



Data Article

16S rRNA gene amplicon dataset of prokaryotic communities from a subantarctic marine ecosystem: Ushuaia Bay and surrounding waters

Clara Natalia Rodríguez-Flórez^a, Andrea Malits^a, Mariana Lozada^{b,*}

^a Laboratorio de Oceanografía Biológica (CADIC-CONICET), Ushuaia, Argentina

^b Laboratorio de Microbiología Ambiental (IBIOMAR-CONICET), Puerto Madryn, Argentina

ARTICLE INFO

Article history:

Received 19 June 2020

Revised 27 July 2020

Accepted 6 August 2020

Available online 12 August 2020

Keywords:

Bacteria

Archaea

Seawater

Subantarctic

16S rRNA

Gene amplicon sequencing

ABSTRACT

We provide a 16S rRNA gene dataset of prokaryotic assemblages of a subantarctic marine ecosystem. Samples were collected at 2 stations (one near Ushuaia Bay and the other close to Bridges islands in the Beagle Channel). At each station, 2 depths (subsurface and bottom waters) were sampled in february, march, may and september during 2018, giving a total of 10 samples. 16S rRNA gene amplicon sequencing (V4 hypervariable region) was performed with the Illumina MiSeq platform. A total of 1116 amplicon sequence variants (ASVs) were recovered from the dataset. The sequences were taxonomically assigned to Alphaproteobacteria ($23 \pm 2\%$, mean \pm standard error), Gammaproteobacteria ($17 \pm 1.5\%$), Flavobacteriia ($8 \pm 2\%$), Deltaproteobacteria ($3.7 \pm 0.5\%$), Acidimicrobiia ($1.7 \pm 0.1\%$), Planctomycetia ($1.9 \pm 0.4\%$), and AB16 group ($1.7 \pm 0.3\%$). Sequences affiliated with Archaea were abundant, reaching one third of analyzed sequences, mainly Thaumarchaeota ($22 \pm 3\%$), and Thermoplasmata ($10 \pm 1\%$). Together, sequences assigned to all these groups accounted for more than 90% of the sequences. This dataset constitutes a valuable resource for future scientific research aiming to unveil the role of these communities in ecosystem services such as carbon and nutrient cycling, and pollutants degradation. This will turn into benefits for future

* Corresponding author.

E-mail address: lozada@cenpat-conicet.gob.ar (M. Lozada).

environmental monitoring and preservation actions, considering the tangible heritage of Ushuaia Bay and surrounding waters.

© 2020 Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Environmental Science (General)
Specific subject area	Marine microbial diversity and ecology
Type of data	fastq file
How data were acquired	16S rRNA gene amplicon sequencing Instruments: Illumina MiSeq platform, QIIME2 software
Data format	Raw Filtered Analyzed
Parameters for data collection	Conditions considered for data collection were: seawater samples from subsurface and bottom, at two stations: one in Ushuaia Bay and the other in surrounding waters. <i>Subsurface waters</i> : -20m. <i>Bottom waters</i> : ~ -120m (Ushuaia Bay station) and ~-148m (surrounding waters station).
Description of data collection	Seawater samples were collected from a boat in Niskin bottles, prefiltered and passed through a Sterivex filter (0.2µm). Metagenomic DNA extraction was performed from filters, and after PCR amplification of 16S rRNA gene V4 hypervariable region, amplicon sequencing was done on Illumina MiSeq platform. Preprocessing and taxonomic profiling was performed in QIIME2.
Data source location	City/Region: Ushuaia/Tierra del Fuego Country: Argentina Latitude and longitude: <i>Ushuaia bay station</i> : 54° 48.728S, 68° 14.388 W, close to Ushuaia city (Tierra del Fuego, Argentina). <i>Surrounding waters station</i> : 54° 52.939S, 68° 10.927W, close to Bridges islands and near kelp forest, in the Beagle Channel. Maritime traffic: the sampling stations are located along a sea route. Possible environmental impact. High nutrient and organic matter load to Ushuaia Bay through glacial runoff and urban and industrial activities.
Data accessibility	The sequence data from this work is available at the NCBI Genbank Sequence Read Archive (SRA) as <i>Bioproject ID</i> : PRJNA622742 and <i>SRA accession no.</i> PRJNA637656. Repository name: Genbank Sequence Read Archive Data identification number: PRJNA637656 (Bioproject) SRR11941686-SRR11941695 (Runs) Direct URL to data: Bioproject: https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA637656&o=acc_s%3Aa Runs: https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941686 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941687 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941688 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941689 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941690 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941691 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941692 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941693 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941694 https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR11941695

Value of the Data

- This is the first checklist of amplicon sequence variants and their taxonomic classification from Ushuaia Bay and surrounding waters in the Beagle Channel, a subantarctic marine environment exposed to anthropogenic impact and suffering rapid shifts due to climate change.
- This first 16S rRNA gene profiling of prokaryotic assemblages of a subantarctic marine ecosystem is a valuable resource for the scientific community as well as for institutions performing environmental monitoring and preservation actions.
- This dataset is useful for upcoming scientific research covering the role of these communities in ecosystem services.

1. Data description

The raw sequencing dataset contained 550,960 sequences from a total of 10 samples. After preprocessing, the dataset contained 302,986 sequences, 150 bp on average. **A total of 1116 amplicon sequence variants (ASVs) were recovered from this environment.** Of the total analyzed sequences, **68 ± 3 % corresponded to Bacteria** (mean ± standard error) and **32 ± 3 % to Archaea.** **The major phyla** (out of 24, B_: Bacteria and A_: Archaea) **were:** B_*Proteobacteria* (45 ± 2 %), A_*Crenarchaeota* (22 ± 3%), A_*Euryarchaeota* 10 ± 1%, B_*Bacteroidetes* 9 ± 2% and B_*Planctomycetes* 4 ± 0.6%. **The most abundant classes (out of 40) were:** B_*Alphaproteobacteria* 23 ± 2%, A_*Thaumarchaeota* 2 ± 3%, B_*Gammaproteobacteria* 17 ± 1%, A_*Thermoplasmata* 10 ± 1%, B_*Flavobacteriia* 8 ± 2%, B_*Deltaproteobacteria* 4 ± 0.5%, B_*Acidimicrobiia* 2 ± 0.1%, B_*Planctomycetia* 2 ± 0.4%, B_*AB16* 1.71 ± 0.2%. **The most representative orders (out of 60) were:** A_*Cenarchaeales* 22 ± 3%, B_*Oceanospirillales* 10 ± 0.7%, A_*E2* 10 ± 1%, B_*Rhodobacteriales* 9 ± 2%, B_*Rickettsiales* 9 ± 1%, and B_*Flavobacteriales* 8 ± 2%. **The major families (out of 78) were:** A_*Cenarchaeaceae* 22 ± 3%, A_*Marine group II* 9 ± 1%, B_*Rhodobacteraceae* 9 ± 2%, B_*Pelagibacteraceae* 9 ± 1%, B_*Flavobacteriaceae* 5 ± 1%, B_*Halomonadaceae* 4 ± 0.6%, B_*Alteromonadaceae* 1.8 ± 0.5% and B_*Nitrospinaceae* 1.8 ± 0.3% (Fig. 1).

The major genera (out of 82) were: A_*Nitrosopumilus* 22 ± 3%, B_*Pelagibacter* 7 ± 1%, B_*Candidatus Portiera* 4 ± 0.6%, B_*Sulfitobacter* 4 ± 1%, 1.7 ± 0.3%, B_*Colwellia* 1.4 ± 0.7%, B_*Polaribacter* 1.3 ± 0.4% and B_*Glaciecola* 0.9 ± 0.5%.

Supplementary File 1. Bioinformatic script used to preprocess the raw sequences. QIIME2 v 2019.1 (<https://qiime2.org/>) was used to process the data.

2. Experimental design, materials and methods

Seawater was collected in Niskin bottles (5L) and prefiltered by a mesh of 213 µm after which each water sample (~4L) was concentrated in a Sterivex filter unit of 0.22 µm (Millipore). The Sterivex units were stored without excess liquid into sterile sampling bags (Microclar) at -20°C, until cell lysis and nucleic acid extraction. The sampling took place during February (2 samples), March (1), May (4) and September (3) 2018.

Cell lysis and nucleic acid extraction were carried out following the protocol modified from Somerville and collaborators [1]. DNA quantification was done using Lambda Phage DNA (Promega). Samples were sequenced at INDEAR (Argentina). Quantification of initial DNA samples was performed using the Quant-iT™ PicoGreen® DNA Assay Kit (Invitrogen). Amplicons of V4 region were obtained by using primers 515F (GTGCCAGCMGCCGCGTA)[2] and 806R (GGACTACNVGGGTWTCTAAT)[3]. A second round of PCR was performed with standard Illumina barcodes and adapters. Libraries were pooled at the same concentration and 1ul of the pool was run in the 2100 Bioanalyzer (Agilent Technologies) using the DNA 12000 chip Kit. The library pool was quantified using DeNovix. Libraries were sequenced on an Illumina MiSeq platform.

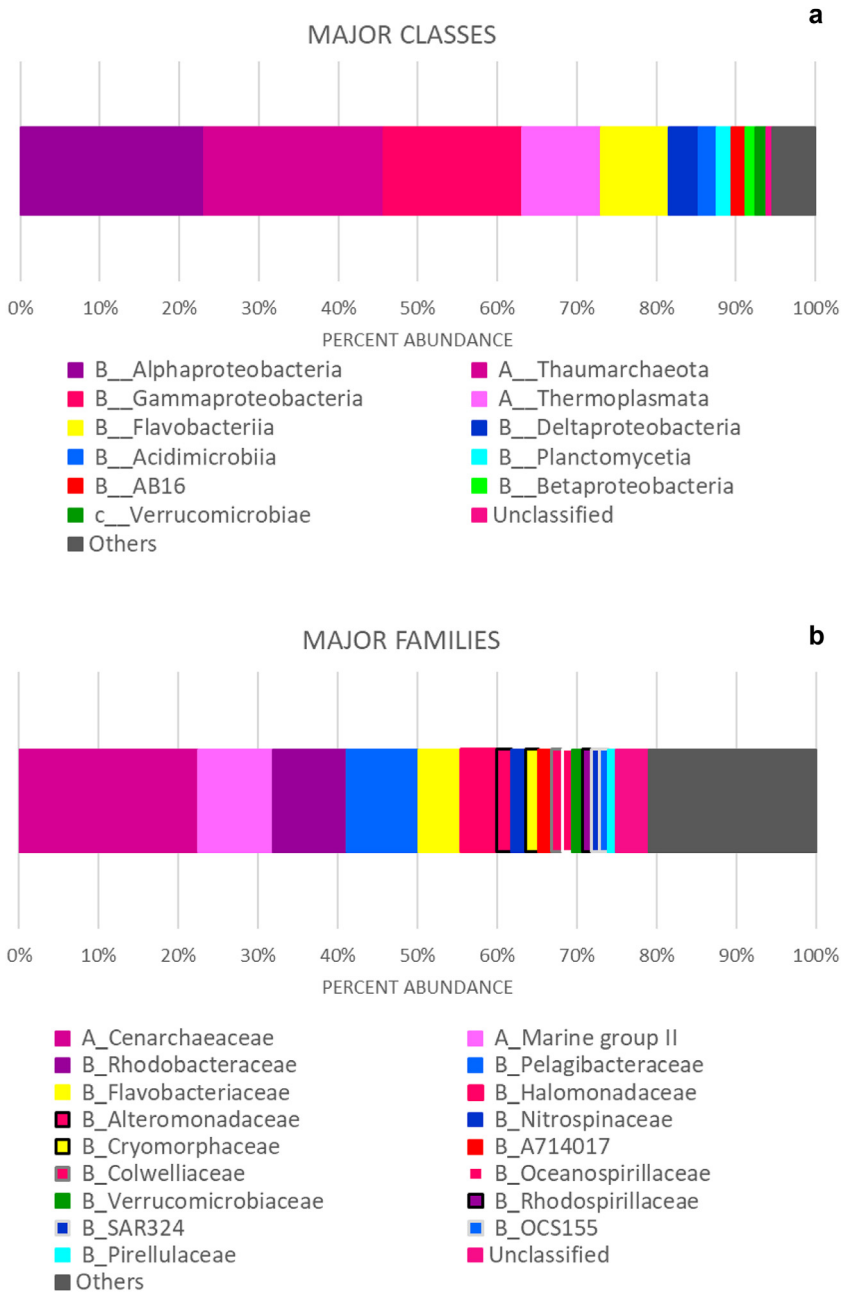


Fig. 1. Percent abundance of major (a) classes and (b) families, identified in prokaryotic communities from Ushuaia Bay and surrounding waters in the Beagle Channel. Only classes and families with more than 1% relative abundance are shown. A_: Archaea. B_: Bacteria.

The generated reads were preprocessed with the QIIME2 package [4]. Briefly, demultiplexed paired-end reads were trimmed to 150 bp, merged and the resulting sequences were denoised in Deblur [5] in QIIME2 environment, in order to identify amplicon sequence variants (ASV). ASVs were taxonomically classified in QIIME2 using Greengenes classifier (<https://data.qiime2.org/2019.1/common/gg-13-8-99-515-806-nb-classifier.qza>). The resulting feature table and taxonomic assignments were analyzed in QIIME2 and phyloseq [6].

Supplementary File 1: Bioinformatic script used in QIIME2.

Ethics statement

The work did not involve the use of human subjects or animal experiments.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

AM and ML are members of the Argentinean National Research Council (CONICET). CNRF is a doctoral fellow (CONICET program). This work was supported by ANPCyT (Grants PICT-2015-0384 and PICT-2018-0903).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2020.106171](https://doi.org/10.1016/j.dib.2020.106171).

References

- [1] C.C. Somerville, I.T. Knight, W.L. Straube, R.R. Colwell. Simple, rapid method for direct isolation of nucleic acids from aquatic environments, *Appl. Environ. Microbiol.* 55 (1989) 548–554.
- [2] A.E. Parada, D.M. Needham, J.A. Fuhrman, Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples, *Environ. Microbiol.* 18 (2016) 1403–1414, doi:[10.1111/1462-2920.13023](https://doi.org/10.1111/1462-2920.13023).
- [3] A. Apprill, S. McNally, R. Parsons, L. Weber, Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton, *Aquat. Microb. Ecol.* 75 (2015) 129–137, doi:[10.3354/ame01753](https://doi.org/10.3354/ame01753).
- [4] E. Bolyen, J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, G.A. Marotz, B.D. Martin, D. McDonald, L.J. McIver, A.V. Melnik, J.L. Metcalf, S.C. Morgan, J.T. Morton, A.T. Naimy, J.A. Navas-Molina, L.F. Nothias, S.B. Orchanian, T. Pearson, S.L. Peoples, D. Petras, M.L. Preuss, E. Pruesse, L.B. Rasmussen, A. Rivers, M.S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S.J. Song, J.R. Spear, A.D. Swafford, L.R. Thompson, P.J. Torres, P. Trinh, A. Tripathi, P.J. Turnbaugh, S. Ul-Hasan, J.J.J. van der Hooft, F. Vargas, Y. Vázquez-Ghalith, H. Alexander, E.J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J.E. Bisanz, K. Bittinger, A. Brejnrod, C.J. Brislawn, C.T. Brown, B.J. Callahan, A.M. Caraballo-Rodríguez, J. Chase, E.K. Cope, R.D. Silva, C. Diener, P.C. Dorrestein, G.M. Douglas, D.M. Durall, C. Duvallet, C.F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J.M. Gauglitz, S.M. Gibbons, D.L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G.A. Huttley, S. Janssen, A.K. Jarmusch, L. Jiang, B.D. Kaehler, K.B. Kang, C.R. Keefe, P. Keim, S.T. Kelley, D. Knights, I. Koester, T. Kosciolk, J. Kreps, M.G.I. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Löffel, C. Lozupone, M. Maher, C. uez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K.C. Weber, C.H.D. Williamson, A.D. Willis, Z.Z. Xu, J.R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, J.G. Caporaso, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, *Nat. Biotechnol.* 37 (2019) 852–857, doi:[10.1038/s41587-019-0209-9](https://doi.org/10.1038/s41587-019-0209-9).
- [5] A. Amir, D. McDonald, J.A. Navas-Molina, E. Kopylova, J.T. Morton, Z.Z. Xu, E.P. Kightley, L.R. Thompson, E.R. Hyde, A. Gonzalez, R. Knight, Deblur rapidly resolves single-nucleotide community sequence patterns, *MSystems* 2 (2017), doi:[10.1128/mSystems.00191-16](https://doi.org/10.1128/mSystems.00191-16).
- [6] P.J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, *PLOS ONE* 8 (2013) e61217, doi:[10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217).