Check for updates

Draft Genome Sequence of *Enterobacter hormaechei* Strain MHSD6, a Plant Endophyte Isolated from Medicinal Plant *Pellaea calomelanos*

Khuthadzo Tshishonga,[a] (ID) Mahloro H. Serepa-Dlamini[a]

[a]Department of Biotechnology and Food Technology, University of Johannesburg, Johannesburg, South Africa

ABSTRACT We describe here the draft genome sequence of *Enterobacter hormaechei* strain MHSD6, a bacterial endophyte isolated from the medicinal plant *Pellaea calomelanos*. The *Enterobacter hormaechei* strain MHSD6 draft genome is 4,817,102 bp in length, with a G+C content of 55.50%.

Endophytes are a group of microorganisms, mostly bacteria and fungi, which colonize internal plant tissues without causing any harm (1). Most plant species host a variety of endophytes, and therefore novel endophytic microorganisms exist within plant species growing in unique environments (2). Endophytes have a symbiotic relationship with their host plants. Plants provide the endophyte with carbon for energy, as well as with an environment for growth and survival, and in turn, endophytes function as biological defense for the plant through production of secondary metabolites with toxicity against infected host cells and phytopathogens (2, 3). In addition, endophytes promote plant growth through production of phytohormones and nitrogen fixation, as well as by solubilization of minerals such as phosphorus (4). Endophytes can produce the same or similar secondary metabolites as their plant hosts; thus, endophytes associated with medicinal plants offer alternative sources for isolation of bioactive compounds (4). Endophyte genomes will enhance understanding of their symbiotic relationships with plants and further reveal information on how their genomic profiles differ from those of clinical or pathogenic strains.

Enterobacter hormaechei strain MHSD6 was isolated from sterilized leaves of the medicinal plant *Pellaea calomelanos*, which was obtained from Botlokwa (23°2934.8′S, 29°4211.2′E) in Limpopo Province, South Africa. The bacterial endophyte was isolated from sterilized leaves and cultivated as described by Mahlangu and Serepa-Dlamini (5). Briefly, following serial sterilization with 70% ethanol and 1% sodium hypochlorite, the plant leaves were ground in 2 ml of saline, and the homogenate was streaked onto nutrient agar plates, followed by incubation at 28°C for 48 h. The chromosomal genome was extracted from solid bacterial colonies using a NucleoSpin microbial DNA extraction kit following the manufacturer's protocol. The concentration and quality of isolated DNA were determined using a NanoDrop ND-2000 UV-visible (UV-Vis) spectrophotometer. Paired-end libraries were generated using a NEBNext Ultra II DNA kit (Illumina) and sequenced with a paired-end sequencing strategy (2 × 300 bp) using an Illumina MiSeq instrument v3 at a commercial service provider (Agricultural Research Council [ARC] Biotechnology Platform, Pretoria, South Africa).

Quality control and assembly of the raw sequence reads were performed on the Galaxy web platform (https://usegalaxy.org) (6). FastQC v0.69 was used to assess the quality of the raw reads (7). The sequence reads were *de novo* assembled with Unicycler v0.4.1.1 (8) and assessed with Quast v4.6.3 (9), using default parameters. The complete genome sequence was submitted to NCBI and annotated using Prokaryotic Genome Annotation Pipeline (PGAP) (10) and the Rapid Annotations using Subsystems Tech-

TABLE 1 Genome data of *Enterobacter hormaechei* strain MHSD6

Attribute	Value
Genome size (bp)	4,817,102
G+C content (%)	55.50
Total no. of genes	4,779
No. of protein coding genes	4,691
No. of RNAs	88
No. of rRNA genes	6
No. of tRNA genes	76
No. of genomic islands	33
No. of CRISPRs	2

nology (RAST) server (11–13). The genomic islands were identified by screening the PGAP annotation file generated from NCBI on the IslandViewer 4 website (http://www .pathogenomics.sfu.ca/islandviewer/) (14). Clustered regularly interspaced short palindromic repeats (CRISPR) were predicted by CRISPRCasFinder software (15–17).

The Illumina MiSeq platform generated 250 Mb of sequence reads. The draft genome of strain MHSD6 produced 60 contigs with a total genomic length of 4,817,102 bp and a G+C content of 55.50% (Table 1). It consists of a total of 4,779 genes; 4,691 of the genes are protein coding genes, 88 are RNA genes, 143 are pseudogenes, and 6 are noncoding RNA (ncRNA) genes. The RNA coding genes predicted include 76 tRNA and 6 rRNA (5S, 16S, and 23S) genes. We identified genes responsible for nitrogen fixation, phytohormone production, transport proteins, and transcriptional regulators, all of which may play a part in symbiosis with the plant and promote plant growth.

Data availability. This whole-genome shotgun project and associated data have been deposited at DDBJ/ENA/GenBank under the accession number VHQJ00000000, BioProject accession number PRJNA550302, BioSample accession number SAMN12124847, and SRA accession number SRR10237366. The version described in this paper is version VHQJ02000000.

ACKNOWLEDGMENT

REFERENCES

1. Bacon CW, Hinton DM. 2007. Bacterial endophytes: the endophytic niche, its occupants, and its utility, p 155–194. *In* Gnanamanickam SS (ed), Plant-associated bacteria. Springer, Dordrecht, Netherlands.
2. Strobel GA. 2003. Endophytes as sources of bioactive products. Microbes Infect 5:535–544. https://doi.org/10.1016/S1286-4579(03)00073-X.
3. Berg G, Hallmann J. 2006. Control of plant pathogenic fungi with bacterial endophytes, p 53–69. *In* Schulz B, Boyle C, Sieber T (ed). Microbial root endophytes. Springer, Berlin, Germany.
4. Alvin A, Miller KI, Neilan BA. 2014. Exploring the potential of endophytes from medicinal plants as sources of antimycobacterial compounds. Microbiol Res 169:483–495. https://doi.org/10.1016/j.micres.2013.12.009.
5. Mahlangu SG, Serepa-Dlamini MH. 2018. First report of bacterial endophytes from the leaves of *Pellaea calomelanos* in South Africa. S Afr J Sci 114:55–63. https://doi.org/10.17159/sajs.2018/4235.
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. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman F. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-

scale datasets. Nucleic Acids Res 45:30–35. https://doi.org/10.1093/nar/gkx343.

15. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 3:52–57. https://doi.org/10.1093/nar/gkm360.

16. Abby SS, Néron B, Ménager H, Touchon M, Rocha E. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726.

17. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for *cas* proteins. Nucleic Acids Res 46:246–251. https://doi.org/10.1093/nar/gky425.