



Data Article

Whole-genome sequence data of cellulase-producing fungi *Trichoderma asperellum* PK1J2, isolated from palm empty fruit bunch in Riau, Indonesia

Fela Laila Nur Hidayati^a, Dian Anggraini Suroto^a, Sardjono^a, Muhammad Nur Cahyanto^a, Jaka Widada^{b,*}

^aDepartment of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Jl. Flora, Bulaksumur, Yogyakarta 55281, Indonesia

^bDepartment of Agricultural Microbiology, Faculty of Agriculture, Universitas Gadjah Mada, Jl. Flora, Bulaksumur, Yogyakarta 55281, Indonesia

ARTICLE INFO

Article history:

Received 6 July 2022

Revised 2 September 2022

Accepted 9 September 2022

Available online 15 September 2022

Dataset link: [richoderma asperellum strain:PK1J2 \(Original data\)](#)

Keywords:

Whole-genome sequence

Trichoderma asperellum

Palm empty fruit bunch

Genomic

ABSTRACT

Trichoderma asperellum PK1J2 is a promising cellulase-producing fungus isolated from a palm empty fruit bunch in Riau, Indonesia. Presented here is the genome assembly of *T. asperellum* PK1J2. The whole genome of the fungi was sequenced using Illumina NovaSeq PE150. The genome assembly was performed using SOAPdenovo, SPAdes, and Abyss software, and the assembly results of the three types of software were integrated with CISA software. *T. asperellum* PK1J2 has 6,835 protein-coding genes with a length of 9,233,597 bp. The final genome assembly was approximately 36 Mbp with a GC content of 48.45%. This whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession JAGJIK000000000.

© 2022 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: jwidada@ugm.ac.id (J. Widada).

Specifications Table

Subject	Biological science
Specific subject area	Genomics, Microbiology
Type of data	Genome sequencing in FASTA format Table Figure
How the data were acquired	Genomic DNA was isolated from <i>Trichoderma asperellum</i> PK1J2. The sequencing libraries were generated using the NEBNext® ULtra™ DNA Prep Kit for Illumina (NEB, USA). Illumina NovaSeq PE150 was used for whole genome sequencing. The genome assembly was performed with SOAPdenovo, SPAdes and Abyss software and integrated with CISA software.
Data format	Raw Analyzed
Description of data collection	Genomic DNA was isolated from <i>Trichoderma asperellum</i> PK1J2. The sequencing libraries were generated using the NEBNext® ULtra™ DNA Prep Kit for Illumina (NEB, USA). Illumina NovaSeq PE150 was used for whole genome sequencing. The genome assembly was performed with SOAPdenovo, SPAdes and Abyss software and integrated with CISA software.
Data source location	Institution: Faculty of Agricultural Technology, Universitas Gadjah Mada City/Town/Region: Sleman, Yogyakarta Country: Indonesia Latitude and longitude for samples/data collection: 7° 46' 14.5" S, 110° 22' 39.8" E
Data accessibility	Repository name: Data identification number: This whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession JAGJIK000000000. The version described in this paper is version JAGJIK000000000. Direct URL to data: www.ncbi.nlm.nih.gov/assembly/GCA_022817925.1/ The raw sequence data of this paper are accessible under SRA accession number SRR19762116. Direct URL to data: www.ncbi.nlm.nih.gov/sra/?term=pk1j2 All data in this paper are available at NCBI with BioProject number PRJNA699105. Direct URL to data: www.ncbi.nlm.nih.gov/bioproject/PRJNA699105

Value of the Data

- The genome data of *Trichoderma asperellum* PK1J2 isolated from Indonesia provide insight into the genetic diversity of *T. asperellum* and essential genetic information to reveal important details of effector proteins, metabolites and enzymes production.
- The data can be useful for researchers working on fungal microbiology, biotechnology, genomics, and genetic engineering.
- This genome information can be used for genome mining to discover the genes involved in metabolites and enzymes biosynthesis pathways.
- Stakeholders, including industry, can use *T. asperellum* PK1J2 as a biocontrol agent, biofertiliser, and producer of metabolites and enzymes, especially cellulase, through this genetic information.

1. Data Description

T. asperellum is a mycoparasitic species widely used for its ability to inhibit the growth of plant pathogens [1]. *T. asperellum* has been shown to produce hydrolytic enzymes such as cellulase and xylanase [2,3]. *T. asperellum* has also been reported to hydrolyse wheat bran, wheat straw, paper, sawdust, corncob, duckweed, and agave by secreting cellulases [2–7]. Strain PK1J2 has been proven to be capable of producing high cellulase. The cellulase from this fungi can hydrolyse cassava stem and sago waste into fermentable sugar [8,9]. In a previous study, strain PK1J2 produced highest cellulase activity among the examined fungi isolated from Indonesia and was further selected to characterize its genome.

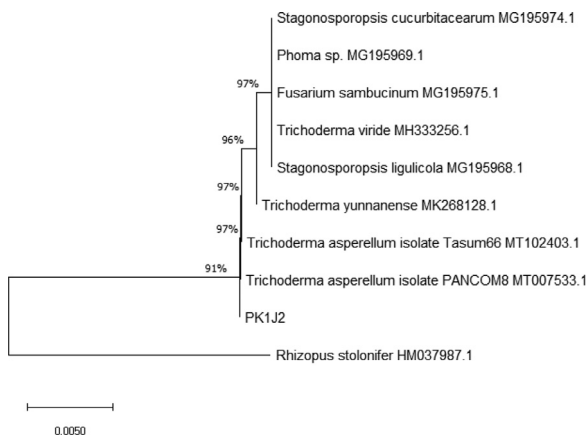


Fig. 1. Phylogenetic tree of strain PK1J2 ITS region compared with the other fungi.

Table 1

Genome features of *T. asperellum* PK1J2.

Genome feature	Value
Genome size	36,156,613
Gene number	6835
Gene length	9,233,597
tRNA genes	236
rRNA genes (5 s)	49
sRNA genes	2
snRNA genes	22

Fig. 1 shows a phylogenetic tree of strain PK1J2 comparing its internal transcribed spacer (ITS) region with the other fungi. As can be seen in the figure, the ITS gene of strain PK1J2 showed the highest similarity with *Trichoderma asperellum* species.

T. asperellum PK1J2 had 6835 protein-coding genes with 9,233,597 bp in length, as seen in **Table 1**. The final assembly for the *T. asperellum* PK1J2 genome was approximately 36 Mbp with a GC content of 48.45%. The genome consisted of 249 scaffolds with a total length of 36,156,613 bp (N_{50} , 724,251 bp; N_{90} , 188,633 bp; and N_{max} , 2,963,926 bp) and 636 contigs with a total length of 36,152,186 bp (N_{50} , 143,266 bp; N_{90} , 33,810 bp; and N_{max} , 506,821 bp). This whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession JAGJIK000000000. The version described in this paper is version JAGJIK000000000 [10].

Functional gene annotation predicted about 4759 genes using GO, 6398 genes using KEGG, 1946 genes using KOG, 4759 genes using Pfam, 2783 genes using SWISS-PROT, and 6544 genes using NR database. Gene coding for protein possibly involved in secondary metabolite production revealed the presence of T1PKS cluster, NRPS cluster, NRPS-like cluster, T1PKS-NRPS hybrid cluster, and terpene cluster. A carbohydrate-active enzyme analysis showed that *T. asperellum* PK1J2 was dominated by GH18, GH3, GH16, GH2, and GH5.

2. Experimental Design, Materials and Methods

2.1. Fungal Strain and DNA Extraction

Strain *Trichoderma asperellum* PK1J2 was obtained from the Laboratory of Biotechnology, Faculty of Agricultural Technology, Universitas Gadjah Mada. Strain PK1J2 was isolated from a rotten

palm empty fruit bunch, Pekanbaru, Riau, Indonesia. The strain was grown on PDA agar at 30 °C for a period of seven days. ZymoBIOIMICS™ DNA Mini Kit (Zymo Research, California) was used for extracting genomic DNA. The harvested DNA was detected by agarose gel electrophoresis and quantified by Qubit® 2.0 Fluorometer.

2.2. Species Identification

The DNA fragment was amplified using universal primer set ITS1 (forward primer) 5'-TCCGTAGGTTAACCTGCGG-3' and ITS4 (reverse primer) 5'-TCCTCCGCTTATTGATATGC-3'. The PCR product was sequenced using Bi-directional Sequencing. The sequence was analyzed by BLAST and then compared to the NCBI database. The phylogenetic tree was constructed using the Neighbor-Joining method (Unrooted Tree) by NCBI BLAST.

2.3. Genome Sequencing and Assembly

Sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's recommendations. The whole genome sequencing of the fungi was performed using an Illumina NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. The genome assembly was done using SOAPdenovo, SPAdes, and Abyss software. The assembly results from all three software were integrated with CISA software. The assembly result with the least scaffolds was selected.

2.4. Genome Component Prediction

Transfer RNA (tRNA) genes were predicted by tRNAscan-SE [11]. Also, ribosome RNA (rRNA) genes were analyzed by rRNAmmer [12], and small nuclear RNAs (snRNA) were predicted by BLAST against the Rfam database [13].

2.5. Genome Annotation

Genome functional annotation was based on the BLASTP with GO (Gene Ontology) [14], KEGG (Kyoto Encyclopedia of Genes and Genomes) [15], COG (Clusters of Orthologous Groups) [16], NR (Non-Redundant Protein Database) [17], and SWISS-PROT [18]. Carbohydrate-active enzymes were predicted by the Carbohydrate-Active Enzymes Database (CAZy) [19]. Genes coding for proteins that were possibly involved in secondary metabolite production were predicted by anti-SMASH v.5.0 [20].

Ethics Statements

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

[richoderma asperellum strain:PK1J2 \(Original data\)](#) (National Center for Biotechnology Information (NCBI)).

CRedit Author Statement

Fela Laila Nur Hidayati: Investigation, Formal analysis, Writing – original draft; **Dian Anggraini Suroto:** Methodology, Software, Data curation; **Sardjono:** Resources; **Muhammad Nur Cahyanto:** Supervision, Project administration, Funding acquisition, Writing – review & editing; **Jaka Widada:** Conceptualization, Methodology, Validation, Writing – review & editing.

Acknowledgments

This work was supported by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia [grant number [3192/UN1.DITLIT/DIT-LIT/PT/2020](#)].

References

- [1] W.Z. Go, P.S. H'ng, M.Y. Wong, K.L. Chin, S. Ujang, Evaluation of *Trichoderma asperellum* as a potential biocontrol agent against rigidoporus microporus hevea brasiliensis, Arch. Phytopathol. Plant Prot. 52 (2019) 639–666, doi:[10.1080/03235408.2019.1587821](#).
- [2] Q. Wang, H. Lin, Q. Shen, X. Fan, N. Bai, Y. Zhao, Characterization of cellulase secretion and Cre1-mediated carbon source repression in the potential lignocellulose-degrading strain *Trichoderma asperellum* T-1, PLoS One 10 (2015) 1–15, doi:[10.1371/journal.pone.0119237](#).
- [3] L. Bech, P.K. Busk, L. Lange, Cell wall degrading enzymes in trichoderma asperellum grown on wheat bran, Fungal Genom. Biol 04 (2014) 1–11, doi:[10.4172/2165-8056.1000116](#).
- [4] S. Pandey, M. Srivastava, M. Shahid, V. Kumar, A. Singh, S. Trivedi, Y.K. Srivastava, Trichoderma species cellulases produced by solid state fermentation, J. Data Min. Genom. Proteom. 6 (2015) 1–4, doi:[10.4172/2153-0602.1000170](#).
- [5] L. Bech, F.A. Herbst, M.N. Grell, L. Lange, On-site enzyme production by *Trichoderma asperellum* for the degradation of duckweed, Fungal Genom. Biol 05 (2015) 1–10, doi:[10.4172/2165-8056.1000126](#).
- [6] N.Y. Nava-Cruz, J.C. Contreras-Esquivel, M.A. Aguilar-González, A. Nuncio, R. Rodríguez-Herrera, C.N. Aguilar, Agave atrovirens fibers as substrate and support for solid-state fermentation for cellulase production by *Trichoderma asperellum*, 3, Biotech 6 (2016) 1–12, doi:[10.1007/s13205-016-0426-6](#).
- [7] Y. Zapata, A. Galviz-quezada, J.C. Salcedo-reyes, Cellulases production on paper and sawdust using native Trichoderma asperellum, Univ. Sci. 23 (2018) 419–436, doi:[10.11144/Javeriana.SC23-3.cpop](#).
- [8] P. Dewi, R. Indrati, R. Millati, S. Sardjono, Effect of lime pretreatment on microstructure of cassava stalk fibers and growth of aspergillus niger, Biosaintifika J. Biol. Biol. Educ. 10 (2018) 205–212, doi:[10.15294/biosaintifika.v10i1.13802](#).
- [9] M. Iqbal, K. Rianse, R. Millati, R. Indrati, Enhanced fermentable sugars production from sago waste by *Trichoderma reesei* Pk1J2 and aspergillus niger FNCC 6114 fermentation, Int. J. Sci. Res. 9 (2020) 1200–1204, doi:[10.21275/SR2061711902](#).
- [10] F.L.N. Hidayati, D.A. Suroto, S. Sardjono, M. Nur Cahyanto, J. Widada, Trichoderma Asperellum Strain PK1J2, v 1, National Center for Biotechnology Information, 2022 www.ncbi.nlm.nih.gov/bioproject/PRJNA699105.
- [11] T.M. Lowe, S.R. Eddy, TRNAScan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964, doi:[10.1093/nar/25.5.0955](#).
- [12] K. Lagesen, P. Hallin, E.A. Rødland, H.H. Stærfeldt, T. Rognes, D.W. Ussery, Rfam: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108, doi:[10.1093/nar/gkm160](#).
- [13] P.P. Gardner, J. Daub, J.G. Tate, E.P. Nawrocki, D.L. Kolbe, S. Lindgreen, A.C. Wilkinson, R.D. Finn, S. Griffiths-Jones, S.R. Eddy, A. Bateman, Rfam: updates to the RNA families database, Nucleic Acids Res. 37 (2009) 136–140, doi:[10.1093/nar/gkn766](#).
- [14] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology : tool for the unification of biology, Nat. Genet. 25 (2000) 25–29, doi:[10.1038/75556](#).
- [15] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori, The KEGG resource for deciphering the genome, Nucleic Acids Res. 32 (2004) 277–280, doi:[10.1093/nar/gkh063](#).
- [16] M.Y. Galperin, K.S. Makarova, Y.I. Wolf, E.V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database, Nucleic Acids Res. 43 (2015) 261–269, doi:[10.1093/nar/gku1223](#).
- [17] W. Li, L. Jaroszewski, A. Godzik, Tolerating some redundancy significantly speeds up clustering of large protein databases, Bioinformatics 18 (2002) 77–82, doi:[10.1093/bioinformatics/18.1.77](#).
- [18] A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, Nucleic Acids Res. 28 (2000) 45–48, doi:[10.1093/nar/28.1.45](#).

- [19] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics, *Nucleic Acids Res.* 37 (2009) 233–238, doi:[10.1093/nar/gkn663](https://doi.org/10.1093/nar/gkn663).
- [20] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, Y. Lee, M.H. Medema, T. Weber, Antismash 5. 0 : updates to the secondary metabolite genome mining pipeline, *Nucleic* 47 (2019) 81–87, doi:[10.1093/nar/gkz310](https://doi.org/10.1093/nar/gkz310).