

Draft Genome Sequence of *Amycolatopsis decaplanina* Strain DSM 44594^T

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

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

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

This is IMTECH communication no. 011/2013.

We report the 8.5-Mb genome sequence of *Amycolatopsis decaplanina* strain DSM 44594^T, isolated from a soil sample from India. The draft genome of strain DSM 44594^T consists of 8,533,276 bp with a 68.6% G+C content, 7,899 protein-coding genes, and 57 RNAs.

Received 27 February 2013 Accepted 4 March 2013 Published 4 April 2013

Citation Kaur N, Kumar S, Bala M, Raghava GPS, Mayilraj S. 2013. Draft genome sequence of *Amycolatopsis decaplanina* strain DSM 44594^T. *Genome Announc.* 1(2):e00138-13. doi:10.1128/genomeA.00138-13.

Copyright © 2013 Kaur 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.

The genus *Amycolatopsis* was proposed by Lechevalier et al. in 1986 (1). *Amycolatopsis decaplanina* was proposed by Wink et al. in 2004 (2). *A. decaplanina* strain DSM 44594^T, which was obtained from the open collection of the German Collection of microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany, deposited by Joachim M. Wink, is a Gram-positive, aerobic, nonmotile, catalase-positive actinomycete that forms extensively branched substrate mycelia. The genome of *A. decaplanina* strain DSM 44594^T was sequenced using the Illumina-HiSeq 1000 paired-end technology and produced a total of 76,665,954 paired-end reads (insert size of 350 bp) of 101 bp. We used the Next-Generating Sequencing Quality Control (NGS QC) toolkit v2.3 (3) to filter the data for high-quality (HQ) (cutoff read length for HQ = 70%, cutoff quality score = 20) vector- and adaptor-free reads for genome assembly. A total of 70,826,953 high-quality vector-filtered reads (~794× coverage) were used for assembly with SOAPdenovo v1.05 (at a hash length of 73), followed by Gap-Closer (at a hash length of 27) software (4). The final assembly contains 85 contigs of total size 8,533,276 bp, with an N₅₀ contig length of 302.8 kb; the largest contigs assembled measure 791.1 kb.

The draft genome (85 contigs) comprising 8,533,276 bp was annotated with the help of the Rapid Annotations using Subsystems Technology (RAST) server (5). A total of 7,899 predicted coding regions (CDSs), 3 rRNAs, and 54 tRNAs were predicted. RAST annotation indicates that the strains *Streptomyces* sp. AA4 (score 530), *Actinosynnema mirum* DSM 43827 (score 448), *Saccharomonospora viridis* DSM 43017 (score 419), and *Saccharopolyspora erythraea* NRRL 2338 (score 415) are the closest neighbors of strain DSM 44594^T. Annotation available at the RAST server also indicates that the strain DSM 44594^T has the genes coding for glycolysis and gluconeogenesis, the Entner-Doudoroff pathway, and the tricarboxylic acid (TCA) cycle. The genes coding for alkaline phosphatase, galactosidase (alpha and beta subunits), mannosidase (alpha and beta subunits), urease (alpha, beta, and gamma), catalase (EC 1.11.1.6), and ferroxidase (EC 1.16.3.1) are also present in the genome annotation of strain DSM 44594^T. We

have mapped all predicted 7,899 CDSs to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (6) with the help of KEGG Automatic Annotation Server (KAAS) (7) and found the genes coding for glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24), dTDP-glucose 4,6-dehydratase (EC 4.2.1.46), and dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) mapped on the pathway of polyketide sugar unit biosynthesis. The genes coding for salicylate biosynthesis isochorismate synthase (EC 5.4.4.2), isochorismate pyruvate lyase (EC 4.2.99.21), bifunctional isochorismate lyase/aryl carrier protein (EC 3.3.2.1), 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28), 2,3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58), and pyochelin synthetase mapped on the pathway of the biosynthesis of siderophore group nonribosomal peptides.

Nucleotide sequence accession numbers. The draft genome sequence of *A. decaplanina* strain DSM 44594^T has been included in the GenBank Whole-Genome Shotgun (WGS) database under the accession no. [AOHO00000000](https://www.ncbi.nlm.nih.gov/nuccore/AOHO00000000). The version described in this paper is the first version, accession no. [AOHO01000000](https://www.ncbi.nlm.nih.gov/nuccore/AOHO01000000).

ACKNOWLEDGMENTS

This work was funded by CSIR-IMTECH. M.B. is supported by a University Grant Commission (UGC) fellowship. N.K. and S.K. are supported by a research fellowship from CSIR.

We thank the C-CAMP (<http://www.ccamp.res.in/>) next-generation genomics facility for help in obtaining the genome sequence.

Genome assembly and annotation data can be downloaded from our genomics web portal at <http://crdd.osdd.net/raghava/genomesrs/>.

REFERENCES

1. Lechevalier MP, Prauser H, Labeda DP, Ruan JS. 1986. Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *Int. J. Syst. Bacteriol.* 36:29–37.
2. Wink J, Gandhi J, Kroppenstedt RM, Seibert G, Sträubler B, Schumann P, Stackebrandt E. 2004. *Amycolatopsis decaplanina* sp. nov., a novel member of the genus with unusual morphology. *Int. J. Syst. Evol. Microbiol.* 54:235–239.
3. 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>.

- 4. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
- 5. 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.
- 6. 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.
- 7. 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.