Supplementary Methods: Large, old pools of carbon and microbial communities are present deep in soils under a temperate planted forest

Authors: Alexa K. Byers[1], Loretta G. Garrett[2], Charlotte Armstrong[3], Fiona Dean[2], Steve A Wakelin[3]

[1]Bioprotection Aotearoa, PO Box 85084, Lincoln University, Lincoln, 8014, New Zealand

[2]Scion, Private Bag 3020, Rotorua, 3046, New Zealand

[3]Scion, PO Box 29237, Riccarton, Christchurch, 8440, New Zealand

Corresponding author: Alexa K. Byers, alexa.byers@lincoln.ac.nz

- Supplementary methods for section "**2.1 Study site and soil collection"**

A permanent sample plot (PSP) within the Rua sub-catchment of Puruki Forest was targeted for sampling. Soil was collected just outside of the plot to avoid disturbance for longer term monitoring at the site. For sampling, 10 points spaced 2 metre (m) apart were set out along an 18 m transect, which was located along a 12° slope. At each sampling point, two paired 1 m soil cores were extracted using a motorised percussion soil sampler capable of collecting intact cores within a plastic sleeve; one core was taken for bulk density analysis and the used for DNA and chemistry analysis. Following extraction, core pairs were visually compared to check for considerable differences in soil colour and texture by depth (Figure A.2).

Soil cores protected by a plastic sleeve were transported back to the lab in cool, dark conditions. Care was taken to minimise disruption during transport (cores all remained intact from field to lab). Following transport, soil cores were divided into 10 cm increment samples. Incidences of compaction were adjusted for during the division of cores into depth increments. Cores were measured from the top (0 cm) and each 10 cm increment working towards the compaction zone was marked, stopping once reaching the compacted area. Then beginning at the bottom of the core, 10 cm increments were marked towards the compacted zone. What was left between the upper and lower measurements was classified as the compacted zone and was divided evenly between the remaining samples.

- Supplementary methods for section "**2.4 Microbial Amplicon Sequencing"**

Following the standard Earth Microbiome Project (EMP) protocols, the barcoded forward and reverse primers 515F-806R were used to amplify the 16S rRNA V4 region (Caporaso et al., 2012). Each 25uL PCR reaction consisted of 17.375 ul ddH_2O, 2.5 ul ExTaq Buffer, 5mM dNTPs, 10nM forward primer, 10nM reverse primer, 10ng DNA and 0.125uL TaKaRa Ex Taq polymerase. Thermocycling conditions were 3 mins at 94°C, then 35 cycles of 94°C for 45 seconds, 50°C for 60 seconds, 72°C for 90 seconds followed by a final extension of 72°C for 10 minutes. Nested PCRs were performed on DNA samples which initially failed to amplify using the non-barcoded forward and reverse primers 27F and 1492R. Thermocycling conditions for the nested PCR were 3 mins at 94°C, then 35 cycles of 95°C for 30 seconds, 57°C for 30 seconds, 72°C for 90 seconds followed by a final extension of 72°C for 10 minutes. Nested PCR products were then amplified using the barcoded 515F and 806R primers. The EMP's standard ITS amplicon protocol was followed to amplify the fungal ITS gene region using the barcoded primers ITS1f and ITS2 (Bokulich & Mills, 2013; Hoggard et al., 2018). PCR reactions followed the same

recipe as outlined above for 16S rRNA PCR amplification. Thermocycling conditions were 1 minute at 94°C, then 35 cycles of 94°C for 30 seconds, 52°C for 30 seconds, 72ºC for 60 seconds followed by a final extension of 72ºC for 10 minutes. DNA samples which failed to amplify following the first round of PCR were targeted for nested PCR using forward and reverse primers NSA3 and ITS4. Nested PCR products were then PCR amplified using the barcoded primers ITS1f and ITS2. Thermocycling conditions for the nested PCR were 3 minutes at 94°C, then 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 60 seconds followed by a final extension of 72ºC for 5 minutes. PCR products were visualised on 1.5% agarose gel for 20 minutes, cleaned using a magnetic bead cleaning kit and quantified using Qubit (Invitrogen). PCR amplicons were sent for Illumina 500 MiSeq 250bp (16S) and Illumina 600 MiSeq 300 bp (ITS) paired end sequencing at the Australian Genome Research Facility (AGRF; Melbourne, Australia). For the 16S amplicon libraries, eight no template control (NTC) samples which showed bands on agarose gel were sent for sequencing to check for contaminants.

- Supplementary methods for section "**2.5 Quantitative PCR"**

The absolute concentration of bacterial and fungal DNA in each soil DNA sample was determined by standard curves using AriaMx SYBR Green qPCR (Agilent Technologies). Broad range forward and reverse primers 338F/518R and ITS1F/5.8s were selected to target the 16S rRNA (V3-V4) and ITS gene region for quantification (Shahsavari et al., 2016). Standard curves were prepared using 5-fold 1:10 serial dilutions of 16S and ITS plasmid DNA which had an initial concentration (Qubit, Invitrogen) of 9.67E+09 (16S) and 1.87E+09 (ITS) copies per reaction. Prior to performing the qPCR, DNA extracts were normalised to 10ng/ul. Each 25ul reaction consisted of 17.375 uL ddH_2O, 2.5 uL ExTaq Buffer, 5mM dNTPs, 10µM forward primer, 10µM reverse primer, 10ng DNA and 0.125uL TaKaRa Ex Taq polymerase. Thermocycling conditions were an initial denaturation of 95ºC for 20 seconds, followed by 40 cycles of 95°C for 60 seconds and 60ºC for 20 seconds, and a final melting curve of 95°C for 15 seconds, 60°C for 60 seconds, and 95°C for 15 seconds. Following quantification, 16S and ITS rRNA copy numbers were standardised to copy number per ng DNA.

- Supplementary methods for section "**2.6.2 Microbial network analyses"**

Prior to performing network analysis, ASV tables were filtered to retain only those with a total relative abundance of at least 0.05% in a minimum of 3 samples. Prior to filtering, ASV tables were split into samples from topsoil (0 to 30 cm) and subsoil (30 to 100 cm) layers to obtain core communities specific to

each soil layer. Interkingdom (16S and ITS rRNA), bacterial (16S rRNA), and fungal (ITS rRNA) topsoil and subsoil ASV tables were then summarized to genus level prior to network construction. SPIEC-EASI (SParse InversE Covariance Estimation for Ecological Association Inference) networks were constructed using the Meinshausen–Buhlmann (MB) neighbourhood algorithm by *SpiecEasi* R (Kurtz et al., 2015). Network visualisation was performed using *igraph* R (Csárdi & Nepusz, 2006) and calculation of network statistics using *Cytoscape* 3.8.2 (Shannon et al., 2003). Negative edge weights were set to zero to retain only positive associations. Significant differences ($p < 0.05$) in network statistics (degree, average shortest path length, betweenness centrality, clustering coefficient, and neighbourhood connectivity) between topsoil and subsoil communities were tested using Wilcoxon signed rank tests (Mundra et al., 2021). The *degree* of each node (genus) in the network describes how many correlations each node forms with other nodes. The *average shortest path length* is the average number of edges shared between each node to each other node in the network (Faust & Raes, 2012). *Clustering coefficient* describes whether the network contains clusters of highly interconnected organisms which share a high degree of interactions and associations (Mundra et al., 2021). The *neighbourhood connectivity* of a node describes how many correlations its neighbouring nodes shared with other nodes in the network. *Betweenness centrality* describes the amount of control a node exerts over the interactions of other nodes in the network; this measure favours nodes that link subnetworks rather than those which lie inside the community (Yoon et al., 2006).

References

Bokulich, N.A., Mills, D.A., 2013. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl. Environ. Microbiol. 79, 2519-26. https://doi.org/10.1128/AEM.03870-12

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., & Bauer, M., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621-1624. https://doi.org/10.1038/ismej.2012.8

Csárdi, G., Nepusz, T., 2005. The igraph software package for complex network research. InterJournal Complex Systems 1695.

Faust, K., & Raes, J., 2012. Microbial interactions: from networks to models. Nat. Rev. Microbiol. 10, 538-550. https://doi.org/10.1038/nrmicro2832

Hoggard, M., Vesty, A., Wong, G., Montgomery, J. M., Fourie, C., Douglas, R. G., ... & Taylor, M. W., 2018. Characterizing the human mycobiota: a comparison of small subunit rRNA, ITS1, ITS2, and large subunit rRNA genomic targets. Front. Microbiol. 9, 2208. https://doi.org/10.3389/fmicb.2018.02208

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A., 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Comput. Biol. 11, e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Mundra, S., Kjonaas, O. J., Morgado, L. N., Krabberod, A. K., Ransedokken, Y., & Kauserud, H., 2021. Soil depth matters: shift in composition and inter-kingdom co-occurrence patterns of microorganisms in forest soils. FEMS Microbiol. Ecol. 97. https://doi.org/10.1093/femsec/fiab022

Shahsavari, E., Aburto-Medina, A., Taha, M., & Ball, A. S., 2016. A quantitative PCR approach for quantification of functional genes involved in the degradation of polycyclic aromatic hydrocarbons in contaminated soils. MethodsX 3, 205-211. https://doi.org/https://doi.org/10.1016/j.mex.2016.02.005

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498-2504. https://doi.org/10.1101/gr.1239303

Smith, D. P., & Peay, K. G., 2014. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. PLOS ONE 9, e90234. https://doi.org/10.1371/journal.pone.0090234

Yoon, J., Blumer, A., & Lee, K., 2006. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics 22, 3106-3108. https://doi.org/10.1093/bioinformatics/btl533