Metagenomes, Metatranscriptomes, and Metagenome-Assembled Genomes from Chesapeake and Delaware Bay (USA) Water Samples

Mir Alvee Ahmed,a Shen Jean Lim,a,* Barbara J. Campbella

aDepartment of Biological Sciences, Clemson University, Clemson, South Carolina, USA

ABSTRACT Here, we present 36 metagenomes, 59 metatranscriptomes, and 373 metagenome-assembled genomes (MAGs) from Chesapeake and Delaware Bay water samples. This data set will be useful for studying microbial biogeochemical cycling in estuaries.

Estuaries are productive aquatic environments harboring diverse flora, fauna, and microbial communities important for global carbon and nutrient cycling (1, 2). Omics analyses of estuary-associated bacteria will reveal associations between bacterial communities, functional composition, and environmental variation. Three cruises and two cruises along longitudinal transects of the Delaware Bay (DEBay) and Chesapeake Bay (CPBay), two geographically close estuaries with contrasting environmental gradients, were completed in 2014 and 2015, respectively, aboard R/V Sharp. Surface (~1.5 m below the seafloor [mbsf]) water samples were collected using a rosette sampler with associated conductivity-temperature-depth (CTD) profiles. The sampling scheme (3), environmental measurements, and bacterial production measurements (4–7) were described previously (4–7) and archived (https://www.bco-dmo.org/dataset/565451).

Cells were separated as large- and small-cell-size fractions by passing water samples through 0.8- and 0.22-μm-pore-size filters. Nucleic acids were extracted from size-fractionated cells using the Allprep DNA/RNA minikit (Qiagen, Valencia, CA, USA) (3). A total of 36 (12 from CPBay and 24 from DEBay) metagenomic and 59 (24 from CPBay and 35 from DEBay) metatranscriptomic libraries were prepared using the TruSeq library preparation kit (Illumina) and sequenced by the Joint Genome Institute on the Illumina HiSeq 2000 platform at 2 × 150 bp, as described previously (3). Two DEBay metagenomes (DEBay_Spr_30_<0.8_DNA and DEBay_Sum_22_D_<0.8_DNA) were also sequenced in-house with the Nanopore rapid sequencing kit (Oxford Nanopore Technologies, Kidlington, Oxfordshire, UK) on a MinION flow cell (9.4 nanopores) with a MinION Mk1B sequencer. MinKNOW was used for basecalling (8). Sequencing statistics are in Table 1 (metagenomes) and online at https://doi.org/10.6084/m9.figshare.14173664 (metatranscriptomes).

Prior to assembly, Cutadapt v1.11 and Sickle v1.33 were used to remove adapters from and quality trim (Q = 30) Illumina-sequenced reads (9). Nanopore-sequenced reads were not error corrected or trimmed prior to hybrid assembly, as recommended by hybridSPAdes v3.11.1, because they were used only for gap closure and repeat resolution (10). Read qualities pre- and posttrimming were assessed with FastQC v0.11.5 (Babraham Bioinformatics, 2010). Metagenomic assemblies were performed using the default parameters of metaSPAdes v3.11.1 (11) with increased memory allocation (--meta --m 450) and evaluated using MetaQUAST, v5.0.2 (12) (Table 1).
For binning, trimmed reads from each Illumina-sequenced library were mapped to contigs ≥2,000 bp in the corresponding metagenome using the default parameters (end-to-end mode) of Bowtie 2 v2.2.7 (13). Alignments converted to binary alignment map (BAM) format with SAMtools v0.1.19 (13, 14) were binned into 373 metagenome-assembled genomes (MAGs) using the default parameters of MetaBAT2 v2.10.2 (15). MAG statistics, including GC content, size, completeness, and contamination, were assessed by CheckM v1.0.16 (16) and Anvi’o v6.2 and v7 (17–19). Coassembled sequences of both size fractions from the same water sample were binned when separate binning did not give useful MAGs. A subset of 364 MAGs (https://doi.org/10.6084/m9.1.figshare.14179448) with >80% completeness and <5% contamination were taxonomically annotated using Anvi’o and Microbial Genome Atlas (MiGA) v0.7.26.2 (20). They belonged to bacterial orders Actinomycetales (n = 7), Burkholderiales (n = 31), Flavobacteriales (n = 55), Microtrichales (n = 39), Nanopolagales (n = 11), Pelagibacteriales (n = 31), Pseudomonadales (n = 26), Rhodobacterales (n = 28), and Synechococcales (n = 13), as well as the archaeal phyla Crenarchaeota (n = 5) and Euryarchaeota (n = 2).

**Data availability.** The metagenomes, metatranscriptomes, and MAGs are available on NCBI under the umbrella project PRJNA432171.

### Table 1: Accession numbers and characteristics of the Chesapeake and Delaware Bay water samples

<table>
<thead>
<tr>
<th>Metagenome</th>
<th>NCBI BioSample no.</th>
<th>Collection date (yr-mo-day, time)</th>
<th>No. of raw reads</th>
<th>No. of paired reads</th>
<th>No. of contigs</th>
<th>Total length (bp)</th>
<th>N50 (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPBay_Spr_0.1_0.8_DNA</td>
<td>SAMN06266052</td>
<td>2015-4-11, 8:54</td>
<td>281,517,088</td>
<td>270,649,460</td>
<td>38,004</td>
<td>64,955,483</td>
<td>1,606</td>
</tr>
<tr>
<td>CPBay_Spr_15.0_0.2_DNA</td>
<td>SAMN06264333</td>
<td>2015-4-13, 7:00</td>
<td>120,796,538</td>
<td>106,955,332</td>
<td>67,010</td>
<td>175,958,607</td>
<td>3,173</td>
</tr>
<tr>
<td>CPBay_Spr_15.0_0.8_DNA</td>
<td>SAMN06266053</td>
<td>2015-4-13, 7:00</td>
<td>48,340,218</td>
<td>45,866,608</td>
<td>102,184</td>
<td>307,564,961</td>
<td>4,092</td>
</tr>
<tr>
<td>CPBay_Spr_31.0_0.2_DNA</td>
<td>SAMN06264351</td>
<td>2015-4-16, 7:25</td>
<td>74,037,960</td>
<td>72,685,680</td>
<td>95,167</td>
<td>260,760,381</td>
<td>3,425</td>
</tr>
<tr>
<td>CPBay_Spr_31.0_0.8_DNA</td>
<td>SAMN06264995</td>
<td>2015-4-16, 7:25</td>
<td>173,012,398</td>
<td>158,595,630</td>
<td>147,394</td>
<td>376,160,656</td>
<td>3,055</td>
</tr>
<tr>
<td>CPBay_Sum_0.6_0.2_DNA</td>
<td>SAMN06266060</td>
<td>2015-8-17, 10:50</td>
<td>141,437,090</td>
<td>123,369,312</td>
<td>244,783</td>
<td>609,304,882</td>
<td>2,860</td>
</tr>
<tr>
<td>CPBay_Sum_0.6_0.8_DNA</td>
<td>SAMN06265975</td>
<td>2015-8-17, 10:50</td>
<td>306,249,226</td>
<td>288,883,072</td>
<td>165,685</td>
<td>439,027,102</td>
<td>3,214</td>
</tr>
<tr>
<td>CPBay_Sum_15.0_0.2_DNA</td>
<td>SAMN06265905</td>
<td>2015-8-20, 10:54</td>
<td>52,780,740</td>
<td>50,636,990</td>
<td>86,423</td>
<td>198,558,727</td>
<td>2,459</td>
</tr>
<tr>
<td>CPBay_Sum_15.0_0.8_DNA</td>
<td>SAMN06266062</td>
<td>2015-8-20, 10:54</td>
<td>37,888,140</td>
<td>36,283,452</td>
<td>94,006</td>
<td>245,336,553</td>
<td>3,092</td>
</tr>
<tr>
<td>CPBay_Sum_20.0_0.8_DNA</td>
<td>SAMN06264994</td>
<td>2015-8-21, 10:58</td>
<td>49,252,868</td>
<td>46,975,502</td>
<td>96,381</td>
<td>226,955,816</td>
<td>2,597</td>
</tr>
<tr>
<td>CPBay_Sum_27.0_0.2_DNA</td>
<td>SAMN06265909</td>
<td>2015-8-22, 12:05</td>
<td>51,275,568</td>
<td>49,528,260</td>
<td>89,461</td>
<td>204,236,447</td>
<td>2,450</td>
</tr>
<tr>
<td>CPBay_Sum_27.0_0.8_DNA</td>
<td>SAMN06265908</td>
<td>2015-8-22, 12:05</td>
<td>44,125,638</td>
<td>42,165,180</td>
<td>104,303</td>
<td>252,367,039</td>
<td>2,714</td>
</tr>
</tbody>
</table>

a Nanopore sequences.

b NA, not available.
ACKNOWLEDGMENTS

The research cruises were supported by a National Science Foundation grant (OCE-082546) to B.J.C. Metagenomic and metatranscriptomic sequencing was supported by a DOE/JGI grant (CSP-1621) to B.J.C. A National Science Foundation grant (EF-2025541) to B.J.C. supported data processing.

REFERENCES