Complete Genome Sequence of *Staphylococcus haemolyticus* Type Strain SGAir0252

Balakrishnan N. V. Premkrishnan,[a] Ana Carolina M. Junqueira,[a]* Akira Uchida,[a] Rikky W. Purbojati,[a] James N. I. Houghton,[a] Caroline Chénard,[b] Anthony Wong,[a] Sandra Kolundžija,[a] Megan E. Clare,[a] Kavita K. Kushwaha,[a] Deepa Panicker,[a] Alexander Putra,[a] Nicolas E. Gaultier,[a] Cassie E. Heinle,[a] Vineeth Kodengil Vettath,[a] Daniela I. Drautz-Moses,[a] Stephan C. Schuster[a]

aSingapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore
bAsian School of the Environment, Nanyang Technological University, Singapore

ABSTRACT *Staphylococcus haemolyticus* is a coagulase-negative staphylococcal species that is part of the skin microbiome and an opportunistic human pathogen. The strain SGAir0252 was isolated from tropical air samples collected in Singapore, and its complete genome comprises one chromosome of 2.63 Mb and one plasmid of 41.6 kb.

The *Staphylococcus* genus is associated with several bacterial infections in humans and other animals and is described as part of the human skin microbiome (1). *Staphylococcus haemolyticus* is included in the coagulase-negative staphylococci (CoNS) group and is one of the most important human-associated pathogens (2) belonging to the *Staphylococcus* genus. *S. haemolyticus* has been implicated in native valve endocarditis, septicemia, peritonitis, and urinary tract infections, and is occasionally associated with wound, bone, and joint infections (3).

Air samples were collected in Singapore using an Andersen single-stage impactor (SKC BioStage). This air sampler collects bioaerosol particles at a flow rate of 28.3 liter/min. Viable air particles were collected on Trypticase soy agar (TSA; Becton, Dickinson) media plates placed below the sampler and incubated overnight at 30°C. Further isolation was conducted in Luria-Bertani (LB) broth incubated at room temperature overnight. DNA was extracted using a Wizard Genomic DNA purification kit (Promega), following the manufacturer's protocol. A whole-genome shotgun (WGS) library was prepared with the SMRTbell template prep kit 1.0 (Pacific Biosciences), using a 15-kb cutoff. Single-molecule real-time (SMRT) sequencing was performed on a PacBio RS II (Pacific Biosciences) sequencer. Additional WGS libraries were prepared with a TruSeq Nano DNA library preparation kit and sequenced on an Illumina MiSeq instrument to generate 300-bp paired-end reads. The total number of long reads obtained with the PacBio RSII platform was 23,264 (71.33-fold coverage), whereas the Illumina MiSeq platform provided 653,569 (121.27-fold coverage) short reads.

De novo genome assembly of long reads was performed using the Hierarchical Genome Assembly Process version 3 (HGAP3) implemented in the SMRT analysis algorithm version 2.3.0 (4). Different stages of the process involved preassembly, *de novo* assembly with Celera Assembler, and polishing with Quiver (4). Illumina reads were used to improve the genome contiguity, with the help of the error correction software Pilon version 1.16 (5). The complete circular genome contained two contigs, consisting of one chromosome of 2,632,932 bp (71.3-fold coverage) and the plasmid pSGAir0252 of 41,662 bp (18.63-fold coverage). The overall G+C content was 32.78%.

The Phylum-specific Automated Phylogenomic Inference Pipeline (Phyla_AMPHORA) (6) predicted 100% identity at the species taxonomic level. Further taxonomic assignment was investigated with the average nucleotide identity (ANI) method, using the Micro-

Address correspondence to Stephan C. Schuster, SCSchuster@ntu.edu.sg.

* Present address: Ana Carolina M. Junqueira, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

B.N.V.P. and A.C.M.J. contributed equally to this article.

bial Species Identifier (MiSI) (7), which showed 99.56% identity with the available reference genome sequence of *Staphylococcus haemolyticus* JCSC1435 (NCBI reference sequence number NC_007168).

Multilevel genome annotation was conducted by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (8). The complete genome consists of 2,528 genes, of which 2,445 were protein-coding genes, 19 rRNA subunits (7 genes for 5S, 6 for 16S, and 6 for 23S), 60 tRNAs, and 4 noncoding RNAs. A total of 108 pseudogenes were predicted. Functional annotation using the Rapid Annotations using Subsystems Technology (RAST) server (9–11) revealed that 75 of the genes were associated with virulence, disease, and defense, 140 genes with RNA metabolism, 196 genes with carbohydrate metabolism, and 259 genes with metabolism of amino acids and derivatives.

Accession number(s). The complete genome sequence of *Staphylococcus haemolyticus* strain SGAir0252, including the plasmid pSGAir0252, is available in GenBank under the accession numbers CP025031 (chromosome) and CP025033 (pSGAir0252).

REFERENCES

1. Lina G, Etienne J, Vandenesch F. 2000. Biology and pathogenicity of staphylococci other than *Staphylococcus aureus* and *Staphylococcus epidermis*. *In* Fischetti VA, Nivick RP, Ferretti JJ, Portnoy DA, and Rood JI (ed), Gram-positive pathogens. ASM Press, Washington, DC.

2. Ing MB, Baddour LM, Bayer AS. 1997. Bacteremia and infective endocarditis: pathogenesis, diagnosis, and complications, p 331–354. *In* Crossley KB, Archer GL (ed), The staphylococci in human disease. Churchill Livingstone, New York, NY.

3. Takeuchi F, Watanabe S, Baba T, Yuzawa H, Ito T, Morimoto Y, Kuroda M, Cui L, Takahashi M, Ankai A, Baba S, Fukui S, Lee JC, Hiramatsu K. 2005. Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. J Bacteriol 187:7292–7308. https://doi.org/10.1128/JB.187.21.7292-7308.2005.

4. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.

5. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

6. Wang Z, Wu M. 2013. A phylum-level bacterial phylogenetic marker database. Mol Biol Evol 30:1258–1262. https://doi.org/10.1093/molbev/mst059.

7. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771. https://doi.org/10.1093/nar/gkv657.

8. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

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, Arrello 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. https://doi.org/10.1186/1471-2164-9-75.

10. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

11. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. *RASTtk*: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.