



Whole-Genome Sequence of Endophytic Plant Growth-Promoting *Escherichia coli* USML2

Munirah Tharek, Kee-Shin Sim, Dzulaikha Khairuddin, Amir Hamzah Ghazali, Nazalan Najimudin

School of Biological Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia

ABSTRACT *Escherichia coli* strain USML2 was originally isolated from the inner leaf tissues of surface-sterilized phytopathogenic-free oil palm (*Elaeis guineensis* Jacq.). We present here the whole-genome sequence of this plant-endophytic strain. The genome consists of a single circular chromosome of 4,502,758 bp, 4,315 predicted coding sequences, and a G+C content of 50.8%.

Escherichia coli is a versatile harmless inhabitant of the gastrointestinal tract that has the ability to survive, adapt, and actively grow in extraintestinal environments, including plants (1–3). Their occurrence *in planta* offers a less competitive niche due to a gain in relative protection against adverse conditions present *ex planta* (2). In fact, interior plant tissues have become a favorable niche for bacteria capable of plant invasion due to the availability of abundant nutrients and stable environment (4). Here, we report the complete genome sequence of the plant-origin *E. coli* strain USML2, isolated from inner leaf tissues of surface-sterilized phytopathogenic-free oil palm (*Elaeis guineensis* Jacq.). Interestingly, this *E. coli* strain has the capability of *in planta* ascending migration and growth promotion of its host plant.

The whole genome was sequenced using the PacBio RSII system (Pacific Biosciences, Menlo Park, CA), with a 10-kbp single-pass SMRTbell library and C2-P4 chemistry, yielding ~74× coverage. PacBio genome data were assembled using the Hierarchical Genome Assembly Process (HGAP) version 3 (SMRT Portal version 2.3.0) workflow, including consensus polishing using the Quiver algorithm. The circular genome was validated by dot plot, BLAST, and trimmed by script. The circularized genome data were annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP). Additionally, automatic annotation was enriched using the RASTtk module from Rapid Annotations using Subsystems Technology (RAST) version 2.0 (5–7).

The complete genome of *E. coli* strain USML2 is composed of a single circular chromosome of 4,502,758 bp, with an average G+C content of 50.8%. The genome annotation predicted 4,442 genes, 4,315 coding sequences (CDSs), 22 rRNAs, 86 tRNAs, 19 noncoding RNAs, and 87 pseudogenes.

In relation to its *in planta* existence, genes putatively involved in its endophytic establishment were identified; they included genes for flagellar biosynthesis (*flgLKJIH-GFEDCBAMN*, *flirQPONMLKJIHGFE*, *flitSDCZY*, *flhDC*, *motAB*, and *flhBAE*), chemotaxis activity (*cheA*, *cheW*, *cheR*, *cheB*, *cheY*, and *cheZ*), pilus production (*pilABC* and *pilM-NOPQ*) for root adhesion, and cellulose and pectin degradation enzymes for colonization and invasion. Genes involved in plant growth promotion were also recognized, including the gene for solubilization of phosphate and potassium, production of 1-aminocyclopropane-1-carboxylate (ACC) deaminase (which enhances plant growth by lowering plant ethylene levels), and those coding for enzymes in the biosynthesis of the auxinic phytohormone indole-3-acetic acid (IAA). The presence of these genes

Received 16 March 2017 Accepted 20 March 2017 Published 11 May 2017

Citation Tharek M, Sim K-S, Khairuddin D, Ghazali AH, Najimudin N. 2017. Whole-genome sequence of endophytic plant growth-promoting *Escherichia coli* USML2. Genome Announc 5:e00305-17. <https://doi.org/10.1128/genomeA.00305-17>.

Copyright © 2017 Tharek 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 Amir Hamzah Ghazali, amirhg@usm.my.

highlighted the potential of *E. coli* strain USML2 as a plant growth-promoting endophyte.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [CP011124](#).

ACKNOWLEDGMENTS

This work was supported by the Ministry of Higher Education (MOHE), Malaysia, Fundamental Research Grant Scheme (FRGS) to Amir Hamzah Ghazali under grant number 203/PBIOLOGI/6711414.

We declare no conflicts of interest in the submission of this work. All authors have read and approved the manuscript.

REFERENCES

- Blount ZD. 2015. The unexhausted potential of *E. coli*. *eLife* 4:e05826. <https://doi.org/10.7554/eLife.05826>.
- Méric G, Kemsley EK, Falush D, Saggars EJ, Lucchini S. 2013. Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environ Microbiol* 15:487–501. <https://doi.org/10.1111/j.1462-2920.2012.02852.x>.
- Nautiyal CS, Rehman A, Chauhan PS. 2010. Environmental *Escherichia coli* occur as natural plant growth-promoting soil bacterium. *Arch Microbiol* 192:185–193. <https://doi.org/10.1007/s00203-010-0544-1>.
- Ikeda S, Okubo T, Anda M, Nakashita H, Yasuda M, Sato S, Kaneko T, Tabata S, Eda S, Momiyama A, Terasawa K, Mitsui H, Minamisawa K. 2010. Community- and genome-based views of plant-associated bacteria: plant-bacterial interactions in soybean and rice. *Plant Cell Physiol* 51: 1398–1410. <https://doi.org/10.1093/pcp/pcq119>.
- 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>.
- 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 genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
- 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>.