Draft Genome Sequences of Four Bacterial Species as Part of an Experiential Microbiology Project at SUNY Geneseo

Lauren N. Ellis,a Enrico Amato, Jr.,a Christopher A. Lepore,a Lucas B. Sutton,a Bea Dipzinski,a Brianna Cunneen,a Julian Wisun,a Kayla A. Cannon,a Kyle O’Byrne,a Isidro Bosch,a Elizabeth Hutchison,a Logan M. Peoplesa

aDepartment of Biology, SUNY Geneseo, Geneseo, New York, USA

ABSTRACT We report four draft genome sequences related to the genera Bacillus and Escherichia, recovered from surfaces associated with human interaction, and Sediminibacterium, recovered from an aquatic environment. This study was part of an undergraduate microbial bioinformatics course at the State University of New York at Geneseo.

As part of a capstone project of an undergraduate bioinformatics course at the State University of New York (SUNY) at Geneseo, students sought to obtain genome sequences from microbes from human-associated environments. Students gained experience with key microbiological techniques consistent with proposed curricular guidelines (1), including microbial culturing, taxonomic identification, hypothesis-driven laboratory practices, and scientific communication. We report four draft genome sequences related to the genera Bacillus, Escherichia, and Sediminibacterium.

Samples were collected from the Integrated Science Center (ISC), located on the SUNY Geneseo campus, and Conesus Lake in Livingston County, New York, a source of drinking water and a recreational area prone to potentially toxic cyanobacterial blooms (2) (Table 1). Three samples were taken aseptically using sterile swabs from the following locations within the ISC: a computer keyboard in the chemistry stockroom, an elevator button, and a phone. The swabs were streaked onto Reasoner’s 2A agar and incubated at 37°C for 1 week. Colonies of interest were streaked onto tryptic soy agar and incubated at 37°C for 1 week. Colonies of interest were streaked onto tryptic soy agar and incubated at 37°C. After repeated subculturing, a single colony was inoculated into tryptic soy broth and incubated for 24 h at 37°C with shaking at 150 rpm to obtain sufficient biomass for DNA extraction.

To isolate potential secondary metabolite-producing Cyanobacteria, freshwater was collected from Conesus Lake at a depth of 3 m using a Van Dorn discrete-depth sampling bottle during a picoplankton bloom in July 2019. To remove eukaryotic algae, a 100-ml subsample was serially filtered through 3.0-μm and 1.0-μm Whatman Nuclepore filters. The filtrate was inoculated into Alga-Gro freshwater medium (Carolina Biological Supply Company, Burlington, NC) and incubated at room temperature under a GrowLite fluorescent light (Barron Lighting Group, Glendale, AZ). The successful enrichment of Cyanobacteria was confirmed using epifluorescence microscopy. After repeated subculturing, a culture identified morphologically as Synechococcus was selected for whole-genome sequencing.

Genomic DNA was extracted from all cultures using a Mo Bio Ultraclean microbial DNA isolation kit (Qiagen, Hilden, Germany). Genomic DNA libraries were prepared using a Nextera XT kit (Illumina, San Diego, CA), and 150-bp paired-end reads were sequenced on a NextSeq 550 instrument at the Microbial Genome Sequencing Center (Pittsburgh, PA). Reads were quality trimmed using Trimmomatic (3) with the parameters LEADING:3 TRAILING:10 SLIDINGWINDOW:4:15. Paired-end reads of ≥100 bp were assembled using SPAdes v3.14.1 with the flag --careful (4). Contigs longer than 1 kb
| Isolate | Related strain (ANI [%]) | Source | Size (Mbp) | Completeness (%) | Contamination (%) | Coverage (×) | No. of contigs | N50 value (bp) | GC content (%) | No. of genes | Genome accession no. | SRA accession no. |
|---------|--------------------------|--------|------------|------------------|------------------|--------------|---------------|----------------|----------------|--------------|-----------------|------------------|
| Gen1    | *Escherichia coli* pK19EC149 (99.98) | Phone  | 4.54       | 99.96            | 0.08             | 73           | 87            | 110,270        | 50.72          | 4,378        | JACBFE0000000000 | SRX8344491       |
| Gen2    | *Bacillus licheniformis* DSM 13 (99.99) | Elevator button | 4.17       | 98.82            | 0.00             | 77           | 30            | 309,949        | 46.14          | 4,270        | JACBFU0000000000 | SRX8344492       |
| Gen3    | *Bacillus oleronius* DSM 9356 (98.85) | Computer keyboard | 5.45       | 98.30            | 2.04             | 77           | 59            | 183,139        | 34.99          | 5,471        | JACBFM0000000000 | SRX8344493       |
| Gen4    | *Sediminibacterium goheungense* | Conesus Lake enrichment | 3.52       | 98.03            | 0.52             | 69           | 59            | 85,556         | 39.32          | 3,187        | JACBFP0000000000 | SRX8344494       |
were retained. Genome statistics were analyzed using QUAST v5.0.2 (5). Completeness and contamination were determined using CheckM v1.0.13 with the \textit{--reduced\_tree} flag (6). Based on the CheckM output, we determined that the Conesus Lake cyanobacterial enrichment was not axenic and contained members related to \textit{Synechococcus}, \textit{Sediminibacterium}, and \textit{Hydrogenophaga}. The contig depth of coverage was estimated using Bowtie 2 v2.4.1 and SAMtools v1.10 (7, 8), and contigs within this minimetagenome were binned using MetaBAT v0.26.3 with the flag \textit{--very\_sensitive} (9). We report a genome bin obtained related to the genus \textit{Sediminibacterium}. All four draft genomes were initially annotated using Prokka v1.14.6 (10), and final genome annotations are reported here using the NCBI Prokaryotic Genome Annotation Pipeline (11, 12). Average nucleotide identity (ANI) comparisons were performed using orthoANI (13).

While three of the genomes are closely related to previously reported strains, \textit{Sediminibacterium} sp. strain Gen4 from Conesus Lake represents a novel species. We add to a growing body of evidence that members of the genus \textit{Sediminibacterium} may be part of the cyanobacterial phycosphere (14, 15).

Data availability. This genome sequencing project has been deposited in GenBank under the accession number PRJNA631923.

ACKNOWLEDGMENTS

We thank the SUNY Geneseo Biology Department for financial support and Tom Reho and Shawn Austin (SUNY Geneseo) for technical support.

REFERENCES

1. Merkel S, the ASM Task Force on Curriculum Guidelines for Undergraduate Microbiology. 2012. The development of curricular guidelines for introductory microbiology that focus on understanding. J Microbiol Biol Educ 13:32–38. https://doi.org/10.1128/jmbe.v13i1.363.
2. Bosch I, Makarewicz JC, Lewis TW, Bonk EA, Finiguerra M, Groveman B. 2009. Management of agricultural practices results in declines of filamentous algae in the lake littoral. J Great Lakes Res 35:90–98. https://doi.org/10.1016/j.jglr.2008.10.007.
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolaenko 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.
5. Gurevich A, 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.
6. 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.168072.114.
7. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
8. 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.
9. Kang DO, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
10. Seemann T. 2014. Prokka: rapid prokaryote genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
11. Tatusova T, DiCuccio M, Badreddin A, Chetverin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostnell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
12. Haft DH, DiCuccio M, Badreddin A, Brover V, Chetverin V, O’Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissl F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.
13. Yoon SH, Ha SM, Lim JM, Kwon SJ, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4.
14. Pinto F, Tett A, Armanini F, Asnicar F, Boscai A, Pasolli E, Zolfo M, Donati C, Salmaso D, Segata N. 2018. Draft genome sequences of novel \textit{Pseudomonas}, \textit{Flavobacterium}, and \textit{Sediminibacterium} species strains from a freshwater ecosystem. Microbiol Resour Announc 6:e00009-18. https://doi.org/10.1128/microbiolresources.0009-18.
15. Xu Z, Te SH, He Y, Gin KY-H. 2018. The characteristics and dynamics of cyanobacteria–heterotrophic bacteria between two estuarine reservoirs—tropical versus sub-tropical regions. Front Microbiol 9:2531. https://doi.org/10.3389/fmicb.2018.02531.