Dear Author,

Please, note that changes made to the HTML content will be added to the article before publication, but are not reflected in this PDF.

Note also that this file should not be used for submitting corrections.

Journal of Biotechnology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Genome announcement

Complete genome sequence of '*Mycobacterium neoaurum*' NRRL B-3805, an androstenedione (AD) producer for industrial biotransformation of sterols

Antonio Rodríguez-García[a,b], Estela Fernández-Alegre[a], Alejandro Morales[a], Alberto Sola-Landa[a], Jess Lorraine[c], Sandy Macdonald[c], Dmitry Dovbnya[d], Margaret C.M. Smith[c], Marina Donova[d], Carlos Barreiro[a,*]

[a] *INBIOTEC (Instituto de Biotecnología de León), Parque Científico de León, Avda. Real 1, 24006 León, Spain*
[b] *Área de Microbiología, Departamento de Biología Molecular, Universidad de León, Campus de Vegazana, 24071 León, Spain*
[c] *Department of Biology, University of York, Wentworth Way, YO10 5DD, United Kingdom*
[d] *G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Prospekt Nauki 5, 142290 Pushchino, Moscow Region, Russia*

ARTICLE INFO

Article history:
Received 23 February 2016
Accepted 10 March 2016
Available online xxx

Keywords:
Mycobacterium
Steroids
4-androstene-3
17-dione
Actinobacteria
Genome

ABSTRACT

Microbial bioconversion of sterols into high value steroid precursors, such as 4-androstene-3,17-dione (AD), is an industrial challenge. Genes and enzymes involved in sterol degradation have been proposed, although the complete pathway is not yet known. The genome sequencing of the AD producer strain '*Mycobacterium neoaurum*' NRRL B-3805 (formerly *Mycobacterium* sp. NRRL B-3805) will serve to elucidate the critical steps for industrial processes and will provide the basis for further genetic engineering. The genome comprises a circular chromosome (5 421 338 bp), is devoid of plasmids and contains 4844 protein-coding genes.

The microbial biotransformation (or bioconversion) of sterols into steroid precursors provides an economical and sustainable pipeline for the production of steroid drugs. The sterol carbon skeleton is found in cholesterol or the so-called phytosterols (e.g., sitosterol, stigmasterol, campesterol, and brassicasterol). Phytosterols are a low cost and abundant source for the synthesis of steroid drugs by microbial bioconversions that degrade their side chain. As a result valuable steroid precursors, such as 4-androstene-3,17-dione (AD) that can be chemically converted to almost all kinds of pharmaceutically steroids, are obtained. There are 300 approved steroids representing, after antibiotics, the second largest class of drugs marketed for medical applications (Donova and Egorova, 2012; García et al., 2012; Shtratnikova et al., 2014).

Sterol bioconversion is efficiently carried out by *Mycobacterium* species at industrial level. *Mycobacterium* sp. NRRL B-3805 (CECT 3018; DSM 2967; NCIMB 11678) (hereafter B-3805) was obtained from a screening and mutagenesis program for efficient bioconversion strains (Marsheck et al., 1972; patent no. US3759791). The parental strain was a soil isolate characterized as an androsta-1,4-diene-3,17-dione (ADD) producer. In contrast, the B-3805 strain, selected after two rounds of ultraviolet irradiation, partly lost the ability to 1-dehydrogenate steroids; consequently, it transforms sterols mainly into AD. This strain is also a better AD producer and generates less of the by-product ADD than its closely-related strain *Mycobacterium neoaurum* VKM Ac-1815D (Shtratnikova et al., 2014)(hereafter Ac-1815D).

The complete genome sequence of *Mycobacterium* sp. NRRL B-3805 was determined with Ion Torrent and Sanger sequencing technologies. Next-generation sequencing of B-3805 total DNA yielded 6 165 546 reads. Sequences were trimmed or discarded using the sickle program (Joshi and Fass, 2011) The remaining sequences (93.8%; average length of 250 bp) were mapped with segemehl (Hoffmann et al., 2009) at a 269-fold coverage to the Ac-1815D reference genome sequence (Shtratnikova et al., 2014). Analysis of the alignment of reads revealed 13 regions with low (<4) or null coverage. PCR products were obtained and used as templates for Sanger sequencing reactions on both DNA strands to

* Corresponding author.
E-mail addresses: c.barreiro@unileon.es, carlos.barreiro.mendez@gmail.com (C. Barreiro).

http://dx.doi.org/10.1016/j.jbiotec.2016.03.021
0168-1656/

Please cite this article in press as: Rodríguez-García, A., et al., Complete genome sequence of '*Mycobacterium neoaurum*' NRRL B-3805, an androstenedione (AD) producer for industrial biotransformation of sterols. J. Biotechnol. (2016), http://dx.doi.org/10.1016/j.jbiotec.2016.03.021

Table 1
General features of the '*Mycobacterium neoaurum*' NRRL B-3805 complete genome.

Genome size (bp)	5 421 338
G + C content (%)	66.9
Coding region (%)	88.6
Total genes	5 065
Protein coding genes	4 844
Average CDS length (bp)	992
Pseudogenes	168
rRNA (operons)	6 (2)
tRNA	46
ncRNA	1
GenBank accession no.	CP011022

close these gaps. A *de novo* assembly of the Ion Torrent and Sanger reads using SPAdes, which allows hybrid assemblies (Bankevich et al., 2012), confirmed the collinearity between the reference (Ac-1815D) and the assembled sequences. The chromosome sequence of the B-3805 strain was nearly identical to that of Ac-1815D strain. The differences comprise 10 SNPs, a point deletion and five 36-nt tandem repeats in B-3805 compared with three equivalent repeats in Ac-1815D. How these sequence differences relate to the variations in steroid precursor (AD, ADD) production described above is not clear.

The genomic sequence of *Mycobacterium* sp. B-3805 comprises a circular chromosome of 5 421 338 bp and no plasmids. Gene prediction and annotation by the NCBI prokaryotic genome annotation pipeline found 5 065 genes, including 4 844 protein-coding genes and 6 rRNA genes arranged in two operons (Table 1).

These genome data permit the species-level identification of the strain using three marker sequences. In addition to the well-known 16S rRNA gene, internal regions of both *rpoB* and *hsp65* genes were used as *Mycobacterium* inter-species markers as proposed by previous researchers (Kim et al., 1999, 2005). The 16S rRNA gene sequence of B-3805, following a search in the EzTaxon database (Kim et al., 2012), showed 100% identity to that of the type strain *M. neoaurum* ATCC 25795^T. A BLAST search against the GenBank's non-redundant nucleotide database (http://www.ncbi.nlm.nih.gov/) showed the highest similarity of B-3805 *hsp65* gene (MyAD_05080) with the Ac-1815D strain (100% identity; accession no. CP006936) and subsequently with *M. neoaurum* ATCC 25795^T (98.28% identity; accession no. AY299156). Finally, *M. neoaurum* ATCC 23072 was the best match against a 638-nt fragment of the B-3805 *rpoB* gene (99.2% identity; accession FJ172338). These results indicate that *Mycobacterium* sp. NRRL B-3805 belongs to the *M. neoaurum* group (Simmon et al., 2009).

In conclusion, the genome sequence of the '*M. neoaurum*' NRRL B-3805 strain will serve to improve the current understanding of the sterol degradation pathway in *Mycobacterium* (Donova and Egorova, 2012) and will help in the future rational development of industrially important bioconversion strains.

Nucleotide sequence accession number

The complete genome sequence of '*M. neoaurum*' NRRL B-3805 has been deposited in the GeneBank under the accession number CP011022.

Acknowledgments

This work was fully supported by a grant of the European Union program ERA-IB [MysterI (EIB.12.010)] through the APCIN call of the Spanish Ministry of Economy and Competitiveness (MINECO, Spain) (PCIN-2013-024-C02-01). Dmitry Dovbnya and Marina Donova acknowledge Russian Science Foundation (Grant No. 14-24-00169) for the supporting of their work. The authors want to thanks the European Union program ERA-IB; the Spanish Ministry of Economy and Competitiveness (MINECO, Spain); the Biotechnology and Biological Sciences Research Council (BBSRC, United Kingdom); the Foundation for Assistance to Small Innovative Enterprises (FASIE, Russia) and the MysterI Consortium (INBIOTEC, Pharmins Ltd., University of York, SINTEF, Technische Universität Dortmund, Gadea Biopharma S.L.). We thank J. Merino, B. Martín and A. Casenave for their excellent technical assistance.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 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.

Donova, M.V., Egorova, O.V., 2012. Microbial steroid transformations: current state and prospects. Appl. Microbiol. Biotechnol. 94, 1423–1447, http://dx.doi.org/10.1007/s00253-012-4078-0.

García, J.L., Uhía, I., Galán, B., 2012. Catabolism and biotechnological applications of cholesterol degrading bacteria. Microb. Biotechnol. 5, 679–699, http://dx.doi.org/10.1111/j.1751-7915.2012.00331.x.

Hoffmann, S., Otto, C., Kurtz, S., Sharma, C.M., Khaitovich, P., Vogel, J., Stadler, P.F., Hackermüller, J., 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput. Biol. 5, e1000502, http://dx.doi.org/10.1371/journal.pcbi.1000502.

N. Joshi J. Fass, 2011 Sickle: A sliding-window, adaptive quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Kim, B.J., Lee, S.H., Lyu, M.A., Kim, S.J., Bai, G.H., Chae, G.T., Kim, E.C., Cha, C.Y., Kook, Y.H., 1999. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). J. Clin. Microbiol. 37, 1714–1720.

Kim, H., Kim, S.-H., Shim, T.-S., Kim, M., Bai, G.-H., Park, Y.-G., Lee, S.-H., Chae, G.-T., Cha, C.-Y., Kook, Y.-H., Kim, B.-J., 2005. Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (hsp65). Int. J. Syst. Evol. Microbiol. 55, 1649–1656, http://dx.doi.org/10.1099/ijs.0.63553-0.

Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon, Y.S., Lee, J.-H., Yi, H., Won, S., Chun, J., 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int. J. Syst. Evol. Microbiol. 62, 716–721, http://dx.doi.org/10.1099/ijs.0.038075-0.

Marsheck, W.J., Kraychy, S., Muir, R.D., 1972. Microbial degradation of sterols. Appl. Microbiol. 23, 72–77.

Shtratnikova, V.Y., Bragin, E.Y., Dovbnya, D.V., Pekov, Y.A., Schelkunov, M.I., Strizhov, N., Ivashina, T.V., Ashapkin, V.V., Donova, M.V., 2014. Complete genome sequence of sterol-Transforming *Mycobacterium neoaurum* strain VKM Ac-1815D. Genome Announc. 2, 12–13, http://dx.doi.org/10.1128/ genomeA.01177-13.

Simmon, K.E., Low, Y.Y., Brown-Elliott, B.A., Wallace, R.J., Petti, C.A., 2009. Phylogenetic analysis of *Mycobacterium aurum* and *Mycobacterium neoaurum* with redescription of *M. aurum* culture collection strains. Int. J. Syst. Evol. Microbiol. 59, 1371–1375, http://dx.doi.org/10.1099/ijs.0.007799-0.

Please cite this article in press as: Rodríguez-García, A., et al., Complete genome sequence of '*Mycobacterium neoaurum*' NRRL B-3805, an androstenedione (AD) producer for industrial biotransformation of sterols. J. Biotechnol. (2016), http://dx.doi.org/10.1016/j.jbiotec.2016.03.021