

High-Quality Draft Genome Sequence of the *Opitutaceae* Bacterium Strain TAV1, a Symbiont of the Wood-Feeding Termite *Reticulitermes flavipes*

Jantiya Isanapong,^{a,b} Lynne Goodwin,^c David Bruce,^c Amy Chen,^d Chris Detter,^c James Han,^d Cliff S. Han,^c Brittany Held,^c Marcel Huntemann,^d Natalia Ivanova,^d Miriam L. Land,^e Konstantinos Mavromatis,^d Matt Nolan,^d Amrita Pati,^d Len Pennacchio,^d Sam Pitluck,^d Ernest Szeto,^d Roxanne Tapia,^c Tanja Woyke,^d and Jorge L. M. Rodrigues^a

Department of Biology^a and Department of Earth and Environmental Sciences,^b University of Texas, Arlington, Texas, USA; Los Alamos National Laboratory, Los Alamos, New Mexico, USA^c; DOE Joint Genome Institute, Walnut Creek, California, USA^d; and Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA^e

Microbial communities in the termite hindgut are essential for degrading plant material. We present the high-quality draft genome sequence of the *Opitutaceae* bacterium strain TAV1, the first member of the phylum *Verrucomicrobia* to be isolated from wood-feeding termites. The genomic analysis reveals genes coding for lignocellulosic degradation and nitrogen fixation.

The *Opitutaceae* bacterium strain TAV1 was isolated from the hindgut of the wood-feeding termite *Reticulitermes flavipes* (10). This termite-associated *Verrucomicrobia* (TAV) isolate is a Gram-negative, coccoid-shaped, microaerophilic bacterium. Owing to the importance of microbial symbionts in the termite hindgut for the degradation of cellulose and hemicellulose into acetate, hydrogen, and methane (2, 3), we investigated the genetic potential of strain TAV1 for the degradation of lignocellulosic material and overall functional attributes associated with its ecological role.

The genomic DNA of strain TAV1 was isolated using a cetyltrimethylammonium bromide method (<http://my.jgi.doe.gov/general/>). Genome sequence was generated with a combination of Illumina and 454 pyrosequencing platforms. The individual reads were assembled with the Newbler assembler (Roche) and generated 82 contigs with the largest being 590 kb and the smallest contig being 530 bp. All contigs span up to the length of 7.1 Mbp, and the average GC content of the genome is approximately 63.2%. Genes were identified using Prodigal (5) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline (9). The predicted protein-coding genes (coding sequences [CDS]) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to produce a product description for each predicted protein. Noncoding genes and miscellaneous features were predicted using tRNAscan-SE (8), RNAMMer (7), Rfam (4), TMHMM (6), and signalP (1).

The draft genome contains 6,051 genes with 5,987 CDS. A total of 64 structural RNAs were identified in the genome, with the presence of one rRNA operon. Protein coding genes were classified according to the cluster of orthologous groups (COG categories). The top five functional groups were as follows: (i) general function prediction only (526 genes), (ii) transcription (476 genes), (iii) carbohydrate metabolism (406 genes), (iv) amino acid metabolism (403 genes), and (v) energy production and conversion (310 genes). Further inspection using the carbohydrate-active enzyme database (<http://www.cazy.org>) revealed that the TAV1 genome contains a large number of genes encoding glycoside hydrolases (GH), which are involved in the breakage of bonds

between two or more carbohydrate moieties. These GH enzymes contained both catalytic and carbohydrate-binding modules, such as glycoside hydrolase family 5, cellulase, endo-1,4-beta-xylanase, N-acetylglucosamine-6-phosphate deacetylase, peptidoglycan glycosyltransferase, and 1,4-alpha-glucan branching enzyme. Moreover, 13 genes associated with nitrogen fixation were identified, such as nitrogen iron reductase protein (*nifH*), nitrogenase molybdenum-iron (*nifD*), FeS assembly protein (*nifU*), and nitrogenase MoFe cofactor biosynthesis protein (*nifE*) genes, among others. In addition, the TAV1 genome contains the *cbb*₃-type cytochrome oxidase, which has high affinity for oxygen. Effective removal of O₂ is essential for the homoacetogenic and methanogenic process to occur. The presence of genes associated with lignocellulose degradation, nitrogen fixation, and oxygen consumption implies an important ecological role for strain TAV1 in the functioning of the hindgut ecosystem.

Nucleotide sequence accession number. The high-quality draft genome sequence of the *Opitutaceae* bacterium was deposited in GenBank under the accession number [AHKS00000000](https://www.ncbi.nlm.nih.gov/nuccore/AHKS00000000).

ACKNOWLEDGMENT

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

REFERENCES

1. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783–795.
2. Breznak JA, Brune A. 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entomol.* **39**:453–487.
3. Brune A, Ohkuma M. 2011. Role of the termite gut microbiota in symbiotic digestion biology of termites: a modern synthesis, p 439–475. *In* Bignell DE, Roisin Y, Lo N (ed), *Biology of termites: a modern synthesis*. Springer, Amsterdam, Netherlands.
4. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res.* **31**:439–441.

Received 23 February 2012 Accepted 28 February 2012

Address correspondence to Jorge L. M. Rodrigues, Jorge@uta.edu.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00264-12

5. Hyatt D, et al. 2010. Prodigal prokaryotic dynamic programming gene finding algorithm. *BMC Bioinformatics* 11:119.
6. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567–580.
7. Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
8. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
9. Pati A, et al. 2010. GenePRIMP: a gene prediction improvement pipeline for microbial genomes. *Nat. Methods* 7:455–457.
10. Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA. 2004. New strategies for cultivation and detection of previously uncultured microbes. *Appl. Environ. Microbiol.* 70:4748–4755.