Environlink 2433-ORC012-Summerfield-PR32213

The impact of land use and land management on soil biological health

Andy Nilsen and Tina Summerfield
Department of Botany
University of Otago

Summary

This pilot study explored the potential of using soil DNA to assess biological indicators of soil health across a range of land uses in Otago, New Zealand. By targeting bacterial, fungal, and invertebrate communities, the study demonstrated that soil DNA can detect land-use-specific shifts in microbial diversity and community composition, with macroporosity and pH emerging as key drivers of microbial structure. While bacterial and fungal markers successfully identified multiple potential indicator taxa and functional differences between sites, the invertebrate data were more limited, likely due to patchy distribution and technical constraints. Despite the challenges, the results highlight the value of soil DNA in capturing ecologically meaningful patterns that are not evident through conventional chemical or physical assessments. These findings provide a foundation for a larger-scale study with increased sampling across land uses to better characterise biological indicators of soil health. Such work would strengthen the case for integrating DNA-based methods into regional and national soil monitoring frameworks.

Introduction

Soil is the most biodiverse habit on Earth, hosting approximately 59% of all known species (Anthony et al., 2023). The vast communities of soil organisms, across multiple phyla, play a crucial role in nutrient cycling, carbon storage and contaminant processing, making them integral to ecosystem functioning and agricultural productivity (Aislabie & Deslippe, 2013). Soil-dwelling organisms are often highly sensitive to changes in their environment, for example, shifts in nutrients, changes in pH and compaction, making them excellent indicators for soil health (Pankhurstab et al., 1995). In New Zealand, soil bacterial community composition has been shown to predict land and soil conditions with high accuracy (Hermans et al., 2020), however, non-bacterial communities remain under studied.

Despite the ecological importance of organisms, biological indicators are currently missing from the national-scale soil quality monitoring programmes, which focus solely on the physical and chemical measures (National Environmental Monitoring Standard, 2022). Soil DNA offers a powerful, non-invasive approach for assessing microbial and invertebrate communities, similar those seen in monitoring aquatic ecosystems through the use of environmental DNA (eDNA). However, challenges such as large data volumes, the need for reference databases and the lack of standardised benchmarks require the collaboration with experts in soil DNA analysis. *This project acts as a pilot study to assess the potential of soil DNA in evaluating soil biodiversity and condition across multiple phyla, with the aim of informing future soil DNA monitoring projects.*

This pilot study analysed soil biodiversity across 15 sites in Otago, representing a range of land uses – including plantation forestry, extensive and intensive pasture, and horticulture – spanning different soil types and climates. DNA markers for bacteria, fungi and invertebrates were used to determine patterns in microbial and invertebrate communities with the aim to assess the potential for development of biological indicators into regional monitoring frameworks.

Discussion

This pilot study demonstrated the application of DNA markers to identify differences in soil community composition across various land use types. The DNA markers targeting the fungal and bacterial community captured the diversity of the different soil communities and highlighted land use specific composition of the soil microbial communities. The third marker, targeting the invertebrate community, did not yield the same level of information likely due to a combination of technical challenges.

Variation in alpha and beta diversity observed in the fungal and bacterial communities were more similar in soils from managed land uses compared to the unmanaged land use types. This may reflect that land management practices, such as compaction, fertiliser use, and irrigation, are exerting significant pressures on soil microbial communities, potentially altering ecosystem functions such as nutrient cycling and organic matter decomposition. This would be consistent with previous studies that have found changes in microbial communities in response to management practises in other countries with a variety of land uses (De Carvalho et al., 2016; Hendgen et al., 2018; Zhou & Fong, 2021).

The indicators of soil health, macroporosity and pH, were found to be the main contributors to differences in community structures in both the fungal and bacterial communities.

This pilot study, with a limited dataset, identified multiple land-use-specific fungal and bacterial indicator ASVs. This approach has been used in other contexts, using larger data sets, such as in different maize farming systems, where over 400 bacterial indicator species have been recorded, with high numbers being identified in traditional farm practises compared to organic systems (Matteoli et al., 2022). Similarly, another study identified ca. 340 bacterial genera associated with one, or multiple farming practices (Wilhelm et al., 2023). In particular, ASVs in the archaeal family Nitrososphaeraceae were associated with poor soil health. While ASVs in the Nitrososphaeraceae were detected in our study, they were not identified as indicator species, possibly due to limited sample size. Given the relatively small sample sizes for the different land uses, the putative indicator ASVs should be confirmed in a broader study.

Community functional predictions were performed using DNA data to assign functional guilds to fungal taxa and metabolic functions to bacterial communities. Functional analyses of bacteria indicated enrichment of pathways in some soils, most marked was in plantation forests including atrazine degradation pathways. While these functional annotations are possible for fungal and bacterial datasets through tools like PICRUSt2 and Funguild, it should be noted that these functional annotation pipelines are only as good as their reference datasets which are, by in large, incomplete. Further samples and additional approaches such as metagenome sequencing would provide useful data where functional differences are predicted.

The invertebrate data yielded only four indicator ASVs that were restricted to dairy, horticulture and wetlands. This low number of ASVs is likely due to the disparate distribution of invertebrates in the samples. While functional annotations are possible for fungal and bacterial datasets through tools like PICRUSt2 and Funguild, no such tools are available for invertebrates.

Direct comparison of the invertebrate ASVs with the physical invertebrate survey data was not possible. This is in part due the challenges with reference databases (see methods), expertise required for invertebrate identification in field surveys and cryptic morphology at different lifecycle stages further adds to the difficulties. However, invertebrates identified in the survey were not detected in the DNA despite sequences being present in the database (see Table 4); this may reflect the uneven distribution of invertebrates with the soil.

The study highlights the potential of integrating DNA-based biological indicators into soil monitoring frameworks. In particular, the fungal and bacterial communities showed more potential with land use specific community composition. These findings provide a foundation for a larger-scale study with increased sampling across land uses to better characterise biological indicators of soil health. Such work would strengthen the case for integrating DNA-based methods into regional and national soil monitoring frameworks. Soil DNA has the potential to transform how we assess and monitor soil health—providing land managers and policymakers with timely, biologically informed insights to guide sustainable land use decisions in the face of environmental change.

Methods

DNA extraction and amplicon library preparation

Soil samples were provided frozen by the Otago Regional Council. DNA was extracted using the Qiagen DNeasy Powersoil following the manufactures instructions. Amplicon libraries were performed using a two-step method for Illumina sequencing. Different target regions were used for the three target communities with commonly used primer sets selected from literature. For fungi, PCR amplification of internal transcribed spacer 2 (ITS2) was performed using the forward primer fITS7 (Ihrmark et al., 2012) and the reverse prime ITS4 (White et al., 1990).

The bacterial 16S rRNA gene region was amplified using 515F (Parada et al., 2016) and 806R (Apprill et al., 2015). The cytochrome *c* oxidase subunit 1 (COI) region from invertebrates was amplified using mlCOIintF (Leray et al., 2013) and HCO2198 (Folmer et al., 1994). All first round primers had Illumina linkers. Each reaction included 0.25 µL of PCRbio HS Verifi (PCR Biosystems), 5 µL of 5x PCRbio buffer, 10 pmol of each primer, 1 µL of 5 mg/mL recombinant albumin (New England Biolabs), 1 µL of 1:10 diluted DNA and 8.5 µL of autoclaved ultrapure water.

The amplification protocols were: for fungi 95˚C for two minutes, followed by, 30 cycles of 95˚C for 15 sec, 57˚C for 15 sec and 72˚C for 30 sec, and a final elongation of 72˚C for 1 min; for bacteria 95˚C for two minutes, followed by, 25 cycles of 95˚C for 15 sec, 60˚C for 15 sec and 72˚C for 30 sec, and a final elongation of 72˚C for 1 min; for insects 95˚C for two minutes, followed by, 30 cycles of 95˚C for 15 sec, 56˚C for 15 sec and 72˚C for 30 sec, and a final elongation of 72˚C for 1 min. Negative controls were included for all PCR reactions. PCR products were purified using AMPure XP beads (Beckman Coulter) following the manufactures instructions and quantified using Qubit dsDNA HS kit (ThermoFisher). The PCR products were diluted to 2 ng/µL and a second round of PCR was performed to attach sample specific barcodes. Each reaction included 0.25 µL of PCRbio HS Verifi, 5 µL of PCRbio 5x buffer, 10 pmol of each barcoded primer, 1 µL of purified PCR product and 9.5 µL of autoclave ultrapure water. The amplification protocol was 95˚C for two minutes, followed by 10 cycles of 95˚C for 20 sec, 60˚C for 30 sec and 72˚C for 30 sec, and a final elongation of 72˚C for 1 min. Amplicons were purified and quantified as above. Samples were pooled at 10 nmol for sequencing at the Otago Genomics Facility on an Illumina MiSeq 2 x 300 bp using V3 chemistry.

Bioinformatics

The different amplicon regions were processed separately. Raw sequences were processed using the DADA2 1.22 pipeline (Callahan et al., 2016) in R 4.1.3. (Team, 2019). Fungal primers were removed using cutadapt 4.5 (Martin, 2011) with up to 20% mismatches allowed. Bacterial and invertebrate primers were truncated using the expected primer lengths. Reads were trimmed and filtered using the "filterAndTrim" function. Reads were truncated at the first instance of a quality score representing two or more expected errors. Reads shorter than 50 bp after trimming were discarded. Sequences were denoised into amplicon sequences variants (ASVs). Taxonomy was assigned using the "assignTaxonomy" function using the UNITE database 10.0 general release (Kõljalg et al., 2020; Nilsson et al., 2019) for fungi, and Silva 138.2 for bacteria (Quast et al., 2013). Databases available for COI were not compatible with dada2 and BOLDigger3 was used to classify sequences against the BOLD database (Buchner & Leese, 2020; Ratnasingham et al., 2024). Only ASVs that were identified at the phylum level and above were retained. The data were rarefied to the sample with the lowest read number for the fungal and bacterial datasets. Because the invertebrate dataset had many samples with low read depth, the data was transformed to presence and absence data. Analyses were performed using the packages vegan 2.6-10 and phyloseq 1.38.0 in R 4.1.3 (McMurdie & Holmes, 2013; Oksanen et al., 2019; Team, 2019) and plotted using ggplot2 (Wickham, 2016). Upset plots were produced using the UpsetR 1.4.0 package on 999 randomly subsampled samples from the different land uses (Conway et al. 2017). Indicator species analyses were conducted using indicspecies 1.8.0 in R (De Cáceres & Legendre, 2009). Fungal ASVs were classified into guilds using FUNGuildR 1.1 retaining only classifications at “Highly Probable” and “Probable” for guild analyses (Brendanf, 2021; Nguyen et al., 2016). Functional profiles from the 16S rRNA communties were predicted using PICRUSt2 2.6.1 and analysed in ggpicrust2 1.7.1 (Douglas et al., 2020; Yang et al., 2023). Physiochemical variables were tested using distance-based Redundancy Analysis in the vegan package.

Results

The amount and quality of the data received was very good. One sample, ORCSQ024 from the horticultural land use, had low diversity, despite the adequate sequencing depth for all the three DNA markers that were investigated.

Fungal Community Analysis (ITS DNA barcode)

Read depth and rarefaction

The fungal dataset had sufficient sequencing depth for analyses and the rarefaction curves indicated that the majority of the species in communities were captured (

Appendix *1*). A total of 8340 ASVs were classified past kingdom and the samples were rarefied to the common depth of 34852 reads per sample (purple line in

Appendix *1*, Figure S1).

Alpha diversity

Alpha diversity indicated that wetlands were the most species rich of the soil types (See Figure 1). Forestry (non native) and horticulture had the lowest richness. High Shannon diversity in the wetlands indicates high richness and evenness. As opposed to the Forestry category, the low Shannon diversity may indicate that there is low evenness and the communities are dominated by a few ASVs. There was a significant difference detected in land use types (kruskal.test: Richness p = 0.003171, Shannon p = 0.006934), however, no significant differences were detected between individual types (pairwise.wilcox.test using FDR correction). This is likely due to the small sample size.

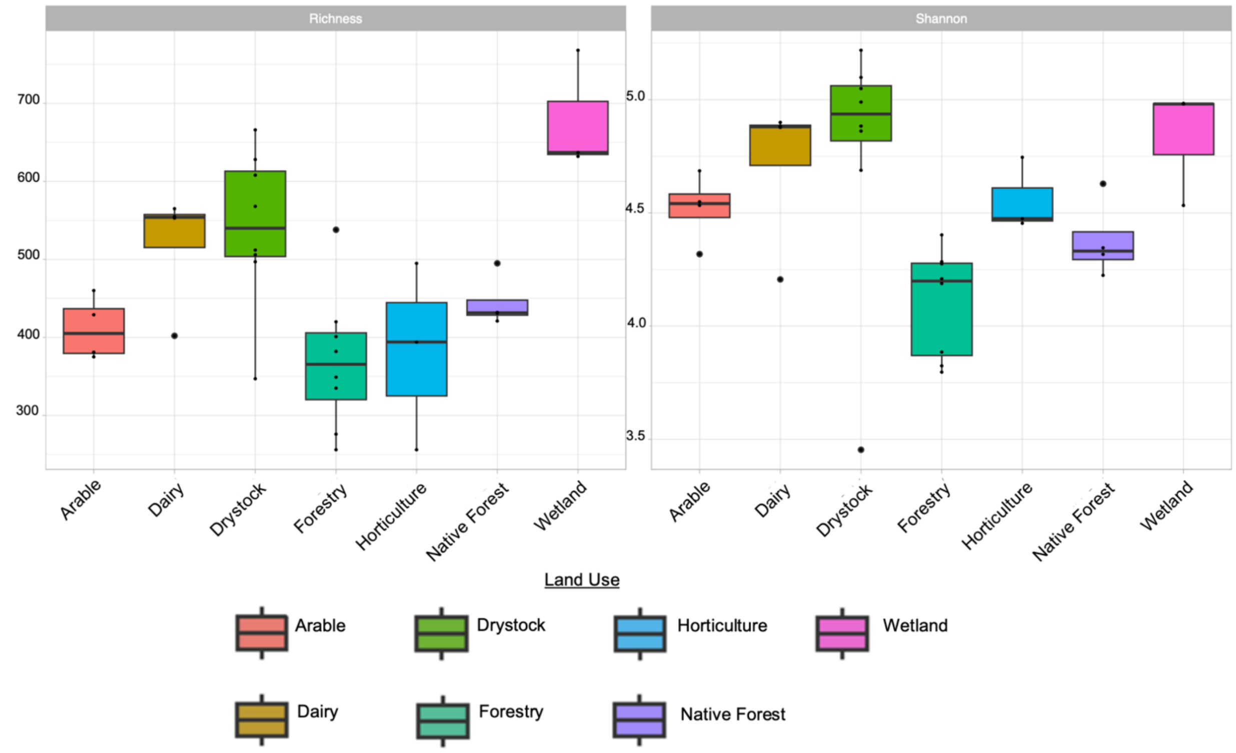


Figure 1 Alpha diversity metrics for fungal communities (based on ITS2 amplicon data) in soil across different land uses. The panel on the left shows fungal richness (total number of observed ASVs), while the panel on the right shows Shannon diversity, which accounts for both richness and evenness of taxa.

At the level of phylum, the Mucoromycota and Entorrhizomycota had high diversity in exotic forests and wetlands, respectively (Figure 2). These phyla had lower relative abundance in all other land uses. The phyla Ascomycota and Basidiomycota were found in all land uses with some variation in richness.

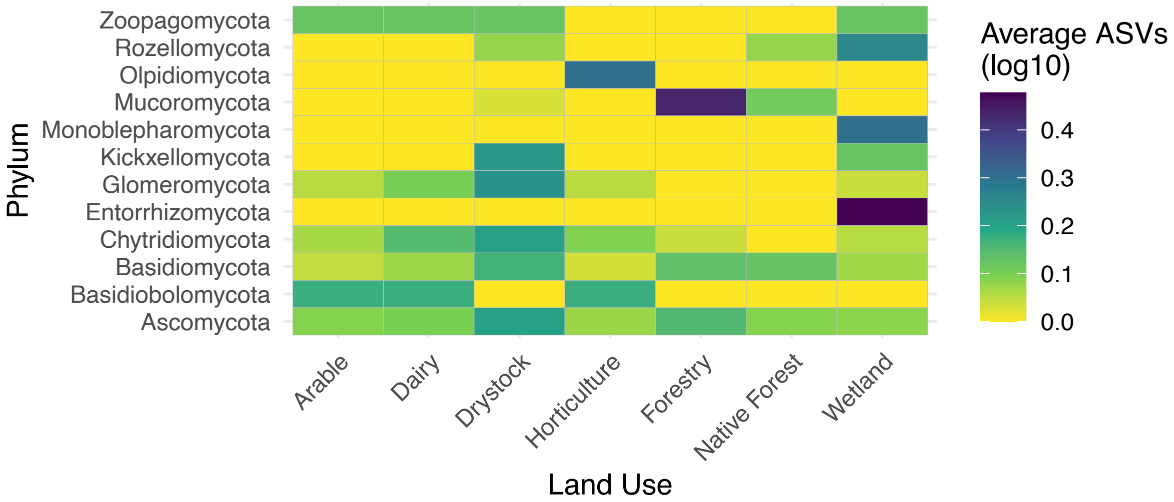


Figure 2 Average number of fungal ASVs (log$_{10}$) detected across different soil types and land uses, summarised at the phylum level.

Beta diversity

Beta diversity analysis showed separate groupings of fungal communities in wetland, native and exotic forests (See Figure 3). Clustering of arable, dairy, dry stock and horticulture was evident, particularly for the shared pallic soil order. Permanova analysis indicated that there was a significant difference in fungal communities between the land uses (p = 0.001). A significant test of homogeneity dispersion result (p = 0.002) may suggest that unequal group variances might be contributing to the significant result. From the physiochemical data (Appendix 5), a distance-based Redundancy Analysis (db-RDA) identified macroporisity (p = 0.001), pH (p = 0.012), total carbon (p = 0.015) and arsenic (p = 0.038) as significant factors influencing community structure. However, a large proportion of the variance in the fungal data remained unexplained (72.4%).

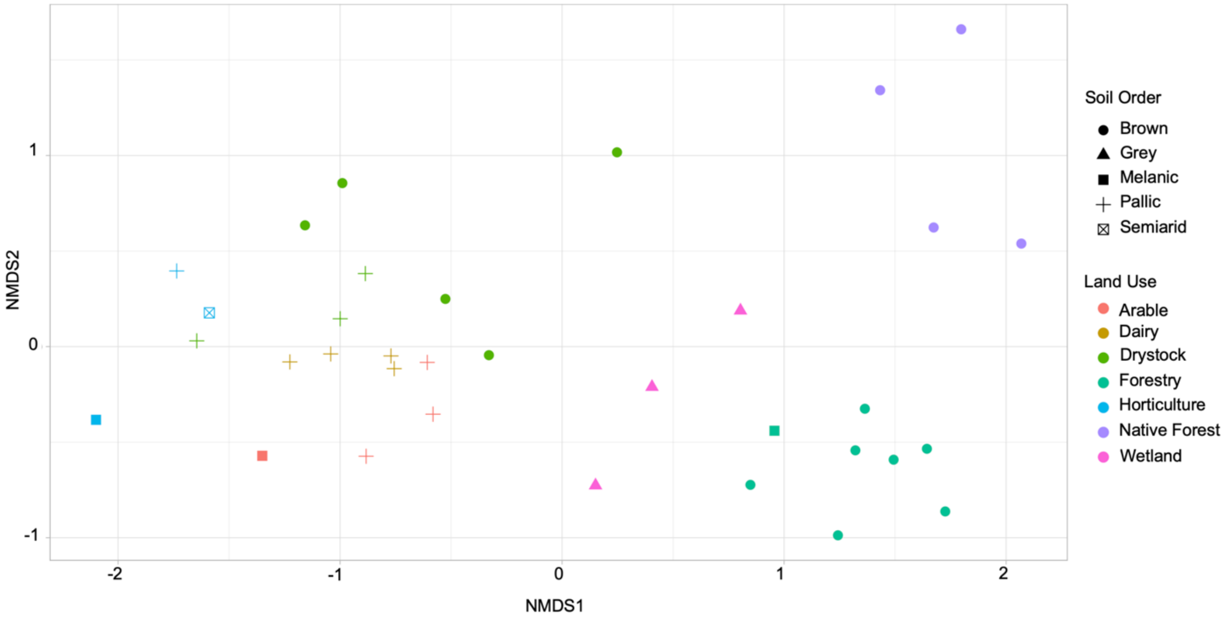


Figure 3 *Non-metric multidimensional scaling (NMDS) plot showing beta diversity of fungal communities across land uses, based on a Bray–Curtis dissimilarity matrix. Each point represents the fungal community composition of a single soil sample, with spatial proximity indicating greater similarity between samples. Colours correspond to different land use types and shapes correspond to different soil orders. The ordination captures variation in community structure among sites, with a stress value of 0.1148 indicating a good fit between the ordination and the original distance matrix.*

ASVs shared across land uses and indicator species

The vast majority of fungal ASVs were only found in one type of land use (Figure 4**Error! Reference source not found.**). Of which, 2–3 times more ASVs were recorded in dry stock (1731), forestry (1415), wetland (1333) and native forest (1298), compared to dairy, arable and horticulture. In addition, a proportionally high number of ASVs were shared between dairy and arable land uses (102). Other land uses with higher numbers of shared ASVs (>30) were several other combinations of managed land uses as well as native and exotic forests.

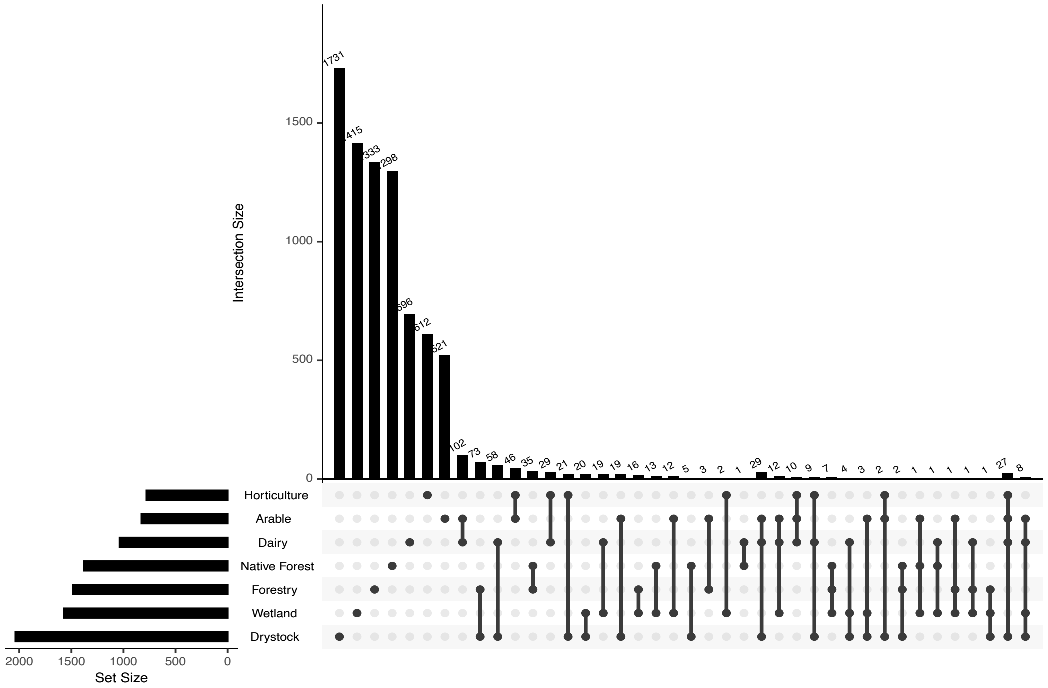


Figure 4 UpSet plot showing the number of fungal ASVs unique to each land use and those shared between different land use types. Horizontal bars represent the total number of ASVs detected per land use, while vertical bars indicate the number of ASVs found exclusively in or shared among the land uses shown in the connected dot matrix below.

The species indicator test identified 21 ASV that were statistically significantly found exclusively in all samples of land use types: native forest, wetlands, dairy and arable. The conservative approach of selecting only ASVs that were exclusively found in all samples of a land use type was adopted due to some of the low replication within some groups.

Predicting fungal functional types

Fungal trophic modes describe the different ways fungi obtain nutrients and interact with their environment, such as decomposing organic matter, forming mutualistic relationships, or acting as pathogens. In this analysis, trophic modes were predicted using FUNGuild based on taxonomic identity and published ecological associations. Multiple different trophic modes were detected in the different land uses, with many ASVs being classified as having multiple different trophic states (Figure 5). Drystock land use contained high diversity of almost all trophic modes. ASVs classified as pathogens were prevalent across all land use types but highest in drystock. Saprotrophic richness was also dominant across all land uses.

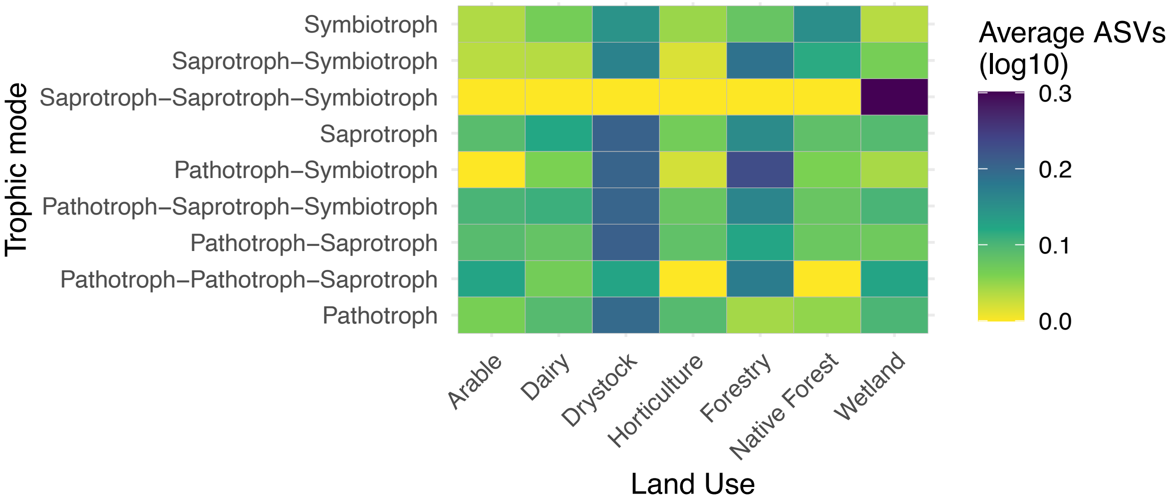


Figure 5 Average number of fungal ASVs (log10) in the different land uses grouped by trophic mode.

Bacterial Community Analysis (16S rDNA barcode)

Read depth and rarefaction

The bacterial data was sufficiently sampled as indicated in the rarefaction curve (Appendix 3). The samples were rarefied to 18894. At this depth, the number of ASVs per sample is beginning to plateau. After rarefaction, a total of 18059 ASVs were retained for downstream analyses.

Alpha diversity

There was no significant difference was detected in alpha diversity for land use.

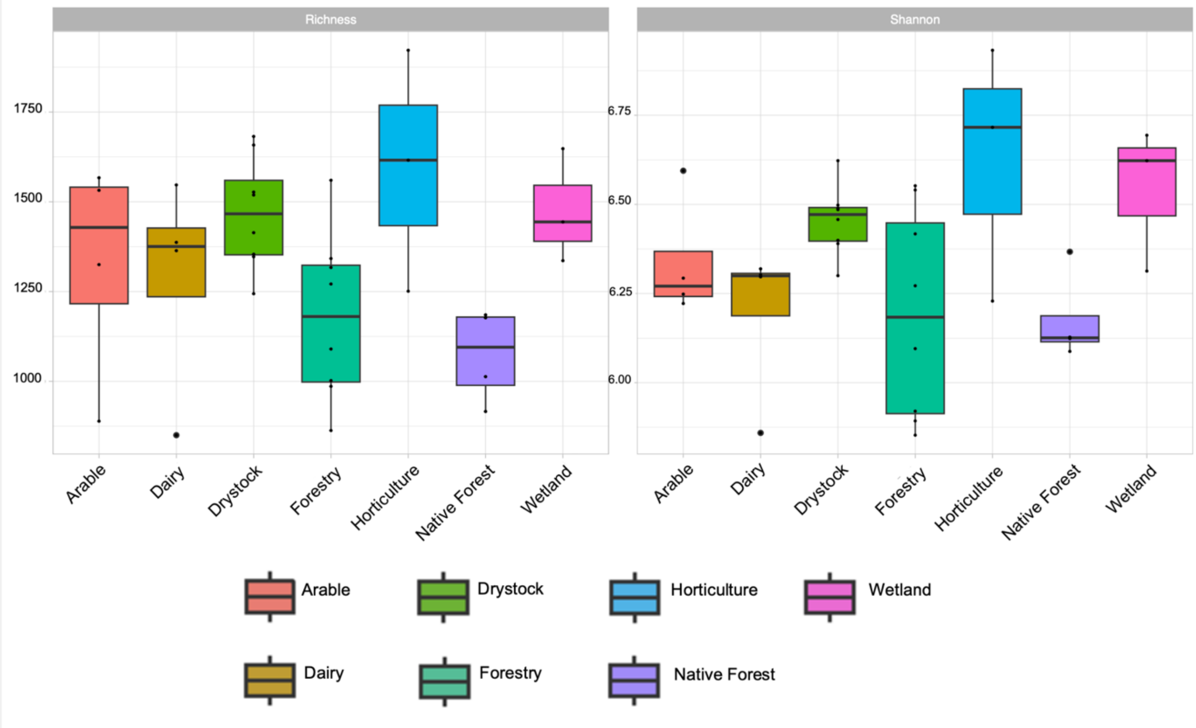


Figure 6 Alpha diversity metrics for bacterial communities (based on 16S rRNA amplicon data) in soil across different land uses. The panel on the left shows bacterial richness (total number of observed ASVs), while the panel on the right shows Shannon diversity, which accounts for both richness and evenness of taxa.

Bacterial phyla had a large variance in richness across the land uses (Figure 7). In particular, multiple phyla were largely restricted to wetlands. Conversely, phyla like Verrucomicrobiota, Pseudomonadota, Bacillota and Actinomycetota were found in all land use types. Some phyla were more restricted to managed land use types i.e. Entotheonellaeota and Deinococcota.

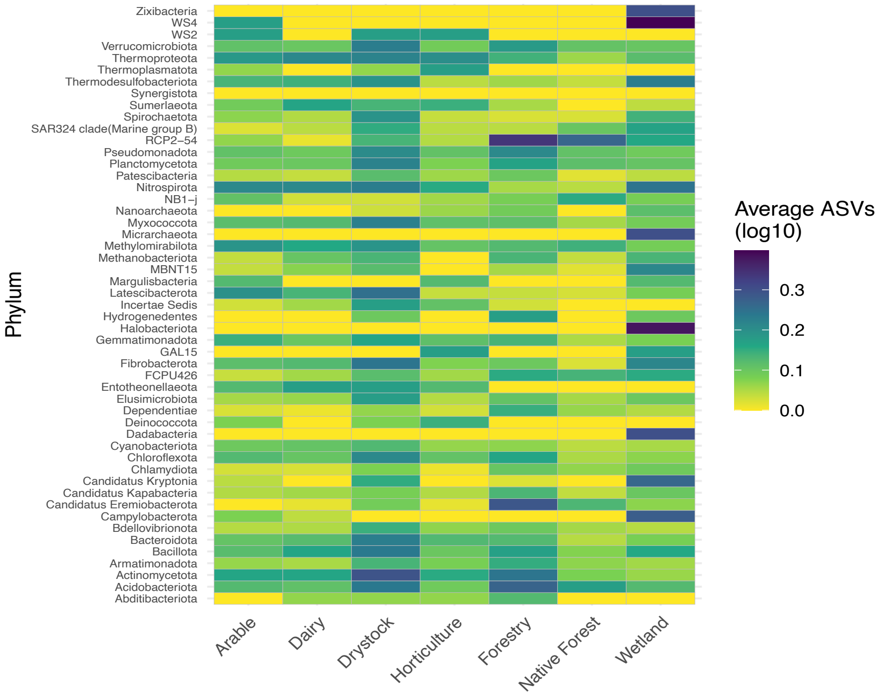


Figure 7 Average number of bacterial ASVs ($\log_{10}$) in the different land uses at the phylum level.

Beta diversity

Beta diversity yielded a similar pattern as the fungal data with arable, dairy, drystock and horticulture mostly grouping together. In contrast to the fungal data, the native forest and forest categories overlapped. A statistically significant difference was observed for land use (permanova p = 0.001, betadisp p = 0.22). The distance-based Redundancy Analysis (db-RDA) of physiochemical data (Appendix 5) identified pH (p = 0.001), Macroporosity (p = 0.004) and Bulk density (p = 0.011) being significant factors influencing community structure with 36% of the variance being explained in the bacterial data.

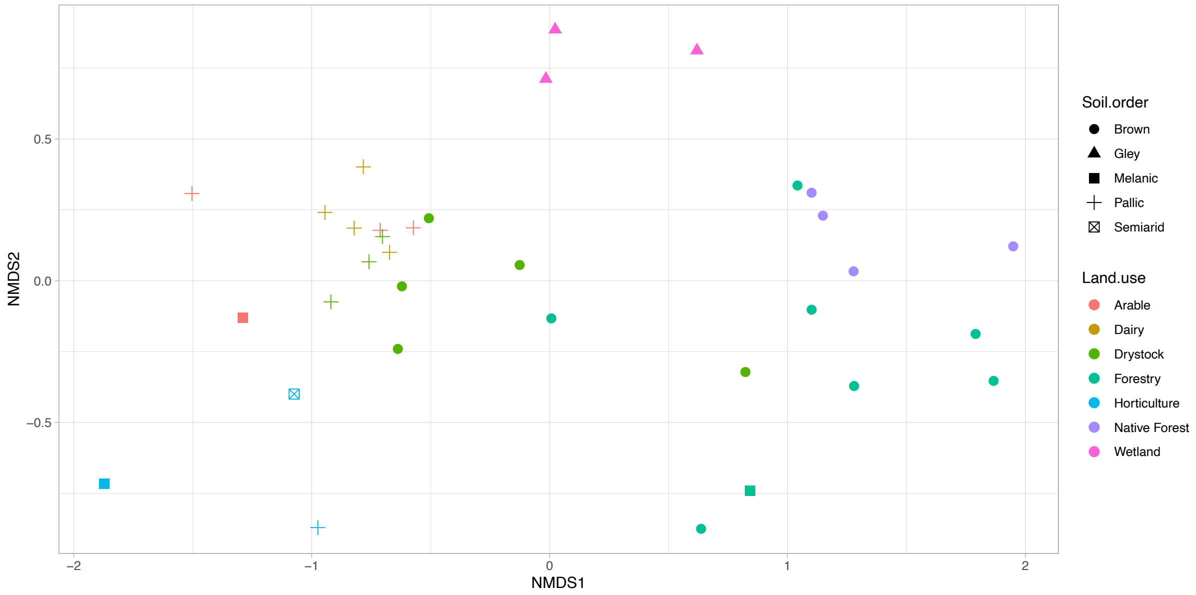


Figure 8 Beta diversity from bacterial data. NMDS of Bray Curtis distance matrix. Stress 0.08425575.

ASVs shared across land uses and indicator species

Similar to the fungal data, the majority of bacterial ASVs were only found in one type of land use (Figure 9). Exotic forest and Dry stock supported the highest total number of ASVs. Compared to the fungal data, there was a more precipitous decrease in total ASVs across the land uses as opposed to 2–3 × drop observed in Dairy, Arable, Horticulture. There was proportionally high numbers of ASVs shared between Dairy, Arable and Drystock land uses. Overall, Dairy had the fewest total ASVs, closely followed by Native Forest, Arable, Wetland and Horticultural land uses.

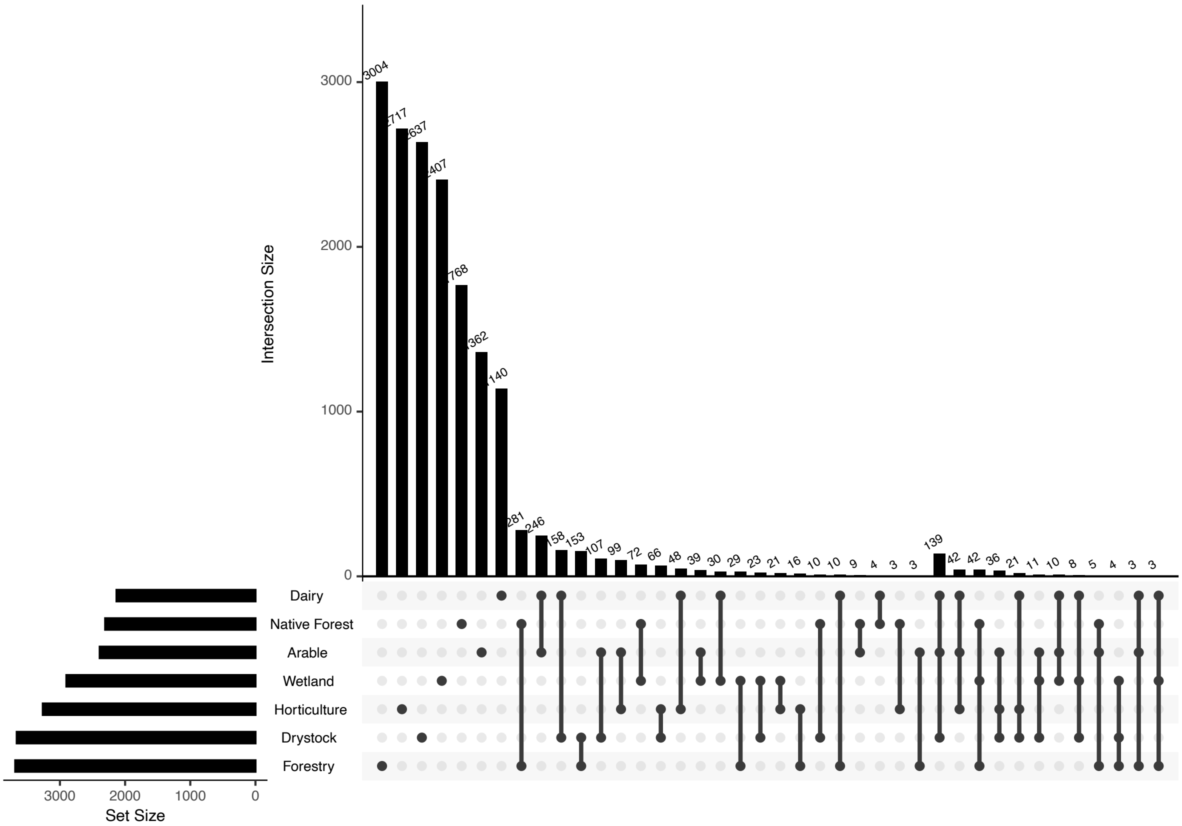


Figure 9 UpSet plot of bacterial ASVs exclusively found in, and shared between land uses

Testing for indicator species, a total of 44 ASVs were found to be exclusively present in all samples within a land use type, or combination of land use types (Table A2). Five different land use types had exclusive ASVs and three ASVs were found to be present in all samples across all land use types. Several pathways in different land uses were found to be statistically significantly overrepresented (Figure 10). Pathways associated with sulfur metabolism and plant-pathogen interactions were increased in plantation forestry, native forest and wetlands. Notably atrazine degradation is increased in forestry.

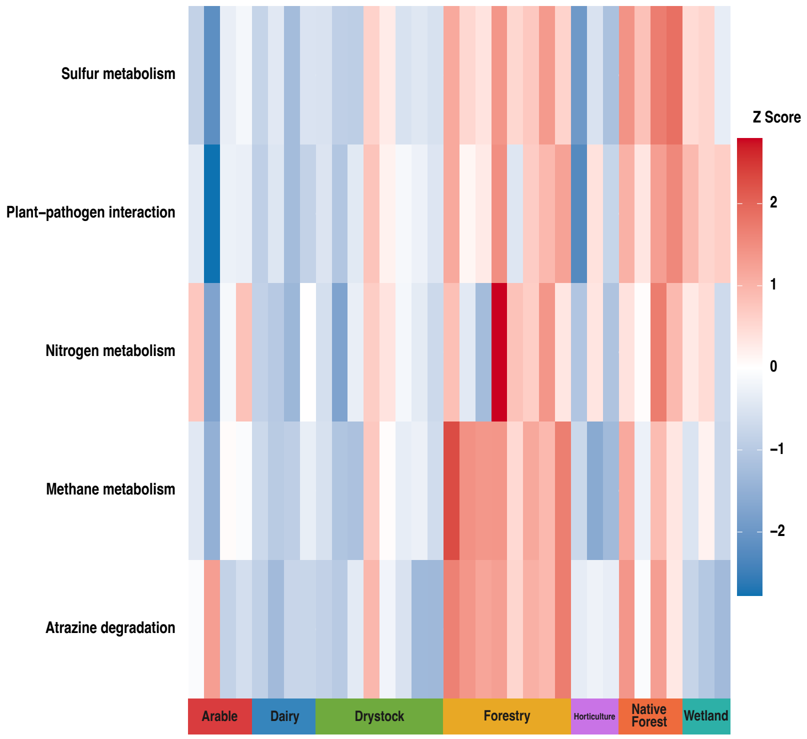


Figure 10 Heatmap of pathways enriched in different land uses based on 16S rRNA sequence.

Invertebrate dataset (COI without taxonomically unclassified ASVs)

The COI dataset presented some challenges for downstream analyses because of the taxonomic assignments. A total of 9665 ASVs were detected. These were classified against the BOLD database using boldigger3. A conservative approach was used and only ASVs classified in the phyla Annelida, Arthropoda, Cnidaria, Nematoda, Rotifera, Mollusca, Nemertea were retained. ASVs classified in the phyla Heterokontophyta, Rhodophyta, Ascomycota, Bacillariophyta, Cryptophyta, Ochrophyta and Chordata were excluded from downstream analyses. After all the non-target ASVs were removed, only 675 remained. Sample ORCSQ024 contained no invertebrate ASVs and was removed from further analyses.

Sampling taxonomically classified ASVs

Retaining only ASVs that were classified to at least phylum level resulted in large differences between samples; the minimum number of reads per sample being 225, and a median of 3160 across all samples (Appendix 4). The COI data containing all ASVs had comparatively even read coverage with a median of ~38 000 read per sample (*Appendix 3*). The large discrepancy between total COI ASVs and the invertebrate ASVs may reflect the distribution of invertebrates in the soil column. Of the 675 ASVs, ca. 75% were recorded only once.
Due multiple samples having very low read depth, rarefaction was not suitable. However, because most of the rarefaction curves were beginning to plateau, indicating most species were captured, the data was transformed to presence/absence data. This was sufficient for alpha diversity metrics, but the very high proportion of singletons meant beta diversity statistics could not be performed.

Alpha diversity

Richness was highest in dairy pasture with native forest and wetlands having the fewest ASVs. This is also reflected in the Shannon index and indicates that a small number of ASVs dominate in the community. However, there are large variances in most of the land use types with the exception of horticulture that only contains two data points since sample ORCSQ024 was removed. None of the statistical tests were significant for the alpha diversity metrics.

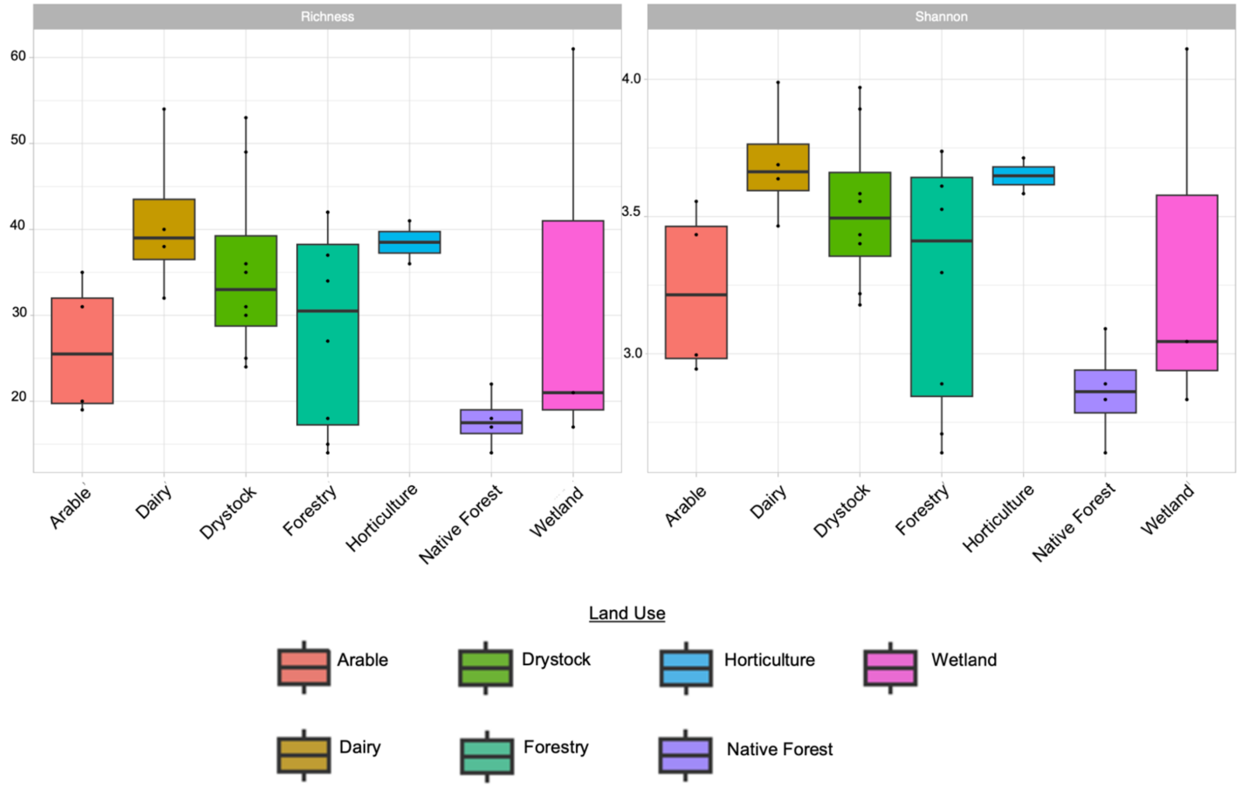


Figure 11 Alpha diversity metrics for invertebrate COI ASVs. Richness on the left and Shannon on the right.

The majority of invertebrate ASVs were in the phyla Arthropoda (invertebrates with exoskeletons) and Annelida (segmented worms) which were detected in all land uses. Nematoda (roundworms) were proportionally higher in managed land uses and wetlands and Nemertea (ribbon worms) were only detected in horticulture. ASV identified in Cnidaria (freshwater and marine invertebrates) were also mostly restricted to managed land uses (Figure 12).

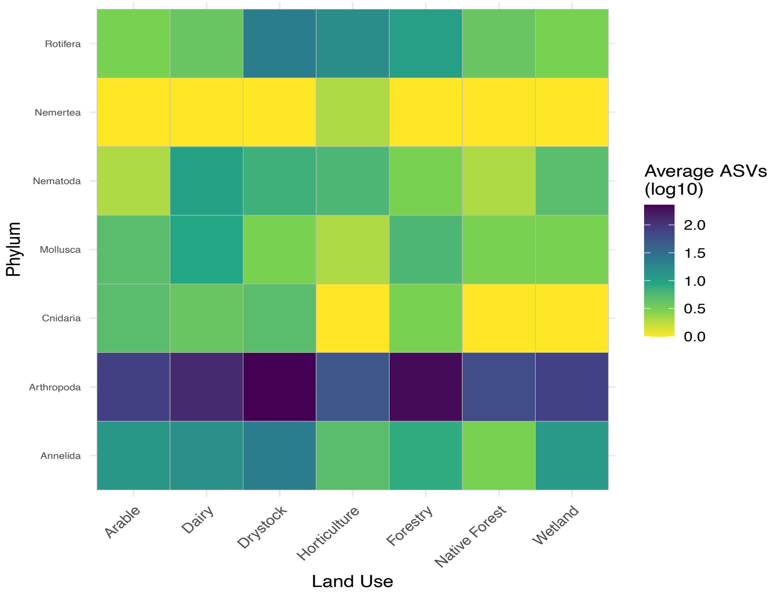


Figure 12 Average number of invertebrate ASVs ($\log_{10}$) in the different land uses at the phylum level

ASVs shared across land uses and indicator species

As with the fungal and bacterial data, Drystock followed by Exotic Forests carried the highest number of ASVs, both exclusively and shared (Figure 13). Overall, Native Forest, Arable and Horticulture had the fewest ASVs. Unlike the fungal and bacterial data, there were far fewer ASVs that were shared between land uses.

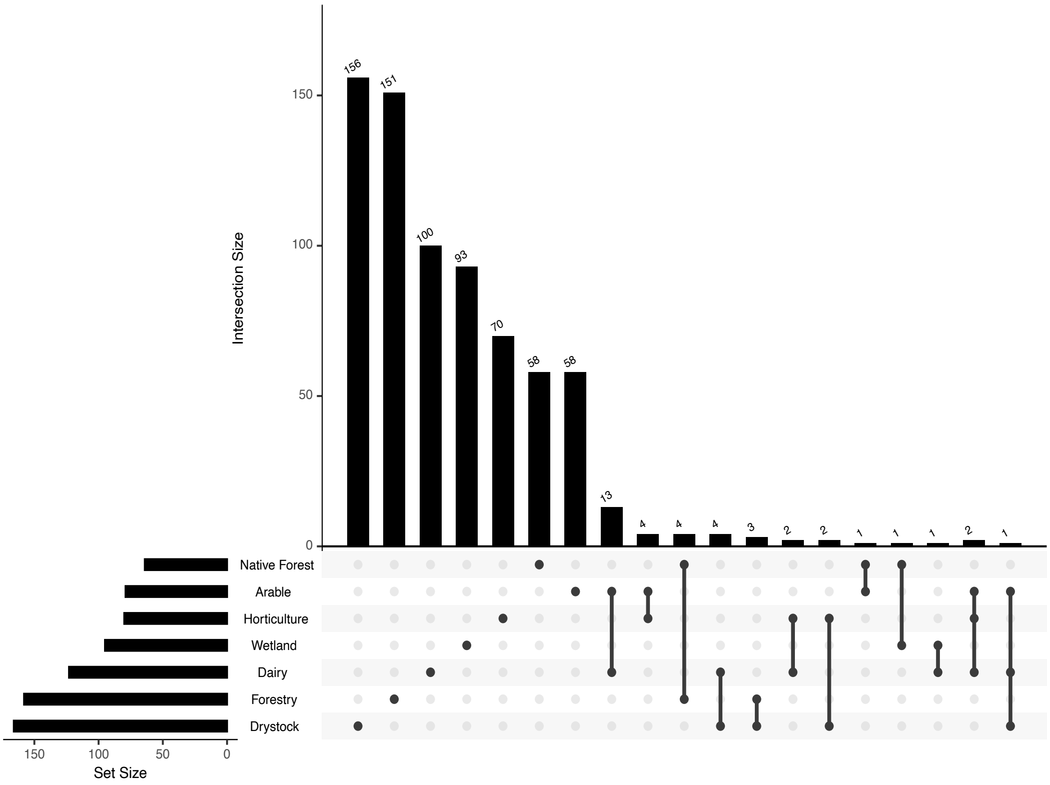


Figure 13 UpSet plot of invertebrate ASVs found exclusively and shared between land use.

Very few ASVs were detected exclusively in all samples of a land use. Additionally, with the removal of sample ORCSQ024 from the horticulture land use, it reduces the total number of samples to two. This might make the detection of COIv3102 only in horticulture somewhat tentative.

Table 3 Invertebrate ASVs that were exclusively found in all samples (B value = 1) of one group (A value = 1)

ASV	group_combination	A	B	Lineage
COIv_0587	Dairy	1	1	Metazoan;Arthropoda;Arachnida;Araneae;Araneidae;NA;NA
COIv_2221	Dairy	1	1	Metazoan;Arthropoda;Insecta;Hymenoptera;Tenthredinidae;NA;NA
COIv_3102	Horticulture	1	1	Metazoan;Rotifera;Bdelloidea;Philodinida;Habrotrochidae;NA;NA
COIv_4514	Wetland	1	1	Metazoan;Arthropoda;Insecta;Hymenoptera;Tenthredinidae;NA;NA

Several taxa of note had been previously been identified from invertebrate surveys in the ORC 2024 Land & Soil Monitoring report were not detected in this study (Table 4). These included invertebrates that live at the soil interface like giant springtails and millipedes, and soil dwelling taxa like giant earthworms. More ubiquitous invertebrates, like poina moth and earth worms were detected.

Table 4 Presence/absence of ASVs from taxa that were noted in the taxa in the ORC 2024 report

Common name	Species	Present in current DNA study	Present in BOLD database	Notes
Giant springtail	*Holacanthella* spp.	No	Yes	There are three ASVs that are in the Neanuridae, but none of these are in the genus *Holacanthella*
Giant native earth worm	Megascolecidae spp.	No	Yes	
Porina moth	Wiseana spp.	Yes	Yes	Three ASVs all as *Wiseana copularis*
Millipedes	Diplopoda spp.	No	Yes	No matches even at the subphylum level Myriapoda
Earth worms	Many	Yes	Yes	53 ASVs classified as Annelida. These were in the classes Crassiclitellata, Enchytraeida and Lumbriculida
Mānuka scarab beetle	*Pyronota* spp.	No	Yes	

References

Aislabie, J., & Deslippe, J. R. (2013). Soil microbes and their contribution to soil services. In *Ecosystem services in New Zealand – conditions and trends* (pp. 143–161). Manaaki Whenua Press.

Anthony, M. A., Bender, S. F., & van der Heijden, M. G. A. (2023). Enumerating soil biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(33). https://doi.org/10.1073/pnas.2304663120

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*(2), 129–137. https://doi.org/10.3354/ame01753

Bernard, K., Burdz, T., Wiebe, D., Alfa, M., & Bernier, A. M. (2018). Clostridium neonatale sp. Nov. linked to necrotizing enterocolitis in neonates and a clarification of species assignable to the genus clostridium (prazmowski 1880) emend. Lawson and Rainey 2016. *International Journal of Systematic and Evolutionary Microbiology*, *68*(8), 2416–2423. https://doi.org/10.1099/ijsem.0.002827

brendanf. (2021). *FUNGuildR* (1.1). https://github.com/brendanf/FUNGuildR

Buchner, D., & Leese, F. (2020). BOLDigger - a Python package to identify and organise sequences with the Barcode of Life Data systems. *Metabarcoding and Metagenomics*, *4*, 19–21. https://doi.org/10.3897/mbmg.4.53535

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*(12), 3566–3574. https://doi.org/10.1890/08-1823.1

De Carvalho, T. S., Da Conceição Jesus, E., Barlow, J., Gardner, T. A., Soares, I. C., Tiedje, J. M., & De Souza Moreira, F. M. (2016). Land use intensifcation in the humid tropics increased both alpha and beta diversity of soil bacteria. *Ecology*, *97*(10), 2760–2771. https://doi.org/10.1002/ecy.1513

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 669–688. https://doi.org/10.1038/s41587-020-0550-z

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Hendgen, M., Hoppe, B., Döring, J., Friedel, M., Kauer, R., Frisch, M., Dahl, A., & Kellner, H. (2018). Effects of different management regimes on microbial biodiversity in vineyard soils. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-27743-0

Hermans, S. M., Buckley, H. L., Case, B. S., Curran-Cournane, F., Taylor, M., & Lear, G. (2020). Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00858-1

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K. E., & Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, *82*(3), 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Kõljalg, U., Nilsson, H. R., Schigel, D., Tedersoo, L., Larsson, K. H., May, T. W., Taylor, A. F. S., Jeppesen, T. S., Frøslev, T. G., Lindahl, B. D., Põldmaa, K., Saar, I., Suija, A., Savchenko, A., Yatsiuk, I., Adojaan, K., Ivanov, F., Piirmann, T., Pöhönen, R., … Abarenkov, K. (2020). The taxon hypothesis paradigm—On the unambiguous detection and communication of taxa. *Microorganisms*, *8*(12), 1–24. https://doi.org/10.3390/microorganisms8121910

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*(1). https://doi.org/10.1186/1742-9994-10-34

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*, 10–12. http://www-huber.embl.de/users/an-

Matteoli, F. P., Silva, A. M. M., Feiler, H. P., de Araújo, V. L. V. P., & Cardoso, E. J. B. N. (2022). Predicting soil farming system and attributes based on soil bacterial community. *Applied Soil Ecology*, *171*. https://doi.org/10.1016/j.apsoil.2021.104335

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

National Environmental Monitoring Standard. (2022). *Soil Quality and Trace Elements Sampling, Measuring, and Managing Soil Quality and Trace Element Data*.

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

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*(D1), D259–D264. https://doi.org/10.1093/nar/gky1022

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., O’hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Maintainer, H. W. (2019). Community Ecology Package: ordination, diversity and dissimilarities. In *vegan: Community ecology package* (2.5-6; Vol. 2, Issue 9, pp. 1–297). https://cran.r-project.org/web/packages/vegan/vegan.pdf

Pankhurstab, C. E., Krkbyab, C. A., & Doubeab, B. M. (1995). Evaluation of soil biological properties as potential bioindicators of soil health. In *Australian Journal of Experimental Agriculture* (Vol. 35).

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. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Philippot, L., Chenu, C., Kappler, A., Rillig, M. C., & Fierer, N. (2024). The interplay between microbial communities and soil properties. In *Nature Reviews Microbiology* (Vol. 22, Issue 4, pp. 226–239). Nature Research. https://doi.org/10.1038/s41579-023-00980-5

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1219

Ratnasingham, S., Wei, C., Chan, D., Agda, J., Agda, J., Ballesteros-Mejia, L., Boutou, H. A., Bastami, Z. M. El, Ma, E., Manjunath, R., Rea, D., Ho, C., Telfer, A. C., McKeowan, J., Rahulan, M., Steinke, C., Dorsheimer, J., Milton, M. A., & Hebert, P. D. N. (2024). BOLD v4: A Centralized Bioinformatics Platform for DNA-based Biodiversity Data. *Methods in Molecular Biology*, *2744*, 403–441. https://api.semanticscholar.org/CorpusID:269005881

Team, R. C. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. Innis, D. Gelfand, J. Sninsky, & T. White (Eds.), *PCR protocols: a guide to methods and applications* (pp. 315–324). Academic Press.

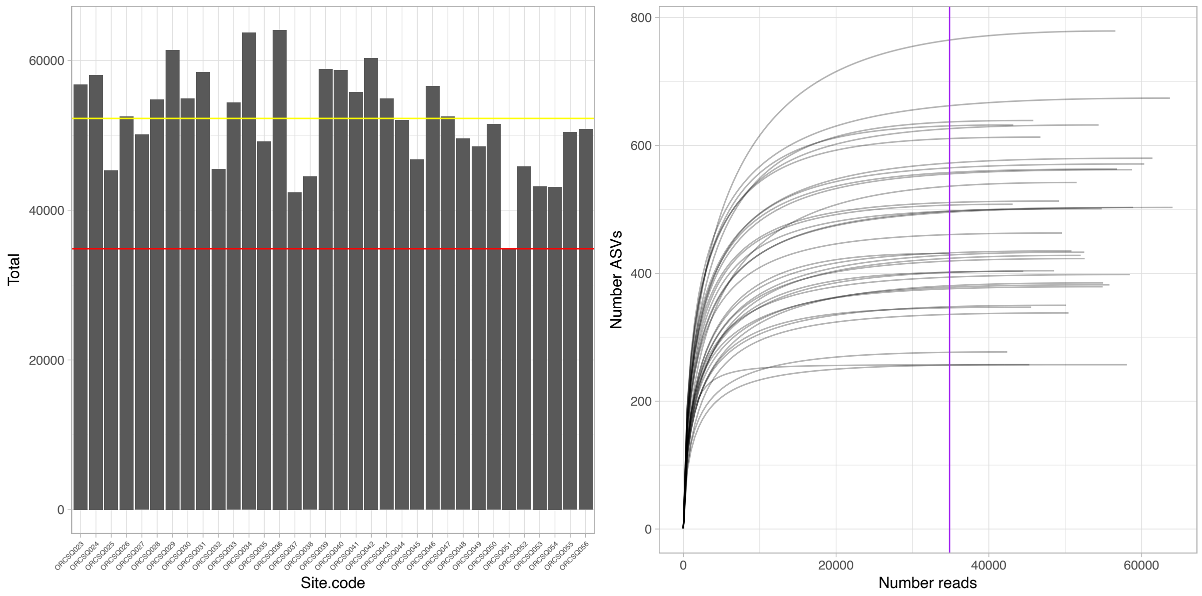
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilhelm, R. C., Amsili, J. P., Kurtz, K. S. M., van Es, H. M., & Buckley, D. H. (2023). Ecological insights into soil health according to the genomic traits and environment-wide associations of bacteria in agricultural soils. *ISME Communications*, *3*(1). https://doi.org/10.1038/s43705-022-00209-1

Yang, C., Mai, J., Cao, X., Burberry, A., Cominelli, F., & Zhang, L. (2023). ggpicrust2: an R package for PICRUSt2 predicted functional profile analysis and visualization. *Bioinformatics*, *39*(8). https://doi.org/10.1093/bioinformatics/btad470

Zhou, J., & Fong, J. J. (2021). Strong agricultural management effects on soil microbial community in a non-experimental agroecosystem. *Applied Soil Ecology*, *165*. https://doi.org/10.1016/j.apsoil.2021.103970

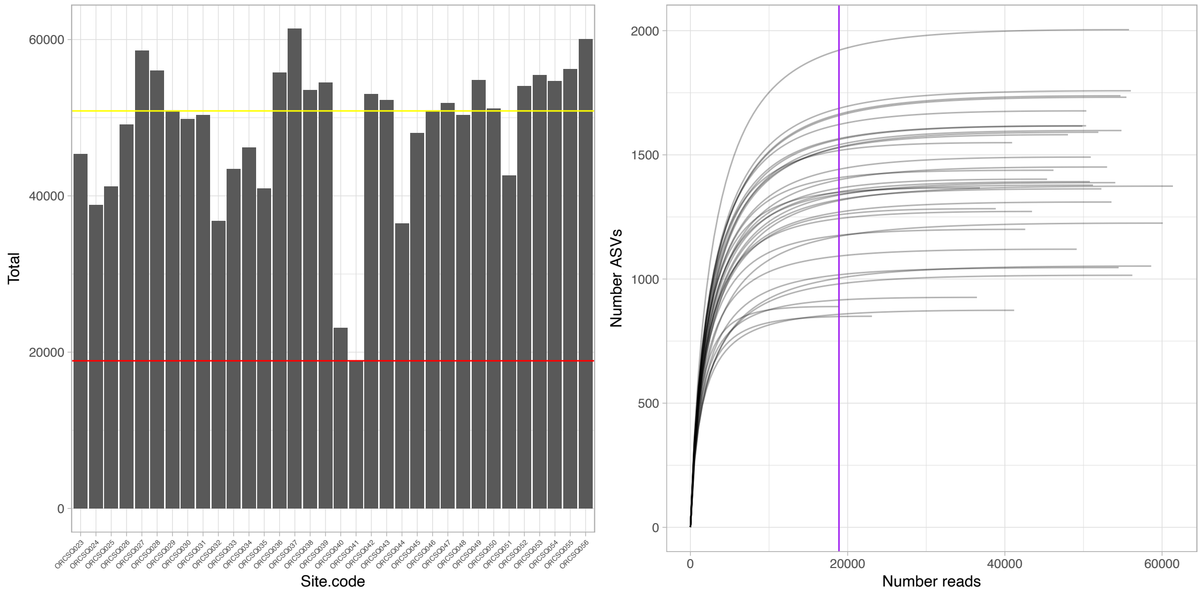
Appendices



Appendix 1 Raw reads from fungi (ITS) per sample (left) and rarefaction curves (right). The red line denotes the minimum number of reads for sample, the yellow denotes the median number of reads per sample, and the purple line is the minimum number of reads that the samples were rarefied to.

Table A1 Fungal ASVs that were exclusively found in all samples for particular land use types (B value = 1) of one group (A value = 1)

ASV	group_combination	A	B	Lineage
ITSv_0092	Native Forest	1	1	k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Cantharellales;f__Hydnaceae;g__Clavulina;s__subrugosa
ITSv_0118	Wetland	1	1	k__Fungi;p__Ascomycota;c__Sordariomycetes;o__Sordariales;f__Lasiosphaeriaceae;g__Schizothecium;NA
ITSv_0121	Dairy	1	1	k__Fungi;p__Basidiomycota;c__Tremellomycetes;o__Trichosporonales;f__Trichosporonaceae;g__Apiotrichum;NA
ITSv_0196	Wetland	1	1	k__Fungi;p__Ascomycota;c__Saccharomycetes;o__Saccharomycetales;f__Phaffomycetaceae;g__Cyberlindnera;s__saturnus
ITSv_0533	Arable	1	1	k__Fungi;p__Basidiomycota;c__Atractiellomycetes;o__Atractiellales;f__Hoehnelomycetaceae;g__Hoehnelomycetaceae_gen_Incertae_sedis;NA
ITSv_0584	Dairy	1	1	k__Fungi;p__Ascomycota;c__Sordariomycetes;o__Sordariales;f__Neoschizotheciaceae;g__Cercophora;NA
ITSv_0840	Wetland	1	1	k__Fungi;p__Ascomycota;c__Dothideomycetes;o__Pleosporales;f__Massarinaceae;g__Stagonospora;NA
ITSv_0847	Native Forest	1	1	k__Fungi;p__Ascomycota;c__Leotiomycetes;o__Helotiales;f__Lachnaceae;g__Lachnellula;NA
ITSv_0853	Native Forest	1	1	k__Fungi;p__Ascomycota;c__Leotiomycetes;o__Helotiales;f__Myxotrichaceae;g__Oidiodendron;s__chlamydosporicum
ITSv_0942	Wetland	1	1	k__Fungi;p__Basidiomycota;c__Cystobasidiomycetes;o__Cystobasidiomycetes_ord_Incertae_sedis;f__Microsporomycetaceae;g__Microsporomycetaceae_gen_Incertae_sedis;NA
ITSv_1093	Arable	1	1	k__Fungi;p__Ascomycota;c__Sordariomycetes;NA;NA;NA;NA
ITSv_1257	Dairy	1	1	k__Fungi;p__Mortierellomycota;c__Mortierellomycetes;o__Mortierellales;f__Mortierellaceae;g__Mortierella;s__indohii
ITSv_1447	Wetland	1	1	k__Fungi;p__Ascomycota;c__Sordariomycetes;NA;NA;NA;NA
ITSv_1562	Native Forest	1	1	k__Fungi;p__Ascomycota;c__Leotiomycetes;o__Helotiales;NA;NA;NA
ITSv_1595	Wetland	1	1	k__Fungi;p__Ascomycota;c__Sordariomycetes;o__Sordariomycetes_ord_Incertae_sedis;f__Junewangiaceae;g__Sporidesmiella;NA
ITSv_1619	Dairy	1	1	k__Fungi;p__Basidiomycota;c__Tremellomycetes;o__Trichosporonales;f__Trichosporonaceae;g__Cutaneotrichosporon;s__moniliiforme
ITSv_1635	Wetland	1	1	k__Fungi;p__Ascomycota;c__Dothideomycetes;o__Minutisphaerales;f__Minutisphaeraceae;g__Minutisphaera;s__aspera
ITSv_2045	Wetland	1	1	k__Fungi;p__Ascomycota;c__Leotiomycetes;o__Helotiales;f__Discinellaceae;g__Varicosporium;NA
ITSv_2495	Wetland	1	1	k__Fungi;p__Basidiomycota;c__Microbotryomycetes;o__Kriegeriales;f__Camptobasidiaceae;g__Camptobasidiaceae_gen_Incertae_sedis;NA
ITSv_4273	Wetland	1	1	k__Fungi;p__Chytridiomycota;c__Rhizophydiomycetes;o__Rhizophydiales;f__Rhizophydiales_fam_Incertae_sedis;g__Rhizophydiales_gen_Incertae_sedis;NA
ITSv_5392	Wetland	1	1	k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Agaricales;NA;NA;NA

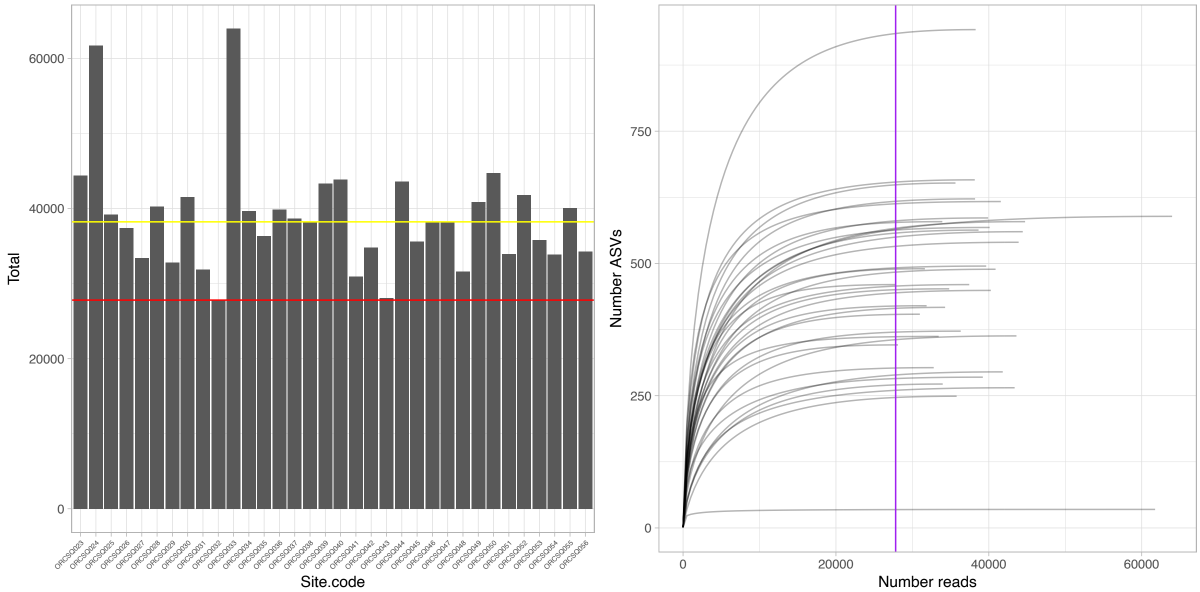


Appendix 2 Raw reads from bacteria (16s) per sample (left) and rarefaction curves (right). The red line denotes the minimum number of reads for sample, the yellow denotes the median number of reads per sample, and the purple line is the minimum number of reads that the samples were rarefied to.

Table A2 Bacterial ASVs that were exclusively found in all samples (B value = 1) of one group (A value = 1)

ASV	group_combination	A	B	Lineage
16Sv_0003	Arable+Dairy+Drystock+Forestry+Horticulture+Native Forest+Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Xanthobacteraceae;Bradyrhizobium;NA
16Sv_0005	Arable+Dairy+Drystock+Forestry+Horticulture+Native Forest+Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Xanthobacteraceae;Bradyrhizobium;NA
16Sv_0035	Arable+Dairy+Drystock+Forestry+Horticulture+Native Forest+Wetland	1	1	Bacteria;Planctomycetota;Planctomycetes;Isosphaerales;Isosphaeraceae;Aquisphaera;NA
16Sv_0460	Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae;Roseiarcus;fermentans
16Sv_0740	Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Hyphomicrobiaceae;Hyphomicrobium;NA
16Sv_1029	Native Forest	1	1	Bacteria;Planctomycetota;Planctomycetes;Isosphaerales;Isosphaeraceae;Aquisphaera;NA
16Sv_1248	Wetland	1	1	Bacteria;Verrucomicrobiota;Verrucomicrobiia;Pedosphaerales;Pedosphaeraceae;ADurb.Bin063-1;NA
16Sv_1273	Native Forest	1	1	Bacteria;Acidobacteriota;Vicinamibacteria;Vicinamibacterales;NA;NA;NA
16Sv_1672	Wetland	1	1	Bacteria;Acidobacteriota;Acidobacteriae;Bryobacterales;Bryobacteraceae;Paludibaculum;NA
16Sv_1699	Native Forest	1	1	Bacteria;Planctomycetota;Planctomycetes;Gemmatales;Gemmataceae;NA;NA
16Sv_1830	Horticulture	1	1	Bacteria;Myxococcota;Myxococcia;Myxococcales;Myxococcaceae;NA;NA
16Sv_1848	Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Hyphomicrobiaceae;Hyphomicrobium;NA
16Sv_1849	Wetland	1	1	Bacteria;Verrucomicrobiota;Verrucomicrobiia;Pedosphaerales;Pedosphaeraceae;NA;NA
16Sv_1914	Horticulture	1	1	Bacteria;Pseudomonadota;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;NA
16Sv_1976	Horticulture	1	1	Bacteria;Pseudomonadota;Gammaproteobacteria;Burkholderiales;Comamonadaceae;NA;NA
16Sv_2303	Wetland	1	1	Bacteria;Chloroflexota;Anaerolineae;Aggregatilineales;Aggregatilineaceae;NA;NA
16Sv_2322	Horticulture	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Ellin6055;NA
16Sv_2373	Wetland	1	1	Bacteria;Actinomycetota;Actinobacteria;Frankiales;Acidothermaceae;Acidothermus;NA
16Sv_2377	Horticulture	1	1	Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;Chitinophagaceae;NA;NA
16Sv_2442	Wetland	1	1	Bacteria;Actinomycetota;Acidimicrobiia;Acidimicrobiales;Acidimicrobiaceae;Acidiferrimicrobium;NA
16Sv_2503	Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Incertae Sedis;Incertae Sedis;Hypericibacter;NA
16Sv_2907	Wetland	1	1	Bacteria;Candidatus Kryptonia;Kryptonia;Kryptoniales;BSV26;NA;NA
16Sv_3074	Wetland	1	1	Bacteria;Pseudomonadota;Gammaproteobacteria;Burkholderiales;Methylophilaceae;NA;NA
16Sv_3404	Wetland	1	1	Bacteria;RCP2-54;NA;NA;NA;NA;NA
16Sv_3801	Dairy	1	1	Bacteria;Bacillota;Clostridia;Clostridiales;Clostridiaceae;Clostridium;neonatale
16Sv_3855	Horticulture	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingopyxis;NA
16Sv_3989	Dairy	1	1	Bacteria;Actinomycetota;Actinobacteria;Streptosporangiales;Thermomonosporaceae;Actinocorallia;NA
16Sv_4204	Horticulture	1	1	Bacteria;Pseudomonadota;Gammaproteobacteria;Lysobacterales;Rhodanobacteraceae;NA;NA
16Sv_4235	Arable	1	1	Bacteria;Gemmatimonadota;Gemmatimonadia;Gemmatimonadales;Gemmatimonadaceae;NA;NA
16Sv_4256	Horticulture	1	1	Bacteria;Verrucomicrobiota;Verrucomicrobiia;Opitutales;Opitutaceae;Opitutus;NA
16Sv_4258	Horticulture	1	1	Bacteria;Actinomycetota;Thermoleophilia;Solirubrobacterales;67-14;NA;NA
16Sv_4311	Wetland	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;KF-JG30-B3;NA;NA
16Sv_4327	Horticulture	1	1	Bacteria;Pseudomonadota;Gammaproteobacteria;Burkholderiales;Comamonadaceae;Caenimonas;NA
16Sv_4437	Wetland	1	1	Bacteria;Planctomycetota;Planctomycetes;Isosphaerales;Isosphaeraceae;Singulisphaera;NA

16Sv_4806	Horticulture	1	1	Bacteria;Actinomycetota;Actinobacteria;Pseudonocardiales;Pseudonocardiaceae;Lentzea;NA
16Sv_5463	Horticulture	1	1	Bacteria;Bacteroidota;Bacteroidia;Cytophagales;Microscillaceae;NA;NA
16Sv_5641	Wetland	1	1	Bacteria;Verrucomicrobiota;Verrucomicrobiia;Chthoniobacterales;Terrimicrobiaceae;FukuN18 freshwater group;NA
16Sv_5643	Wetland	1	1	Bacteria;Candidatus Kryptonia;Kryptonia;Kryptoniales;BSV26;NA;NA
16Sv_5778	Wetland	1	1	Bacteria;Nitrospirota;4-29-1;NA;NA;NA;NA
16Sv_5779	Wetland	1	1	Bacteria;Planctomycetota;Planctomycetes;Gemmatales;Gemmataceae;NA;NA
16Sv_5996	Wetland	1	1	Bacteria;Bacteroidota;Bacteroidia;Sphingobacteriales;KD3-93;NA;NA
16Sv_6785	Horticulture	1	1	Bacteria;Pseudomonadota;Alphaproteobacteria;Hyphomicrobiales;Beijerinckiaceae;Enterovirga;NA
16Sv_7333	Wetland	1	1	Archaea;Halobacteriota;Methanocellia;Methanocellales;Methanocellaceae;Methanocella;NA
16Sv_9554	Wetland	1	1	Bacteria;Chloroflexota;Anaerolineae;NA;NA;NA;NA

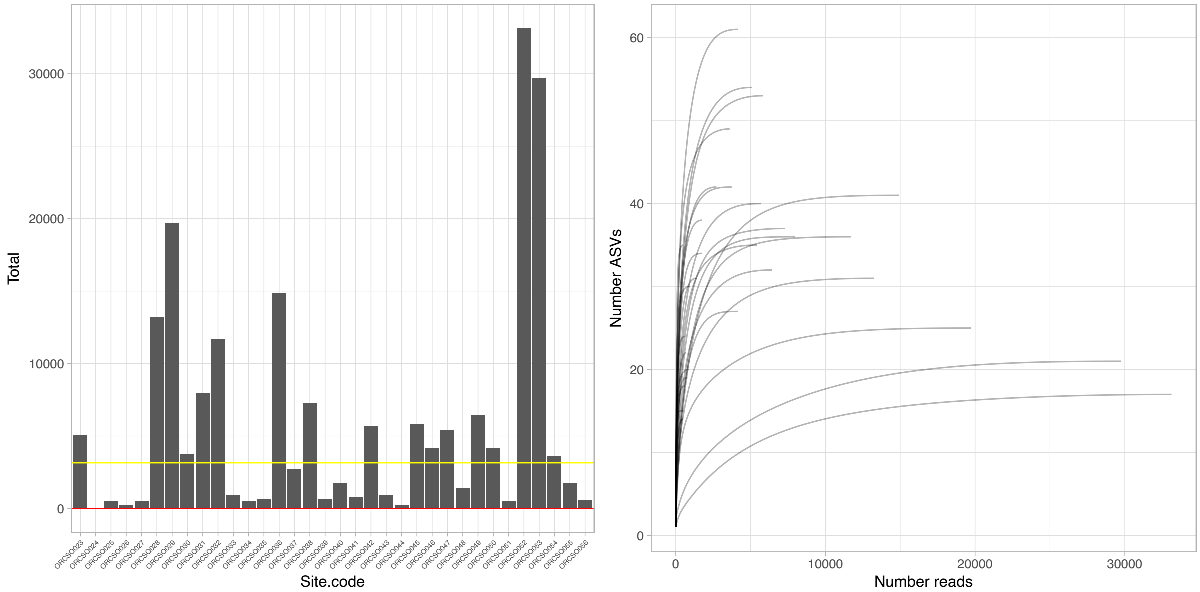


Appendix 3 Raw reads from all eukaryote ASVs (all COI) per sample (left) and rarefaction curves (right). The red line denotes the minimum number of reads for sample, the yellow denotes the median number of reads per sample, and the purple line is the minimum number of reads that the samples were rarefied to.

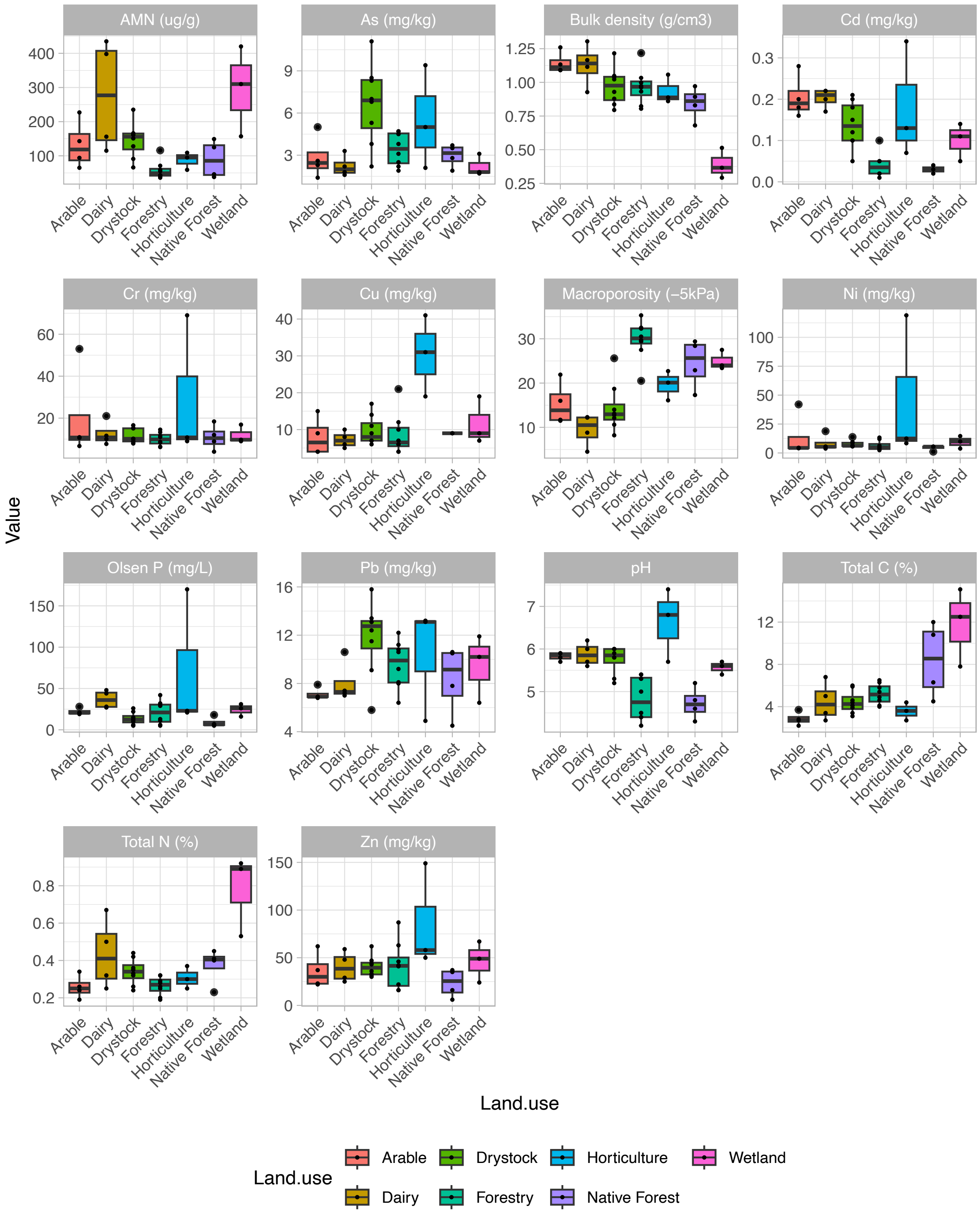
Eukaryote (COI) data – includes all ASVs from animals, fungi and plants

Because there was a large portion of non-target organism sequence data, the unfiltered COI dataset was not pursued further than analysis of read depth and rarefaction curves. The plots have been included for brevity and comparison to the invertebrate only data.

The COI data had good read coverage and rarefaction curves indicated that most taxa were discovered, with the exception of sample 24 which had low richness despite the high read coverage (*Appendix 3*). Over 90% of the ASVs were unassigned when compared to the BOLD database. This was largely due to non-target organisms being amplified during PCR, with Ascomycota ASVs making up the majority.



Appendix 4 Raw reads from ASVs classified to phylum per sample (left) and rarefaction curves (right). The red line denotes the minimum number of reads for sample, the yellow denotes the median number of reads per sample. This was transformed to presence/absence data as rarefaction is not suitable due to the patch coverage.



Appendix 5 Physiochemical properties by land use.