

# Genome Sequence of *Amycolatopsis* sp. Strain ATCC 39116, a Plant Biomass-Degrading Actinomycete

Jennifer R. Davis,<sup>a</sup> Lynne A. Goodwin,<sup>b,c</sup> Tanja Woyke,<sup>b</sup> Hazuki Teshima,<sup>c</sup> David Bruce,<sup>b,c</sup> Chris Detter,<sup>b,c</sup> Roxanne Tapia,<sup>b,c</sup> Shunsheng Han,<sup>c</sup> James Han,<sup>b</sup> Sam Pitluck,<sup>b</sup> Matt Nolan,<sup>b</sup> Natalia Mikhailova,<sup>b</sup> Miriam L. Land,<sup>d</sup> and Jason K. Sello<sup>e</sup>

Department of Molecular Pharmacology, Physiology, and Biotechnology, Brown University, Providence, Rhode Island, USA<sup>a</sup>; DOE Joint Genome Institute, Walnut Creek, California, USA<sup>b</sup>; Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA<sup>c</sup>; Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA<sup>d</sup>; and Department of Chemistry, Brown University, Providence, Rhode Island, USA<sup>e</sup>

**We announce the availability of a high-quality draft of the genome sequence of *Amycolatopsis* sp. strain 39116, one of few bacterial species that are known to consume the lignin component of plant biomass. This genome sequence will further ongoing efforts to use microorganisms for the conversion of plant biomass into fuels and high-value chemicals.**

There is growing interest in the use of microorganisms for the conversion of plant biomass into fuels and high-value chemicals (13). While most research has focused on fungi, there have been studies of bacteria that catabolize plant biomass (9), among them *Amycolatopsis* sp. strain ATCC 39116 (previously known as *Streptomyces setonii*). In the 1980s, Crawford and others established that this soil-dwelling actinobacterium can depolymerize lignin (1, 12) and catabolize the resulting aromatic compounds, including benzoate, catechol, gentisate, guaiacol, *p*-coumarate, protocatechuate, ferulate, and vanillin (14). There has been a resurgence of interest in the biotechnological applications of *Amycolatopsis* sp. ATCC 39116, as evidenced by the recent use of next-generation sequencing and functional proteomics to discover enzymes that depolymerize lignin (3).

A high-quality draft of the *Amycolatopsis* sp. ATCC 39116 genome sequence was generated at the DOE Joint Genome Institute (JGI) using a combination of Illumina (2) and 454 technologies (10). The Illumina GAII shotgun library generated 41,277,361 reads totaling 3,137.1 Mb. The 454 Titanium standard library yielded 272,738 reads, and a paired-end 454 library (with an average insert size of 7 kb) generated 936,887 reads, totaling 258.5 Mb. The Illumina sequencing data were assembled with Velvet, version 1.0.13 (15), while the combined 454 data were assembled with Newbler, version 2.3. The 454 and Illumina assemblies were integrated using parallel Phrap, version SPS-4.24 (High Performance Software, LLC). Illumina data were used to correct potential base errors and increase consensus quality using the software Polisher, developed at the JGI (Alla Lapidus, unpublished). Possible misassemblies were corrected using Gap Resolution (Cliff Han, unpublished) or Dupfinisher (7) or by sequencing cloned bridging PCR fragments via subcloning. Gaps between contigs were closed by editing in the software Consed (4, 5, 6), by PCR, and by bubble PCR (J.-F. Cheng, unpublished) primer walks. The final assembly is based on 153.3 Mb of 454 draft data, which provides an average of 18.2 $\times$  coverage of the genome, and 3,160.6 Mb of Illumina draft data, which provides an average of 376.3 $\times$  coverage of the genome.

The total genome size is 8,442,518 bp with a G+C content of 71.9%. Prodigal software (8) and the JGI GenePRIMP pipeline (11) were used to identify 8,264 candidate protein-encoding genes. Annotations using the NCBI nonredundant database and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and Inter-

Pro databases were completed, and the results can be found at <http://img.jgi.doe.gov>.

Consistent with the reported lignin catabolism of *Amycolatopsis* sp. ATCC 39116 is the presence of genes encoding putative lignin-depolymerizing enzymes, such as heme peroxidases, laccases, catalases, and oxidases. Likewise, there are genes encoding canonical pathways for catabolism of catechol, benzoate, protocatechuate, phenylacetate, and methylated aromatic compounds. Curiously, the organism lacks genes encoding cellulases but has many others encoding carbohydrate-degrading enzymes. We anticipate that mining of this genome will yield new insights into bacterial catabolism of plant biomass and the identities of genes that can be used in the engineering of a lignocellulose biorefinery.

**Nucleotide sequence accession number.** The genome sequence of *Amycolatopsis* sp. ATCC 39116 has been deposited in GenBank under accession no. [AFWY000000000](https://www.ncbi.nlm.nih.gov/nuccore/AFWY000000000).

## ACKNOWLEDGMENTS

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. In addition, this work was generously supported by National Science Foundation research grants (MCB-09020713 and MCB-1053319) and a SEED award from the Office of the Vice President for Research at Brown University to J.K.S. Support for J.R.D. comes from a National Science Foundation Graduate Research Fellowship.

## REFERENCES

1. Antai SP, Crawford DL. 1981. Degradation of softwood, hardwood, and grass lignocelluloses by two *Streptomyces* strains. *Appl. Environ. Microbiol.* 42:378–380.
2. Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* 5(4):433–438.
3. Brown ME, Walker MC, Nakashige TG, Iavarone AT, Chang MCY. 2011. Discovery and characterization of heme enzymes from unsequenced bacteria: application to microbial lignin degradation. *J. Am. Chem. Soc.* 133(45):18006–18009.
4. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186–194.

Received 6 February 2012 Accepted 13 February 2012

Address correspondence to Jason K. Sello, [jason\\_sello@brown.edu](mailto:jason_sello@brown.edu).

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

doi:10.1128/JB.00186-12

5. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175–185.
6. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8:195–202.
7. Han C, Chain P. 2006. Finishing repeat regions automatically with Dupfinisher, p 141–146. *In* Arabnia HR, Valafar H (ed), Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology, June 26–29, 2006. CSREA Press, Las Vegas, NY.
8. Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
9. Kirby R. 2006. Actinomycetes and lignin degradation. *Adv. Appl. Microbiol.* 58:125–168.
10. Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
11. Pati A, et al. 2010. GenePRIMP: a gene prediction improvement pipeline for microbial genomes. *Nat. Methods* 7:455–457.
12. Pometto AL, Crawford DL. 1986. Catabolic fate of *Streptomyces viridosporus* T7A-produced, acid-precipitable polymeric lignin upon incubation with ligninolytic *Streptomyces* species and *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 51:171–179.
13. Rubin EM. 2008. Genomics of cellulosic biofuels. *Nature* 454:841–845.
14. Sutherland JB. 1986. Demethylation of veratrole by cytochrome P-450 in *Streptomyces setonii*. *Appl. Environ. Microbiol.* 52:98–100.
15. Zerbino D, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.