Title

Reconstruction of Metagenome-Assembled Genomes from Aquaria.

Journal

Microbiology resource announcements, 10(31)

ISSN

2576-098X

Authors

Ettinger, Cassandra L
Bryan, Jordan
Tokajian, Sima
et al.

Publication Date

2021-08-01

DOI

10.1128/mra.00557-21

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Reconstruction of Metagenome-Assembled Genomes from Aquaria

Cassandra L. Ettinger,[a,b,c] Jordan Bryan,[d] Sima Tokajian,[e] Guillaume Jospin,[a,f] David Coil,[a] Jonathan A. Eisen[a,b,g]

[a]Genome Center, University of California, Davis, California, USA
[b]Department of Evolution and Ecology, University of California, Davis, California, USA
[c]Department of Microbiology & Plant Pathology, University of California, Riverside, California, USA
[d]College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, USA
[e]Department of Natural Sciences, Lebanese American University, Byblos, Lebanon
[f]AnimalBiome, Oakland, California, USA
[g]Department of Medical Microbiology and Immunology, University of California, Davis, California, USA

ABSTRACT Here, we report 11 metagenome-assembled genomes (MAGs) reconstructed from freshwater and saltwater aquaria, including representatives of *Polynucleobacter*, *Anaerolinea*, *Roseobacter*, *Flavobacteriia*, *Octadecabacter*, *Mycobacterium*, and Candidate Phyla Radiation (CPR) members. These MAGs can serve as a resource for aquatic research and elucidating the role of CPR taxa in the built environment.

Microbial communities play critical roles in aquarium health. Aquaria support complex multitrophic interactions between fish, invertebrates, plants, and microbial communities that occur in an enclosed built environment. Understanding the genomics of aquarium microbial communities is critical for understanding the health of other enclosed aquatic systems.

Samples were collected prior to the start of an undergraduate research project that investigated microbial community assembly of multiple aquaria in the fall of 2012 at the University of California, Davis (1). Tropical tank sediment (n = 3), cold reef tank sediment (n = 1), freshwater tank sediment (n = 3), cold reef tank water (n = 3), freshwater wipes (n = 3), and freshwater tank water (n = 3) were collected and processed for DNA extraction as described by Bik et al. (1). Libraries were made using a Nextera XT DNA library sample preparation kit (Illumina, Inc.) and were sequenced on an Illumina MiSeq instrument (paired end, 150-bp reads).

Reads were not quality filtered prior to assembly. All raw reads from all samples were coassembled using MEGAHIT (2) v.1.0.6. Metagenome-assembled genomes (MAGs) were generated using anvi'o v.2.3.2 (3). First, a contig database was produced using "anvi-gen-contigs-database," and open reading frames were identified with Prodigal (4) v.2.6.2. We then used "anvi-run-hmms" to run HMMER v.3.1b2 (5) to identify bacterial (6) and archeal (7) single-copy genes. Contig taxonomy was inferred using Kaiju v.1.5.0 (8) with the NCBI BLAST nonredundant protein database, including fungi and microbial eukaryotes v.2016-09-18. Reads were mapped using Bowtie 2 v.2.2.8 (9) and SAMtools v.1.4.1 (10). Using "anvi-profile" and "anvi-merge," contigs of >2.5 kbp were mapped to samples and then profiles were combined. On average, 780,565 reads per sample mapped to the contig database with the majority of mapped reads from cold reef tank water (57.3%) and freshwater tank water (41.7%). Contigs were clustered using "anvi-cluster-with-concoct" to automatically bin MAGs (11). MAG completeness and contamination were assessed in anvi'o using "anvi-summarize" and confirmed with CheckM v.1.0.7 (12). PhyloSift v.1.0.1 (13) was used to place MAGs in a phylogenetic context to provide additional information about taxonomic assignments. Candidate Phyla Radiation (CPR) genomes were identified with "anvi-script-gen-cpr-classifier" and "anvi-script-predict-cpr-genomes" using the Brown et al. (14) and Campbell et al.

Citation Ettinger CL, Bryan J, Tokajian S, Jospin G, Coil D, Eisen JA. 2021. Reconstruction of metagenome-assembled genomes from aquaria. Microbiol Resour Announc 10:e00557-21. https://doi.org/10.1128/MRA.00557-21.

Editor J. Cameron Thrash, University of Southern California

Address correspondence to Cassandra L. Ettinger, cassande@ucr.edu.

Received 1 June 2021
Accepted 16 July 2021
Published 5 August 2021

TABLE 1 Genomic feature summary for metagenome-assembled genomes identified from aquaria

Bin identifier	Draft sequence quality[a]	Putative taxonomy	Genome size (bp)	No. of contigs	N_{50} (bp)	GC content (%)	No. of genes[b]	CheckM % completion	CheckM % redundancy	anvi'o % completion	anvi'o % redundancy	CPR % completion[c]	GenBank accession no.
AQU-01	High	*Polynucleobacter* sp.	1,676,117	78	37,546	45.38	1,729	97.69	0.16	98.56	0.72	NA[d]	JAHBAK000000000
AQU-02	High	*Anaerolinea* sp.	5,354,340	574	12,300	53.21	4,854	91.36	1.73	95.68	2.88	NA	JAHBAL000000000
AQU-03	Medium	*Roseobacter* sp.	2,693,064	264	12,778	60.69	2,712	88.8	0.63	75.54	2.16	NA	JAHBAM000000000
AQU-04	Medium	*Flavobacteriia* sp.	1,846,950	228	9,847	41.36	1,796	86.62	0.07	92.81	1.44	NA	JAHBAN000000000
AQU-05	Medium	*Octadecabacter* sp.	2,609,161	434	6,891	55.83	2,896	84.33	1.96	76.98	1.44	NA	JAHBAO000000000
AQU-06	Medium	*Mycobacterium* sp.	3,351,002	625	5,981	66.59	3,663	70.94	1.74	71.94	2.16	NA	JAHBAP000000000
AQU-07	High	Candidatus *Shapirobacteria* sp.	863,951	104	11,221	35.61	941	77.27	1.72	76.98	3.6	93.02	JAHBAQ000000000
AQU-08	High	Candidatus *Kerfeldbacteria* sp.	1,070,839	48	32,721	46.2	1,043	72.94	0.31	84.89	2.88	93.02	JAHBAR000000000
AQU-09	High	Candidatus *Uhrbacteria* sp.	1,115,436	46	43,390	51.31	1,073	68.81	0.5	77.7	0	93.02	JAHBAS000000000
AQU-10	High	Candidatus *Moranbacteria* sp.	892,370	24	56,939	43.56	901	67.95	0.99	82.01	0	93.02	JAHBAT000000000
AQU-11	High	Candidatus *Saccharibacteria* sp.	1,090,569	48	32,149	47.69	1,151	56.94	1.03	71.22	0.72	90.7	JAHBAU000000000

[a] Quality estimates were based on CheckM values for non-CPR members and the CPR-specific completion estimates for CPR members.
[b] No. of genes predicted by Prodigal.
[c] Completion estimates were generated using 43 single-copy markers for CPR members following Brown et al. (14).
[d] NA, not applicable.

(6) databases. CPR genome completion was then re-estimated using 43 single-copy marker genes, as CPR members are known to have missing single-copy genes (14). CPR is putatively a diverse group of uncultured bacterial lineages with poorly understood metabolic functions known mostly from metagenomic sequencing work. Representatives of CPR have been previously found in a wide range of aquatic habitats, including bioreactors, ocean, lakes, groundwater, and waterways (15–21).

Using the standards suggested by Bowers et al. (22), we report two high-quality draft MAG sequences with >90% completion and four medium-quality draft MAG sequences with >70% completion (Table 1). Additionally, we report five high-quality draft MAG sequences that were identified as potential CPR genomes with >90% completion (Table 1). These metagenome-assembled genomes will enable deeper insights into the ecology of aquarium microbial communities and also into the possible functional roles of understudied lineages (e.g., CPR members) in the built environment.

Data availability. The raw sequencing reads, the coassembly, and the individual MAGs were deposited at DDBJ/ENA/GenBank under BioProject accession number PRJNA728121. Contigs identified as possible contaminants or adaptors by the NCBI Contamination Screen were subsequently trimmed or removed from the coassembly or individual MAGs during deposition.

ACKNOWLEDGMENTS

Illumina sequencing was performed at the DNA Technologies Core facility in the Genome Center at the University of California, Davis, Davis, California (UCD). We acknowledge the contributions of multiple people for help in various stages of the project, especially Thein Mai and Matt Wein for access to and assistance with the aquaria used in this study.

Funding for this study was provided by the Alfred P. Sloan Foundation through their program in the "Microbiology of the Built Environment."

REFERENCES

1. Bik HM, Alexiev A, Aulakh SK, Bharadwaj L, Flanagan J, Haggerty JM, Hird SM, Jospin G, Lang JM, Sauder LA, Neufeld JD, Shaver A, Sethi A, Eisen JA, Coil DA. 2019. Microbial community succession and nutrient cycling responses following perturbations of experimental saltwater aquaria. mSphere 4:e00043-19. https://doi.org/10.1128/mSphere.00043-19.
2. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
3. Eren AM, Murat Eren A, 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.
4. 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.
5. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7: e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
6. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, Söll 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.
7. 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.
8. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257.
9. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
10. 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.
11. 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.
12. 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.
13. Darling AE, Jospin G, Lowe E, Matsen FA, IV, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243. https://doi.org/10.7717/peerj.243.
14. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523:208–211. https://doi.org/10.1038/nature14486.
15. Cabello-Yeves PJ, Zemskaya TI, Zakharenko AS, Sakirko MV, Ivanov VG, Ghai R, Rodriguez-Valera F. 2020. Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. Limnol Oceanogr 65:1471–1488. https://doi.org/10.1002/lno.11401.
16. Ruuskanen MO, Colby G, St Pierre KA, St Louis VL, Aris-Brosou S, Poulain AJ. 2020. Microbial genomes retrieved from High Arctic lake sediments encode for adaptation to cold and oligotrophic environments. Limnol Oceanogr 65:S233–S247. https://doi.org/10.1002/lno.11334.
17. Thrash JC, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN, Henrissat B, Mason OU. 2017. Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico "dead zone." mBio 8: e01017-17. https://doi.org/10.1128/mBio.01017-17.
18. Zhao Y, Liu S, Jiang B, Feng Y, Zhu T, Tao H, Tang X, Liu S. 2018. Genome-centered metagenomics analysis reveals the symbiotic organisms possessing ability to cross-feed with anammox bacteria in anammox consortia. Environ Sci Technol 52:11285–11296. https://doi.org/10.1021/acs.est.8b02599.
19. Vigneron A, Cruaud P, Langlois V, Lovejoy C, Culley AI, Vincent WF. 2020. Ultra-small and abundant: candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. Limnol Oceanogr 5:212–220. https://doi.org/10.1002/lol2.10132.
20. Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ. 2017. Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. Microbiome 5:112. https://doi.org/10.1186/s40168-017-0331-1.

21. León-Zayas R, Peoples L, Biddle JF, Podell S, Novotny M, Cameron J, Lasken RS, Bartlett DH. 2017. The metabolic potential of the single cell genomes obtained from the Challenger Deep, Mariana Trench within the candidate superphylum Parcubacteria (OD1). Environ Microbiol 19:2769–2784. https://doi.org/10.1111/1462-2920.13789.
22. 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.