High-Quality Whole-Genome Sequence of an Estradiol-Degrading Strain, Novosphingobium tardaugens NBRC 16725

J. Ibero,a D. Sanz,a B. Galán,a E. Díaz,a J. L. García,a,b

aDepartment of Microbial and Plant Biotechnology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain
bDepartment of Applied Biotechnology, Institute for Integrative Systems Biology (I2SysBio), Universidad de Valencia-CSIC, Valencia, Spain

ABSTRACT In this work we report the complete sequence and assembly of the estradiol-degrading bacterium Novosphingobium tardaugens NBRC 16725 genome into a single contig using the Pacific Biosciences RS II system.

Estrogens are C18 steroid hormones that play several roles in animal physiology (1) but also are important endocrine-disrupting compounds (EDCs) affecting aquatic fauna, even at extremely low concentrations (2). The occurrence and abundance of estrogens in aquatic environments present a serious risk to public health (3), and they have been classified by the World Health Organization as group 1 carcinogens (https://monographs.iarc.fr/list-of-classifications-volumes/).

Although the presence and removal of estrogens via biodegradation have been studied to some extent (4), current knowledge of microbial estrogen degradation pathways is still very limited (5). The bacterial degradation of estrogens is an uncommon trait, and only a few bacteria are known to completely metabolize them (6–8).

The alphaproteobacterium Novosphingobium tardaugens NBRC 16725 (described as strain ARI-1) was isolated at a sewage treatment plant in Tokyo, Japan, and is able to use estradiol, estrone, and estriol as its only carbon source (9, 10). It is a Gram-negative, aerobic, rod-shaped, and nonmotile bacterium. The genome of this strain was previously sequenced using 454 GS-FLX Titanium and Illumina HiSeq 1000 technologies and assembled in 54 contigs (contig \( N_{50} \) 486,386 bp; contig \( L_{50} \) 4 bp) (BioProject number PRJDB314).

Here, we report the complete sequence and assembly of the N. tardaugens NBRC 16725 genome into a single contig using the Pacific Biosciences (PacBio) RS II technology that offers very long and unbiased reads that are uniquely suited for closing genome assemblies.

N. tardaugens NBRC 16725 was purchased from the Leibniz Institute DSMZ—German Collection of Microorganisms and Cell Cultures. The strain was grown at 30°C in nutrient broth medium in an orbital shaker. Genomic DNA was isolated using the phenol chloroform extraction method (11). The G-tube method was used to construct a 10-kb SMRTbell template library of 9,172 kb, which was checked on an Agilent Technologies 2100 Bioanalyzer instrument. The sequencing reagent used was single-molecule real-time (SMRT) cell 8Pac v3 with the DNA polymerase binding kit P6. Sequencing yielded 1,248,591,484 total bases and 205,778 total reads with a mean subread length of 6,067 bp and an \( N_{50} \) value of 7,616 bp. All raw data were deposited in the SRA database under accession number SRR8271516. The PacBio reads were assembled de novo using RS Hierarchical Genome Assembly Process (HGAP) v3.0 software (12) with default options, yielding a single contig with a total length of 4,358,096 bp and a 61.2% GC content (Fig. 1). Pairwise alignment between that single contig sequence and a previous assembly using Geneious v11.0.5 software revealed 99.8% base pair identity, and 175,514 additional nucleotides were sequenced that were
distributed in 26 gaps existing between previously available contigs. The whole assembled nucleotide sequence was deposited in GenBank under the accession number CP034179.

Protein-, rRNA-, and tRNA-coding genes were annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (13) and Rapid Annotations using Subsystems Technology (RAST) (14). This genome contained 4,097 open reading frames (ORFs) with 4,040 coding DNA sequences (CDS), 48 tRNA-coding sequences, and 6 rRNA-coding sequences.

The *N. tardaugens* NBRC 16725 genome (4.36 Mbp) was found to be highly similar (70.79%) to that of *Novosphingobium subterraneum* strain DSM 12447 (15) using average nucleotide identity based on BLAST (ANIb) analysis (16).

The *oeCA, oecB, and oecC* genes involved in estradiol degradation (17) are located in two distinct gene clusters as in *Sphingomonas* sp. strain KC8 (7), covering a 76-kb region that also contains the testosterone degradation cluster (18), which was segregated in the previous assembly. The genome includes a third cluster containing the *sal2, scd2AB*, and *stdA2* genes involved in cholic acid degradation (19, 20).

**Data availability.** This genome sequencing project has been deposited in GenBank under the accession number PRJNA505257. The version described in this paper is version CP034179. All raw data have been deposited in the SRA database under accession number SRR8271516.

**FIG 1** Graphical map of the circular genome of *Novosphingobium tardaugens* NBRC 16725 obtained using the GCView Server (21) comparative genomic tool. Colors show CDS (blue), ORFs (red), GC content (black), GC skew + (green) and GC skew – (purple).
ACKNOWLEDGMENTS

The bioinformatic support of L. Agudo is greatly appreciated. We acknowledge the financial support provided by the Ramón Areces Foundation.

REFERENCES

1. Ryan KJ. 1982. Biochemistry of aromatase: significance to female reproductive physiology. Cancer Res 42:3342s–3344s.
2. Teles M, Gravato C, Paccheco M, Santos MA. 2004. Juvenile sea bass bioavailability, genotoxic and endocrine responses to β-naphthoflavone, 4-nonylphenol and 17β-estradiol individual and combined exposures. Chemosphere 57:147–158. https://doi.org/10.1016/j.chemosphere.2004.02.023.
3. Barbosa MO, Moreira NFF, Ribeiro AR, Pereira MFR, Silva AMT. 2016. Occurrence and removal of organic micropolllutants: an overview of the watch list of EU Decision 2015/495. Water Res 94:257–279. https://doi.org/10.1016/j.watres.2016.02.047.
4. Khanal SK, Xie B, Thompson ML, Sung S, Ong S-K, van Leeuwen J. 2006. Transport and biodegradation of natural estrogens in the environment and engineered systems. Environ Sci Technol 40:6537–6546. https://doi.org/10.1021/es0607739.
5. Yu C-P, Deeb RA, Chu K-H. 2013. Microbial degradation of steroidal estrogens. Chemosphere 91:1225–1235. https://doi.org/10.1016/j.chemosphere.2013.01.112.
6. Kurisu F, Ogura M, Saitoh S, Yamazoe A, Yagi O. 2010. Degradation of natural estrogen and identification of the metabolites produced by soil isolates of Rhodococcus sp. and Sphingomonas sp. J Biosci Bioeng 109:576–582. https://doi.org/10.1016/j.jbiosc.2009.11.006.
7. Yu C-P, Roh H, Chu K-H. 2007. 17β-Estradiol-degrading bacteria isolated from activated sludge. Environ Sci Technol 41:486–492. https://doi.org/10.1021/es060923f.
8. Yu Q, Wang P, Liu D, Gao R, Shao H, Zhao H, Ma Z, Wang D, Hua H. 2016. Degradation characteristics and metabolic pathway of 17β-estradiol (E2) by Rhodococcus sp. DS201. Biotechnol Bioprocess Eng 21:804–813. https://doi.org/10.1007/s12257-016-0283-5.
9. Horiouchi M, Hayashi T, Kudo T. 2012. Steroid degradation in Comamonas testosteroni. J Steroid Biochem Mol Biol 129:4–14. https://doi.org/10.1016/j.jsbmb.2010.10.008.
10. Horinouchi M, Hayashi T, Kudo T. 2012. Steroid degradation in Comamonas testosteroni. J Steroid Biochem Mol Biol 129:4–14. https://doi.org/10.1016/j.jsbmb.2010.10.008.
11. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
12. Chen Y-L, Yu C-P, Lee T-H, Goh K-S, Chu K-H, Wang P-H, Ismail W, Shih C-J, Chiang Y-R. 2017. Biochemical mechanisms and catabolic enzymes involved in bacterial estrogen degradation pathways. Cell Chem Biol 24:712–724.e7. https://doi.org/10.1016/j.chembiol.2017.05.012.
13. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
14. Horinouchi M, Hayashi T, Kudo T. 2012. Steroid degradation in Comamonas testosteroni. J Steroid Biochem Mol Biol 129:4–14. https://doi.org/10.1016/j.jsbmb.2010.10.008.
15. Barrientos Á, Merino E, Casabon I, Rodríguez J, Crowe AM, Holert J, Philipp B, Eltis LD, Olivera ER, Luengo JM. 2015. Functional analyses of three acyl-CoA synthetases involved in bile acid degradation in Pseudomonas putida. Environ Microbiol 17:47–63. https://doi.org/10.1111/1462-2920.12395.
16. Grant JR, Stothard P. 2008. The CGView server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184. https://doi.org/10.1093/nar/gkn179.