Draft Genome Sequence of *Komagataeibacter rhaeticus* Strain AF1, a High Producer of Cellulose, Isolated from Kombucha Tea

Renato Augusto Corrêa dos Santos,[a] Andresa A. Berretta,[b,c] Hernane da Silva Barud,[d] Sidney José Lima Ribeiro,[d] Laura Natalia González-García,[e] Tiago Domingues Zucchi,[f] Gustavo H. Goldman,[a,b] Diego M. Riaño-Pachón[a]

Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Campinas, SP, Brazil[a]; Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil[b]; Laboratório de Pesquisa, Desenvolvimento e Inovação, Apis Flora Indl. Coml. Ltda., Ribeirão Preto, SP, Brazil[c]; Instituto de Química, Universidade Estadual Paulista (UNESP), Araraquara, SP, Brazil[d]; Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá D.C., Colombia[e]; Laboratório de Microbiologia Ambiental, EMBRAPA Meio Ambiente, Jaguariúna, SP, Brazil[f]

Here, we present the draft genome sequence of *Komagatabaeicter rhaeticus* strain AF1, which was isolated from Kombucha tea and is capable of producing high levels of cellulose.

Address correspondence to Diego M. Riaño-Pachón, diego.riano@bioetanol.org.br, or Gustavo H. Goldman, ggoldman@usp.br.

*K*omagataeibacter rhaeticus AF1, previously known as *Gluconacetobacter rhaeticus* (1), is a Gram-negative rod isolated from Kombucha tea. Briefly, for the isolation of *K. rhaeticus* AF1, 1 ml of Kombucha tea was subjected to serial 10-fold dilutions in 0.85% sterile NaCl solution. Aliquots of each dilution were plated on petri dishes containing Hestrin and Schramm (HS) medium (2). The plates were incubated aerobically at 30°C for 5 days. Suspected colonies were seeded in HS medium and incubated again as described above. After the incubation period, the colonies were transferred to test tubes (20×150 mm) containing HS broth and incubated aerobically at 30°C for five to seven days, in order to evaluate the production of cellulose, which can be easily observed on the surface of the culture medium. The AF1 isolate produced 3.0 g/liter of cellulose.

Here, we present the genome sequence of *Komagataeibacter rhaeticus* strain AF1. This genome was sequenced on the Illumina HiSeq2000 system, generating 44,413,164 paired-end reads of 100 bp (insert size, 250 bp). The reads were preprocessed with the Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), resulting in 43,205,995 paired-end reads. The genome size was estimated to be 4.98 Mbp based on k-mer count statistics (3), with an estimated coverage of 1,561×. Chromosomal assembly was optimized by eliminating reads coming from plasmids, using BLAST (4) and Bowtie (5). The resulting data set was subsampled to a genome coverage of approximately 100×, and this subset was assembled using SPAdes (6, 7). The presence of typical bacterial marker genes was assessed using Amphora2 (8). The remaining reads (raw data set without plasmid-originated reads) were used to extend the contigs and perform scaffolding using SSPACE Basic (9). The Post Assembly Genome Improvement Toolkit was used in order to close gaps and correct substitution and insertion/deletion errors (10). The resulting assembly has 219 scaffolds, with a total length of 3,944,291 bp and an N_{50} of 73,183 bp. The average G+C content of the genome is 62.44%, which is similar to those of related species: *K. xylinus* G+C content, 62.1% (11); *K. hansenii* G+C content, 59.5% (12); *K. europaeus* 5P3 G+C content, 61.5%,

and *K. oboediens* 174Bp2 G+C content, 61.3% (13); and *Gluconacetobacter diazotrophicus* G+C content, 66.4% (14). Gene prediction was carried out with the NCBI Prokaryotic Genome Annotation Pipeline. A total of 3,460 genes were identified. Of these, 3,358 are protein-encoding genes, and there are 33 pseudogenes, 10 rRNA genes, 58 tRNA genes, and 1 noncoding RNA (ncRNA) gene. The gene content is similar to those of related species, i.e., *K. xylinus* (3,674 genes), *K. hansenii* (3,308 genes), *K. europaeus* 5P3 (3,939 genes), *K. oboediens* 174Bp2 (4,076 genes), and *G. diazotrophicus* (3,472 genes). A search against the UniProt database revealed 3,294 protein-encoding genes with strong sequence similarity hits to proteins in that database. The current genome assembly provides a preliminary landscape of the genomic and metabolic capabilities of *K. rhaeticus*.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JDTI00000000. The version described in this paper is version JDTI01000000.

REFERENCES

1. **Yamada Y, Yukphan P, Lan Vu HT, Muramatsu Y, Ochaikul D, Tanasupawat S, Nakagawa Y.** 2012. Description of Komagataeibacter gen. nov., with proposals of new combinations (Acetobacteraceae). J. Gen. Appl. Microbiol. **58:**397–404. http://dx.doi.org/10.2323/jgam.58.397.
2. **Hestrin S, Schramm M.** 1954. Synthesis of cellulose by Acetobacter xylinum. II. Preparation of freeze-dried cells capable of polymerizing glucose to cellulose. Biochem. J. **58:**345–352.
3. **Marçais G, Kingsford C.** 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27:**764–770. http://dx.doi.org/10.1093/bioinformatics/btr011.
4. **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.

5. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10:**R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25.

6. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

7. **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. http://dx.doi.org/10.1089/cmb.2012.0021.

8. **Wu M, Scott AJ.** 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics **28:**1033–1034. http://dx.doi.org/10.1093/bioinformatics/bts079.

9. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.

10. **Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD.** 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat. Protoc. **7:**1260–1284. http://dx.doi.org/10.1038/nprot.2012.068.

11. **Ogino H, Azuma Y, Hosoyama A, Nakazawa H, Matsutani M, Hasegawa A, Otsuyama K, Matsushita K, Fujita N, Shirai M.** 2011. Complete genome sequence of NBRC 3288, a unique cellulose-nonproducing strain of Gluconacetobacter xylinus isolated from vinegar. J. Bacteriol. **193:**6997–6998. http://dx.doi.org/10.1128/JB.06158-11.

12. **Iyer PR, Geib SM, Catchmark J, Kao TH, Tien M.** 2010. Genome sequence of a cellulose-producing bacterium, Gluconacetobacter hansenii ATCC 23769. J. Bacteriol. **192:**4256–4257. http://dx.doi.org/10.1128/JB.00588-10.

13. **Andres-Barrao C, Falquet L, Calderon-Copete SP, Descombes P, Ortega Pérez R, Barja F.** 2011. Genome sequences of the high-acetic acid-resistant bacteria Gluconacetobacter europaeus LMG 18890T and G. europaeus LMG 18494 (reference strains), G. europaeus 5P3, and Gluconacetobacter oboediens 174Bp2 (isolated from vinegar). J. Bacteriol. **193:**2670–2671. http://dx.doi.org/10.1128/JB.00229-11.

14. **Bertalan M, Albano R, de Pádua V, Rouws L, Rojas C, Hemerly A, Teixeira K, Schwab S, Araujo J, Oliveira A, França L, Magalhães V, Alquéres S, Cardoso A, Almeida W, Loureiro MM, Nogueira E, Cidade D, Oliveira D, Simão T, Macedo J, Valadão A, Dreschsel M, Freitas F, Vidal M, Guedes H, Rodrigues E, Meneses C, Brioso P, Pozzer L, Figueiredo D, Montano H, Junior J, de Souza Filho G, Flores M, Ferreira B, Branco A, Gonzalez P, Guillobel H, Lemos M, Seibel L, Macedo J, Alves-Ferreira M, Sachetto-Martins G, Coelho A, Santos E, Amaral G, Neves A, Pacheco AB, Carvalho D, Lery L, Bisch P, Rössle SC, Urményi T, Rael Pereira A, Silva R, Rondinelli E, von Krüger W, Martins O, Baldani JI, Ferreira PC.** 2009. Complete genome sequence of the sugarcane nitrogen-fixing endophyte Gluconacetobacter diazotrophicus Pal5. BMC Genomics **10:**450. http://dx.doi.org/10.1186/1471-2164-10-450.