Draft Genome Sequence of *Micromonospora* sp. Strain HK10, Isolated from Kaziranga National Park, India

Madhumita Talukdar, Dhrubajyoti Das, Chiranjeeta Borah, Hari Prasanna Deka Boruah, Tarun Chandra Bora, Anil Kumar Singh

Department of Biotechnology, CSIR-North East Institute of Science and Technology, Jorhat, Assam, India

M.T. and D.D. contributed equally to this work.

We report the 6.92-Mbp genome sequence of *Micromonospora* sp. HK10, isolated from soil samples collected from Kaziranga National Park, Assam, India. The full genome of strain *Micromonospora* sp. strain HK10 consists of 6,911,179 bp with 73.39% GC content, 6,196 protein-coding genes, and 86 RNAs.

Address correspondence to Anil Kumar Singh, anil@rrljorhat.res.in.

Members of the *Micromonospora* genus are widely distributed in terrestrial, aquatic, and marine environments (1–4), but few of them have been isolated from nitrogen-fixing root nodules (5–7). *Micromonospora* species are well known for synthesizing antibiotics, especially aminoglycoside, enediyne, and oligosaccharide antibiotics. Despite the production of various useful antibiotics by *Micromonospora* species, the full genomes of only six strains have been sequenced to date. To enhance the knowledge base of *Micromonospora*, we contribute here a new genome sequence of *Micromonospora* sp. strain HK10, which was isolated from a soil sample of Kaziranga National Park, Assam, India.

The *Micromonospora* sp. HK10 genome was sequenced with the Illumina NextSeq 500 format using a paired-end 2×150-bp library. The raw data were trimmed using Trimmomatic version 0.30 (8) for high-quality read length (cutoff quality score, 20). A total of 8,896,194 high-quality vector-filtered reads were used for assembly with CLC Genomics Workbench version 6 (CLC bio, Denmark). The final assembly contains 294 contigs with a total size of 6,911,179 bp, a GC content of 73.39%, and an N_{50} contig length of 61,860 bp; the longest contig assembled measures 352,445 bp. The draft genome of *Micromonospora* sp. HK10 was annotated for protein-coding genes with the help of Prodigal version 2.60 (9), and a total of 6,282 CDSs were predicted.

All of the CDSs were further analyzed using BLASTx and the Kyoto Encyclopedia of Genes and Genomes (KEGG). BLASTx analysis was performed for predicted protein-coding genes using the nonredundant protein database (E value = 10^{-5}); as a result, 6,282 CDSs were observed and annotated, including 3 rRNAs and 58 tRNAs. The assignment of orthologous genes and the mapping of the CDSs to the biological pathways were performed using the KEGG server; a total of 1,052 genes were annotated. The gene ontology mapping was performed using Blast2GO version 2.0 (10), and the overall gene ontology distribution was 43.1% for molecular function, 43.4% for biological process, and 13.5% for cellular function. A maximum number of 2,086 out of 6,282 CDSs were matched with orthologs in *Micromonospora* sp. L5 (5), which is the closest member to HK10; the second closest is *Micromono-spora aurantiaca* ATCC 27029 (NC_014391) with 1,688 ortholog matches.

Our annotation results indicated that HK10 has genes for carbohydrate metabolism (tricarboxylic acid cycle, pentose phosphate pathway, glycolysis, and glyconeogesis), energy metabolism (oxidative phosphorylation), lipid metabolism (fatty acid biosynthesis and degradation), amino acid metabolism, glycan metabolism (peptidoglycan, lipopolysaccharide), and cofactor and vitamin metabolism. Various pigments (porphyrin, carotenoid), antibiotic coding clusters (ansamycins, tetracycline, vancomycin group of antibiotics, neomycin, novobiocin, penicillin, and cephalosporin), and polyketide synthase coding genes have been annotated. Several genes have also been annotated related to xenobiotic biodegradation and their metabolism (benzoate, chloroalkane, toluene, xylene, naphathalene, polycyclic aromatic hydrocarbon, etc).

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JTGL00000000. The version described in this paper is the first version.

ACKNOWLEDGMENTS

REFERENCES

1. **Kawamoto I**. 1989. Genus *Micromonospora* Ørskov 1923, 147[AL], p 2442–2450. *In* Williams ST, Sharpe ME, Holt JG (ed), Bergey's manual of systematic bacteriology, vol. 4. Williams & Wilkins, Baltimore, MD.
2. **Zhao H, Kassama Y, Young M, Kell DB, Goodacre R**. 2004. Differentiation of *Micromonospora* isolates from a coastal sediment in Wales on the basis of Fourier transform infrared spectroscopy, 16S rRNA sequence analysis, and the amplified fragment length polymorphism technique. Appl Environ Microbiol **70**:6619–6627. http://dx.doi.org/10.1128/AEM.70.11.6619-6627.2004.

3. **Maldonado LA, Stach JE, Pathom-aree W, Ward AC, Bull AT, Goodfellow M**. 2005. Diversity of cultivable actinobacteria in geographically widespread marine sediments. Antonie Van Leeuwenhoek **87:**11–18. http://dx.doi.org/10.1007/s10482-004-6525-0.

4. **Talukdar M, Duarah A, Talukdar S, Gohain MB, Debnath R, Yadav A, Jha DK, Bora TC**. 2012. Bioprospecting *Micromonospora* from Kaziranga National park of India and their anti-infective potential. World J Microbiol Biotechnol **28:**2703–2712. http://dx.doi.org/10.1007/s11274-012-1080-8.

5. **Hirsch AM, Alvarado J, Bruce D, Chertkov O, De Hoff PL, Detter JC, Fujishige NA, Goodwin LA, Han J, Han S, Ivanova N, Land ML, Lum MR, Milani-Nejad N, Nolan M, Pati A, Pitluck S, Tran SS, Woyke T, Valdés M**. 2013. Complete genome sequence of *Micromonospora* strain L5, a potential plant-growth-regulating actinomycete, originally isolated from *Casuarina equisetifolia* root nodules. Genome Announc **1**(5): e00759-13. http://dx.doi.org/10.1128/genomeA.00759-13.

6. **Valdés M, Pérez N-O, Estrada-de los Santos P, Caballero-Mellado J, Peña-Cabriales JJ, Normand P, Hirsch AM**. 2005. Non-*Frankia* actinomycetes isolated from surface-sterilized roots of *Casuarina equisetifolia* fix

nitrogen. Appl Environ Microbiol **71:**460–466. http://dx.doi.org/10.1128/AEM.71.1.460-466.2005.

7. **Trujillo ME, Kroppenstedt RM, Schumann P, Carro L, Martínez-Molina E**. 2006. *Micromonospora coriariae* sp. nov., isolated from root nodules of *Coriaria myrtifolia*. Int J Syst Evol Microbiol **56:**2381–2385. http://dx.doi.org/10.1099/ijs.0.64449-0.

8. **Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B**. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res **40:**W622–W627. http://dx.doi.org/10.1093/nar/gks540.

9. **Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.

10. **Conesa A, Götz S, García-Gómez J, Terol J, Talon M, Robles M** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **15:**3674–3676. http://dx.doi.org/10.1093/bioinformatics/bti610.