JOURNAL OF BACTERIOLOGY, Dec. 2010, p. 6317–6318 Vol. 192, No. 23
0021-9193/10/$12.00 doi:10.1128/JB.00859-10

Draft Genome Sequence of *Streptomyces clavuligerus* NRRL 3585, a Producer of Diverse Secondary Metabolites[▽]

Ju Yeon Song,[1,7] Haeyoung Jeong,[1] Dong Su Yu,[1] Michael A. Fischbach,[2] Hong-Seog Park,[1] Jae Jong Kim,[3] Jeong-Sun Seo,[4] Susan E. Jensen,[5] Tae Kwang Oh,[1,6] Kye Joon Lee,[7]* and Jihyun F. Kim[1]*

Korea Research Institute of Bioscience and Biotechnology (KRIBB), 111 Gwahangno, Yuseong-gu, Daejeon 305-806, Republic of Korea[1]; Department of Bioengineering and Therapeutic Sciences and California Institute of Quantitative Biosciences, University of California San Francisco, San Francisco, California 94158[2]; GenoTech Corporation, Research and Development Center, 59-5 Jang-Dong, Yuseong-gu, Daejeon 305-343, Republic of Korea[3]; Macrogen Inc., Seoul 153-023, Republic of Korea[4]; Department of Biological Sciences, University of Alberta, Edmonton T6G 2E9, Alberta, Canada[5]; 21C Frontier Microbial Genomics and Applications Center, Yuseong-gu, Daejeon 305-806, Republic of Korea[6]; and School of Biological Sciences, Seoul National University, Seoul 151-747, Republic of Korea[7]

Received 22 July 2010/Accepted 23 September 2010

Streptomyces clavuligerus **is an important industrial strain that produces a number of antibiotics, including clavulanic acid and cephamycin C. A high-quality draft genome sequence of the *S. clavuligerus* NRRL 3585 strain was produced by employing a hybrid approach that involved Sanger sequencing, Roche/454 pyrosequencing, optical mapping, and partial finishing. Its genome, comprising four linear replicons, one chromosome, and four plasmids, carries numerous sets of genes involved in the biosynthesis of secondary metabolites, including a variety of antibiotics.**

Streptomyces clavuligerus is a bacterium of industrial and clinical importance producing the β-lactamase inhibitor clavulanic acid (4), as well as cephamycin C (a β-lactam), clavams, tunicamycin, and holomycin (5, 9, 14, 17). *S. clavuligerus* is intriguing in regard to its genetic structure and mechanism of cephamycin C, clavulanic acid, and clavam biosynthesis (19, 20, 21). Here, we present a draft genome sequence of the type strain NRRL 3585 (ATCC 27064). The specific clone used, SC2, is a single-spore isolate from the stock of NRRL 3585 that showed good sporulation and metabolite production, a typical example of the wild-type strain.

The genome sequence was determined by Sanger paired-end sequencing (7) and Roche/454 pyrosequencing (12). Sanger reads at 4.9-fold coverage were produced from 4- or 40-kb genomic libraries, followed by 454 reads at 61.6-fold coverage. Sanger reads (14-fold coverage) provided by the Broad Institute were also utilized. The Sanger paired-end reads and the Newbler-assembled 454 contigs were assembled with a PCAP assembler (8). Optical mapping (Opgen, Inc.) was performed to confirm the assembly output and to assign contigs into each replicon. Gap closure was attempted using gap-spanning clones and PCR products. Coding sequences were predicted by the combined use of Glimmer (6), GeneMark (3), and CRITICA (2). Automatic functional annotation results obtained by AutoFACT (10) and the Rapid Annotation using Subsystem Technology (RAST) server (1) were compiled and validated with Artemis.

The genome consists of one linear chromosome (58 contigs in 4 scaffolds, 6,736,475 bp, 72.69% G+C) and four linear plasmids, pSCL1 (3 contigs in 2 scaffolds, 10,266 bp, 71.96% G+C), pSCL2 (2 contigs in 2 scaffolds, 149,326 bp, 70.07% G+C), pSCL3 (15 contigs in 2 scaffolds, 442,792 bp, 70.77% G+C), and pSCL4 (11 contigs in 3 scaffolds, 1,796,117 bp, 71.85% G+C). The 6.7-Mb chromosome is the smallest of the completely sequenced *Streptomyces* species. At least six rRNA operons and 66 tRNA genes as well as 7,898 protein-coding genes were annotated. Recently, a draft sequence of *S. clavuligerus* ATCC 27064 describing the 1.8-Mb megaplasmid was reported (13). The sequences of the chromosome and pSCL4 were nearly identical to our sequences. However, three other plasmids (15, 22) are present only in our data, suggesting that our clone has preserved the genome in its intact form.

A plethora of genes related to biosynthesis of secondary metabolites were discovered. The super-cluster for cephamycin C and clavulanic acid (21) is located on the chromosome, and the clavam cluster (19) is located approximately 1.4 Mb away. In contrast, a paralogue cluster (19) for clavulanic acid and clavam production lies on pSCL4. Dozens of gene sets for nonribosomal peptide synthetases, polyketide synthases, and the hybrids were detected from sequences of pSCL3, pSCL4, and the chromosome. Notably, two gene clusters that can synthesize enediyne-containing compounds, potent antitumor agents (11), were found. The genome also bears many gene clusters for secondary metabolites, such as staurosporine (18), moenomycin (16), terpenes, pentalenenes, phytoenes, siderophores, and lantibiotics.

This work could provide a platform to exploit bioactive com-

* Corresponding author. Mailing address for Jihyun F. Kim: Industrial Biotechnology and Bioenergy Research Center, Korea Research Institute of Bioscience and Biotechnology, 111 Gwahangno, Yuseong-gu, Daejeon 305-806, Republic of Korea. Phone: 82 42 860 4412. Fax: 82 42 879 8595. E-mail: jfk@kribb.re.kr. Mailing address for Kye Joon Lee: School of Biological Sciences, Seoul National University, Seoul 151-747, Republic of Korea. Phone: 82 2 780 6339. Fax: 82 2 785 6339. E-mail: lkj12345@snu.ac.kr.

▽ Published ahead of print on 1 October 2010.

pounds produced by *S. clavuligerus* NRRL 3585 and subsequently to develop strains better producing the molecules.

Nucleotide sequence accession numbers. The assembled genome sequence of *S. clavuligerus* NRRL 3585 was deposited in GenBank under accession number ADWJ00000000, and the one described in this paper is the first version, ADWJ01000000. The sequence and annotation are also available from the Genome Encyclopedia of Microbes (GEM; http://www.gem.re.kr).

We are grateful to members of GEM and the KRIBB sequencing team for technical assistance. We are also thankful to the Broad Institute of Harvard and MIT for providing the Sanger reads of *S. clavuligerus* ATCC 27064 (project identifier 28551, accession number ABJH00000000). Computer resources for this research were provided in part by the Korean BioInformation Center (KOBIC) Research Support Program.

This work was funded by the 21C Frontier Microbial Genomics and Applications Center Program (to J.F.K.) and the KICOS grant (K2072000001-07B0100-00110 to K.J.L.) of the Ministry of Education, Science and Technology.

REFERENCES

1. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75.
2. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
3. **Besemer, J., and M. Borodovsky.** 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. **33:**W451–W454.
4. **Brown, A. G., D. Butterworth, M. Cole, G. Hanscombe, J. D. Hood, C. Reading, and G. N. Rolinson.** 1976. Naturally occurring β-lactamase inhibitors with antibacterial activity. J. Antibiot. **29:**668–669.
5. **Brown, D., J. R. Evans, and R. A. Fletton.** 1979. Structures of three novel β-lactams isolated from *Streptomyces clavuligerus.* J. Chem. Soc. Chem. Commun. (Camb.) **1979:**282–283.
6. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
7. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. Fitzhugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
8. **Huang, X., J. Wang, S. Aluru, S. P. Yang, and L. Hillier.** 2003. PCAP: a whole-genome assembly program. Genome Res. **13:**2164–2170.
9. **Kenig, M., and C. Reading.** 1979. Holomycin and an antibiotic (MM19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. J. Antibiot. **32:**549–554.
10. **Koski, L. B., M. W. Gray, B. F. Lang, and G. Burger.** 2005. AutoFACT: an automatic functional annotation and classification tool. BMC Bioinformatics **6:**151.
11. **Liu, W., J. Ahlert, Q. Gao, E. Wendt-Pienkowski, B. Shen, and J. S. Thorson.** 2003. Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. Proc. Natl. Acad. Sci. U. S. A. **100:**11959–11963.
12. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
13. **Medema, M. H., A. Trefzer, A. Kovalchuk, M. van den Berg, U. Müller, W. Heijne, L. Wu, M. T. Alam, C. M. Ronning, W. C. Nierman, R. A. L. Bovenberg, R. Breitling, and E. Takano.** 2010. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol. Evol. **2:**212–224.
14. **Nagarajan, R., L. D. Boeck, M. Gorman, R. L. Hamill, C. E. Higgens, M. M. Hoehn, W. M. Stark, and J. G. Whitney.** 1971. β-Lactam antibiotics from *Streptomyces*. J. Am. Chem. Soc. **93:**2308–2310.
15. **Netolitzky, D. J., X. Wu, S. E. Jensen, and K. L. Roy.** 1995. Giant linear plasmids of β-lactam antibiotic producing *Streptomyces*. FEMS Microbiol. Lett. **131:**27–34.
16. **Ostash, B., E. H. Doud, C. Lin, I. Ostash, D. L. Perlstein, S. Fuse, M. Wolpert, D. Kahne, and S. Walker.** 2009. Complete characterization of the seventeen step moenomycin biosynthetic pathway. Biochemistry **48:**8830–8841.
17. **Pruess, D. L., and M. Kellett.** 1983. Ro-22-5417, a new clavam antibiotic from *Streptomyces clavuligerus*. I. Discovery and biological activity. J. Antibiot. **36:**208–212.
18. **Salas, J. A., and C. Méndez.** 2009. Indolocarbazole antitumour compounds by combinatorial biosynthesis. Curr. Opin. Chem. Biol. **13:**152–160.
19. **Tahlan, K., H. U. Park, and S. E. Jensen.** 2004. Three unlinked gene clusters are involved in clavam metabolite biosynthesis in *Streptomyces clavuligerus*. Can. J. Microbiol. **50:**803–810.
20. **Townsend, C. A.** 2002. New reactions in clavulanic acid biosynthesis. Curr. Opin. Chem. Biol. **6:**583–589.
21. **Ward, J. M., and J. E. Hodgson.** 1993. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. FEMS Microbiol. Lett. **110:**239–242.
22. **Wu, X., and K. L. Roy.** 1993. Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. J. Bacteriol. **175:**37–52.