genomeA
Journals.ASM.org

Genome Sequence of *Streptomyces viridosporus* Strain T7A ATCC 39115, a Lignin-Degrading Actinomycete

Jennifer R. Davis,[a] Lynne Goodwin,[b] Hazuki Teshima,[b] Chris Detter,[b] Roxanne Tapia,[b] Cliff Han,[b] Marcel Huntemann,[c] Chia-Lin Wei,[c] James Han,[c] Amy Chen,[c] Nikos Kyrpides,[c] Kostas Mavrommatis,[c] Ernest Szeto,[c] Victor Markowitz,[c] Natalia Ivanova,[c] Natalia Mikhailova,[c] Galina Ovchinnikova,[c] Ioanna Pagani,[c] Amrita Pati,[c] Tanja Woyke,[c] Sam Pitluck,[c] Lin Peters,[c] Matt Nolan,[c] Miriam Land,[d] Jason K. Sello[e]

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

We announce the availability of the genome sequence of *Streptomyces viridosporus* strain T7A ATCC 39115, a plant biomass-degrading actinomycete. This bacterium is of special interest because of its capacity to degrade lignin, an underutilized component of plants in the context of bioenergy. It has a full complement of genes for plant biomass catabolism.

Received 15 May 2013 **Accepted** 24 May 2013 **Published** 5 July 2013

Citation Davis JR, Goodwin L, Teshima H, Detter C, Tapia R, Han C, Huntemann M, Wei C-L, Han J, Chen A, Kyrpides N, Mavrommatis K, Szeto E, Markowitz V, Ivanova N, Mikhailova N, Ovchinnikova G, Pagani I, Pati A, Woyke T, Pitluck S, Peters L, Nolan M, Land M, Sello JK. 2013. Genome sequence of *Streptomyces viridosporus* strain T7A ATCC 39115, a lignin-degrading actinomycete. Genome Announc. 1(4):e00416-13. doi:10.1128/genomeA.00416-13.

Address correspondence to Jason K. Sello, jason_sello@brown.edu.

In the search for fossil fuel alternatives, much effort has been devoted to the discovery and development of microorganisms that can efficiently convert plant biomass into biofuels and commodity chemicals (1, 2). Many microorganisms have been identified that can degrade the carbohydrate components in plants. However, few are known that also depolymerize and consume lignin, a structural polymer in plant cell walls. One such organism is *Streptomyces viridosporus* strain T7A ATCC 39115, a soil-dwelling actinomycete (3–7). Importantly, it is the source of the first bacterial peroxidase that depolymerizes lignin (8–12). The absence of information on the genes encoding this lignin peroxidase and other enzymes responsible for lignin degradation and their biotechnological potential motivated our efforts to sequence the *S. viridosporus* genome.

The draft genome sequence of *S. viridosporus* was generated at the Department of Energy (DOE) Joint Genome Institute (JGI) using a combination of Illumina (13) and 454 technologies (14). The Illumina GAii shotgun library generated 67,837,180 reads totaling 5,155.6 Mb. The 454 Titanium standard library generated 228,388 reads, and the paired-end 454 library (with an average insert size of 8 kb) generated 635,872 reads, totaling 223.4 Mb of 454 data. The 454 data were assembled together with Newbler, version 2.3, while Illumina sequencing data were assembled with Velvet, version 1.0.13 (15). 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 (A. Lapidus, unpublished). Possible misassemblies were corrected using gapResolution (C. Han, unpublished) or Dupfinisher (16), or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed (17–19), by PCR, and by bubble PCR (J.-F. Cheng, unpublished) primer walks. A total of 442 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The estimated genome size is 8.3 Mb and the final assembly is based on 163.6 Mb of 454 draft data, which provide an average 19.7× coverage of the genome, and 5,006.8 Mb of Illumina draft data, which provide an average 603.2× coverage of the genome.

The total genome size is 8,278,598 bp with a G+C content of 72.5%. Prodigal software (20) and the JGI GenePRIMP pipeline (21) were used to identify 7,552 candidate protein-encoding genes. Annotations using the NCBI nonredundant database, UniProt, TIGRFam, Pfam, Priam, KEGG, COG, and InterPro databases were completed and the results can be accessed at http://img.jgi.doe.gov.

The genome contains numerous genes encoding homologs of enzymes that deconstruct plant biomass. COG annotation showed that 8.35% of the predicted proteins are involved in carbohydrate transport and metabolism. Genes encoding putative lignin-degrading enzymes, such as heme peroxidases, Dyp-type peroxidases, and catalases, were identified and are currently being analyzed. Pathways for the catabolism of lignin-derived aromatic compounds, such as protocatechuate, benzoate, and catechol, were also identified. Based on these findings, we anticipate that *S. virdosporus* or its complement of genes for plant catabolism could constitute a lignocellulose biorefinery.

Nucleotide sequence accession number. The genome sequence of *Streptomyces viridosporus* has been deposited in GenBank under accession no. AJFD00000000.

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. came from a National Science Foundation Graduate Research Fellowship.

REFERENCES

1. **Rubin EM.** 2008. Genomics of cellulosic biofuels. Nature **454:**841–845.
2. **Bugg TD, Ahmad M, Hardiman EM, Singh R.** 2011. The emerging role for bacteria in lignin degradation and bio-product formation. Curr. Opin. Biotechnol. **22:**394–400.
3. **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.
4. **Borgmeyer JR, Crawford DL.** 1985. Production and characterization of polymeric lignin degradation intermediates from two different *Streptomyces* spp. Appl. Environ. Microbiol. **49:**273–278.
5. **Antai SP, Crawford DL.** 1981. Degradation of softwood, hardwood, and grass lignocelluloses by two *Streptomyces* strains. Appl. Environ. Microbiol. **42:**378–380.
6. **Davis JR, Sello JK.** 2010. Regulation of genes in *Streptomyces* bacteria required for catabolism of lignin-derived aromatic compounds. Appl. Microbiol. Biotechnol. **86:**921–929.
7. **Crawford DL, Pometto AL, Crawford RL.** 1983. Lignin degradation by *Streptomyces viridosporus*: isolation and characterization of a new polymeric lignin degradation intermediate. Appl. Environ. Microbiol. **45:** 898–904.
8. **Gottschalk LM, Bon EP, Nobrega R.** 2008. Lignin peroxidase from *Streptomyces viridosporus* T7A: enzyme concentration using ultrafiltration. Appl. Biochem. Biotechnol. **147:**23–32.
9. **Nascimento HJ, Silva JG, Jr.** 2008. Purification of lignin peroxidase isoforms from *Streptomyces viridosporus* T7A by hydrophobic based chromatographies. World J. Microbiol. Biotechnol. **24:**1973–1975.
10. **Le Roes-Hill M, Khan N, Burton SG.** 2011. Actinobacterial peroxidases: an unexplored resource for biocatalysis. Appl. Biochem. Biotechnol. **164:** 681–713.
11. **Ramachandra M, Crawford DL, Hertel G.** 1988. Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete *Streptomyces viridosporus*. Appl. Environ. Microbiol. **54:**3057–3063.
12. **Thomas L, Crawford DL.** 1998. Cloning of clustered *Streptomyces viridosporus* T7A lignocellulose catabolism genes encoding peroxidase and endoglucanase and their extracellular expression in *Pichia pastoris*. Can. J. Microbiol. **44:**364–372.
13. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433–438.
14. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
15. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.
16. **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 and Computational Biology, Jun 26–29. CSREA Press, Las Vegas, NV.
17. **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.
18. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.
19. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
20. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
21. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat. Methods **7:**455–457.