Glen Prosen soil biodiversity baseline survey

Andrea Britton, Victoria Buswell, Debbie Fielding, Richard Hewison, Isla Morton, Tom Parker, Lucinda Robinson, Duncan White, Alon Zuta & Andy Taylor

The James Hutton Institute

March 2025

SEFARI

Contents

Executive summary 3

Introduction 6
- Glen Prosen 6
- Purpose of the baseline soil biodiversity survey 6

Methods 7
- Survey design 7
- Field Sampling methodology 10
- Laboratory analyses 10
  - Soil chemical analysis 10
  - DNA extraction and sequencing 11
- Bioinformatics 12
- Statistical analyses 12

An introduction to eDNA approaches for exploring soil biodiversity 13

Survey results 15
- Total soil biodiversity at Glen Prosen 15
- Taxonomic composition of soil biodiversity 17
  - Prokaryote kingdoms 17
  - Eukaryote kingdoms 18
- Variation in soil community richness and composition across land use categories 22
- Relationship between soil biodiversity and the environment 26
  - The environmental setting 26
  - Soil biodiversity response to the environment 29
    - Soil community taxon richness 29
    - Soil community composition 29
- Focus on plant root associated fungi 33
  - Significance of mycorrhizal associations 33
  - Plant root associated fungi at Glen Prosen 33

Conclusions 37

References 38

Appendix 40

This report should be cited as:
Britton A, Buswell V, Fielding D, Hewison R, Morton I, Parker T, Robinson L, White D, Zuta A and Taylor A (2025) Glen Prosen soil biodiversity baseline survey. Research Report for Forestry and Land Scotland. The James Hutton Institute, Aberdeen. Zenodo. DOI: 10.5281/zenodo.15744474

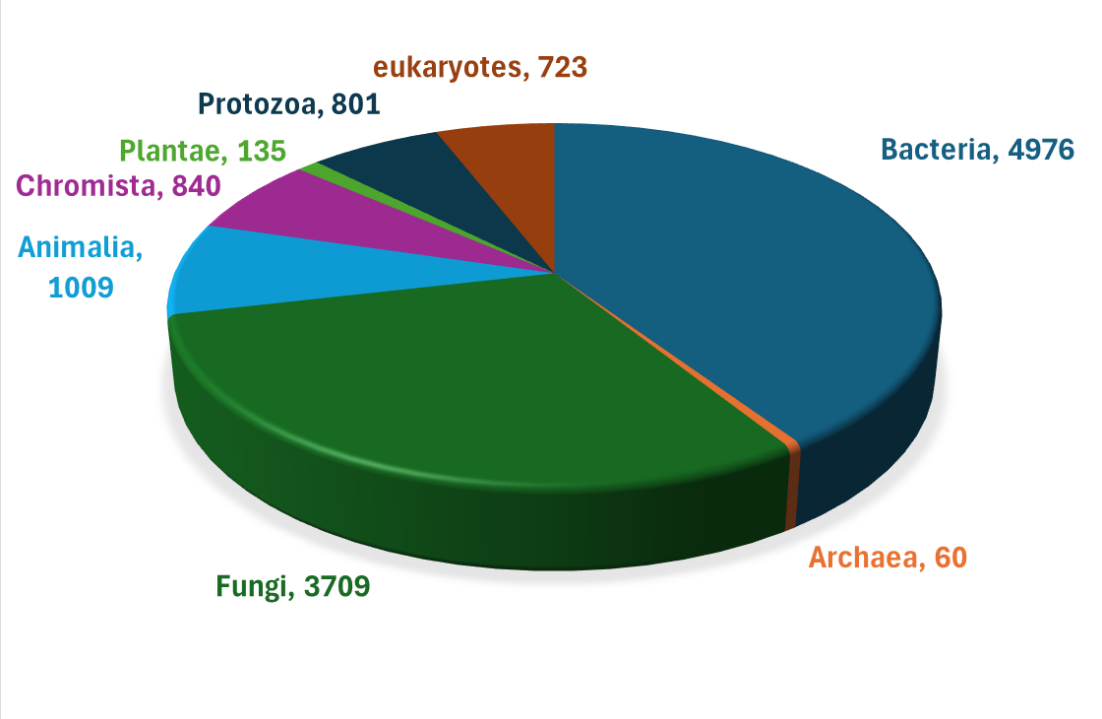
Executive summary

The Glen Prosen Estate was purchased by Forestry and Land Scotland in 2022 to create an opportunity for landscape-scale restoration of native woodland, peatlands and open-ground habitats. Before changes in habitat management commence, a range of biodiversity surveys are being undertaken to inform future management decisions.

This report describes an eDNA survey of soil biodiversity in Glen Prosen which is intended to provide a baseline against which future changes in soil biodiversity could be measured. Details of the approaches and methodologies used are provided, such that the survey could be repeated at a future date to investigate the progress of soil communities in response to the changing management envisaged for the glen.

For the survey we used three DNA markers (16S, 18S and ITS2) which allowed us to detect organisms from all seven kingdoms of life (Animalia, Archaea, Bacteria, Chromista, Fungi, Plantae, Protozoa). We put a particular emphasis on exploring fungal diversity, as many fungi form symbiotic relationships with plants, particularly trees, which are essential to plant survival and therefore have a crucial role to play in regeneration of native woodland cover at Glen Prosen.

Soil diversity was examined in 48 sample plots distributed across the glen, contrasting open and forested habitats, natural regeneration of trees or tree planting, and areas where no change is planned. During the survey we recovered more than 3.6 million individual DNA sequences from the 48 sample plots. A total of 12,253 taxa were detected in the soil, representing diverse groups from all seven kingdoms of life.



Proportion of organisms in the total soil biodiversity at Glen Prosen assigned to each kingdom of life. Numbers are the total number of taxa detected in each kingdom, 'eukaryotes' refers to 18S eukaryote sequences which could not be assigned to a kingdom.

Soil biodiversity includes a vast array of different organisms from bacteria and archaea through unicellular microfauna to larger mesofauna and macro-organisms such as vascular plants and vertebrates which live in and on the soil. At Glen Prosen, the incredible richness present in the soil was dominated by bacteria (40% of all taxa) and fungi (30%) but a huge range of other organisms was also detected. Many of the most familiar soil organism groups fall within the kingdom Animalia. Diversity in this group was dominated by nematode worms as is typical for the types of acidic upland soils present at Glen Prosen. Arthropods, including mites, springtails, millipedes and insects, and annelids (worms), tardigrades and rotifers also represented significant proportions of the animal diversity. Many of these organisms feed on bacteria, fungi and organic detritus in the soil or predate on each other. In the kingdom Chromista it was notable that there was a high diversity of Apicomplexa, which is a group of unicellular organisms that are primarily obligate parasites of animals, usually arthropods, reflecting the complexity of soil food webs.

Spatial heterogeneity of soil biodiversity is very high, and it would not be possible to exhaustively sample every taxon present within an area such as Glen Prosen. This was reflected in the analysis of the richness data, which showed that, for every kingdom, more taxa would be found if the number of samples collected was increased. The level of sampling was however sufficient to reveal clear and distinct patterns in taxon richness and community composition across the glen. The number of taxa of each kingdom detected in each sample did not differ in a consistent way between different land use categories at Glen Prosen. However, soil taxon richness in all kingdoms except the Bacteria was significantly related to topographic and/or soil chemical conditions at the sampling plots. For the majority of kingdoms, greatest richness was found on higher pH soils and richness declined within increasing elevation and soil C:N ratio.

In contrast to richness, the community composition of the soil biodiversity did differ significantly between land use categories, with the primary contrast being between open habitats and those which currently or previously supported trees. As for richness, community composition was also affected by topography and soil chemistry. Open habitats with high pH soils in the valley bottom were differentiated from higher elevation plots with more acidic soils. These patterns were clearest for the kingdom Fungi but were also seen to varying degrees in all other kingdoms.

Plant root associated fungi showed clear patterns of differential distribution among the land use categories. Ectomycorrhizal fungi which are essential for the establishment and survival of many tree species were found primarily in current plantation areas or where trees had recently been felled or windblown and surprisingly also in areas that had not supported trees for many years, albeit in much lower numbers. They had very limited occurrence in open habitats, especially those identified for tree planting and natural regeneration of tree cover, which could limit or slow down tree regeneration until suitable fungal partners are able to colonise.

This eDNA survey of the soil biodiversity at Glen Prosen marks one of the first occasions that this relatively new technique has been used to establish a soil biodiversity baseline before a landscape-scale change in management. The survey has been extremely successful in showing that this technique can be used effectively to identify patterns and land use

associations in soil biodiversity that are relevant to the future management of the glen. The survey also represents one of the first times that soil biodiversity has been explored at scale across an upland landscape in Scotland and demonstrates that the vast majority of biodiversity in this type of landscape is to be found in the soil.

The dataset generated by this soil biodiversity baseline is huge and the analyses presented in this report have only scratched the surface of what could be investigated using these data. The dataset illustrates very effectively how changes in land use impact not only on the visible above ground plant communities but also result in massive associated changes below ground in the much richer communities of soil organisms. The tight group of the alpine and many of the open area (OP) plots highlights the similarities of the communities in these plots under similar vegetations. This contrasts with the much greater heterogeneity amongst communities at lower elevations, particularly between forested and open areas in the valley bottom. These differences below ground will strongly influence important ecosystem services, including soil nutrient and carbon cycling and sequestration – functions which are dependent on different components of soil biodiversity, particularly the fungi. The generated dataset would enable an examination of these ecosystem services in relation to the identified communities of soil biodiversity. The Glen Prosen data are also unique because they cover all seven kingdoms of biodiversity and offer a remarkable opportunity for the construction of complex soil food webs and investigation of how these are related to land use type and land use change. In addition, other aspects of soil biology could be explored in more detail including predator/prey or parasite/host relationships, co-occurrence or avoidance between different groups of soil organisms, and network analyses of plant communities and soil biodiversity. All of these would be worthwhile avenues for further research.

These baseline data set the stage for future monitoring of change in Glen Prosen, with sample plots which are geopositioned and photographed to aid relocation in the future. At present we do not know how quickly changes above ground are accompanied by changes below ground or how quickly soil functions change. We also do not know the dynamics of natural cycles in soil biodiversity associated with many of the land use categories at Glen Prosen. These are questions that could be addressed using the present study as a baseline. This survey has highlighted that even with a relatively small number of samples it is possible, with a well-designed sampling strategy, to capture sufficient detail of the megadiversity of soil communities to enable visualisation of meaningful spatial patterns, which can form the baseline for future assessments of the impacts of changes in land use.

Introduction

Glen Prosen

In November 2022, Forestry and Land Scotland purchased 3,500 ha of land at Glen Prosen with the broad aim of linking existing public land holdings in the Angus glens and creating an opportunity for landscape-scale habitat restoration within the Cairngorms National Park. The concept for the future management of Glen Prosen is to create biodiverse and resilient ecosystems by increasing native woodland cover, restoring peatland habitats, improving the condition of riparian and open-ground habitats and managing herbivore densities to benefit a range of wildlife in the glen. Restoration of native woodland habitats within the glen is envisaged to be implemented through a combination of planting of native tree species and natural regeneration.

In the initial phase of the project, a range of biodiversity and habitat surveys are being undertaken to provide data which can inform land management planning for the subsequent restoration and recreation of native habitats within the glen.

Purpose of the baseline soil biodiversity survey

To develop an understanding of the belowground biodiversity at Glen Prosen, including the species present and the distribution of biodiversity across the landscape, a baseline survey of soil biodiversity using eDNA methods was conducted during early summer 2024. The purpose of this baseline sampling was to determine what soil biodiversity was present in open and forested habitats at Glen Prosen, prior to the start of tree planting and native woodland restoration activities. This baseline could then be revisited at a future date to monitor how tree planting and natural regeneration of woodland species have altered belowground biodiversity over the longer term. In addition, baseline soil biodiversity in habitats at Glen Prosen could be compared with soil biodiversity in established native woodland habitats in north-east Scotland, where these are in good condition, giving an indication of potential targets for soil biodiversity restoration.

Methods

Survey design

The soil biodiversity sampling scheme for Glen Prosen was designed to cover the main habitat transitions likely to occur following tree planting and natural regeneration of native woodland in the glen. Since there were potentially many possible habitat type transitions, sampling effort was focused on dry open habitats and currently or formerly forested habitats where tree establishment is most likely to occur. We did not sample wet habitats (peatland, marsh and fen) which would be less suitable for woodland regeneration. The habitat type transitions investigated are referred to throughout as land use categories for ease of reference.

Based on topographic information, habitat maps and provisional planting plans supplied by Forestry and Land Scotland (Figure 1), six land use categories were identified for sampling (Table 1). These included currently open habitats above and below 650m (contrasting upland and alpine habitat types), areas of previous plantation forestry now felled, and areas of extant plantation forest. Land use categories were defined by their current habitat type and planned future management. The number of samples in each category was selected to reflect their relative extent, with 48 plots sampled in total. The distribution of sampling locations is shown in Figure 2.

Soil biodiversity at each sampling location was determined using an eDNA approach (DNA metabarcoding) as described below. Surveyors also collected basic information on vegetation composition and the environment, and some simple soil chemical parameters were determined to allow some insight into factors which might influence the distribution of soil biodiversity within the glen.

Table 1. Land use categories sampled during the soil biodiversity survey at Glen Prosen.

Land use category	Code	Number of samples	General location of sampling areas
> 650m Open alpine habitat (some natural regeneration may occur)	A	12	Hill of Strone, Hunt Hill, Kilbo Path and Bawhelps/Broom Hill
< 650 Open habitat + natural regeneration	OR	12	Middle slopes and valley bottoms throughout the site
<650m Open habitat + native tree planting	OP	6	Lower slopes of Hunt Hill and area to the south of Glen Prosen Lodge
Plantation felled + native tree planting	PP	6	Lower and middle slopes around Kilbo Bothy and Craig Tillelet
Plantation felled + natural regeneration	PR	6	Upper reaches of Glen Prosen forest
Plantation retained	P	6	Glen Prosen forest

Figure 1. Map of Glen Prosen showing extant forest and proposed planting and regeneration areas as of April 2024

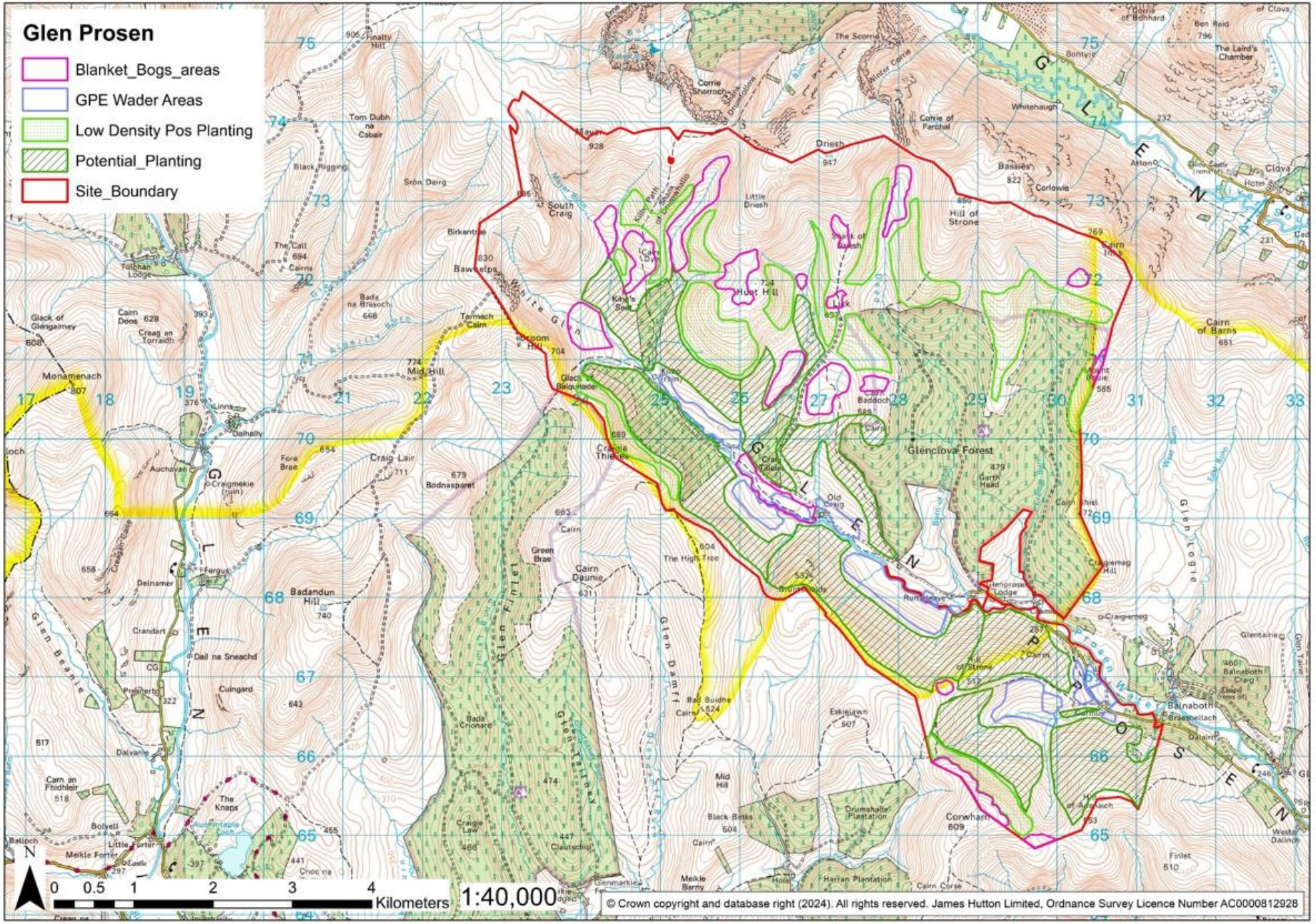
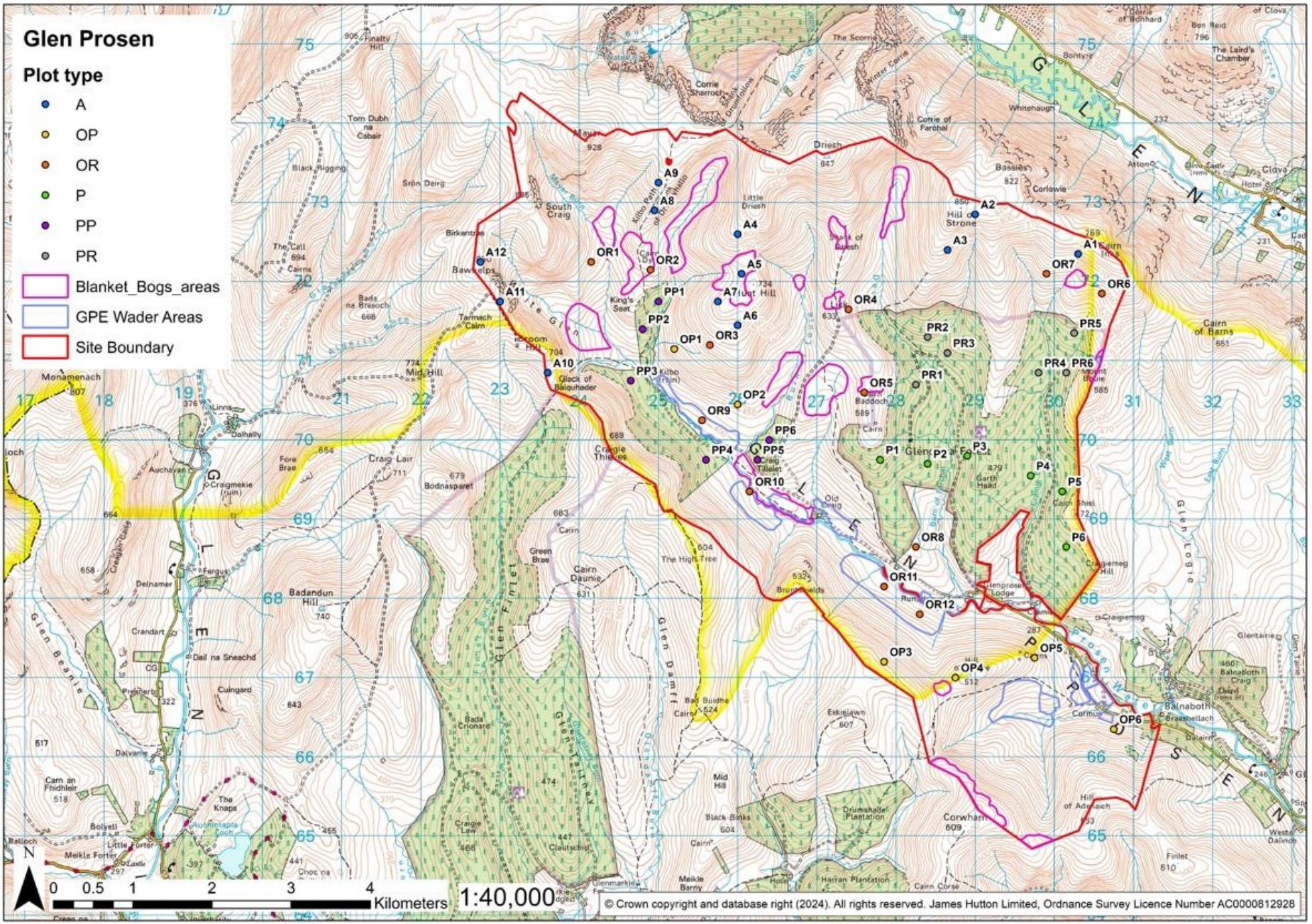


Figure 2. Map of Glen Prosen showing soil biodiversity survey sampling locations. For explanation of plot type codes see Table 1.



Field sampling methodology

At each sample point, information was collected for a representative 10 m x 10 m area. Sample points were located using GPS, with the grid reference of the sample point defining the southern point of a square with the diagonal aligned north south. Plots were temporarily marked out with a tape and canes, but no permanent markers of the sample points were left in situ. For each sample point, the elevation of the plot was measured using GPS and the slope and aspect was recorded. Each plot was then photographed with the marker canes in view to aid any subsequent relocation efforts.

After walking over the plot, surveyors recorded the percent cover of mosses, lichens, grasses, forbs, dwarf-shrubs, shrubs and trees, to the nearest 5% for cover 5-100% or to the nearest 1% for cover <5%. If there were any trees (including saplings or seedlings) present this was noted together with the species. The presence of any (non-tree) species likely to host ectomycorrhizal (ECM) fungi (which play a significant role in the establishment and health of trees) such as *Arctostaphylos uva-ursi*, *Bistorta vivipara* and *Dryas octopetala* was also noted.

Soil samples were collected using a 5 cm diameter corer, with one core taken from the centre of each quarter of the plot. Before collecting the cores, any loose plant material or litter was brushed to one side and samples were taken from the organic horizon downwards to a depth of 5 cm (thus sampling the most biologically active portion of the soil). All four cores were pooled into a single sample per plot and any large stones (>5 mm) were removed. Any plant roots present were retained in the sample as these are important sources of soil biodiversity. All sampling equipment was cleaned thoroughly between plots using 70% ethanol and samples were kept cool during sampling and then stored at -20 °C until they were processed.

Laboratory analyses

The four pooled cores from each sample plot were thawed and passed through an 8 mm sieve to remove coarse roots (> 2 mm diameter) and large stones. The sieved sample was then thoroughly homogenised before a 10 ml subsample was removed, frozen and cryomilled in a liquid nitrogen cooled mill. This subsample was used for DNA extraction, soil chemistry and pH determination. Three additional aliquots of cryomilled soil (ca. 2 g) were archived at minus 80 °C for future use.

Soil chemical analysis

The cryomilled soil was oven dried (2 days @ 60 °C) before being Retch milled (3 mins) prior to determination of total carbon and nitrogen content by Dumas combustion using a Flash EA1112 Elemental Analyser (Thermofinnigan, Milan, Italy). Soil pH was measured using 5g soil in 45 ml water.

DNA extraction and sequencing

DNA extraction

Total nucleic acids were extracted from homogenised, cryomilled soil following the method of Griffiths et al. (2000), as modified by Freitag and Prosser (2009). Briefly, liquid N_2-chilled 250 mg subsamples were placed in lysing matrix 'B' tubes (MP Biomedicals), preloaded with 0.5 ml 5% CTAB in 0.7M NaCl/120mM K_2HPO_4 extraction buffer, brought to ice temperatures, before the addition of 250 µL of ice-cold Tris-buffered phenol at pH 6.7±0.2 and 250 µL chloroform/isoamyl alcohol (24:1). Samples were homogenized for 2x 15 s at 5000 rpm on a Precellys 24 tissue homogenizer with a five-minute incubation on ice after each homogenization cycle. After centrifugation for 10 min at 16,000 x g, 4 °C, the upper, aqueous layer was transferred to ice-cold 2 mL MaXtract High Density gel barrier tubes (Qiagen), pre-loaded with 100 µL of each phenol and chloroform:isoamyl alcohol (24:1). Tubes were gently inverted for ca. 30 s and centrifuged at 20817 x g, 10 min, 4 °C, to allow gel barrier separation of aqueous and organic layers. Traces of phenol were removed with the addition of 500 µL of ice-cold chloroform:isoamyl alcohol. Tubes were gently mixed by inversion and centrifuged at 20817 x g for 10 min. The aqueous phase was transferred to 1.5ml tubes and 1 mL of 30 % (w/v) PEG 6000 in 1.6 M NaCl was added and mixed gently by inversion before resting on ice for 2 hours. Nucleic acids were pelleted by centrifugation for 20 min at 20817 x g and the supernatant was removed. Pellets were washed twice with 1 mL of 70 % (v/v) ethanol and centrifuged for 10 min at 20817 x g. Ethanol was removed by pipette aspiration and pellets dried at 55 °C, before resuspension in 30 µL of nuclease-free water. Tubes were then incubated for 5- 15 min at 55 °C to remove any residual ethanol. Resuspended extracts were flash frozen in liquid N_2 and stored at -80 °C until further analysis.

Contamination during the extraction procedure was examined using control extraction blanks that were mock samples of 250µL molecular grade sterile nuclease free water (Sigma-Aldrich) and sterile sand, both of which were extracted and cleaned up alongside field samples. Extraction controls were included in a ratio of 1:24 samples. The control extracts were diluted in a similar ratio as the experimental samples and submitted for sequencing along with the sample set.

Sequencing

For sequencing library production and sequencing, 200 ng of DNA from each sample and control extract were shipped to LGC Biosearch Technologies, Berlin, Germany. Amplicons were generated using the 515F-806R 16S rRNA gene primers (Apprill et al. 2015; Parada et al. 2016) for bacteria and Archaea, fITS7 - ITS4 primers for fungi (Ihrmark et al. 2012), and 18S: TAReuk454FWD1 / TAReukREV3mod (modified from Stoeck et al. 2010 - ACTTTCGTTCTTGATYRATGA) for eukaryotes. As expected, controls failed to amplify and were excluded from sequencing. Library preparation and sequencing was performed according to the Ovation Rapid DR Multiplex System (Tecan Trading AG, Switzerland) according to the manual 15027617 MiSeq System Guide (15027617), on an Illumina MiSeq V3 (2x300bp) sequencer.

Bioinformatics

All data were processed using the Qiime2 conda environment version amplicon-2024.10 (Bolyen et al. 2019). Raw data for all amplicons were assessed for read count and quality. ITS amplicon data were trimmed using the ITSxpress plugin to remove the conserved regions surrounding the ITS region to improve taxonomic classification (Nilsson et al. 2009; Rivers et al. 2018).

The Dada2 plugin was used to denoise all datasets (Callahan et al. 2016). Specifically, 16SrRNA and 18SrRNA reads were truncated based on average quality dropping below a Phred score of 20 and adaptors were trimmed. Due to the variable length of the ITS region, no truncation took place, but adaptors were removed. Dada2 was then used to dereplicate, merge forward and reverse reads and remove chimeric reads for all amplicon data sets.

The downstream analysis of the 16SrRNA and 18SrRNA data used amplicon sequence variants (ASVs), while ITS data were clustered into operational taxonomic units (OTUs) at a threshold of 97.5% similarity using the VSEARCH plugin (Rognes et al. 2017).

Taxonomic assignments of ASVs and OTUs were performed using Qiime2's Naive Bayes classifier trained using the Silva (16SrRNA, 18SrRNA) and UNITE (ITS) databases (Quast et al. 2012; Nilsson et al. 2019). To improve classification, the taxonomic classifier for the 16SrRNA and 18SrRNA data was trained on target regions based on primer sequences (Werner et al. 2012). The UNITE database did not undergo read extraction as it is not seen to improve classification (Qiime2 documentation 2023). Finally, ASV and OTU abundance tables were output into the BIOM format and reformatted into a tab separated file for downstream analysis.

Statistical analyses

Data analysis was conducted in R version 4.4.1 (R Core Team 2024) with data visualisation using 'ggplot2' (Wickham 2016). Differences in environmental parameters (soil chemistry and topography) among land use categories were assessed using linear models. For ASV/OTU richness, differences among land use categories and relationships with environmental factors were assessed using Generalized Linear Models (function *glm*) with a quasi-Poisson distribution and a log-link function to account for the use of counts and overdispersion in the data. Taxon composition of soil communities was investigated separately for each kingdom using the packages 'BiodiversityR' (Kindt & Coe 2005) and 'vegan' (Oksanen et al. 2024). Prior to analysis, ASV read abundance matrices for 16S and 18S data were log transformed. Due to the high variability in region length and copy number, ITS2 read numbers do not necessarily reflect taxon abundance, so ITS2 read abundance matrices were transformed to presence-absence before further analysis. For all kingdoms, variation in community composition was visualised using Non-metric Multi-Dimensional Scaling (function *metaMDS*) and significance of differences between land use categories was assessed using PERMANOVA (function *adonis2*). Canonical Correspondence Analysis (function *cca*) was then used to investigate the relationships between soil biodiversity community composition and environmental factors.

An introduction to eDNA approaches for exploring soil biodiversity

Soil has been termed the 'poor man's rainforest' as a reflection of the incredible species richness – biodiversity - which can be found packed into even small volumes of soil. Unknown numbers of viruses, millions of bacteria and archaea, thousands of fungi and other organisms (e.g. worms, protozoa, nematodes, mites) inhabit every cubic centimetre of the upper horizon of most soils. This plethora of organisms carry out many of the major processes essential for the ecosystem services provided by healthy soils, including carbon and nutrient cycling and supporting diverse plant communities. Much of this incredibly rich biodiversity remains unknown and its functions poorly understood. A major first step in closing these huge knowledge gaps is determining which organism groups are associated with different land use types and the environmental factors that drive these associations.

The extraordinary advances over the last 30 years in our ability to extract DNA from environmental samples and then to selectively characterise different components of soil biodiversity has revolutionised our understanding of the complexities of the interactions between below- and above-ground components of terrestrial ecosystems. Although the laboratory methodologies involved in characterising soil communities are highly complex, the basic premise is straight forward. DNA extracted from samples is used to create unique barcodes (small fragments of DNA sequences) for the species in the group(s) of organisms of interest and these barcodes are then compared to databases containing reference barcodes from identified species. Matching environmental barcodes with known reference barcodes allows us to build up a picture of the species present in the communities being sampled. This is the same principle we use when checking out at a supermarket: you scan the unique barcodes on produce and your receipt is a printout of your community of purchases.

Unfortunately, the supermarket analogy is too simplistic. While the supermarket reference database contains all the barcodes of the individual items of produce available on the shelves, the reference databases for biodiversity barcodes are often very incomplete. For example, it is likely that we have identified only about 3-5% of the millions of fungi believed to exist on earth, and of those identified, only a small percentage are in DNA sequence reference databases. A consequence of this is that when we characterise fungal (or any other) communities with sequence barcodes, it is common that we can only get a species name for about 10-15% of the total number of unique barcodes that we can distinguish in the datasets. This does not mean that the other 85% of the barcodes are worthless. On the contrary, each unique barcode represents a population of organisms making up part of the vast richness of soil biodiversity, and because these barcodes are unique, they can be tracked between samples, across habitats, and indeed across the planet – even if we don't know what species they represent. Over time, as the coverage of reference data improves, so will our ability to put species names on environmentally derived DNA barcodes.

The enormous diversity of organisms found in soils come from all of the highest levels or kingdoms that we use to classify known biodiversity. This high-level classification of biodiversity is still somewhat uncertain, and different classification schemes have been proposed, but for this study we use the seven-kingdom scheme of Ruggiero et al. (2015),

namely Animalia, Archaea, Bacteria, Chromista, Fungi, Plantae, Protozoa. The genetic (DNA) diversity represented in soil is therefore vast. It is therefore not surprising that there is no single barcode (DNA sequence region) common to all groups that can be used to characterise communities within all kingdoms. Barcodes used for distinguishing communities in environmental samples are chosen for their ability to detect and distinguish between species within groups and, equally importantly, for the ease with which they can be obtained from environmental samples. It doesn't matter if a particular stretch of DNA can distinguish between every single species in a group when the DNA is obtained from physical specimens (e.g. worms) of each species, if it is very difficult to generate the barcodes from soil samples full of inhibitory compounds like plant phenols. The three barcodes we have used for this study of soil biodiversity in Glen Prosen, the 18S, 16S and the ITS2, all have properties which make them incredibly useful for our purposes. All three are fundamental components of the 'household' apparatus that keep organisms functioning and as such they exist multiple times, sometimes hundreds of times, within each cell of an organism. This means that they are common, and they are also (usually) easily generated from environmental samples. In addition, each barcode or sequence is 300-400 bases in length with sufficient sequence variation to allow us to differentiate between groups of organisms.

Five of the kingdoms examined in this project are termed eukaryotes, which have all of their genetic code contained in a nucleus within each cell. We used the 18S barcode to detect all of these eukaryotic organisms. However, the utility of the 18S for distinguishing between species varies between and within kingdoms. This means that in some groups, the 18S can distinguish between organisms at the species level, while in others it can only separate between different genera or families. Since soil fungi were of particular interest in this project, we also used the ITS2 barcode region, as this is an excellent species-level marker for fungi. Bacteria and Archaea are very different from the other five kingdoms as they lack a nucleus, and for these groups the 16S is used to distinguish between 'species'.

The barcodes we have used are, like all parts of genetic codes, subject to evolutionary change which involves changes in the DNA sequences or mutations. Since the 18S and 16S DNA regions encode very critical functions, the rate of mutation is very slow and strictly controlled and even a single base change in sequences 300-400 bases long is considered ecologically significant. The ITS2 used for the Fungi differs from the other two barcode regions because it is not directly related to critical functions in organisms, the mutation rate is much higher, and the ITS2 region is really only conserved at the species level – this makes it a great barcode for fungal species. These differences between the 16S, 18S and the ITS2 affect how their data are interpreted and reported. In both 16S and 18S data sets, every sequence which is unique, even if it only differs from another by one base, is treated as significant and counted as part of the community richness. The unique entities distinguished using 16S and 18S barcodes are called Amplicon Sequence Variants (ASVs). In contrast, we know that the ITS2 barcode varies within fungal species, even within individuals, and to account for this, data for ITS2 barcodes are clustered together at 97.5% similarity. This means that each entity/species distinguished with the ITS2 barcode is therefore a cluster of sequences which share 97.5% similarity. These are called Operational Taxonomic Units (OTUs). The numbers of OTUs in a sample is treated as the fungal community richness of the sample.

Total soil biodiversity at Glen Prosen

A total of 3,680,064 individual DNA sequences (reads) were recovered from the 48 soil samples collected at Glen Prosen. These comprised 843,976 reads for the 16S marker used to detect Bacteria and Archaea, 1,465,250 reads for the 18S marker used to detect eukaryotes (all organisms with a cell nucleus) and 1,370,838 reads for the ITS2 fungal marker. Assignment of sequence reads to taxa revealed a total of 12,253 taxa present across all the samples. These represent all 7 kingdoms of life: Bacteria (4976 ASVs), Archaea (60 ASVs), Fungi (3709 OTUs), Animalia (1009 ASVs), Chromista (840 ASVs), Plantae (135 ASVs) and Protozoa (801 ASVs), plus an additional 723 ASVs which could only be identified as eukaryotes (Figure 3).

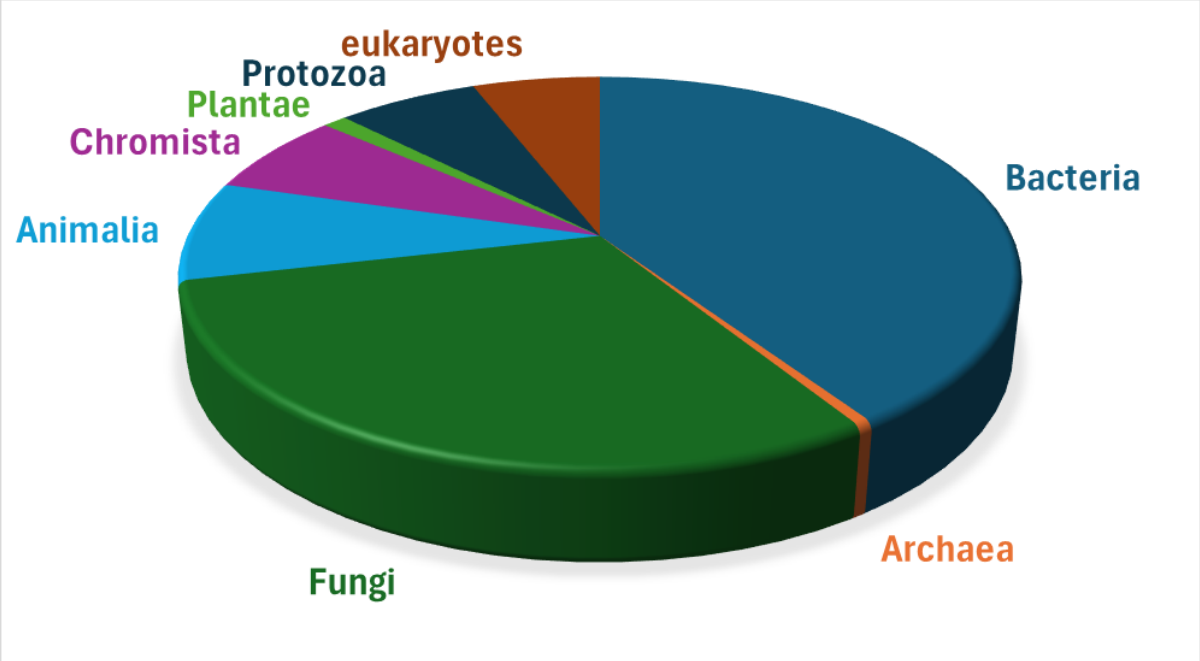


Figure 3. Proportions of organisms in the total soil biodiversity at Glen Prosen assigned to each kingdom of life. The group labelled 'eukaryotes' denotes 18S eukaryote sequences which could not be assigned to a kingdom.

For each kingdom, species accumulation curves showed that the number of taxa detected increased as the number of samples increased, with little sign of the number of taxa found reaching a maximum (Figure 4). This indicates that the soil biodiversity present at Glen Prosen has not been exhaustively sampled in this study, and additional sampling would reveal more taxa. This pattern was the same across all kingdoms suggesting that additional taxa could be found for all forms of soil life. This finding is common in studies of soil biodiversity, where spatial heterogeneity of soil communities is extremely high, and additional taxa will always be found as the number of samples increases. The results of any study will therefore only capture a proportion of the total diversity present in any area.

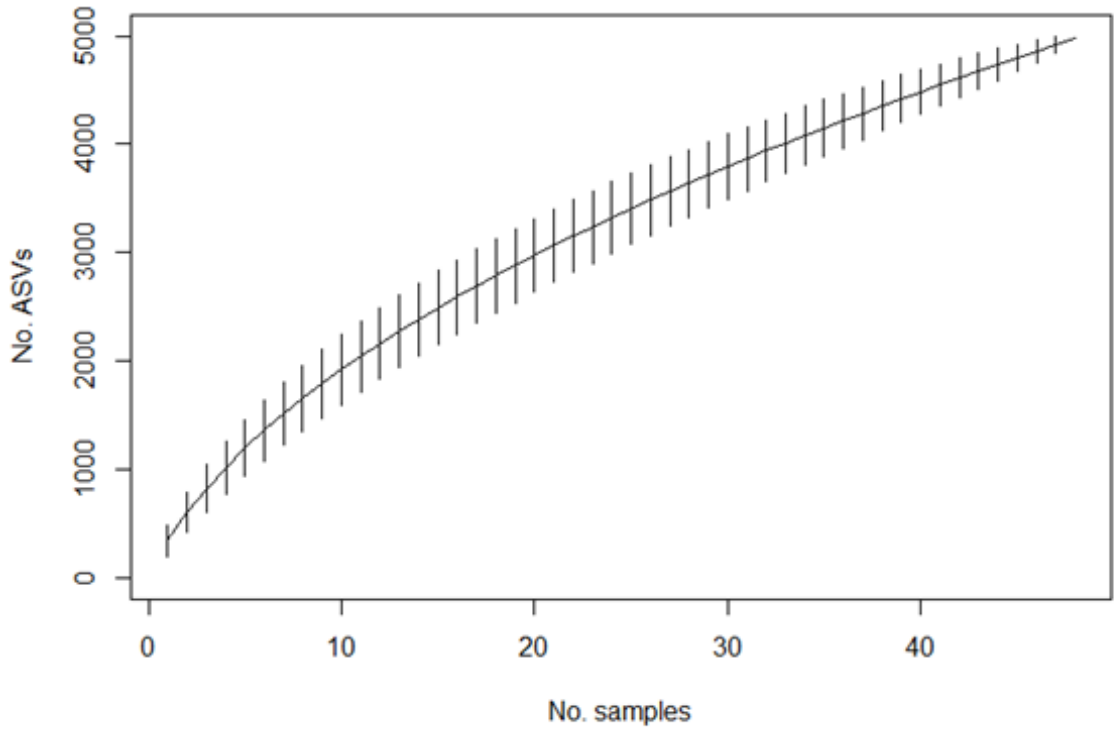


Figure 4. Species accumulation curve for kingdom Bacteria. The line shows the mean number of taxa detected from the specified number of samples when drawn at random from the total sample pool. Error bars show the 95% Confidence Interval.

Taxonomic composition of soil biodiversity

Prokaryote kingdoms

The reads assigned to each kingdom represent a diverse range of different types of organisms. For Bacteria, the 4976 ASVs detected represented 33 phyla (Figure 5), with an additional 133 ASVs which could only be assigned as Bacteria. Five key phyla comprised a significant proportion of the total diversity. These were the Proteobacteria with 1212 ASVs, followed by Planctomycetota (686 ASVs), Acidobacteriota (612 ASVs), Verrucomicrobiota (509 ASVs) and Actinobacteriota (470 ASVs). These phyla have all been previously found to be widespread in soil, especially under aerobic and acidic conditions. Proteobacteria is the largest and most diverse bacterial phylum, containing species occupying a wide variety of ecological niches including involvement in carbon, nitrogen and sulphur cycling. Planctomycetota includes many anaerobic ammonium oxidisers and Acidobacteriota are a physiologically diverse group especially associated with low pH conditions, while Actinobacteriota play a key role in decomposition of organic matter.

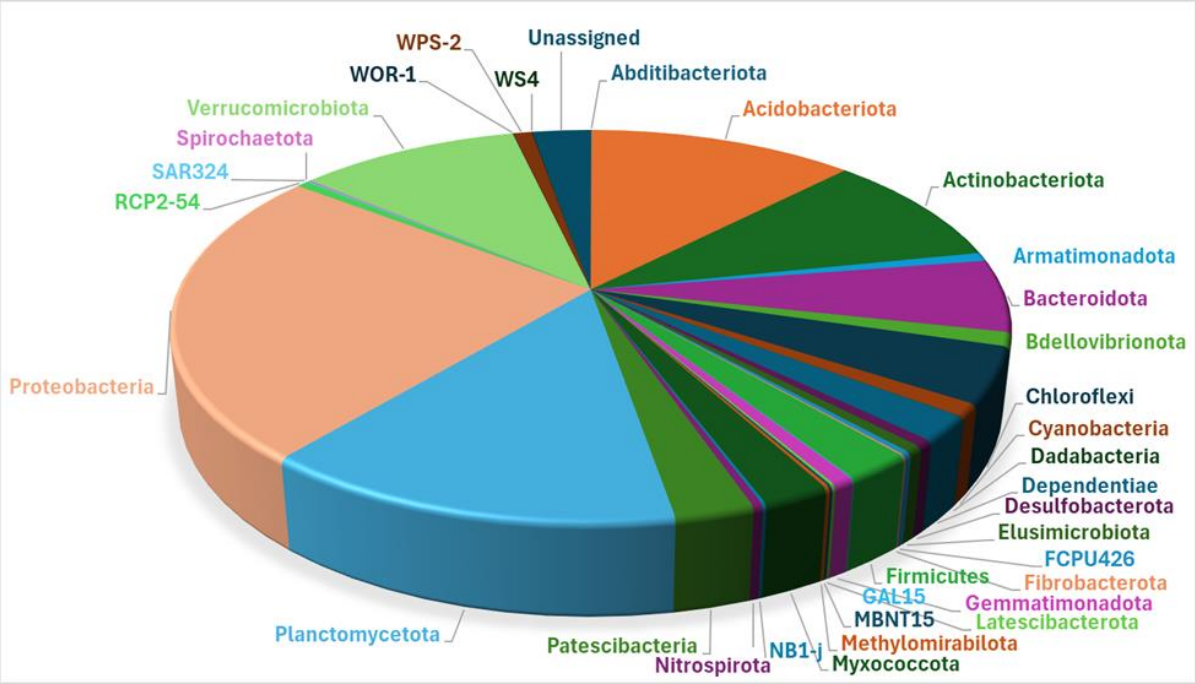


Figure 5. Phylum composition as proportions of total reads assigned to kingdom Bacteria.

Archaea are prokaryotic organisms similar to bacteria, which utilise a diverse range of energy sources. They were first detected living in extreme habitats such hot springs or mine wastes but are now known to be widespread across many habitats including in the soil, where they typically make up around 1% of soil prokaryote (organisms lacking a nucleus) richness. They can fulfil a variety of ecologically important roles including nitrogen cycling, methane production and turnover of organic compounds. Sixty ASVs belonging to Archaea were detected in total, representing 6 phyla (Figure 6). Crenarchaeota (29 ASVs) and Thermoplasmatota (21 ASVs) were the most taxon-rich phyla, with all remaining groups having only 2-3 ASVs. Crenarchaeota are known to be one of the most abundant archaeal

groups in terrestrial environments and are frequently found in soils. Thermoplasmatota are also previously known from soil and are associated with acidic environments.

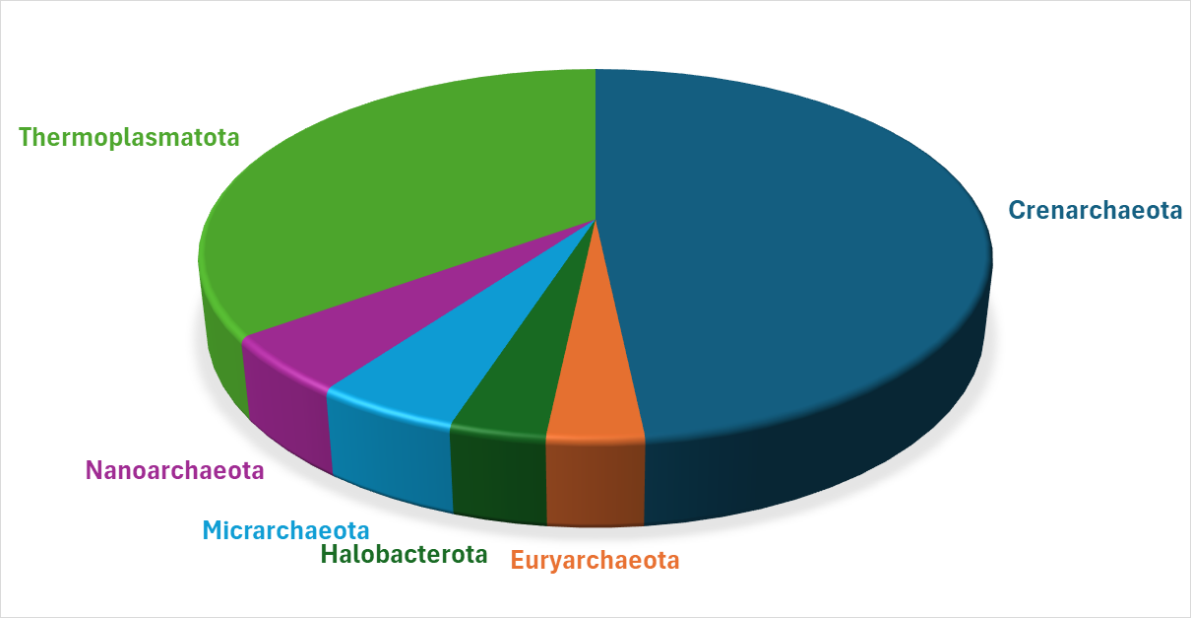


Figure 6. Phylum composition of reads assigned to the kingdom Archaea.

Eukaryote kingdoms

The 1009 ASVs detected in the kingdom Animalia included 12 phyla (Figure 7). The Nematozoa (nematodes) was by far the most taxon-rich phylum with 562 ASVs. Arthropoda (arthropods including insects, millipedes, mites, springtails and spiders) and Rotifera (rotifers) were the next richest phyla with 186 and 76 ASVs respectively. Other important groups of soil fauna which were detected included Annelida (worms), Tardigrada (tardigrades) and Platyhelminthes (flat worms).

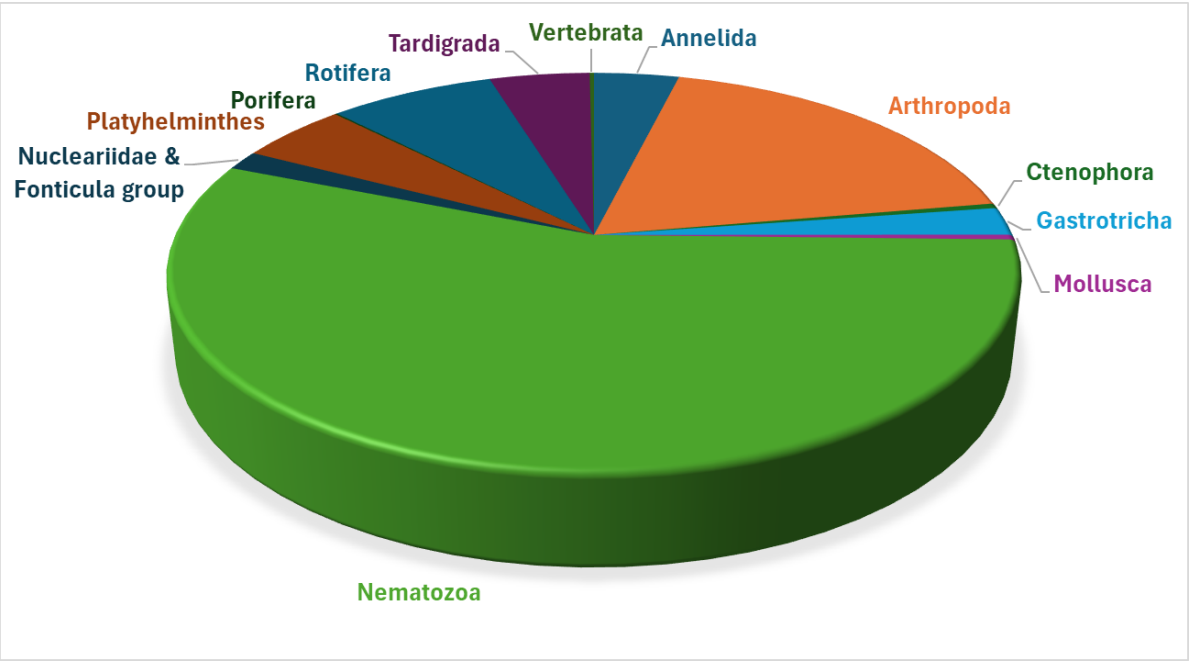


Figure 7. Phylum composition of reads assigned to the kingdom Animalia.

The kingdom Chromista comprises a diverse range of eukaryotic organisms ranging from photosynthetic organisms such as algae to unicellular parasites of arthropods. The taxonomy of organisms in this group is currently quite uncertain, and their position in the tree of life is fluid. In total, 14 phyla of Chromista were detected (Figure 8), with Apicomplexa and Ciliophora being by far the most taxon-rich groups with 302 and 288 ASVs, respectively. Apicomplexa are unicellular and are mainly obligate parasites of animals, while Ciliophora are unicellular ciliate organisms which can be free living, feeding on bacteria and other unicellular organisms, or parasitic. The next most taxon-rich groups were SAR (65 ASVs) which is a very diverse group of photosynthetic and non-photosynthetic organisms and Ochrophyta (59 ASVs) which are a group of mostly photosynthetic algae.

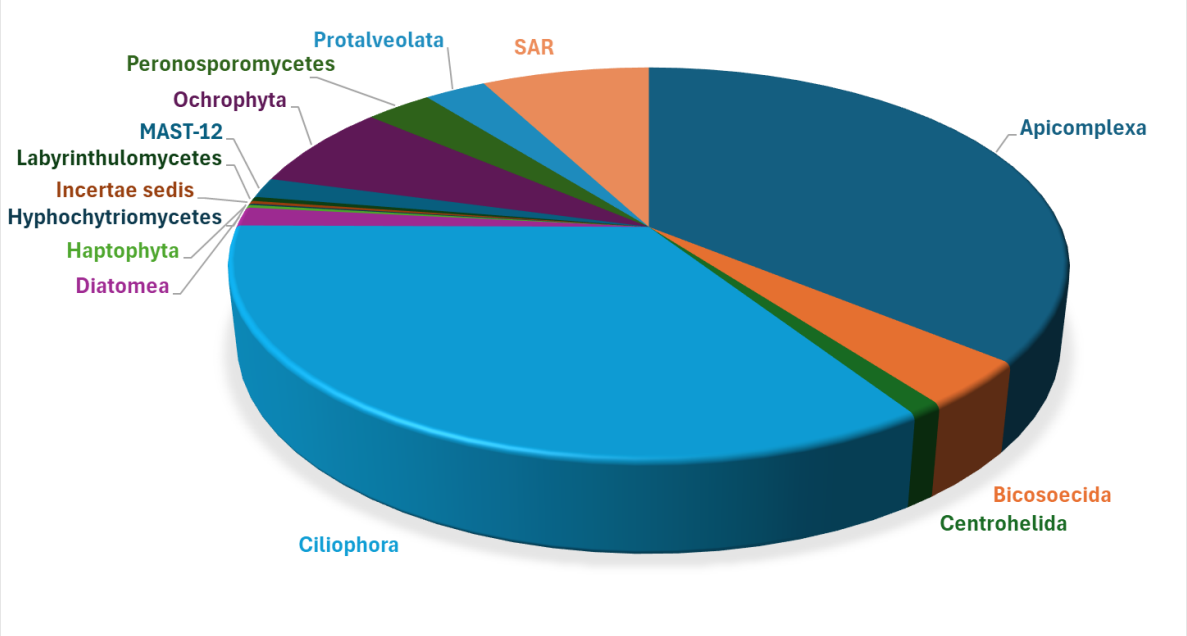


Figure 8. Phylum composition of reads assigned to the kingdom Chromista.

Four phyla of Plantae were detected in total (Figure 9). Phragmoplastophyta, which includes all vascular plants and bryophytes was the most taxon-rich phylum (96 ASVs) followed by Chlorophyta which contains the green algae (36 ASVs). Charophyta and Klebsormidiophyceae (both phyla of charophyte algae) had only 1 and 2 ASVs, respectively.

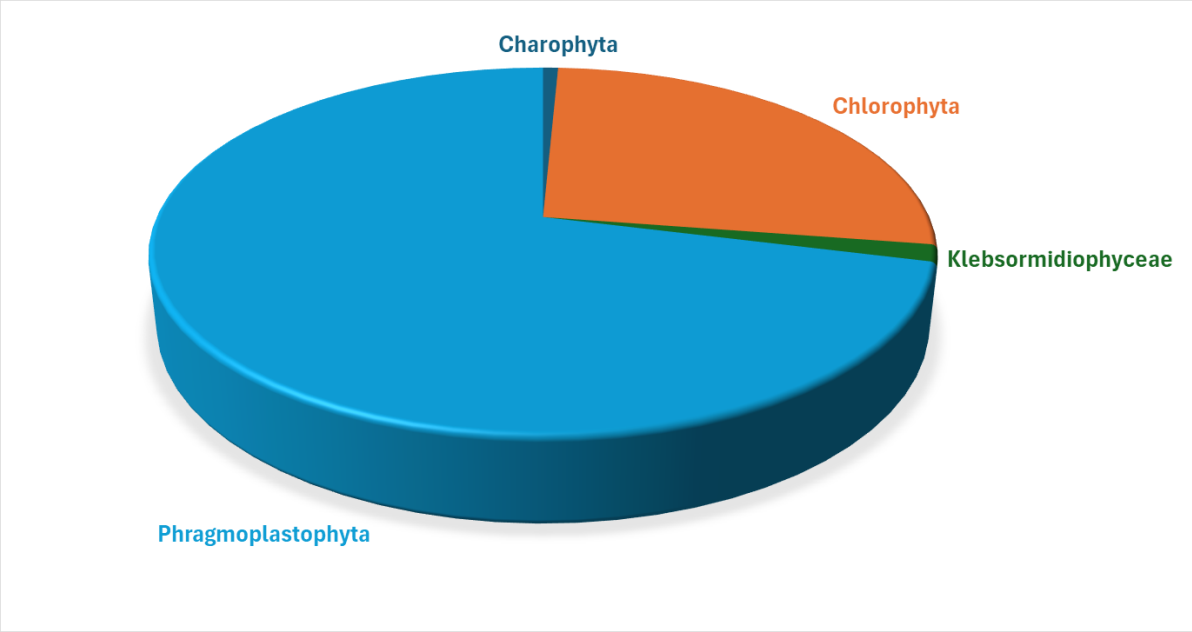


Figure 9. Phylum composition of reads assigned to the kingdom Plantae.

Protozoa are single-celled eukaryotic organisms which are either-free living or parasitic and feed on organic matter including other micro-organisms. The kingdom Protozoa was represented by 12 Phyla (Figure 10). Cercozoa (471 ASVs), Schizoplasmodiida (137) and Amoebozoa (115 ASVs) were the most taxon-rich phyla. Cercozoa are a diverse group of unicellular species which mainly feed on bacteria, although some species feed on fungi or are parasites on plants and soil invertebrates. Both Amoebozoa and Schizoplasmodiida are amoeba-type single celled organisms which occur on wet surfaces and feed on bacteria, fungal hyphae or other soil microorganisms.

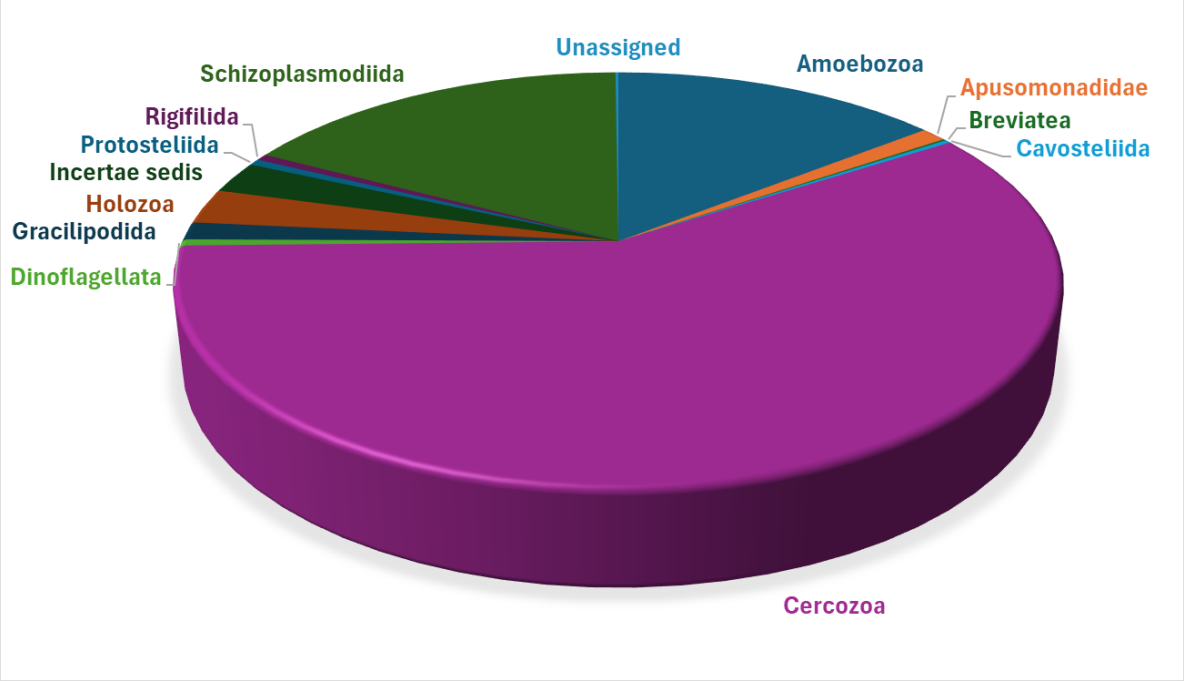


Figure 10. Phylum composition of reads assigned to the kingdom Protozoa.

Fungi are extremely abundant in soils and comprise a large part of the total species richness detected (Figure 3). Thirteen phyla of fungi were detected using the ITS2 marker (Figure 11), with an additional 424 OTUs defined only as fungi. Ascomycota (1909 OTUs) was by far the most taxon-rich phylum, followed by Rozellomycota (559 OTUs) and Basidiomycota (552 OTUs). Fungi play a wide range of ecologically important roles in soil, especially in carbon and nutrient cycling and turnover, and through formation of a variety of symbiotic relationships with plants. They can also be parasites or pathogens. The Ascomycota is the largest phylum of fungi and includes saprotrophs, pathogens and symbionts including lichen fungi and ericoid mycorrhizal fungi. The Basidiomycota is also a large and diverse phylum which notably contains fungi which form ectomycorrhizal associations with plants. Glomeromycota, the phylum of fungi which form arbuscular mycorrhizal associations with plants were also detected.

There were some notable species in the fungal community identified from Glen Prosen. Examples of these include: the ascomycete *Extremus caricis* which was only formally described in 2023, and is a new record to the UK; the ectomycorrhizal ascomycete *Wilcoxina mikolae*, also new to the UK; *Cordyceps militaris* which is a parasite of moth larvae, and the agaric *Tricholomopsis pteridicola* described in 2014 from Spain - this was found in 6 plots and is another first record for UK. The Butt rot fungus, *Heterobasidion annosum*, a serious pathogen of many conifers was also detected in 24 of the plots, most commonly in current or former plantations but somewhat surprisingly also in 7 of the 12 alpine plots. However, a closer examination of the ecology of *H. annosum* in the literature (Hansen & Knudsen 1997), from expert knowledge (Neville Kilkenny, pers. comm) and in our own previous eDNA metabarcoding studies (Britton & Taylor unpublished) found that the mycelium of *H. annosum* is actually widespread, and that the fruit bodies can also be found with deciduous trees and shrubs. In light of this, its occurrence in the alpine soils may therefore not be as unusual as first thought.

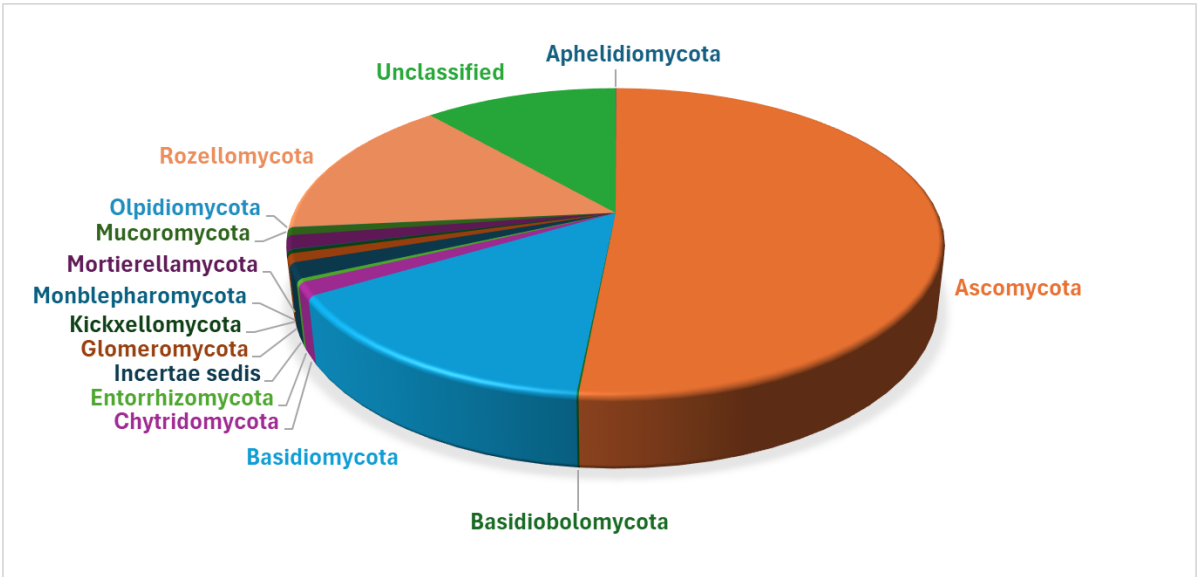


Figure 11. Phylum composition of reads from the ITS2 marker assigned to the kingdom Fungi.

Variation in soil community richness and composition across land use categories

The number of taxa from each kingdom that was detected in each soil sample was quite variable (Figures 12 & 13) and for most kingdoms there was no clear pattern of richness differences between land use categories. Significant differences in taxon richness between land use categories were only found for Chromista (Figure 13d, P=0.0201) where the OR category had greatest richness and Plantae (Figure 13e, P=0.0084) where richness was greatest in the OR and PP categories.

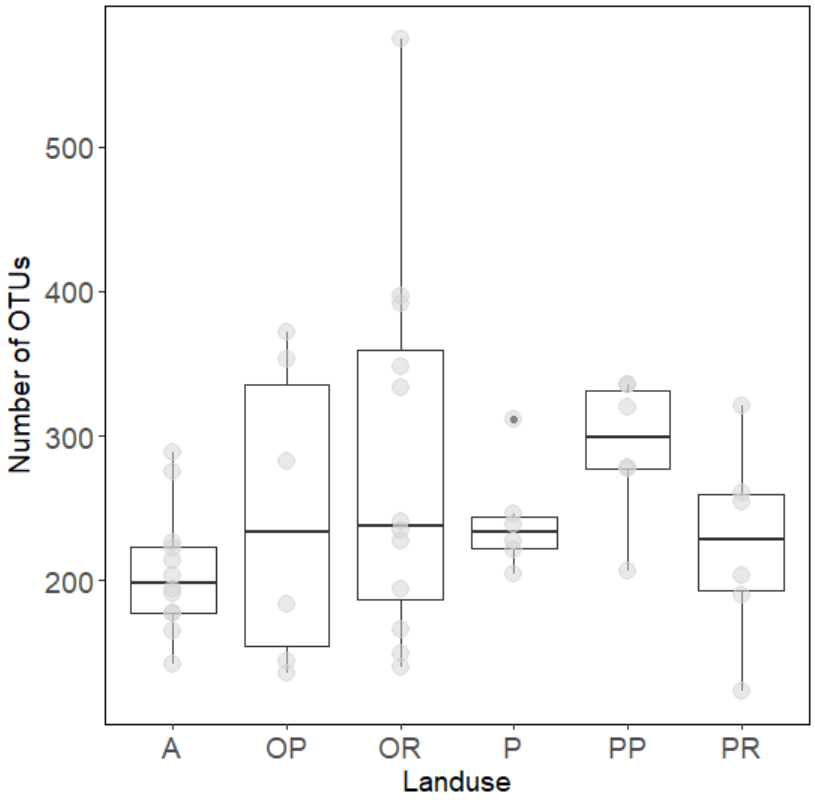


Figure 12. Taxon richness (number of OTUs) by land use category for kingdom Fungi based on ITS2 data. Boxes show the median and interquartile range and whiskers show the 10th and 90th centile for each category. Points show data for individual samples. See Table 1 for explanation of land use category codes.

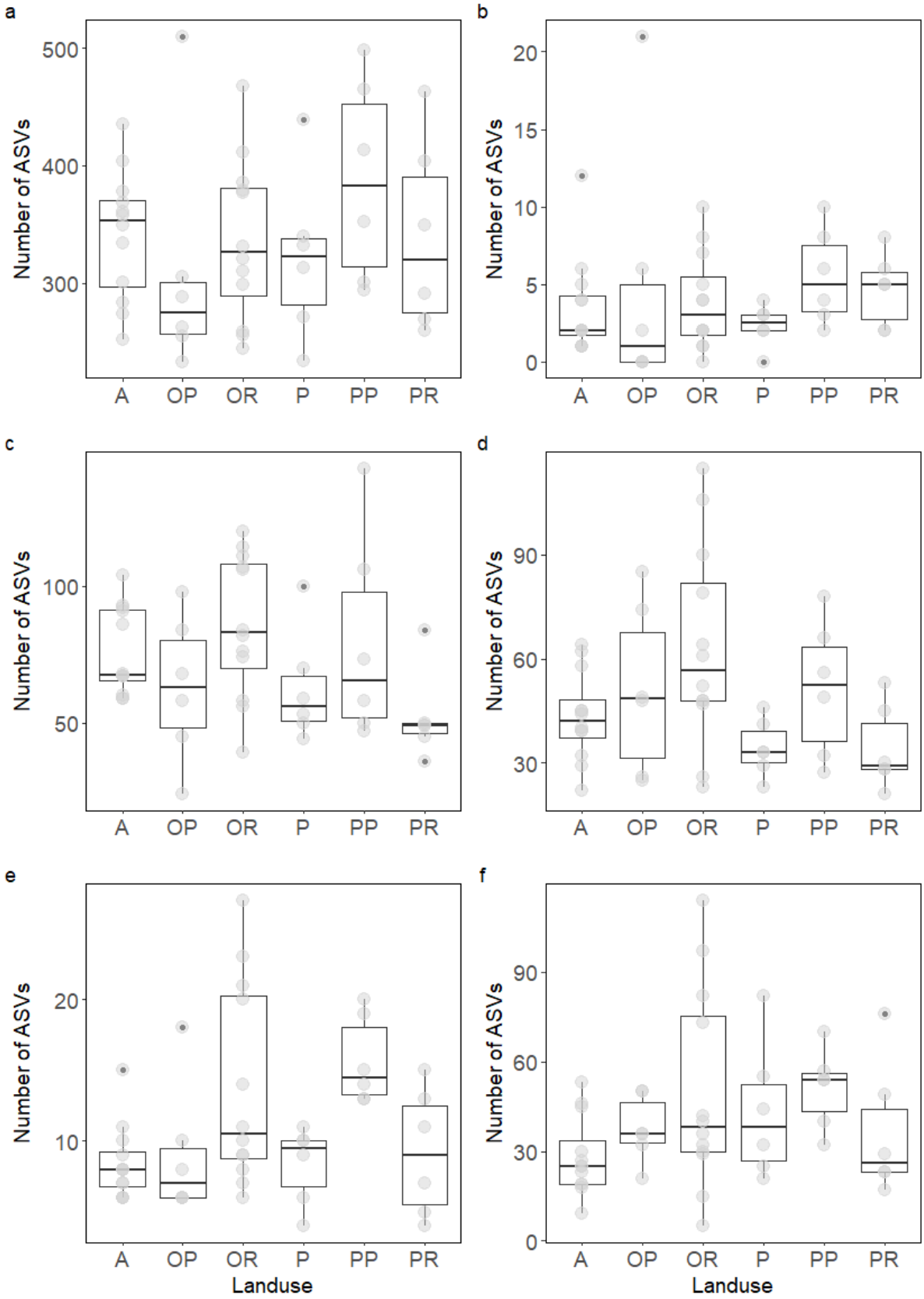


Figure 13. Taxon richness (number of ASVs) by land use category for a) Bacteria, b) Archaea, c) Animalia, d) Chromista, e) Plantae, f) Protozoa. Boxes show the median and interquartile range and whiskers show the 10th and 90th centile for each category. Points show data for individual samples. See Table 1 for explanation of land use category codes.

While there were few clear differences in kingdom richness between the land use categories, multivariate analysis of data for each kingdom showed clear patterns of differentiation of community composition between land uses. Patterns were similar for all seven kingdoms, with the primary contrast being that soil communities from open habitats (A, OP and OR) were differentiated from those currently or previously supporting trees (P, PP and PR). The differences between land use types were highly significant, with P<0.0001 for all kingdoms except Archaea (P=0.0021).

The differences in community composition were most pronounced for the fungi (Figure 14). On the first axis, plots in open habitats in riverside and valley bottom locations are separated from those on the mid and upper slopes. On the second axis the open habitat plots primarily dominated by ericaceous heath are separated from the currently and formerly wooded plots where trees have been present.

Community composition is also shown for the Bacteria and Animalia kingdoms (Figures 15 & 16). Although the pattern for these kingdoms is not as distinct as in the fungi, there is a clear differentiation between wooded and non-wooded land use types for both communities and this pattern was seen consistently across all other kingdoms.

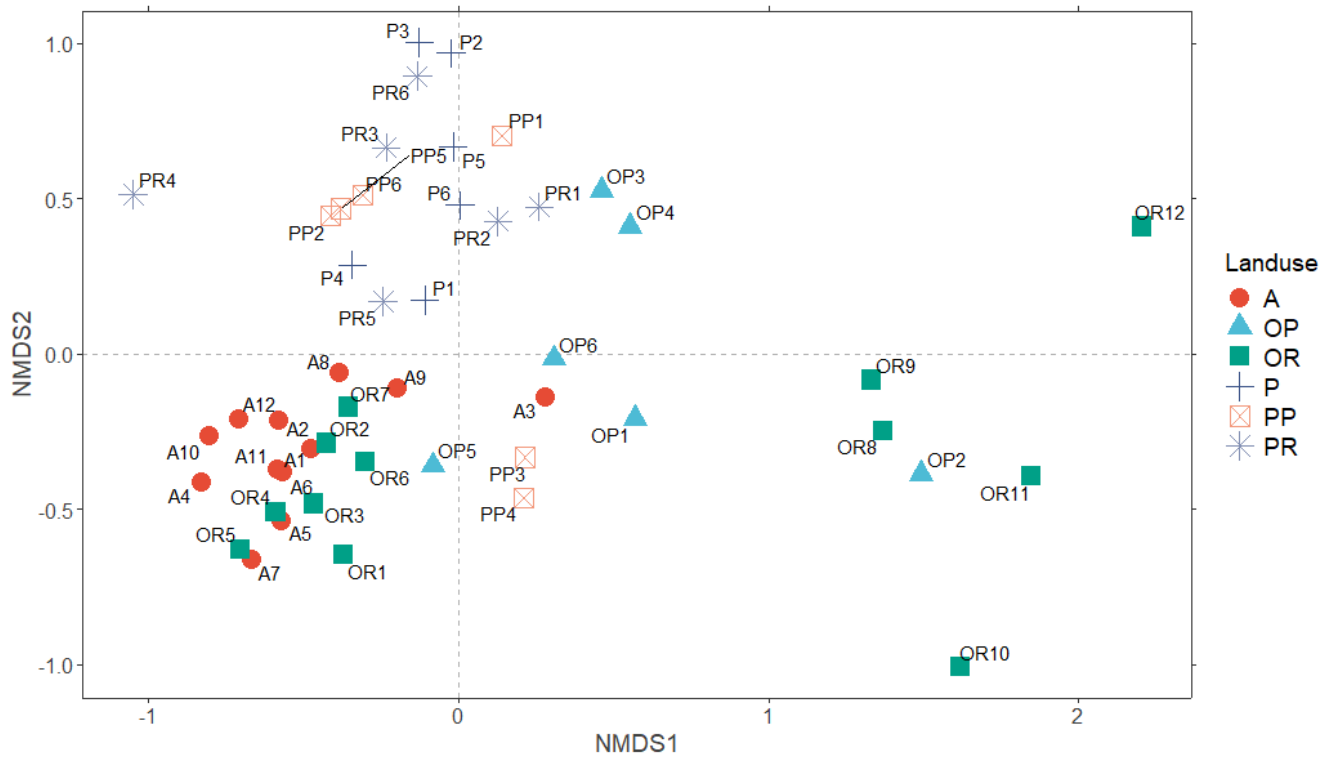


Figure 14. Non-metric Multi-Dimensional Scaling (NMDS) of Fungi community composition. NMDS model has 2 axes, stress = 0.13.

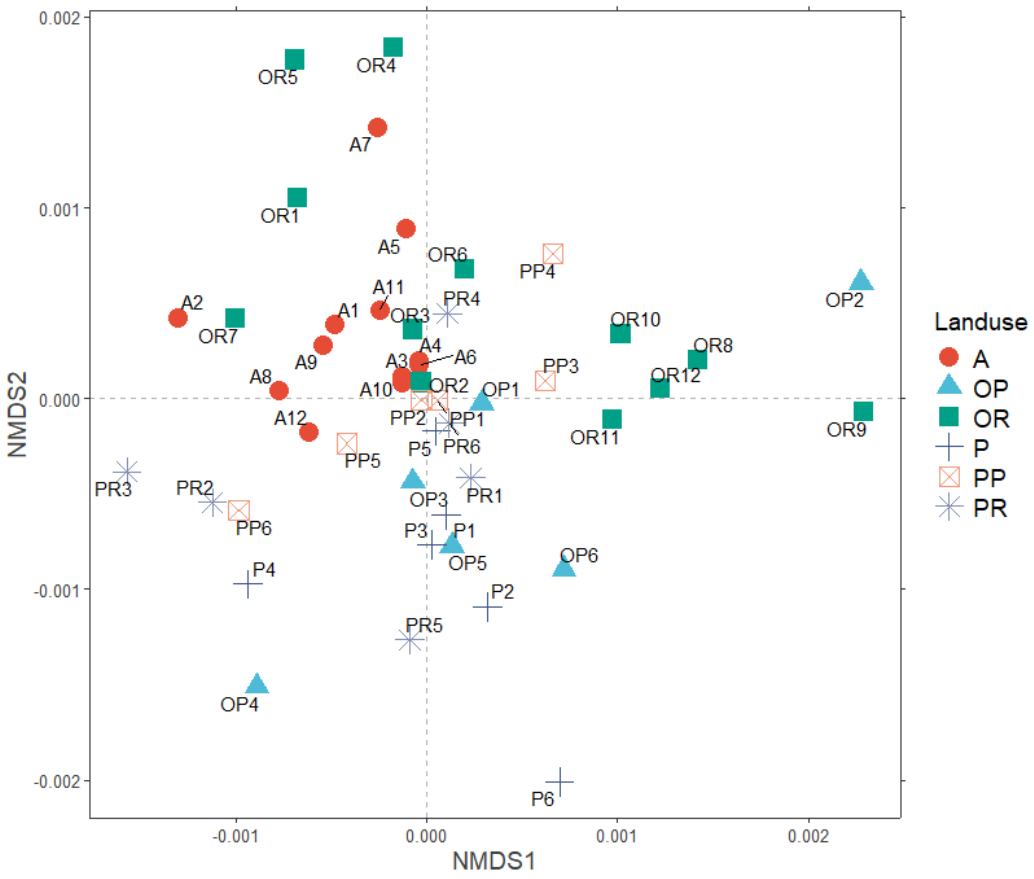


Figure 15. Non-metric Multi-Dimensional Scaling (NMDS) of Bacteria community composition. NMDS model has 3 axes (axes 1 & 2 shown), stress = 0.19

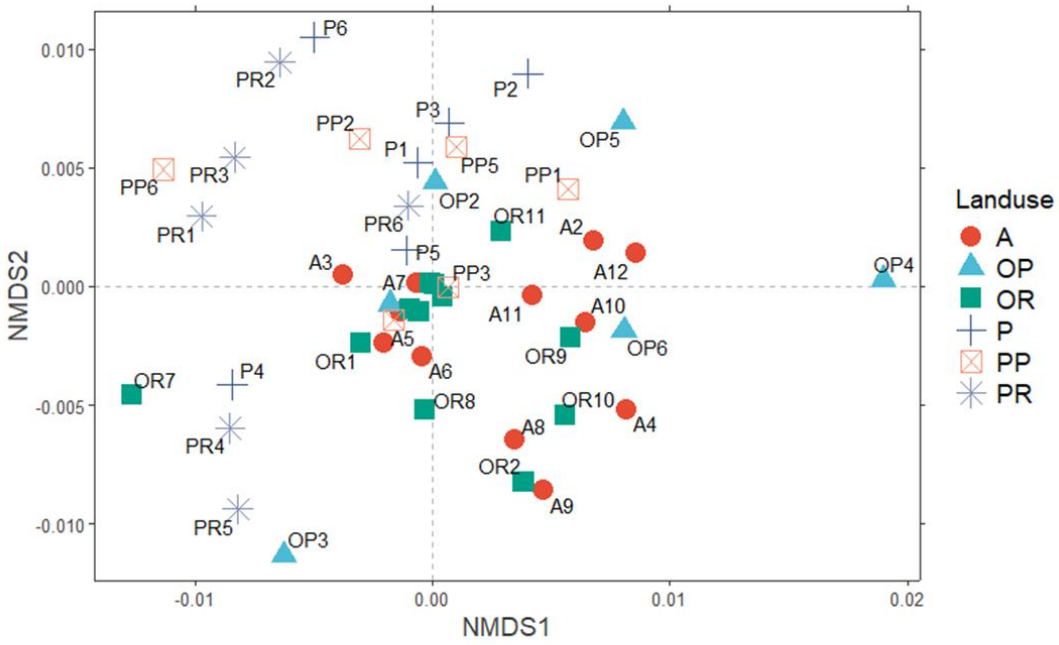


Figure 16. Non-metric Multi-Dimensional Scaling (NMDS) of Animalia community composition. NMDS model has 3 axes (axes 1 & 2 shown), stress = 0.20

Relationships between soil biodiversity and the environment

The environmental setting

The soil sampling scheme used at Glen Prosen was based on the likely shifts in land use expected to occur during ecological restoration activities in the glen. Within some of these land use categories the vegetation composition (which would be expected to influence soil biodiversity) was found to be quite variable. The results of the basic vegetation survey carried out during soil sampling (Figure 17) show a distinction on Axis 1 between open ground habitats dominated by ericaceous dwarf-shrubs (on the left of the diagram, including most alpine plots 'A' and a majority of those identified as open with natural regeneration 'OR') and those dominated by graminoids and forbs (most open areas identified for planting 'OP', and the remaining OR plots, generally in valley bottom locations). On Axis 2 most of the plots from current or former areas of plantation (P, PP and PR) are separated from the open habitats and are associated with greater cover of trees, litter (including brash and stumps), bare ground and ferns.

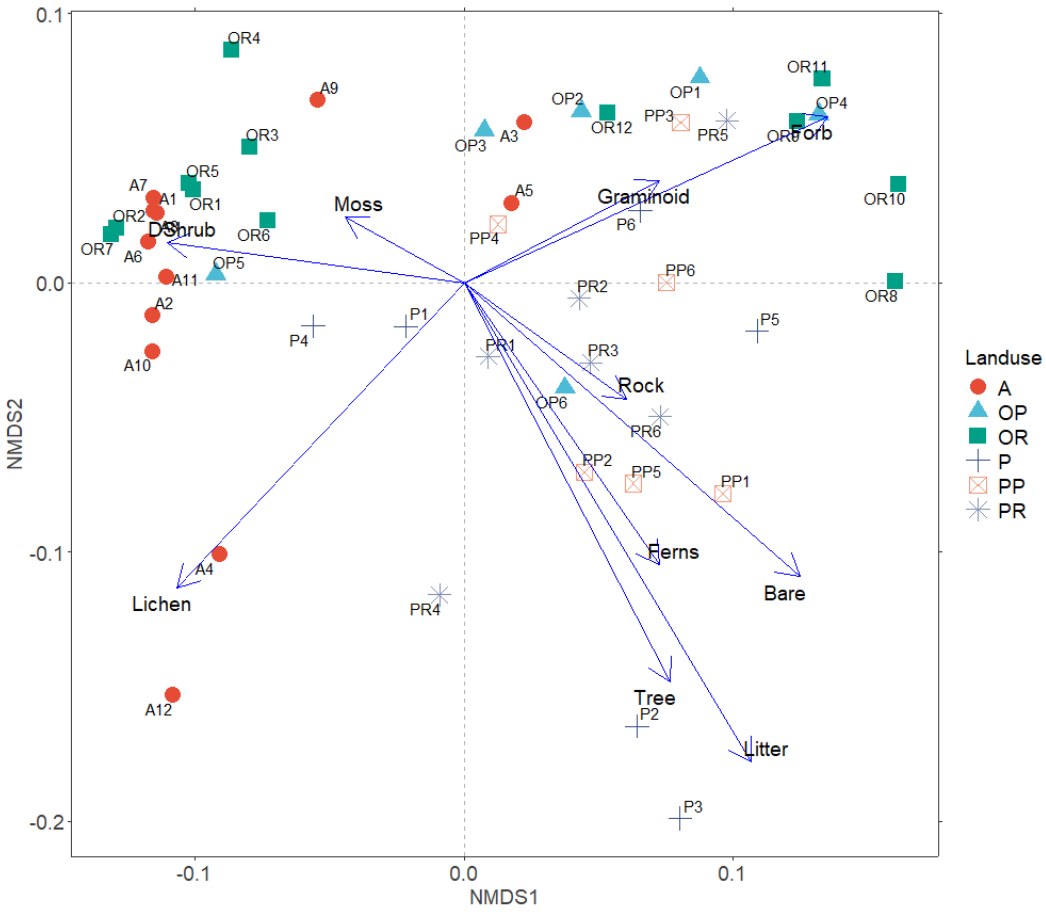


Figure 17. Non-metric Multi-Dimensional Scaling (NMDS) of vegetation community composition. Symbols show vegetation composition of sample plots, arrows indicate the location of individual 'species' within the ordination. NMDS model has 2 axes, stress = 0.13.

In addition to the variation in vegetation composition, the land use categories also differ in terms of their topography and soil chemistry (Figures 18 & 19). Alpine plots occurred at a significantly higher elevation (650-850 m asl) than the other land use categories (OP and OR range between 350 – 650 m asl, while P, PP and PR are focussed between 400-500m asl), but there were no consistent differences in slope or aspect.

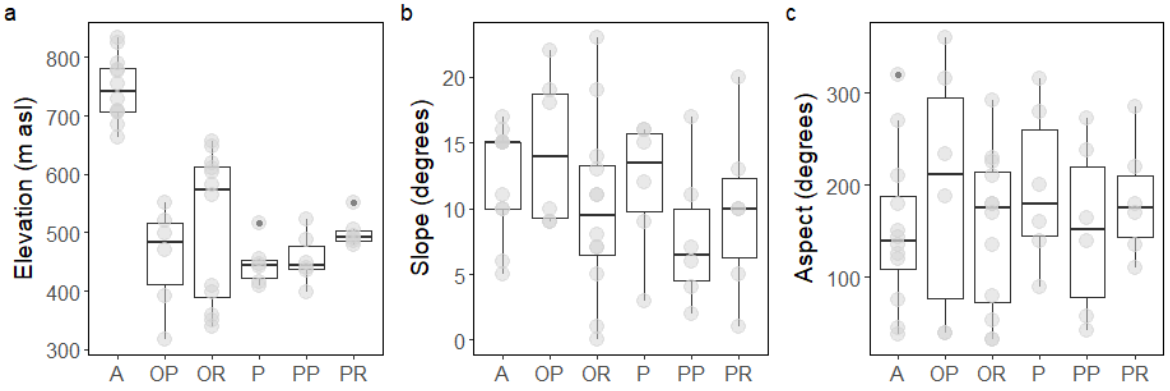


Figure 18. Topographic characteristics of each land use category; a) elevation, b) slope, c) aspect. Boxes show median and inter-quartile range, whiskers show 10th and 90th centiles, points show individual sample values.

Soil chemistry varied between land use categories, with significant differences seen in all of the soil chemical parameters measured (Table 2, Figure 19). Soil from the OP, P and PR plots had lower carbon and nitrogen content than soils from the A, OR and PP plots and the C:N ratio was generally higher in alpine (A) plots compared with all the other groups. Soil pH varied between 4 and 6 and was variable both within and between groups. Plots with high soil pH (5.25 and above) were found in the OP, OR and PP groups. Alpine (A) and retained plantation (P) groups had fairly uniform and low pH of around 4.5. Correlations between the topographic and soil chemistry variables (Figure 20) showed that soil carbon and nitrogen content and soil C:N ratio were all positively correlated with elevation, while soils with low carbon and nitrogen content, low C:N and high pH were found at low elevation in valley bottom locations.

Table 2. Summary of significance of differences in environmental parameters between land use categories (based on ANOVA).

Environmental factor	F value	P value
Elevation	20.348	<0.0001 ***
Slope	1.1661	0.3418
Aspect	0.4678	0.7980
Soil %N	4.2735	0.0031 **
Soil %C	4.7452	0.0016 **
Soil C:N	2.5951	0.0392 *
Soil pH	2.4944	0.0458 *

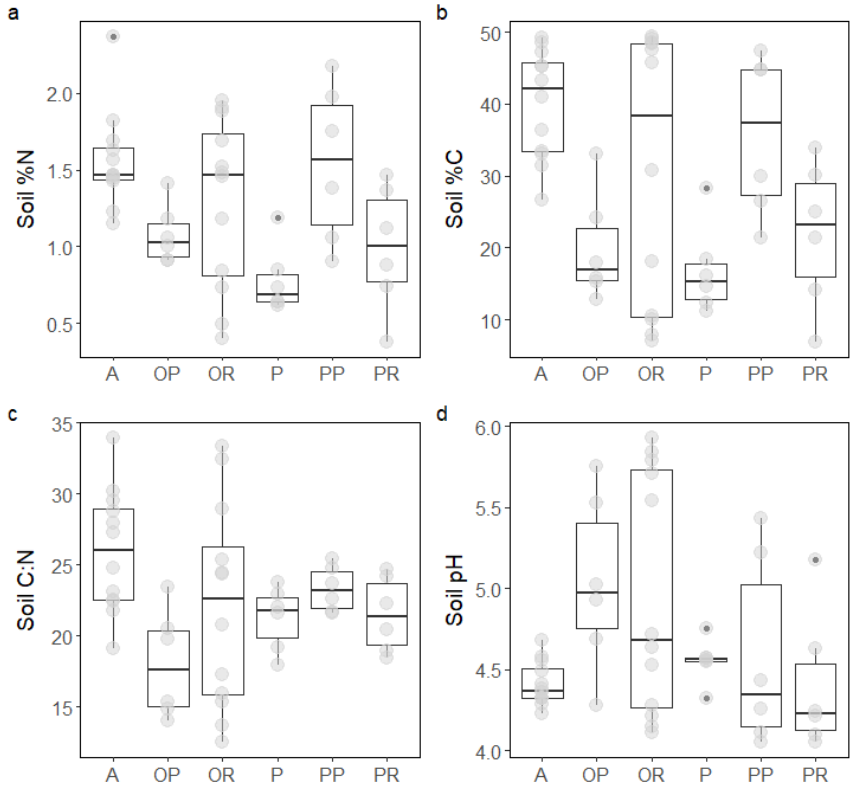


Figure 19. Differences in a) soil %N, b) soil %C, c) soil C:N, d) soil pH among land use categories. Boxes show median and inter-quartile range, whiskers show 10th and 90th centiles, points show individual sample values. See Table 2 for significance of differences between categories.

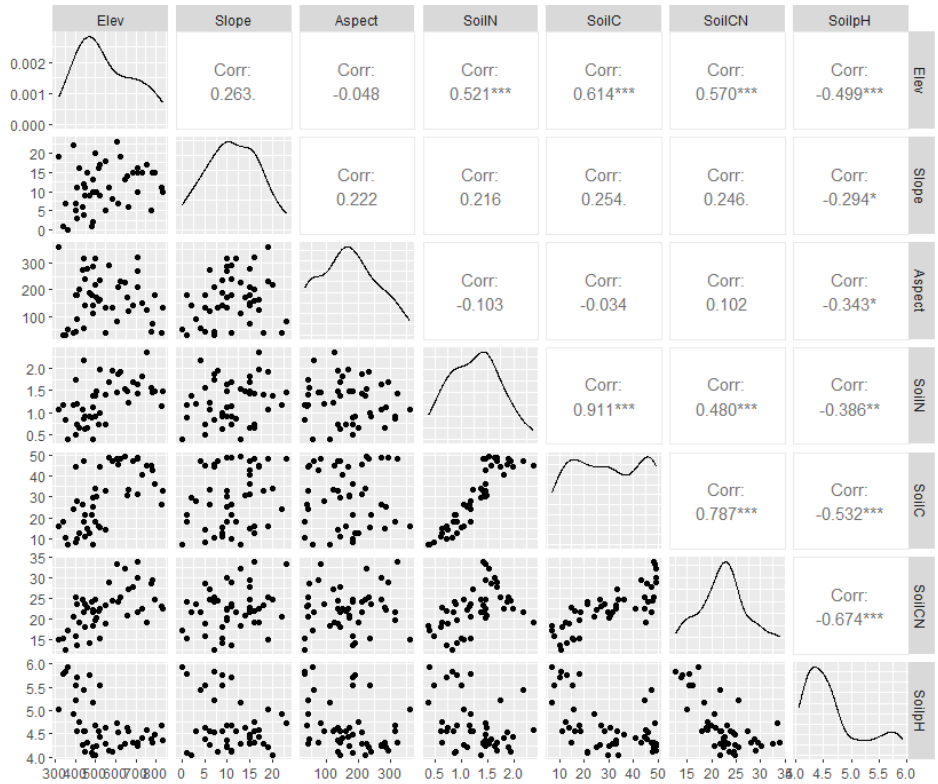


Figure 20. Correlations between topographic and soil chemical parameters across all sample sites at Glen Prosen.

Soil biodiversity response to the environment

Soil community taxon richness

Looking across all of the survey plots together, the taxon richness of each kingdom showed some clear responses to topography and especially soil chemistry (Table 3). Bacteria were the only group where taxon richness was not affected by any of the environmental parameters. Archaea were similarly insensitive to most parameters but were significantly affected by soil pH, being the only group to show a negative response to this parameter, perhaps reflecting their preference for acidic soil conditions. The richness of fungi showed a decline with increasing elevation and fungal richness was also lower in soils with high carbon content and C:N ratio. Fungal richness was positively associated with soil pH. Animalia were the only group to have higher richness in more organic soils, as shown by their positive relationship with soil C and N content, but their richness was unaffected by topography or by soil pH. Chromista, Plantae and Protozoa all had similar negative richness relationships with elevation and soil C:N ratio and positive relationships with soil pH. Plantae were the only group to have a relationship between taxon richness and slope, possibly resulting from greater richness in the more gently sloping plots found in the valley bottom. Protozoa richness showed a negative relationship with soil C and N content indicating greater richness in the more mineral (and higher pH) soils found in the valley bottoms. Some of these soils were also noted to be wetter than those on slopes and summits which could favour Protozoa which utilise water films in soils.

Table 3. Summary of the significance of effects of environmental factors on the richness of taxa from each kingdom found in each plot (P values based on Generalized Linear Models). Significant effects shown in bold. Blue text indicates a negative relationship, red text indicates a positive relationship.

Kingdom	Elev. (masl)	Slope (°)	Aspect (°)	Soil %N	Soil %C	Soil C:N	Soil pH
Bacteria	0.6957	0.2216	0.8639	0.1139	0.2914	0.9775	0.8610
Archaea	0.5998	0.9314	0.1823	0.1074	0.1347	0.2682	**0.0226**
Fungi	**<0.0001**	0.0595	0.5009	0.2102	**0.0116**	**<0.0001**	**<0.0001**
Animalia	0.7766	0.8491	0.9978	**0.0054**	**0.0033**	0.1457	0.1084
Chromista	**0.0212**	0.1101	0.0600	0.8043	0.2644	**0.0034**	**<0.0001**
Plantae	**<0.0001**	**<0.000**	0.0854	0.2098	0.0577	**0.0075**	**<0.0001**
Protozoa	**<0.0001**	0.0700	0.1040	**0.0268**	**0.0017**	**<0.0001**	**<0.0001**

Soil community composition

The taxon composition of the soil communities in each plot were also strongly influenced by topographic and soil chemical parameters (Table 4). Statistical models of community composition were constructed for each kingdom to test which environmental parameters significantly influenced taxon distributions. Soil pH had the greatest influence on soil biodiversity, being significantly linked to taxon composition across all seven kingdoms. Acidic soils supported fewer species. Elevation also significantly influenced community composition in all kingdoms except Archaea. Slope had no impact on soil communities for any kingdom, but aspect significantly influenced Bacteria, Archaea, Animalia and Chromista, this could reflect an effect of unmeasured parameters such as soil temperature regimes which could be influenced by aspect.

Table 4. Summary of significance of effects of environmental factors on taxon composition in each Kingdom. P values are shown for each significant factor in Canonical Correspondence models of community composition, based on permutation tests with 999 permutations. Blank cells indicate a parameter is non-significant and not included in final model. % Var. shows the total percentage of variation in community composition explained by the model. All models were significant at P<0.001.

Kingdom	Elev. (masl)	Slope (°)	Aspect (°)	Soil %N	Soil %C	Soil C:N	Soil pH	% Var.
Bacteria	0.001		0.009	0.001	0.016	0.001	0.001	18.13
Archaea			0.033				0.002	9.93
Fungi	0.001				0.011	0.026	0.001	13.37
Animalia	0.001		0.003	0.007	0.016		0.001	18.93
Chromista	0.001		0.003	0.043			0.001	13.65
Plantae	0.001					0.008	0.001	12.13
Protozoa	0.001					0.027	0.002	9.68

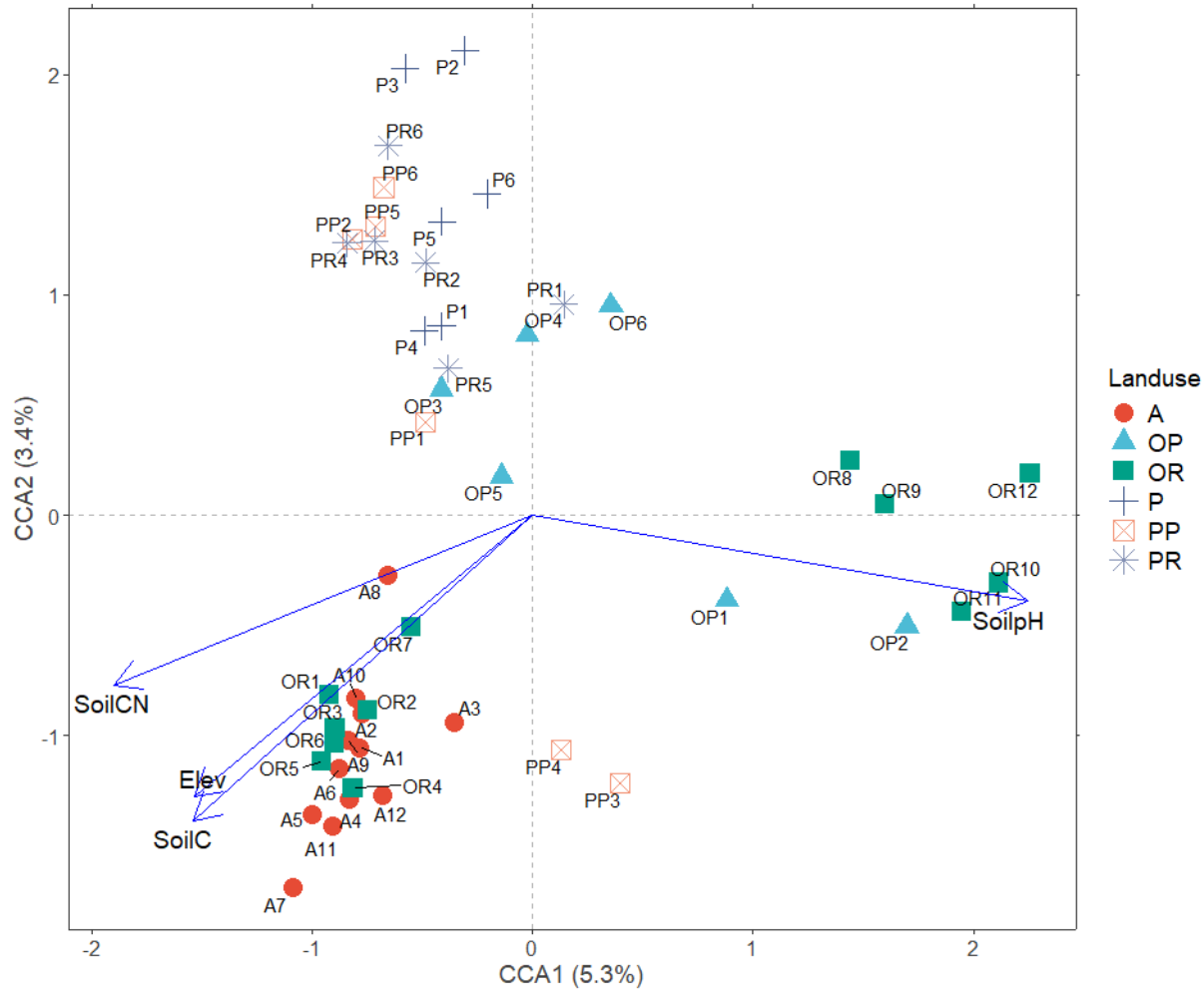


Figure 21. Canonical Correspondence Analysis (CCA) of soil fungal community composition. Symbols represent fungal communities in sample plots and arrows show the influence of environmental drivers. See Table 4 for the significance of individual drivers and the total amount of fungal community variation explained by the model.

Each kingdom responded differently to soil carbon and nitrogen content and soil C:N (Table 4). Despite bacterial community richness being insensitive to soil chemical parameters, the composition of bacterial communities was significantly influenced by both C and N content, C:N ratio and pH. Archaea communities in contrast were only influenced by soil pH. Soil C:N ratio had a significant influence on community composition of Fungi, Plantae and Protozoa, while N content influenced Animalia and Chromista and C content influenced Fungi and Animalia.

The model of environmental variable influences on fungal community composition is shown in Figure 21. Communities in alpine plots (A) are strongly clustered together at the lower left, together with some plots from open habitats earmarked for natural regeneration (OR). This group is associated with higher elevations and higher soil C content and C:N ratio. A second cluster of plots on the right comprises the lower elevation plots with high soil pH in the OP and OR categories. The communities from plots on the current and former plantations form a third distinct group associated with low pH, but with contrasting composition to the higher elevation alpine plots. Similar patterns of plot groupings are seen in the models for all other kingdoms (examples are shown for Animalia and Protozoa, Figures 22 & 23). All of the models are dominated by the gradient in community composition associated with soil pH but also show varying degrees of distinction between plots from open and currently/previously forested habitats associated with differences in soil chemistry and elevation.

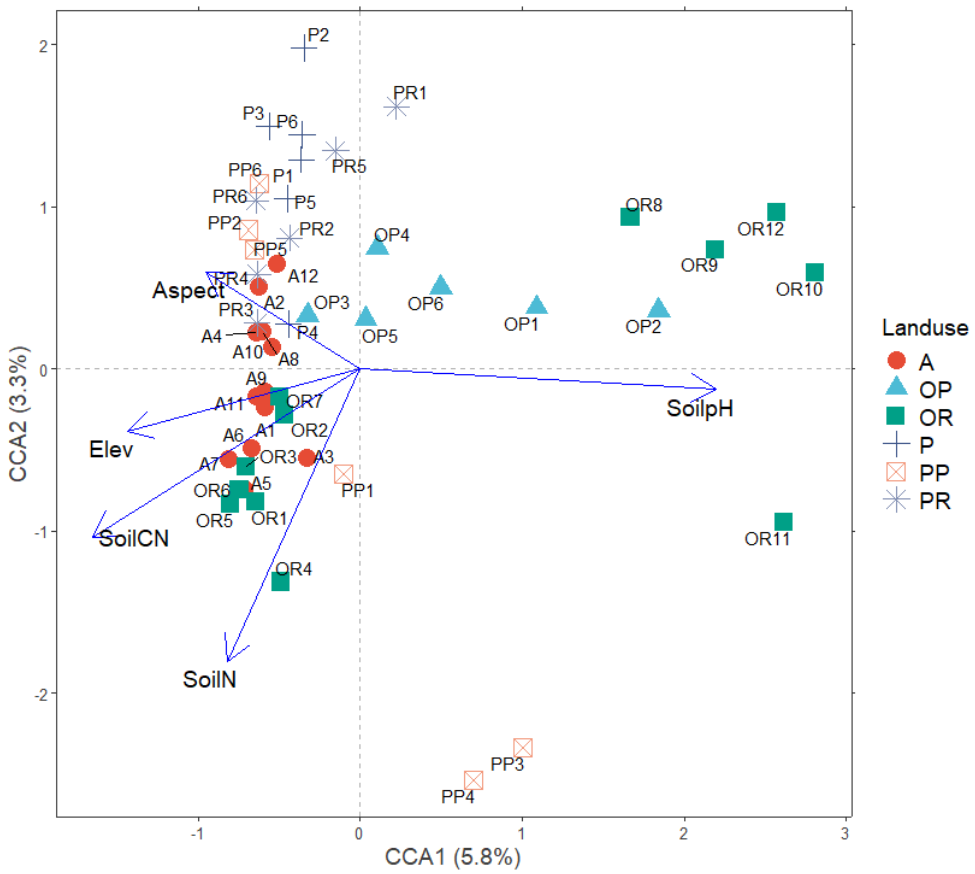


Figure 22. Canonical Correspondence Analysis (CCA) of Animalia community composition. Symbols represent animal communities in sample plots and arrows show the influence of environmental drivers. See Table 4 for the significance of individual drivers and the total amount of animal community variation explained by the model.

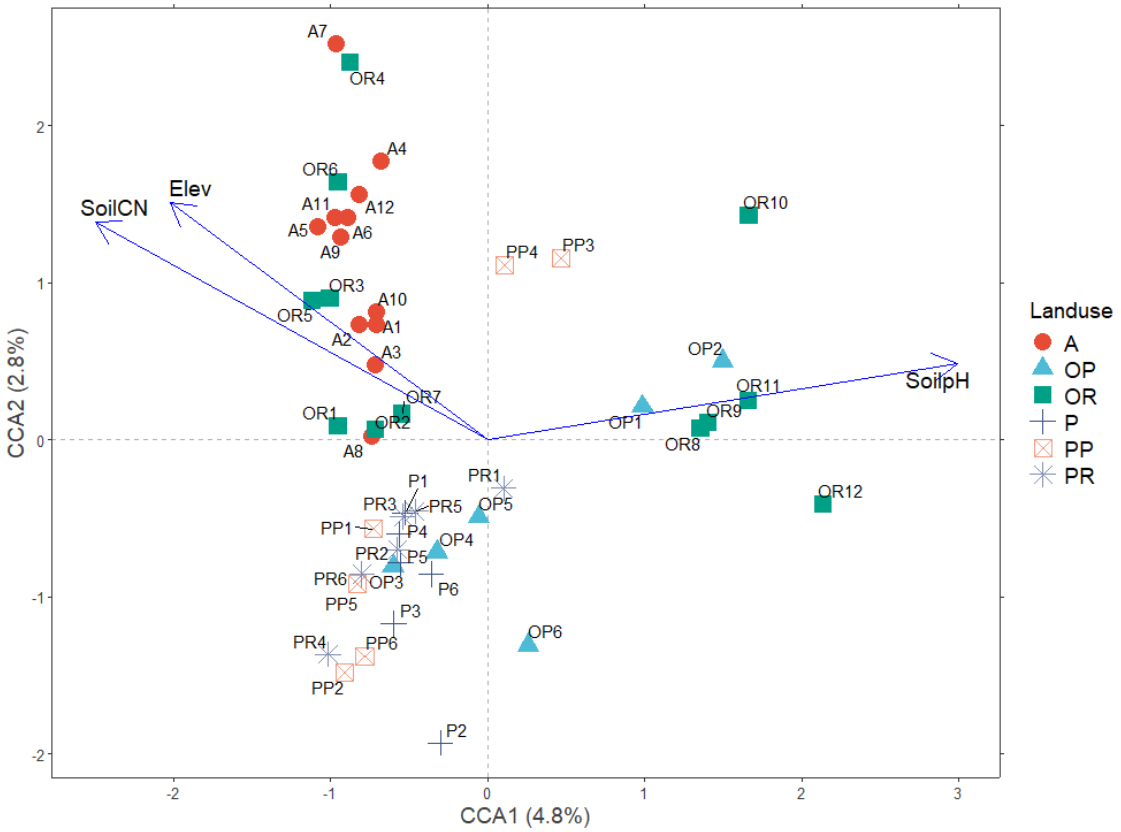


Figure 23. Canonical Correspondence Analysis (CCA) of Protozoa community composition. Symbols represent protozoan communities in sample plots and arrows show the influence of environmental drivers. See Table 4 for the significance of individual drivers and the total amount of protozoan community variation explained by the model.

Focus on plant root associated fungi

Significance of mycorrhizal associations

Soil fungi carry out a wide range of functions in natural ecosystems. They are the major recyclers of organic matter, breaking down complex organic molecules and particularly recalcitrant plant polymers like cellulose and lignin. Many are parasites or pathogens of different organisms including plants, animals and other fungi but, importantly, a select group of fungi form symbiotic associations with plant roots and are critical for plant survival and nutrient uptake. The latter group includes a group of soil fungi called mycorrhizal fungi that form obligate associations with virtually all terrestrial plants including trees. The three main mycorrhizal associations relevant to the growth of trees at Glen Prosen - arbuscular mycorrhizas (AM), ectomycorrhizas (ECM) and ericoid mycorrhizas (ERM) - are described in Figure 24.

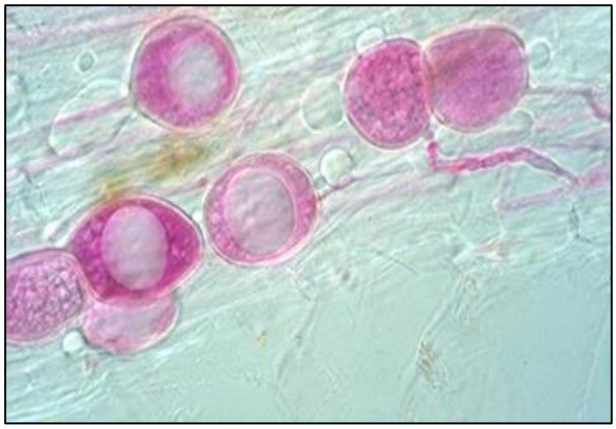
Many of our native trees in Scotland including Pine, Birch, Oak, Hazel and Lime form ECM associations. A smaller number of trees form AM associations, including Rowan, Holly and Elm. Willows and Aspen are unique in that they can form both ECM and AM associations. The fungal species involved in each of the major mycorrhizal associations are unique to that association and can only associate with certain plant taxa. The obligate nature of mycorrhizal associations means that ECM or AM trees would find it difficult to colonise heathlands dominated by ericaceous plants and ericoid mycorrhizal fungi.

Coevolution of plants and their mycorrhizal fungi over the past 400 million years has resulted in many ecosystems where a particular type of mycorrhizal association dominates and where conditions favouring that association are maintained and enforced by above/below ground interactions. This can have serious consequences for attempts to change land use (e.g. from heathland to woodland) as ecosystems dominated by one association can hinder the establishment of another. Due to the obligate nature of most mycorrhizal associations, both symbionts (plant and fungus) must be present in order for either partner to have any chance of surviving.

Plant root associated fungi at Glen Prosen

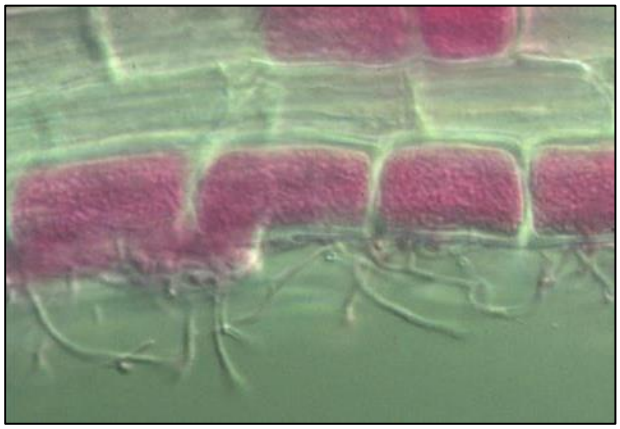
Investigation of the distribution of plant root associated fungi across the land use categories at Glen Prosen gives an indication of the areas where natural regeneration of trees is likely to be successful (e.g. where there is a legacy of ECM fungi formerly associated with plantations) and areas where natural regeneration of trees may take a long time due to a lack of suitable fungal partners.

Fungal OTUs belonging to four groups of root-associated fungi were identified from the ITS2 fungal dataset. These were the AM, ECM and ERM mycorrhizal species and the Archaeorhizomycetes. The Archaeorhizomycetes are an ancient group of root-associated fungi but were only recently discovered and described. They have been found to be globally distributed and are likely to be important in plant ecology, but the exact nature of their relationship with the plants is not yet clear.



Arbuscular mycorrhizal (AM) associations: the fungi invade the root and penetrate the cells forming vesicles (shown) and tree-like structures called arbuscules, which are the sites of transfer of nutrients and organic carbon. This is by far the most common association and the most ancient. Often associated with more nutrient rich soils where phosphorus may be limiting. Plants often with more easily decomposed litter.

Ectomycorrhizal (ECM) associations: the fungi colonise fine feeder roots, completely covering the surface with a mantle of hyphae. This may be hydrophobic, effectively isolating the root from the soil. Hyphae penetrate between the root cells, forming a net-like structure – the Hartig net, which is the site of transfer. Hosts are usually long-lived woody perennials (trees), occurring on soils where organic nutrient sources are important.



Ericoid mycorrhizal (ERM) associations: the fungi (shown in pink) invade the cells of the fine hair-roots, forming dense coils. Formed on ericaceous plants, whose litter is usually very recalcitrant, containing high levels of phenolics. However, the ERM fungi can produce a wide range of enzymes capable of degrading the litter. Plants with ericoid mycorrhizas often form mono-dominant stands on organic soils.

Figure 24. The three main mycorrhizal associations important for woodland expansion

The mean richness of the root associated fungal groups in each of the land use categories is shown in Figure 25. Archaeorhizomycetes are the most species-rich group across all of the land use categories with up to 35 OTUs per plot. The three main groups of mycorrhizal species are less species rich but show clear differences in their distributions between the land use categories. Arbuscular mycorrhizal fungi are almost absent from the alpine (A) and retained plantation (P) plots with both categories having only one plot where these species are present. The greatest richness of AM species is seen in the OR category which includes the valley bottom plots on higher pH soils. Ericoid mycorrhizal species were found in all of the alpine (A) plots and all of the current/former plantation plots (P, PP, PR). They were less frequent in the OP and OR categories where they were present in about half of the plots.

A total of 44 ECM fungi were found in the samples collected from Glen Prosen, with their distribution and frequency amongst the land use categories as follows: Alpine 5/12 plots and 3 ECM taxa; OR 4/12 plots and 4 ECM taxa; OP 0/6 plots; PP 6/6 plots with 9 ECM taxa; PR 5/6 plots with 27 ECM taxa, and P 6/6 plots with 32 ECM taxa. The occurrence and frequency of the ECM fungi was therefore very tightly related to the current or past presence of conifer plantations.

The three ECM fungi found in the alpine plots were common components of the plantation ECM fungal communities, suggesting that their occurrence in the alpine plots may be influenced by the availability of local inoculum (spore rain). There are a number of potential plant hosts which do occur in the alpine zone at Glen Prosen, including dwarf willow species, alpine bistort and Mountain avens, but none of these hosts were recorded in any of the vegetation descriptions from the survey. In addition, if the roots of one of these ECM plants had been included in one of the samples, a wider diversity of ECM fungi would be expected in the alpine samples, as alpine ECM plants can all harbour species-rich ECM communities. It is possible that the three ECM fungi detected in the alpine plots could be present as spores in the soil. However, if this was the case, then it would be logical to assume that they would have also been detected in more of the plots in OP and OR areas below 600m. At the moment, the presence of these three ECM fungi in alpine samples is unexplained.

The presence of ECM fungi in OP and OR plots was very rare and restricted to the 4 OR plots located along the bottom of the glen, where 3 plots had a single taxon and 1 plot had two ECM fungi. Again, no ECM host plants were recorded in the vegetation survey, but the vegetation in these plots was quite rank, and it is possible that there may have been small willow plants hidden amongst the dense vegetation. The virtual absence of ECM fungi in currently open areas designated for woodland establishment would suggest that the availability of suitable ECM fungi to colonise non-mycorrhizal planted or naturally occurring tree seedlings is low, and that tree establishment will potentially be patchy and for the natural regeneration areas, very slow.

Unsurprisingly, the ECM species exhibited greatest richness in the P and PR plots which are in the areas of current plantation forest and in the upper reaches of Glen Prosen Forest where most trees have only recently been felled. All the P plots and five out of six PR plots had ECM fungi present – the only PR plot without ECM fungi appeared to have been felled/windblown some time ago. A lower diversity of ECM species was seen in the PP plots which comprise areas of former plantation, some of which were felled up to 25 years previously. All PP plots had at least one ECM present, some of these may have persisted on tree seedlings within the felled areas.

The great majority – 38 out of the 44 ECM fungi detected in Glen Prosen, were found exclusively in samples from existing or former plantations. Only 2 taxa occurred solely outwith these areas, and an additional 4 were also rarely recorded in other areas in addition to the plantations. The ECM fungi in the plantations are primarily generalists that can associate with a range of ECM host plants. The exceptions to this are the two *Suillus* species*, S. luteus* and *S. grevillei*, which can only associate with either pine or larch, respectively, and were found in plots with these host trees. It is interesting to note that amongst the 44 taxa, there were 3 truffle species, 1 cup fungus, 22 species which form

typical above ground mushrooms, and 18 which do not form any kind of macroscopic fruiting structure. The high proportion of the latter compared to the typical mushroom formers is unusual.

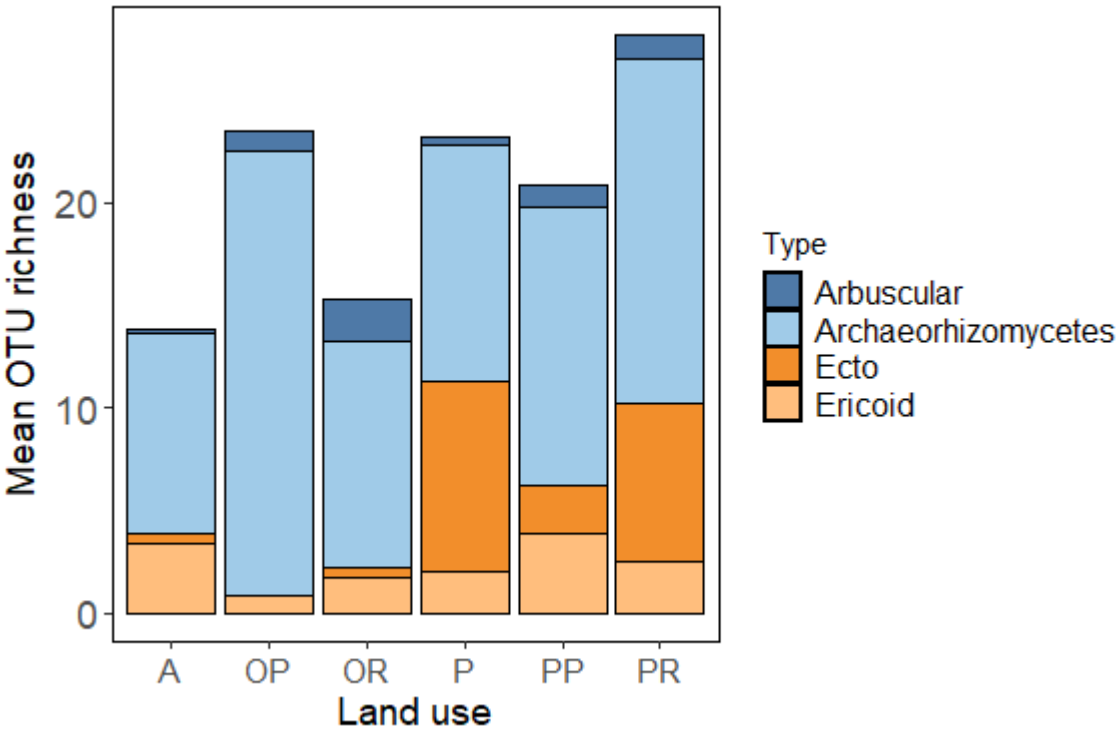


Figure 25. Mean richness per plot of root associated fungal groups by land use category.

Conclusions

This baseline soil biodiversity survey has successfully demonstrated the enormous richness of soil biodiversity present at Glen Prosen with 12,253 taxa from 7 kingdoms of life detected in only 48 samples. There will undoubtedly be many more soil taxa present in the glen, some of which could be captured by additional sampling, but the baseline survey is sufficient to reveal patterns of taxon distribution around the glen and responses of soil biodiversity to environmental conditions. Much of the biodiversity revealed by this study is not well known and our understanding of the functional roles of these organisms is still in its infancy.

Although there were few consistent and significant differences in soil organism richness across the land use categories, richness was strongly influenced by environmental factors, especially soil pH, elevation and soil C:N ratio. Each kingdom responded to a different combination of environmental drivers and drilling down to lower-level taxonomic groupings (e.g. orders and families) would undoubtedly reveal more details about the responses of soil biodiversity to environmental parameters.

Unlike richness, soil community composition was clearly related to the different land use categories, with clear distinctions between forested and unforested habitats and between open plots in alpine, mid slope and valley bottom locations. There were also strong influences of elevation and soil chemistry, especially soil pH. The influence of soil pH on both richness and composition of the soil community highlights the importance of higher pH soils as a reservoir of unique soil biodiversity in a landscape dominated by acidic soil conditions.

Distribution of plant root associated symbiotic fungi and especially ECM fungi which associate with trees will be particularly important to consider when planning future management at Glen Prosen. Clear distribution patterns were seen across the land use categories, with low frequency and limited richness of ECM fungi present in plots outside of currently or formerly tree planted areas. The lack of ECM fungi in open areas should be taken into account when planning for tree regeneration in these areas as appropriate fungi may need to be introduced along with the trees.

References

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. Aquatic Microbial Ecology, 75, 129–137. http://doi.org/10.3354/ame01753

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology, 37(8), pp.852-857.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), pp.581-583.

Freitag TE, Prosser JI (2009) Correlation of methane production and functional gene transcriptional activity in a peat soil. Applied and Environmental Microbiology, 75, 6679-6687. https://doi.org/10.1128/AEM.01021-09

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Applied and Environmental Microbiology, 66, 5488-5491. DOI: 10.1128/AEM.66.12.5488-5491.2000

Hansen L, Knudsen H (1997) Nordic macromycetes Vol. 3. Nordsvamp, Copenhagen.

Ihrmark KB, Bödeker ITM, Cruz-Martinez K, Friberg, H, Kubartova, A, et al. (2012) New primers to amplify the fungal Its2 region - evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiology Ecology, 82, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Kindt R, Coe R (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF). ISBN 92-9059-179-X. http://www.worldagroforestry.org/output/tree-diversity-analysis.

Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS et al. (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, 47, D259-D264.

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2024) Vegan: Community Ecology Package. R package version 2.6-8, https://CRAN.R-project.org/package=vegan.

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field

samples. Environmental Microbiology, 18, 1403–1414. http://doi.org/10.1111/1462-2920.13023

Qiime2 documentation (2023) docs.qiime2.org. Training feature classifiers with q2-feature-classifier — QIIME 2 2023.9.2 documentation. [online] Available at: https://docs.qiime2.org/2023.9/tutorials/feature-classifier/

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research, 41(D1), pp.D590-D596.

R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rivers AR, Weber KC, Gardner TG, Liu S, Armstrong SD (2018) ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. F1000Research, 7, 1418. doi: 10.12688/f1000research.15704.1

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open-source tool for metagenomics. PeerJ, 18, e2584. doi: 10.7717/peerj.2584.

Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, Cavalier-Smith T, Guiry MD, Kirk PM (2015) A higher level classification of all living organisms. PLoS ONE, 10(4): e0119248. doi:10.1371/journal.pone.0119248.

Stoeck T, Bass D, Nebel M, Christen R, Jones MD et al. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology, 19, 21–31.

Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R, Ley RE (2012) Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME journal, 6(1), pp.94-103.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Appendix

Several datasets which were collected during the baseline soil biodiversity survey at Glen Prosen accompany this report and are described below:

Survey plot information
GlenProsen-2024-Soil-Biodiversity-Baseline-Survey-plot-data.xlsx
Sheet 'Plot Info' contains sample plot locations, topographic information (elevation, slope and aspect), the results of the vegetation survey and additional notes recorded by surveyors in the field. Sheet 'Soil chemistry' includes the results of the soil chemical analyses for each survey plot – see the methods section of this report for details of the methods used.

Plot photographs
Photographs of each survey plot taken by the surveyors are supplied as .JPG files labelled with the plot code. See the surveyors notes in the survey plot information excel sheet for descriptions of what is shown in the plot photographs and the directions from which they are taken.

Soil DNA sequence data
GlenProsen-2024-Soil-Biodiversity-Baseline-Survey-DNA-data.xlsx
This excel work book includes three sheets, one for each DNA marker used (16S, 18S, ITS2). Each sheet is arranged as a table of OTUs/ASVs (rows) by samples (columns). For each OTU we include: the unique identifier code, the DNA sequence, the taxon assignment string, the OTU number code used during data analysis, the Frequency (number of samples where the OTU was present) and the total number of reads. The number of reads present in each sample is then shown in subsequent columns.

Aberdeen

The James Hutton Institute
Craigiebuckler
Aberdeen AB15 8QH
Scotland
UK

Dundee

The James Hutton Institute
Invergowrie
Dundee DD2 5DA
Scotland
UK

Contact

Tel: +44 (0) 344 928 5428
Fax: +44 (0) 344 928 5429

info@hutton.ac.uk

Farms

Balruddery Research Farm
Invergowrie
Dundee DD2 5LJ

Glensaugh Research Farm
Laurencekirk
Aberdeenshire AB30 1HB