

Bacterial community structure in a sympagic habitat expanding with global warming: brackish ice brine at 85–90 °N by Fernández-Gómez et al. (The ISME Journal 2018)

Supplementary material and methods

Analyses of basic sample characteristics

For determination of dissolved inorganic nutrient concentrations (NO_3^- , NO_2^- , PO_4^{3-} , and SiO_2), 12 mL of water were filtered through a 0.45 μm pore membrane filter and frozen at -20°C for later analysis at Stockholm University (Sweden). The NH_4^+ concentration was measured immediately after sampling onboard the icebreaker with a Perkin Elmer LS 55 fluorometer according to the method of Holmes *et al.* (1999). For the determination of C, N and P in particulate matter and chlorophyll *a* (Chla), 500–2000 mL of water were filtered on pre-combusted Whatman® GF/F glass fiber filters for later analysis at Stockholm University (Sweden). All samples were filtered and stored at -80°C within three hours after field sampling. Particulate C and N were measured with a CHN-900 analyzer (Leco, USA). The samples for particulate P analysis were combusted for 2 h at 500°C in 15 mL glass tubes and digested in 6 mL persulfate solution ($50\text{ g K}_2\text{S}_2\text{O}_8\text{ L}^{-1} + 30\text{ mL H}_2\text{SO}_4\text{ L}^{-1}$ for 1 h at 120°C). After addition of 4 mL 1.6 M NaOH, the samples were analysed for molybdate-reactive orthophosphate with flow injection (Lachat Instruments, slightly modified Quick-Chem Method 31-115-01-3-A). Since inorganic fractions are negligible in these samples, C, N and P in particulate matter are equivalent to particulate organic fractions (POC, PON and POP, respectively). Chlorophyll *a* was measured with high performance liquid chromatography according to Andersson *et al.* (2003). For the determination of bacterial cell abundance, 1.35 mL of water was fixed with glutaraldehyde (0.1% final concentration) and stored at -80°C for later analysis with a FACSCalibur flow cytometer (Becton-Dickinson) at the University of Concepción (Chile).

DNA extraction

DNA was extracted using a modification of a phenol:chloroform:IAA protocol. Before extraction, the filters were placed in 2 mL Lysing Matrix E tubes (Qbiogene Inc., Carlsbad, CA, USA) containing sterile glass beads (1 mm) and 500 μL of Buffer X (0.63 M of xantogenase (Sigma-Aldrich), 0.8 M ammonium acetate, 0.1 M Tris-HCl 1M pH=7.4, 0.0162 M EDTA pH=8) and cells were homogenized by bead beating (4.0 ms^{-1} for 30 s; 3 times and 1 min on ice in between). The lysate was mixed with 50 μL of 10% SDS and gently agitated before incubation at 65°C for 2 h and placed on ice for 30 min. The debris was removed by centrifugation for 10 min at 13 000 rpm (4°C). Concentration and purity of DNA was verified with a Nanodrop® ND-1000 spectrophotometer. DNA integrity was confirmed on a 1% agarose gel.

High-throughput sequencing protocol

The 16S rRNA gene was amplified using the universal bacterial primers 358Fgc and 907R (Muyzer *et al.*, 1995). Each PCR reaction contained 1x PCR reaction buffer (Invitrogen), 2 mM MgCl_2 , dNTPs 0.2 mM, primers 0.48 μM , DMSO 3%, Taq DNA polymerase 1U (Invitrogen), and 1 μL of DNA in a final reaction volume of 25 μL . The PCR program consisted of 1 min denaturation at 95°C , 1 min annealing at 55°C , 1 min extension at 72°C , and a 10 min final extension step at 72°C for a total 30 amplification cycles. The hypervariable region 4 (V4) of the 16S rRNA was amplified using a high-fidelity polymerase (Phusion High-Fidelity DNA Polymerase, New England BioLab) and primers PE_16S_V4_515F (5'-YR YRGTGCCAGCMGCCGCGGTAA-3') and PE_16S_V4_786R (5'-GGACTACHVGGGTWTCTAAT-3'), which amplify approximately 270 bp. Amplification was verified on a 1% agarose gel.

Sequencing was carried out on a MiSeq® Illumina sequencer at the Massachusetts Institute of Technology (USA). Two IB samples, IB12 at the North Pole and IB23 at the marginal ice zone were deep-sequenced. The DNA concentration for each sample was determined by qPCR

(LightCycler® 480 Real time, Roche). Each reaction contained 1x HF Buffer (New England), dNTPs 0.2 mM, PE16S_V4_U515F and PE16S_V4_E786R 0.3 μ M, DMSO 3%, 0.5x SybrGreen (Life Technologies), Phusion DNA Polymerase 2.5U (New England), 5 μ L of DNA, in a 25 μ L reaction. The amplification conditions were denaturation at 98 °C for 30 s, annealing at 52 °C for 30 s, extension at 72 °C for 30 s, and a second final extension at 72 °C for 5 min, for a total of 45 cycles of amplification. Samples were normalized according to the less concentrated one, according with Ct (Ct = 20).

Another qPCR was performed with the normalized DNA, using the same conditions as the previous one, but in this case using only 20 cycles of amplification, according to the normalization. Purification of the PCR products was performed with magnetic beads SPRI. Barcodes were added to the PCR products and the following PCR was performed in quadruplicate to obtain enough amplicons for Illumina sequencing. Each PCR reaction contained 1x HF Buffer (New England), dNTPs 0.2 mM, PE16S_IV_F and PE16S_IV_XXX 0.3 μ M, Phusion DNA Polymerase 2.5U (New England), 4 μ L of DNA, in a 25 μ L reaction. The PCR program consisted of initial denaturation at 98 °C for 30 s, annealing at 70 °C for 9 s, extension at 72 °C for 30 s, and a second final extension at 72 °C for 2 min, for a total of 9 cycles of amplification. Purification of PCR products was performed with magnetic beads SPRI. In the final step, all samples were mixed in a single Eppendorf tube according to the last qPCR, ensuring the same DNA concentration in the Illumina plate. This concentration was calculated with a final qPCR that consisted on 1x Master mix Quantitec (Qiagen), PE_seq_F and PE_seq_R 0.2 μ M, 5 μ L of DNA, in a 25 μ L reaction. The PCR program consisted of denaturation at 95 °C for 10 s, annealing at 60 °C for 20 s, extension at 72 °C for 30 s, and a second final extension at 72 °C for 5 min, for 45 cycles of amplification.

Phylogenetic analyses

To estimate Faith's phylogenetic diversity (PD) (Faith, 1992) and the standardized effect sizes of the mean pairwise distance (MPD) (Webb *et al.*, 2003) we reconstructed a phylogenetic tree using the 2715 consensus 16S rRNA sequences of all the identified OTUs. The sequence alignment was performed with MAFFT software version 7 (Katoh & Standley, 2013), strategy G-INS-i and default parameters. We used a bacterial 16S rRNA model of nucleotide substitution and a maximum likelihood approach implemented in 'Treefinder' (Jobb *et al.*, 2004). The reconstructed topology was used to estimate PD and MPD with the package 'Picante' (Kembel *et al.*, 2010) in the software environment 'R' (R Development Core Team, 2014).

For further reconstruction of phylogenetic trees only OTUs with RA \geq 0.01% across all samples were used. This arbitrary percentage enabled use of a relatively high number of OTUs while at the same time obtaining a good visualization of the resulting tree. Of the 180 OTUs with RA \geq 0.01%, 161 belonging to the three dominant phyla, Proteobacteria, Bacteroidetes, and Actinobacteria, were used. The remaining 19 OTUs with RA \geq 0.01% belonged to nine other phyla and they were not included in the analyses. To avoid losing information due to the less sequenced sites and to take advantage of the information provided by the ones most deeply sequenced, we used non-rarefied data normalized to the total number of reads for each sample. The OTUs were aligned with known 16S rRNA downloaded from SILVA and one Archaeal 16S rRNA sequence as root. Alignment was performed with MAFFT (v.7.215) using the algorithm FFT-NS-2 (Katoh & Standley, 2013) and gaps were removed using UGENE (v.1.22.0) (Okonechnikov *et al.*, 2012). Phylogenies were constructed using maximum likelihood implemented in 'FastTree' (v.2.1.7) SSE3, OpenMP (Price *et al.*, 2010) with GTR nucleotide substitution model. The generated trees were visualized and edited in the Interactive Tree Of Life (Letunic & Bork, 2011).

Reconstruction of probabilistic graphical models (PGMs)

To explore the structure of the sympagic bacterial metacommunity (defined as a set of local communities linked by dispersal of multiple potentially interacting species (Gilpin & Hanski,

1991; Leibold *et al.*, 2004) in the CAO, the machine learning approach implemented with the ‘huge’ package (Zhao *et al.*, 2012) in the software environment ‘R’ (R Development Core Team, 2014) was used. Probabilistic Graphical Models (PGMs) were reconstructed for two different distribution abundance thresholds: all 2715 OTUs and the 180 OTUs with RA $\geq 0.01\%$, expecting to obtain different patterns of modularity.

For the network analysis we used rarefied OTU abundances as input for the machine learning approach implemented with the ‘huge’ package (Zhao *et al.*, 2012) in the software environment ‘R’ (R Development Core Team, 2014) was used. The algorithm, based on the maximum entropy principle, estimates the correlation matrix and then its inverse (Lezon *et al.*, 2006; Stein *et al.*, 2015). An element of the inverse matrix corresponds to an edge of the network and a zero element implies that the OTU is conditionally independent. We used the algorithm ‘glasso’ to estimate a set of matrices with different levels of sparsity (number of zero elements in the matrix). The algorithm ‘stars’ allowed us to select the matrix with higher stability. Each non-zero element of the inverse correlation matrix represented conditional dependence. We estimated basic network properties, including the degree distribution, the clustering coefficient (defined as $C_n = 2e_n/(k_n(k_n-1))$ where k_n is the number of neighbours of n and e_n is the number of connected pairs between all neighbours of n), network centralization (showing whether a network has a star-like topology or whether the nodes have on average the same connectivity), characteristic path length (expected average distance between two connected nodes), average number of neighbours, network density (number of edges divided for all potential edges, reflecting connectance), network heterogeneity (tendency of a network to contain hub nodes), and modularity, a measure showing which groups of nodes are more connected among themselves than with other nodes in the network (Pavlopoulos *et al.*, 2011). All parameters were estimated in Cytoscape (Shannon *et al.*, 2003), except modularity which was estimated in R as implemented in the Walktrap community finding algorithm (Pons & Latapy, 2006).

References

- Andersson M, Van Nieuwerburgh L, Snoeijs P (2003). Pigment transfer from phytoplankton to zooplankton with emphasis on astaxanthin production in the Baltic Sea food web. *Marine Ecology Progress Series* 254:213–224.
- Faith DP (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61:1–10.
- Gilpin ME, Hanski IA (1991). *Metapopulation Dynamics: Empirical and Theoretical Investigations*. Academic Press, London.
- Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* 4:18. *BMC Evolutionary Biology* 15:243.
- Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.
- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464.
- Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7:601–613.
- Letunic I, Bork P (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 39:W475–W478.
- Lezon TR, Banavar JR, Cieplak M, Maritan A, Fedoroff NV (2006). Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proceedings of the National Academy of Sciences of the USA* 103:19033–19038.
- Muyzer G, Teske A, Wirsen CO, Jannasch HW (1995). Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by

- denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology* 164:165–172.
- Okonechnikov K, Golosova O, Fursov M. Unipro (2012). UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167.
- Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, *et al.* (2011). Using graph theory to analyze biological networks. *BioData Mining* 4:10.
- Pons P, Latapy M (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications* 10:191–218.
- Price MN, Dehal PS, Arkin AP (2010). FastTree 2 approximately maximum likelihood trees for large alignments. *PLoS ONE* 5(3):e9490.
- R Development Core Team (2014). R: A Language and Environment for Statistical Computing Vienna, Austria: the R Foundation for Statistical Computing. ISBN: 3-900051-07-0, Available online at [http://wwwR-project.org/](http://www.R-project.org/)
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504.
- Stein RR, Marks DS, Sander C (2015). Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Computational Biology* 11(7):e1004182.
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ (2003). Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Zhao T, Liu H, Roeder K, Lafferty J, Wasserman L (2012). The huge package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research* 13:1059–1062.