AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

OMICS DATA SETS

Coastal Ocean Metagenomes and Curated Metagenome-Assembled Genomes from Marsh Landing, Sapelo Island (Georgia, USA)

Julian Damashek,[a]* Christian F. Edwardson,[a]* Bradley B. Tolar,[a]* Scott M. Gifford,[b] Mary Ann Moran,[a] James T. Hollibaugh[a]

[a]Department of Marine Sciences, University of Georgia, Athens, Georgia, USA
[b]Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

ABSTRACT Microbes play a dominant role in the biogeochemistry of coastal waters, which receive organic matter from diverse sources. We present metagenomes and 45 metagenome-assembled genomes (MAGs) from Sapelo Island, Georgia, to further understand coastal microbial populations. Notably, four MAGs are archaea, with two *Thaumarchaeota* and two marine group II *Euryarchaeota*.

Coastal oceans receive carbon and nutrients from rivers and marshes, driving high productivity. The metabolism of coastal microbes largely determines how much of the resulting organic matter (OM) is exported (1). Metagenomic data can provide insights into how microbial diversity relates to metabolic potential and drives OM processing (2). Coastal microbial biogeochemistry has been well studied at Sapelo Island, Georgia (3–5). Furthermore, these waters host a summer "bloom" of *Thaumarchaeota* and have been studied extensively to understand thaumarchaeal ecology (e.g., references 6–9). The metagenomic data presented here will guide an understanding of the microbial taxa in these waters and complement existing data for the same communities.

Seawater was collected at Marsh Landing (31°25′4.08″N, 81°17′34.26″W) as part of the Sapelo Island Microbial Carbon Observatory (http://www.simco.uga.edu/) by filtering through a 3.0-μm-pore-size prefilter and a 0.2-μm-pore-size Supor filter (Pall), which was frozen in liquid nitrogen (10). Duplicate filters were collected in August 2008 and 2009, 1 h before both day and night high tide on consecutive days (11). DNA extraction was done using the PowerSoil kit (Mo Bio), as described previously (7). DNA was sheared to ~225 bp, and libraries were constructed with the TruSeq DNA kit (Illumina) at the Georgia Genomics and Bioinformatics Core. Replicates from day and night samples on consecutive days were pooled to make 4 libraries (08N, 08D, 09N, and 09D; see Table 1), which were sequenced on 25% of an Illumina HiSeq 2500 platform rapid lane (paired-end, 150-bp reads) at the HudsonAlpha Institute for Biotechnology.

Default software parameters were used, unless otherwise stated. The reads had adapters removed with Trim Galore (https://github.com/FelixKrueger/TrimGalore), were trimmed with PRINSEQ v.0.20.4 (12), and were joined using PEAR v.0.9.10 (13), using parameters described previously (14) (Table 1). Paired and high-quality orphaned/singleton reads were coassembled using metaSPAdes ("--meta") within SPAdes v.3.7.0 (15), producing 83,626 contigs of >1,000 bp (N_{50}, 718 bp; L_{50}, 152,728; calculated with QUAST v.4.2 [16]).

Reads were mapped and indexed using Bowtie2 v.2.2.9 (17) and SAMtools v.1.3.1 (18), and contigs of >2.5 kbp (n = 18,714) were binned using anvi'o v.3 (19), following published protocols (20) (http://merenlab.org/data/tara-oceans-mags/). An anvi'o contig database was built to calculate k-mer frequencies, determine genes using Prodigal

Citation Damashek J, Edwardson CF, Tolar BB, Gifford SM, Moran MA, Hollibaugh JT. 2019. Coastal ocean metagenomes and curated metagenome-assembled genomes from Marsh Landing, Sapelo Island (Georgia, USA). Microbiol Resour Announc 8:e00934-19. https://doi.org/10.1128/MRA.00934-19.

Editor J. Cameron Thrash, University of Southern California

Address correspondence to Julian Damashek, judamash@utica.edu.

* Present address: Julian Damashek, Biology Department, Utica College, Utica, New York, USA; Christian F. Edwardson, Taconic Biosciences, Rensselaer, New York, USA; Bradley B. Tolar, Department of Earth System Science, Stanford University, Stanford, California, USA.

Received 1 August 2019
Accepted 16 September 2019
Published 3 October 2019

TABLE 1 Sampling, pooling, and quality control of metagenomic libraries

Library	SRA BioSample no.	Pooled samples (filter IDs)[a]	Samples, collection date (mo/day/year), collection time[b]	No. of raw reads	No. of high-quality reads, with adapters removed	No. of trimmed reads	No. of paired reads
08D	SAMN12211998	FN64, FN65, FN74, FN75	FN64 and FN65, 8/6/08, 11:47; FN74 and FN75, 8/7/08, 11:03	5,569,551	5,393,758	4,496,340	3,590,532
08N	SAMN12212006	FN59, FN60, FN69, FN70	FN59 and FN60, 8/6/08, 00:15; FN69 and FN70, 8/7/08, 00:50	6,495,098	6,266,198	5,284,070	5,005,127
09D	SAMN12212021	FN143, FN144, FN153, FN154	FN143 and FN144, 8/12/09, 11:26; FN153 and FN154, 8/13/09, 13:59	6,258,053	6,028,226	5,049,333	4,240,004
09N	SAMN12212029	FN148, FN149, FN159, FN160	FN148 and FN149, 8/13/09, 01:14; FN159 and FN160, 8/14/09, 02:30	6,324,184	6,090,086	5,084,547	4,444,614
Total				24,646,886	23,778,268	19,914,290	17,280,277

[a] IDs, identifiers.
[b] For sampling details, see Gifford et al. (11).

v.2.6.3 (21), and identify single-copy genes (22, 23) using HMMER v.3.1b2 (24). Bins generated by CONCOCT v.1.0.0 (25) were refined using the anvi'o interactive interface (26). Completeness and redundancy were assessed using anvi'o and CheckM v.1.0.12 (27); bins with <10% redundancy and ≥50% completeness were rerefined to minimize redundancy. Their resulting completeness and redundancy were estimated using anvi'o, CheckM, and the Microbial Genome Atlas (MiGA) Web server (28) (last accessed 18 August 2018). The resulting bins with completion of ≥50% were considered metagenome-assembled genomes (MAGs; $n = 45$) and were taxonomically annotated with MiGA. MAGs annotated below the order (genus) level included *Thaumarchaeota* (*Nitrosopumilus* spp., $n = 2$), marine group II *Euryarchaeota* ($n = 2$), *Synechococcaceae* (strain WH 8109, *Cyanobium* sp., $n = 2$), *Rhodobacteraceae* (*Phaeobacter* spp., $n = 5$), *Pelagibacteraceae* ($n = 2$), *Flavobacteriia* ($n = 3$), *Acidimicrobiaceae* (*Ilumatobacter* spp., $n = 2$), and *Halieaceae* ($n = 1$) (see https://figshare.com/articles/SIMO_MAG_table_v2/9791465/1).

Data availability. The reads, coassembly, and MAGs were deposited under GenBank BioProject number PRJNA552566. The reads are under SRA accession numbers SRX6421373 to SRX6421376. The coassembly and MAGs are under whole-genome sequencing (WGS) project numbers VMBT00000000 to VMDM00000000.

ACKNOWLEDGMENTS

Logistical support in the field was provided by the staff of the University of Georgia Marine Institute (UGAMI) and the Georgia Coastal Ecosystems Long Term Ecological Research (GCE-LTER) program. Shalabh Sharma kindly provided advice on bioinformatics.

This work was funded by National Science Foundation (NSF) grants OCE1538677 and OPP1643466 to J.T.H. and OCE1356010 to M.A.M. and was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

REFERENCES

1. Smith SV, Hollibaugh JT. 1993. Coastal metabolism and the oceanic organic carbon balance. Rev Geophys 31:75–89. https://doi.org/10.1029/92RG02584.
2. Moran MA, Kujawinski EB, Stubbins A, Fatland R, Aluwihare LI, Buchan A, Crump BC, Dorrestein PC, Dyhrman ST, Hess NJ, Howe B, Longnecker K, Medeiros PM, Niggemann J, Obernosterer I, Repeta DJ, Waldbauer JR. 2016. Deciphering ocean carbon in a changing world. Proc Natl Acad Sci U S A 113:3143–3151. https://doi.org/10.1073/pnas.1514645113.
3. Hanson RB, Synder J. 1979. Microheterotrophic activity in a salt-marsh estuary, Sapelo Island, Georgia. Ecology 60:99–107. https://doi.org/10.2307/1936472.
4. Kirchman D, K'nees E, Hodson R. 1985. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. Appl Environ Microbiol 49:599–607.
5. Mou X, Sun S, Edwards RA, Hodson RE, Moran MA. 2008. Bacterial carbon processing by generalist species in the coastal ocean. Nature 451:708–711. https://doi.org/10.1038/nature06513.
6. Hollibaugh JT, Gifford S, Sharma S, Bano N, Moran MA. 2011. Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. ISME J 5:866–878. https://doi.org/10.1038/ismej.2010.172.
7. Hollibaugh JT, Gifford SM, Moran MA, Ross MJ, Sharma S, Tolar BB. 2014.

Seasonal variation in the metratranscriptomes of a Thaumarchaeota population from SE USA coastal waters. ISME J 8:685–698. https://doi.org/10.1038/ismej.2013.171.
8. Schaefer SC, Hollibaugh JT. 2017. Temperature decouples ammonium and nitrite oxidation in coastal waters. Environ Sci Technol 51:3157–3164. https://doi.org/10.1021/acs.est.6b03483.
9. Damashek J, Tolar BB, Liu Q, Okotie Oyekan AO, Wallsgrove NJ, Popp BN, Hollibaugh JT. 2019. Microbial oxidation of nitrogen supplied as selected organic nitrogen compounds in the South Atlantic Bight. Limnol Oceanogr 64:982–995. https://doi.org/10.1002/lno.11089.
10. Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME J 5:461–472. https://doi.org/10.1038/ismej.2010.141.
11. Gifford SM, Sharma S, Moran MA. 2014. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. Front Microbiol 5:185. https://doi.org/10.3389/fmicb.2014.00185.
12. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
13. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. https://doi.org/10.1093/bioinformatics/btt593.
14. Edwardson CF, Hollibaugh JT. 2017. Metatranscriptomic analysis of prokaryotic communities active in sulfur and arsenic cycling in Mono Lake, California, USA. ISME J 11:2195–2208. https://doi.org/10.1038/ismej.2017.80.
15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
16. Gurevich AA, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
17. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
19. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.
20. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lücker S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol 3:804–813. https://doi.org/10.1038/s41564-018-0176-9.
21. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
22. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, Soll D, Podar M. 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci U S A 110:5540–5545. https://doi.org/10.1073/pnas.1303090110.
23. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352.
24. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
25. 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.
26. Delmont TO, Eren AM. 2016. Identifying contamination with advanced visualization and analysis practices: metagenomic approaches for eukaryotic genome assemblies. PeerJ 4:e1839. https://doi.org/10.7717/peerj.1839.
27. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
28. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288. https://doi.org/10.1093/nar/gky467.