Complete Genome Sequence of the Lignocellulose-Degrading Actinomycete *Streptomyces albus* CAS922

Anna Tippelt,a Markus Nett,a M. Soledad Vela Gurovicb,c

aDepartment of Biochemical and Chemical Engineering, TU Dortmund University, Dortmund, Germany  
bCenter for Renewable Natural Resources of the Semi-Arid Zone, National Scientific and Technical Research Council, Bahía Blanca, Argentina  
cDepartment of Biology, Biochemistry, and Pharmacy, Universidad Nacional del Sur, Bahía Blanca, Argentina

**ABSTRACT** *Streptomyces albus* CAS922 was isolated from sunflower seed hulls. Its fully sequenced genome harbors a multitude of genes for carbohydrate-active enzymes, which likely facilitate growth on lignocellulosic biomass. Furthermore, the presence of 27 predicted biosynthetic gene clusters indicates a significant potential for the production of bioactive secondary metabolites.

Actinomycete bacteria are important contributors to the natural degradation of lignocellulosic biomass (1, 2). An illustrative example is *Streptomyces albus* CAS922, which was isolated from sunflower seed hulls discarded by the sunflower oil industry (Cargill S.A.C.I., Bahía Blanca, Argentina) by using ISP-2 agar plates (3) and has been found to grow on diverse lignocellulosic substrates. The strain produces an earthy odor and shows typical streptomycete morphology, including aerial mycelia and substrate hyphae penetrating the agar. Since many actinomycetes are versatile producers of clinically used antibiotics (4, 5), *S. albus* CAS922 might possess the potential to convert lignocellulosic biomass into value-added products. To obtain data to test this hypothesis, its genome was sequenced.

For DNA isolation, a glycerol stock of *S. albus* CAS922 was incubated in ISP-2 broth at 30°C and 250 rpm for 4 days. Genomic DNA was then isolated according to a protocol adapted from a report by Neumann et al. with minor modifications (6). Briefly, cells were harvested from liquid ISP-2 cultures and ground in liquid nitrogen in a mortar. The sample was incubated with lysozyme for 2 h at 300 rpm and proteinase K for 3 h at 300 rpm and was extracted with a mixture of phenol, chloroform, and isoamyl alcohol (25:24:1) followed by RNase treatment. Genome sequencing was executed by CD Genomics (New York, NY, USA) using the Oxford Nanopore sequencing platform. The SQK-LSK109 kit was used for library preparation following the 1D genomic DNA by ligation protocol (7). Adapters were filtered with the software Porechop v0.2.3. Low-quality and short-fragment (<2,000-bp) reads were removed with the MinKNOW software package to obtain a total of 1,543,556,228 bp of clean data from 260,187 clean reads with an N50 value of 7,942 bp. The clean reads were assembled using Canu v1.5 software (8), followed by a subsequent data-polishing step with Pilon software (9) to increase accuracy. This resulted in a linear chromosome of 8.06 Mb with 191-fold genome coverage and a G+C content of 72.59%. The Prokaryotic Genome Annotation Pipeline (PGAP) (10) predicted 6,776 protein-coding genes, 80 RNAs, 59 tRNAs, 3 noncoding RNAs, and 312 pseudogenes. A repetitive sequence content of 2.45% was predicted by RepeatMasker (11). Assignment of the strain to the species *S. albus* was verified based on alignments with the nonredundant database (12). Default parameters were used for all software.

A total of 232 carbohydrate-active enzymes were predicted using the CAZy database (13) and dbCAN2 (14). Three of these proteins belong to the AA10 family, recently
recognized as copper-dependent lytic polysaccharide monoxygenases involved in the lysis of chitin, cellulose, and xylan (15).

An antiSMASH analysis (16) revealed 27 putative biosynthetic gene clusters. According to this prediction, S. albus CAS922 produces the siderophore desferrioxamine E (17), the compatible solute ectoine (18), and the terpenes cyaslabdan (19) and geosmin (20). Moreover, the actinomycete harbors the genes for biosynthesis of the antibiotics xantholin (21) and pseudouridimycin (22). Overall, S. albus CAS922 possesses a high potential for the utilization of lignocellulose and the biosynthesis of secondary metabolites.

Data availability. This whole-genome project has been deposited in GenBank with the accession number CP048875. The raw data are available under accession number SRR11069776.

ACKNOWLEDGMENTS

M.S.V.G. acknowledges Cristian Gallo for technical assistance.

This study was supported by grants PUE2017 CERZOS-CONICET and PGI 24/B294 SeGCyT-UNS (to M.S.V.G.). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Antunes PL, Martins FL, Pereira RV, Thomas AM, Barbosa D, Lemos LN, Machado Silva GM, Silva Moura LM, Condomitti Epamino GW, Digiampietri LA, Lombardi KC, Locosque Ramos P, Bento Quaggio R, Franco de Oliveira JC, Castiglioni Pascon R, da Cruz JB, da Silva AM, Setubal JC. 2016. Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics. Sci Rep 6:38915. https://doi.org/10.1038/srep38915.

2. Větrovský T, Steffen KT, Baldrian P. 2014. Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil actinobacteria. PLoS One 9:e89108. https://doi.org/10.1371/journal.pone.0089108.

3. Shirling EB, Gottlieb D. 1966. Methods for characterization of actinobacteria. PLoS One 9:e89108. https://doi.org/10.1371/journal.pone.0089108.

4. Oxford Nanopore Technologies. 2020. Genomic DNA by ligation (SQK-LSK109)-minion.pdf.

5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/412158.

6. Neumann B, Pospiech A, Schairer HU. 1992. Rapid isolation of genomic DNA from Gram-negative bacteria. Trends Genet 8:332–333. https://doi.org/10.1016/0168-9525(92)90269-A.

7. Oxford Nanopore Technologies. 2020. Genomic DNA by ligation (SQK-LSK109). Oxford Nanopore Technologies, Oxford, United Kingdom. https://store.nanoporetech.com/us/media/wysiwyg/pdfs/SQK-LSK109/Genomic DNA_by_Ligation_SQK-LSK109_-_minion.pdf.

8. Harris MN, Squares R, Searle MG, Zurtseh R, Kosuri S, Meese JS, et al. 2015. Canu: scalable and accurate contig assembly for long-read genomes. Genome Res 25:258–268. https://doi.org/10.1101/gr.192210.115.

9. Wang K, Burton CJ, Ruan J, Li Y, Zhang G, Bibikova M, et al. 2009. Canu: error-tolerant de novo genome assembly. Genome Res 20:175–187. https://doi.org/10.1101/gr.088932.107.

10. Ferragina P, Manzini E. 2005. Compressed text indexing. ACM Comput Surv 37:193–218. https://doi.org/10.1145/1089984.1089989.

11. Ferragina P, Manzini E. 2007. Opportunistic data structures with applications. ACM Trans Database Syst 32:71–74. https://doi.org/10.1145/1240723.1240725.

12. Deng YY, Li JQ, Wu SF, Zhu YP, Chen YW, He FC. 2006. Integrated nr database in protein annotation system and its localization. Comput Eng 32:71–74.