Title page:

Short communication for *Soil Biology & Biochemistry*

Title: High-throughput sequencing shows inconsistent results with a microscope-based analysis of the soil prokaryotic community

Authors: Masayuki Ushio[1] *, Kobayashi Makoto[2, 3, 4], Jonatan Klaminder[2], Hiroyuki Takasu[1, 5], Shin-ichi Nakano[1]

Affiliations: [1]Center for Ecological Research, Kyoto University, 2-509-3 Hirano, Otsu, Shiga 520-2113, Japan, [2]Climate Impact Research Centre, Umeå University, 981-07, Abisko, Sweden, [3]Soil Ecology Research Group, Graduate School of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Yokohama, Kanagawa 240-8501, Japan, [4]Nakagawa Experimental Forest, Field Science Center for Northern Biosphere, Hokkaido University, Otoineppu 098-2501, Japan, [5]Atmosphere and Ocean Research Institute (AORI), The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan

*Correspondence to: ong8181@gmail.com

Tel: +81-77-549-8215, Fax: +81-77-549-8201

Keywords: Archaea; Bacteria; CARD-FISH; High-throughput sequencing; Soil prokaryotic community; Tundra ecosystem

Abstract:

In the present study, we perform the first direct analysis on how the composition of the prokaryotic soil community differs depending on whether high-throughput sequencing or fluorescent *in situ* hybridization (FISH) coupled with catalyzed reporter deposition (CARD) is used. Soil samples were collected along short (< 3 m) tundra vegetation gradients from Northern Sweden. Relative abundances of *Acidobacteria* and *Bacteroidetes* estimated by the high-throughput sequencing were higher than those estimated by CARD-FISH, while relative abundances of *Archaea* and *α-Proteobacteria* estimated by high-throughput sequencing were lower than those estimated by CARD-FISH. The results indicated that the high-throughput sequencing overestimates/underestimates the relative abundance of some microbial taxa if we assume that CARD-FISH can provide potentially more quantitative data. Great caution should be taken when interpreting data generated by molecular technologies (both of high-throughput sequencing and CARD-FISH), and supports by multiple approaches are necessary to make a robust conclusion.

Text

When investigating the composition of microbial communities, high-throughput sequencing technologies are becoming the most commonly applied method in microbial ecology. The rationale for this is a cost-effective means of identifying thousands of microbial phylotypes that are present in samples (Lauber et al., 2009; Sogin et al., 2006). Without these technologies, it is almost impossible to reveal the very high diversity of soil microbial communities, and thus, they currently constitute the most important tools for our understanding about soil microbes.

However, many experimental steps in the high-throughput sequencing analysis could potentially produce biases/artifacts that significantly influence biological interpretations of the dataset (Engelbrektson et al., 2010; Gomez-Alvarez et al., 2009; Zhou et al., 2011). For example, Zhou et al. (2011) examined the quantitative capacity of high-throughput sequencing of 16S rRNA gene amplicons by adding a known quantity of extracted DNA from a cultured microbial strain. They found that the percentages of the strain OTU varied substantially among different samples, from 0.00% to 5.34% (theoretically it should have been 0.1%), and thus they questioned the quantitative capacity of high-throughput sequencing. However, studies that compared results of other potentially more quantitative approaches (e.g., microscope-based investigations) in complex soil matrixes have been very limited so far.

To compare the results of high-throughput sequencing and a potentially more quantitative microscope-based analysis, we analyzed soil samples with IonPGM high-throughput sequencer (Rothberg et al., 2011; Ion Torrent by Life Technologies, Guilford, CT, USA) and fluorescent *in situ* hybridization (FISH) coupled with catalyzed reporter deposition (CARD) (Eickhorst and Tippkötter, 2008; Pernthaler et al., 2002), targeting *Bacteria* and *Archaea*. In the present study, we focused on the methodological comparison because the results of CARD-FISH analysis of the soil prokaryotic community were already reported in Ushio et al. (2013). Soil samples were collected from the north-facing slope of Mt. Suorooaivi in Abisko, Northern Sweden. In this area, patterned ground also referred to as non-sorted circles (Fig. 1a; Klaus et al., 2013) occurs frequently because of the soil-frost process. Within these features, a dramatic vegetation change from lichen-dominated plant communities to dense shrub communities occurs over distances of less than 3 m (Fig. 1b; Makoto and Klaminder, 2012).

Soil samples were collected along the vegetation gradients (6 individual circles × 7 locations = 42 samples; Fig. 1b) and taken back to the laboratory immediately. After the sorting, CARD-FISH was conducted as described previously (Ushio et al., 2013; Supplementary Methods), and the soil subsamples were stored at −20 °C until further DNA analyses. For the CARD-FISH analysis, nine probes (*Eubacteria*, *Archaea*, *α-Proteobacteria*, *β-Proteobacteria*, *γ-Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *CFB*, and Nonsense probes), which target potentially dominant soil microbial groups (e.g., Lauber et al 2009), were applied to quantify the abundance of microbial groups in the soil samples (see Table 1 in Ushio et al. 2013). By following Bates et al. (2011), soil DNA extraction, polymerase chain reaction (PCR), and purification were conducted (see detail in Supplementary Methods). The multiple PCR products were pooled, and the single composite 6-bp-barcoded sample was sent for sequencing at Life Technologies Japan (Tokyo, Japan) on IonPGM. After sequencing, the raw sequence data were processed using QIIME (Caporaso et al., 2010). Quality filtering, chimera identifications, and operational taxonomic unit (OTU) clustering (≥ 97% similarity) were performed using the USEARCH option (Edgar, 2010; Edgar et al., 2011) in QIIME. After filtering and clustering, taxonomies were assigned to the OTUs. Detailed experimental protocols and data handling procedures are described in the Supplementary Information. The sequences data is archived in DNA Data Bank of Japan

(DDBJ) Sequence Read Archives (DRA), and the accession number is DRA001218 for the submission data.

After processing, 620,345 sequences from a total of 42 soil samples were obtained (14,770 ± 4564 [± standard deviation] sequences per sample; Table S1), and approximately 2,200 OTUs were identified for each sample (Table S1 and Fig. S1). According to the high-throughput sequencing analysis, *Acidobacteria* was the most dominant microbial group, followed by *Bacteroidetes*, *α-Proteobacteria*, and *γ-Proteobacteria* (Fig. 1c). The relative abundance of *Archaea* was less than 1% (Fig. 1c). Conversely, the relative abundance of *Archaea* estimated by CARD-FISH exceeded 30% in the soil samples (Fig. 1d), which is similar in range to that previously reported for farmland and paddy soils by CARD-FISH (Eickhorst and Tippkötter, 2008). Relative abundances of *Acidobacteria* and *Bacteroidetes* estimated by the high-throughput sequencing were higher than those estimated by CARD-FISH, while relative abundances of *Archaea* and *α-Proteobacteria* estimated by the high-throughput sequencing were lower than those estimated by CARD-FISH (Fig. 2). Data handling procedures often have significant influences on the results of high-throughput sequencing analysis (Edgar, 2013), but qualitatively the same result was obtained even if UPARSE, a recently published new pipeline (Edgar, 2013), was used (Fig. S2). These results suggested that the high-throughput sequencing analysis could overestimates/underestimates the relative abundance of some microbial taxa.

To investigate the overall differences in community composition among the samples, principle coordinate analysis was performed using unweighted UniFrac distance (Lozupone and Knight, 2005). The result showed that the prokaryotic community compositions of locations 0 and 1 were significantly different from those of locations 2–5 and H, and that of location H was different from those of locations 3–5 (Fig. S3a). The result is inconsistent with the CARD-FISH result that community composition was not different among locations

(Ushio et al., 2013). This inconsistency is probably because our CARD-FISH analysis investigated microbial community at phylum or sub-phylum level (i.e., phylum or sub-phylum specific probes were used). If we had conducted the CARD-FISH analysis at genus or species level, we could have detected significant differences in community composition among locations. Soil pH was a good predictor of the prokaryotic community composition estimated by the high-throughput sequencing (Fig. S3b), which is in accordance with previous studies (e.g., Lauber et al., 2009).

In conclusion, our study showed for the first time that high-throughput sequencing could not reliably estimate the relative abundance of soil microbial taxa based on 16S rRNA gene amplicon sequencing, if we assume that CARD-FISH can potentially provide more quantitative data. This assumption may be reasonable considering its technical basis (Pernthaler et al., 2002), though we should be careful because CARD-FISH also includes experimental steps that might produce biases (Amann and Fuchs, 2008). The fixation procedure of samples, probe specificity, and probe hybridization conditions are potential factors producing biases for CARD-FISH (Amann and Fuchs, 2008), while DNA extraction methods, primer sets, PCR conditions and data processing procedures are factors contributing to biases for high-throughput sequencing (e.g., Engelbrektson et al., 2010). Direct capturing and genomic analysis of the FISH-positive cells (Pernthaler et al., 2008), which is not a common technique yet, would provide strong support for the quantitative capacity of CARD-FISH. We would like to emphasize that great caution should be taken when interpreting data generated by molecular techniques (i.e., both of CARD-FISH and high-throughput sequencing), and that supports by multiple approaches are necessary to make a robust conclusion about microbial communities in a field condition. The finding seems crucial to be aware of, considering that high-throughput sequencing data currently drives a large part of the scientific perception about soil microbial communities.

Acknowledgments

We thank A.J. Nagano, M.N. Honjo, and H. Kudoh for their support for the DNA sequencing experiments. We also thank Abisko Scientific Research Station for allowing us to use the research facility. M.U. and K.M. are financially supported by Research Fellowships of Japan Society for the Promotion of Science (JSPS; 23-586 and 24-3240, respectively).

References

Amann, R., Fuchs, B.M., 2008. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nature Reviews Microbiology 6, 339–348.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., Fierer, N., 2011. Examining the global distribution of dominant archaeal populations in soil. ISME Journal 5, 908–917.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, 335–336.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10, 996–998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

Eickhorst, T., Tippkötter, R., 2008. Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). Soil Biology and Biochemistry 40, 1883–1891.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME Journal 4, 642–647.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME Journal 3, 1314–1317.

Klaus, M., Becher, M., Klaminder, J., 2013. Cryogenic Soil Activity along Bioclimatic Gradients in Northern Sweden: Insights from Eight Different Proxies. Permafrost and Periglacial Processes 24, 210–223.

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and Environmental Microbiology 75, 5111–5120.

Lozupone, C., Knight, R., 2005. UniFrac: A new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71, 8228–8235.

Makoto, K., Klaminder, J., 2012. The influence of non-sorted circles on species diversity of vascular plants, bryophytes and lichens in Sub-Arctic Tundra. Polar Biology 35, 1659–1667.

Pernthaler, A., Dekas, A.E., Brown, C.T., Goffredi, S.K., Embaye, T., Orphan, V.J., 2008. Diverse syntrophic partnerships from deep-sea methane vents revealed by direct cell capture and metagenomics. Proceedings of the National Academy of Sciences of the United States of America 105, 7052–7057.

Pernthaler, A., Pernthaler, J., Amann, R., 2002. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. Applied and Environmental Microbiology 68, 3094–3101.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–352.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proceedings of the National Academy of Sciences 103, 12115–12120.

Ushio, M., Makoto, K., Klaminder, J., Nakano, S.I., 2013. CARD-FISH analysis of prokaryotic community composition and abundance along small-scale vegetation gradients in a dry arctic tundra ecosystem. Soil Biology and Biochemistry 64, 147–154.

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.H., Tu, Q., Xie, J., Van Nostrand, J.D., He, Z., Yang, Y., 2011. Reproducibility and quantitation of amplicon sequencing-based detection. ISME Journal 5, 1303–1313.

Figure legends

Fig. 1. (a) Images of a non-sorted circle (NSC), where a 1-m measure is placed between the center of the NSC and the edge of the inner domain. (b) Soil sampling design (left panel). Here, six sampling points were assigned between the center of NSC and the edge of the outer domain. Aboveground vegetation is absent or relatively poor at locations 0, 1, and 2, while plants are densely colonized at locations 3, 4, and 5. The humus layer, which is present only at locations 3, 4, and 5, was also collected. This sampling design is identical to that of Makoto and Klaminder (2012). Plant community composition changed dramatically along this transect because of the soil-frost process (cryoturbation) in the system (right panel). (c) Composition of the prokaryotic community along the NSC vegetation gradients estimated by high-throughput sequencing, or (d) CARD-FISH. Numbers on the *x*-axis indicate distance from the center of the NSC, and "H" indicates humus layer samples. Each bar represents the mean value of the relative abundance of each microbial group. "Others" includes prokaryotic microbes other than the listed microbial groups, and unidentified sequence reads. (a), (b), and (d) are reproduced and modified from Ushio et al. (2013).

Fig. 2. Direct comparison of relative abundance values estimated by high-throughput sequencing and CARD-FISH. The *x*-axis indicates relative abundance estimated by CARD-FISH analysis, and the *y*-axis indicates that estimated by high-throughput sequencing analysis. The solid line indicates 1:1 in the figure. "*α*", "*β*", and "*γ*" indicate *α-Proteobacteria*, *β-Proteobacteria*, and *γ-Proteobacteria*, respectively. Bars indicate standard deviation.

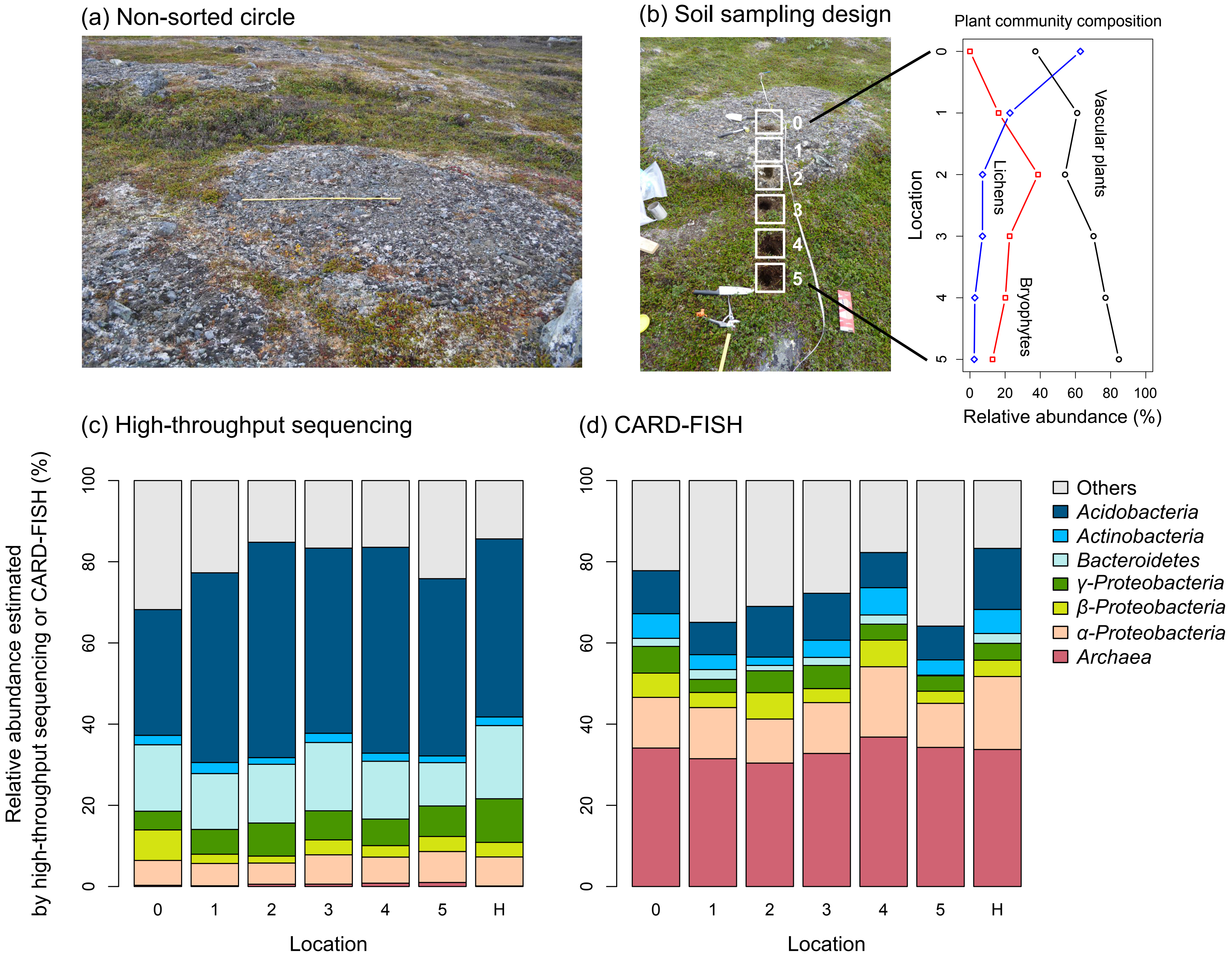

Fig. 1 Ushio et al.

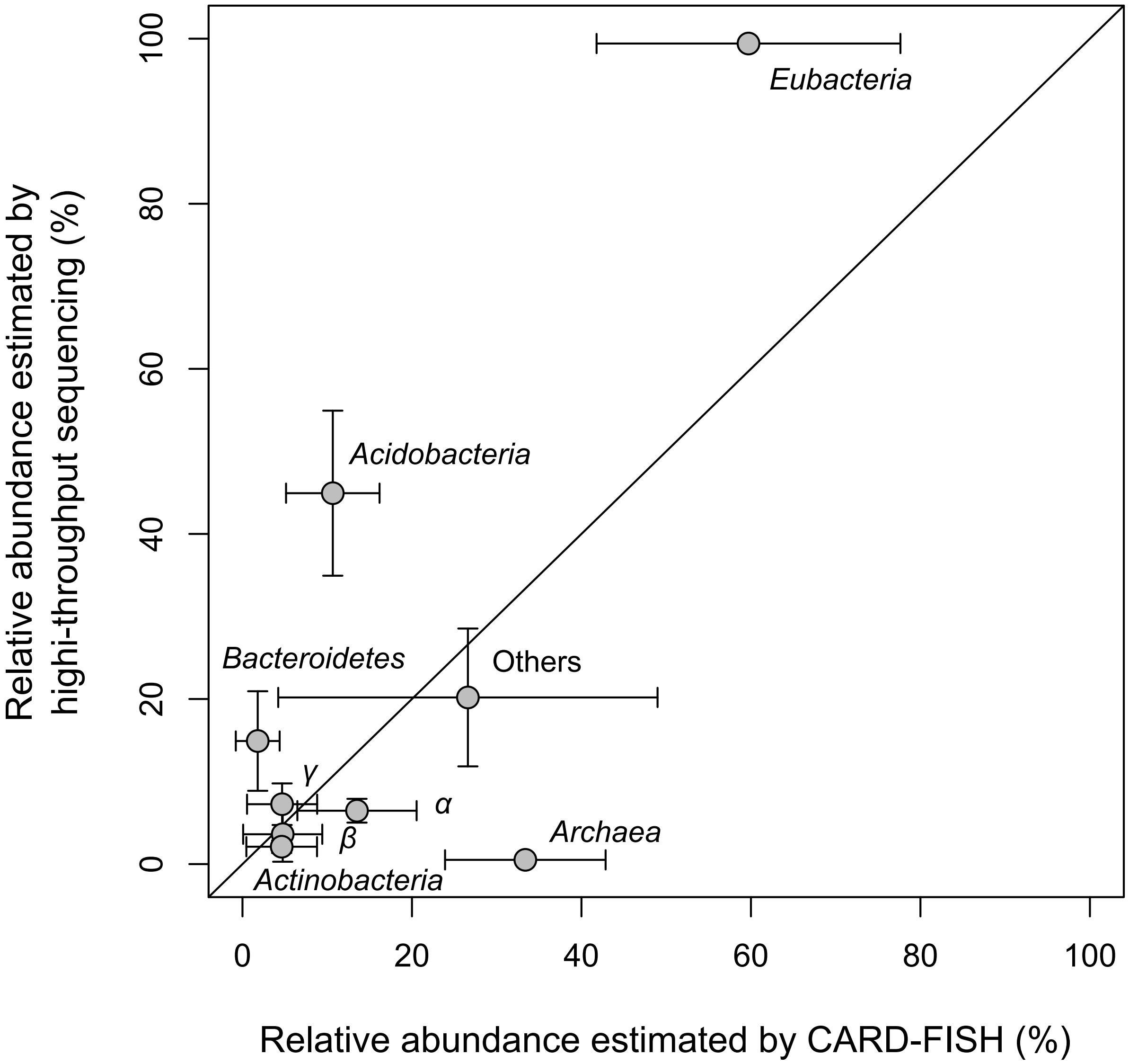

Fig. 2 Ushio et al.

Supplementary Information

Soil Biology & Biochemistry

Title: High-throughput sequencing shows inconsistent results with a microscope-based analysis of the soil prokaryotic community

Authors: Ushio, M., Makoto, K., Klaminder, J., Takasu, H., Nakano, S-I.

Contents:

Supplementary Methods

Descriptions of the study site, soil sampling, coupled with catalyzed reporter deposition fluorescent *in situ* hybridization (CARD-FISH), sample processing, DNA extraction, polymerase chain reaction (PCR), sequence data handling procedures, and data accessibility.

Supplementary Table S1

Number of sequences and OTUs per sample obtained by the high-throughput sequencing.

Supplementary Figure S1

Rarefaction results for each sample and each location in a non-sorted circle (NSC).

Supplementary Figure S2

Results of UPARSE pipeline analysis.

Supplementary Figure S3

Principle coordinate analysis (PCoA) using unweighted UniFrac distance.

Supplementary Methods

Study site description

Field research was conducted from September to October in 2011, on the north-facing slope of Mt. Suorooaivi (1193 m a.s.l., 68°16′ N, 19°06′ E; Fig. 1a, b), located approximately 20 km south of Abisko, Northern Sweden (68°18′ N, 19°10′ E). The research site is situated at an elevation of 700–750 m. Between 1981 and 2010, the mean annual temperature and precipitation in Abisko were –0.1 °C and 335 mm, respectively.

Within the arctic region, sharp vegetation gradients are present over small spatial scales within patterned ground systems. One example of these sharp vegetation gradients can be found within patterned ground systems referred to as non-sorted circles (NSCs), where soil frost processes (cryoturbation) generate sparsely vegetated circle-like features of approximately 1–3 m surrounded by densely vegetated shrub communities (see Makoto and Klaminder 2012, Ushio et al 2013, Klaue et al. 2013 and references therein for more detailed site description). Consequently, the sharp vegetation gradient generated within NSCs, consisting of lichen-dominated plant communities in the center to shrub-dominated communities in the outer domain (Fig. 1a, b; Makoto and Klaminder 2012), are representative of vegetations of a large part of the dry arctic landscape. The species diversity of the vascular plants and the density of the bryophyte community increased with increasing distance from the center of the NSCs (Makoto and Klaminder 2012). Therefore, the NSC system provides a good opportunity to study the extent to which microbial communities co-vary with the aboveground plant community over a considerable proportion of arctic tundra ecosystems.

Soil sampling

In each NSC, samples from seven locations (i.e., location 0–5 and H) were collected in late September 2011. At locations 0–5, mineral soil samples (~ 10 cm depth) were collected. At location H, humus layer samples, which occur at location 3–5, were collected and combined as one humus sample. Replicate soil samples were taken from six individual NSCs, and therefore, we obtained 42 soil samples in total (7 locations × 6 individual NSCs). The soil samples were immediately taken back to the laboratory, sieved to remove stones and plant roots, and homogenized thoroughly for the subsequent analyses. Soil pH (soil:water = 1:10), carbon, and nitrogen content as well as aboveground vegetation were previously determined. Briefly, ranges of physico-chemical properties of mineral soils in our study site are as follows: soil pH 4.5–6.2, organic carbon content 0.3–24.1%, total soil nitrogen 0.02–1.12%, and soil water content 9.5–54.3%.

For the collected soil samples, coupled with catalyzed reporter deposition fluorescent *in situ* hybridization (CARD-FISH) and high-throughput sequencing analyses were conducted. As described in subsequent sections, the primer set used in the high-throughput sequencing analysis can amplify 16S rRNA genes from *Bacteria* and *Archaea* (Bates et al 2011), and the probes used in the CARD-FISH analysis can also detect *Bacteria* and *Archaea* (Ushio et al. 2013). Therefore, both methods used in our study targeted the same microbial groups, which allowed us to compare the results of these two methods.

CARD-FISH

To investigate prokaryotic community composition along the NSC vegetation gradients, the CARD-FISH method was applied to each sample. We generally followed the experimental protocol described in Eickhorst and Tippkötter (2008) with several modifications. Approximately 0.5 cm^3 of fresh bulk soil from each soil sample was weighed and transferred to 2-mL plastic micro-tubes. The samples were fixed with 4% (w/v) freshly prepared particle-free paraformaldehyde solution, and the suspension was stored at 4 °C overnight (~16 h). The fixed samples were washed twice with 1× PBS, centrifuged at 10,000 g for 5 min after each washing, and stored in PBS/ethanol (1:1) at −20 °C for further processing. Then, 100 µL of the fixed sample was diluted with 900 µL of PBS/ethanol and dispersed by ultrasonic dispersion at minimum power for 20 s (10%; UR-21P; TOMY, Tokyo, Japan).

A volume of 20 µL of the dispersed sample was diluted in 10 mL of sterilized water, and the suspension was filtered on a polycarbonate filter (0.2 µm pores, φ25 mm; Millipore, Billerica, MA USA) placed on a nitrocellulose filter (0.45 µm pores, φ25 mm; Millipore), which were mounted in a glass holder for the filtration. After the filtration, the filters were dipped in 0.2% low melting point agarose (Sigma, St. Louis, MO USA) and dried in an incubator at 46 °C. The agarose-embedded filters were then incubated with a lysozyme solution (10 mg lysozyme, 100 µL 0.5 M EDTA [pH 8.0], 100 µL 1 M Tris-HCl [pH 8.0], 800 µL of sterilized water), which were placed in a sealed petri dish at 37 °C for 1 h. After washing with MQ water, the filters were incubated with methanol containing 0.15% H_2O_2 for 30 min at room temperature to inactivate endogenous peroxidase activity. The filters were washed with MQ water and dehydrated by dipping them in 98% ethanol and subsequently air drying. After this step, filters were stored at −20 °C until further processing.

For the CARD-FISH analysis, nine probes (*Eubacteria*, *Archaea*, *α-Proteobacteria*, *β-Proteobacteria*, *γ-Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *CFB*, and Nonsense probes), which target potentially dominant soil microbial groups (e.g., Lauber et al 2009), were applied to quantify the abundance of various microbial groups in the soil samples (see Table 1 in Ushio et al. 2013). For the *in situ* hybridization, the filter was cut into small pieces and incubated with 300 μL of hybridization buffer (900 mM NaCl, 20 mM Tris-HCl [pH 8.0], 10% [w/v] dextran sulfate, 2% [w/v] blocking reagent [Roche, Mannheim, Germany], 0.1% [w/v] sodium dodecyl sulfate, and formamide (its concentration depending on the probe)) and 2 μL of horseradish peroxidase-labeled probe solution (50 ng μL^{-1}; Greiner Bio-One, Frickenhausen, Germany). The stock solution of the blocking reagent (10% w/v) was prepared in maleic acid buffer (100 mM maleic acid, 150 mM NaCl, pH 7.5). The optimal formaldehyde concentration and hybridization temperature were determined by testing a series of formamide concentrations (0–60%) and two hybridization temperatures (35 and 46 °C) to produce maximum detection rates. The hybridization reaction was conducted in a 24-well microplate overnight (up to 18 h) with mild agitation (10 rpm). The microplate was sealed carefully with parafilm to prevent evaporation of the hybridization buffer. After hybridization, the filter pieces were washed with 0.05% (v/v) Triton X-100 amended with PBS for 15 min. Stringent washing was omitted, as Wendeberg (2010) reported that it did not make a significant difference to CARD-FISH results because CARD-FISH works with lower concentrations of the probe than does FISH using fluorochrome-labeled probes. After removing excess liquid from the filters, they were incubated in 30 μL of amplification mixture (1× amplification diluent [PerkinElmer, Waltham, MA USA]: 40% [w/v] dextran sulfate:fluorescein-tyramide reagent [PerkinElmer] = 25:25:1) in a 24-well microplate and incubated in the dark for 45 min at 37 °C. The filters were subsequently dipped in 0.05% (v/v) Triton X-100 amended with PBS for 5–10 min in the dark, washed in MQ water, and dehydrated with 98% ethanol. The filters were then mounted on a glass slide with an anti-fading reagent (Citifluor [Citifluor, Leicester, UK]: Vectashield [Vector Laboratories, Burlingame, CA, USA] = 4:1) containing approximately 1 $\mu g\ mL^{-1}$ of DAPI.

For each sample, two or three microscope pictures of DAPI-positive cells and the corresponding FISH-positive cells with UV (330-350 nm excitation by U-MWU, Olympus, Tokyo, Japan) and blue (460–-490 nm excitation by U-MWB2, Olympus) excitation, respectively, were taken at 400× magnification using an epifluorescence microscope (BX60; Olympus) and an attached digital-camera (EOS Kiss X5; Canon, Tokyo, Japan). The microscope pictures were then processed with an automated image processing program, the "EBImage" package of the software R (Sklyar et al 2012).

DNA extraction, polymerase chain reaction (PCR), and purification

DNAs were extracted from 0.4 g of homogenized soil by use of a PowerSoil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) following the manufacturer's instructions, with an additional incubation step at 65 °C for 10 min followed by 3 min of bead beating. Eluted DNAs were stored at −20 °C until further processing.

Amplification, purification, and pooling were conducted by following Bates et al. (2011). Briefly, the method includes targeted amplification of a portion of the 16S small-subunit ribosomal gene, triplicate PCR-product pooling (per sample) to mitigate reaction-level PCR biases, and IonPGM high-throughput sequencing (Rothberg et al 2011; Ion Torrent by Life Technologies, Guilford, CT, USA). PCR amplification used the primers F515 (5′-GTGCCAGCMGCCGCGGTAA-3′) and R806 (5′-GGACTACVSGGGTA TCTAAT-3′). This primer set is *in silico* shown to amplify 16S rRNA genes from a broad range of archaeal and bacterial groups with few biases or excluded taxa (Bates et al 2011). Although a broad range of archaeal and bacterial groups is "PCR amplifiable" using the primer set (Bates et al. 2011), amplification efficiency may be different among sequences, which could be one of potential factors that contributed to the inconsistency between the two methods shown in this study.

For the primer set, IonPGM sequencing adaptors and 6-bp barcode sequence (unique to each individual sample) were included. PCR was performed in 25 μL reactions, each containing 1 μL of 10 μM of forward and reverse primers, 10 μL of 5Prime HotMasterMix (Eppendorf-5Prime Inc., Gaithersburg, MD, USA), and 2 μL of extracted soil DNA as a template. The PCR were performed following cycling parameters: 35 cycles (95 °C, 30 s; 50 °C, 1 min; 72 °C, 1 min) after an initial denaturation 95 °C, 3 min. Pooled triplicate PCR products were combined and purified using the UltraClean PCR clean-up kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) and then quantified using Qubit dsDNA HS Assay Kits (Invitrogen, Life Technologies, Carlsbad, CA, USA). The single composite barcoded PCR product was normalized in equal amounts to produce equivalent sequencing depth from all samples. The sample was sent for sequencing at Life Technologies Japan (Tokyo, Japan) on IonPGM.

Sequence data handling and downstream statistical analyses

A single FASTQ file containing raw sequences and quality scores was obtained as a result of IonPGM sequencing. The high-throughput sequence dataset was processed using the QIIME pipeline (Caporaso et al 2010). The FASTQ file was converted to FASTA and QUAL files using convert_fastaqual_fastq.py script. After the conversion, the initial filtering and demultiplexing was conducted according to their unique 6-bp barcode by split_libraryies.py script with default parameter settings, and then operational taxonomic units (OTUs) were picked by pick_otus.py script based on ≥ 97% similarity with the USEARCH option (Edgar 2010). In the pick_otu.py script, chimera sequences were removed using UCHIME (Edgar et al 2011) and singleton OTUs were also removed. After taxa were assigned to the sequences, OTUs associated with non-prokaryotic sequences (i.e., OTUs identified as chloroplast and mitochondria) were removed filter_taxa_from_otu_table.py script. The output file (i.e., BIOM format file) was summarized by summarize_taxa_through_plots.py script, and rarefaction curves were drawn using the BIOM file by alpha_rarefaction.py script (Fig. S1).

In addition to the common pipeline (QIIME), UPAESE was used for the same sequence dataset to check the dependency of our results on data processing pipeline because data handling procedures often have significant impacts on results and interpretations of high-throughput sequencing analysis (Edgar 2013). UPARSE pipeline is a recently published new pipeline, which allows an accurate OTU identification (Edgar 2013). We generally followed a data handling manual described in Edgar (2013) and its website (http://drive5.com/usearch/manual/uparse_cmds.html). Briefly, the raw FASTQ file was processed by fastq_strip_barcode_relabel2.py script. Then, quality filtering (i.e. global trimming of 150 bp length and minimum Phred score 15), dereplication, abundance sort, singleton removal, OTU clustering, and chimera filtering were conducted by following the manual in the website. Taxa were assigned using assign_taxonomy.py script implemented in QIIME. UPARESE analyses under different conditions (e.g., global trimming of 230 bp length) were also tested, and qualitatively similar results were obtained. As a result of UPARSE pipeline analysis, a total of 187,093 sequences (4,454 ± 1,433 [± standard deviation] sequences per sample) passed the filtering processes and a total of 1,702 OTUs were identified. Although the number of OTU was significantly reduced compared with QIIME analysis (5643 OTUs; Table S1), the general trend of the QIIME analysis (Figs. 1c and 2) was also kept for the UPARSE analysis (Fig. S2).

The BIOM file generated by the QIIME processing was exported to the “phyloseq” package (McMurdie and Holmes 2013) in free statistical environment R (R Development Core Team 2013). Unweighted UniFrac distance matrix (Lozupone and Knight 2005) was calculated because the comparison between CARD-FISH results and high-throughput sequencing results suggested that high-throughput sequencing data could not be reliably quantitative. Then, principle coordinate analysis (PCoA) was performed on the unweighted UniFrac distance matrix (Fig. S3). Principle coordinate values were compared among NSC locations using analysis of variance (ANOVA) and the Tukey's HSD test.

Sequence data accessibility

The sequence data is deposited in DDBJ (DNA Data Bank of Japan) Sequence Read Archive (DRA). The accession numbers are DRA001218 for the submission data, DRP001283 for the study data, DRS012794–DRS012835 for the sample data, DRX012986–DRX013027 for the experiment data, and DRR014466–DRR014507 for the run data. Please note that barcode tags registered in DRX012987–DRX013027 (experiment xml files) are incorrect. Correct barcode tags are shown in Table S1 as well as the experimental design description in the xml files.

References in Supplementary methods

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME Journal* **5:** 908-917.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7:** 335-336.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Edgar RC (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*. **10**:996-998

Eickhorst T, Tippkötter R (2008). Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biology and Biochemistry* **40:** 1883-1891.

Klaus, M., Becher, M., Klaminder, J., 2013. Cryogenic Soil Activity along Bioclimatic Gradients in Northern Sweden: Insights from Eight Different Proxies. *Permafrost and Periglacial Processes* 24, 210–223.

Lauber CL, Hamady M, Knight R, Fierer N (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* **75:** 5111-5120.

Lozupone C, Knight R (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71:** 8228-8235.

Makoto K, Klaminder J (2012). The influence of non-sorted circles on species diversity of vascular plants, bryophytes and lichens in Sub-Arctic Tundra. *Polar Biology* **35:** 1659-1667.

McMurdie PJ, Holmes S (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS one* **8:** e61217.

R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M *et al* (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*

475: 348-352.

Sklyar O, Pau G, Smith M, Huber W (2012). EBImage: Image processing toolbox for R. R package version 3.11.0.

Ushio M, Makoto K, Klaminder J, Nakano SI (2013). CARD-FISH analysis of prokaryotic community composition and abundance along small-scale vegetation gradients in a dry arctic tundra ecosystem. *Soil Biology and Biochemistry* **64:** 147-154.

Wendeberg A (2010). Fluorescene in situ hybridization for the identification of environmental microbes. *Cold Spring Harbor Protocols* **5**. doi:10.1101/pdb.prot5366

Table S1. Sample description, sequence counts per sample and barcode sequence.

Sample ID	Barcode	Location	Individual NSC ID	Sequences/sample		
				after the initial filtering and demultiplexing[1]	after the second filtering[2]	OTU count
NO.1	ATGTGG	Location 0	A2	19,101	12,433	2,231
NO.2	TTCCGA	Location 1	A2	17,835	11,309	2,224
NO.3	CCGGCC	Location 2	A2	29,976	18,458	2,539
NO.4	CATTAG	Location 3	A2	15,612	9,169	1,886
NO.5	GCTCTG	Location 4	A2	21,049	12,540	2,310
NO.6	TCACGG	Location 5	A2	12,502	7,293	1,726
NO.7	TATATA	Location H	A2	17,470	11,320	1,849
NO.8	CGGAGA	Location 0	A3	25,050	15,193	2,348
NO.9	CGATGC	Location 1	A3	20,044	10,455	1,833
NO.10	CGACTT	Location 2	A3	23,884	16,994	2,429
NO.11	CTGCGC	Location 3	A3	20,533	11,401	2,233
NO.12	GTAAGC	Location 4	A3	39,494	27,749	2,952
NO.13	CCACCA	Location 5	A3	16,097	11,832	2,107
NO.14	GCCAAC	Location H	A3	26,151	18,506	2,445
NO.15	GACCGC	Location 0	A4	21,859	13,918	2,171
NO.16	GTTCGC	Location 1	A4	26,143	16,092	2,039
NO.17	CGAAGG	Location 2	A4	21,527	15,918	2,271
NO.18	CTCCAG	Location 3	A4	23,047	16,236	2,057
NO.19	TGATGT	Location 4	A4	36,170	27,034	2,575
NO.20	TAGCCA	Location 5	A4	21,227	15,371	2,298
NO.21	TCCGTC	Location H	A4	24,512	17,983	2,186
NO.22	TATTGT	Location 0	A6	19,241	13,666	2,110
NO.23	AGGAGT	Location 1	A6	22,080	12,844	2,285
NO.24	GGATAT	Location 2	A6	32,581	19,914	2,088
NO.25	GCTTGT	Location 3	A6	27,797	20,496	2,756
NO.26	GCCTTC	Location 4	A6	13,133	9,534	2,097
NO.27	CTACAC	Location 5	A6	17,100	11,257	2,002
NO.28	GCCAGT	Location H	A6	21,911	15,377	2,150
NO.29	GGCTGT	Location 0	B2	16,312	9,680	1,911
NO.30	CAATCG	Location 1	B2	23,883	11,344	1,805
NO.31	GTGTAC	Location 2	B2	22,438	12,428	2,278
NO.32	GCGCCG	Location 3	B2	28,655	21,412	2,582
NO.33	TTCTCG	Location 4	B2	20,216	14,353	2,271
NO.34	ATTATC	Location 5	B2	24,666	18,849	2,202
NO.35	TATAAT	Location H	B2	28,040	20,882	2,713
NO.36	GGCTCC	Location 0	B3	10,160	6,422	1,502
NO.37	AACCGT	Location 1	B3	24,597	15,192	2,168
NO.38	ATGTAA	Location 2	B3	24,092	13,950	2,236
NO.39	CGTTAT	Location 3	B3	22,785	15,506	2,391
NO.40	AGCATT	Location 4	B3	20,867	14,240	2,219
NO.41	GCTAGC	Location 5	B3	15,677	12,048	1,968
NO.42	GCGTCT	Location H	B3	20,486	13,747	1,570
			Total	**936,000**	**620,345**	**5,643**
			Mean	**22,286**	**14,770**	**2,191**
			SD	**5,885**	**4,564**	**298**

[1] The initial filtering was done by the QIIME command splity_libraryies.py, which includes several quality filtering steps based on sequence length, quality scores and so on.

[2] The second filtering includes processes of chimera filtering (UCHIME) and remove singleton OTUs or associated with non-prokaryotic sequences such as chloroplast and mitochondria.

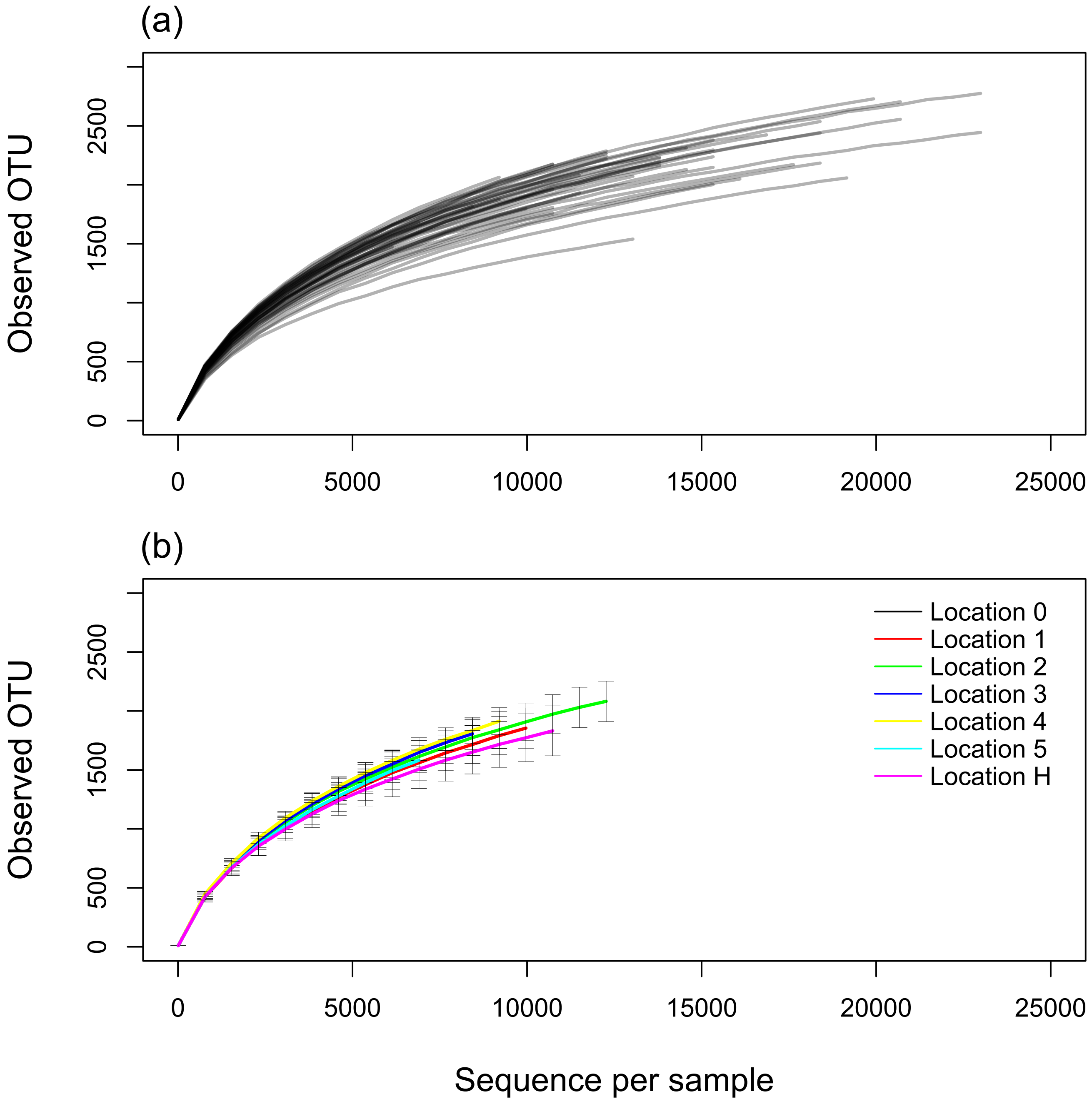

Figure S1 Ushio et al.

Rarefaction results for each sample (a) and each location in a NSC (b). Bars in the bottom figure indicate standard deviation of six replicates in each location.

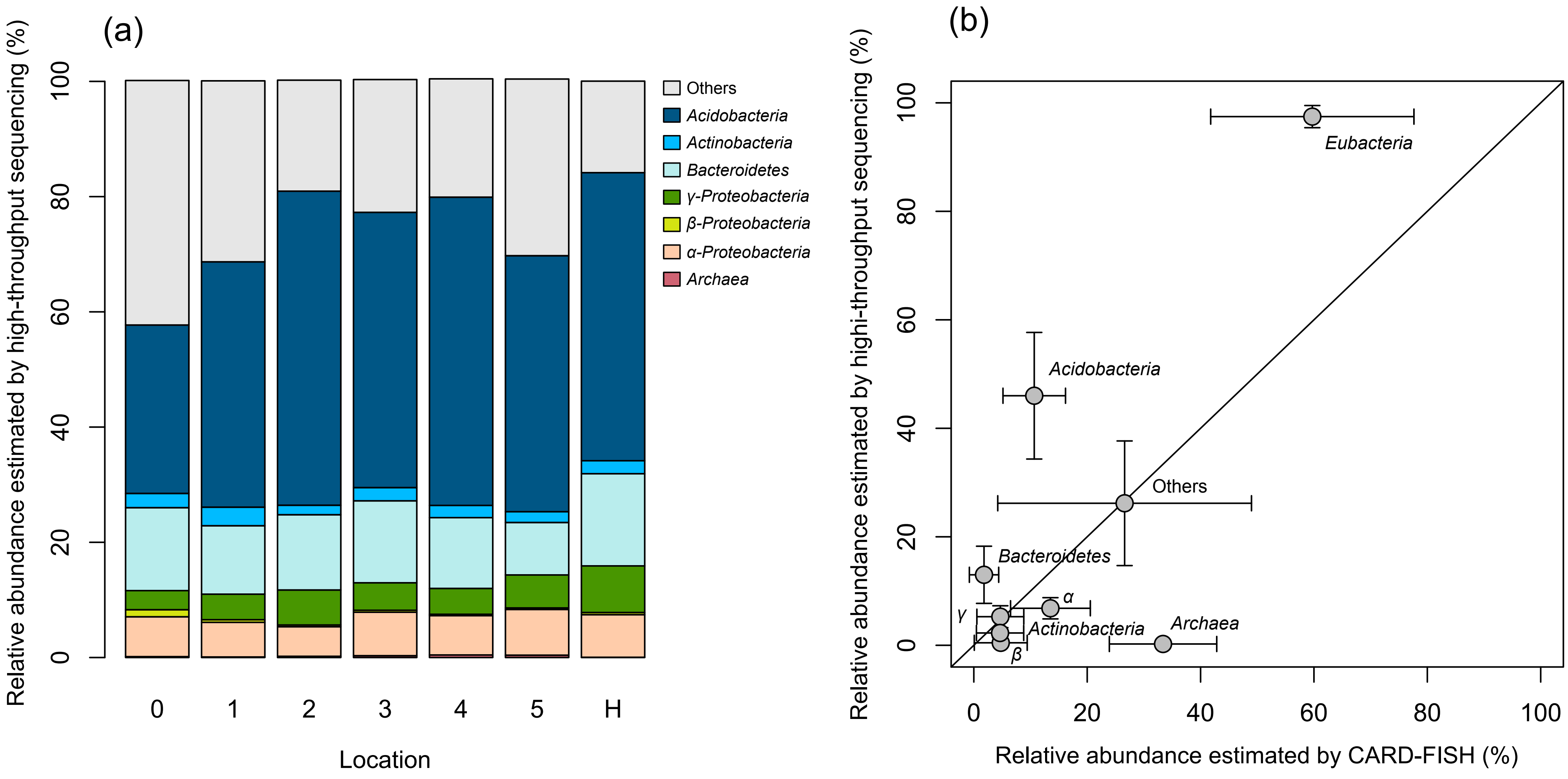

Figure S2 Ushio et al.

Results of UPARSE pipeline analysis. (a) Prokaryotic community composition estimated by high-throughput sequencing analysis followed by UPARSE data processing (corresponds to Fig. 1c). "Others" indicates prokaryotic microbes other than the listed microbial groups, and unidentified sequence reads. (b) Comparison of relative abundance estimated by high-throughput sequencing analysis followed by UPAESE data processing and those with CARD-FISH analysis (corresponds to Fig. 2).

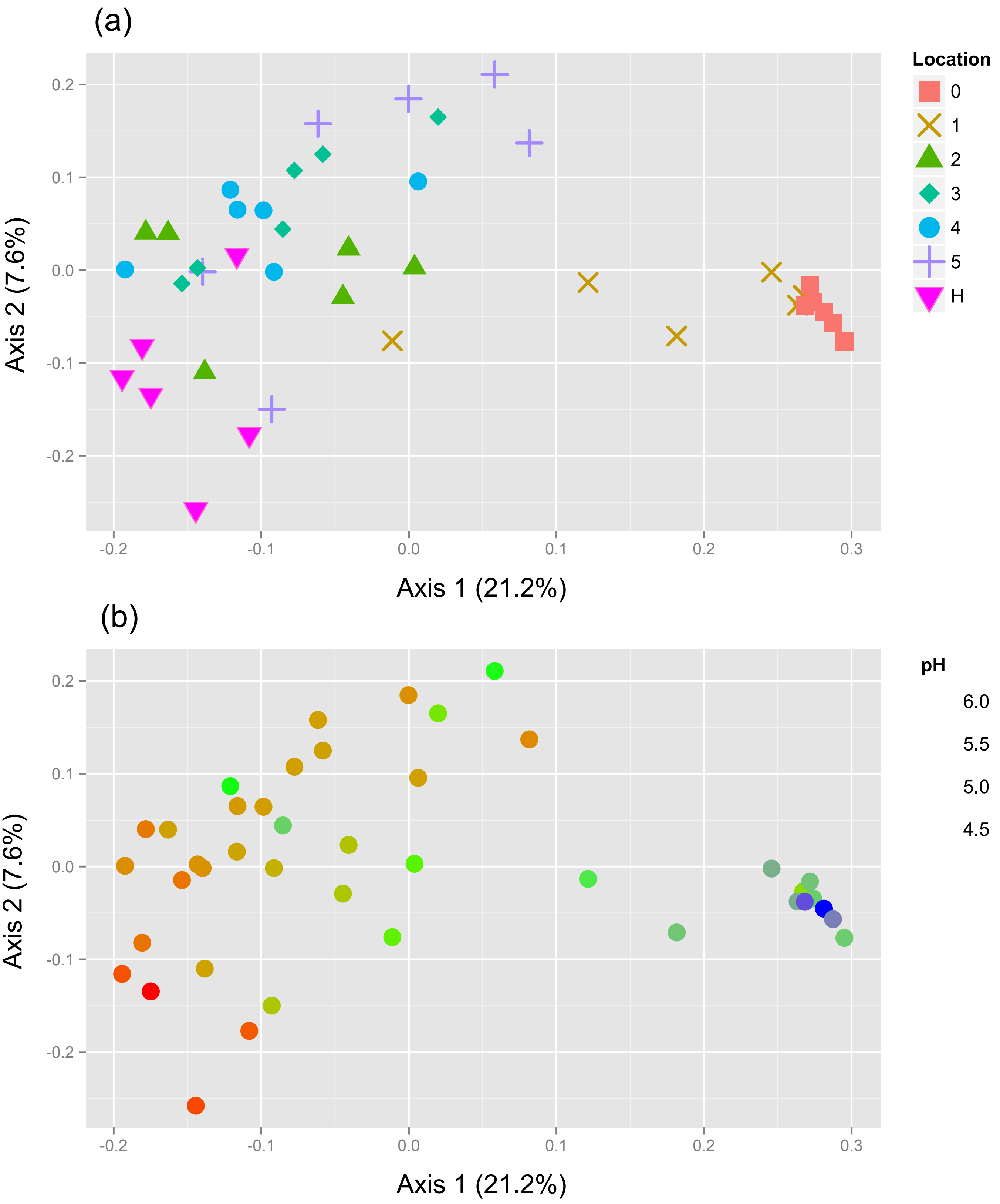

Figure S3 Ushio et al.

Principle coordinate analysis (PCoA) of high-throughput sequence data processed by QIIME pipeline. Unweighted UniFrac distance was used in this analysis. Symbols coded by locations (a) and soil pH (b). Principle component (PC) 1 (*x*-axis) explains 15.9% of total variation, while PC2 (*y*-axis) explains 7.5% of total variation. In terms of PC1 values, location 0 and 1 are significantly different from other locations ($P < 0.0005$), and location H is significantly different from location 5 ($P < 0.05$). In terms of PC2 values, location 0 is significantly different from location 5 ($P < 0.05$), and location H is significantly different from location 3, 4 and 5 ($P < 0.005$).