Complete Genome Sequence of *Streptomyces venezuelae* ATCC 15439, Producer of the Methymycin/Pikromycin Family of Macrolide Antibiotics, Using PacBio Technology

Jingxuan He,[a] Anitha Sundararajan,[b] Nicholas P. Devitt,[b] Faye D. Schilkey,[b] Thiruvarangan Ramaraj,[b,c] Charles E. Melançon III[a,c,d]

Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA[a]; National Center for Genome Resources (NCGR), Santa Fe, New Mexico, USA[b]; Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA[c]; Center for Biomedical Engineering, University of New Mexico, Albuquerque, New Mexico, USA[d]

Here, we report the complete genome sequence of *Streptomyces venezuelae* ATCC 15439, a producer of the methymycin/pikromycin family of macrolide antibiotics and a model host for natural product studies, obtained exclusively using PacBio sequencing technology. The 9.03-Mbp genome harbors 8,775 genes and 11 polyketide and nonribosomal peptide natural product gene clusters.

Address correspondence to Charles E. Melançon III, cemelanc@unm.edu.

Actinobacteria are well known for their ability to produce a variety of structurally complex, often bioactive natural products that are useful as drugs, drug leads, and chemical probes (1, 2). Among Actinobacteria, members of the genus *Streptomyces* are some of the most frequently observed species in nature, with >575 validly published species names (3) and >18,000 publicly deposited 16S rRNA gene sequences reported to date. Since 2001, when the first *Streptomyces* genomes, those of the model organisms *S. coelicolor* A3(2) (4) and *S. avermitilis* MA-4680 (5), were sequenced, the genomes of nearly 700 *Streptomyces* strains have been sequenced to at least the draft stage and made publicly available. However, among these, only 35% (245 genomes) are high-quality (<100 scaffold) assemblies, and only 8% (56 genomes) are complete genomes.

The full complement of genes required for the biosynthesis of bacterial natural products are almost invariably found at specific genomic loci called natural product biosynthetic gene clusters (BGCs) (6), which typically range in size from ~10 kb to >100 kb. A sufficiently high-quality (typically <100 scaffolds) genome assembly is an important prerequisite for obtaining the intact natural product BGC sequences needed for accurate bioinformatics-guided natural product discovery (7, 8) and synthetic biology-based natural product production (9) efforts.

There is an extensive collection of molecular genetic tools available for use in *Streptomyces*, and several model *Streptomyces* hosts, including *S. coelicolor* A3(2) (4, 10), *S. lividans* strains, *S. avermitilis* MA-4680 (5, 11), *S. albus* J1074 (12), *S. venezuelae* ATCC 10712 (13), and *S. venezuelae* ATCC 15439, have been developed. *S. venezuelae* ATCC 15439, a producer of the methymycin/pikromycin family of macrolide antibiotics (14), has been a model host for studying and manipulating deoxysugar and polyketide biosynthesis and macrolide glycosylation (15), for heterologous production of natural products

(16), and recently for unnatural amino acid incorporation (17). *S. venezuelae* ATCC 15439 is an advantageous model host because it is among the fastest growing *Streptomyces* strains (doubling time, ~60 min) (16), it grows in a dispersed manner in liquid culture, and it can be transformed efficiently. The complete nearly error-free genomes of model *Streptomyces* strains have been invaluable guides in the effort to understand and manipulate secondary metabolism.

To extend the capabilities afforded by a high-quality genome sequence to the model host *S. venezuelae* ATCC 15439, we sequenced its genome using PacBio next-generation technology. Genome sequencing was carried out using the Pacific Biosciences RSII (Menlo Park, CA) sequencing platform. PacBio long reads (two single-molecule real-time [SMRT] cells, ≈80× coverage) were assembled using the Hierarchical Genome Assembly Process 2 (HGAP2) protocol from SMRT Analysis version 2.0 package (18), resulting in the complete linear 9,034,396-bp *S. venezuelae* ATCC 15439 genome.

Gene prediction and annotation were carried out using RAST (19), incorporating the Glimmer (20) algorithm, and identified 8,682 putative protein-coding genes, 7 rRNA operons, and 72 tRNAs. Eleven polyketide and nonribosomal peptide natural product biosynthetic gene clusters, including the nearly error-free pikromycin cluster, were identified using *Dynamite* (8) and confirmed using antiSMASH (21). The *S. venezuelae* ATCC 15439 genome sequence will be a valuable resource for the continued development of the strain as a model host for natural product biosynthesis and synthetic biology studies.

Nucleotide sequence accession numbers. This genome sequence was deposited in EMBL/GenBank under accession no. LN881739. The version described in this paper is the first version, LN881739.1.

ACKNOWLEDGMENTS

FUNDING INFORMATION

REFERENCES

1. **Walsh CT, Fischbach MA**. 2010. Natural products version 2.0: connecting genes to molecules. J Am Chem Soc **132**:2469–2493. http://dx.doi.org/10.1021/ja909118a.
2. **Bologa CG, Ursu O, Oprea TI, Melançon CE, III, Tegos GP**. 2013. Emerging trends in the discovery of natural product antibacterials. Curr Opin Pharmacol **13**:678–687. http://dx.doi.org/10.1016/j.coph.2013.07.002.
3. **Labeda DP**. 2011. Multilocus sequence analysis of phytopathogenic species of the genus *Streptomyces*. Int J Syst Evol Microbiol **61**:2525–2531. http://dx.doi.org/10.1099/ijs.0.028514-0.
4. **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 C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajadream M-A, 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. http://dx.doi.org/10.1038/417141a.
5. **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.
6. **Fischbach M, Voigt CA**. 2010. Prokaryotic gene clusters: a rich toolbox for synthetic biology. Biotechnol J **5**:1277–1296. http://dx.doi.org/10.1002/biot.201000181.
7. **Fedorova ND, Moktali V, Medema MH**. 2012. Bioinformatics approaches and software for detection of secondary metabolic gene clusters. Methods Mol Biol **944**:23–45. http://dx.doi.org/10.1007/978-1-62703-122-6_2.
8. **Ogasawara Y, Yackley BJ, Greenberg JA, Rogelj S, Melançon CE, III**. 2015. Expanding our understanding of sequence-function relationships of type II polyketide biosynthetic gene clusters: bioinformatics-guided identification of frankiamicin A from *Frankia* sp. EAN1pec. PLoS One **10**:0121505. http://dx.doi.org/10.1371/journal.pone.0121505.
9. **Breitling R, Takano E**. 2015. Synthetic biology advances for pharmaceutical production. Curr Opin Biotechnol **35**:46–51. http://dx.doi.org/10.1016/j.copbio.2015.02.004.
10. **Gomez-Escribano JP, Bibb MJ**. 2014. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol **41**:425–431. http://dx.doi.org/10.1007/s10295-013-1348-5.
11. **Ikeda H, Kazuo SY, Omura S**. 2014. Genome mining of *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biotechnol **41**:233–250. http://dx.doi.org/10.1007/s10295-013-1327-x.
12. **Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A**. 2014. Insights into naturally minimised *Streptomyces albus* J1074 genome. BMC Genomics **15**:97. http://dx.doi.org/10.1186/1471-2164-15-97.
13. **Pullan ST, Chandra G, Bibb MJ, Merrick M**. 2011. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. BMC Genomics **12**:175. http://dx.doi.org/10.1186/1471-2164-12-175.
14. **Xue Y, Zhao L, Liu H-W, Sherman DH**. 1998. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. Proc Natl Acad Sci USA **95**:12111–12116. http://dx.doi.org/10.1073/pnas.95.21.12111.
15. **Thibodeaux CJ, Melançon CE, III, Liu H-W**. 2007. Natural product sugar biosynthesis and enzymatic glycodiversification. Ang Chem Int Ed Engl **47**:9814–9859.
16. **Kim EJ, Yang I, Yoon YJ**. 2015. Developing *Streptomyces venezuelae* as a cell factory for the production of small molecules used in drug discovery. Arch Pharm Res **38**:1606–1616. http://dx.doi.org/10.1007/s12272-015-0638-z.
17. **He J, Van Treeck B, Nguyen HB, Melançon CE, III**. 2016. Development of an unnatural amino acid incorporation system in the actinobacterial natural product producer *Streptomyces venezuelae* ATCC 15439. ACS Synth Biol **5**:125–132. http://dx.doi.org/10.1021/acssynbio.5b00209.
18. **Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10**:563–569. http://dx.doi.org/10.1038/nmeth.2474.
19. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.
20. **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.
21. **Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH**. 2015. antiSMASH 3.0-a comprehensive the genome mining of biosynthetic gene clusters. Nucleic Acids Res **43**:W237–W243. http://dx.doi.org/10.1093/nar/gkv437.