Draft Genome Sequence of Uncultured SAR324 Bacterium lautmerah10, Binned from a Red Sea Metagenome

Mohamed F. Haroon,[a] Luke R. Thompson,[a,b] Ulrich Stingl[a]

Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia[a]; Department of Pediatrics, University of California San Diego, San Diego, California, USA[b]

A draft genome of SAR324 bacterium lautmerah10 was assembled from a metagenome of a surface water sample from the Red Sea, Saudi Arabia. The genome is more complete and has a higher G+C content than that of previously sequenced SAR324 representatives. Its genomic information shows a versatile metabolism that confers an advantage to SAR324, which is reflected in its distribution throughout different depths of the marine water column.

Address correspondence to Mohamed F. Haroon, fauziharoon@gmail.com, or Ulrich Stingl, ulistingl@gmail.com.

Members of SAR324 (also known as marine group B) represent a deeply branched clade within the *Deltaproteobacteria*. They have been identified via 16S rRNA marker gene surveys at various depths of the marine water column but are most abundant in oxygen minimum zones (1). Despite their ubiquity in the marine environments, there is still no successful report of cultured isolates from this clade. Previous studies have retrieved SAR324 genomes from the sequencing of single cells and environmental DNA, but these genomes are highly fragmented and incomplete (<80%) (2–4). Genomic information indicates that SAR324 has a flexible metabolism, which includes sulfur oxidation, carbon fixation, hydrocarbon utilization, and heterotrophy (4). None of the sequenced genomes from SAR324 have originated from the Red Sea, although abundance profiles have shown that SAR324 comprises a significant proportion of its resident microbial population (L. R. Thompson, G. J. Williams, M. F. Haroon, A. Shibl, P. Larsen, J. Shorenstein, R. Knight, and U. Stingl, unpublished data). Here, we describe a near-complete SAR324 bacterium lautmerah10 genome sequence retrieved from shotgun metagenomic sequencing of the Red Sea.

Surface water samples (10 m depth) were collected from the Red Sea (18.58°N, 39.799°E) during the 2011 KAUST Red Sea expedition in the summer (L. R. Thompson et al., unpublished data). The seawater was filtered through different filter pore sizes (5.0 μm, 1.2 μm, and 0.1 μm), and genomic DNA was isolated from the 0.1-μm-pore-size filters (0.1- to 1.2-μm size fraction) using phenol-chloroform extraction, as previously described (L. R. Thompson et al., unpublished data; see also reference 5). Paired-end libraries (2 × 100 bp) were prepared using the Nextera DNA library prep kit (Illumina, Inc.) and sequenced on a HiSeq 2000 (Illumina, Inc.). Reads were quality checked and trimmed using PRINSEQ version 0.20.4 (6). Trimmed metagenome reads were assembled using IDBA-UD version 1.1.1 (7) with the pre-correction option. The SAR324 lautmerah10 genome was extracted based on tetranucleotide frequency and coverage using MetaBAT version 0.26.1 (8), using default parameters.

RAST (9) and the NCBI Prokaryotic Genome Annotation Pipeline were used for genome annotation. The final SAR324 lautmerah10 genome comprised 290 scaffolds, with a total length of 3.5 Mb, and contained 3,832 protein-coding genes, 31 tRNAs, and 1 rRNA operon. The completeness (>96%) and contamination (0%) of the draft genome were assessed using CheckM version 1.0.3 (10). The 16S rRNA gene showed 99% similarity to uncultured environmental clones retrieved from the Sargasso Sea (11). The G+C content (47%) is higher than that of previously sequenced SAR324 representatives (<42%). The annotation showed a high number of coding sequences (3,832 is the largest number of genes observed in an SAR324 genome to date) and supported a metabolically diverse metabolism consistent with other SAR324 bacteria, which includes sulfur oxidation, alkane oxidation, and the utilization of organic carbon compounds as electron donors. However, in contrast to other SAR324 genomes, the SAR324 lautmerah10 genome presented here does not contain the nitrite reductase gene and thus is likely not able to reduce nitrite.

Nucleotide sequence accession number. The SAR324 lautmerah10 draft genome has been deposited at GenBank under the accession no. LNZD00000000.

REFERENCES

1. **Wright JJ, Konwar KM, Hallam SJ.** 2012. Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol **10**:381–394. http://dx.doi.org/10.1038/nrmicro2778.

2. **Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland EDP, Gomez ML, Sieracki ME, DeLong EF, Herndl GJ, Stepanauskas R**. 2011. Potential for chemo-lithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science **333**:1296–1300. http://dx.doi.org/10.1126/science.1203690.

3. **Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo M-J, Dupont CL, Badger JH, Novotny M, Rusch DB, Fraser LJ, Gormley NA, Schulz-Trieglaff O, Smith GP, Evers DJ, Pevzner PA, Lasken RS**. 2011. Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol **29**:915–921. http://dx.doi.org/10.1038/nbt.1966.

4. **Sheik CS, Jain S, Dick GJ**. 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. Environ Microbiol **16**:304–317. http://dx.doi.org/10.1111/1462-2920.12165.

5. **Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg J, Utterback T, Rogers Y-H, Falcon L, Souza V, Bonilla-Rosso G, Eguiarte L, Karl D, Sathyendranath S**. 2007. The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. **5**:e77. http://dx.doi.org/10.1371/journal.pbio.0050077.

6. **Schmieder R, Edwards R**. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**:863–864. http://dx.doi.org/10.1093/bioinformatics/btr026.

7. **Peng Y, Leung HC, Yiu S-M, Chin FY**. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28**:1420–1428. http://dx.doi.org/10.1093/bioinformatics/bts174.

8. **Kang DD, Froula J, Egan R, Wang Z**. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PEER J **3**:e1165. http://dx.doi.org/10.7717/peerj.1165.

9. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.

10. **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 http://dx.doi.org/10.1101/gr.186072.114.

11. **Wright TD, Vergin KL, Boyd PW, Giovannoni SJ**. 1997. A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. Appl Environ Microbiol **63**:1441–1448.