![American Society for Microbiology logo] ![Microbiology Resource Announcements logo]

Complete Genome Sequence of an Efficient Vitamin D$_3$-Hydroxylating Bacterium, *Pseudonocardia autotrophica* NBRC 12743

Keitaro Yoshida,[a] **Yoshiaki Yasutake,**[a] **Tomohiro Tamura**[a,b,c]

[a]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo, Japan

[b]Laboratory of Molecular Environmental Microbiology, Graduate School of Agriculture, Hokkaido University, Sapporo, Japan

[c]Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL), AIST, Sapporo, Japan

ABSTRACT *Pseudonocardia autotrophica* NBRC 12743 contains a cytochrome P450 vitamin D$_3$ hydroxylase, and it is used as a biocatalyst for the commercial production of hydroxyvitamin D$_3$, a valuable compound for medication. Here, we report the complete genome sequence of *P. autotrophica* NBRC 12743, which could be useful for improving the productivity of hydroxyvitamin D$_3$.

Hydroxylated vitamin D$_3$ (VD$_3$) and its derivatives are useful as pharmaceuticals for treating VD$_3$ deficiency-related diseases in clinical settings. Several actinomycete species capable of hydroxylating VD$_3$ have been isolated to date (1–4). Among these isolated bacteria, *Pseudonocardia autotrophica* NBRC 12743 showed the highest VD$_3$-hydroxylating activity, and the strain has thus been applied to one-pot commercial production of calcitriol (1α,25-dihydroxyvitamin D$_3$) (1, 5). The gene encoding VD$_3$ hydroxylase (Vdh) was successfully cloned and sequenced (GenBank accession number AB456955) (2), showing that Vdh is a cytochrome P450 monooxygenase. Directed-evolution studies together with three-dimensional structure analyses of Vdh have also been performed to improve its enzymatic activity (2, 6–8). A draft genome sequence of this strain with 117 contigs was reported recently (9).

Here, we sequenced and assembled the genome of *P. autotrophica* strain NBRC 12743. The *P. autotrophica* strain was cultured in Luria-Bertani broth for 72 h at 28°C, and genomic DNA was isolated by the conventional phenol-chloroform extraction method. The genome was sequenced using PacBio RS II and Illumina HiSeq 2500 sequencers with 100-bp paired-end systems (Hokkaido System Science Co., Ltd., Sapporo, Japan). The PacBio raw reads were filtered using the SMRT Analysis program (version 2.3.0), and 115,874 subreads (886,044,835 bp) were obtained. These subreads were assembled with the Hierarchical Genome Assembly Process (HGAP) (version 2) (10), resulting in a circular chromosomal sequence with a mean coverage of 67-fold. The HiSeq paired-end reads were filtered with Trimmomatic (version 0.38) to remove low-quality reads with the parameters SLIDINGWINDOW:20:20 and MINLEN:50 (11), and 44,241,534 reads (4,438,360,972 bp) were obtained. These filtered reads were then mapped to the chromosome sequence using the BWA-MEM algorithm (version 0.7.12) with a seed length of 19, and sequencing errors were corrected using the Genome Analysis Toolkit (GATK) pipeline (version 4.0.6.0) with default parameters (12, 13). The plasmid sequence was assembled from the hybrid reads (both the filtered PacBio and HiSeq reads), using the Unicycler hybrid assembler (version 0.4.6) with default parameters (14). Finally, we identified another circular sequence which does not overlap the chromosome sequence. Genes were annotated using the DDBJ Fast Annotation and Submission Tool (DFAST) pipeline (15).

The genome of NBRC 12743 was inferred to contain two circular sequences with a length of 7,246,130 bp with 72.9% G+C content and a length of 289,155 bp with 70.2% G+C content. The genome contains 7,132 putative coding sequences, 12 rRNAs (4 23S, 16S, and 5S rRNAs each), and 70 tRNAs. We searched for genes encoding cytochrome P450s using the Microbial Genome Annotation Pipeline (http://www.migap.org/) and found that NBRC 12743 possesses 28 cytochrome P450 genes, including the gene encoding Vdh. We found that all of the cytochrome P450 genes exist in the chromosome. The genomic information reported in this study will be helpful for engineering this bacterium as a recombinant expression host of P450 and for improving expression levels of Vdh in order to achieve more efficient bioproduction of hydroxyvitamin D_3.

Data availability. The genome sequence has been deposited in DDBJ/ENA/GenBank under the accession numbers AP018920 and AP018921. The raw sequencing reads have been deposited in the DDBJ Sequence Read Archive under the accession numbers DRX139961 and DRX139962.

REFERENCES

1. Sasaki J, Miyazaki A, Saito M, Adachi T, Mizoue K, Hanada K, Omura S. 1992. Transformation of vitamin D_3 to 1α,25-dihydroxyvitamin D_3 via 25-hydroxyvitamin D_3 using *Amycolata* sp. strains. Appl Microbiol Biotechnol 38:152–157. https://doi.org/10.1007/BF00174460.
2. Fujii Y, Kabumoto H, Nishimura K, Fujii T, Yanai S, Takeda K, Tamura N, Arisawa A, Tamura T. 2009. Purification, characterization, and directed evolution study of a vitamin D_3 hydroxylase from *Pseudonocardia autotrophica*. Biochem Biophys Res Commun 385:170–175. https://doi.org/10.1016/j.bbrc.2009.05.033.
3. Abdulmughni A, Jóźwik IK, Brill E, Hannemann F, Thunnissen A-MWH, Bernhardt R. 2017. Biochemical and structural characterization of CYP109A2, a vitamin D_3 25-hydroxylase from *Bacillus megaterium*. FEBS J 284:3881–3894. https://doi.org/10.1111/febs.14276.
4. Sawada N, Sakaki T, Yoneda S, Kusudo T, Shinkyo R, Ohta M, Inouye K. 2004. Conversion of vitamin D_3 to 1α,25-dihydroxyvitamin D_3 by *Streptomyces griseolus* cytochrome P450SU-1. Biochem Biophys Res Commun 320:156–164. https://doi.org/10.1016/j.bbrc.2004.05.140.
5. Takeda K, Asou T, Matsuda A, Kimura K, Okamura K, Okamoto R, Sasaki J, Adachi T, Omura S. 1994. Application of cyclodextrin to microbial transformation of vitamin D_3 to 25-hydroxyvitamin D_3 and 1α,25-dihydroxyvitamin D_3. J Ferment Bioeng 78:380–382. https://doi.org/10.1016/0922-338X(94)90285-2.
6. Yasutake Y, Fujii Y, Nishioka T, Cheon W-K, Arisawa A, Tamura T. 2010. Structural evidence for enhancement of sequential vitamin D_3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D_3 hydroxylase. J Biol Chem 285:31193–31201. https://doi.org/10.1074/jbc.M110.147009.
7. Imoto N, Nishioka T, Tamura T. 2011. Permeabilization induced by lipid II-targeting lantibiotic nisin and its effect on the bioconversion of vitamin D_3 to 25-hydroxyvitamin D_3 by *Rhodococcus erythropolis*. Biochem Biophys Res Commun 405:393–398. https://doi.org/10.1016/j.bbrc.2011.01.038.
8. Yasutake Y, Nishioka T, Imoto N, Tamura T. 2013. A single mutation at the ferredoxin binding site of P450 Vdh enables efficient biocatalytic production of 25-hydroxyvitamin D_3. Chembiochem 14:2284–2291. https://doi.org/10.1002/cbic.201300386.
9. Grumaz C, Vainshtein Y, Kirstahler P, Luetz S, Kittelmann M, Schroer K, Eggimann FK, Czaja R, Vogel A, Hilberath T, Worsch A, Girhard M, Urlacher VB, Sandberg M, Sohn K. 2017. Draft genome sequences of three *Actinobacteria* strains presenting new candidate organisms with high potentials for specific P450 cytochromes. Genome Announc 5:e00532-17. https://doi.org/10.1128/genomeA.00532-17.
10. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
11. 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.
12. Li H. 2012. Exploring single-sample SNP and INDEL calling with whole-genome *de novo* assembly. Bioinformatics 28:1838–1844. https://doi.org/10.1093/bioinformatics/bts280.
13. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.
14. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
15. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039. https://doi.org/10.1093/bioinformatics/btx713.