Supplementary materials for the manuscript

"Streambed microbial communities in the transition zone between groundwater and a first-order stream as impacted by bidirectional water exchange"

Zhe Wang[1,2,8], Oscar Jimenez-Fernandez[3,4], Karsten Osenbrück[3,9], Marc Schwientek[3], Michael Schloter[5,7], Jan H. Fleckenstein[4,6], Tillmann Lueders[1,*]

[1] Chair of Ecological Microbiology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Germany

[2] Institute of Groundwater Ecology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany

[3] Center for Applied Geoscience, University of Tübingen, Tübingen, Germany

[4] Department of Hydrogeology, Helmholtz-Zentrum für Umweltforschung GmbH - UFZ, Leipzig, Germany

[5] Chair of Soil Science, Technical University of Munich, Freising, Germany

[6] Hydrologic Modelling Unit, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany

[7] Research Unit for Comparative Microbiome Analyses, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany

[8] School of Life Sciences, Technical University of Munich, Freising, Germany

[9] Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany

[*]Corresponding authors. E-mail address: tillmann.lueders@uni-bayreuth.de

Content of supplementary materials

Supplementary materials and methods

PCR reactions for 16S rRNA gene sequencing

Each 25 µl PCR reaction contained 1x KAPA HiFi GC buffer, 0.3 mM dNTPs, 0.3 µM of each forward and reverse primer, 0.03 U KAPA HiFi Hot Start DNA polymerase, and ~2 ng template DNA. PCR involved 30 cycles of denaturation at 95 °C for 30 s, annealing at 57 °C for 30 s, and extension at 72 °C for 60 s. Products of the PCR amplification were purified with the MicroElute® Cycle-Pure Kit (Omega Bio-Tek, Georgia, USA), following the manufacturer's instructions. The quality and concentration of purified PCR products was confirmed on a Fragment Analyzer 5200 System (Agilent, Santa Clara, CA, USA), using a DNF-473 NGS Fragment Analysis Kit (1-6000bp) (Advanced Analytical Technologies, Ames, IA, USA). The

subsequent second PCR reaction included, also in 25 µl, 1x KAPA HiFi GC buffer, 0.3 mM dNTPs, 0.24 µM of PacBio barcoded Universal F/R primers (PacBio, Menlo Park, CA, USA), 0.02 U KAPA HiFi Hot Start DNA polymerase, and 0.05 ng template DNA. The cycling conditions were identical to the first PCR reaction, but with 20 cycles only. The PCR products from the second amplification were purified with AMPure PB beads (PacBio, Menlo Park, CA, USA) according to the manufacturer's instructions. The quality and concentration of purified PCR products was checked on an Agilent 2100 Bioanalyzer system (Agilent, Santa Clara, CA, USA).

Calculation of alpha diversity indices

Alpha diversity indices, including the Shannon Index and Shannon diversity based evenness index, were calculated using the phyloseq package (version 1.28.0) (McMurdie and Holmes, 2013). Differences in alpha diversity were assessed by non-parametric Kruskal-Wallis analysis in combination with Dunn’s tests for multiple comparisons, and Benjamini-Hochberg correction for multiple comparisons using DescTools package (version 0.99.41) (Signorelli, 2020).

Data visualization

Major packages used for data visualization include phyloseq (version 1.34.0), ggplot2 (version 3.3.2), tidyverse (version 1.3.0), ComplexHeatmap (version 2.7.6.1002), ggpubr (version 0.4.0), ComplexUpset (version 1.1.0) (Ginestet, 2011; Gu et al., 2016; Kassambara, 2020; Krassowski, 2021; Li, 2020; McMurdie and Holmes, 2013; Wickham et al., 2019). Phylogenetic trees for *Sulfuricurvum* spp. and *Thiobacillus* spp. were then visualized and analyzed with iTol v6 (Letunic and Bork, 2019).

Supplementary Tables

Table S1.

Sediment sample codes according to the standardized Schönbrunnen project database.

Project site code for hydrology monitoring	Sediment sample code
GNS0	*Up-A*
GNS1	*Up-B*
SB25	*Mid-A*
SB26	*Mid-B*
SB40	*Down*
KB2	*Conf*

Table S2.

Universal primer set used for amplifying the full-length 16S rRNA gene from each sediment sample.

Primer	5’ Block	PacBio universal sequence adapters	16S rRNA gene primer
*27F:	/5AmMC6/	gcagtcgaacatgtagctgactcaggtcac	AGRGTTYGATYMTGGCTCAG
**1492R:	/5AmMC6/	tggatcacttgtgcaagcatcacatcgtag	RGYTACCTTGTTACGACTT

Table S3.

Primer sets and annealing temperatures used for gene abundance measurements through qPCR.

Genes	**Primers**	**Primer Sequences (5'–3')**	**Length of the primer**	**Annealing temperature**
Bac. 16S rRNA (Schwieger and Tebbe, 1998)	Ba519f	CAGCMGCCGCGGTAATA	17	54 °C
	Ba907r	CCGTCAATTCCTTTGAGTTT	20	
nirS (Throbäck et al., 2004)	nirS-Cd3aF	GTSAACGTSAAGGARACSGG	20	58 °C
	nirS-R3cd	GASTTCGGRTGSGTCTTSAYGAA	24	
nirK (Henry et al., 2004)	nirK876	ATYGGCGGVCAYGGCGA	17	58 °C
	nirK1040	GCCTCGATCAGRTTRTGGTT	20	

Table S4.

Read statistics on PacBio sequences processed by SMRTLink (version 6.0) from subreads (raw reads) to Circular Consensus Sequences (CCS) reads.

Amplicon	Sample code	Raw sample name	Subreads	CCS reads	Number of passes
1	*Up-A*	17ss_01	791749	29732	16
2	*Up-A*	17ss_02	589002	23244	17
3	*Up-A*	17ss_03	510993	18640	16
4	*Up-A*	17ss_04	94450	5040	14
5	*Up-A*	17ss_05	89193	3644	14
6	*Up-A*	17ss_06	41969	3214	14
7	*Up-B*	17ss_07	313924	13382	16
8	*Up-B*	17ss_08	574149	24542	14
9	*Up-B*	17ss_09	367691	15846	16
10	*Up-B*	17ss_10	165619	7081	14
11	*Up-B*	17ss_11	716356	30923	17
12	*Up-B*	17ss_12	402309	17500	17
13	*Mid-A*	17ss_13	140353	5931	17
14	*Mid-A*	17ss_14	400455	15121	17
15	*Mid-A*	17ss_15	127435	4838	17
16	*Mid-A*	17ss_16	256302	9429	17
17	*Mid-A*	17ss_17	215721	8171	17
18	*Mid-A*	17ss_18	428174	16173	17
19	*Mid-B*	17ss_19	142209	5281	17
20	*Mid-B*	17ss_20	214753	7952	17
21	*Mid-B*	17ss_22	188779	6966	17
22	*Mid-B*	17ss_23	527820	19611	17
23	*Mid-B*	17ss_24	403187	15211	17
24	*Down*	17ss_25	195288	7235	17
25	*Down*	17ss_26	599195	22000	17
26	*Down*	17ss_27	523254	19411	17
27	*Down*	17ss_28	434595	4735	17
28	*Down*	17ss_29	394688	16036	17
29	*Down*	17ss_30	18685	14703	17
30	*Conf*	17ss_31	261526	9865	17
31	*Conf*	17ss_34	150575	4032	17
32	*Conf*	17ss_35	165553	5470	17
33	*Conf*	17ss_36	161400	5376	17

Table S5.

Summary of phylum-level taxa. The mean relative abundance of a phylum in all samples, and the number of ASVs detected within a given phylum are shown in the table. All samples were dominated by three phyla (highlighted in red color): *Proteobacteria*, *Bacteroidota*, and *Acidobacteriota*.

Rank	Phylum	Number of ASVs	Mean relative abundance
1	Proteobacteria	2173	27.7293%
2	Acidobacteriota	1160	9.2575%
3	Bacteroidota	1125	15.9077%
4	Planctomycetota	752	4.9452%
5	Chloroflexi	656	6.2494%
6	unclassified_Bacteria	518	7.6183%
7	Desulfobacterota	389	3.5120%
8	Actinobacteriota	262	1.9312%
9	Firmicutes	233	2.2400%
10	NB1-j	201	1.7083%
11	Sva0485	174	2.4328%
12	Patescibacteria	159	1.5301%
13	MBNT15	145	2.6433%
14	Myxococcota	139	0.5590%
15	Nitrospirota	124	2.2556%
16	Gemmatimonadota	116	1.4032%
17	Campilobacterota	83	3.2710%
18	RCP2-54	76	0.8137%
19	Verrucomicrobiota	65	0.2230%
20	Spirochaetota	64	0.8254%
21	Zixibacteria	46	0.4349%
22	Dependentiae	45	0.2280%
23	Methylomirabilota	43	0.4740%
24	WS2	38	0.4036%
25	Armatimonadota	37	0.1784%
26	Nitrospinota	34	0.1553%
27	WS1	20	0.0985%
28	Sumerlaeota	19	0.0600%
29	Bdellovibrionota	14	0.1079%
30	Hydrogenedentes	13	0.0281%
31	Entotheonellaeota	11	0.0357%
32	Fibrobacterota	11	0.0435%
33	Dadabacteria	10	0.1896%
34	Elusimicrobiota	8	0.0316%
35	TA06	7	0.0696%
36	Cyanobacteria	6	0.0226%
37	DTB120	6	0.0492%
38	LCP-89	6	0.0396%
39	Cloacimonadota	5	0.0970%
40	Calditrichota	4	0.0622%
41	WS4	4	0.0152%
42	SAR324 clade(Marine group B)	3	0.0062%
43	Abditibacteriota	2	0.0023%
44	Caldatribacteriota	2	0.0172%
45	Deferrisomatota	2	0.0097%
46	Fusobacteriota	2	0.0101%
47	GAL15	2	0.0084%
48	Schekmanbacteria	2	0.0076%
49	WPS-2	2	0.0027%
50	Deinococcota	1	0.0044%
51	Edwardsbacteria	1	0.0135%
52	FCPU426	1	0.0044%
53	Fermentibacterota	1	0.0070%
54	GN01	1	0.0216%
55	Modulibacteria	1	0.0044%

Supplementary Figures

Fig. S1

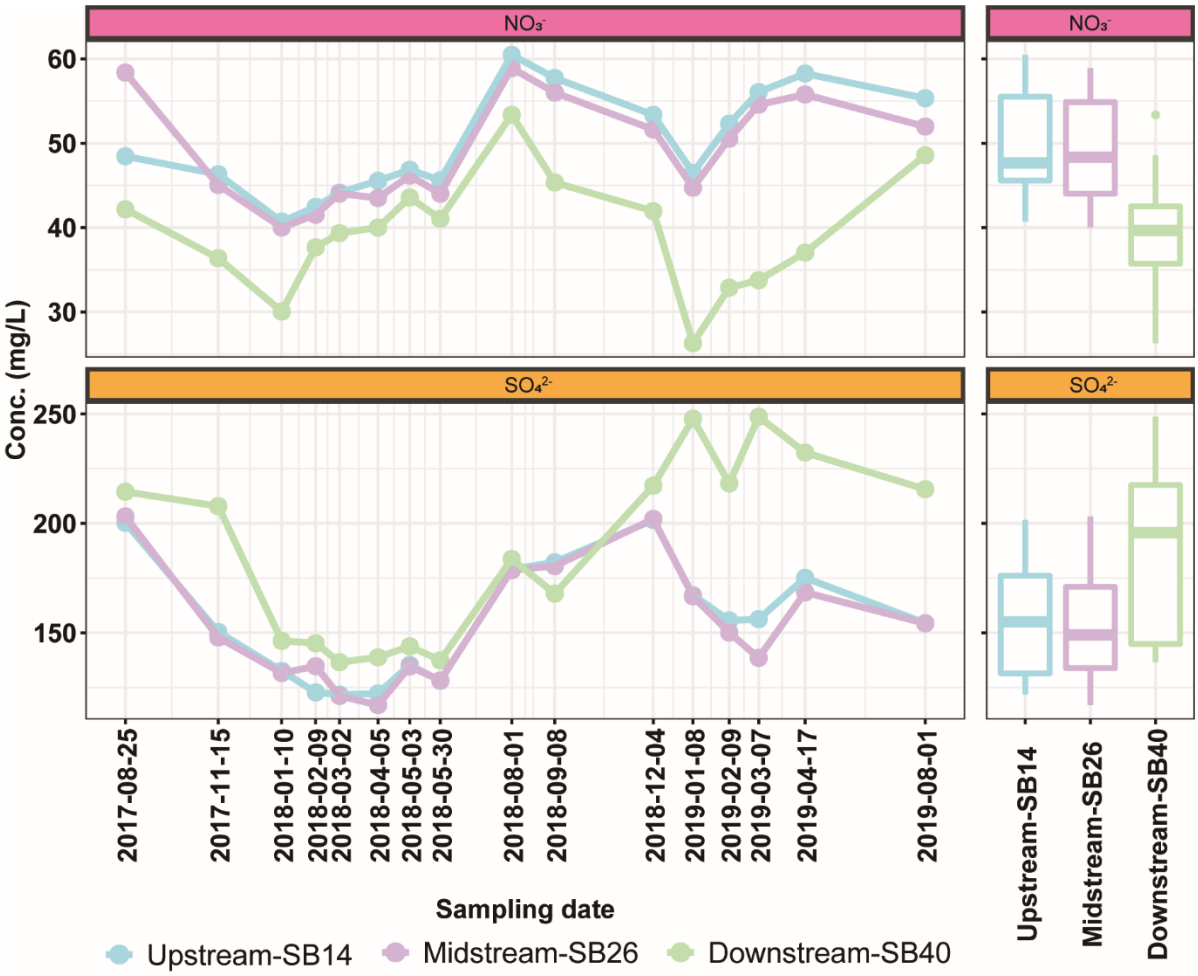

Fig. S1. Concentrations of nitrate and sulfate in the Schönbrunnen between the sampling years 2017 and 2019. Water samples were obtained from upstream to downstream. Box plots on the right side provide a summary of either nitrate or sulfate concentrations from three sampling locations between the sampling years 2017 and 2019.

Fig. S2

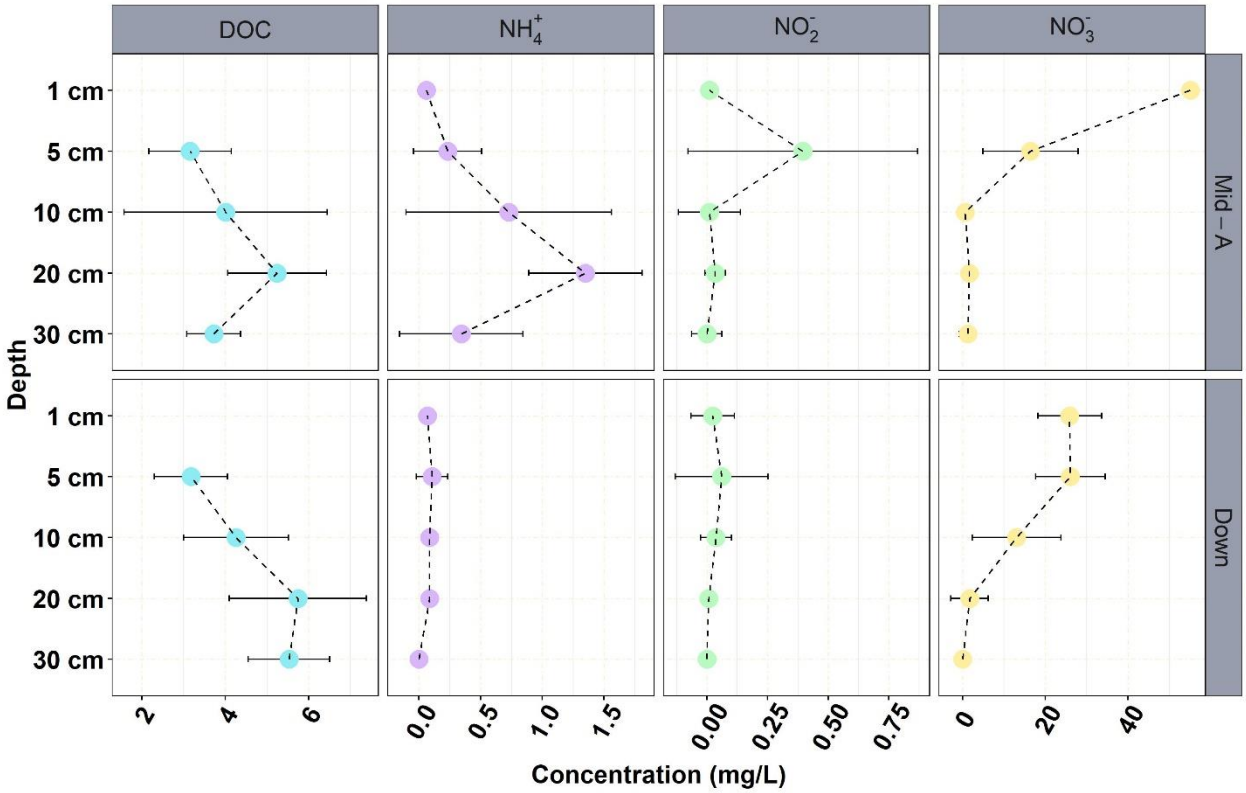

Fig. S2. Depth profiles of dissolved organic carbon (DOC) and nitrogen species from exemplary porewater samples collected from *Mid-A* and *Down* of the Schönbrunnen. Error bars indicate measurement variation (mean ± standard deviation) over the summer season (duplicate samples from 4 measurement dates between July 15 and August 15).

Fig. S3

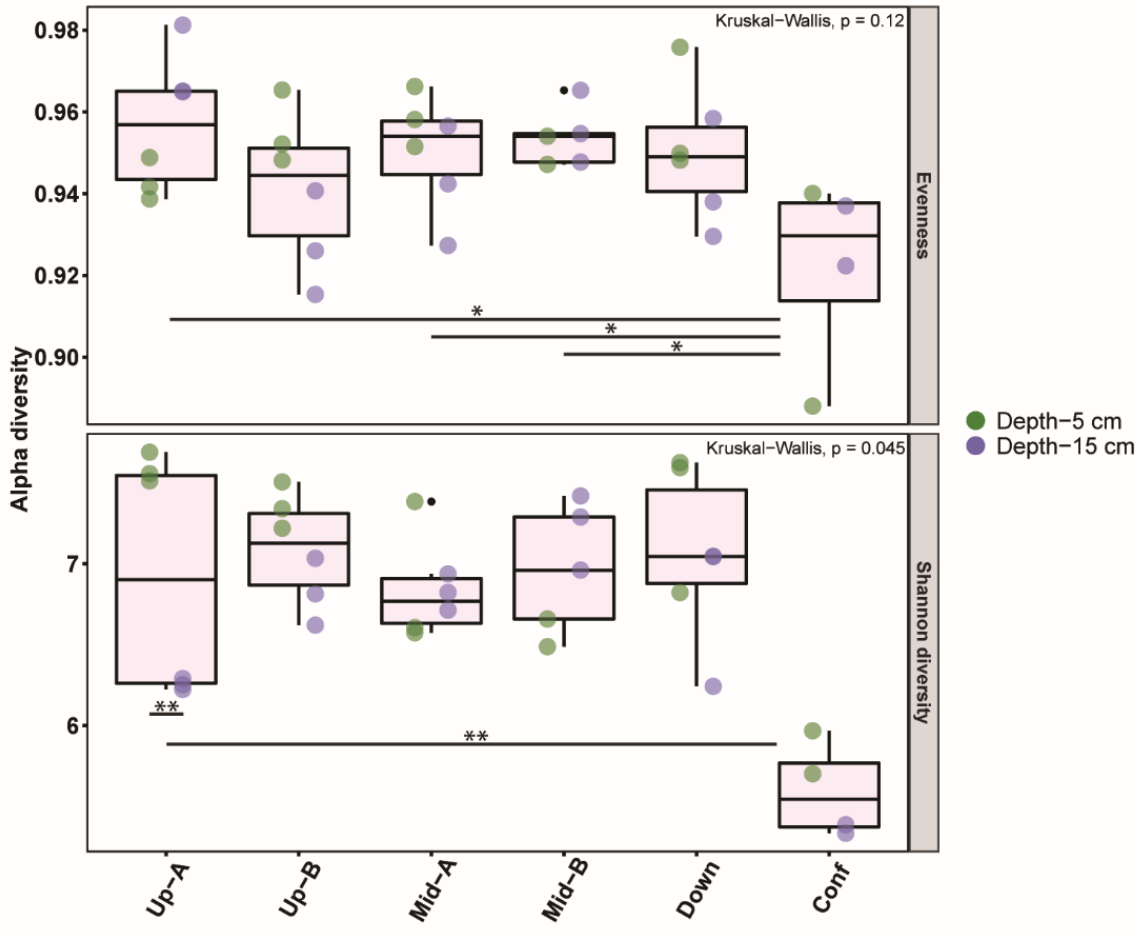

Fig. S3. Alpha diversity of sediment bacterial communities in the Schönbrunnen streambed. Shannon diversity and Shannon diversity based evenness for 16S rRNA gene sequencing data are plotted for each sampling depth and sampling location. Boxplots indicate the mean Shannon diversity and evenness at each sampling location. Asterisks indicate significant differences in Kruskal-Wallis tests with Dunn's Multiple Comparison post-tests (*: $p < 0.05$; **: $p < 0.01$).

Fig. S4

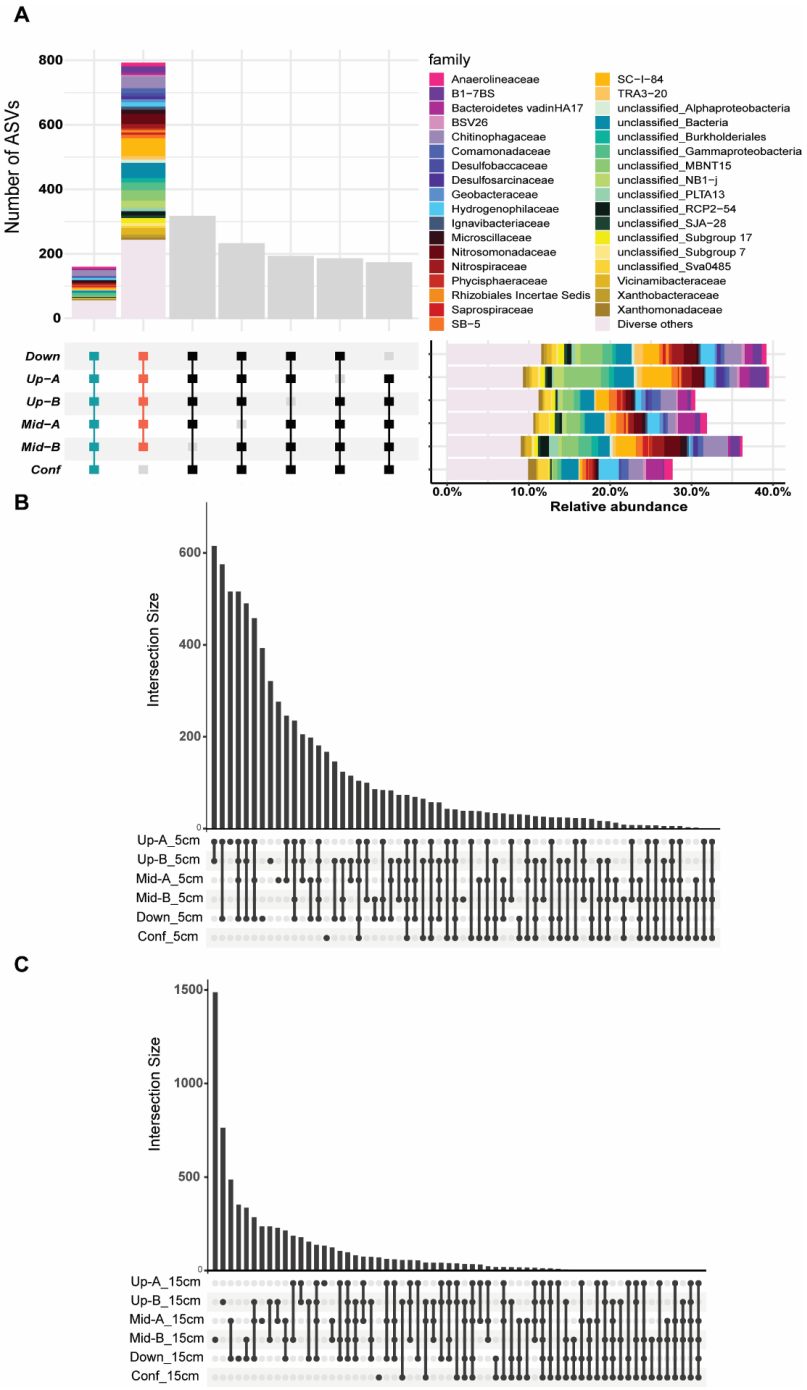

Fig. S4. (A) Upset plot on ASVs shared by at least five sampling locations. In total, in total, 1083 ASVs in 167 families were plotted here. However, only the top 35 families, which have most of the shared ASVs in at least five sampling locations, were plotted with different color codes. Other ASVs presented in at least five sampling locations were merged into “Diverse others”. The bottom bar plot showed the relative abundance of these 1083 ASVs in all ASVs of a sampling location. (B) and (C) upset plots show the number of ASVs shared by 5 cm samples and 15 cm samples, respectively.

Supplementary References

Ginestet, C., 2011. ggplot2: Elegant Graphics for Data Analysis. J. R. Stat. Soc. Ser. A (Statistics Soc. 174. https://doi.org/10.1111/j.1467-985x.2010.00676_9.x

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32. https://doi.org/10.1093/bioinformatics/btw313

Henry, S., Baudoin, E., López-Gutiérrez, J.C., Martin-Laurent, F., Brauman, A., Philippot, L., 2004. Quantification of denitrifying bacteria in soils by nirK gene targeted real-time PCR. J. Microbiol. Methods 59. https://doi.org/10.1016/j.mimet.2004.07.002

Kassambara, A., 2020. ggpubr: "ggplot2" Based Publication Ready Plots. R Packag. version 0.4.0.

Krassowski, M., 2021. krassowski/complex-upset: v1.0.2. https://doi.org/10.5281/ZENODO.4416064

Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 47. https://doi.org/10.1093/nar/gkz239

Li, C., 2020. GitHub - cxli233/customized_upset_plots [WWW Document]. https://doi.org/https://github.com/cxli233/customized_upset_plots

McMurdie, P.J., Holmes, S., 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8. https://doi.org/10.1371/journal.pone.0061217

Schwieger, F., Tebbe, C.C., 1998. A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. Appl. Environ. Microbiol. 64. https://doi.org/10.1128/aem.64.12.4870-4876.1998

Signorelli, A., 2020. DescTools: Tools for Descriptive Statistics. R package version 0.99.37. Cran.

Throbäck, I.N., Enwall, K., Jarvis, Å., Hallin, S., 2004. Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiol. Ecol. 49. https://doi.org/10.1016/j.femsec.2004.04.011

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4. https://doi.org/10.21105/joss.01686