



Draft Genome Sequence of *Grammothele lineata* SDL-CO-2015-1, a Jute Endophyte with a Potential for Paclitaxel Biosynthesis

Avizit Das,^a Oly Ahmed,^a A. K. M. Abdul Baten,^b Samira Bushra,^a
M. Tariqul Islam,^{a*} Ahlan Sabah Ferdous,^a Mohammad Riazul Islam,^a
Haseena Khan^a

Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh^a; Southern Cross Plant Science, Southern Cross University, Lismore, New South Wales, Australia^b

ABSTRACT *Grammothele lineata* strain SDL-CO-2015-1, a basidiomycete fungus, was identified as an endophyte from a jute species, *Corchorus olitorius* var. 2015, and found to produce paclitaxel, a diterpenic polyoxygenated pseudoalkaloid with anti-tumor activity. Here, we report the draft genome sequence (42.8 Mb with 9,395 genes) of this strain.

Grammothele lineata strain SDL-CO-2015-1 is a basidiomycote belonging to the family *Polyporaceae*; it has been isolated as an endophyte of *Corchorus olitorius* var. 2015. Current studies in our lab have revealed that this strain has the potential to produce a diterpenic polyoxygenated pseudoalkaloid-paclitaxel (taxol), an antitumor drug, in culture conditions (1); however, the molecular mechanisms and pathways of taxol production in fungus have not been fully elucidated (2). In order to attain the genomic support for the existence of an independent paclitaxel biosynthesis pathway, we report here the draft genome sequence of *G. lineata* SDL-CO-2015-1.

The strain was cultured in potato dextrose agar (PDA) medium containing 50 ng/mL of ampicillin. Total genomic DNA was isolated using a modified CTAB-SDS DNA extraction protocol developed in our laboratory. Genomic DNA was sequenced using Illumina HiSeq and PacBio RSII with one SMRT cell (1st BASE, Malaysia). The quality of the reads was checked using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and the adapter trimming was done with Cutadapt (3). Both Illumina reads and PacBio reads were utilized to maximize the contiguity of the resulting genome assembly sequences. The genome was assembled by the SPAdes assembler version 3.9.1 (4), which involved the following steps: multi-*k*-mer-based assembly; long-read error correction; and assembly and scaffolding of the contigs. Pilon version 1.16 (5) was used to polish the assembled data in order to remove small mismatch or insertion/deletion errors that are caused by the inherent biases of the sequencing platforms. The assembled genome has a total size of 42,856,078 bp, with 1,276 scaffolds and an N_{50} value of 169,521 bp. The longest scaffold has a length of 810,186 bp.

We used proteins and expressed sequence tags from three closely related species as references for evidence-based gene prediction. MAKER version 2-32 (6) identified 9,395 high-confidence gene models with an average gene length of 2,271.457 bp. BUSCO version 2 (7) was used to assess the completeness of the genome, which reported the genome to be 93.10% complete. InterProScan version 5.17-56.0 software (8) was used to search and assign gene ontology (GO) terms to these gene models; out of the 9,395 genes, 6,519 were found to have at least 1 GO term assigned to them (in total, 13,534 GO terms were assigned to 6,519 genes). The GC content of the genome was deter-

Received 4 July 2017 Accepted 7 July 2017 Published 17 August 2017

Citation Das A, Ahmed O, Baten AKMA, Bushra S, Islam MT, Ferdous AS, Islam MR, Khan H. 2017. Draft genome sequence of *Grammothele lineata* SDL-CO-2015-1, a jute endophyte with a potential for paclitaxel biosynthesis. *Genome Announc* 5:e00825-17. <https://doi.org/10.1128/genomeA.00825-17>.

Copyright © 2017 Das 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 Haseena Khan, haseena@du.ac.bd.

* Present address: M. Tariqul Islam, Pennsylvania State University, University Park, Pennsylvania, USA.

mined to be 56.8%. RepeatMasker (<http://www.repeatmasker.org>), with the latest version of the repbase database, was used to identify the known repetitive elements; 3.61% of the genome was found to be repetitive, and 2.8% of those repeats were unclassified, as they are newly identified repeats (*de novo*) specific to this species.

As the expected biosynthetic pathway for paclitaxel in fungi is expected to be different from that of plants (9), our reported genomic data of *G. lineata* SDL-CO-2015-1 will be of colossal importance in understanding taxol production and metabolic manipulation in this strain in particular and fungi in general.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [NDFF0000000](https://doi.org/10.1093/nar/gki442). The version described in this paper is the second version, NDFF02000000.

ACKNOWLEDGMENTS

We acknowledge the use of bioinformatics resources of Southern Cross University, Lismore, NSW, Australia, in this project.

We also acknowledge the Higher Education Quality Enhancement Project (HEQEP) funded by the World Bank for providing financial support (HEQEP CP-3250).

REFERENCES

1. Das A, Rahman MI, Ferdous AS, Amin A, Rahman MM, Nahar N, Uddin MA, Islam MR, Khan H. 2017. An endophytic basidiomycete, *Grammothele lineata*, isolated from *Corchorus olitorius*, produces paclitaxel that shows cytotoxicity. PLoS One 12:e0178612. <https://doi.org/10.1371/journal.pone.0178612>.
2. Heinig U, Scholz S, Jennewein S. 2013. Getting to the bottom of Taxol biosynthesis by fungi. Fungal Divers 60:161–170. <https://doi.org/10.1007/s13225-013-0228-7>.
3. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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>.
5. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
6. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18: 188–196. <https://doi.org/10.1101/gr.6743907>.
7. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
8. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Res 33: W116–W120. <https://doi.org/10.1093/nar/gki442>.
9. Schmidt-Dannert C. 2014. Biosynthesis of terpenoid natural products in fungi. Adv Biochem Eng Biotechnol 148:19–61.