Draft genome sequence of *Gordonia neofelisaeic* NRRL B-59395, a cholesterol-degrading actinomycete

Fanglan Ge, Wei Li*, Guiying Chen, Yuchang Liu, Guangxiang Zhang, Bin Yong, Qiong Wang, Nan Wang, Zhumei Huang, Weitian Li, Jing Wang, Cheng Wu, Qian Xie, Gang Liu

College of life Sciences, Sichuan Normal University, Chengdu, 610061, P.R.China

Corresponding author, Wei Li, weecle201@yahoo.com.cn

We report a draft sequence of the genome of *Gordonia neofelisaeic* NRRL B-59395, a cholesterol-degrading actinomycete isolated from fresh faeces of a clouded leopard (*Neofelis nebulosa*). As predicted, the reported genome contains several gene clusters for cholesterol degradation. This is the second available genome sequence of the family *Gordoniaceae*.

The genus *Gordonia* was originally proposed by Tsukamura (in 1971) and belongs to the mycolic acid group of the *actinomycetes* (15). In recent years, it has attracted much interest for its ability to degrade xenobiotics, environmental pollutants (1), and recalcitrant natural polymers, such as rubber (9), dibenzothiophene (6), explosive hexogen (14), butyl benzyl phthalates, and cholesterol (4, 10). And quite a number of *Gordonia* species are opportunistic pathogens (5). However, in the genome database, the only available genome sequence is *Gordonia bronchialis* DSM 43247 at this time (7). Recently, our laboratory isolated and characterized *Gordonia neofelisaeic* (NRRL B-59395) (11), a novel species of *Gordonia* which shows high ability to
degrade steroidal compounds, and a draft genome sequence of this species is presented here.

The genome sequence was determined using Illumina Genome Analyzer II at the Beijing Genomics Institute (BGI; Shenzhen, China). Draft assemblies were based on 500-Mb reads. All reads provided about 117-fold coverage of the genome. The GAII paired-end reads were assembled into 168 contigs in 47 scaffolds with the SOAPdenovo program. Putative open reading frames with more than 30 amino acid residues were predicted using Glimmer 3.0 (3). rRNA gene were identified by RNAmmer (8) and the tRNAscan-SE server (12). The scaffolds were searched against the KEGG and COG (Clusters of Orthologous Groups) databases to annotate the gene descriptions.

The draft genome of *G. neofelisae* NRRL B 59395 consists of 4,257,286 bases with a G+C content of 68.62%, and there are 4,032 putative coding sequences with average length of 951 bp. The coding percentage is 90.3%. There are 46 genes for tRNAs and 5 rRNA loci.

A plethora of genes related to catabolism of cholesterol were discovered. There are two genes encoding putative 3-ketosteroid-delta-1-dehydrogenase (KSTD), with 60-76% identity to those of orthologs in *R. erythropolis* SQ1, they catalyses transhydrogenation of 3-keto-4-ene-steroid to 3-keto-1,4-diene-steroid. Three open read frames encoding 3-ketosteroid 9-alpha-hydroxylases (KshA and KshB) were found. A gene cluster SCNU_09391 to SCNU_09186, including about 30 genes, was discovered, most of these gene products share greatest identity to ro04482-ro04705
cluster in *Rhodococcus jostii* RHA1 and rv3492c - rv3574 in *Mycobacterium tuberculosis* H37Rv (16), which encode all of the enzymes necessary to perform one full cycle of β-oxidation and involve in the cholesterol side-chain shorting. Similar to the *igr* locus Rv3545c to Rv3540c in *M. tuberculosis* (2, 16), the second cluster SCNU_16768 to SCNU_16798 consists of a single operon containing genes for three acyl-CoA dehydrogenas, a lipid-transfer protein and a putative TetR-family transcriptional regulatory protein. A third cluster SCNU_09181 to SCNU_09146 includes eight genes that predicted to encode a multicomponent cholesterol uptake system: *mce4ABCDEF* and *supAB* (called *mce* cluster) (16). The genome sequence of *G. neofelisaecis* and its annotation will provide further insight into the physiology and metabolic potential of *Gordonia*.

**Nucleotide sequence accession numbers.** This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AEUD00000000. The version described in this paper is the first version, AEUD01000000.

**Acknowledgements**

This work was supported by the Sichuan provincial Science & Technology Department (2008JY0103-1, 2009JY0067). We thank Dr. Dongxia Li for her careful reading of the manuscript.
REFERENCES


16. Van der Geize R., et al. 2007. A gene cluster encoding cholesterol catabolism in...