




# Draft Genome Sequence of FK506-Producing *Streptomyces tsukubensis* Strain VKM Ac-2618D

Veronika Y. Poshekhontseva,<sup>a,b</sup> Eugeny Y. Bragin,<sup>a</sup>  Victoria V. Fokina,<sup>a</sup> Victoriya Y. Shtratnikova,<sup>c</sup> Irina P. Starodumova,<sup>a</sup> Sergey V. Tarlachkov,<sup>a,d</sup> Marina V. Donova<sup>a,e</sup>

<sup>a</sup>G. K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Federal Research Center, Pushchino, Russia

<sup>b</sup>Pushchino State Natural Science Institute, Pushchino, Russia

<sup>c</sup>Department of Bioengineering and Bioinformatics, M. V. Lomonosov Moscow State University, Moscow, Russia

<sup>d</sup>Branch of Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Pushchino, Russia

<sup>e</sup>Pharmins, Ltd., R&D, Pushchino, Russia

**ABSTRACT** The 23-membered macrolide tacrolimus (FK506) is an important immunosuppressant that is widely used in the prevention of graft rejection and in the treatment of inflammatory skin diseases and immune diseases. We report here the draft genome sequence of the FK506 producer *Streptomyces tsukubensis* VKM Ac-2618D.

Tacrolimus (FK506) is widely used as an immunosuppressant after medulla, kidney, and heart transplantation (1). The FK506 yield from *Streptomyces tsukubensis* VKM Ac-2618D was higher (2) than that of *S. tsukubensis* NRRL 18488<sup>T</sup>, a wild-type progenitor of the industrially used high-FK506-producing strains (3).

*S. tsukubensis* VKM Ac-2618D is a Gram-positive soil-dwelling bacterial strain. It was previously identified as *S. tsukubaensis* based on 16S rRNA gene sequence analysis and chemotaxonomic, morphological, and physiological characteristics (2, 4).

Genomic DNA extraction from *S. tsukubensis* VKM Ac-2618D was carried out as described (5) with the following modifications. The cells from one colony were subcultured from agar into 50 ml of nutrient medium containing 20 g/liter starch, 20 g/liter glucose, and 18 g/liter peptone at pH 7.0 on a rotary shaker (220 rpm) at 25°C for 48 h. Then, 50 ml of the same medium was inoculated with 10% of the cell suspension obtained, and a second cultivation step was performed for 24 h. Bacterial biomass was harvested by centrifugation at 8,000 × *g* for 10 min and washed twice with sterile saline. Then, the cells were suspended in 3 ml Tris-EDTA (TE) buffer (25 mM Tris-HCl [pH 8], 10 mM EDTA) and 1 mg/ml lysozyme. SDS and proteinase K were added to final concentrations of 1% and 400 μg, respectively, and the suspension was incubated for 1 h at 54°C. RNase was added to a final concentration of 50 μg/ml, and the mixture was incubated for 30 min at 54°C. DNA was purified by sequential extractions of the liquid phase with phenol and the phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol/vol]) and chloroform-isoamyl alcohol (24:1 [vol/vol]) mixtures and precipitated with 0.7 volume of isopropanol at room temperature at 8,000 × *g* for 15 sec. Precipitated DNA was collected by centrifugation (4,000 × *g*, 4°C, 10 min), washed twice with 70% ethanol, air dried, and resuspended in TE buffer.

Fragmentation of the genomic DNA was done by sonication with a Covaris S220 instrument. The short-read library containing DNA fragments of 300- to 400-bp insert lengths was prepared with a NEBNext Ultra II DNA library prep kit (Illumina). The library was sequenced on an Illumina HiSeq 2500 platform with a HiSeq rapid PE cluster kit v2 and a HiSeq rapid SBS kit v2 (500 cycles).

Default parameters were used for all software unless otherwise specified. The read quality was checked with FastQC v0.11.8 (6). Raw Illumina sequencing data were

**Citation** Poshekhontseva VY, Bragin EY, Fokina VV, Shtratnikova VY, Starodumova IP, Tarlachkov SV, Donova MV. 2019. Draft genome sequence of FK506-producing *Streptomyces tsukubensis* strain VKM Ac-2618D. Microbiol Resour Announc 8:e00510-19. <https://doi.org/10.1128/MRA.00510-19>.

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

**Copyright** © 2019 Poshekhontseva et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Victoria V. Fokina, [2vvfokina@gmail.com](mailto:2vvfokina@gmail.com).

**Received** 2 May 2019

**Accepted** 19 May 2019

**Published** 13 June 2019

adapter trimmed using BBDuk v38.35 (7) (with the parameters ktrim, r; k, 23; mink, 11; hdist, 1; tpe; minlen, 30; and ref, adapters), and then filtered to remove the known Illumina artifacts, and PhiX using BBDuk v38.35 (7) (with the parameters k, 31; ref, artifacts, phix; and cardinality). To remove possible contamination by human DNA, the reads were mapped to masked versions of a human reference genome (hg19) and discarded if the identity exceeded 95% using BMap v38.35 (7) with the following parameters: minid, 0.95; maxindel, 3; bwr, 0.16; bw, 12; quickmatch; fast; and minhits, 2. The remaining reads were trimmed by quality scores using BBDuk v38.35 (7) with the following parameters: qtrim, r; trimq, 18; and minlen, 30.

Clean sequencing data were assembled using SPAdes v3.13.0 (8) with the parameters --careful and --cov-cutoff auto, and the resulting contigs were discarded if their length was <200 bp. Genome annotation was performed with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (9). The average nucleotide identities based on BLAST (ANIb) and digital DNA-DNA hybridization (dDDH) values were calculated using JSpecies v1.2.1 (10) and the Genome-to-Genome Distance Calculator (GGDC) v2.1 (11), respectively.

A total of  $3.6 \times 10^6$  paired-end reads with 251-bp lengths (1.8 Gb) were obtained from the sequencing; 91.9% of the bases had a quality score of over 30. As a result,  $3.5 \times 10^6$  clean paired-end reads with an average 226-bp length (1.6 Gb) were assembled into 79 scaffolds with 202-fold coverage. The scaffold  $N_{50}$  value is 971,906 bp, and the largest scaffold is 1,270,971 bp. The genome size is 7.93 Mb with an average G+C content of 71.9%. A total of 6,264 protein-coding genes (1,227 of which are hypothetical proteins), 66 tRNAs, 23 complete or partial rRNAs, and 3 noncoding RNAs (ncRNAs) were predicted.

The ANI and dDDH values calculated between the genome sequence of VKM Ac-2618D and the genome sequence of the type strain of *S. tsukubensis* NRRL 18488 (AJSZ00000000) were 99.97% and 99.9%, respectively, which are significantly higher than the threshold values (95 to 96% ANI and 70% dDDH) for species delineation (10–12). The results confirm that strain VKM Ac-2618D belongs to the species *S. tsukubensis*.

As expected, the tacrolimus biosynthesis cluster was found in the genome of the strain VKM Ac-2618D. According to the alignment, 8 nucleotide substitutions and 2 indels were observed between the gene clusters of VKM Ac-2618D (GenBank accession no. SGFG01000008; region of interest, nucleotides 83700 to 166586) and the type strain NRRL 18488 (GenBank accession no. AJSZ01000908; complementary strand of region of interest, nucleotides 93310 to 10426), while 100% identity was found for VKM Ac-2618D with the corresponding sequence of the *Streptomyces* sp. strain KCTC 11604BP (GenBank accession no. HM116537; region of interest, nucleotides 8712 to 91598).

The data obtained are important for the development of genome-based improvement of the *S. tsukubensis* strain to achieve higher tacrolimus production rates.

**Data availability.** The raw reads have been deposited in the NCBI SRA under the accession no. SRR8649924, and the whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession no. SGFG00000000. The version described in this paper is the first version, SGFG01000000.

## ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Higher Education of Russia (unique project identifier RFMEFI58818X0008).

## REFERENCES

1. Ordóñez-Robles M, Santos-Beneit F, Martín JF. 2018. Nutritional unraveling regulation of tacrolimus biosynthesis in *Streptomyces tsukubaensis* through *omic* approaches. *Antibiotics* (Basel) 7:39. <https://doi.org/10.3390/antibiotics7020039>.
2. Sukhodolskaya GV, Lobastova TG, Gulevskaya SA, Fokina VV, Shutov AA, Nikolaeva VM, Donova MV. 2013. Method of obtaining tacrolimus by microbiological synthesis. Russian patent 2495937.
3. Goranovič D, Blažič M, Magdevska V, Horvat J, Kuščer E, Polak T, Santos-Aberturas J, Martínez-Castro M, Barreiro C, Mrak P, Kopitar G, Kosec G, Fujs Š, Martín JF, Petković H. 2012. FK506 biosynthesis is regulated by

- two positive regulatory elements in *Streptomyces tsukubaensis*. BMC Microbiol 12:238. <https://doi.org/10.1186/1471-2180-12-238>.
4. Poshekhontseva VY, Gulevskaya SA, Donova MV. 2014. Comparative study of pheno- and genotype features and identification of *Streptomyces* spp. strains producing tacrolimus. J Sci Foundation Biol 3:119–122. (In Russian). <http://biologyfond.ru/files/14-15.11.2014/bf3.pdf#page=120>.
  5. Marmur J. 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. Mol Biol 3:208–216. [https://doi.org/10.1016/S0022-2836\(61\)80047-8](https://doi.org/10.1016/S0022-2836(61)80047-8).
  6. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
  7. Bushnell B. 2014. BBTools software package. <https://sourceforge.net/projects/bbmap/>.
  8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
  9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
  10. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106: 19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
  11. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. <https://doi.org/10.1186/1471-2105-14-60>.
  12. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. <https://doi.org/10.1099/ijsem.0.002516>.