Spatially and Temporally Explicit Metagenomes and Metagenome-Assembled Genomes from the Comau Fjord (42°S), Patagonia

Eduardo Castro-Nallar,[a,b] Valentín Berríos-Farías,[a,b] Beatriz Díez,[c,d,e] Sergio Guajardo-Leiva[a,b]

[a]Departamento de Microbiología, Facultad de Ciencias de la Salud, Universidad de Talca, Talca, Chile
[b]Centro de Ecología Integrativa, Universidad de Talca, Talca, Chile
[c]Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Santiago, Chile
[d]Center for Climate and Resilience Research (CR)², Santiago, Chile
[e]Millennium Institute Center for Genome Regulation (CGR), Santiago, Chile

ABSTRACT Microbes play an important role in coastal and estuarine waters. We present 93 metagenomes and 677 metagenome-assembled genomes (MAGs) from Comau Fjord, Patagonia (42°S), to further understand the microbial dynamics and their response to anthropogenic disturbances. These data represent a spatially (35-km transect) and temporally (2016 to 2019) explicit data set.

Coastal and estuarine waters not only receive carbon and nutrients from rivers and other freshwater sources but also are a hot spot of disturbances of anthropogenic origin (1–5). Comau Fjord in northern Patagonia, Chile, is surrounded by national parks and privately protected land. However, the fjord is part of commercial and passenger transport routes, as well as open-cage aquaculture centers, which affect biodiversity by pouring hydrocarbons, antibiotics, and excess nutrients into the ecosystem (6–8). To date, no studies have addressed the extent and impact of these disturbances on the microbial communities of the Comau Fjord. The metagenomic data presented here will contribute to our understanding regarding the consequences of anthropogenic activities in the fjord.

We collected seawater over a period of 3 years at depths of 5 and 20 m, from the mouth to the end of the fjord (40 L per sample, using an electric water pump). The water was prefiltered using a 50-μm nylon mesh (Sefar) to avoid macroscopic organisms and debris, followed by filtering with 3-μm polycarbonate filters (Merck, Millipore) to collect eukaryotes and particle-attached prokaryotes. Free-living prokaryotes were collected in 0.22-μm polyethersulfone (PES) Sterivex filters (Merck, Millipore), and DNA from this fraction was obtained using phenol-chloroform extraction with the xanthogenate protocol (9–11). Sequencing libraries were constructed using Illumina TruSeq DNA kits and sequenced using Illumina instruments (see the Comau sample table at https://doi.org/10.6084/m9.figshare.21960866.v3).

We used default software parameters unless otherwise stated. We removed adapters (–detect_adapter_for_pe) and carried out filtering and trimming (-q 30 -l 100) using fastp 0.20.1 (12). We first estimated the genomic distances using Mash v2.3 (13) and checked whether the metagenomes formed clusters using the Hopkins statistic (>0.7). We then used agglomerative hierarchical clustering (the hclust function in the base R package), which resulted in 4 clusters (Table 1), as suggested by silhouette analysis (14). *De novo* metagenome coassembly was carried out on each of the four clusters using MEGAHIT v1.2.9 (–min-contig-len 1500) (15). To bin the resulting contigs (778,598), we used GroopM v2.0.0, CONCOCT v1.0.0, and MetaBAT v2.15 separately to then refine the bins using the binning refinement module in MetaWRAP v1.3 (-c 50 -x 10) (16–19). The

TABLE 1 Assembly statistics for metagenomes and MAGs

Cluster	No. of contigs	Total length (bp)	Minimum length (bp)	Maximum length (bp)	Mean length (bp)	No. of refined MAGs
C1	140,400	666,764,307	1,500	95,119	4,749	128
C2	387,383	1,742,624,422	1,500	1,141,308	4,498	266
C3	138,518	666,137,931	1,500	427,687	4,809	160
C4	112,297	581,840,304	1,500	564,936	5,181	123

metagenome-assembled genomes (MAGs) presented here are ≥50% complete and <10% contaminated (see MAG stats at https://doi.org/10.6084/m9.figshare.21960866.v3) (20). Taxonomic assignment was conducted using GTDB-tk 1.5.1 and database version R202 (21). Finally, to obtain the read abundances per MAG, we mapped the reads from each sample against the MAGs using CoverM v0.6.1 (–dereplicate –dereplication-ani 99 –min-read-aligned-percent 30 -m trimmed_mean) (total 517 dereplicated genomes) (https://github.com/wwood/CoverM).

At the phylum level, the most abundant MAGs belonged to *Proteobacteria* (49%), *Bacteroidota* (31%), *Actinobacteriota* (7%), *Cyanobacteria* (4%), *Verrucomicrobiota* (2%), *Planctomycetota* (1.4%), and *Thermoplasmatota* (1.2%). Likewise, at the family level, the most abundant MAGs belonged to *Flavobacteriaceae* (26%), *Rhodobacteraceae* (20%), *Actinomarinaceae* (5%), *Cyanobiaceae* (4%), *Thioglobaceae* (4%), D2472 (3.7%), *Porticoccaceae* (3.2%), *Methylophilaceae* (2.8%), *Pseudohongiellaceae* (2.3%), *Schleiferiaceae* (2.2%), *Halieaceae* (1.8%), HTCC2089 (1.4%), *Akkermansiaceae* (1.3%), TMED25 (1%), and SAR86 (1%).

Data availability. This whole-metagenome shotgun project has been deposited at GenBank under the BioProject accession no. PRJNA729490. The version described in this paper is the first version.

ACKNOWLEDGMENTS

REFERENCES

1. Zhu Y-G, Zhao Y, Li B, Huang C-L, Zhang S-Y, Yu S, Chen Y-S, Zhang T, Gillings MR, Su J-Q. 2017. Continental-scale pollution of estuaries with antibiotic resistance genes. Nat Microbiol 2:16270. https://doi.org/10.1038/nmicrobiol.2016.270.

2. Brakstad OG, Davies EJ, Ribicic D, Winkler A, Brönner U, Netzer R. 2018. Biodegradation of dispersed oil in natural seawaters from western Greenland, and a Norwegian fjord. Polar Biol 41:2435–2450. https://doi.org/10.1007/s00300-018-2380-8.

3. Ortiz P, Quiroga E, Montero P, Hamame M, Betti F. 2021. Trophic structure of benthic communities in a Chilean fjord (45°S) influenced by salmon aquaculture: insights from stable isotopic signatures. Mar Pollut Bull 173:113149. https://doi.org/10.1016/j.marpolbul.2021.113149.

4. Zhang T, Li J, Wang N, Wang H, Yu L. 2022. Metagenomic analysis reveals microbiome and resistome in the seawater and sediments of Kongsfjorden (Svalbard, High Arctic). Sci Total Environ 809:151937. https://doi.org/10.1016/j.scitotenv.2021.151937.

5. Perez-Venegas DJ, Toro-Valdivieso C, Ayala F, Brito B, Iturra L, Arriagada M, Seguel M, Barrios C, Sepúlveda M, Oliva D, Cárdenas-Alayza S, Urbina MA, Jorquera A, Castro-Nallar E, Galbán-Malagón C. 2020. Monitoring the occurrence of microplastic ingestion in otariids along the Peruvian and Chilean coasts. Mar Pollut Bull 153:110966. https://doi.org/10.1016/j.marpolbul.2020.110966.

6. Mayr C, Rebolledo L, Schulte K, Schuster A, Zolitschka B, Försterra G, Häussermann V. 2014. Responses of nitrogen and carbon deposition rates in Comau Fjord (42°S, southern Chile) to natural and anthropogenic impacts during the last century. Cont Shelf Res 78:29–38. https://doi.org/10.1016/j.csr.2014.02.004.

7. Häussermann V, Försterra G, Melzer RR, Meyer R. 2013. Gradual changes of benthic biodiversity in Comau Fjord, Chilean Patagonia—lateral observations over a decade of taxonomic research. Spixiana 36:161–171.

8. Urbina M, Cumillaf J, Paschke K, Gebauer P. 2019. Effects of pharmaceuticals used to treat salmon lice on non-target species: evidence from a systematic review. Sci Total Environ 649:1124–1136. https://doi.org/10.1016/j.scitotenv.2018.08.334.

9. Tillett D, Neilan BA. 2000. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. J Phycol 36:251–258. https://doi.org/10.1046/j.1529-8817.2000.99079.x.

10. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Tara Oceans Coordinators. 2015. Structure and function of the global ocean microbiome. Science 348:1261359. https://doi.org/10.1126/science.1261359.

11. Guajardo-Leiva S, Mendez KN, Meneses C, Díez B, Castro-Nallar E. 2023. A first insight into the microbial and viral communities of Comau Fjord—a unique human-impacted ecosystem in Patagonia (42°S). Microorganisms 11:904. https://doi.org/10.3390/microorganisms11040904.

12. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

13. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation

using MinHash. Genome Biol 17:14. https://doi.org/10.1186/s13059-016-0997-x.

14. R Core Team. 2013. R: a language and environment for statistical computing.

15. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

16. Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. 2014. GroopM: an automated tool for the recovery of population genomes from related metagenomes. PeerJ 2:e603. https://doi.org/10.7717/peerj.603.

17. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.

18. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.

19. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.

20. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.

21. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.