

# Draft Genome Sequence of *Komagataeibacter intermedius* Strain AF2, a Producer of Cellulose, Isolated from Kombucha Tea

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

Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil<sup>a</sup>; Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil<sup>b</sup>; Laboratório de Pesquisa, Desenvolvimento e Inovação, Apis Flora Industrial e Comercial Ltda., Ribeirão Preto, São Paulo, Brazil<sup>c</sup>; Instituto de Química, Universidade Estadual Paulista, Araraquara, São Paulo, Brazil<sup>d</sup>; Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia<sup>e</sup>; AGRIVALLE, Salto, São Paulo, Brazil<sup>f</sup>

Here, we present the draft genome sequence of *Komagataeibacter intermedius* strain AF2, which was isolated from Kombucha tea and is capable of producing cellulose, although at lower levels compared to another bacterium from the same environment, *K. rhaeticus* strain AF1.

Received 8 October 2015 Accepted 16 October 2015 Published 3 December 2015

**Citation** dos Santos RAC, Berretta AA, Barud HDS, Ribeiro SJL, González-García LN, Zucchi TD, Goldman GH, Riaño-Pachón DM. 2015. Draft genome sequence of *Komagataeibacter intermedius* strain AF2, a producer of cellulose, isolated from Kombucha tea. *Genome Announc* 3(6):e01404-15. doi:10.1128/genomeA.01404-15.

**Copyright** © 2015 dos Santos et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Diego M. Riaño-Pachón, [diego.riano@bioetanol.org.br](mailto:diego.riano@bioetanol.org.br), or Gustavo H. Goldman, [ggoldman@usp.br](mailto:ggoldman@usp.br).

*Komagataeibacter intermedius* AF2, previously known as *Gluconacetobacter intermedius*, is a Gram-negative rod isolated from Kombucha tea. Briefly, for the isolation of *K. intermedius* AF2, 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. 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 AF2 isolate produced 1.41 g/L of cellulose, whereas *K. rhaeticus* strain AF1, isolated at the same time from the same environment, produced cellulose at higher levels (1).

Here, we present the genome sequence of *K. intermedius* strain AF2. This genome was sequenced on the Illumina HiSeq2000 system, generating 45,480,010 paired-end reads of 100 bp (insert size, 250 bp). The reads were preprocessed with Trimmomatic (2), resulting in 34,345,171 paired-end reads. Digital normalization was carried using Khmer (3), resulting in 482,978 reads, which were used to assemble contigs with SPAdes (4), using the best *k*-mer size chosen using KmerGenie (5). The post-assembly genome-improvement toolkit (PAGIT) was used to close gaps and correct substitution and insertion/deletion errors (6). An additional scaffolding step was carried out in SSPACE (7), followed by gap filling with GapFiller (8). The final assembly has a total length of 4,465,062 bp and an  $N_{50}$  of 70,565 bp, represented by 268 scaffolds. The average G+C content of the genome is 61.35%, which is similar to related species: *K. rhaeticus* AF1 (GC: 62.44%) (1); *K. xylinus* NBRC 3288 (GC: 60.92%) (9); *K. hansenii* (GC: 59.0%) (10); *K. europaeus* 5P3 (GC: 61.2%); *K. oboediens* 174Bp2 (GC: 61.3%) (11); and *Gluconacetobacter diazotrophicus* Pa1 5 (GC: 66.19%) (12). Gene prediction was carried out with the Prokka

Pipeline (13) using a nonredundant database of proteins in *Ace-tobacteriaceae* as the first annotation source. A total of 4,232 genes were identified, including 4,145 protein-encoding genes, 11 rRNA genes, 64 tRNA genes, 303 signal peptide genes, 1 tmRNA gene, and 11 ncRNA genes. Gene content is similar to related species: *K. rhaeticus* (3,460 genes), *K. xylinus* (3,195), *K. hansenii* (3,308), *K. medellinensis* (3,195), *K. europaeus* 5P3 (3,586), *K. oboediens* 174Bp2 (3,601), and *G. diazotrophicus* (3,864). A search against the UniProt database revealed 3,641 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. intermedius* strain AF2 and will provide insights about the molecular mechanisms involved in high and low cellulose production in this genus.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [JUFEX00000000](https://www.ncbi.nlm.nih.gov/nuccore/JUFEX00000000). The version described in this paper is the second version, [JUFEX02000000](https://www.ncbi.nlm.nih.gov/nuccore/JUFEX02000000).

## ACKNOWLEDGMENTS

This work was financially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). R.A.C.D.S. holds a CNPq scholarship (no. 184722/2014-4).

## REFERENCES

1. Dos Santos RAC, Berretta AA, Barud HDS, Ribeiro SJL, Gonzalez-Garcia LN, Zucchi TD, Goldman GH, Riano-Pachon DM. 2014. Draft genome sequence of *Komagataeibacter rhaeticus* strain AF1, a high producer of cellulose, isolated from Kombucha Tea. *Genome Announc* 2(4): e00731-14. <http://dx.doi.org/10.1128/genomeA.00731-14>.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
3. Crusoe MR, Edverson G, Fish J, Howe A, McDonald E, Nahum J, Nanlohy K, Ortiz-Zuazaga H, Pell J, Simpson J, Scott C, Srinivasan RR, Zhang Q, Brown CT. 2014. The khmer software package: enabling

- efficient sequence analysis. Figshare <http://dx.doi.org/10.6084/m9.figshare.979190>.
4. 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>.
  5. Chikhi R, Medvedev P. 2014. Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics* 30:31–37. <http://dx.doi.org/10.1093/bioinformatics/btt310>.
  6. 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>.
  7. 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>.
  8. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
  9. 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>.
  10. Iyer PR, Geib SM, Catchmark J, Kao T-H, 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>.
  11. Andres-Barrao C, Falquet L, Calderon-Copete SP, Descombes P, Ortega Perez 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.
  12. 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 M, Nogueira E, Cidade D, Oliveira D, Simão T, Macedo J, Valadao A, Dreschel 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 VMQ, Ferreira B, Branco A, Gonzalez P, Guillobel H, Lemos M, Seibel L, Alves-Ferreira M, Sachetto-Martins G, Coelho A, Santos E, Amaral G, Neves A, Pacheco AB, Carvalho D, Lery L. 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>.
  13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.