AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequences of Three *Staphylococcus haemolyticus* Strains Isolated from the Lung of a TGFβ1 Transgenic Mouse with Lung Fibrosis

Ahmed M. Abdel-Hamid,[a,b] Corina N. D'Alessandro-Gabazza,[a,c,d] Taro Yasuma,[c,d] Kimberly K. O. Walden,[e] Christopher J. Fields,[e] Esteban C. Gabazza,[a,c,d] Isaac Cann[a,f,g,h,i]

[a]Carl R. Woese Institute for Genomic Biology (Microbiome Metabolic Engineering), University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[b]Department of Botany and Microbiology, Faculty of Science, Minia University, El-Minia, Egypt
[c]Department of Immunology, Mie University Faculty and Graduate School of Medicine, Tsu, Mie, Japan
[d]Center for Intractable Diseases, Mie University, Tsu, Mie, Japan
[e]Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[f]Department of Animal Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[g]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[h]Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[i]Center for East Asian & Pacific Studies, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Ahmed M. Abdel-Hamid, Corina N. D'Alessandro-Gabazza, and Taro Yasuma contributed equally to this article. Author order was determined alphabetically.

ABSTRACT We report here the complete genome sequences of three *Staphylococcus haemolyticus* strains isolated from a mouse fibrotic lung tissue and exhibiting proapoptotic activity on human lung alveolar epithelial cells. The genomes were obtained from a combination of Illumina MiSeq and Oxford Nanopore MinION sequencing.

Idiopathic pulmonary fibrosis (IPF) is a disease of unknown etiology, and the lung microbiome of IPF patients is characterized by increases in *Staphylococcus* and *Streptococcus* genera (1). We demonstrated that corisin, a peptide in diverse staphylococci, induces apoptosis of lung alveolar epithelial cells, a hallmark of IPF (2). Tissues from IPF patients and from transforming growth factor β1 (TGFβ1) transgenic (TG) mice with lung fibrosis enrich for halophilic bacteria (3). To isolate proapoptotic bacteria associated with IPF, tissue samples from the fibrotic lung of TGFβ1 TG mice (2) were inoculated into the high-salt ATCC medium 1097 and incubated at 37°C with shaking (220 rpm) until growth was visible. Bacterial colonies were obtained by plating the enriched culture on ATCC 1097 agar plates. One colony was further purified by streaking, and three colonies designated 1b, 7b, and 12b were picked, demonstrated to be proapoptotic, and subjected to genome sequencing. After growth in ATCC 1097 liquid medium for 24 h, genomic DNA was extracted using Qiagen Genomic-tip 20/g. Genome sequencing was then carried out by the Carver Biotech Center (UIUC) using paired-end short-read Illumina MiSeq and long-read MinION technologies (ONT). The shotgun genomic libraries, prepared with the HyperPrep Library construction kit from Kapa Biosystems (Roche), were pooled, quantitated, and sequenced on one MiSeq nano flow cell using a MiSeq 500-cycle sequencing kit (v2). Fastq files were demultiplexed with the bcl2fastq conversion software (Illumina; v2.20), and adaptors were trimmed from the 3′ ends of the reads. For ONT long-read sequencing, unsheared genomic DNA (1 μg) was converted into a barcoded Nanopore library with the NBD-114 and 1D library kit SQK-LSK109 from ONT and pooled for sequencing on a SpotON R9.4.1 FLO-MIN106 flow cell (GridIon X5 sequencer).

Illumina MiSeq reads and ONT long reads were checked for quality prior to and after trimming using FastQC (v0.11.8). MiSeq reads were trimmed using Trimmomatic v0.38,

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Address correspondence to Esteban C. Gabazza, gabazza@doc.medic.mie-u.ac.jp, or Isaac Cann, icann@illinois.edu.

The authors declare no conflict of interest.

Received 7 December 2021
Accepted 15 February 2022
Published 7 March 2022

TABLE 1 Genome characteristic of three of *Staphylococcus haemolyticus* strains isolated from lungs of TGF β1 TG mice with lung fibrosis

Characteristic	Data for *Staphylococcus haemolyticus* strain:		
	1b	7b	12b
Whole-genome characteristics			
BioSample accession no.	SAMN18212548	SAMN18212549	SAMN18212550
Total no. of reads (Illumina)	173,026	138,020	146,291
SRA accession no. (Illumina)	SRX13955341	SRX13955342	SRX13955343
Total no. of reads (ONT)	136,577	243,795	256,825
SRA accession no. (ONT)	SRX13955344	SRX13955345	SRX13955346
Oxford Nanopore N_{50} (bp)	11,643	11,730	11,624
Genome size (no. of bp)	2,390,652	2,398,463	2,382,565
GC content (%)	32.83	32.76	32.76
Total no. of genes	2,373	2,382	2,364
No. of chromosomes + plasmids	1 + 3	1 + 3	1 + 2
N_{50} (bp)	2,339,731	2,341,142	2,339,810
Chromosome			
GenBank accession no.	CP071512	CP071508	CP071505
Illumina short-read coverage	30×	30×	36×
Chromosome size (bp)	2,339,731	2,341,142	2,339,810
Chromosome topology	Circular	Circular	Circular
Total no. of chromosomal genes	2,317	2,319	2,316
No. of proteins	2,132	2,133	2,130
No. of rRNAs	19	19	19
No. of tRNAs	62	62	63
No. of other RNAs	4	4	4
No. of pseudogenes	100	101	100
Plasmids			
Plasmid name	pSH_1b_1	pSH_7b_1	pSH_12b_1
Accession no.	CP071513	CP071509	CP071506
Illumina short-read coverage	2,100×	1,700×	1,800×
Size (bp)	41,143	39,811	38,899
GC content (%)	31.7	31.0	30.4
No. of genes	44	44	44
Plasmid name	pSH_1b_2	pSH_7b_2	pSH_12b_2
Accession no.	CP071514	CP071510	CP071507
Illumina short-read coverage	14,600×	5,000×	18,900×
Size (bp)	5,922	13,654	3,856
GC content (%)	52.9	31.0	31.2
No. of genes	8	15	4
Plasmid name	pSH_1b_3	pSH_7b_3	
Accession no.	CP071515	CP071511	
Illumina short-read coverage	22,000×	17,800×	
Size (bp)	3,856	3,856	
GC content (%)	31.2	31.2	
No. of genes	4	4	

retaining reads longer than 30 bp (4). Long reads were adapter trimmed with Porechop (v0.2.3) and length filtered to a minimum of 1 kb with seqtk (v1.3). Unicycler v0.4.8 assembled the trimmed MiSeq and uncorrected ONT reads in a hybrid assembly using the default "normal" mode (5). Within Unicycler, the MiSeq reads were assembled with SPAdes (v3.11.1) (6), and the resulting long-anchor contigs were assembled together with the ONT reads by miniasm (7) and Racon (v0.5.0) (8). Pilon (v1.23) was used within Unicycler to iteratively polish the assembly with the MiSeq reads (9). Assemblies were evaluated for completeness using BUSCO (v3.0.1) (10) and QUAST (11) with GCF_002954055.1 (*Staphylococcus epidermidis*) as a reference. The sequence reads from each of strains 1b and 7b were assembled by Unicycler as a single circular chromosome and three circular plasmids (Bandage assembly graph viewer) (12). The reads from strain 12b were assembled as a

single circular chromosome and two circular plasmids. The circularized chromosomes and plasmids were reoriented automatically by the Unicycler assembler to start with the replication initiation proteins DnaA and RepA, respectively. The genome assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (13). The 16S rRNA genes identified the bacteria as *Staphylococcus haemolyticus*, and they were designated strains 1b, 7b, and 12b, respectively. Detailed information for the genomes is provided in Table 1.

Data availability. The sequences have been deposited in the GenBank database under BioProject accession number PRJNA707584. The Sequence Read Archive (SRA) for the raw reads and the NCBI RefSeq accession numbers are provided in Table 1.

ACKNOWLEDGMENTS

The research reported herein was supported in part by funding support to the Microbiome Metabolic Engineering Theme (Carl R. Woese Institute for Genomic Biology), a gift from the Charles and Margaret Levin Family Foundation in support of research on pulmonary diseases, a grant (20K08564) from the Japan Society for the Promotion of Science, and a grant from the Takeda Science Foundation (2019).

REFERENCES

1. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, Moore BB, White ES, Flaherty KR, Huffnagle GB, Martinez FJ. 2014. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. Lancet Respir Med 2:548–556. https://doi.org/10.1016/S2213-2600(14)70069-4.
2. D'Alessandro-Gabazza CN, Kobayashi T, Yasuma T, Toda M, Kim H, Fujimoto H, Hataji O, Takeshita A, Nishihama K, Okano T, Okano Y, Nishii Y, Tomaru A, Fujiwara K, D'Alessandro VF, Abdel-Hamid AM, Ren Y, Pereira GV, Wright CL, Hernandez A, Fields CJ, Yau PM, Wang S, Mizoguchi A, Fukumura M, Ohtsuka J, Nosaka T, Kataoka K, Kondoh Y, Wu J, Kawagishi H, Yano Y, Mackie RI, Cann I, Gabazza EC. 2020. A *Staphylococcus* pro-apoptotic peptide induces acute exacerbation of pulmonary fibrosis. Nat Commun 11:1539. https://doi.org/10.1038/s41467-020-15344-3.
3. D'Alessandro-Gabazza CN, Méndez-García C, Hataji O, Westergaard S, Watanabe F, Yasuma T, Toda M, Fujimoto H, Nishihama K, Fujiwara K, Taguchi O, Kobayashi T, Mackie RI, Cann I, Gabazza EC. 2018. Identification of halophilic microbes in lung fibrotic tissue by oligotyping. Front Microbiol 9:1892. https://doi.org/10.3389/fmicb.2018.01892.
4. 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.
5. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko 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.
7. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
8. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
9. 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.
10. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. https://doi.org/10.1093/molbev/msx319.
11. 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.
12. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
13. 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.