi.org/10.1016/j.ejsobi.2016.10.005

Chamberlain, S., Oldoni, D., Geffert, L., Desmet, P., Barve, V., Ram, K., Blissett, M., Waller, J., McGlinn, D., Ooms, J., Steven (Siwei) Ye, Oksanen, J., Marwick, B., John, Sumner, M., Sriram, 2022. ropensci/rgbif: rgbif v3.7.0. https://doi.org/10.5281/ZENODO.1045299

Collins, R.A., Bakker, J., Wangensteen, O.S., Soto, A.Z., Corrigan, L., Sims, D.W., Genner, M.J., Mariani, S., 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution 10, 1985–2001. https://doi.org/10.1111/2041-210X.13276

de Souza, G.B.G., Barros, F., 2017. Cost/benefit and the effect of sample preservation procedures on quantitative patterns in benthic ecology. Helgol Mar Res 71, 21. https://doi.org/10.1186/s10152-017-0501-3

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv 081257. https://doi.org/10.1101/081257

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A., Winfield, I.J., 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25, 3101–3119. https://doi.org/10.1111/mec.13660

Hervé, A., Domaizon, I., Baudoin, J.-M., Dejean, T., Gibert, P., Jean, P., Peroux, T., Raymond, J.-C., Valentini, A., Vautier, M., Logez, M., 2022. Spatio-temporal variability of eDNA signal and its implication for fish monitoring in lakes. PLoS ONE 17, e0272660. https://doi.org/10.1371/journal.pone.0272660

Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol Evol 7, 1451–1456. https://doi.org/10.1111/2041-210X.12613

Kelly, R.P., Port, J. a., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F., Mach, M.E., Bennett, M., Prahler, E., Caldwell, M.R., Crowder, L.B., 2014. Harnessing DNA to improve environmental management. Science 344, 1455–1456. https://doi.org/10.1126/science.1251156

Lawson Handley, L., Read, D.S., Winfield, I.J., Kimbell, H., Johnson, H., Li, J., Hahn, C., Blackman, R., Wilcox, R., Donnelly, R., Szitenberg, A., Hänfling, B., 2019. Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. Environmental DNA 1, 26–39. https://doi.org/10.1002/edn3.5

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., Zizka, V.M.A., 2021. Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environmental DNA 3, 261–276. https://doi.org/10.1002/edn3.177

Lejzerowicz, F., Gooday, A.J., Barrenechea Angeles, I., Cordier, T., Morard, R., Apothéloz-Perret-Gentil, L., Lins, L., Menot, L., Brandt, A., Levin, L.A., Martinez Arbizu, P., Smith, C.R., Pawlowski, J., 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, 1–23. https://doi.org/10.3389/fmars.2021.671033

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J., 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, 34. https://doi.org/10.1186/1742-9994-10-34

Li, J., Hatton-Ellis, T.W., Lawson Handley, L.J., Kimbell, H.S., Benucci, M., Peirson, G., Hänfling, B., 2019. Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. Journal of Applied Ecology 56, 1232–1244. https://doi.org/10.1111/1365-2664.13352

Littlefair, J.E., Hrenchuk, L.E., Blanchfield, P.J., Rennie, M.D., Cristescu, M.E., 2021. Thermal stratification and fish thermal preference explain vertical eDNA distributions in lakes. Mol Ecol 30, 3083–3096. https://doi.org/10.1111/mec.15623

Lobo, J., Costa, P.M., Teixeira, M.A., Ferreira, M.S., Costa, M.H., Costa, F.O., 2013. Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecol 13, 34. https://doi.org/10.1186/1472-6785-13-34

Longmire, J.L., Baker, R.J., Maltbie, Mary., Texas Tech University., 1997. Use of "Lysis Buffer" in DNA isolation and its implication for museum collections /. Museum of Texas Tech University, Lubbock, TX. : https://doi.org/10.5962/bhl.title.143318

Manu, M., Băncilă, R.I., Bîrsan, C.C., Mountford, O., Onete, M., 2021. Soil mite communities (Acari: Mesostigmata) as indicators of urban ecosystems in Bucharest, Romania. Sci Rep 11, 3794. https://doi.org/10.1038/s41598-021-83417-4

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17, 10.

May, L., Taylor, P., Gunn, I.D.M., Thackeray, S.J., Carvalho, L.R., Hunter, P., Corr, M., Dobel, A.J., Grant, A., Nash, G., Robinson, E., Spears, B.M., 2022. Assessing climate change impacts on the water quality of Scottish standing waters.

Miya, M., Gotoh, R.O., Sado, T., 2020. MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries Science. https://doi.org/10.1007/s12562-020-01461-x

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., Iwasaki, W., 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, 150088. https://doi.org/10.1098/rsos.150088

Moore, C.G., 2013. NatureScot Commissioned Report 631: 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.

Murphy, K.J., Beveridge, M., Tippett, R., 1994. The Ecology of Loch Lomond (Developments in Hydrobiology, 101. Springer.

NatureMetrics, 2022. Protocol for Sampling Marine Benthic Sediment for DNA-based Analysis (JNCC Report No. 706) 16.

Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47, D259–D264. https://doi.org/10.1093/nar/gky1022

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples: Primers for marine microbiome studies. Environ Microbiol 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, D590–D596. https://doi.org/10.1093/nar/gks1219

Ratnasingham, S., Hebert, P.D.N., 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., Coissac, E., 2011. EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research 39, 1–11. https://doi.org/10.1093/nar/gkr732

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E., Hellström, M., 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution 8, 635–645. https://doi.org/10.1111/2041-210X.12683

Stork, N.E., Eggleton, P., 2009. Invertebrates as determinants and indicators of soil quality. am j alt ag 7, 38–47. https://doi.org/10.1017/S0889189300004446

van der Heyde, M., Bunce, M., Nevill, P., 2022. Key factors to consider in the use of environmental DNA metabarcoding to monitor terrestrial ecological restoration. Science of The Total Environment 848, 157617. https://doi.org/10.1016/j.scitotenv.2022.157617

van der Loos, L.M., Nijland, R., 2020. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology 1–19. https://doi.org/10.1111/mec.15592

Wang, Y., Naumann, U., Wright, S.T., Warton, D.I., 2012. mvabund– an R package for model-based analysis of multivariate abundance data. Methods in Ecology and Evolution 3, 471–474. https://doi.org/10.1111/j.2041-210X.2012.00190.x

Wangensteen, O.S., Palacín, C., Guardiola, M., Turon, X., 2018. DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. PeerJ 6, e4705. https://doi.org/10.7717/peerj.4705

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: PCR Protocols. Elsevier, pp. 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Xu, J., Morris, P.J., Liu, J., Holden, J., 2018. PEATMAP: Refining estimates of global peatland distribution based on a meta-analysis. CATENA 160, 134–140. https://doi.org/10.1016/j.catena.2017.09.010

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucl. Acids Res. 42, D643–D648. https://doi.org/10.1093/nar/gkt1209

Zhang, S., Lu, Q., Wang, Y., Wang, X., Zhao, J., Yao, M., 2020. Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Molecular Ecology Resources 20, 242–255. https://doi.org/10.1111/1755-0998.13105

NATURE METRICS
DNA-BASED MONITORING

Email
edna-lab@naturemetrics.co.uk

Phone
+44 2038 767350

Website
www.naturemetrics.co.uk