Draft Genome Sequence of a Novel *Streptomyces griseorubens* Strain, JSD-1, Active in Carbon and Nitrogen Recycling

Haiwei Feng,[a,b,c] Yuee Zhi,[a,b,c] Yujing Sun,[a] Xing Wei,[a,b,c] Yanqing Luo,[a,b,c] Pei Zhou[a,b,c]

School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China[a]; Key Laboratory of Urban Agriculture (South), Ministry of Agriculture, Shanghai Jiao Tong University, Shanghai, China[b]; Bor S. Luh Food Safety Research Center, Shanghai Jiao Tong University, Shanghai, China[c]

Streptomyces griseorubens JSD-1, isolated from compost-treated soil, is able to utilize lignocellulose and nitrate as its sole carbon and nitrogen source for growth. Here, we announce the draft genome map of this actinomycete. The genes participating in lignocellulose and nitrate metabolism were picked out and identified.

Address correspondence to Pei Zhou, peizhousjtu@163.com.

Streptomycetes are the most numerous and ubiquitous soil bacteria (1). They are crucial in the soil environment because of their broad range of metabolic processes and biotransformations. These include degradation of the insoluble remains of other organisms, such as lignocellulose and chitin, making streptomycetes central organisms in carbon recycling. In addition, streptomycetes can utilize nitrate as a nitrogen resource for growth, indicating that they also play an important role in nitrogen recycling. Here, *Streptomyces griseorubens* JSD-1 could metabolize 88% rice straw (20.0 g/liter) and 95% nitrate (5.0 g/liter) within 10 days' cultivation. However, little is known about the genomic or genetic background related to lignocellulose and nitrate utilization of this isolate.

The actinomycete designated JSD-1 was isolated from compost-treated soil in Shanghai, China, and identified as *S. griseorubens* through 16S rRNA sequencing as well as its morphological and physiological characteristics. Then genome sequencing of JSD-1 was performed using an Illumina MiSeq platform with insert sizes of 300 bp, 360 bp, and 700 bp in paired-end as well as 3 kb and 8 kb in mate-paired libraries. Assembly of all sequence reads by applying Newbler 2.8 assembler resulted in a draft genome. Glimmer 3.0 (2) was used to predict open reading frames (ORFs) with BLASTp (3) against the Nr proteins database. The functional annotation was determined with the KEGG, COG, and Swiss-Prot databases (4–6). tRNAs, rRNAs, and other nontranslated genes were predicted using tRNAscan-SE, RNAmmer, and Rfams, respectively (7–9). CRISPRFinder was used to determine CRISPR repeats (10). The signal peptide cleavages sites, transmembrane topologies, and lipoproteins were predicted by the online programs Signal 4.0, TMHMM 2.0, and LipoP 1.0, respectively (11–13).

A total of 6,432,848 reads including up to 2,209-Mb clean data were generated, representing a 263.0-fold average coverage of the whole genome. The assembled genome contained 2 scaffolds and 246 contigs. The N_{50} length of contigs was 53,294 bp and that of scaffolds was 7,563,100 bp. Finally, we obtained the draft genome of *S. griseorubens* with a single linear chromosome of 8,463,223 bp

and an average G+C content of 72.42%. The chromosome is smaller than the ones of *S. coelicolor* A3(2) (8.7 Mbp), *S. avermitilis* MA-4680 (9.0 Mbp), or *S. griseus* IFO 13350 (8.5 Mbp) (14–16). Analysis of the genome revealed that it contained 6 rRNA operons, 66 tRNA genes, and 7,159 protein-coding sequences (CDSs). For these CDSs, 4,587 proteins could be assigned to clusters of orthologous groups (COG) families. As for the subcellular localization of the proteins, 452 proteins were identified as secreted proteins, 1,386 proteins as transmembrane proteins, and 299 proteins as transporters.

Through the analysis of RNA-Seq and quantitative real-time PCR, a variety of genes involved in the metabolism of lignocellulose and nitrate were acquired. The genes participating in lignocellulose degradation mainly included those encoding exo-1,4-β-glucanase, endo-1,4-β-glucanase, β-glucosidase, endo-1,4-β-xylanase, β-xylosidase, α-glucuronidase, α-L-arabinofuranosidase, copper oxidase, dyp-type peroxidase, catalase peroxidase, and chloroperoxidase. Meanwhile, genes for nitrate reductase, nitrite reductase, glutamine synthetase, and glutamate synthase, which are related to nitrate assimilation, were identified as well.

Nucleotide sequence accession number. The whole-genome shotgun project of *S. griseorubens* JSD-1 has been deposited at GenBank under the accession no. JJMG00000000.

ACKNOWLEDGMENTS

REFERENCES

1. **Hodgson DA.** 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. Adv. Microb. Physiol. **42:**47–238. http://dx.doi.org/10.1016/S0065-2911(00)42003-5.
2. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.

3. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

4. **Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M.** 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. **38:**D355–D360. http://dx.doi.org/10.1093/nar/gkp896.

5. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36. http://dx.doi.org/10.1093/nar/28.1.33.

6. **Bairoch A, Apweiler R.** 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. **28:**45–48. http://dx.doi.org/10.1093/nar/28.1.45.

7. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33:**W686–W689. http://dx.doi.org/10.1093/nar/gki366.

8. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

9. **Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A.** 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. **33:**D121–D124. http://dx.doi.org/10.1093/nar/gni123.

10. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52–W57. http://dx.doi.org/10.1093/nar/gkm360.

11. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8:**785–786. http://dx.doi.org/10.1038/nmeth.1701.

12. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580. http://dx.doi.org/10.1006/jmbi.2000.4315.

13. **Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC.** 2008. Methods for the bioinformatic identification of bacteria lipoproteins encoded in the genomes of Gram-positive bacteria. World J. Microbiol. Biotechnol. **24:**2377–2382. http://dx.doi.org/10.1007/s11274-008-9795-2.

14. **Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417:**141–147.

15. **Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S.** 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. **21:**526–531. http://dx.doi.org/10.1038/nbt820.

16. **Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S.** 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J. Bacteriol. **190:**4050–4060. http://dx.doi.org/10.1128/JB.00204-08.