AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Complete Draft Genome Sequence of the Actinobacterium *Nocardiopsis sinuspersici* UTMC102 (DSM 45277^T), Which Produces Serine Protease

Bogdan Tokovenko,[a] Christian Rückert,[b] Jörn Kalinowski,[b] Fatemeh Mohammadipanah,[c] Joachim Wink,[d] Birgit Rosenkränzer,[e] Maksym Myronovskyi,[e] Andriy Luzhetskyy[a,e]

Helmholtz Institute for Pharmaceutical Research Saarland, Saarbrücken, Germany[a]; Technology Platform Genomics, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany[b]; Department of Microbial Biotechnology, School of Biology and Center of Excellence in Phylogeny of Living Organisms, College of Science, University of Tehran, Tehran, Iran[c]; Microbial Strain Collection, Helmholtz Centre for Infection Research, Braunschweig, Germany[d]; University of Saarland, Department of Pharmaceutical Biotechnology, Saarbrücken, Germany[e]

ABSTRACT The genome sequence of alkalohalophilic actinobacterium *Nocardiopsis sinuspersici* UTMC102 is provided. *N. sinuspersici* UTMC102 produces a highly active serine alkaline protease, and contains at least 11 gene clusters encoding the biosynthesis of secondary metabolites. The *N. sinuspersici* UTMC102 genome was assembled into a single chromosomal scaffold.

Received 24 March 2017 **Accepted** 27 March 2017 **Published** 18 May 2017

Citation Tokovenko B, Rückert C, Kalinowski J, Mohammadipanah F, Wink J, Rosenkränzer B, Myronovskyi M, Luzhetskyy A. 2017. Complete draft genome sequence of the actinobacterium *Nocardiopsis sinuspersici* UTMC102 (DSM 45277^T), which produces serine protease. Genome Announc 5:e00362-17. https://doi.org/10.1128/genomeA.00362-17.

Address correspondence to Fatemeh Mohammadipanah, fmohammadipanah@ut.ac.ir, or Andriy Luzhetskyy, Andriy.Luzhetskyy@helmholtz-hzi.de.

The genus *Nocardiopsis* harbors the most abundant halophilic and halotolerant species compared to other genera in class *Actinobacteria* (1). Members of the genus *Nocardiopsis* are present in frozen soils, desert sand, compost, saline, or hypersaline habitats (marine systems, salterns, soils), and alkaline places (slag dumps, lake soils, sediments) (2). *Nocardiopsis* species produce enzymes that are cold-adapted (α-amylases), thermotolerant (α-amylases and xylanases), thermoalkalotolerant (cellulases, β-1,3-glucanases), alkalitolerant thermostable (inulinases), acid-stable (keratinase), and alkalophilic (serine proteases). Enzymes derived from *Nocardiopsis* species act on insoluble polymers such as glucans (pachyman and curdlan), keratin (feathers and prion proteins), and polyhydroxyalkanoates (2).

N. sinuspersici UTMC102 was discovered in sandy soil from the banks of the Arvand River, Khoramshahr, Iran (3). *N. sinuspersici* UTMC102 has the ability to produce a highly active serine alkaline protease which effectively hydrolyzes milk protein. The strain also has the genomic potential to produce a spectrum of secondary metabolites.

For genome sequencing, two libraries were constructed: a 7 to 9 kb mate-pair library, and a 450 to 550 bp paired-end library. Both libraries were sequenced using a MiSeq system. Overall, 2 × 2,867,592 MP reads (300 bp long) and 2 × 778,640 PE reads (also 300 bp long) were obtained. Nextera MP reads were processed with NxTrim (4) to separate them into proper MP/PE/single-end reads. All reads were then subjected to Trimmomatic (5) trimming. Processed reads (total coverage 106×) were assembled using SPAdes v3.8.1, discarding fragments either shorter than 1 kbp or with coverage under 50% of scaffold NOSIN_1 coverage. Scaffold NOSIN_1 is the chromosome (6,071,583 bp, 71.7% G+C content). The other 3 short contigs represent unplaced/repetitive fragments possibly belonging to the chromosome. Contig NOSIN_2 (39,728 bp, 63.1% G+C) contains 2 DEAD/DEAH helicase genes, a single *traC* conjugal transfer protein gene, 4 biosynthetic genes (including *queC*, *queE*, *queD*), and 2 transposases. Contig NOSIN_3 (4,414 bp, 67.4% G+C) contains 2 pseudogenes; database search identifies

NOSIN_3 as a fragment of an NRPS gene. Contig NOSIN_4 (3,771 bp, 73.6% G+C) contains a single pseudo-gene, which yields multiple BLAST hits to NRPS genes of *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43111. NOSIN_4 may fill the largest (4,328 bp) of the 5 gaps in the NOSIN_1 chromosome assembly, as that gap lies within an NRPS gene. The other 4 gaps in the chromosome are inside rRNA gene clusters, between the rRNA genes.

The genome sequence was submitted to NCBI Prokaryotic Gene Annotation Pipeline. A total of 5,213 genes were identified, of them 15 rRNA genes (in 5 rRNA operons), 58 tRNA genes, and 5,056 protein coding genes, encompassing 5,052,038 nucleotides of coding regions (83% of scaffold NOSIN_1); 3,502 genes had function assignment after the annotation.

Using antiSMASH 3 (6), 63 gene clusters encoding the biosynthesis of secondary metabolites were identified, 52 of them with the ClusterFinder algorithm. In comparison, the genomes of *S. coelicolor* (7), *S. fulvissimus* (8), and *K. albida* (9) contain 20, 30, and 46 secondary metabolism gene clusters, respectively.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. MCOK00000000. The version described in this paper is version MCOK01000000.

ACKNOWLEDGMENTS

F.M. would like to thank INSF for the grant allocated for analysis of the protease enzyme produced by the UTMC102 strain.

This work was supported by the ERC starting grant EXPLOGEN No. 281623 to A.L.

REFERENCES

1. Hamedi J, Mohammadipanah F, Ventosa A. 2013. Systematic and biotechnological aspects of halophilic and halotolerant actinomycetes. Extremophiles 17:1–13. https://doi.org/10.1007/s00792-012-0493-5.
2. Bennur T, Kumar AR, Zinjarde S, Javdekar V. 2014. *Nocardiopsis* species as potential sources of diverse and novel extracellular enzymes. Appl Microbiol Biotechnol 98:9173–9185. https://doi.org/10.1007/s00253-014-6111-y.
3. Hamedi J, Mohammadipanah F, von Jan M, Pötter G, Schumann P, Spröer C, Klenk HP, Kroppenstedt RM. 2010. *Nocardiopsis sinuspersici* sp. nov., isolated from sandy rhizospheric soil. Int J Syst Evol Microbiol 60:2346–2352. https://doi.org/10.1099/ijs.0.018366-0.
4. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. 2015. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics 31:2035–2037. https://doi.org/10.1093/bioinformatics/btv057.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
7. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a.
8. Myronovskyi M, Tokovenko B, Manderscheid N, Petzke L, Luzhetskyy A. 2013. Complete genome sequence of *Streptomyces fulvissimus*. J Biotechnol 168:117–118. https://doi.org/10.1016/j.jbiotec.2013.08.013.
9. Rebets Y, Tokovenko B, Lushchyk I, Rückert C, Zaburannyi N, Bechthold A, Kalinowski J, Luzhetskyy A. 2014. Complete genome sequence of producer of the glycopeptide antibiotic aculeximycin *Kutzneria albida* DSM 43870T, a representative of minor genus of *Pseudonocardiaceae*. BMC Genomics 15:885. https://doi.org/10.1186/1471-2164-15-885.