Normalization of environmental metagenomic DNA enhances the discovery of under-represented microbial community members

Jean-Baptiste Ramond[1], Thulani P. Makhalanyane[1], Marla I. Tuffin[2], Don A. Cowan[1,*]

[1] Centre for Microbial Ecology and Genomics (CMEG), Genomic Research Institute, University of Pretoria, Pretoria 0002, South Africa

[2] Institute for Microbial Biotechnology and Metagenomics (IMBM), University of the Western Cape, Bellville 7535, South Africa

* **Corresponding author**: Prof. Don A. Cowan

Phone: +27 (0)12 420 587

Email: don.cowan@up.ac.za

Jean-Baptiste Ramond: jean-baptiste.ramond@up.ac.za

Thulani P. Makhalanyane: thulani.makhalanyane@up.ac.za

Abstract

Normalization is a procedure classically employed to detect rare sequences in cellular expression profiles (i.e. cDNA libraries). Here, we present a normalization protocol involving the direct treatment of extracted environmental metagenomic DNA with S1 nuclease; referred to as normalization of metagenomic DNA: NmDNA. We demonstrate that NmDNA, prior to *post hoc* PCR based experiments (16S rRNA gene T-RFLP fingerprinting and clone library), increased the diversity of sequences retrieved from environmental microbial communities by detection of rarer sequences. This approach could be used to enhance the resolution of detection of ecologically relevant rare members in environmental microbial assemblages and therefore is promising in enabling a better understanding of ecosystem functioning.

 Significance and Impact of the Study

This study is the first testing 'normalization' on environmental metagenomic DNA (mDNA).

The aim of this procedure was to improve the identification of rare phylotypes environmental communities. Using hypoliths as model systems, we present evidence that this post-mDNA extraction molecular procedure substantially enhances the detection of less common phylotypes and could even lead to the discovery of novel microbial genotypes within a given environment.

Introduction

Environmental microbial ecology studies have been revolutionized by the advent of molecular biology which has provided access to the environmental "microbial dark matter", i.e. the vast majority of yet-to-culture phylo/genotypes which still include complete bacterial divisions (Coleman and Whitman, 2005; Jansson and Prosser, 2013; Rinke et al., 2013). Total environmental nucleic acid (DNA and/or RNA) extraction in association with 16S rRNA gene (pyro)sequencing has led to an extensive description of microbial community assemblages within Earth's many environmental niches (e.g. Baker and Banfield, 2003; Fierer et al., 2012; Pointing and Belnap, 2012). Such a systemic approach has enabled us to unravel microbial community diversity, colonization mechanisms, functional roles, temporal/spatial evolution *in situ* or in controlled systems as well as the modelling of community evolution and biochemical capacities (e.g. Fierer et al., 2013; Jansson and Prosser, 2013; Makhalanyane et al., 2013a,b; Ramond et al., 2014).

In microbial community phylogenetic studies, particularly those involving the use of high-throughput sequencing technologies, two experimental steps have a major impact on the apparent diversity: the effectiveness of the DNA extraction protocol (Terrat et al., 2012; Vishnivetskaya et al., 2014) and the biases introduced by PCR amplification protocols (Klindworth et al., 2013). In particular, the presence of strongly dominating species in a community, represented in PCR reactions in high amounts in the template DNA, reduces the effectiveness of PCR amplification of rarer template sequences (Ogram, 2000; Hazen et al., 2013).

'Normalization' procedures have been developed as means of reducing the concentration of dominant sequences, generally in cDNA libraries, to equalize the number of gene copies/sequences in libraries, independently from their original abundance (Shcheglov et al.,

2007). Numerous normalization processes have been published (and patented) in the discovery of novel genes (Soares et al., 1994; Bonaldo et al., 1996; Short and Mathur, 1998, 1999, and 2003; Zhulidov et al., 2004; Shcheglov et al., 2007; Bogdanova et al., 2008), with the objective of increasing the chances of identifying rare gene sequences in cellular expression profiles (transcriptomes). Recently, normalization has been used successfully to eliminate evolutionarily young repetitive elements in human genomic DNA (Shagina et al., 2010).

Normalization has, to the best of our knowledge, never been employed in microbial ecology. Despite high-throughput sequencing technologies, which now commonly enable the recovery of over one million reads per run per sample (Di Bella et al., 2013), in environmental microbial assemblages rare sequences or phyla (the 'rare biosphere') may still be ignored and/or eliminated by the *post hoc* bioinformatics analyses (Kunin et al., 2010). This could potentially hamper accurate analyses of microbially mediated ecological processes; particularly when performing meta-transcriptome studies, as despite their very low abundance, rare members of environmental communities may perform important ecosystem functions (Sogin et al., 2006).

In this study, we have 'normalized' total extracted metagenomic DNA (mDNA) preparations by S1 nuclease treatment (Shagina et al., 2010). The normalization process relies on the fact that during re-hybridization of denatured double-stranded DNA, the more common sequence will re-hybridize less accurately, with a higher frequency of S1 nuclease susceptible regions (e.g. hair pin loops). The aim of this procedure was to improve the identification of rare phylotypes in environmental communities, and we present evidence that this post mDNA extraction molecular procedure substantially enhances the detection of less common phylotypes and could even lead to the discovery of novel microbial genotypes within a given environment.

4

Materials and Methods

Hypolith sampling and storage

Hypolithic samples were obtained from the desert pavements in the vicinity of the Gobabeb Training and Research Centre (www.gobabebtrc.org; Namibia) as previously described (Makhalanyane et al., 2013a). Briefly, hypolithic biomass was recovered by scraping adherent material from the rock sub-surface. Samples were transported to the laboratory, homogenized with a sterile spatula, transferred into 2 ml tubes and frozen at -80°C until further use.

DNA extraction

DNA was extracted using the PowerSoil DNA isolation kit (MoBio laboratories, USA) according to the manufacturer's instructions.

PCR amplification, purification and restriction

All polymerase chain reactions (PCRs) were carried out in a Bio-Rad Thermocycler (T100^{TM} Thermal Cycler). Bacterial 16S rRNA encoding genes were amplified using the universal primers sets E9F (5′-GAGTTTGATCCTGGCTCAG-3′) / U1510R (5′-GGTTACCTTGTTACGACTT-3′) or 341F (5′- CCTACGGGAGGCAGCAG -3′) / 908R (5′-'CCGTCAATTCMTTTGAGTTI- -3′) according to the protocols (PCR reactions and conditions) described by Ramond et al (2014) and Makhalanyane et al (2013a), respectively. For T-RFLP analyses, the E9F primer was 5'-end FAM-labelled and the PCR products were purified using the GFX^{TM} PCR DNA and gel band purification kit as directed by the supplier

(GE Healthcare, UK). Purified PCR products (200 ng) were digested with the restriction enzyme *Hae*III at 37°C overnight.

Normalization processes

Two variants of a standardized normalization process were tested in this study (Figure 1). The procedures involved the heat denaturation of either environmental metagenomic DNA or 16S rRNA gene PCR amplicon sets. The denatured DNA preparations were subsequently re-annealed and then digested by S1 nuclease (which degrades single-stranded DNA and double-stranded DNA at single-stranded regions such as hair-pin loops; Shagina et al., 2010).

Normalization of metagenomic DNA (NmDNA)

In a final volume of 30 µL, extracted environmental metagenomic DNA was mixed with a hybridization buffer (0.12M NaH_2PO_4, pH 6.8/ 0.82M NaCl/ 1mM EDTA/0.1% SDS; Short and Mathur 1998, 1999 and 2003), denatured for 12 min at 100°C and allowed to re-hybridize at 68°C for 3h. 3 µL of a 7.5M ammonium acetate solution, 3 µL of glycogen (20 ng.$µL^{-1}$) and 90 µL of 95% ice cold ethanol were then added to the mixture, and incubated overnight at -80°C for DNA precipitation. After centrifugation (13000 rpm, 20 min, 4°C), the pellet was re-suspended in ice cold 70% ethanol, centrifuged (13000 rpm, 20 min, 4°C) and the pellet air-dried before re-suspending in sterile ddH_2O.

300 ng of the resulting DNA was then digested using S1 nuclease (Thermo Scientific, USA) according to the manufacturer's instructions. S1 nucleases were removed using a standard phenol/chloroform extraction protocol, where 1 vol phenol was added to the mixture, the aqueous phase removed after centrifugation (2000 rpm, 5 min, 4°C) and treated with 1 vol of

chloroform/isoamyl alcohol (24:1) solution. After centrifugation (2000 rpm, 5min, 4°C), the aqueous phase was removed and supplemented with 1/10 vol of 7.5M ammonium acetate with 3 vol of ice cold isopropanol. After 1h at -80°C, the mixture was centrifuged (13000 rpm, 20 min, 4°C) and the pellet washed with ice-cold 70% ethanol. After centrifugation (13000 rpm, 20 min, 4°C), the DNA pellet was air-dried and re-suspended in sterile ddH$_2$0 overnight at 4°C.

The normalized metagenomic DNA obtained was used for T-RFLP analysis and the construction of bacterial 16S rRNA gene clone libraries.

Amplicon Normalization (AN)

After PCR amplification of the 16S rRNA gene from the mDNA preparations, amplicons were purified using a NucleoSpin® Extract II kit (Macherey-Nagel, Germany) as directed by the supplier. As for the NmDNA procedure, the purified amplicons were denatured for 12 min at 100°C, and allowed to re-hybridize at 68°C for 2h. After purification using the NucleoSpin® Extraction kit, 500 ng of the purified PCR amplicons were subjected to S1 nuclease digestion according to the manufacturer's instructions and purified with the NucleoSpin® Extract II kit. The normalized PCR amplicon preparation was then used for T-RFLP analysis.

Clone library construction

16S rRNA gene PCR amplicons from non-treated mDNA or the normalized mDNA were purified with NucleoSpin® Extract II kit (Macherey-Nagel, Germany) and cloned into pGEM®-T Easy (Promega) according to the manufacturer's instructions. Ligation reactions

were transformed into electrocompetent *E. coli* DH5αMCR using a BioRad Gene Pulser. Transformants were selected by blue-white screening on LB/ampicillin/IPTG/X-Gal plates according to standard methods. All putative recombinant clones were screened by colony PCR using universal M13 primers.

200 and 192 clones from the mDNA and NmDNA libraries, respectively, were screened by ARDRA. The sequencing of de-replicated clones (34 and 33, respectively) was performed with an ABI 3130 DNA Sequencer (Applied Biosystems, Foster City, California) at the Central Analytical Facility of the University of Stellenbosch (South Africa).

Terminal-Restriction Fragment Length Polymorphism (T-RFLP)

To perform T-RFLP analyses, the E9F primer was 5'-end FAM-labelled and the PCR products were purified using the GFX^{TM} PCR DNA and gel band purification kit as directed by the supplier (GE Healthcare, UK). 200 ng of purified PCR products was digested overnight at 37°C with the restriction enzyme *Hae*III. T-RF size was determined by capillary electrophoresis using an Applied Biosystems DNA Sequencer 3130 (Applied Biosystems, Foster City, California, USA) with an error of ±1 bp.

To analyze the T-RFLP profiles, only the T-RFs between 35 and 1000 bp were taken into account. The T-RFs detected by the software Peak Scanner (Applied Biosystem, USA) were compiled into data matrixes and processed using the online software T-REX (http://trex.biohpc.org/index.aspx; Culman et al., 2009). The relative abundance of a true terminal restriction fragment within a given T-RFLP pattern was generated as a ratio of the respective peak height. The term OTU is used to refer to individual restriction fragments in T-

RFLP patterns (based on variation in the 16S rRNA gene), with recognition that each OTU may comprise more than one distinct bacterial ribotype (Ramond et al., 2014).

Non-metric multidimensional scaling (MDS) plots were created using Bray-Curtis similarity matrices of square-root transformed data with the software Primer6 (Primer-E Ltd, UK). Two-dimensional MDS plots are used, where the distance between points reflects the degree of similarity between the microbial community profiles in the samples. In these plots, the percentage of similarity determined by cluster analysis is displayed by ellipses.

Phylogenetic analysis

Chimeric sequences were filtered using Bellerophon (http://comp-bio.anu.edu.au/bellerophon/bellerophon.pl; Huber et al., 2004). Phylogenetic OTUs (97% similarity cut-off) were determined using CD-Hit (Huang et al., 2010). The libraries were compared using the RDP-based online library compare tool (http://rdp.cme.msu.edu/comparison/comp.jsp; Cole et al., 2009). Sequence data have been submitted to NCBI GenBank database (accession numbers JN714842 - JN714926 and KM010192-KM010223).

Results and Discussion

The NmDNA protocol shown in Figure 1 was used to enrich the diversity of bacterial sequences retrieved from an environmental sample by removing a proportion of the most abundant ones. Briefly, following metagenomic DNA extraction, which leads to the recovery of sheared double stranded DNA, the latter is denatured and re-hybridized. This process is based on the expected differential re-hybridization efficiencies of highly abundant sequences (i.e. in our study, sequences sharing high identities) and rare sequences; the former being more likely to form secondary structures hydrolysable by S1 nuclease (Shagina et al., 2010). S1 nuclease-treated mDNA preparations were subjected to microbial community analysis using T-RFLP fingerprinting (Figures 2 and 3) and clone library construction (Figure 4). A similar S1 nuclease-based procedure on human genomic DNA (which included the ligation of adapter sequences as an additional step) has been shown to eliminate high identity sequences (Shagina et al., 2010).

To assess the impact of NmDNA on environmental microbial community profiles, the resulting diversity was compared to that retrieved from non-normalized mDNA and from a more 'classic' normalization procedure, i.e. amplicon normalization (AN) (Soares et al., 1994; Bonaldo et al., 1996; Short and Mathur, 1998, 1999, and 2003; Zhulidov et al., 2004; Shcheglov et al., 2007; Bogdanova et al., 2008). Hypolithic communities were selected as the mDNA source as these niche communities are relatively simple in trophic structure with aparticularly high abundance of cyanobacterial phyla (Chan et al. 2012; Pointing and Belnap, 2012; Makhalanyane et al., 2013a,b). The application of NmDNA to these communities and the reduction of the frequency of the dominating cyanobacterial sequences may reveal potential keystone taxa which may be poorly represented in 'normal' datasets.

Metagenomic DNA and 16S rRNA gene amplicons sets from two independent hypolithic samples (termed S4 and S5) were normalized. Both the AN and the NmDNA procedures decreased the hypolithic community diversity by 24.5% and 20% (AN) or 24.5% and 15.9% (DmDNA) for S4 and S5,respectively (Figure 2). This was expected, as we used particularly short re-hybridization times (2h for AN and 3h for NmDNA) compared to other published/patented normalization procedures (12h to 36h; Bonaldo et al., 1996; Short and Mathur, 1998, 1999 and 2003). Shorter re-hybridization periods prevent correct (m)DNA re-annealing and favor the formation S1 nuclease-hydrolysable single stranded DNA and hair-pin loops.

Non-normalized bacterial community fingerprints shared more OTUs with the NmDNA-generated fingerprints (S4: 16, Figure 2A; S5: 11, Figure 2B) than those generated by the AN process (S4: 5, Figure 2A; S5: 0, Figure 2B). This result strongly suggests that AN substantially modified hypolithic bacterial community structures when compared to NmDNA. This was corroborated by the nMDS ordination (Figure 3). The NmDNA hypolithic community samples clustered with the non-processed samples while the AN ones were distant. These results imply that NmDNA is more effective than AN in accessing 'novel' bacterial diversities within an environmental community without excessively altering its 'original' composition (and risk studying chimeric sequences).

To further assess the impact of NmDNA on environmental bacterial communities, a control (non-processed) 16S rRNA gene clone library was compared to an NmDNA-generated library (Figure 4). As expected, the control hypolith 16S rRNA gene library was dominated by sequences with a high homology to cyanobacteria (87%, Figure 4; Chan et al. 2012; Makhalanyane et al., 2013a,b; Pointing and Belnap, 2012); and as hypothesized, their relative abundance significantly decreased in the NmDNA-generated library (27.3%, Figure 4). Moreover, the control library contained sequences spanning four bacterial phyla

(Acidobacteria, Cyanobacteria, Actinobacteria, Chloroflexi) and one bacterial class (α-Proteobacteria) already observed in hypolithic environments (Figure 4A; Chan et al., 2012), while the NmDNA-generated library comprised seven bacterial phyla (Acidobacteria, Cyanobacteria, Actinobacteria, Armatimonadetes, Gemmatimonadetes, Verrumicrobia, Bacteroidetes), two classes (α-Proteobacteria, Deinococcus) and 12.1% of sequences originating from unclassified bacteria (Figure 4B). Therefore, the NmDNA process clearly increased the bacterial sequence diversity retrieved from the target environment. Interestingly, these phyla (with the exception of Verrumicrobia and Armatimonadetes phyla) are well known hot desert hypolithic colonists (e.g. Lacap et al., 2011; Azua-Bustos et al., 2011; Caruso et al., 2011), but the phyla Verrumicrobia and Armatimonadetes have, to date, only been observed in cold desert hypolithic communities with pyrosequencing (Lee et al., 2012; Makhalanyane et al., 2013b). The use of a normalization approach has thus extended their known distribution to hot desert hypolithic communities, albeit in low abundance, and these findings support our view that NmDNA can lead to the detection of 'novel' phyla even when applied to low resolution method (e.g. clone library). This finding also suggests that normalization of metagenomic DNA is a promising pre-treatment method in the characterization of members of the 'microbial dark matter' (Rinke et al., 2013); i.e., keystone species which may be numerically rare but perform essential ecosystem function(s) (Shade et al., 2012). This is fundamental in microbial ecology as NmDNA could *in fine* lead to (i) the identification of novel microbial biochemical and functional capacities and, in the process, (ii) better capture the global functioning of Ecosystems (Jansson and Prosser, 2013; Sogin et al., 2006).

We conclude that the findings of this study have significant ecological implications in the evaluation of the relationships between biodiversity and system functionality, through the ability to access previously uncharacterized and numerically less abundant taxa Figure 4).

And, in depauperate environments, such as desert (hypolith) ecosystems, rarer taxa are assumed to play important roles in ecosystem adaptation to changes as sources of functional redundancy (Caron and Countway, 2009). Understanding the quantitative and qualitative characteristics of rare taxa may become critically important under conditions of environmental (climate) change, where extinction events may lead to reduced environmental capacity (Sogin et al., 2006).

Acknowledgements

References

Azua-Bustos, A., C. Urrejola, and R. Vicuña. 2012. Life at the dry edge: microorganisms of the Atacama Desert. FEBS Lett. *586*: 2939-2945.

Baker, B.J. and J.F. Banfield. 2003. Microbial communities in acid mine drainage. FEMS Microbiol. Ecol. *44*:139-152.

Bonaldo, M.F., G. Lennon, and M.B. Soares. 1996. Normalization and substraction: two approaches to facilitate gene discovery. Genome Res. *6*:791-806.

Bogdanova, E.A., D.A. Shagin, and S.A. Lukyanov. 2008. Normalization of full-length enriched cDNA. Molecular BioSystems *4*:205-212.

Caron, D.A., and P.D. Countway. 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquatic Microbial Ecol. *57*:227.

Caruso, T., Y. Chan, D.C. Lacap, M.C. Lau, C.P. McKay, and S.B. Pointing. 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. ISME J. *5*:1406-1413.

Cole, J.R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, and J.M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37 (Database issue): D141-145; doi: 10.1093/nar/gkn879.

Coleman, D.C., and W.B. Whitman. 2005. Linking species richness, biodiversity and ecosystem function in soil systems. Pedobiologia *49*: 479-497.

Chan, Y., D.C. Lacap, M.C.Y. Lau, K.Y. Ha, Warren-Rhodes KA, C.S. Cockell, et al. 2012. Hypolithic microbial communities: Between a rock and a hard place. Environ. Microbiol. 14:2272–2282.

Culman, S.W., R. Bukowski, H.G. Gauch, H. Cadillo-Quiroz, and D.H. Buckley. 2009. T-REX: Software for the Processing and Analysis of T-RFLP data. BMC Bioinformatics *10*:171.

Di Bella, J.M., Y. Bao, G.B. Gloor, J.P. Burton, and G. Reid. 2013.High throughput sequencing methods and analysis for microbiome research. J. Microbiol. Met. *95*:401-414.

Hazen, T.C., A.M. Rocha, and S.M. Techtmann. 2013. Advances in monitoring environmental microbes. Curr. Opinion Biotechnol. *24*:526-533.

Hahn, W.E. and J. Van Ness. 1976. Elimination of double strand nuclease activity from S1 nuclease prepared from crude α amylase. Nucleic Acids Res. *3*:1419-1424.

Huang, Y., B. Niu, Y. Gao, L. Fu, and W. Li. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics *26*:680.

Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics. *20*:2317-2319.

Fierer, N., J.W. Leff, B.J. Adams, U.N. Nielsen, S.T. Bates, C.L. Lauber, J.A. Gilbert, D.H. Wall, and J.G. Caporaso. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. *109*:21390-21395.

Fierer, N., J. Ladau, J.C. Clemente, J.W. Leff, S.M. Owens, Pollard, K. S., R. Knight, J.A. Gilbert, and R.L. McCulley. 2013. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science *342*:621-624.

Jansson, J.K., and J.I. Prosser. 2013. Microbiology: The life beneath our feet. Nature *494*:40-41.

Klindworth, A., E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F.O Glöckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. *41*:e1-e1.

Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. *12*:118-123.

Lacap, D.C., K.A. Warren-Rhodes, C.P. McKay, and S.B. Pointing. 2011. Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. Extremophiles *15*:31-38.

Lee, C.K., B.A. Barbier, E.M. Bottos, I.R. McDonald, and S.C. Cary. 2012. The inter-valley soil comparative survey: the ecology of Dry Valley edaphic microbial communities. ISME J. *6*:1046-1057.

Makhalanyane, T.P., A. Valverde, D.C. Lacap, S.B. Pointing, M.I. Tuffin, and D.A. Cowan. 2013a. Evidence of species recruitment and development of hot desert hypolithic communities. Environ. Microbiol. Rep. *5*:219-224.

Makhalanyane, T.P., A. Valverde, N.K. Birkeland, S.C. Cary, M.I. Tuffin, and D.A. Cowan. 2013b. Evidence for successional development in Antarctic hypolithic bacterial communities. ISME J. *7*:2080-2090.

Pointing, S.B., K.A. Warren-Rhodes, D.C. Lacap, K.L. Rhodes, and C.P. McKay. 2007. Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. Environ. Microbiol. *9*:414–424.

Pointing, S.B. and J. Belnap. 2012. Microbial colonization and controls in dryland systems. Nature Rev. Microbiol. *10*:551–562.

Ramond, J-B., P.J. Welz, M. Le Roes-Hill, M.I. Tuffin, S.G. Burton, and D.A. Cowan. 2014. Selection of *Clostridium* spp. in biological sand filters treating synthetic acid mine drainage. FEMS Microbiol. Ecol. *3*:678- 690.

Rinke, C., P. Schwientek, A. Sczyrba, N.N. Ivanova, I.J. Anderson, J.F. Cheng, et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*:431-4437.

Shade, A., C.S. Hogan, A.K. Klimowicz, M. Linske, P.S. McManus, and J. Handelsman. 2012. Culturing captures members of the soil rare biosphere. Environ. Microbiol. *14*:2247-2252.

Shagina, I., E. Bogdanova, I.Z. Mamedov, Y. Lebedev, S. Lukyanov, and D. Shagin. 2010. Normalization of genomic DNA using duplex-specific nuclease. BioTechniques *48*:455-459.

Shcheglov, A.S., P.A. Zhulidov, E.A. Bogdanova, and D.A. Shagina. 2007. Normalization of cDNA libraries, p97-124. *In* A Buzdin and S Lukyanov (Eds.), Nucleic acids hybridization modern applications. Springer, Netherlands.

Short, J.M and E.J. Mathur. 1998. Production and use of normalized DNA libraries. US patent 5763239.

Short, J.M and E.J. Mathur. 1999. Production and use of normalized DNA libraries. US patent 6001574.

Short, J.M and E.J. Mathur. 2003. Production and use of normalized DNA libraries. US patent 20030104440.

Soares, M.B., M.F. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. *91*:9228-9232.

Sogin, M.L., H.G. Morrison, J.A. Huber, D.M. Welch, S.M. Huse, P.R. Neal, J.M. Arrieta and G.J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. *103*:12115-12120.

Terrat, S., R. Christen, S. Dequiedt, M. Lelièvre, V. Nowak, T. T. Regnier, et al. (2012). Molecular biomass and MetaTaxogenomic assessment of soil microbial communities as influenced by soil DNA extraction procedure. Microbial Biotechnol. *5*:135-141.

Vishnivetskaya, T.A., A.C. Layton, M.C. Lau, A. Chauhan, K.R. Cheng, A.J. Meyers, et al. 2014. Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. FEMS Microbiol. Ecol. *87*:217-230.

Zhulidov, P.A., E.A. Bogdanova, A.S. Shcheglov, L.L. Vagner, G.L. Khaspekov, V.B. Kozhemyako, et al. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. *32*:e37-e37.

Legends to Figures

Figure 1. Schematic protocol of the normalization processes used in this study. The colored lines represent amplified sequences of low-abundant microbial community members. DS: double stranded.

Figure 2. Venn diagram depicting the (un)shared T-RFs between all the S4 (A) and S5 (B) samples. NmDNA: Normalized metagenomic DNA samples / AN: Amplicon Normalized samples.

Figure 3. 2D-Nonmetric multidimensional scaling plot of Bray-Curtis similarity based on square root transformed 16S rRNA gene T-RFLP profiles. ○: S4 non processed metagenomic DNA; ●: S4 Amplicon Normalized (AN); ●: S4 Normalized metagenomic DNA samples (NmDNA); ◇: S5 non processed metagenomic DNA; ◆: S5 Amplicon Normalized (AN) ◆: S5 Normalized metagenomic DNA samples (NmDNA). Ellipses around the samples indicate similarities in bacterial community fingerprints determined by cluster analysis (black: 20%; doted black: 40%; doted grey: 60%).

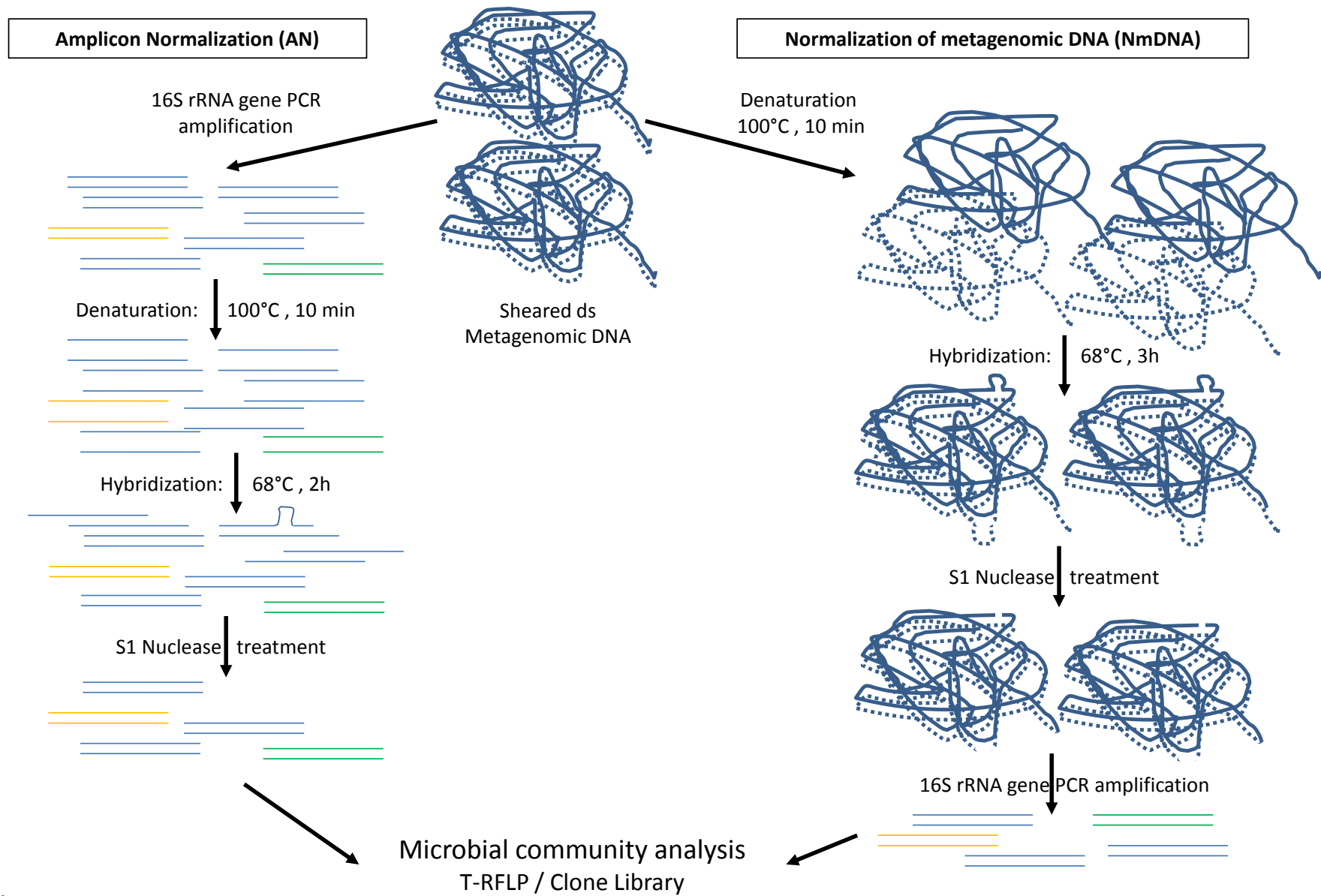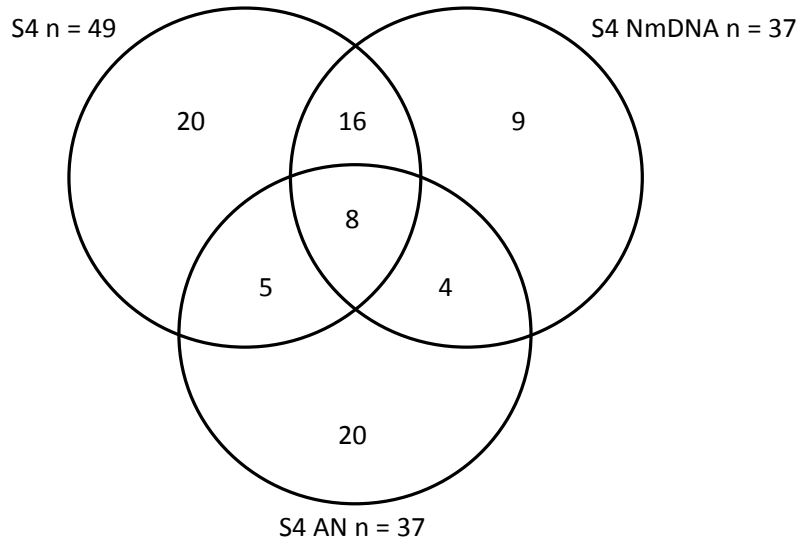**Figure 4**. 16S rRNA bacterial gene clone library composition comparisons.

Amplicon Normalization (AN)
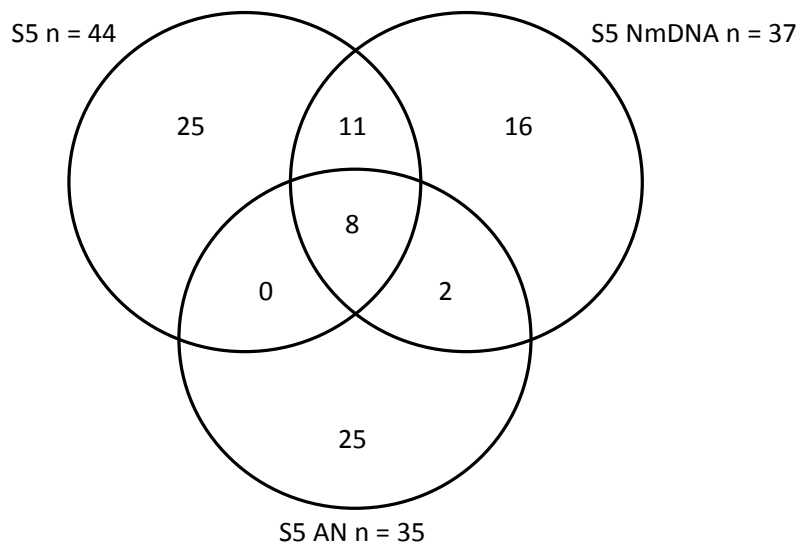
16S rRNA gene PCR amplification

Denaturation: 100°C , 10 min

Hybridization: 68°C , 2h

S1 Nuclease treatment

Normalization of metagenomic DNA (NmDNA)

Denaturation 100°C , 10 min

Hybridization: 68°C , 3h

S1 Nuclease treatment

16S rRNA gene PCR amplification

Sheared ds Metagenomic DNA

Microbial community analysis
T-RFLP / Clone Library

Figure 1

Figure 2

A

S4 n = 49

S4 NmDNA n = 37

20

16

9

8

5

4

20

S4 AN n = 37

B

S5 n = 44

S5 NmDNA n = 37

25

11

16

8

0

2

25

S5 AN n = 35

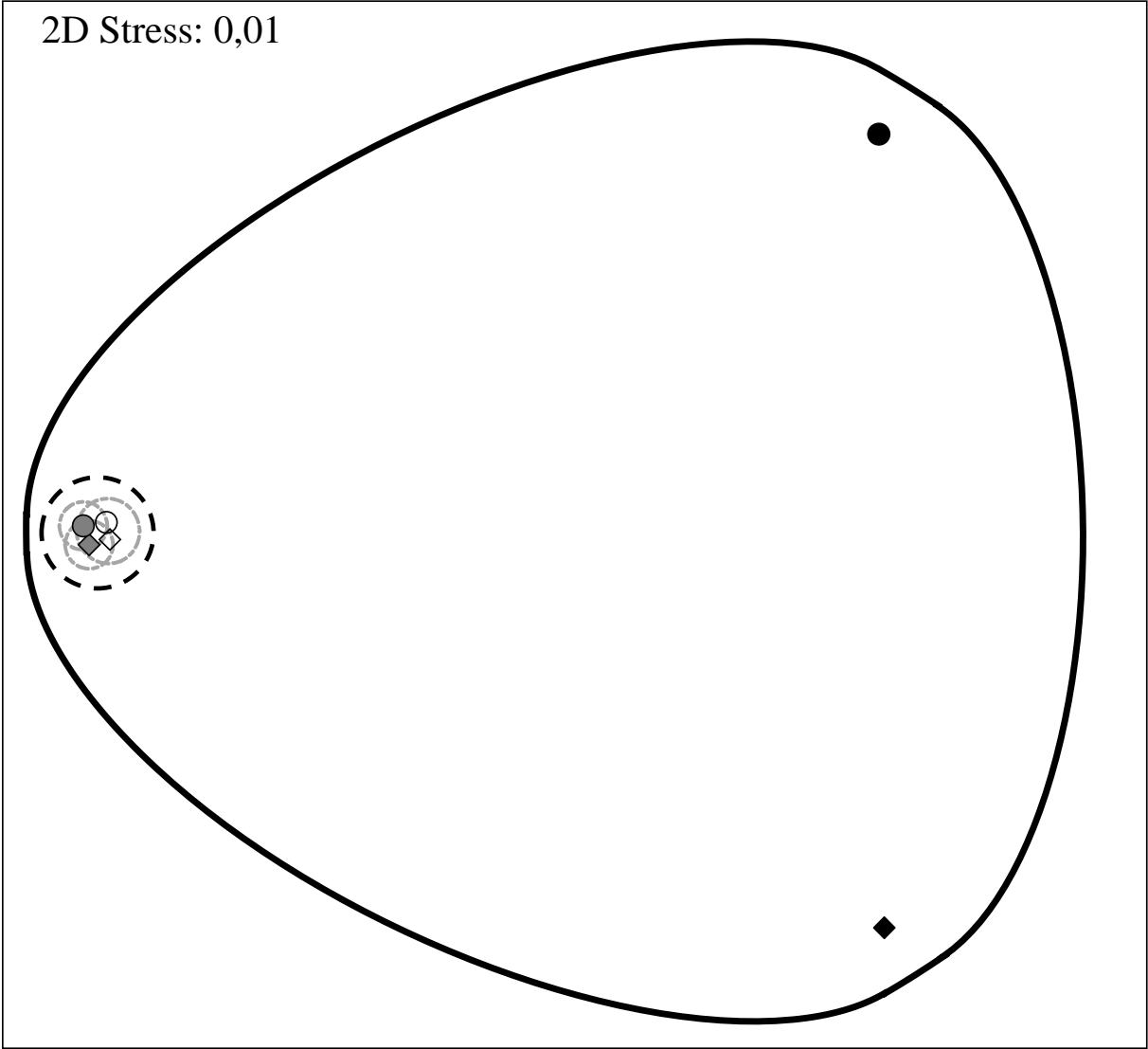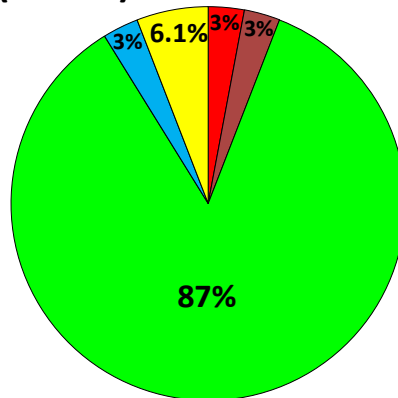**Figure 3**
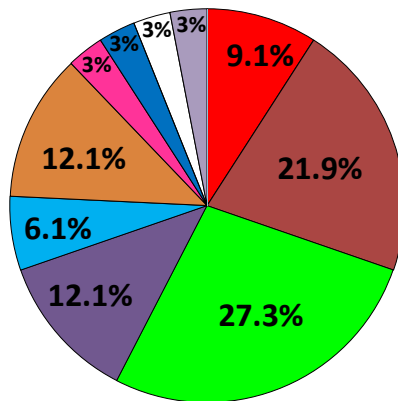

2D Stress: 0,01

Figure 4

Non-processed Library (n = 34)

NmDNA Library (n = 33)

Legend:
- 🟥 Acidobacteria
- 🟫 α-Proteobacteria
- 🟩 Cyanobacteria
- 🟪 Bacteroidetes
- 🟦 Actinobacteria
- 🟧 Unclassified bacteria
- 🟪 Deinococcus
- 🟦 Armatimonadetes
- ⬜ Gemmatimonadetes
- 🟪 Verrumicrobia
- 🟨 Chloroflexi