Appendix S1 Extended methodology for the soil DNA metabarcoding analysis

Samples were sent to AllGenetics & Biology SL (A Coruña, Spain) for DNA metabarcoding analysis. DNA analysis included DNA extraction; library preparation with 2 pairs of different primers specific for barcoding (ITS for fungi and 16S for bacteria); sequencing on an Illumina NovaSeq PE250 platform (assuming an amplicon size < 450 bp); and standard bioinformatic analysis including demultiplexing, quality control and sample preprocessing, taxonomic assignment, number of reads per taxon and generation of rarefaction curves.

DNA isolation and quantification

DNA was isolated from each soil sample using the DNeasy PowerSoil Pro DNA isolation kit (Qiagen) following the manufacturer's instructions, and DNA was eluted in a final volume of 50 L. In one sample the above-mentioned protocol did not yield a DNA extract of enough concentration, so DNA was isolated using the FastDNA Spin Kit for Soil (MPBio) following the manufacturer's instructions, and DNA was eluted in a final volume of 70 L. An extraction blank (Bex) was included in each round of DNA extraction and treated as plot samples to check for contamination. DNA concentration in each extract was quantified using the Qubit High Sensitivity dsDNA Assay (Thermo Fisher Scientific).

DNA metabarcoding library preparation and sequencing

For fungal library preparation, a fragment of the nuclear ITS1 region of around 280-300 bp (including the primer sequences) was amplified using the following primers: forward - ITS1f (5' CTTGGTCATTTAGAG-GAAGTAA 3') (Gardes and Bruns, 1993); reverse - ITS2 (5' GCTGCGTTCTTCATCGATGC 3') (White et al., 1990). For prokaryotic library preparation, a fragment of the 16S rRNA gene of around 300 bp (including primer sequences) was amplified using the following primers: forward - 515F-Y (5' GTGYCAGCMGCCGCG-GTAA 3') (Parada et al., 2016); reverse - 806RB (5' GGACTACNVGGGTWTCTAAT 3') (Apprill et al., 2015). These primers also included the Illumina sequencing primer sequences attached to their 5' ends. In the first amplification step, PCRs were carried out in a final volume of 12.5 L, containing 1.25 L of template DNA (1:2 diluted in PCR round 3; not diluted in PCR round 11), 0.5 M of the primers, 3.13 L of Supreme NZYTaq 2x Green Master Mix (NZYTech), and ultrapure water up to 12.5 L. The fungal reaction mixture was incubated as follows: an initial denaturation at 95 ºC for 5 min, followed by 35 cycles of 95 ºC for 30 s, 49 ºC for 45 s, 72 ºC for 45 s, and a final extension step at 72 ºC for 7 min. The bacterial reaction mixture was incubated as follows: an initial denaturation at 95 ºC for 5 min, followed by 25 cycles of 95 ºC for 30 s, 47 ºC for 45 s, 72ºC for 45 s, and a final extension step at 72 ºC for 7 min. The oligonucleotide indices that are required for multiplexing different libraries in the same sequencing pool were attached in a second amplification step with identical conditions but only 5 cycles and 60 ºC as the annealing temperature. A negative control that contained no DNA (BPCR) was included in every PCR round to check for contamination during library preparation. Library size was verified by running the libraries on 2 % agarose gels stained with GreenSafe (NZYTech) and imaging them under UV light. Then, libraries were purified using the Mag-Bind RXNPure Plus magnetic beads (Omega Bio-tek), following the instructions provided by the manufacturer. Finished libraries were pooled in equimolar amounts according to the results of a Qubit dsDNA HS Assay (Thermo Fisher Scientific) quantification. The pool was sequenced in a fraction of a NovaSeq PE250 flow cell (Illumina) aiming for a total output of 2 gigabases. Any potential traces of adapter dimers were removed

using Cutadapt v3.5 (Martin, 2011). Quality of the sequencing data was assessed with the software FastQC (Andrews, 2010).

Processing of sequencing data and inference of ASVs

Cutadapt v3.5 was used to trim off non-biological DNA sequences (primers, indices, and sequencing adapters). Then, amplicon reads were processed using the DADA2 pipeline (Callahan et al., 2016) implemented in QIIME 2 (release 2023.7) (Bolyen et al., 2019). The first step in the DADA2 pipeline was to remove PCR primers and filter reads according to their quality. After checking the read quality profiles, forward reads were truncated at position 175, and reverse reads at position 131. The next step was to denoise and infer Amplicon Sequence Variants (ASVs). Error rates were learned from the dataset to denoise, using the parametric error model implemented in DADA2. Dereplication of the dataset (i.e. combination of all identical reads into unique reads) was carried out to reduce computational effort. These dereplicated forward and reverse reads were used to infer ASVs with the core sample inference algorithm (Callahan et al., 2016). Subsequently, corresponding R1 and R2 reads were merged into pairs with a minimum overlap of 12 identical base pairs. The final step in the DADA2 pipeline was to reduce the impact of artefacts (chimaeras) in the dataset. The final output of the DADA2 pipeline was a table containing the count of reads of every observed ASV in each sample, as well as a table listing all ASVs and their corresponding representative sequences.

Taxonomic assignment

Taxonomic assignment of each ASV was performed using a pre-trained classifier of the UNITE reference database for fungi (Abarenkov et al., 2023) (release May 2021); and of the SILVA reference database for bacteria (Quast et al., 2012) (release 138.1 August 2020). To compare the representative sequences of the ASVs to the reference database, and compute the taxonomic assignment, the algorithm sklearn (Bokulich et al., 2018) implemented in QIIME 2 was applied, setting a minimum confidence level of 0.8. Several filters were applied to exclude: (1) singletons (i.e. ASVs containing only one member sequence in the whole data set); (2) ASVs occurring at a frequency below 0.01 % in each sample, to reduce the mistagging bias (Bartram et al., 2016); (3) from the bacterial dataset, non-prokaryotic ASVs such as eukaryotic sequences of plastid (Moore et al., 2019) and mitochondrial (Emelyanov, 2001) origin; (4) those assigned only at domain (Bacteria) or kingdom (Fungi) level; (5) ASVs present in the negative controls that were more abundant in the negative controls than in the samples, indicating contamination (Salter et al., 2014).

Alpha rarefaction plots

Alpha rarefaction plots showed that the sequencing depth to reach saturation was 30,000 reads for the fungi and 22,000 reads for the bacteria, above the sequencing depth of all plots for fungi (mean = 114,919 reads; min = 79,045; max = 157,110) and bacteria (mean = 113,341 reads; min = 84,094; max = 171,404).

References

Abarenkov, K., Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., May, T.W., Frøslev, T.G., Pawlowska, J., Lindahl, B., Põldmaa, K., Truong, C., Vu, D., Hosoya, T., Niskanen, T., Piirmann, T., Ivanov, F., Zirk, A., Peterson, M., Cheeke, T.E., Ishigami, Y., Jansson, A.T., Jeppesen, T.S., Kristiansson, E., Mikryukov, V., Miller, J.T., Oono, R., Ossandon, F.J., Paupério, J., Saar, I., Schigel, D., Suija, A., Tedersoo, L., Kõljalg, U., 2023. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. Nucleic Acids Research 52, D791–D797. https://doi.org/10.1093/nar/gkad1039

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac f1000research.

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.

Bartram, J., Mountjoy, E., Brooks, T., Hancock, J., Williamson, H., Wright, G., Moppett, J., Goulden, N., Hubank, M., 2016. Accurate sample assignment in a multiplexed, ultrasensitive, high-throughput sequencing assay for minimal residual disease. The Journal of Molecular Diagnostics 18, 494–506. https://doi.org/10.1016/j.jmoldx.2016.02.008

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., Hooft, J.J.J. van der, Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., Hippel, M. von, Walters, W., others, 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9

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, 581–583. https://doi.org/10.1038/nmeth.3869

Emelyanov, V.V., 2001. Evolutionary relationship of Rickettsiae and mitochondria. FEBS Letters 501, 11–18. https://doi.org/10.1016/S0014-5793(01)02618-7

Gardes, M., Bruns, T.D., 1993. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. Molecular Ecology 2, 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.

Moore, K.R., Magnabosco, C., Momper, L., Gold, D.A., Bosak, T., Fournier, G.P., 2019. An expanded ribosomal phylogeny of Cyanobacteria supports a deep placement of plastids. Frontiers in Microbiology 10, 1612. https://doi.org/10.3389/fmicb.2019.01612

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, 1403–1414. https://doi.org/10.1111/1462-2920.13023

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

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology 12, 87. https://doi.org/10.1186/s12915-014-0087-z

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics, in: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols. Academic Press, San Diego, pp. 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1