

1 **The Complete Genome Sequence of *Streptomyces* cf. *griseus* (XylebKG-1), an Ambrosia**
2 **Beetle-Associated Actinomycete.**

3 Kirk J. Grubbs^{1,2}, Peter H.W. Biedermann^{1,3}, Garret Suen^{1,4}, Sandra M. Adams^{1,4}, Joseph A. Moeller¹,
4 Jonathan L. Klassen¹, Lynne A. Goodwin^{5,6}, Tanja Woyke⁵, A. Christine Munk^{5,6}, David Bruce^{5,6}, Chris
5 Detter^{5,6}, Roxanne Tapia^{5,6}, Cliff S. Han^{5,6} and Cameron R. Currie^{1,4*}

6 Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin,¹ Cellular and Molecular
7 Pathology Graduate Program, University of Wisconsin-Madison, Madison, Wisconsin,² Institute of Ecology and
8 Evolution, Division Behavioral Ecology, University of Berne, Bern, Switzerland,³ DOE Great Lakes Bioenergy
9 Research Center, University of Wisconsin-Madison, Madison, Wisconsin,⁴ DOE Joint Genome Institute, Walnut
10 Creek, California,⁵ Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico⁶

11 **ABSTRACT**

12 *Streptomyces* cf. *griseus* (XylebKG-1) is an insect-associated strain of the well-studied
13 Actinobacterial species *S. griseus*. Here we present the genome of XylebKG-1 and discuss its
14 similarity to the genome of *S. griseus* subsp. *griseus* NBRC13350. XylebKG-1 was isolated from
15 the fungus cultivating *Xyleborinus saxesenii* system. Given its similarity to free-living *S. griseus*
16 subsp. *griseus* NBRC13350, comparative genomics will elucidate critical components of
17 bacterial interactions with insects.

* Corresponding author. Mailing address: DOE Great Lakes Bioenergy Research Center and Department of Bacteriology, 6155 MSB, 1550 Linden Drive, University of Wisconsin-Madison, Madison, WI 53706. Phone: (608)265-8034. Fax: (608)262-9865. E-mail: currie@bact.wisc.edu.

18 *Streptomyces griseus* is a soil bacterium known for its production of secondary
19 metabolites including streptomycin, the first effective antibiotic for tuberculosis (19). Here we
20 present the genome sequence of *Streptomyces cf. griseus* (XylebKG-1), which, to our
21 knowledge, is the first strain of *S. griseus* associated with an insect. XylebKG-1 was isolated
22 from the ambrosia beetle *Xyleborinus saxeseni*, which cultivates a fungus for food (2).
23 Actinobacteria-specific isolations from both beetles and their fungal galleries resulted in isolation
24 of XylebKG-1.

25 DNA from pure isolates was extracted using a bead-beating protocol (10) and the
26 genome was sequenced at the DOE Joint Genome Institute (JGI). A non-contiguous finished
27 genome of XylebKG-1 was generated using a shotgun approach employing a combination of
28 Illumina (3) and 454 sequencing technologies (15). An Illumina shotgun library (69,927,062
29 reads totaling ~5.3 Gbp) and two 454 GS (FLX Titanium) shotgun libraries (688,595 standard
30 reads and 172,566 20 Kbp paired-end reads totaling ~352 Mbp) were sequenced and
31 assembled. The 454 data was assembled using Newbler, version 2.3 (Roche) and Illumina
32 sequencing data was assembled with VELVET, version 0.7.63 (20). All assemblies were
33 integrated using parallel phrap, version SPS D 4.24 (High Performance Software, LLC). Illumina
34 data was used to correct potential base errors and increase consensus quality using the
35 software Polisher developed at the JGI (Alla Lapidus, unpublished). Possible misassembled
36 regions were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher (11), or by
37 sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed
38 by editing in Consed (6-8), by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks.
39 The total size of the genome is 8,727,768 bp and the final assembly is based on 352.4 Mbp of
40 454 sequence (38X coverage) and 5.3 Gbp of Illumina sequence (257x coverage).

41 The overall genome G+C content is 72.1%. Generation
42 (<http://compbio.ornl.gov/generation/>), Glimmer (5) and Critica (v1.05) (1) were used to predict a
43 total of 7,265 candidate protein-encoding gene models. RNAmmer (13) annotated six 16S

44 rRNAs and six 23S rRNAs. A tRNAscan-SE (14) search revealed 66 tRNAs corresponding to all
45 20 standard amino acids. Additional PRIAM (4), KEGG (12) and COG (18) analyses were
46 completed and can be accessed at <http://genome.ornl.gov/microbial/streACT1>.

47 Evidence for XylebKG-1 as a strain of *S. griseus* is based on similarity between the
48 genomes of XylebKG-1 and the type strain *S. griseus* subsp. *griseus* NBRC13350 (16). Both
49 have similar genome sizes (8.7 Mbp to 8.5 Mbp), G+C content (72.1% to 72.2%), six rRNA
50 operons and 66 tRNAs. An average nucleotide identity analysis conducted between both
51 genomes using Jspecies (v1.2.1) (17) revealed a 98.98 ANI_b value (91.68% genome alignment)
52 and a 98.95 ANI_m value (94.13% genome alignment), indicating a species level degree of
53 similarity (9). Given this similarity, XylebKG-1 represents a unique opportunity to study genetic
54 elements involved in Actinobacteria/Insect associations.

55 **Nucleotide sequence accession number:** The genome sequence of *Streptomyces cf. griseus*
56 XylebKG-1 is deposited in GenBank under accession ADFC00000000.

57

58 **Acknowledgements.** This work was funded by the DOE Great Lakes Bioenergy Research
59 Center (DOE BER Office of Science DE-FC02-07ER64494) supporting G.S., S.M.A., J.A.M.
60 and C.R.C. This work was also funded by the National Institute of Food and Agriculture, United
61 States Department of Agriculture, under ID number WISO1321 supporting K.J.G. P.H.W.B. was
62 partly supported by a DOC-fellowship of the Austrian Academy of Sciences at the Department
63 of Behavioral Ecology, University of Bern. The work conducted by the US Department of
64 Energy Joint Genome Institute is supported by the Office of Science of the US Department of
65 Energy under Contract No. DE-AC02-05CH11231. Additional support was provided by the
66 National Science Foundation (MCB-0702025). Support for J.L.K. comes from an NSERC
67 postdoctoral fellowship.

68

69

70

References

- 71 1. **Badger, J. H. and G. J. Olsen.** 1999. CRITICA: coding region identification tool
72 invoking comparative analysis. *Molecular Biology and Evolution* **16**:512-524.
- 73 2. **Batra, L. R.** 1967. *Ambrosia Fungi: A Taxonomic Revision, and Nutritional Studies of*
74 *Some Species.* *Mycologia* **59**:pp. 976-1017.
- 75 3. **Bennett, S.** 2004. *Solexa Ltd. Pharmacogenomics* **5**:433-438.
- 76 4. **Claudel-Renard, C., C. Chevalet, T. Faraut, and D. Kahn.** 2003. Enzyme-specific
77 profiles for genome annotation: PRIAM. *Nucleic Acids Research* **31**:6633-6639.
- 78 5. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying
79 bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673-679.
- 80 6. **Ewing, B. and P. Green.** 1998. Base-Calling of Automated Sequencer Traces
81 UsingPhred. II. Error Probabilities. *Genome Research* **8**:186-194.
- 82 7. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-Calling of Automated
83 Sequencer Traces UsingPhred. I. Accuracy Assessment. *Genome Research* **8**:175-
84 185.
- 85 8. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: A Graphical Tool for
86 Sequence Finishing. *Genome Research* **8**:195-202.
- 87 9. **Goris, J., K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme,**
88 **and J. M. Tiedje.** 2007. DNA-DNA hybridization values and their relationship to whole-
89 genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**:81-91.
- 90 10. **Graff, A. and R. Conrad.** 2005. Impact of flooding on soil bacterial communities
91 associated with poplar (*Populus* sp.) trees. *FEMS Microbiol. Ecol.* **53**:401-415.
- 92 11. **Han, C. and P. Chain.** 2006. Finishing Repetitive Regions Automatically with
93 Dupfinisher. In H. Valafar (ed.), *Proceedings of the 2006 International Conference on*
94 *Bioinformatics Computational Biology, BIOCOMP'06.* CSREA Press, Las Vegas,
95 Nevada, USA.
- 96 12. **Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hiraoka.** 2010. KEGG
97 for representation and analysis of molecular networks involving diseases and drugs.
98 *Nucleic Acids Research* **38**:D355-D360.

- 99 13. **Lagesen, K., P. Hallin, E. A. Rødland, H. Stærfeldt, T. Rognes, and D. W.**
100 **Ussery.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes.
101 *Nucleic Acids Research* **35**:3100-3108.
- 102 14. **Lowe, T. M. and S. R. Eddy.** 1997. tRNAscan-SE: A Program for Improved
103 Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*
104 **25**:0955-964.
- 105 15. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J.**
106 **Berka, M. S. Braverman, Y. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V.**
107 **Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L.**
108 **I. Alenquer, T. P. Jarvie, K. B. Jirage, J. Kim, J. R. Knight, J. R. Lanza, J. H.**
109 **Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E.**
110 **McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P.**
111 **Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M.**
112 **Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y.**
113 **Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome
114 sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
- 115 16. **Ohnishi, Y., J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A.**
116 **Yamashita, M. Hattori, and S. Horinouchi.** 2008. Genome Sequence of the
117 Streptomycin-Producing Microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.*
118 **190**:4050-4060.
- 119 17. **Richter, M. and R. Rosselló-Móra.** 2009. Shifting the genomic gold standard for
120 the prokaryotic species definition. *Proceedings of the National Academy of Sciences*
- 121 18. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG
122 database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic*
123 *Acids Research* **28**:33-36.
- 124 19. **Waksman, S. A.** 1953. Streptomycin: Background, Isolation, Properties, and
125 Utilization. *Science* **118**:pp. 259-266.
- 126 20. **Zerbino, D. R. and E. Birney.** 2008. Velvet: Algorithms for de novo short read
127 assembly using de Bruijn graphs. *Genome Research* **18**:821-829.