Draft Genome Sequence of a Novel SAR11 Clade Species Abundant in a Tibetan Lake

Seungdae Oh,[a] Rui Zhang,[b,c] Qinglong L. Wu,[d] Wen-Tso Liu[a]

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA[a]; State Key Laboratory of Marine Environmental Science[b] and Institute of Marine Microbes and Ecospheres,[c] Xiamen University, Fuijan, China; State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China[d]

SAR11 clade bacteria are abundant and play a key role in the nutrient cycles of marine and, presumably, inland aquatic environments. We report here the draft genome sequence of a novel species in the SAR11 cluster, reconstructed from a metagenomic data set obtained from a Tibetan lake.

Address correspondence to Wen-Tso Liu, wtliu@illinois.edu.

The SAR11 clade is a member of heterotrophic *Alphaproteobacteria* that are widely distributed in marine and inland aquatic ecosystems (1–3). While the SAR11 clade bacteria are abundant and play an important role in marine biogeochemical cycles, their ecological roles in lacustrine ecosystems remain relatively less characterized. Lake Qinghai represents the seventh largest lake on Earth, located on the Tibetan Plateau (4). Our previous 16S rRNA gene clone library analysis of the planktonic microbial communities obtained from the mountain lake discovered SAR11-like sequences, which were abundant (up to 19% of the total bacteria) and phylogenetically distinguishable from the freshwater subclade IIIb/LD12 lineage (5). Here, we report the draft genome sequence of the SAR11 clade species obtained from Lake Qinghai.

The draft genome sequence of the SAR11 clade species alphaproteobacterium strain QL1, was recovered from a metagenomic data set of a surface water sample taken from Lake Qinghai. The DNA was extracted as described previously (4, 5), and the Roche 454 GS-FLX sequencing technology generated 577-Mbp metagenomic sequences. The metagenomic data set was trimmed based on a Phred quality score cutoff of 10 using SolexaQA2 (6) and assembled using Newbler 2.8. Binning of the assembled contigs was carried out based on metagenomic read coverage, tetranucleotide frequency, and the occurrence of unique marker genes using MaxBin (7), which generated 99 clustered contigs. Searching the 99 contigs against all bacterial and archaeal genome sequences available in the GenBank database (as of January 2014 [ftp://ftp.ncbi.nih.gov/]) using BLASTn (8) showed best hits to "*Candidatus* Pelagibacter" IMCC9063 (9). Metagenomic read recruitment on the 99 contigs using BLASTn showed >95% nucleotide sequence identity with >20× coverage. The sequence composition-based binning, the sequence homology search, and the metagenomic read recruitment collectively indicated that the 99 contigs represent well the draft genome sequence of the single-species population.

The draft genome was 952,704 bp, with 31.4% G+C content. Gene prediction and functional annotation on the draft genome were performed using the RAST server (10) and the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). The draft genome contains a total of ≥870 protein-coding genes (>300 bp), a 16S rRNA gene, a 23S rRNA gene, and 20 tRNA genes. Searching the 16S rRNA gene against a reference rRNA sequence database (GenBank) using BLASTn revealed the highest sequence identity to "*Ca.* Pelagibacter" IMCC9063 (98% identity) in subclade IIIa. Searching the 16S rRNA gene against the nucleotide sequence collection database (GenBank) using BLASTn revealed 100% sequence identity to uncultured bacterium clone sequences obtained from the Chesapeake Bay and 99% sequence identity to those obtained from the Baltic Sea, Delaware Bay, and Lake Nam Co (5, 11). These results highlighted the global distribution of the QL1-like species in both lacustrine and marine ecosystems. We are currently exploring the gene content and genomic variability of the draft genome compared to those of other SAR11 clade members.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. JPLS00000000. The version described in this paper is version JPLS01000000.

ACKNOWLEDGMENTS

REFERENCES

1. **Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ.** 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature **420:**806–810. http://dx.doi.org/10.1038/nature01240.
2. **Salcher MM, Pernthaler J, Posch T.** 2011. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria "that rule

the waves" (LD12). ISME J **5**:1242–1252. http://dx.doi.org/10.1038/ is-mej.2011.8.

3. **Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J.** 2009. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ. Microbiol. **11**:2078–2093. http://dx.doi.org/10.1111/j.1462-2920.2009.01929.x.

4. **Wu QL, Zwart G, Schauer M, Kamst-van Agterveld MP, Hahn MW.** 2006. Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. Appl. Environ. Microbiol. **72**:5478–5485. http://dx.doi.org/10.1128/AEM.00767-06.

5. **Zhang R, Wu Q, Piceno YM, Desantis TZ, Saunders FM, Andersen GL, Liu W-T.** 2013. Diversity of bacterioplankton in contrasting Tibetan lakes revealed by high-density microarray and clone library analysis. FEMS Microbiol. Ecol. **86**:277–287. http://dx.doi.org/10.1111/1574-6941.12160.

6. **Cox MP, Peterson Da, Biggs PJ.** 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics **11**:485. http://dx.doi.org/10.1186/1471-2105-11-485.

7. **Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW.** 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. **2**:26. http://dx.doi.org/10.1186/2049-2618-2-26.

8. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215**:403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

9. **Oh H-M, Kang I, Lee K, Jang Y, Lim S-I, Cho J-C.** 2011. Complete genome sequence of strain IMCC9063, belonging to SAR11 subgroup 3, isolated from the Arctic Ocean. J. Bacteriol. **193**:3379–3380. http://dx.doi.org/10.1128/JB.05033-11.

10. **Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. **33**:5691–5702. http://dx.doi.org/10.1093/nar/gki866.

11. **Shaw AK, Halpern AL, Beeson K, Tran B, Venter JC, Martiny JBH.** 2008. It's all relative: ranking the diversity of aquatic bacterial communities. Environ. Microbiol. **10**:2200–2210. http://dx.doi.org/10.1111/j.1462-2920.2008.01626.x.