

High-Quality Draft Genome Sequence of Actinobacterium *Kibdelosporangium* sp. MJ126-NF4, Producer of Type II Polyketide Azicemicins, Using Illumina and PacBio Technologies

Yasushi Ogasawara,^{a*} Norah Torrez-Martinez,^b Anthony D. Aragon,^a Benjamin J. Yackley,^a Jessica A. Weber,^c Anitha Sundararajan,^d Thiruvarangan Ramaraj,^d Jeremy S. Edwards,^{a,b,e,f,g} Charles E. Melançon III^{a,c,g}

Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA^a; Department of Molecular Genetics and Microbiology, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA^b; Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA^c; National Center for Genome Resources (NCGR), Santa Fe, New Mexico, USA^d; Department of Chemical and Biological Engineering, University of New Mexico, Albuquerque, New Mexico, USA^e; Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA^f; Center for Biomedical Engineering, University of New Mexico, Albuquerque, New Mexico, USA^g

* Present address: Yasushi Ogasawara, Graduate School of Engineering, Hokkaido University, Hokkaido, Japan.

Here, we report the high-quality draft genome sequence of actinobacterium *Kibdelosporangium* sp. MJ126-NF4, producer of the type II polyketide azicemicins, obtained using Illumina and PacBio sequencing technologies. The 11.75-Mbp genome contains > 11,000 genes and 22 polyketide and nonribosomal peptide natural product gene clusters.

Received 29 January 2015 Accepted 4 February 2015 Published 2 April 2015

Citation Ogasawara Y, Torrez-Martinez N, Aragon AD, Yackley BJ, Weber JA, Sundararajan A, Ramaraj T, Edwards JS, Melançon CE, III. 2015. High-quality draft genome sequence of actinobacterium *Kibdelosporangium* sp. MJ126-NF4, producer of type II polyketide azicemicins, using Illumina and PacBio technologies. *Genome Announc* 3(2):e00114-15. doi:10.1128/genomeA.00114-15.

Copyright © 2015 Ogasawara et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

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

Actinobacteria are an important source of natural products, which are structurally complex small molecules that possess diverse and often medicinally useful bioactivities (1). Natural products are also increasingly recognized as playing important ecological roles (2). As improvements are made in bioinformatic methods for natural product gene cluster identification and functional prediction (3) and in techniques and hosts for sequence-based gene cluster capture (4) and heterologous expression (5), high-quality actinobacterial genome sequence data are becoming increasingly useful as the genetic template for the native or heterologous production of natural products and engineered analogues.

Although an increasing number of actinobacterial genomes have recently been sequenced, and recent phylogenetics-guided sequencing projects (6) provide a more complete genomic picture of phylum *Actinobacteria*, many actinobacterial genera do not yet have a representative genome sequence. The genomes of organisms from these genera represent an untapped source of natural-product biosynthetic diversity (7). Thus, we are interested in obtaining high-quality draft genome sequences of organisms from actinobacterial genera lacking genome-sequenced representatives, including those from the genus *Kibdelosporangium*.

The genus *Kibdelosporangium*, which is in the suborder *Pseudonocardineae*, is one of the rarest actinobacterial genera, with only 12 unique full-length 16S rRNA sequences from cultured isolates currently available in the NCBI databank. In spite of their rarity, members of the genus *Kibdelosporangium* produce a number of natural products with antibacterial, anticancer, and antiviral activities, including the structurally unusual aziridine-containing type II polyketide azicemicins produced by

Kibdelosporangium sp. MJ126-NF4 (8), whose genome sequence is reported here. This organism, which was initially misclassified as *Amycolatopsis* sp. MJ126-NF4, was isolated from a soil sample collected in Setagaya-ku, Tokyo Prefecture, Japan, in 1993 (9). In 2009, the azicemicin biosynthetic gene cluster was sequenced and annotated, and biochemical studies were carried out to investigate the biosynthetic origin of the aziridine moiety (10).

Genome sequencing was carried out using Illumina MiSeq (San Diego, CA) and Pacific Biosciences RS II (Menlo Park, CA) sequencing platforms. The MiSeq paired-end data (one lane, 216-bp average read length, $\approx 440\times$ coverage) were assembled using MIRA (11) and polished using a custom script, resulting in a 74-contig draft assembly. The polished synthetic reads from this assembly were combined with PacBio long reads (one single-molecule real-time [SMRT] cell, $\approx 38\times$ coverage) and merged using MIRA, resulting in an improved 41-contig assembly. This assembly was further scaffolded with the same PacBio long reads using the A Hybrid Assembler (AHA) protocol from the SMRT Analysis package (12). Finally, the GapCloser (version 1.12-r6) module from the SOAP suite (13) was used to attempt to resolve N-spacers introduced during the assembly scaffolding process, resulting in a high-quality 11.75-Mbp (21-scaffold; N_{50} , 1.6 Mbp) draft genome assembly.

Gene prediction and annotation were carried out using RAST (14), incorporating the Glimmer (15) algorithm, and identified 10,999 putative protein-coding genes, 7 rRNAs, and 63 tRNAs. Twenty-two polyketide and nonribosomal peptide natural-product biosynthetic gene clusters, all of which were intact, including the error-free azicemicin cluster, were identified using Dynamite (16) and confirmed using antiSMASH (17).

Nucleotide sequence accession numbers. This genome sequence has been deposited in EMBL/GenBank under the accession no. [CDME000000000](https://www.ncbi.nlm.nih.gov/nuclink/CDME000000000). The version described in this paper is the first version, CDME01000000.

ACKNOWLEDGMENTS

This work was supported by NIH NM-INBRE grant P20 GM103451 (to C.E.M.), by startup funds from the University of New Mexico (to C.E.M.), by NIH grant R01 HG006876 (to J.S.E.), and by NIH grant P30 CA118100 (to J.S.E.), which supports the University of New Mexico Cancer Center Bioinformatics Core.

We thank the Institute of Microbial Chemistry, Shinagawa-ku, Tokyo, Japan, for providing *Kibdelosporangium* sp. MJ126-NF4.

REFERENCES

- 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>.
- Romero D, Traxler MF, López D, Kolter R. 2011. Antibiotics as signal molecules. *Chem Rev* 111:5492–5505. <http://dx.doi.org/10.1021/cr2000509>.
- 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>.
- Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A*. 111:1957–1962. <http://dx.doi.org/10.1073/pnas.1319584111>.
- Gomez-Escribano JP, Bibb MJ. 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* 4:207–215. <http://dx.doi.org/10.1111/j.1751-7915.2010.00219.x>.
- Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng J-F, Lucas S, Kerfield C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk H-P, Eisen JA. 2009. A phylogeny-driven genomic encyclopaedia of bacteria and archaea. *Nature* 462:1056–1060. <http://dx.doi.org/10.1038/nature08656>.
- Tiwari K, Gupta RK. 2012. Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit Rev Biotechnol* 32:108–132. <http://dx.doi.org/10.3109/07388551.2011.562482>.
- Tsuchida T, Iinuma H, Kinoshita N, Ikeda T, Sawa R, Takahashi Y, Naganawa H, Sawa T, Hamada M, Takeuchi T. 1993. Azicemicin A, a new antimicrobial antibiotic from *Amycolatopsis*. *J Antibiot* 46:1772–1774.
- Tsuchida T, Iinuma H, Kinoshita N, Ikeda T, Sawa T, Hamada M, Takeuchi T. 1995. Azicemicins A and B, a new antimicrobial agent produced by *Amycolatopsis*. I. Taxonomy, fermentation, isolation, characterization and biological activities. *J Antibiot* 48:217–221.
- Ogasawara Y, Liu H-W. 2009. Biosynthetic studies of aziridine formation in azicemicins. *J Am Chem Soc* 131:18066–18068. <http://dx.doi.org/10.1021/ja907307h>.
- Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
- 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>.
- Luo R, Liu B, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Zhu X, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
- 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>.
- 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>.
- Ogasawara Y, Yackley BJ, Greenberg JA, Rogelj S, Melançon CE, III. 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*, in press.
- Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* 41:W204–W212. <http://dx.doi.org/10.1093/nar/gkt449>.