

Draft Genome Sequence of *Streptomyces gancidicus* Strain BKS 13-15

Shailesh Kumar,^b Navjot Kaur,^a Nitin Kumar Singh,^a Gajendra Pal Singh Raghava,^b Shanmugam Mayilraj^a

Microbial Type Culture Collection and Gene Bank (MTCC), CSIR-Institute of Microbial Technology, Chandigarh, India^a; Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India^b

S.K. and N.K. contributed equally to this work.

This is IMTECH communication number 012/2013.

We report the 7.3-Mbp genome sequence of *Streptomyces gancidicus* strain BKS 13-15, isolated from mangrove sediment samples collected from the Bhitara Kanika Mangrove Reserve Forest, Odisha, India. The draft genome of strain *Streptomyces gancidicus* strain BKS 13-15 consists of 7,300,479 bp with 72.6% G+C content, 6,631 protein-coding genes, and 71 RNAs.

Received 1 March 2013 Accepted 13 March 2013 Published 18 April 2013

Citation Kumar S, Kaur N, Singh NK, Raghava GPS, Mayilraj S. 2013. Draft genome sequence of *Streptomyces gancidicus* strain BKS 13-15. *Genome Announc.* 1(2):e00150-13. doi:10.1128/genomeA.00150-13.

Copyright © 2013 Kumar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Shanmugam Mayilraj, mayil@imtech.res.in.

Waksman and Henrici proposed the genus *Streptomyces* in 1943 (1). *Streptomyces* species are well known by a linear chromosome, complex morphological differentiation, and an ability to produce many bioactive secondary metabolites containing important compounds for pharmaceutical and agrochemical uses (2, 3). Suzuki proposed *Streptomyces gancidicus* in 1957 (4). The organism in this study is *Streptomyces gancidicus* strain BKS 13-15, a Gram-positive bacterium isolated from a Bhitara Kanika soil sample from Odisha, India.

The genome of *Streptomyces gancidicus* strain BKS 13-15 was sequenced using the Illumina-HiSeq 1000 paired-end technology, which produced a total of 81,608,084 paired-end reads (insert size of 350 nucleotides [nt]) of 101 nt. We have used NGS QC toolkit v2.3 (5) to filter the data for high-quality (HQ) (cutoff read length for HQ, 70%; cutoff quality score, 20), vector/adaptor-free reads for genome assembly. A total of 75,080,804 high-quality, vector-filtered reads (~1,011× coverage) were used for assembly with SOAPdenovo v1.05 (at a hash length of 79) followed by GapCloser (at a hash length of 17) software (6). The final assembly contains 174 contigs of total size 7,303,225 bp, with an N₅₀ contig length of 79.0 kb; the largest contig assembled measures 351.9 kb. Three contaminated contigs of length 457 nt, 864 nt, and 1,425 nt were removed from assembly, and the genome draft comprising 7,300,479 bp with an N₅₀ contig length of 70 kbp and largest contig size of 351.9 kbp was submitted to GenBank (NCBI).

This draft genome (i.e., 171 contigs) comprising 7,300,479 bp was annotated with the help of the RAST (Rapid Annotation using Subsystem Technology) (7) server. A total of 6,631 protein coding regions, 3 rRNAs, and 68 tRNAs were predicted.

RAST annotation indicated that strain *Streptomyces coelicolor* strain A3 (score 502), *Streptomyces lividans* strain TK24 (score 459), *Streptomyces avermitilis* strain MA-4680 (score 386), and *Streptomyces viridochromogenes* strain DSM 40736 (score 378) are the closest neighbors of strain BKS 13-15. The annotation available from the RAST server also indicated that the strain BKS 13-15 has the genes for glycolysis and gluconeogenesis, the pentose

phosphate pathway, and the tricarboxylic acid (TCA) cycle. The strain BKS 13-25 also has the genes for catalase (EC 1.11.1.6), ferroxidase (EC 1.16.3.1), urease (alpha, beta, and gamma subunit) (EC 3.5.1.5), fructokinase (EC 2.7.1.4), nicotinamidase (EC 3.5.1.19), peroxidase (EC 1.11.1.7), and galactosidase (alpha and beta) (EC 3.2.1.23). Genes for glycine and serine utilization, i.e., 2-amino-3-ketobutyrate coenzyme A ligase (EC 2.3.1.29), aminomethyltransferase (glycine cleavage system T protein) (EC 2.1.2.10), cystathionine beta-synthase (EC 4.2.1.22), cystathionine gamma-lyase (EC 4.4.1.1), glycerate kinase (EC 2.7.1.31), and D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95), are also present in the annotation. Genes for putative polyketide synthase were also found in the annotation. We have mapped all predicted 6,631 protein coding sequences to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (8) with the help of the KEGG Automatic Annotation Server (KASS) server (9) and found that 7 proteins mapped on the streptomycin biosynthesis pathway, i.e., glucokinase (EC 2.7.1.2), phosphoglucosyltransferase (EC 5.4.2.2), myo-inositol-1-phosphate synthase (EC 5.5.1.4), myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25), myo-inositol 2-dehydrogenase (EC 1.1.1.18), dTDP-glucose 4,6-dehydratase (EC 4.2.1.46), and dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AOHP00000000](http://www.ncbi.nlm.nih.gov/nuccore/AOHP00000000). The version described in this paper is the first version, [AOHP01000000](http://www.ncbi.nlm.nih.gov/nuccore/AOHP01000000).

ACKNOWLEDGMENTS

This work was funded by IMTECH-CSIR. S.K. and N.K. are supported by a research fellowship from the Council of Scientific and Industrial Research (CSIR, India), and N.K.S. is supported by a fellowship from the University Grant Commission (UGC, India).

We thank the C-CAMP (<http://www.ccamp.res.in/>) next generation genomics facility for help in obtaining genome sequences. Genome assembly and annotation data can be downloaded from our genomics web portal at <http://crdd.osdd.net/raghava/genomesr/>.

REFERENCES

1. Waksman SA, Henrici AT. 1943. The nomenclature and classification of the actinomycetes. *J. Bacteriol.* 46:337–341.
2. Arakawa K, Sugino F, Kodama K, Ishii T, Kinashi H. 2005. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* 12:249–256.
3. El-Gendy MM, Shaaban M, EL-Bondkly AM, Shaaban KA. 2008. Bioactive benzopyrone derivatives from new recombinant fusant of marine *Streptomyces*. *Appl. Biochem. Biotechnol.* 150:85–96.
4. Suzuki M. 1957. Studies on an antitumor substance, gancidin. Mycological study on the strain AAK-84 and production, purification of active fractions. *J. Chiba Med. Soc.* 33:535–542. (In Japanese.)
5. Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. <http://dx.doi.org/10.1371/journal.pone.0030619>.
6. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
7. 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.
8. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40:D109–D114
9. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.