

# Draft genome sequence of *Trichoderma asperellum* MLT1J1, isolated from coconut husk in Maluku, Indonesia

Ika Octariyani Safitri<sup>1</sup> , Dian Anggraini Suroto<sup>1</sup>, Sardjono Sardjono<sup>1</sup>, Muhammad Nur Cahyanto<sup>1</sup> , and Jaka Widada<sup>2\*</sup> 

<sup>1</sup>Department of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Gadjah Mada University, Jl. Flora, Bulaksumur, Yogyakarta 55281, Indonesia

<sup>2</sup>Department of Agricultural Microbiology, Faculty of Agriculture, Gadjah Mada University, Jl. Flora, Bulaksumur, Yogyakarta 55281, Indonesia

(Received June 13, 2022; Revised September 3, 2022; Accepted September 13, 2022)

A whole genome sequence was performed on *Trichoderma asperellum* MLT1J1 isolated from coconut husk in Maluku, Indonesia. This strain is a white fungus that has glucose-resistant properties in the production of alpha-amylase and glucoamylase. In this study, the genome sequence of *Trichoderma asperellum* MLT1J1 was sequenced using Illumina NovaSeq PE150. The genome assembly has a length of 48.66 Mb with a GC content of 52.32%, 471 scaffolds and 14,103 protein-coding genes. Based on additional analyses, the number of genes encoding carbohydrate-active enzymes and secondary metabolites gene clusters were revealed. The result of this study will provide useful genomic information that can be compared with other *Trichoderma* species.

**Keywords:** *Trichoderma asperellum*, genome sequence, glucose resistant

*Trichoderma* is a thread fungi taxon commonly found in natural habitats (Li *et al.*, 2017). *Trichoderma* is also widely reported to play an essential role in producing enzymes, especially cellulases. In this study, *Trichoderma asperellum* MLT1J1 was isolated from coconut husk in Maluku, Indonesia. The *T. asperellum* MLT1J1 has glucose-resistant properties in the production of alpha-amylase and glucoamylase (Safitri *et al.*, 2022). This condition is closely related to carbon catabolite repression, which is generally owned by fungi and inhibits enzyme production.

The genomic DNA of *Trichoderma asperellum* MLT1J1

was extracted using the SDS method (Lim *et al.*, 2016). Beijing Novogene Bioinformatics Technology Co., Ltd performed the whole genome sequencing. The library for sequencing was generated using NEBNext<sup>®</sup> Ultra<sup>™</sup> DNA preparation kit according to the manufacturer's instructions and sequenced on Illumina NovaSeq PE150 with coverage of 100x. The assembly was first performed using SOAPdenovo (version 2.04) (Luo *et al.*, 2012), SPAdes (Bankevich *et al.*, 2012), AbySS assembling software, and finally using CISA (Lin and Liao, 2013) software for integration. The genome size of *Trichoderma asperellum* MLT1J1 is 48.66 Mb with a total length of 18,073,466 bp with a gene length/genome value of 37.09%. The genome contains 471 scaffolds with a maximum scaffold size of 3,214,281 bp with an N50 value of 996,985 and N90 value of 84,280 and a GC content of 52.32% (Table 1). Among the 14 species of *Trichoderma* reported, this species has the largest genome size (Table 2).

**Table 1.** Draft genome features of *Trichoderma asperellum* MLT1J1

Genome properties	Value
Genome size (bp)	48,664,938
Total length (bp)	18,073,466
GC content (%)	52.32
Scaffolds	471
CDS	14,103
COG-categorized CDS	2,636
N50	996,985
N90	84,280
tRNA	346
rRNA	54

\*For correspondence. E-mail: [jwidada@ugm.ac.id](mailto:jwidada@ugm.ac.id);  
Tel.: +62-818-265-654; Fax: +62-274-563-062

**Table 2.** Comparative genomics of the most common *Trichoderma* species (Kubicek *et al.*, 2019)

No.	Species	Strain	Genome size (Mb)	Gene number
1	<i>Trichoderma reesei</i>	QM6a	32.7	9,877
2		RUT C30	34.2	10,877
3	<i>Trichoderma longibrachiatum</i>	ATCC18648	31.74	10,938
4	<i>Trichoderma citrinoviride</i>	TUCIM 6016	33.2	9,737
5	<i>Trichoderma parareesei</i>	CBS125925	32.07	9,292
6	<i>Trichoderma harzianum</i>	CBS 226.95	40.9	14,095
7		TR257	39.4	13,932
8	<i>Trichoderma afroharzianum</i>	T6776	39.7	11,297
9	<i>Trichoderma guizhouense</i>	NJAU4742	38.8	11,297
10	<i>Trichoderma virens</i>	Gv29-8	40.52	12,427
11	<i>Trichoderma atroviride</i>	IMI 206040	36.4	11,863
12	<i>Trichoderma gamsii</i>	T6085	37.9	10,709
13	<i>Trichoderma asperellum</i>	CBS433.97	37.66	12,586
14	<i>Trichoderma hamatum</i>	GD12	38.43	10,520

The genome prediction was performed by using GeneMarks (Besemer *et al.*, 2001). 22.4% of repetitive sequences were masked by RepeatMasker (Saha *et al.*, 2008) and TRF (Tandem Repeats Finder) (Benson, 1999). Transfer RNA (tRNA) genes were predicted by the tRNAscan-SE (Lowe and Eddy, 1997). Ribosome RNA (rRNA) genes were analyzed by the rRNAmmer (Lagesen *et al.*, 2007). Small nuclear RNAs (snRNA) were predicted by BLAST against the Rfam (Gardner *et al.*, 2009). A total of 14,103 gene numbers were used for making 51,459 valid annotations by carrying out sequence similarity searches against the GO (Ashburner *et al.*, 2000), KEGG (Kanehisa *et al.*, 2004), COG (Galperin *et al.*, 2015), Swissprot (Bairoch and Apweiler, 2000), and CAZyme databases (Cantarel *et al.*, 2008).

The CaZy analysis showed that this strain has 395 enzymes involved in CAZymes (229 glycoside hydrolases, 128 glycosyl transferase, 23 carbohydrate bindings modules, 9 carbohydrate esterase, and 5 polysaccharide lyases). This strain produced the gene of alpha-amylase (shared 81.9% identity with *Trichoderma asperellum* CBS 433.97) and glucoamylase (shared 84.8% identity with *Trichoderma atroviride* IMI 206040). 11 secondary metabolites gene clusters (with 30 biosynthetic enzymes and 17 smCOGs) were found by using antiSMASH (Medema *et al.*, 2011) consisting of NRPS-Like, T1PKS, terpene, beta lactone, ectoine, NRPS, NAPAA, lasso peptide, resorcinol, aryl polyene, phosphonate, and RiPP-like.

#### Nucleotide sequence accession number

The draft sequence of *Trichoderma asperellum* MLT1J1 (FNCC 6189) has been deposited at GenBank under the accession JAGJIL000000000 (BioProject; PRJNA699103). The version described in this paper is version JAGJIL000000000.

#### Acknowledgments

We are grateful to Purwito. Biotechnology Laboratory Staff. Faculty of Agricultural Technology. Gadjah Mada University for generously provides the fungal strain. This research was funded by a PMDSU scholarship from The Ministry of Research and Higher Education of the Republic of Indonesia with grant number 3191/UN1.DITLIT/DIT-LIT/PT/2020.

#### Conflict of Interest

The authors have no conflict of interest to report.

#### References

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler J, Cherry M, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–

- 29.
- Bairoch A and Apweiler R.** 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477.
- Benson G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.
- Besemer J, Lomsadze A, and Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, and Henrissat B.** 2008. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238.
- Galperin MY, Makarova KS, Wolf YI, and Koonin EV.** 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, 261–269.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Jones SG, Eddy SR, et al.** 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res.* **37**, D136–D140.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, and Hattori M.** 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280.
- Kubicek CP, Steindorff AS, Chenthamara K, Manganiello G, Henrissat B, Zhang J, Cai F, Kopchinskiy AG, Kubicek EM, Kuo A, et al.** 2019. Evolution and comparative genomics of the most common *Trichoderma* species. *BMC Genomics.* **20**, 485.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H, Rognes T, and Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108.
- Li WC, Huang CH, Chen CL, Chuang YC, Tung SY, and Wang TF.** 2017. *Trichoderma reesei* complete genome sequence. repeat-induced point mutation. and partitioning of CAZyme gene clusters. *Biotechnol. Biofuels* **10**, 170.
- Lim HJ, Lee EH, Yoon Y, Chua B, and Son A.** 2016. Portable lysis apparatus for rapid single-step DNA extraction of *Bacillus subtilis*. *J. Appl. Microbiol.* **120**, 379–387.
- Lin SH and Liao YC.** 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS ONE* **8**, e60843.
- Lowe TM and Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al.** 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **4**, 18.
- Medema HM, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, and Breiting R.** 2011. antiSMASH: rapid identification. annotation. and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339–W346.
- Safitri IO, Suroto DA, Widada J, Sardjono S, and Cahyanto MN.** 2022. Effects of glucose as carbon catabolite repressor on alpha-amylase and glucoamylase production in Indonesian indigenous fungi. *Malays. J. Microbiol.* **18**, 322–330.
- Saha S, Bridges S, Magbanua ZV, and Peterson DG.** 2008. Empirical comparison of *ab initio* repeat finding programs. *Nucleic Acids Res.* **36**, 2284–2294.