Draft Genome Sequence of *Planomonospora sphaerica* JCM9374, a Rare Actinomycete

Hideo Dohra,[a,b] Tomohiro Suzuki,[c] Yuto Inoue,[d] Shinya Kodani[d]

Instrumental Research Support Office, Research Institute of Green Science and Technology, Shizuoka University, Suruga-ku, Shizuoka, Japan[a]; Department of Science, Graduate School of Integrated Science and Technology, Shizuoka University, Suruga-ku, Shizuoka, Japan[b]; Center for Bioscience Research and Education, Utsunomiya University, Utsunomiya, Tochigi, Japan[c]; Department of Agriculture, Graduate School of Integrated Science and Technology, Shizuoka University, Suruga-ku, Shizuoka, Japan[d]

Planomonospora sphaerica **is a rare actinomycete that is a potential antibiotic producer. Here, we report the draft genome sequence of *P. sphaerica* strain JCM9374. This is the first genome report of a bacterium belonging to the genus *Planomonospora*. The genome information of *P. sphaerica* will contribute to studies on the structure and function of antibiotics.**

Address correspondence to Shinya Kodani, kodani.shinya@shizuoka.ac.jp.

In a previous study, a rare actinomycete *Planomonospora sphaerica* was reported to show antibacterial activity against *Micrococcus luteus* (1). However, the antibacterial principle has not been identified yet. Microbial genome mining on the basis of next-generation sequencing technology is a powerful approach to discover the antibacterial compounds and biosynthetic pathways. The genome sequence of *P. sphaerica* was determined and the genes possibly involved in the antibacterial activity were searched for.

A genomic DNA of *P. sphaerica* strain JCM9374^T was extracted using the DNeasy blood and tissue kit, and then paired-end genome libraries (enzymatic fragmentation for 3 and 5 min) were constructed using the KAPA HyperPlus kit. Genome libraries of *P. sphaerica* were sequenced using the Illumina MiSeq platform at the Instrumental Research Support Office, Research Institute of Green Science and Technology, Shizuoka University. Paired-end raw reads (101-bp) were cleaned using Trimmomatic (2) by trimming adapter sequences and low-quality ends (quality score, <15) and khmer (3) by filtering the reads with a low k-mer coverage (<5) to remove sequence errors and contamination sequences. The resultant 1,305,056 high-quality reads totaling 611 Mb, which corresponds to an approximately 75-fold coverage of the genome, were assembled using SPAdes version 3.6.2 (4) with a k-mer size of 21, 33, 55, 77, 99, and 127 bp with the careful option, and contigs less than 200-bp were eliminated. The draft genome of *P. sphaerica* JCM9374 contains 84 contigs consisting of 8,132,638 bp with a G+C content of 72.73%.

The draft genome sequence was annotated using Prokka version 1.11 (5) with an in-house bacterial database containing 5,399 bacterial genomes (chromosomes and complete genomes only) in the NCBI RefSeq database as of 8 April 2016, and tRNA genes were predicted using tRNAscan-SE version 1.3.1 (6). The annotated genome contains 7,166 protein-coding sequences and 61 tRNA genes. The use of the in-house bacterial database reduced the number of proteins annotated as "hypothetical proteins" compared to the original database included in the Prokka (from 2,146 to 1,198), which provides more information for prediction of the protein function. In the genome of *P. sphaerica*, we found genes encoding proteins possibly involved in antibacterial activity, such as non-ribosomal peptide synthetases, bacteriocin, and lantibiotic biosynthesis proteins. The proteins encoded by the genome were additionally annotated with the COG (Clusters of Orthologous Groups) database (7) and Pfam database of protein families (8). Among 7,166 proteins, 4,843 (67.6%) and 5,570 (77.7%) were assigned to the COG functional categories and Pfam protein families, respectively, including siderophore synthetase component (COG4264), lanthionine synthetase C-like protein (PF05147), and lantibiotic dehydratase, C-terminus (PF04738). The genome information of *P. sphaerica* will contribute to studies on the structure and function of antibacterial peptides and the discovery of novel bioactive compounds and biosynthetic pathways.

Accession number(s). The draft genome sequence of *P. sphaerica* strain JCM9374 has been deposited in the DDBJ/EMBL/GenBank under the accession no. BDCX00000000.

REFERENCES

1. **Sherwood EJ, Hesketh AR, Bibb MJ.** 2013. Cloning and analysis of the planosporicin lantibiotic biosynthetic gene cluster of *Planomonospora alba*. J Bacteriol **195:**2309–2321. http://dx.doi.org/10.1128/JB.02291-12.
2. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.
3. **Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cart-**

wright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, Gonzalez I, Guermond S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Harpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Ranney C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, Stapleton JA, Steinman SR, Stein J, Taylor B, Trimble W, Wiencko HL, Wright M, Wyss B, Zhang Q, Zyme E, Brown CT. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Res **4**:. http://dx.doi.org/10.12688/f1000research.6924.1.

4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

5. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

6. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res **33**:W686–W689. http://dx.doi.org/10.1093/nar/gki366.

7. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res **43**:D261–D269. http://dx.doi.org/10.1093/nar/gku1223.

8. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res **42**:D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.