Draft Genome Sequence of *Stenotrophomonas pavanii* Strain MHSD12, a Bacterial Endophyte Associated with *Dicoma anomala*

Mehabo P. Maela,ᵃ Mahloro H. Serepa-Dlaminiᵃ

ᵃDepartment of Biotechnology and Food Technology, University of Johannesburg, Johannesburg, South Africa

ABSTRACT *Stenotrophomonas pavanii* strain MHSD12 is an endophyte isolated from *Dicoma anomala* surface-sterilized leaves. Here, we report its draft genome sequence (4.3 Mb) assembled to 30 contigs, with 3,923 protein-coding genes. The genome sequence highlights important genes for an endophytic lifestyle.

*S*tenotrophomonas spp. are widespread and were previously isolated from diverse environments, such as soil, sewage, compost, human clinical samples, petrochemical waste, and plants (1). Several *Stenotrophomonas* species have symbiotic interactions with plants and have been isolated as endophytes (2–5). *Stenotrophomonas pavanii* strain MHSD12 was isolated as an endophyte from healthy surface-sterilized leaves of *Dicoma anomala*. *Dicoma anomala* aerial tissues were collected from Eisleben, Limpopo Province, South Africa. Leaf samples were serially sterilized with 70% ethanol and 5% sodium hypochlorite (5) and rinsed 3 times with autoclaved tap water. The surface-sterilized leaves were crushed using a sterile mortar and pestle and macerated with phosphate-buffered saline (8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and KH_2PO_4 at pH 7.4). The homogenate was streaked onto nutrient agar and incubated at 28°C for 2 to 7 days, followed by subculturing single colonies 3 times. *Stenotrophomonas pavanii* strain MHSD12 was initially identified by phylogenetic analysis of its 16S rRNA gene (GenBank accession number MN078164). The genome sequence of *S. pavanii* strain MHSD12 will augment studies on the plant-bacterium interaction and highlight important genes responsible for an endophytic lifestyle.

For genomic DNA extraction, MHSD12 was grown on nutrient broth at 28°C for 48 h. Genomic DNA was extracted using the NucleoSpin microbial DNA extraction kit (Macherey-Nagel, Germany), following the manufacturer's protocol. The obtained DNA was sequenced on an Illumina platform at a commercial service provider (Agricultural Research Council [ARC], Onderstepoort, South Africa). Paired-end libraries (2 × 150 bp) were generated using the Nextera DNA sample preparation kit (Illumina, USA), and sequencing was performed using the HiSeq 2500 platform.

All of the preannotation analyses were performed on the Galaxy Web server (https://usegalaxy.org) (6), using default parameters. Quality control of the raw reads was performed by FastQC version 0.72 (7). Sequence reads were *de novo* assembled using Unicycler version 0.4.6.0 (8), and the assembly quality was assessed with QUAST version 0.4.6.3 (9). The draft genome sequence of strain MHSD12 was submitted to NCBI for automated annotation using Prokaryotic Genome Annotation Pipeline (PGAP) (10) and the Rapid Annotations using Subsystems Technology (RAST) server (11–13). The genome sequence data of *S. pavanii* strain MHSD12 was submitted to the Type (Strain) Genome Server (https://tygs.dsmz.de) for a whole-genome-based taxonomic analysis with other validly published type strains (14). Additionally, the average nucleotide identity (ANI) value with closely related species was determined using the Orthologous Average Nucleotide Identity Tool (OAT) software (15).

The sequencing platform produced 2,587,651 sequence reads and allowed a 57-fold coverage. The genome sequence was assembled to 30 contigs with an N_{50} value of 403,304 bp, a total genome size of 4,385,734 bp, and a G+C content of 67.34%. A total of 4,025 genes, including 3,923 protein-coding genes and 73 RNA genes, were identified. The RAST annotation identified 1,495 subsystems with amino acids and derivatives as the dominant (17%) category, and categories that play a part in plant promotion and growth, such as nitrogen and phosphate metabolism, iron acquisition and virulence, and disease and defense, were identified. The highest similarity was with *Stenotrophomonas pavanii* strain DSM 25135 which is an endophyte isolated from sugar cane (16); MHSD12 had a 93.2% digital DNA-DNA hybridization value and a 98% ANI value with DSM 25135.

Data availability. This whole-genome shotgun project and associated data have been deposited at DDBJ/ENA/GenBank under the accession number JAAKGL000000000, BioProject accession number PRJNA607646, and BioSample accession number SAMN14142486. The version described in this paper is the first version, JAAKGL010000000. The raw sequence reads are available at SRR11955405.

REFERENCES

1. Svensson-Stadler LA, Mihaylova SA, Moore ERB. 2012. *Stenotrophomonas* interspecies differentiation and identification by *gyrB* sequence analysis. FEMS Microbiol Lett 327:15–24. https://doi.org/10.1111/j.1574-6968.2011.02452.x.

2. Feng F, Ge J, Li Y, He S, Zhong J, Liu X, Yu X. 2017. Enhanced degradation of chlorpyrifos in rice (*Oryza sativa L.*) by five strains of endophytic bacteria and their plant growth promotional ability. Chemosphere 184: 505–513. https://doi.org/10.1016/j.chemosphere.2017.05.178.

3. Liu Y, Zuo S, Xu L, Zou Y, Song W. 2012. Study on diversity of endophytic bacterial communities in seeds of hybrid maize and their parental lines. Arch Microbiol 194:1001–1012. https://doi.org/10.1007/s00203-012-0836-8.

4. Rustamova N, Wubulikasimu A, Mukhamedov N, Gao Y, Egamberdieva D, Yili A. 2020. Endophytic bacteria associated with medicinal plant Vernonia anthelmintica: diversity and characterization. Curr Microbiol https://doi.org/10.1007/s00284-020-01924-5.

5. Miguel PSB, de Oliveira MNV, Delvaux JC, de Jesus GL, Borges AC, Tótola MR, Neves JC, Costa MD. 2016. Diversity and distribution of the endophytic bacterial community at different stages of Eucalyptus growth. Antonie Van Leeuwenhoek 109:755–771. https://doi.org/10.1007/s10482-016-0676-7.

6. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. Nucleic Acids Res 44:W3–W10. https://doi.org/10.1093/nar/gkw343.

7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Cambridge, United Kingdom. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

8. 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.

9. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

10. 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.

11. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

12. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:206–214. https://doi.org/10.1093/nar/gkt1226.

13. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.

14. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3.

15. Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760.

16. Ramos PL, Van Trappen S, Thompson FL, Rocha RCS, Barbosa HR, De Vos P, Moreira-Filho CA. 2011. Screening for endophytic nitrogen-fixing bacteria in Brazilian sugar cane varieties used in organic farming and description of *Stenotrophomonas pavanii* sp. nov. Int J Syst Evol Microbiol 61:926–931. https://doi.org/10.1099/ijs.0.019372-0.