Supporting Information for

Atmospheric chemosynthesis may enable colonization of the driest place on Earth

Rachel A. Moore, Diana Boy, Jens Boy, Marcus A. Horn, Georg Guggenberger, Armando Azua-Bustos, and Christopher E. Carr

Rachel A. Moore.
E-mail: rmoore305@gatech.edu

This PDF file includes:

Supporting Information Text

1. Introduction

The file contains supplementary methods and data from a 400 km soil transect study (south-north) in the Atacama Desert, Chile. We elucidated the functional profiles of microbial carbon utilization and energy generation derived from intracellular and extracellular DNA (iDNA and eDNA) using a metagenomic approach, correlating them with various soil parameters. Detailed information on sampling, sample processing, iDNA and eDNA isolation, sequencing, data processing, and statistical methods is included. We also provide data on the sequencing yield and quality. Supplemental figures of measured soil parameters such as SOC, pH, texture size classes, electric conductivity, and soluble salts are provided, along with detailed plots of the metagenomic data and graphical outputs of partial redundancy analyses. Online supplemental datasets include: Dataset S1 (estimated functional gene abundances in the community), and Dataset S2 (all measured soil data).

Materials and Methods

Overall Study Location. We used satellite imagery of Google Earth Pro, Version 7.3.3.7786, as well as climate and weather data ranging from 1970-2016 obtained from various meteorological stations (data sources: Red CEDEUS (1), DGAC (2), Arriagada (3), and $(CR)^2$ (4)) across Chile to pre-select the sampling plots. The climate data indicated that mean annual precipitation decreases from South to North across the Atacama Desert, classifying the desert into semi-arid, arid, and hyper-arid regions based on latitude. Our five sampling plots followed the South to North aridity gradient approximately 40 km from the Pacific Ocean in the Central Valley and approximately 100 km from each other (Fig. 1). Four of the five plots were located in a flat, broad landscape with no significant exposures or slopes, except for the MES plot, which was located at the foot of a small hill covered with some rocks (Fig. S1). None of these locations were located in Wadi/Aguadas. GPS coordinates of the plots were between S22.79°S and S29.14°S and between W69.69°S and W70.88°S, ranging from UNEP-defined arid to hyperarid climate zones

Field Sampling Methodology. At each of the five plots, we opened a soil profile up to 60 cm of depth in February 2022, except for plot MES, where soil profiles in triplicate were opened (Fig. S2). In order to extrapolate the results into a larger area at each sampling plot and to test for the heterogeneity of the soil chemical parameters per plot, we added some soil data from three additional soil profiles of the plots DGA (SN-4), TAL (SN-5), YUN (SN-6), and NOA (SN-7) stemming from a sampling campaign in February 2016 (5). Every profile was sampled for soil chemistry at four different soil depths: 0-1 cm, 5–10 cm, 10-20 cm, and 50-60 cm. These soil samples were sieved in situ through a 2 mm mesh Retsch sieve and stored in sterile bags and sterile 50 mL plastic tubes at ambient temperature. Bulk density was sampled at each soil depth and in each soil profile using a soil core with a fixed volume of 100 cm^3. Soils were classified according to the IUSS Working Group World Reference Baseline (WRB; (6)).

Soil samples for microbiological analysis were collected in duplicate from each of the soil profiles at depths of 0-1, 5-10, 10-20, 30-40, and 50-60 cm below the surface. These samples were collected into sterile Whirl-Pak bags and stored at ambient temperature in insulated coolers until processed back in the lab. All sampling instruments such as sieves, hand shovels, and spoons, were cleaned using 70% Ethanol between every sampling step. In addition, we wore sterilized gloves and FFP-2 mouth protection to avoid contamination.

Specific Sampling Plots. Plot DGA (S26°23´15.2“, W70°04´17.6“) lies in the transition between the arid and the hyper-arid zone roughly 2.5 km uphill in the north-west of a small town called Diego de Amagro. Between depths of 10-20 cm a 5 cm thick hard layer of salt accumulation was observable, however, the salt layer was not continuously distributed across the soil, since we observed patches in the profiles without salt. At 30 cm of soil depth, a very rigid salt layer was found.

Plot TAL (S25°32´58.9“, W70°12´31.9“) lies in the hyper-arid zone, roughly 32 km air-line distance landwards from Taltal. Vegetation is completely absent. The structure of these soils suggests sedimentary processes, which were also shaped by hydrological processes in the distant past. We observed layers of lamellae not broader than 1 cm and not totally parallel pointing towards wind-driven sedimentation (water-driven sedimentation results in parallel deposition of material).

Plot YUN (S24°3´35.1“, W69°50´23.7“) lies 72 km southeast of the city of Antofagasta in the hyper-arid core of the Atacama Desert, Chile. Large nitrate deposits occur here, indicating the dryness of this region. Vegetation is totally absent and large, dark rocks are scattered on the surface. Between 10-40 cm of depth, we found rigid crusts of salts in all four soil profiles collected here (i.e., from our 2016 and present sampling campaigns).

Plot NOA (S22°47´48.2“, W69°36´18.5“) lies 125 km northeast of the city of Antofagasta. On the surface and up to a depth of 0.5-20 cm skeleton form a visible black overlay, however, the soil at the surface is loose. At some profiles, below 20 cm, the soil is free of skeleton, very soft, and appears to be of aeolian origin, although the skeleton on the surface indicates another source. In a looser layer soft, destructive salt nodules were found. We also found broad layers of soft, laminated-looking gypsum-like concretions with rosette-like crystal intergrowth as well as deep soil tongues that stretched downward up to 40 cm depth.

Plot MES (S22°15´38.82“, W63°43´28.78“) lies on the foot of some hills and is scattered with bigger rounded boulders. In between, there are patches where the surface is covered only with tiny stones or pebbles. The first cm of a soil layer are crusts. We did not encounter any salt layers at MES and generally, the soil layering resembled that of plot TAL.

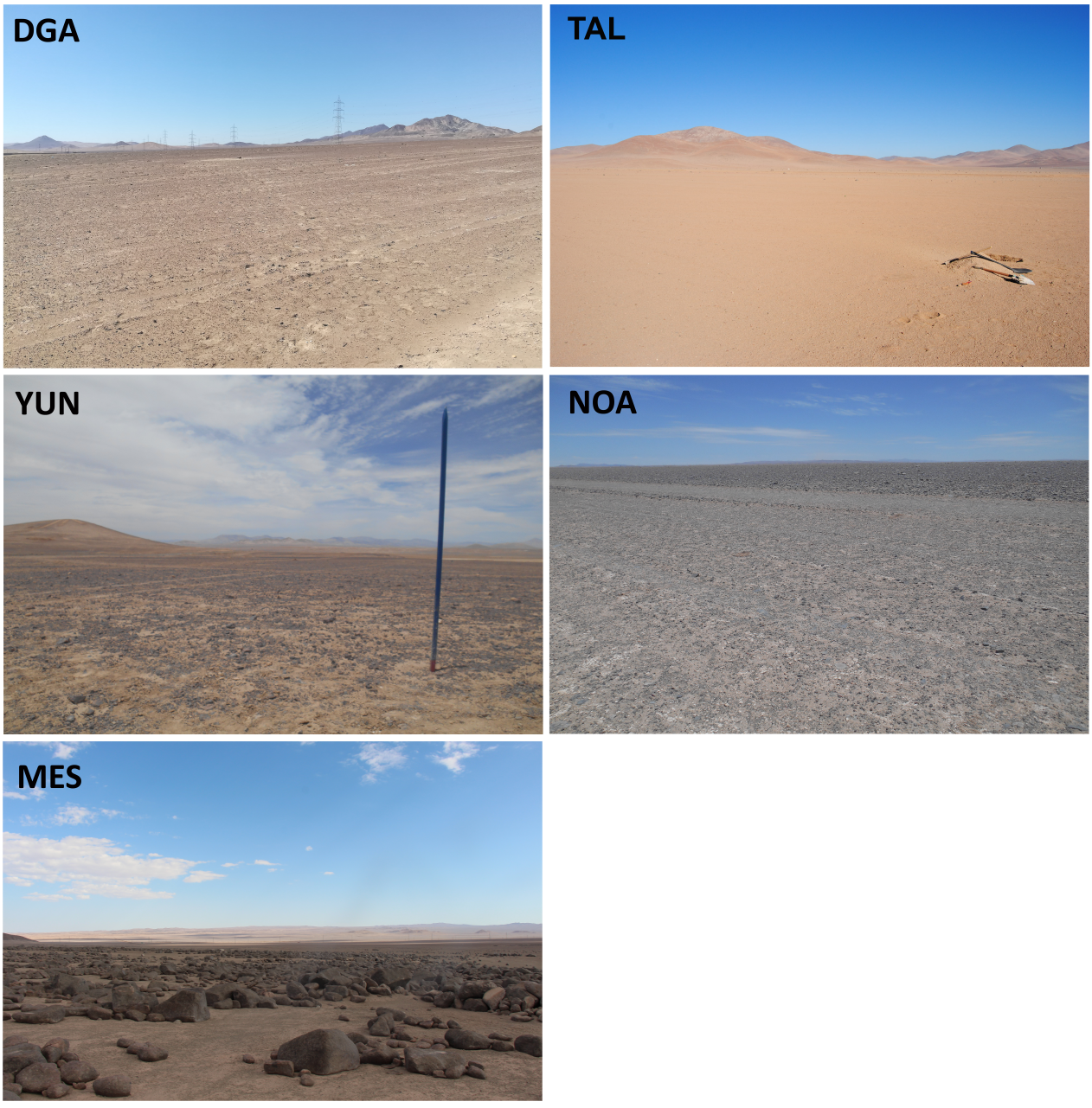

Fig. S1. Representative images taken at each of the five locations along the aridity transect (from left to right): DGA, TAL, YUN, NOA, and MES.

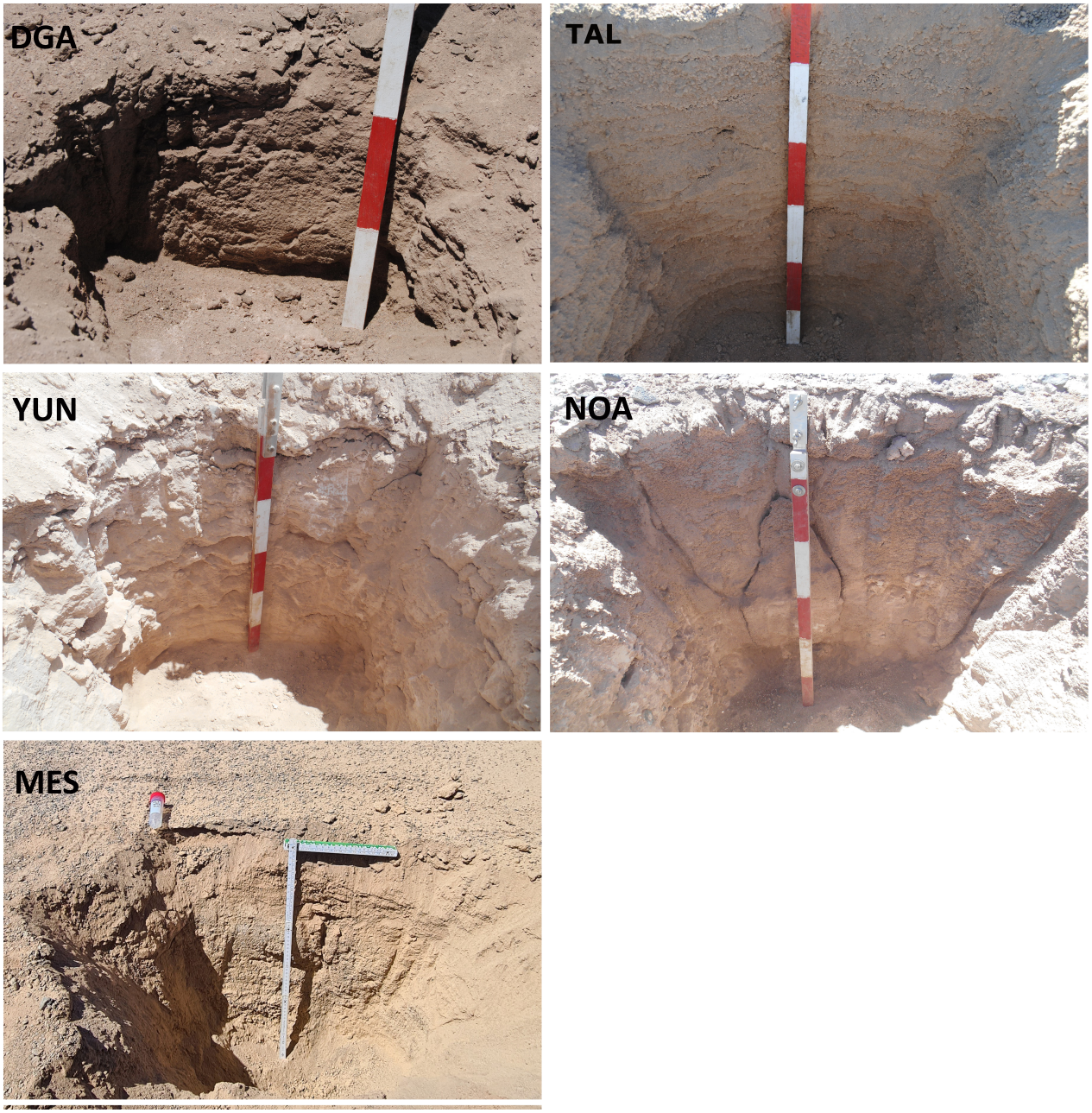

Fig. S2. The soil profiles opened at each of the five plots (from left to right): DGA, TAL, YUN, NOA, and MES. Each soil profile was opened up to 60 cm in depth.

iDNA and eDNA Isolation, NGS Preparation, and Sequencing

All steps, except for bucket rotor centrifugation, were carried out in a 1300 Series A2 Biological Safety Cabinet with several extra precautions taken for these low biomass samples. These precautions included wiping down the biological safety cabinet with 70% ethanol, running the UVC light in the cabinet for a full cycle, and running the cabinet empty for 10 minutes with the sash open before and after use. Soil sample bags, DNA extraction kits, reagents, and tip boxes were only opened within the biological safety cabinet. Filtered tips were used for all pipetting steps, and pipettors were washed with 70% ethanol and exposed to a full UVC cycle in the cabinet prior to use.

iDNA and eDNA isolation. DNA isolations were carried out using a modified version of the protocol from Schulze-Makuch et al. (7) as follows. The primary modification to the Schulze-Makuch protocol was the use of 31 grams of soil per sample, as opposed to the 6.5 grams utilized in the original protocol. Soil samples were added to 50 mL sterile conical tubes, and mixed with 20 mL of 1:1 m/v NaPP/Tween 80 (30 mM NaPP and 0.1% Tween 80. The NaPP solution was filtered through a 0.2 μM filter and autoclaved prior to mixing with sterile Tween 80). The conical tubes were gently shaken to wet all of the soil with the NaPP/Tween 80 solution and then chilled on ice for 1 minute.

Next, the tubes were horizontally mixed on a vortex for five minutes at 150 rpm, chilled on ice for 1 minute, and shaken again horizontally for five minutes at 100 rpm. The conical tubes were then centrifuged at 500 × g for 10 minutes in a bucket rotor. The supernatants were transferred to new sterile 50 mL conical tubes and kept chilled on ice. The soil was then resuspended in 10 mL of the NaPP/Tween 80 solution, and the mixing, chilling, and centrifugation steps from above were repeated. The supernatants were pooled with the prior supernatants. The final combined supernatants, containing both the iDNA and eDNA portions, were then centrifuged at 3,000 × g for 1 hour in a bucket rotor to pellet the intact cells. After pelleting, the supernatants were transferred to new 50 mL conical tubes and individually concentrated to 150 μl using an INNOVAPREP concentrating pipette CP Select™System and INNOVAPREP irradiated ultrafilter tips (SKU CC08003-10).

iDNA extraction. The intact cell pellets were resuspended in 1 mL of sodium phosphate buffer (MP Bio, cat. no. 6560205), transferred to a 1.5 mL LoBind tube, and centrifuged for 20 minutes at 12,000 × g to remove any residual eDNA, as described in Schulze-Makuch et al. (7). The resulting washed intact cell pellets were extracted using a ZymoBIOMICS DNA microprep kit with modifications for low biomass. Specifically, steps 4, 12, and 13 were omitted from the ZymoBIOMICS DNA microprep protocol (Zymo-Spin™III-F Filter and Zymo-Spin™II-μHRC Filter steps, respectively). The DNase/RNase-free water used to elute the DNA was warmed to 55 °C prior to use. Procedural controls were generated following the same methods above but without the addition of soil into the conical tubes.

NGS library preparation and sequencing . All samples, including the procedural controls, were prepared and sequenced at the Parker H. Petit Institute for Bioengineering and Bioscience Molecular Evolution Core at the Georgia Institute of Technology in Atlanta, GA. DNA sequencing libraries were generated using the NEBNext®Ultra™II FS DNA Library Prep Kit. The pooled libraries were evaluated by Bioanalyzer prior to sequencing. Libraries were sequenced on the Illumina NovaSeq6000 as a standalone run (Instrument: A01113; flow cell ID: HCLM3DRX3; SP flow cell; run type: PE150; cycles: 151 | 10 | 10 | 151).

Metagenomics Data Processing.

Trimming and taxonomic classification. FASTQ files were uploaded to Google Drive and classified using Kaiju (8) on Google Colab, and the Google Cloud using an e2-highmem-16 machine. First, Trimmommatic (9) was used to remove adapter sequences and quality trim paired-end raw sequencing reads. Quality trimming was performed with the code: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30. This trimmed any low-quality bases from the start and end of the read with a quality threshold of three. Sliding window trimming cut once the average quality within the window fell below a threshold of 15 over a window size of four bases. The MINLEN specification dropped reads that were shorter than 30 bases after trimming. Following Trimmomatic, FastQC (10) was used to perform quality control checks.

Kaiju classification was performed on the trimmed reads using the default run mode *Greedy* with three allowed mismatches. The cutoff for the E-value was set to 0.01. The cutoffs for minimum required match length and score were set to 11 and 65, respectively. The database used was *nr_euks* which is a subset of the NCBI BLAST *nr* database that contains all proteins belonging to bacteria, Archaea, fungi, and microbial eukaryotes. SEG filtering was enabled. The outputs were converted to raw data tables using the *kaiju2table* function for further preprocessing (below).

Metagenomics data preprocessing and statistical methods. All statistical analyses were conducted in R studio v.2021.09.0+351 "Ghost Orchid". The data tables produced using Kaiju were converted into one phyloseq object in R using libraries *phyloseq* and *tidyverse*. Certain names of bacterial phyla were changed (e.g., Firmicutes to Bacillota) to reflect the updated nomenclature (11) using base R functions *gsub* and *grepl*. The R library *decontam* (12) and our procedural control data were used to identify and remove contaminating DNA features statistically. The combined *decontam* methods of identifying contaminating features, frequency and prevalence, and a conservative threshold of 0.5 were used. Metagenomics data were analyzed for diversity and richness (Observed, Chao1, and Shannon) using *vegan*. Statistical methods PcoA, RDA, and ANOVA were also carried out using the R library *vegan*.

The function *decostand* from *vegan* was used to standardize species data with the Hellinger method (i.e., all values in a row were divided by the row sum, and then the square root was applied), and soil chemistry data using the log method. The *anova.cca* function from *vegan* was used to calculate the significance of the RDA model in figure 4 with step = 1000. Statistical

significance of soil chemistry variables to community composition (Pr(>F)) was determined using analysis of variance with the function *anova* from the *stats* package with 999 permutations.

RDA was applied using the *vegan* library to correlate matrices of Hellinger-transformed estimated functional gene abundances with the normalized soil data. For Fig. S21, the normalized K^+ and abundances of FeFe- and CooS-genes were removed from the dataset as they were not correlated with the other parameters. The *anova.cca* function from *vegan* was used to calculate the significance of the RDA models of the estimated functional genes with 999 permutations.

The Venn diagrams in figure 2 and figure S4 were generated using R library *MicEco*. The heatmap in figure 3 was produced using R libraries *pheatmap* and *ecodist*. All other plots were produced using a combination of *ggplot2*, *ggord*, *ggpubr*, *microViz*, and *patchwork*.

Functional gene classification and estimation of abundances in the community. We followed the methods of Ortiz et al. (13) to estimate the potential metabolic capability of Atacaman communities (i.e., percent of organisms estimated to have at least one copy of the given gene). We searched the metagenomes from this study against the Greening Lab metabolic marker gene databases (14) (https://doi.org/10.26180/c.5230745) using DIAMOND (v2.1.8) in blastx mode (15). The marker gene databases that we searched against were comprised of 25 conserved marker genes (Fig. 3), including those that are involved in carbon fixation (*AclB*, *AcsB*, *HbsC*, *HbsT*, *Mcr*, and *RbcL*), respiration (*AtpA*, *CoxA*, *CcoN*, *CyoA*, *CydA*, *NuoF*, *SdhA*, *FrdA*), phototrophy (*PsaA*, *PsbA*, and *RHO*), and trace gas metabolism (NiFe-, FeFe-, and Fe-hydrogenase large subunits; *CooS*, *CoxL*, *MmoA*, *PmoA*, *McrA*).

Like Ortiz et al. (13), we filtered the above results based on an identity threshold ($\geq$ 50% identity). We also screened for 14 universal single-copy ribosomal marker genes (S2 rpsB, L1 rplA, L2 rplB, S9 rplC, L14b/L23 e rplN, S5, S19 rps, S7, L16/L10E rpiPL6 rplF, L11 rplK, L5 rplE, S15p/S13e used by tools PhyloSift and SingleM (13, 16)). We used those results ($\geq$ 50% identity) to determine the estimated abundance of the functional genes within the microbial community. Specifically, the actual read count for each functional gene was divided by the total number of reads and then divided by the length of that gene in kilobases. The read counts for the universal single-copy marker genes were 'normalized' in the same way. Then, the normalized read count for the functional gene was divided by the mean of the normalized read counts of the universal single-copy genes (both in reads per kilobase per million, RPKM). This resulted in the estimated average copy number of the functional gene in the microbial community (**Dataset S1**).

Physico-chemical Soil Parameters. We analyzed the following basic soil parameters: Inorganic carbonates, soil organic carbon (SOC), (δ^{13}C), electric conductivity (EC), pH, water-extractable anions and cations, and the volumetric fraction of different grain sizes (**Dataset S2**) Inorganic carbonates ($CaCO_3$) were measured following the DIN-ISO 10693 norm. Milled soil samples were treated with 10% HCl which induced the dissolution of carbonates and led to the release of carbon dioxide which was kept in a closed calcimeter system. The volume of released gas was detected and the carbonate content (%) was calculated considering temperature and gas pressure.

SOC and the ratio of the ^{13}C to ^{12}C (δ^{13}C) were measured on milled samples by combustion with an Isotope Cube®, Elementar Analysensysteme GmbH, Germany, coupled to an Isoprime Mass Spectrometer®, Isoprime Ltd., Cheadle Hulme, UK. Present carbonates were dissoluted prior to the SOC measurement using the acid fumigation method of Harris et al. (17). Carbon contents were calculated for each of the different soil depths.

EC and pH were measured in a 1:5 volumetric ratio with pure water. The suspension was shaken overhead for 1 hour to allow the dissolution of ions and kept in peace for another hour. After calibration, the electric conductivity was measured in the supernatant and the pH in the same, but freshly stirred suspension.

Aliquots of the EC and pH solution were used to measure the water-extractable anions of NO^{3-}, Cl^-, PO_4^{3-} and SO_4^{2-} and cations such as Na^+, K^+ Mg^{2+}, Ca^{2+} via ion chromatography (IC; Dionex ion chromatograph, Thermo Fisher Scientific with DS5 detection stabilizer and CCRS 500 suppressor). Suspensions were centrifuged at 5 min at 739 g and filtered through a <0.45 µm cellulose acetate filter prior to measurement.

The volumetric fraction of the different grain sizes was measured using laser diffraction (Coulter LS 200, Beckman Coulter GmbH, Krefeld, Germany). The principle of this method is that large particles scatter light at low angle, while small particles scatter at high angles, which enables the particle size determination. Inorganic carbonates and iron oxides were removed prior to the analysis using HCl for the carbonates and dithionite for the iron oxides.

Supplemental Results and Discussion.

Sequencing yield and quality. Sequencing of the library with the Illumina NovaSeq6000 resulted in a yield of 246.25 Gbp, 88 GB of data in total. 140 samples were sequenced (seven plots, two of which were not a part of the N-S transect and not included in this paper × five depths × two replicates at each depth × two treatments: iDNA and eDNA = 140 sequenced samples). 88 GB / 140 samples ≈ 0.63 GB per sample. The percent of reads greater than or equal to a Quality Score of 30 (Q30) was 82.7%, indicating that the majority of the base calls were inferred to be 99.9% accurate.

Preprocessing with decontam and final read counts. The *decontam* R library statistically identified 2564/27688 eDNA and 5328/24924 iDNA taxa as contaminants within the samples using a conservative threshold of 0.5 and the 'combined' method. In these samples, the contaminant sequences represented a mean of 2.7±0.35% and 6.3±0.65% of the total sequences in eDNA and iDNA samples (±SEM), respectively. The final read counts (i.e., reads that were assigned) with contaminating sequences removed are depicted in figure S3.

The largest number of assigned reads were observed in samples from DGA (13,074,038 total reads) and TAL (12,986,254 total reads). Samples from northern plots NOA and MES had 80% and 74% fewer assigned reads when compared to DGA, with 2,524,039 and 3,329,298 total assigned reads, respectively. Plot YUN had the fewest assigned reads with 2,427,416 in total, an 81% decrease from DGA. In general, there were more assigned reads in the first three depths (i.e., 0-1, 5-10, and 10-20 cm below the surface; figure S3), but this difference was not significant (ANOVA, $Pr(>F) > 0.05$).

Microbial identities and abundances. The phyla by plot are described in the main text and depicted in Fig. 2. Fig. S6 depicts the same data but is additionally separated by depth. Figs. S7-11 display the relative abundances at the genera level for DGA, TAL, YUN, NOA, and MES.

Functional Analysis. While we filtered the results for sequence identity ($\geq 50\%$), the mean identity across all hits was 73%. The sequence identity varied among marker genes, with photoautotrophy-related genes showing mean identities of 72% (PsaA), 74% (PsbA), and 55% (RHO). For respiration-related genes, the mean sequence identities were 78% (AtpA), 92% (CcoN), 72% (CoxA), 79% (CydA), 64% (CyoA), 58% (FrdA), 67% (NuoF), and 67% (SdhA). In carbon fixation, the genes AclB, AcsB, HbsC, HbsT, Mcr, and RbcL had sequence identities of 54%, 50%, 50%, 50%, 51%, and 81%, respectively. For trace gas metabolism, the sequence identities were 51% (CooS), 57% (CoxL), 50% (Fe), 58% (FeFe), 51% (McrA), 51% (MmoA), 61% (NiFe), and 70% (PmoA). The mean overall bitscore and E-value were 60 and 3.66×10^{-5}, respectively.

Supplemental Figures

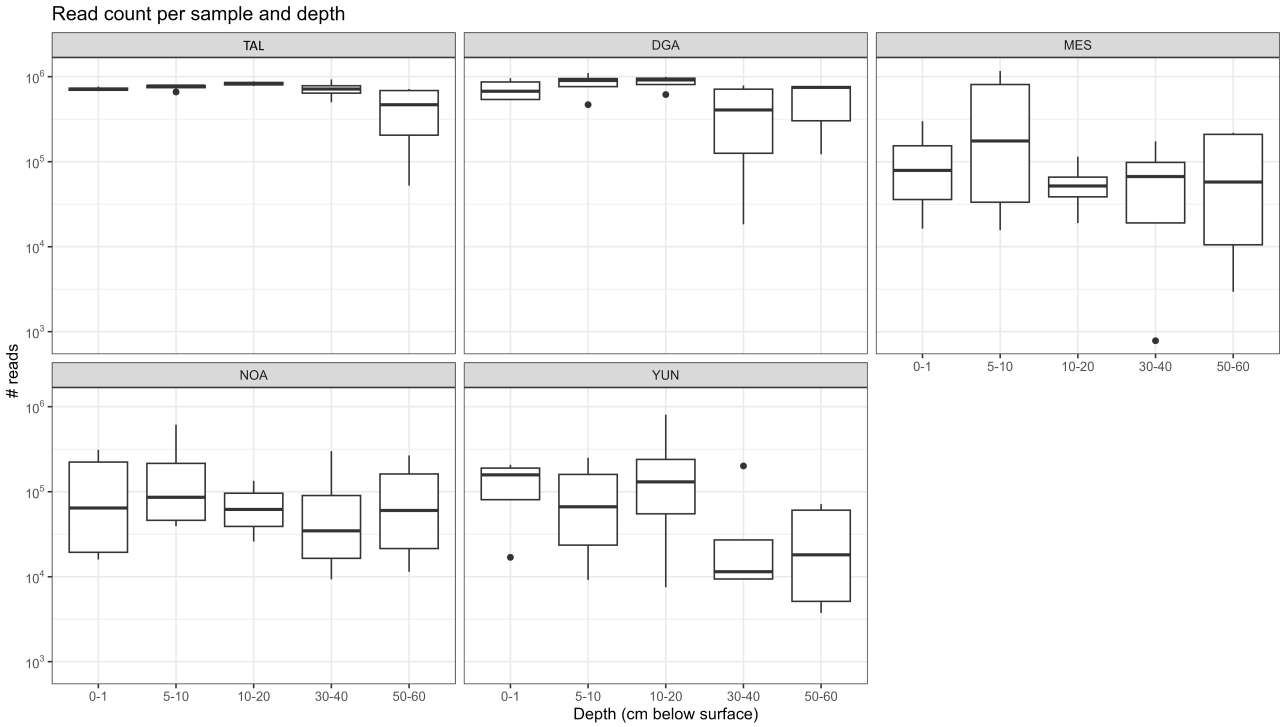

Fig. S3. Total assigned read counts by plot and sample depth.

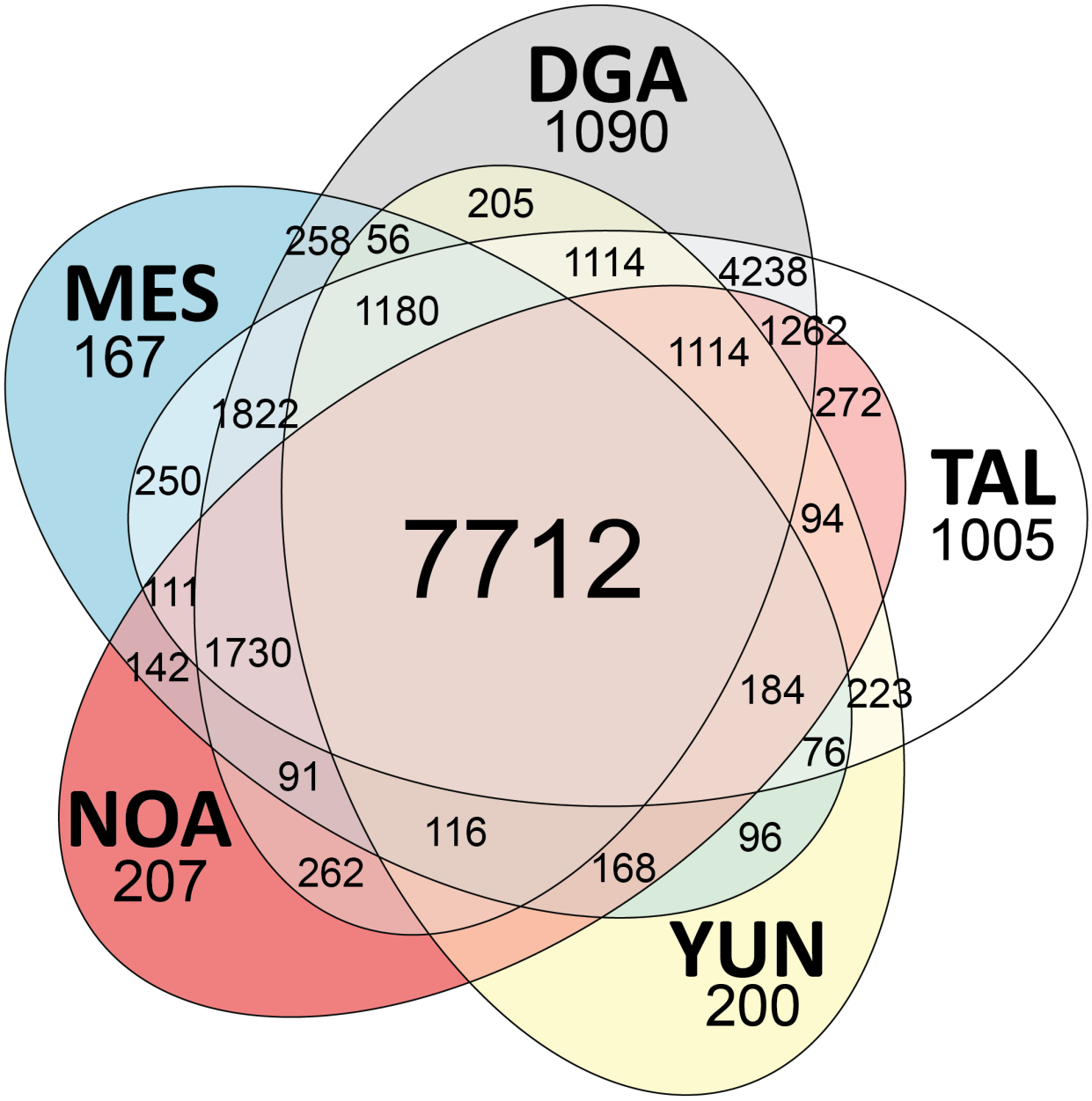

Fig. S4. Total (not agglomerated at any level, contains species and subspecies) classified iDNA taxa, including bacteria, Archaea, Fungi, and microbial Eukaryota, shared between plots.

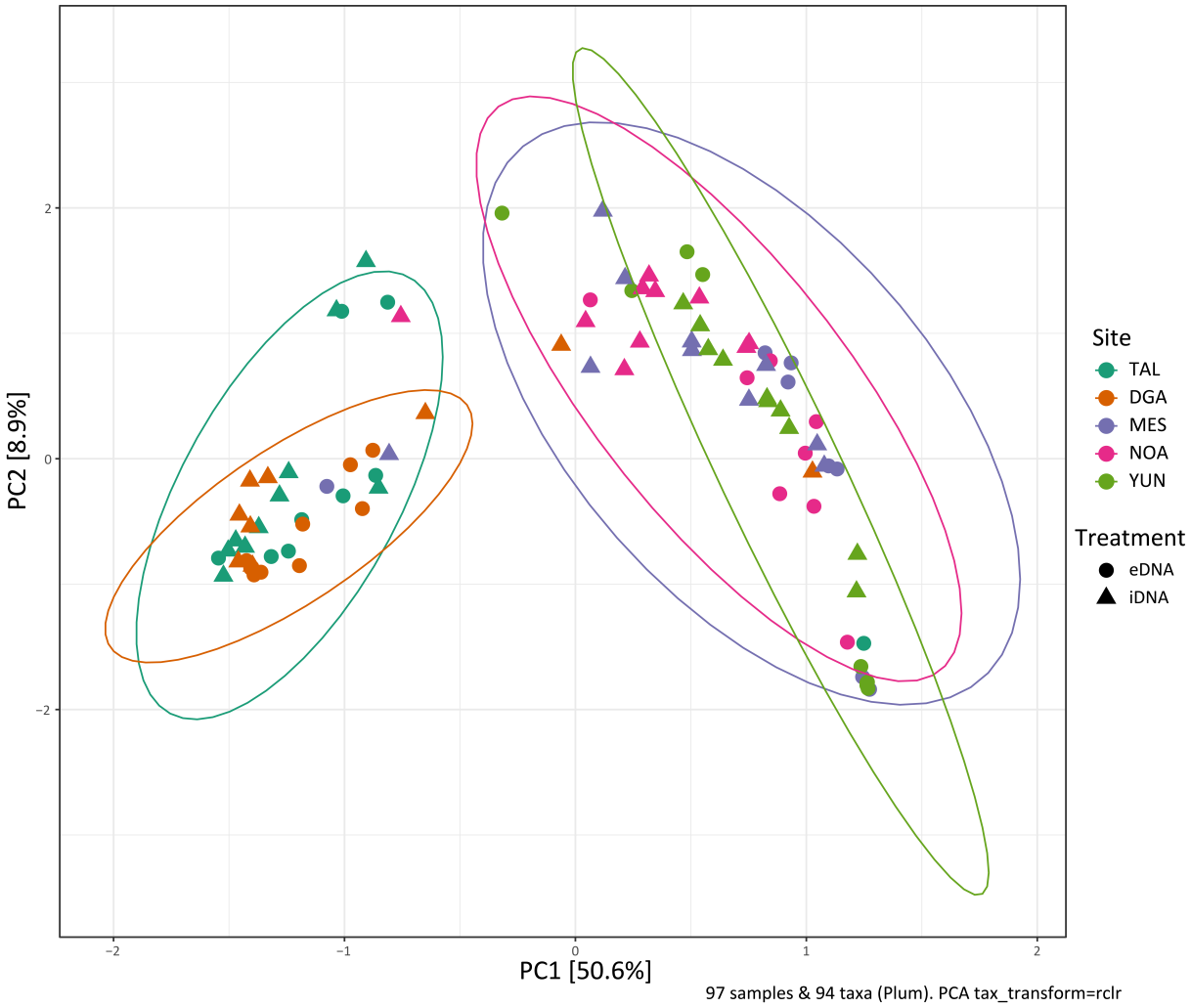

Fig. S5. PCA plot of the microbial communities at each plot based on the phylum level. PC1 (50.6%) and PC2 (8.9%) are the first two principal coordinates, together explaining 59.5% of the total variance in the data. Data at the phylum level were transformed using the robust centered log ratio transformation (rclr) from R library microViz. Plots are indicated by color while treatment (iDNA and eDNA) is indicated by shape. Ellipses indicate 95% confidence intervals around the centroid for each plot, calculated using the stat_ellipse function from the ggplot2 library.

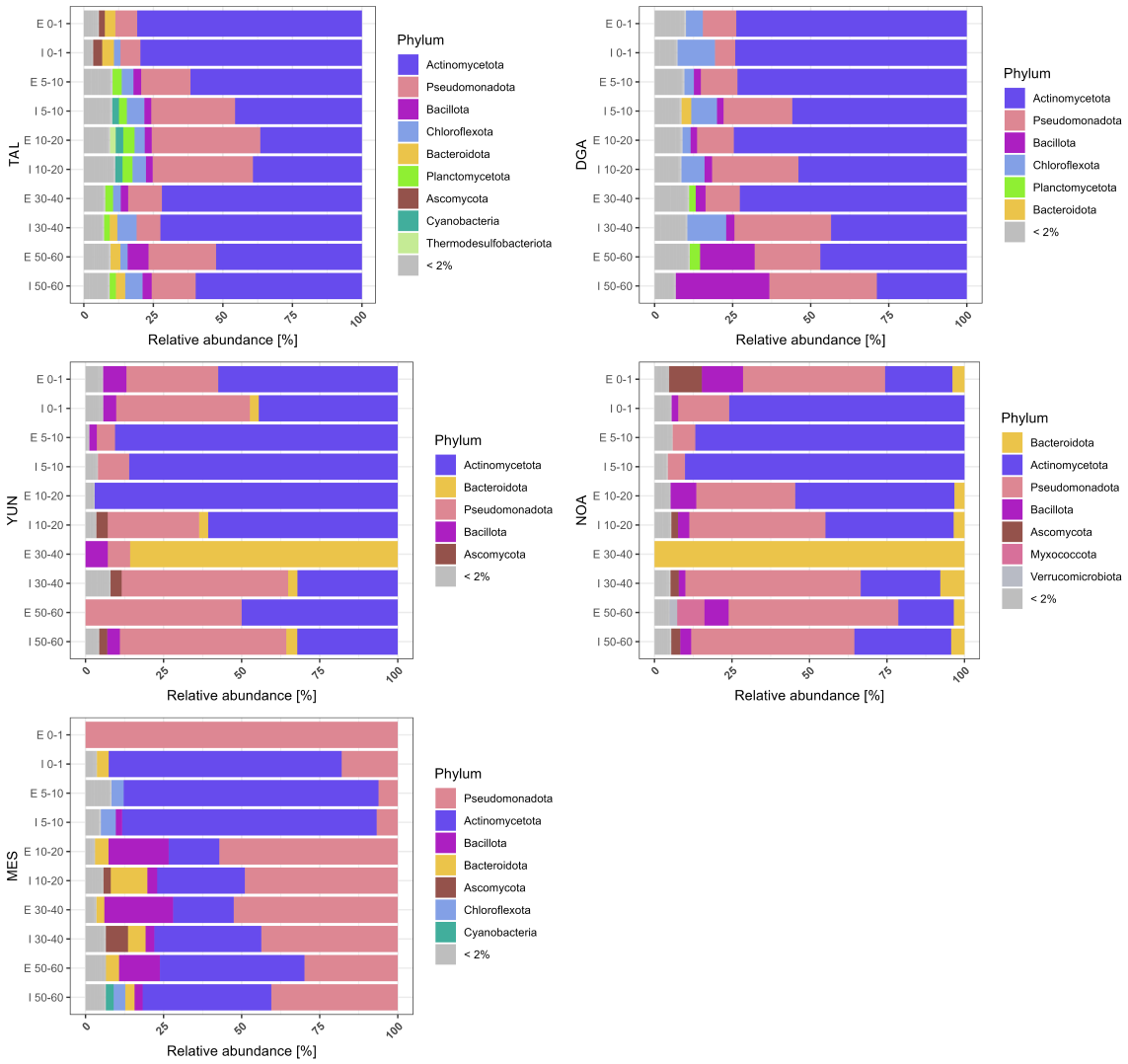

Fig. S6. Relative abundance (%) of all phyla by plot and depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 2%" category (light gray) combines all phyla whose relative abundance was less than 2% of the total for that sample and plot. Phyla colors are consistent across the plots.

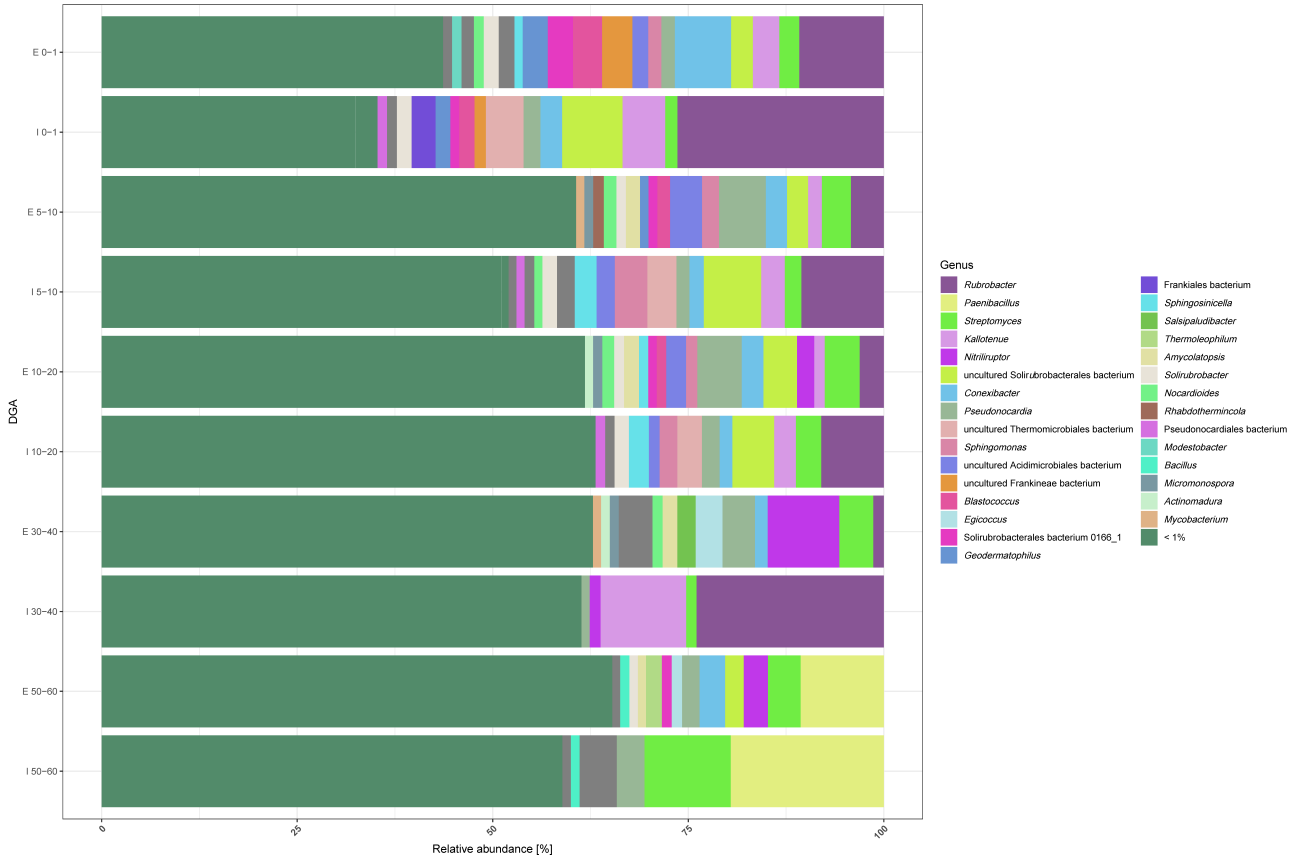

Fig. S7. Relative abundance (%) of all genera at DGA by depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 1%" category (forest green) combines all phyla whose relative abundance was less than 1% of the total for that sample.

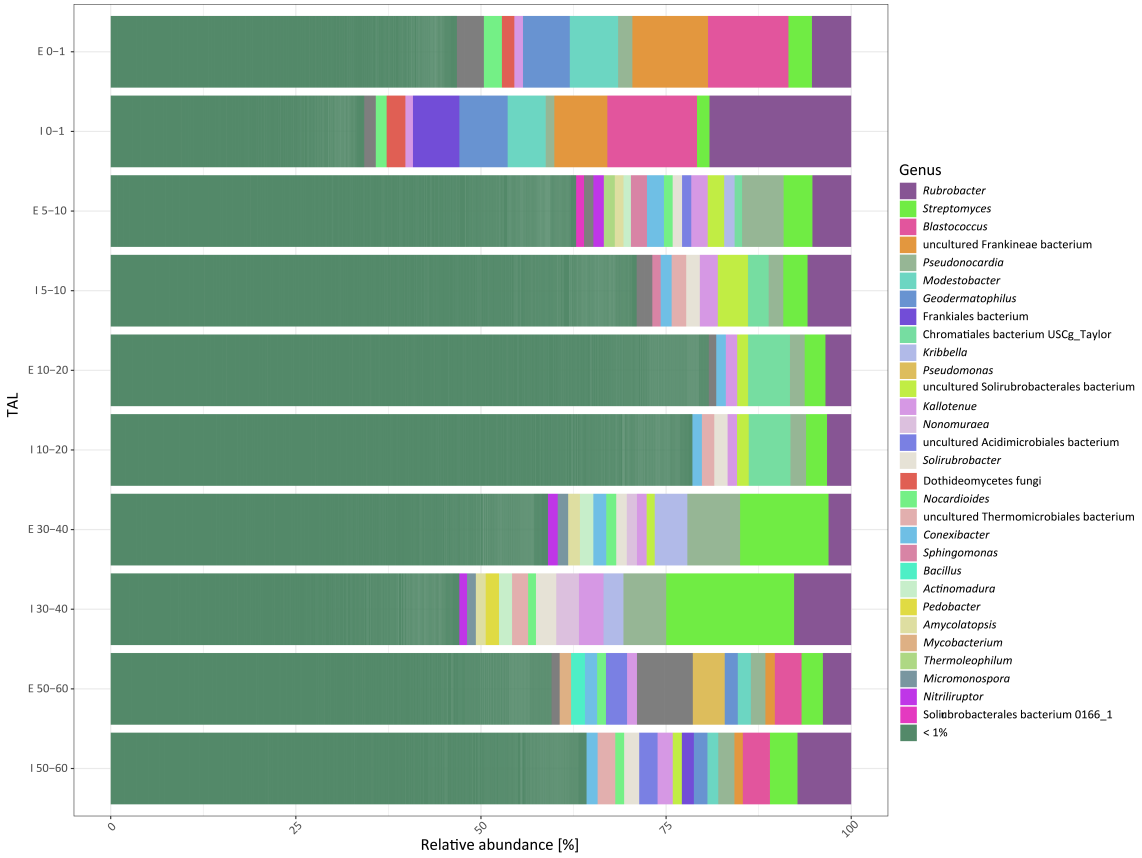

Fig. S8. Relative abundance (%) of all genera at TAL by depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 1%" category (forest green) combines all phyla whose relative abundance was less than 1% of the total for that sample.

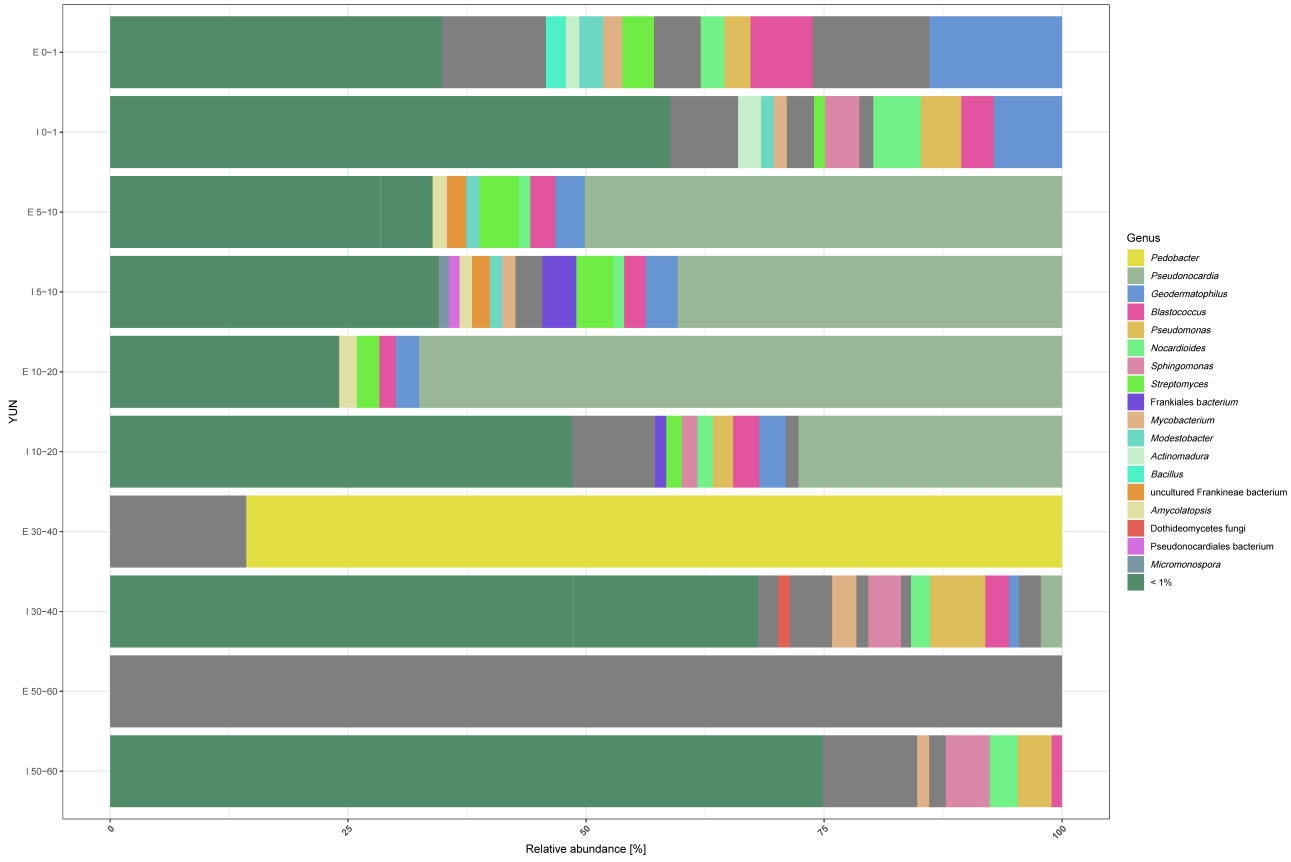

Fig. S9. Relative abundance (%) of all genera at YUN by depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 1%" category (forest green) combines all phyla whose relative abundance was less than 1% of the total for that sample.

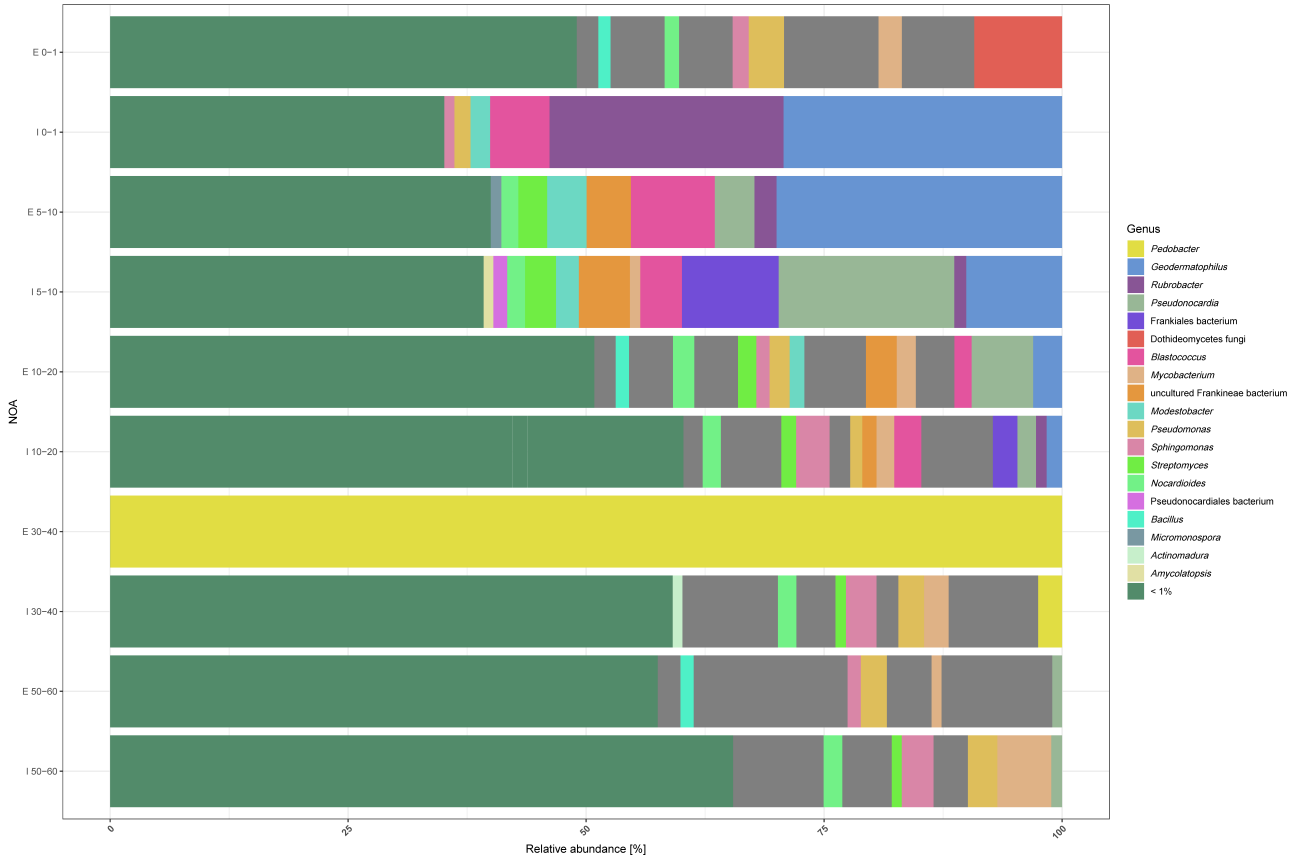

Fig. S10. Relative abundance (%) of all genera at NOA by depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 1%" category (forest green) combines all phyla whose relative abundance was less than 1% of the total for that sample.

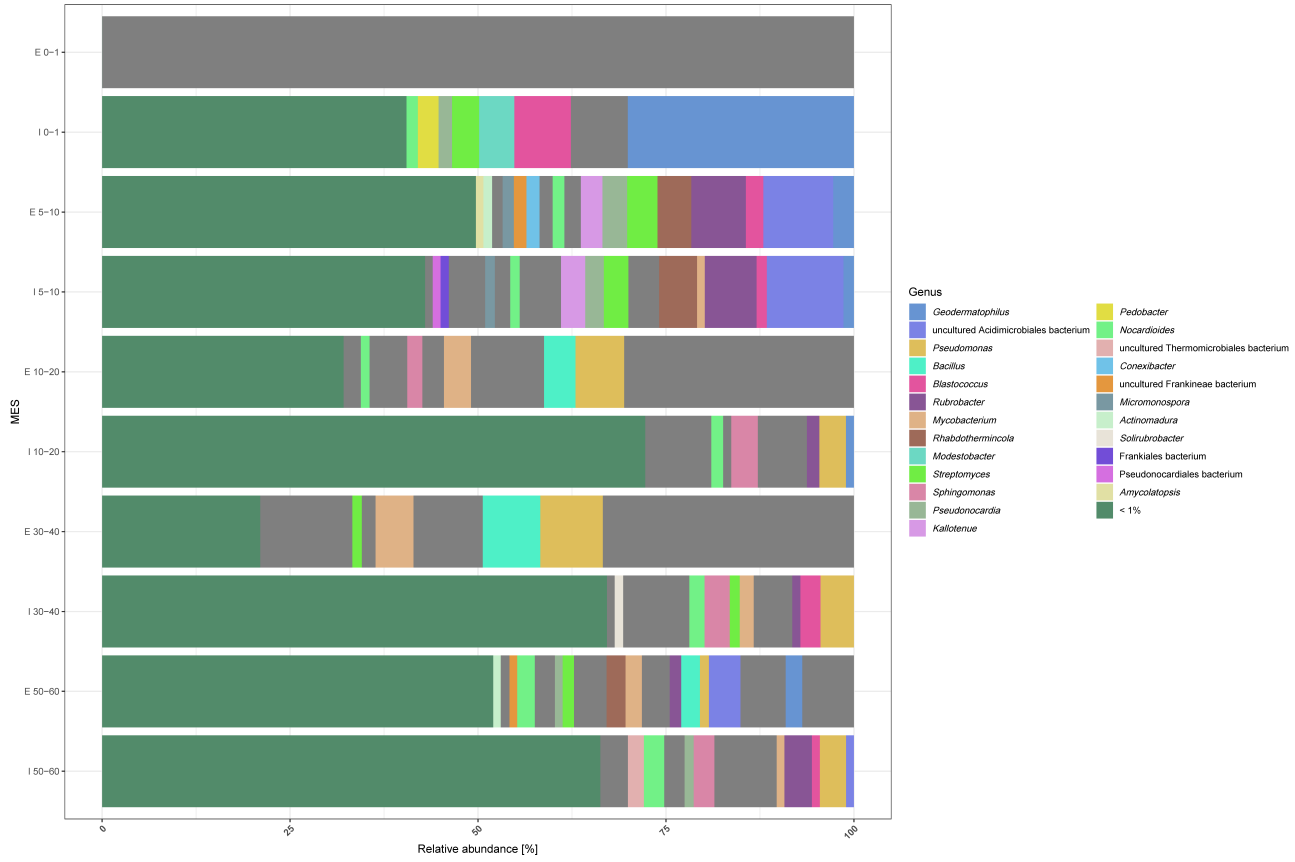

Fig. S11. Relative abundance (%) of all genera at MES by depth. iDNA and eDNA are denoted by an I or E on the Y-axis, respectively. Depths are listed as 0-1, 5-10, 10-20, 30-40, and 50-60 (cm below the surface) on the Y-axis. The "< 1%" category (forest green) combines all phyla whose relative abundance was less than 1% of the total for that sample.

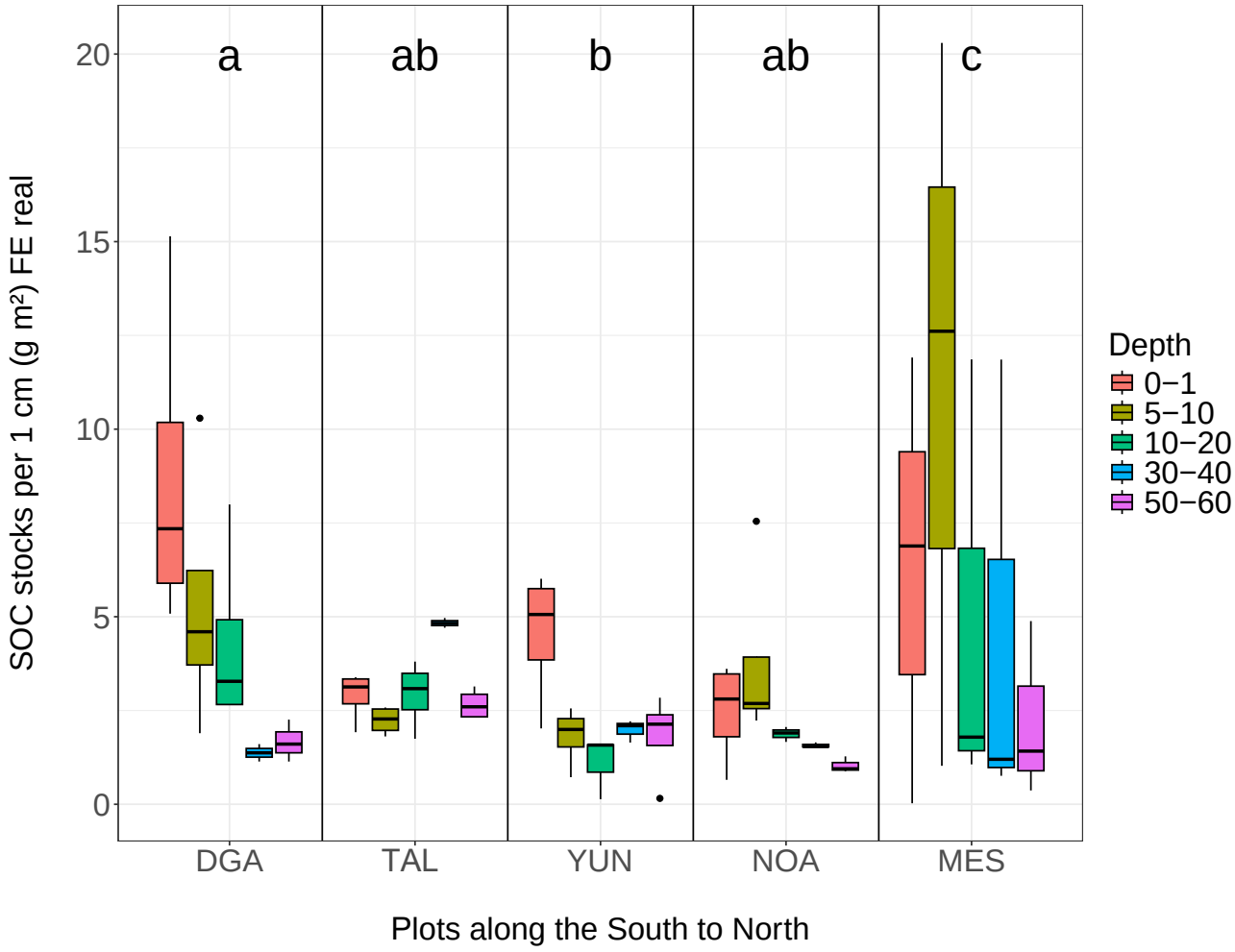

Fig. S12. Soil organic carbon (SOC) stocks along the gradient from South to North in the different depth increments. Largest stocks occurred at plot MES in 5-10 cm of depth (11.31 $\pm$ 7.92 per 1 cm [g m^{-2}] FE real) and lowest at plot NOA in 50-60 cm of depth with 1.04 $\pm$ 0.17 per 1 cm [g m^{-2}] FE real.

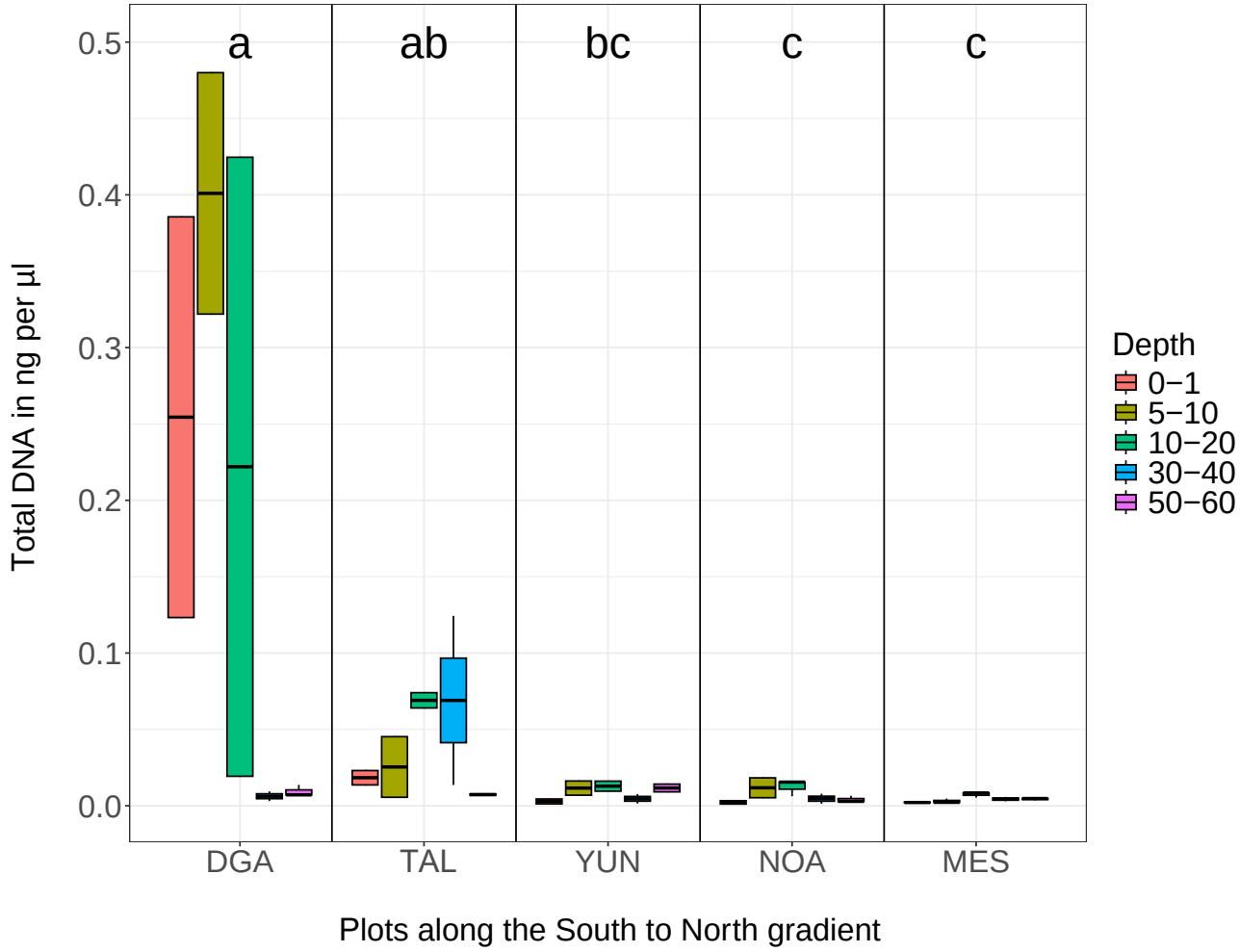

Fig. S13. DNA concentration along the South-North gradient. The concentration of extracted DNA was highest at plot DGA, which is the southernmost and "wettest" plot, and decreased towards plot MES, the northernmost plot ($p = 2.356 \times 10^{-7}$). Highest mean concentrations were found at DGA with 0.18 ± 0.077 ng μl^{-1} and lowest at plot MES with an average 0.0043 ± 0.0005 ng μl^{-1}. No statistical differences were found for the various depths of the plots and we also did not detect an overall decrease of DNA concentration with depth.

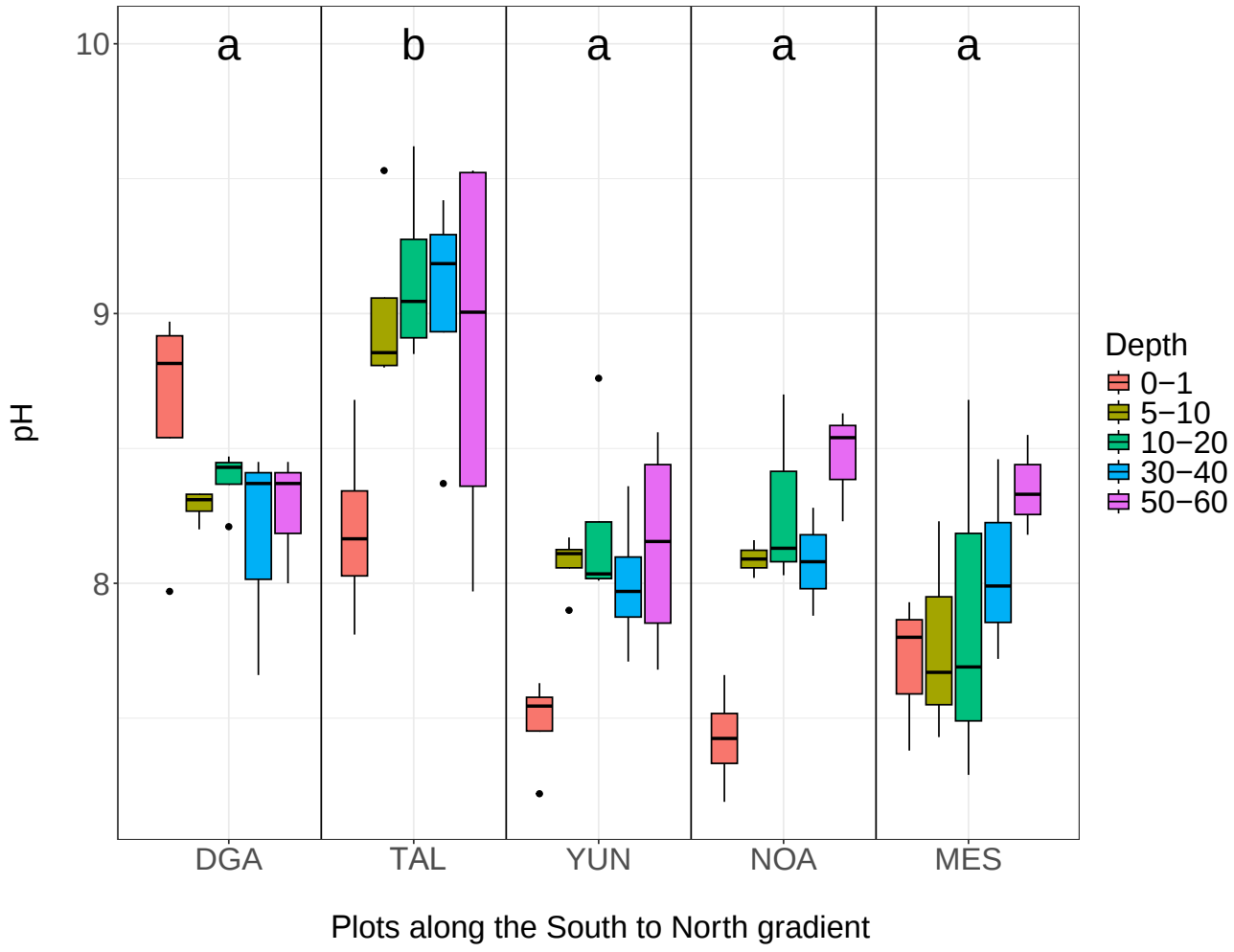

Fig. S14. pH along the South to North gradient including the different soil depths. pH ranged along the transect between 7.43 $\pm$ 0.17 at NOA at the surface and 9.14 $\pm$ 0.30 at TAL in 10-20 cm of soil depth. Between the soil depths, we found a difference at plots YUN and NOA, where pH was significantly lower at the surface compared to the deeper depths.

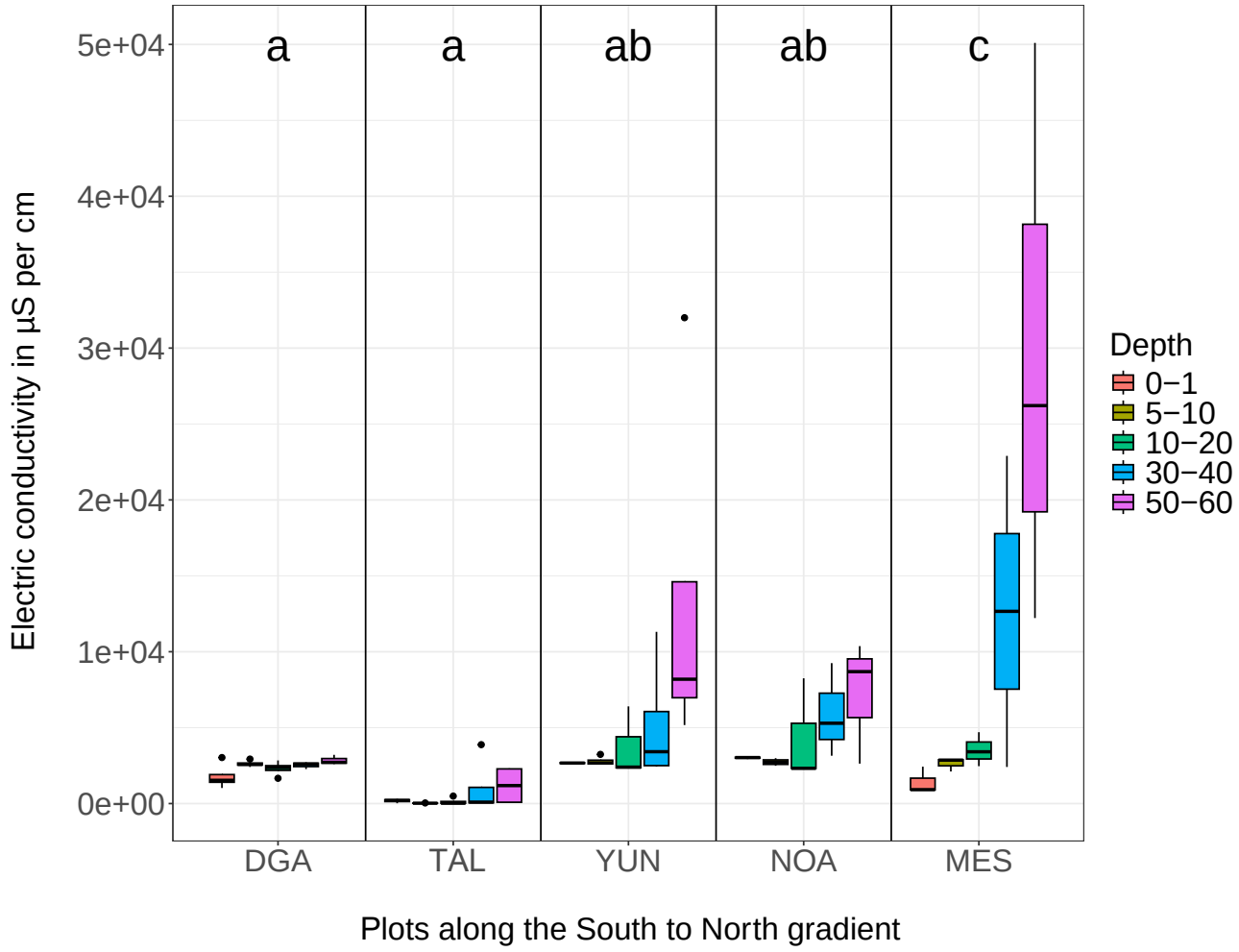

Fig. S15. Electric conductivity (EC) at our plots from south to north. EC differed significantly on plot level between TAL and all other plots and on depth level at YUN, NOA, and MES, however EC was not different along the transect in the single depths ($p > 0.05$).

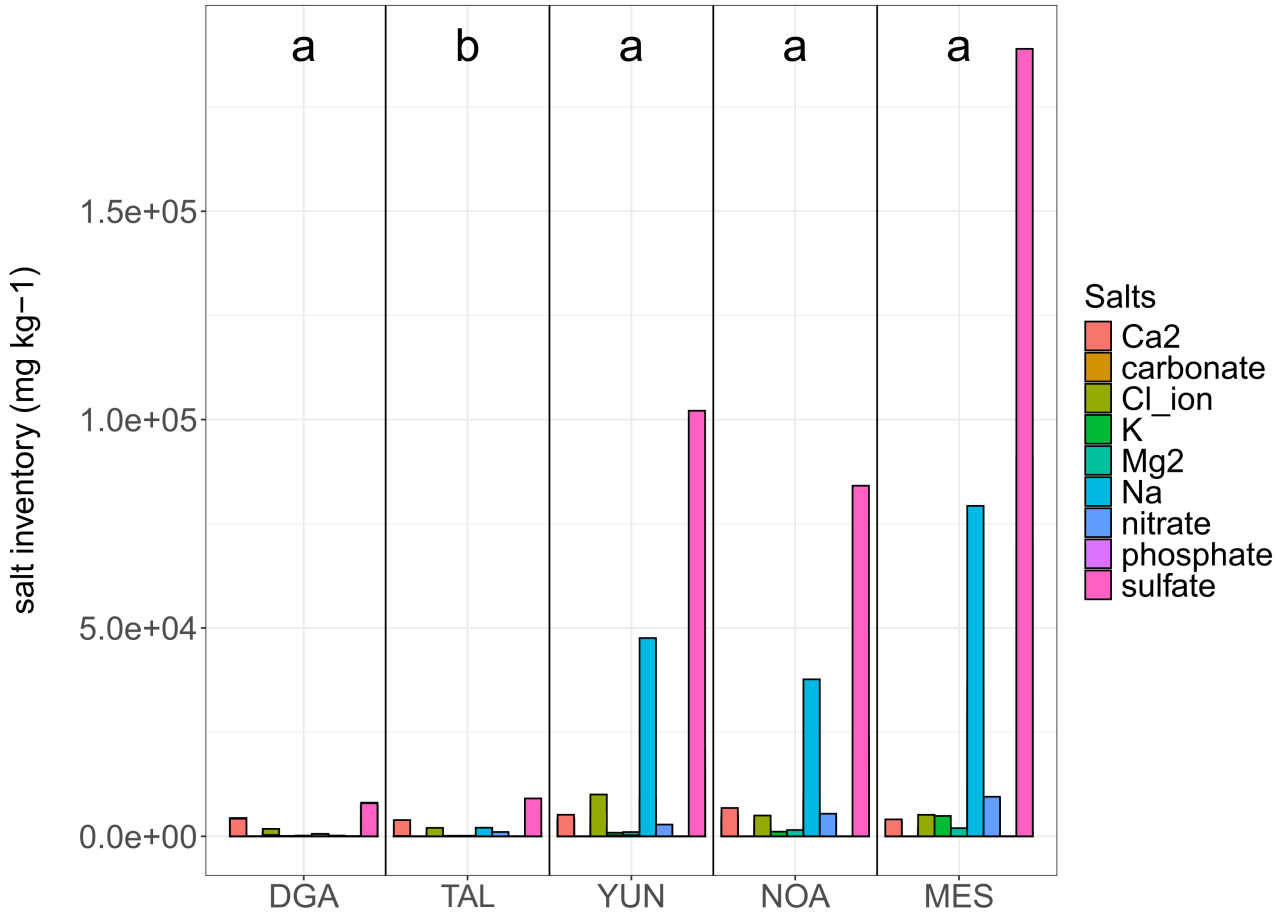

Fig. S16. Salt concentrations per plot along the south to north transect (mg kg^{-1}). Each bar represents the sum of the different depths per salt species. The concentration of the Na+ was low at all plots in the upper soil depths along the gradient, but did increase towards the northerly plots in the deeper soil depths. We did not find a statistical difference, when testing for plot alone ($p>0.05$), but there was a significant difference at the sites of YUN, NOA, and MES when testing for the depths at these plots. Largest inventories of Na+ were found in MES with 58.81 $\pm$ 27.81 mg kg-1 and smallest at TAL with 1.40 $\pm$ 0.86 mg kg-1. Highest concentrations of carbonates per sample were found at DGA in 30-40 cm with 2.4 % $\pm$ 1.9 and lowest at YUN in 50-60 cm of depth with 0.01% ± 0.01%. In tendency, carbonates decreased with soil depth but this was not significant ($p > 0.05$). K+ occurred very similar to Na+ with increasing concentrations towards the north and into the depth at the three northerly located plots, although much smaller. Largest inventory was found at MES with 4.30 ± 1.61 mg kg-1 and smallest at plot TAL with 1.21 $\pm$ 0.48 mg kg-1. Inventories of Ca2+ did not differ much between the plots and ranged between 14.2 $\pm$ 1.05 mg kg-1 at plot MES and 20.45 ± 1.41 mg kg-1 at plot NOA. The only exception is plot TAL where roughly tenfold less Ca2+ was detected (1.76 ± 1.68 mg kg-1). Concerning the occurrence of Ca2+ in the depth increments along the transect, there was no statistical difference, neither when considering the depths at the individual plots. Mg2+ occurred randomly along the gradient and the depths. While there were only minor concentrations at plot DGA (0.38 $\pm$ 0.04 mg kg-1) and TAL (0.08 $\pm$ 0.06 mg kg-1), we found Mg2+ in high concentration at NOA and MES in 5-10 cm of soil depth. Chlorides (Cl-) increased from south to north ($p<0.001$), although concentrations were comparable at DGA, YUN, and MES. In the depth increments, the upper soil depths of 0-1 cm, 5-10 cm, and 10-20 cm were similar and increased to the deeper depths ($p< 0.001$). Chloride concentrations were highest at plot YUN (1896 ± 2038 mg kg-1) and lowest at plot TAL (267.7 $\pm$ 198.5 mg kg-1). Nitrates were significantly more abundant in the north with a strong increase between TAL and YUN ($p< 0.001$) as well as at soil depths below 30 cm ($p<0.001$). Nitrate concentrations ranged between 64.0 $\pm$ 13.7 mg kg-1 at DGA and 2677 $\pm$ 3344 mg kg-1 at plot MES, thus were roughly 40times larger in the north depending on the depth. Phosphate concentrations were quite similar along the gradient with TAL being statistically different ($p<0.01$) and were largest at plot DGA (3.5 $\pm$ 4.48 mg kg-1) and smallest at plot YUN (1.17 $\pm$ 1.09 mg kg-1). Phosphates occurred at all plots at the surface, but showed random occurrence in deeper soil depths ($p>0.05$). Sulfate concentrations increased 5fold from south to north ($p<0.001$) and from the surface to the deeper soil depths in the soil profiles of the hyper-arid plots ($p<0.01$). Concentrations ranged between 933.9 $\pm$ 1082 mg kg-1 at plot TAL and 32833 $\pm$ 36034 mg kg-1 at plot MES.

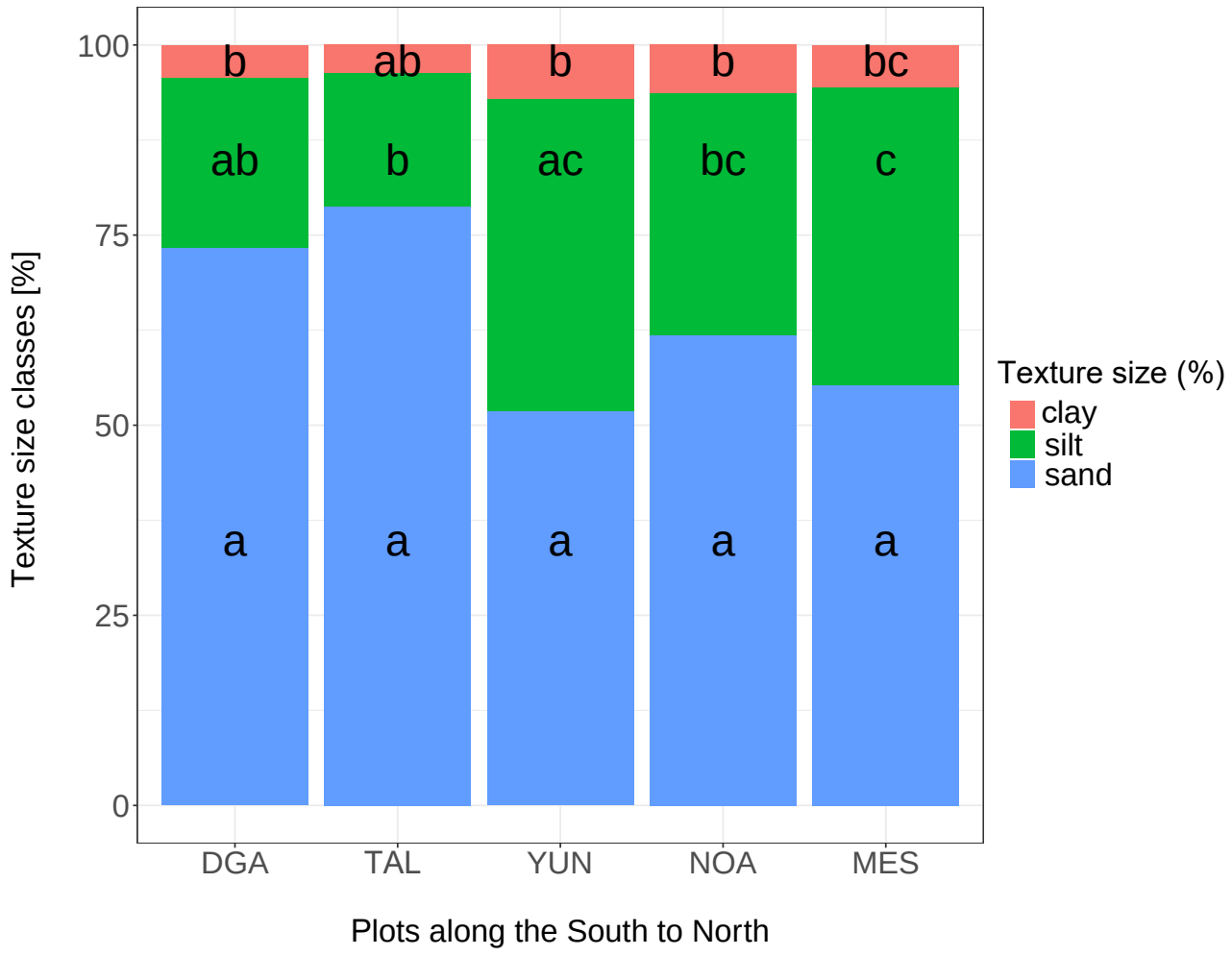

Fig. S17. The relative distribution of the three texture size classes: clay, silt, and sand. The fraction of clay was in tendency higher at the northerly plots YUN, NOA, and MES, but did not differ statistically between the plots as the variability was relatively high between the depths. Clay content ranged between 3.73 $\pm$ 1.07 vol% at plot TAL and 6.17 $\pm$ 3.89 vol% at plot YUN. At the plots YUN and NOA, clay content was highest on the surface reaching more than 15 vol%. Silt did not differ between the depths, but its volumetric concentration increased significantly from south to north ($p<0.05$) with 41.7 $\pm$ 2.3 vol% at plot YUN and 17.58 $\pm$ 2.55 vol% at plot TAL. As a consequence of the occurrence of clay and silt, the fraction of sand did not differ between the plots and ranged between 78.7 $\pm$ 3.4 vol% at plot TAL and 52.51 $\pm$ 2.64 vol% at plot YUN.

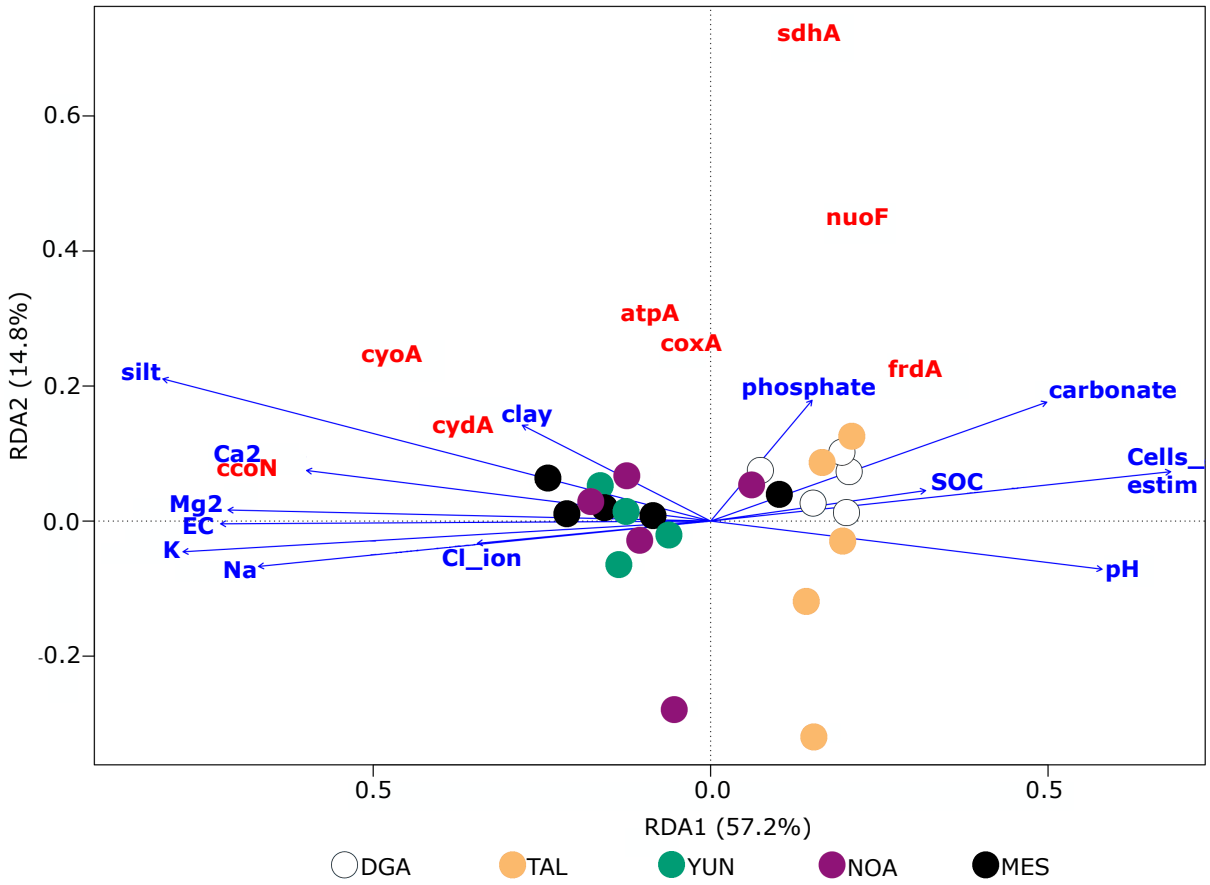

Fig. S18. Genes involved in aerobic respiration shown in a triplot of RDA (R^2=0.54, p=0.003** using a permutation test for RDA with 999 number of permutations). The RDA correlates matrices of the Hellinger transformed functional gene abundances with the normalized soil data (performed in the vegan package version 2.6.4 (Oksanen et al. 2022) in Rstudio. In general, we observed a site specifity of the functional genes associated with aerobic respiration either towards the arid plots DGA and TAL, e.g., *frdA* and *nuoF* or towards the hyperarid plots YUN, NOA, and MES. The gene *ccoN* correlates well with the highly abundant salt fractions such as Ca^{2+} and Mg^{2+}, while the abundance of *cydA* and *cyoA* is impacted by the occurrence of clay.

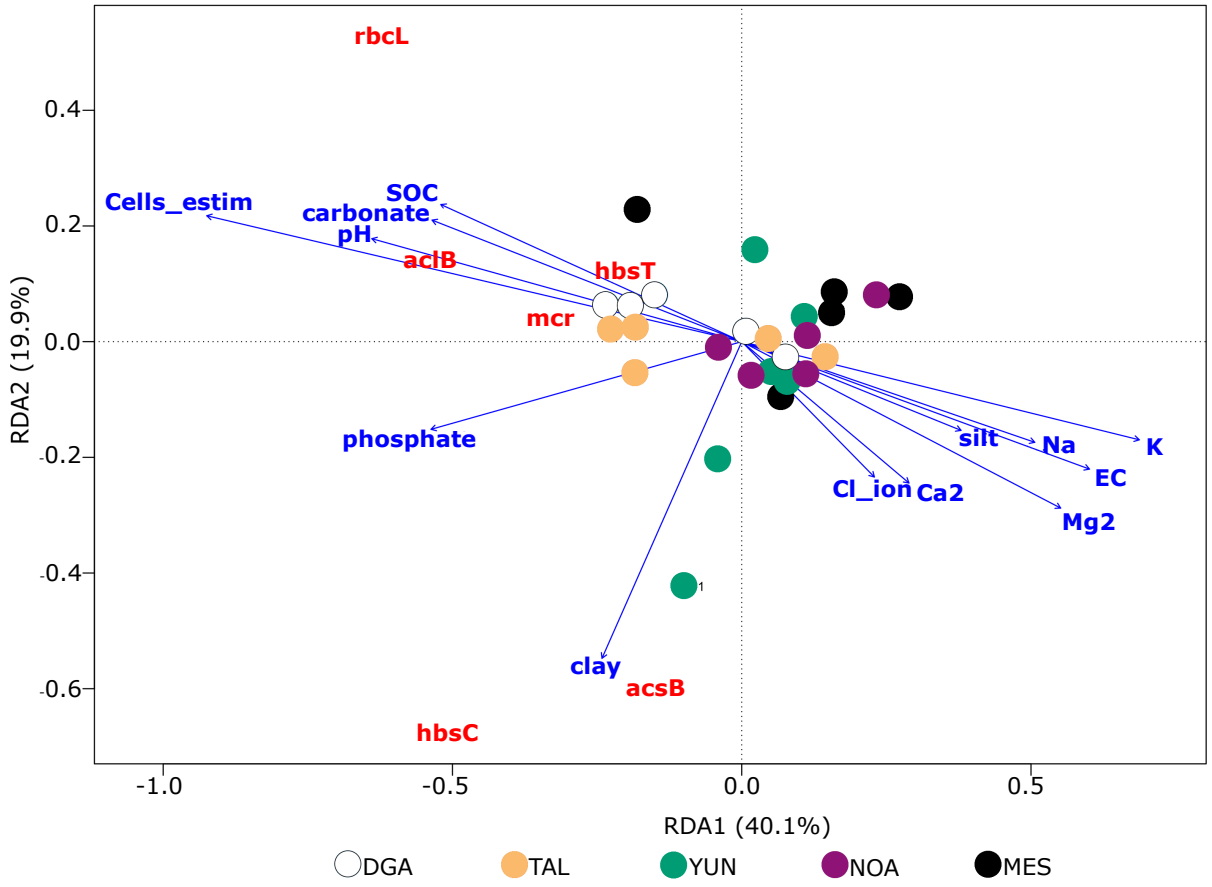

Fig. S19. Genes involved in carbon fixation shown in a triplot of RDA (R^2=0.42, p=0.008**). Carbon fixings genes such as *aclB*, *mcr*, *hbsT*, and *rbcL* were more abundant in the south and were strongly affected by the SOC, carbonates, pH, and the estimated cell counts, while *hbsC* and *acsB* were affected by clay.

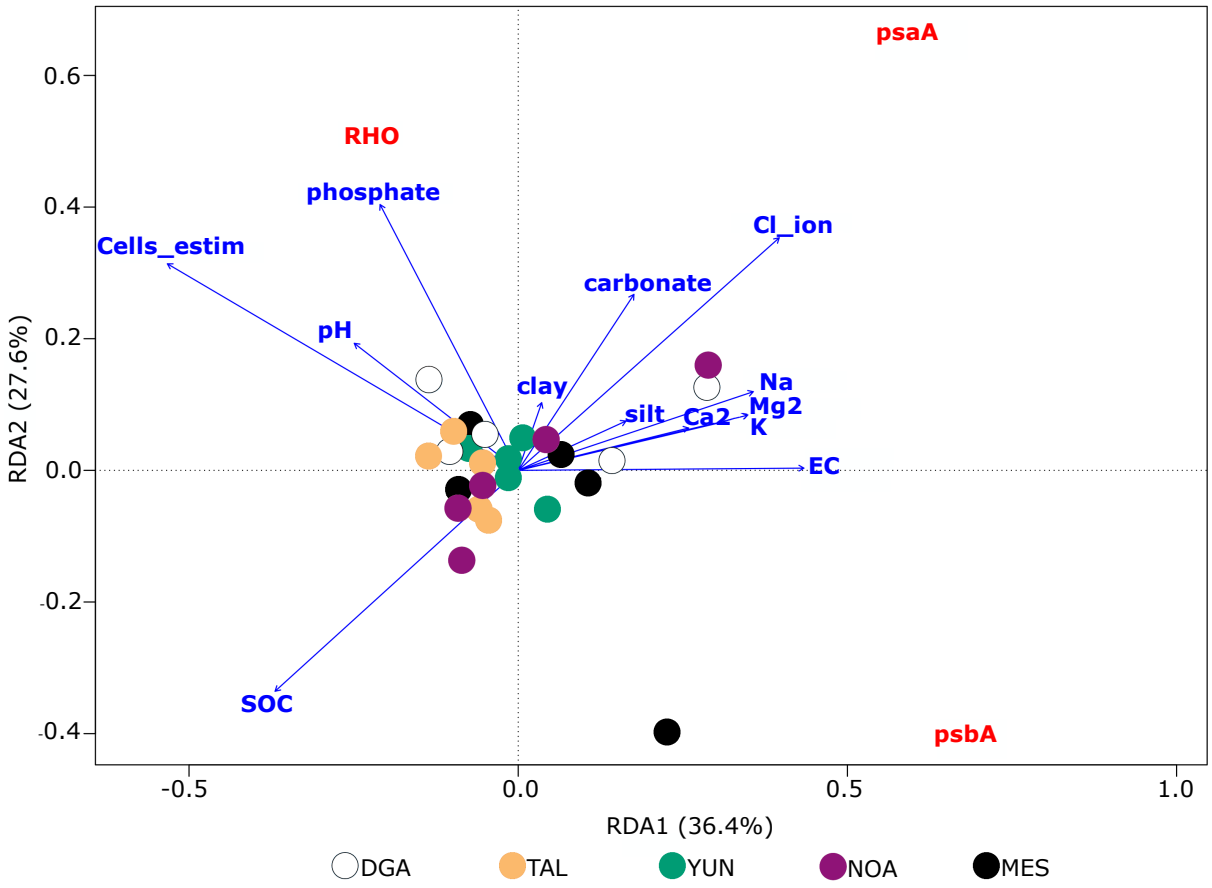

Fig. S20. Genes involved in phototrophy shown in a triplot of RDA (R^2=0.45, p=0.005**). There was no clear pattern of occurrence along the gradient of none of the genes in the statistical comparison between the two ($p > 0.05$) plot clusters, as visible also by the overlap of the samples in the RDA, however, *RHO* correlated slightly positive with the occurrence of phosphates (R^2=0.39), *psaA* negatively with SOC (R^2=-0.48), while *psbA* was uncorrelated with any soil parameter.

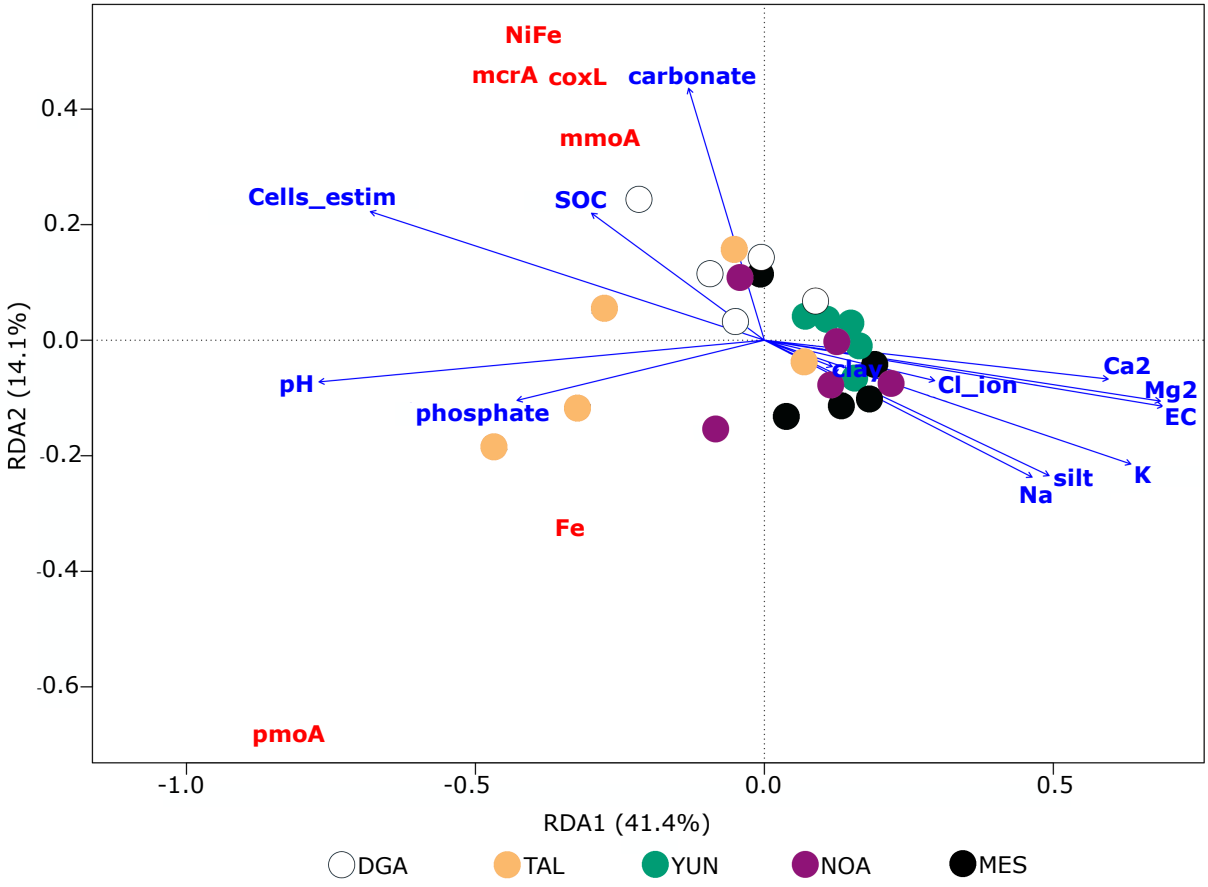

Fig. S21. Genes involved in trace gas metabolism shown in a triplot of RDA (R^2=0.28, p=0.04*). We deleted the data for K^+, the *FeFe* and *CooS* gene abundances, as they did not correlate with any other parameter and made the model therefore insignificant. Genes involved in trace gas metabolism clearly cluster in the southerly located plots, thus, have a higher abundance here as compared to the hyperarid plots in the north. *NiFe*, *mcrA*, *coxL* correlated with SOC, carbonates, and estimated cell counts, *Fe* with phosphates (R^2=0.53).

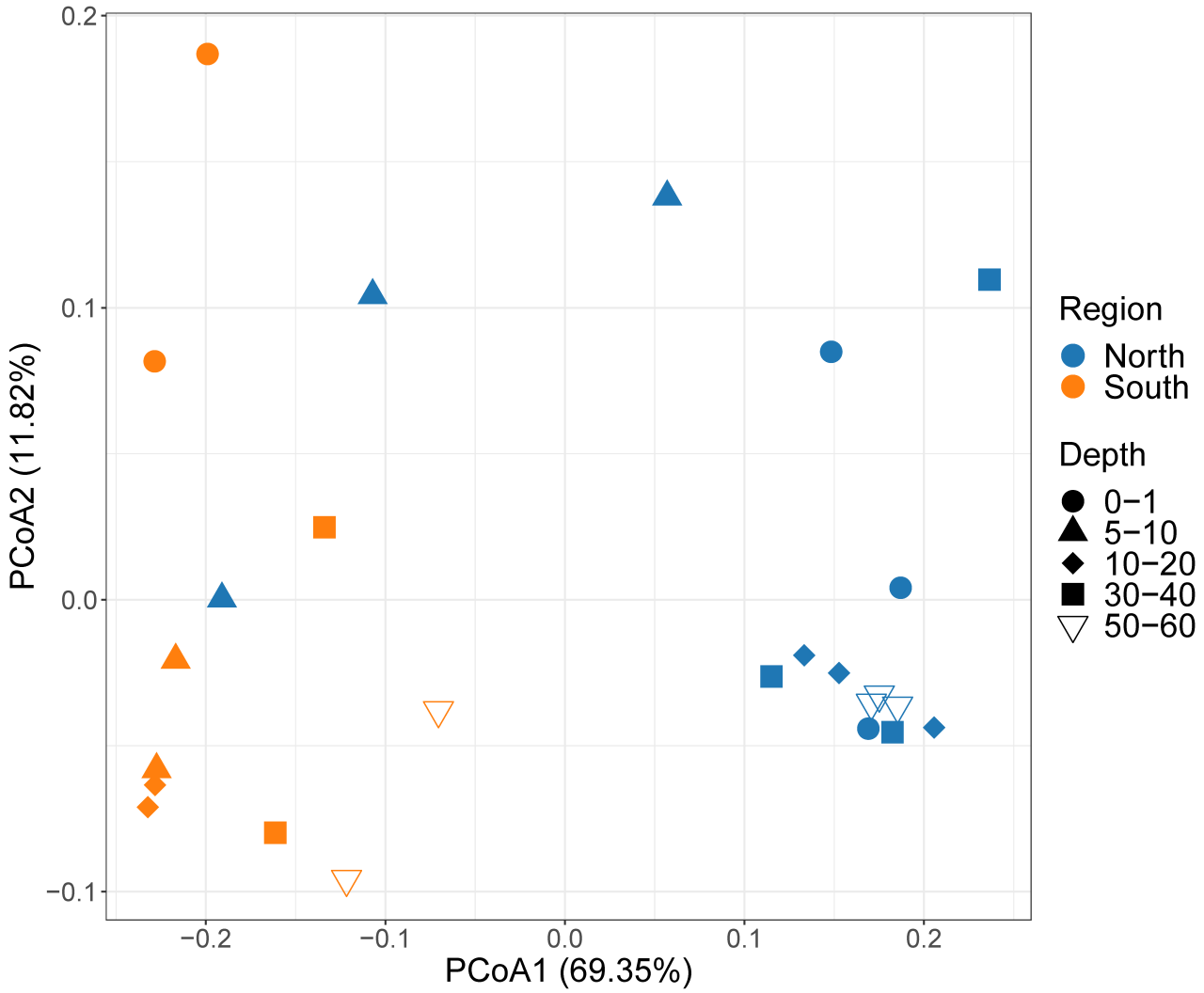

Fig. S22. PCoA plot of microbial community function based solely on the estimated functional gene abundance at each plot/depth (iDNA only). PCoA1 (69.35%) and PCoA2 (11.82%) are the first two principal coordinates, together explaining 81.17% of the total variance in the data. Regions (north and south) are indicated by color (blue and orange) while depths in cm below the surface are indicated by shape.

References

1. Red Cedeus. By GeoNode (2018) Version 2.0.1. Retrieved from http://datos.cedeus.cl/layers/ (2021).
2. DGAC. Dirección Meteorológica de Chile. Av. Portales 3450, Estación Central, Santiago de Chile, Chile. (2021).
3. J Arriagada, Elaboración de una base digital del clima comunal de Chile: Línea base (1980–2010) y proyección al año 2050. Estudio encargado por el Ministerio del Medio Ambiente Departamento de Cambio Climático. Informe final. With the help of K.-P. Muck. Ministerio del Medio Ambiente Departamento de Cambio Climático. (2016).
4. (CR)2, Valores mensuales de precipitación medidos desde estaciones nacionales, desde 1900 al presente año. CR2_pramon_2018.zip. Santiago de Chile, Chile: Center for Climate and Resilience Research. (2018).
5. D Boy, et al., Gradient Studies Reveal the True Drivers of Extreme Life in the Atacama Desert. *J. Geophys. Res. Biogeosciences* **127**, e2021JG006714 (2022) _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JG006714.
6. I Wrb, World reference base for soil resources 2014. *Update* **201** (2015).
7. D Schulze-Makuch, et al., Transitory microbial habitat in the hyperarid Atacama Desert. *Proc. Natl. Acad. Sci.* **115**, 2670–2675 (2018) Publisher: Proceedings of the National Academy of Sciences.
8. P Menzel, KL Ng, A Krogh, Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016) Number: 1 Publisher: Nature Publishing Group.
9. AM Bolger, M Lohse, B Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
10. S Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data [Online] (2010).
11. A Oren, GM Garrity, Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **71**, 005056 (2021) Publisher: Microbiology Society,.
12. NM Davis, DM Proctor, SP Holmes, DA Relman, BJ Callahan, Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
13. M Ortiz, et al., Multiple energy sources and metabolic strategies sustain microbial diversity in Antarctic desert soils. *Proc. Natl. Acad. Sci.* **118**, e2025322118 (2021) Publisher: Proceedings of the National Academy of Sciences.
14. PM Leung, C Greening, Greening lab metabolic marker gene databases (2020).
15. B Buchfink, K Reuter, HG Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021) Publisher: Nature Publishing Group.
16. AE Darling, et al., PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**, e243 (2014) Publisher: PeerJ Inc.
17. D Harris, WR Horwáth, C van Kessel, Acid fumigation of soils to remove carbonates prior to total organic carbon or CARBON-13 isotopic analysis. *Soil Sci. Soc. Am. J.* **65**, 1853–1856 (2001) _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2001.1853.