Complete Genome Sequence of *Megasphaera stantonii* AJH120T, Isolated from a Chicken Cecum

Joel J. Maki,[a,b] Torey Looft[a]

aFood Safety and Enteric Pathogens Research Unit, National Animal Disease Center, Agricultural Research Service, United States Department of Agriculture, Ames, Iowa, USA

bInterdepartmental Microbiology Graduate Program, Iowa State University, Ames, Iowa, USA

ABSTRACT A novel bacterium, *Megasphaera stantonii* (strain AJH120T), was isolated from the cecum of a chicken during a screen for microorganisms using a brain heart infusion (BHI) medium. PacBio and Illumina MiSeq sequencing technologies were used to produce a single contiguous chromosome. The resulting genome sequence is 2,652,760 bp long with a 52.62% GC content.

The genus *Megasphaera* comprises butyrate-producing organisms within the *Firmicutes* phylum. Members of the *Megasphaera* genus have been isolated from a variety of anaerobic environments, including the intestinal tracts of pigs, cows, and humans (1–5). Here, we describe the complete genome sequence of a recently characterized *Megasphaera* species, isolated from the cecal contents of a healthy white leghorn chicken. The isolation, cultivation, and biochemical properties of *Megasphaera stantonii* AJH120T were previously described (6). Anaerobic commensal bacteria, such as *Megasphaera* spp., may play an important role in host energy acquisition, and beneficial functions, such as butyrate production, could be exploited for poultry nutrition (7).

Genomic DNA of strain AJH120T was extracted from a 48-h culture grown in peptone-yeast-fructose broth at 42°C using the PureLink genomic DNA extraction minikit (Invitrogen) according to the manufacturer's instructions. A DNA library (350-bp insertion size) was generated from 100 ng of extracted DNA using the TruSeq Nano DNA library preparation kit (Illumina) and sequenced on a MiSeq machine utilizing v3 chemistry (Illumina). High-molecular-weight genomic DNA was extracted using the Qiagen Genomic-tip 20/G genomic DNA extraction kit according to the manufacturer's instructions. From that extracted DNA, 10 μg was submitted to the Yale University genomic facility for single-molecule real-time (SMRT) sequencing (Pacific Biosciences) to generate long sequencing reads. A total of 458,687 PacBio subreads were obtained with an N_{50} value of 20,052 bp. For *de novo* assembly of the PacBio reads, Canu v1.5 (8) was used on the raw concatenated reads using an estimated genome size of 2.8 Mb and the default settings. Canu corrected, trimmed, and assembled the PacBio reads to generate one 2,666,587-bp contig with a 52.60% GC content. Circularization of the contig was conducted with the program Circlator v1.5 using default settings (9). Circlator identified and trimmed the overlap at the contig ends, resulting in a genome that was 2,652,026 bp long and had a GC content of 52.62%.

To polish the genome sequence, the short MiSeq reads were used to correct errors in the PacBio data. Illumina MiSeq reads were processed with Trimmomatic v0.36 (10) using the default settings for paired-end mode to remove adaptors and sequences that had a quality score of <15 and a length of <36 bp. The Burrows-Wheeler aligner "mem" algorithm (BWA-mem) v0.7.12 (11) was used to map the Illumina reads to the genome contig, producing an alignment file. SAMtools v1.4.1 (12) was used to sort and index the alignment file. Pilon v1.18 (13) was used to correct the genome contig using the option

"–fix bases," resulting in a final genome sequence of 2,652,760 bp with a 52.62% GC content.

The corrected reads were annotated using the PathoSystems Resource Integration Center (PATRIC) online platform (14) to generate an annotated contiguous genome sequence. The PATRIC annotation of the genome identified 2,568 protein-coding sequences, 851 of which are hypothetical proteins. There are 57 tRNA genes and 6 rRNA operons in the genome. AJH120 has a complete butyrate production pathway and several antimicrobial resistance genes, including those for tetracycline, lincosamide, macrolide, and fosfomycin resistance.

Data availability. This complete genome sequence has been deposited in GenBank under the accession no. CP029462. The version described in this report is the first version.

ACKNOWLEDGMENTS

REFERENCES

1. Chaplin AV, Polyakova SI, Kochetkova TO, Sokolova SR, Goltsov AY, Trofimov DY, Efimov BA. 2018. Noncontiguous finished genome sequence of Megasphaera sp. ASD88, isolated from feces of child with autism spectrum disorder. New Microbes New Infect 22:13–16. https://doi.org/10.1016/j.nmni.2017.12.010.

2. Jeon BS, Kim S, Sang B-I. 2017. *Megasphaera hexanoica* sp. nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. Int J Syst Evol Microbiol 67:2114–2120. https://doi.org/10.1099/ijsem.0.001888.

3. Lanjekar VB, Marathe NP, Ramana VV, Shouche YS, Ranade DR. 2014. *Megasphaera indica* sp. nov., an obligate anaerobic bacteria isolated from human faeces. Int J Syst Evol Microbiol 64:2250–2256. https://doi.org/10.1099/ijs.0.059816-0.

4. Rogosa M. 1971. Transfer of *Peptostreptococcus elsdenii* Gutierrez et al. to a new genus, *Megasphaera* [*M. elsdenii* (Gutierrez et al.) comb. nov.]. Int J Syst Bacteriol 21:187–189. https://doi.org/10.1099/00207713-21-2-187.

5. Stanton TB, Humphrey SB. 2003. Isolation of tetracycline-resistant *Megasphaera elsdenii* strains with novel mosaic gene combinations of *tet*(O) and *tet*(W) from swine. Appl Environ Microbiol 69:3874–3882. https://doi.org/10.1128/AEM.69.7.3874-3882.2003.

6. Maki JJ, Looft T. 2018. *Megasphaera stantonii* sp. nov., a butyrate-producing bacterium isolated from the cecum of a healthy chicken. Int. J Syst Evol Microbiol https://doi.org/10.1099/ijsem.0.002991.

7. Eeckhaut V, Van Immerseel F, Croubels S, De Baere S, Haesebrouck F, Ducatelle R, Louis P, Vandamme P. 2011. Butyrate production in phylogenetically diverse *Firmicutes* isolated from the chicken caecum. Microb Biotechnol 4:503–512. https://doi.org/10.1111/j.1751-7915.2010.00244.x.

8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

9. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.

10. 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.

11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

13. 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.

14. Snyder EE, Kampanya N, Lu J, Nordberg EK, Karur HR, Shukla M, Soneja J, Tian Y, Xue T, Yoo H, Zhang F, Dharmanolla C, Dongre NV, Gillespie JJ, Hamelius J, Hance M, Huntington KI, Jukneliene D, Koziski J, Mackasmiel L, Mane SP, Nguyen V, Purkayastha A, Shallom J, Yu G, Guo Y, Gabbard J, Hix D, Azad AF, Baker SC, Boyle SM, Khudyakov Y, Meng XJ, Rupprecht C, Vinje J, Crasta OR, Czar MJ, Dickerman A, Eckart JD, Kenyon R, Will R, Setubal JC, Sobral BWS. 2007. PATRIC: the VBI PathoSystems Resource Integration Center. Nucleic Acids Res 35:D401–D406. https://doi.org/10.1093/nar/gkl858.