The Highly Contiguous Genome Resource of *Trichoderma semiorbis* FJ059, a Biological Control Agent for Litchi Downy Blight

Zhigang Li and Tong Liu[†]

College of Plant Protection/Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests, Ministry of Education, Hainan University, Haikou, 570228, China

Funding

This work was funded by the National Natural Science Foundation of China (grant numbers 31760510 and 32001844).

Genome Announcement

Trichoderma spp. are well known for their multifarious actions to suppress plant diseases and improve plant growth. However, the biological mechanism and evolution of these traits are still unclear. The genus *Trichoderma* is one of the most commonly used microbial biopesticides, widely used in agriculture and industry (Mukhopadhyay and Kumar 2020; Samuels et al. 1994). A number of registered biopesticides are *Trichoderma*-based, which may not only protect plants by killing other fungi and nematodes but also induce resistance against plant pathogens (Hermosa et al. 2012; Shores et al. 2010; Verma et al. 2007). Some genome sequences of *Trichoderma* spp. have been released (Mukherjee et al. 2013), with estimated genome sizes and chromosome numbers ranging from 31 to 39 Mb and 3 to 7, respectively (El-Bondkly 2014). However, the distinction of species in the genus remained problematic and needed to be analyzed at molecular levels. High-quality genome sequencing for *Trichoderma* spp. will help us understand the evolution of the genus and the effective utilization of its biocontrol potential against plant diseases. In this study, we identified *Trichoderma semiorbis* strain FJ059 as a biological control agent for litchi downy blight and sequenced the first genome of *T. semiorbis* FJ059.

T. semiorbis FJ059 was isolated from the pumpkin rhizosphere (latitude 26.193563°N; longitude 119.360027°E; Fujian Province, China) as described by Chen and Zhuang (2017) and identified by morphological characteristics and sequence analysis of translation elongation factor 1 and RNA polymerase 2 (Jaklitsch and Voglmayr 2015). The inhibition assay of the strain to plant pathogens was described by Marques et al. (2018) and first identified as a potential biological control agent for litchi downy blight with the highest inhibition rate (100%) against *Peronophythora litchi* in this study (Fig. 1A). Here, we sequenced the first *T. semiorbis* FJ059 genome. We extracted total genomic DNA of *T. semiorbis* FJ059 using the cetyltrimethylammonium bromide method (Weiland 1997). To ensure that the strain was free from contamination, the *Trichoderma* strain was purified by single-spore isolation as described by Choi et al. (1999). The purified *Trichoderma* strain was inoculated into potato dextrose medium in a shaker at 28°C for 6 days, and hyphae were collected and water removed by a freeze-drying process. The quality and concentration of DNA were evaluated by a NanoDrop 2000C Spectrophotometer (Thermo Fisher Scientific) and Qubit 2.0 fluorometer (Life Technologies).

For the short-reads sequencing, the genomic DNA of this strain was used to generate a 250- to 300-bp library with the TrueLib DNA Library Rapid Prep Kit (Nanjing Novizan Biotechnology Co., Ltd., Nanjing, China), which was sequenced in the Illumina Xten platform (2 × 150 bp). Long reads were obtained by preparing libraries using the rapid barcoding kit (catalog number

Keywords

Trichoderma semiorbis, genomics, fungal antagonistic, biological control, litchi downy blight, biological control, bioinformatics

[†]Corresponding author: T. Liu; liutongamy@sina.com

The author(s) declare no conflict of interest.

Accepted for publication 13 December 2021.

Fig. 1. Characteristics of *Trichoderma semiorbis*. **A**, Inhibition rate of fermentation broth against plant pathogens. The Y-axis indicates inhibition rate of fermentation broth, which was measured by the radial growth of plant pathogen. Inhibition rate = $\{(\text{control colony diameter [mm]} - \text{treated colony diameter [mm]}) / \text{control colony diameter [mm]}\} \times 100\%$. Error bars refer to standard error. Abbreviations: Cg, *Colletotrichum gloeosporioides*; Fg, *Fusarium graminearum*; Fo, *F. oxysporum*; Pl, *Peronophythora litchi*; Bc, *Botrytis cinerea*; and Db, *Didymella bryoniae*. **B**, Phylogenetic relationship of *T. semiorbis* and 12 publicly available *Trichoderma* genomes. Genomes were obtained from the GenBank database; that is, *T. arundinaceum* (GCA_003012105.1), *T. asperellum* (GCA_003025105.1), *T. atroviride* (GCA_000171015.2), *T. citrinoviride* (GCA_003025115.1), *T. gamsii* (GCA_001481775.2), *T. guizhouense* (GCA_002022785.1), *T. harzianum* (GCA_000988865.1), *T. lentiforme* (GCA_011066345.1), *T. longibrachiatum* (GCA_003025155.1), *T. parareesei* (GCA_001050175.1), *T. reesei* (GCA_000513815.1), and *T. virens* (GCA_000170995.2). The species tree was inferred from all orthogroups by STAG and rooted by STRIDE. Scale bar indicates the number of substitutions per site. **C**, Gene annotation of carbohydrate-active enzymes. **D**, An illustration of nonribosomal peptide synthase-polyketide synthase hybrid gene clusters at contig 6: 4,022,870 to 4,110,963.

SQK-RBK004) and sequencing these libraries using an R9.4.1 (FLO-MIN106) in an Oxford Nanopore PromethION sequencer (Oxford Nanopore Technologies [ONT]). Both Nanopore Technologies (ONT) long-read sequencing and Illumina short-read sequencing were performed at Biomarker Company, Beijing, China. We obtained a total of 634,450 ONT reads

Table 1. Assembly statistics of the *Trichoderma semiorbis* FJ059 genome

Features ^a	FJ059
ONT long reads (Gb)	4.43
ONT reads N ₅₀ (kb)	10.59
Illumina short reads (Gb)	6.92
RNA-sequencing reads (Gb)	7.66
Assembly size (bp)	42,023,549
Contig number	7
Contig 1 size (bp)	8,305,737
Contig 2 size (bp)	7,794,183
Contig 3 size (bp)	6,906,462
Contig 4 size (bp)	4,956,843
Contig 5 size (bp)	4,939,019
Contig 6 size (bp)	4,720,324
Contig 7 size (bp)	4,400,981
GC content (%)	46.62
BUSCO assessment in fungi (<i>n</i> = 759)	99.1
Number of predicted genes	10,251

^a ONT = Oxford Nanopore Technologies and BUSCO = benchmarking universal single-copy orthologs.

(total bases = 4.43 Gb, N₅₀ = 10.59 kb, mean read length = 6.98kb) and 6.92 Gb of paired-end Illumina reads (46,173,374 reads with an insert length of 350 bp and 150-bp read length) (Table 1).

RNA-sequencing (RNA-Seq) data were used to improve the gene annotation. Total RNAs were extracted and sequenced. Conidia of this strain were collected from the sterile colony for RNA extraction after cultivation for approximately 2 weeks. The samples were immediately frozen in liquid nitrogen. RNA was then extracted using the Quick RNA isolation Kit (TaKaRa Biotech Co., Ltd., Dalian, China). The extracted RNA was treated with RNase-free DNaseI (TaKaRa Biotech Co., Ltd.) to remove residual DNA. RNA integrity was checked on a 1.2% agarose gel by electrophoresis, and RNA concentration was estimated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, U.S.A.). The sequencing library was sequenced by a 150-bp paired-end Illumina HiSeq 4000 platform. In total, 7.66 Gb of RNA-Seq data was obtained. The clean data were obtained after removing the adaptor sequences, duplicated sequences, ambiguous reads ('N'), and low-quality reads. In all, 2,000 reads were randomly selected from the clean data and compared with the NT database with Blast software to ensure no contamination in this sequence sample.

To assemble the genome, short ONT reads (<1 kb) were filtered by custom program script, and Nextdenovo v2.4.0 (<https://github.com/Nextomics/NextDenovo>) was employed to generate a raw assembly using ONT reads only. RACON v1.4.20 (Vaser et al. 2017) and pilon v1.24 (Walker et al. 2014) were employed to polish and correct raw assembled contigs using ONT reads and Illumina clean reads, respectively (removing reads with adaptor or low-quality reads by Trimmomatic) (Bolger et al. 2014). Finally, we obtained a highly contiguous assembly of 42.02 Mb, containing seven contigs. GC content of FJ059 determined by seqkit (Shen et al. 2016) was 46.62%. The completeness of the genome assembly was evaluated through the benchmarking universal single-copy orthologs (BUSCO) tool v5.0.0 (Seppey et al. 2019) with "fungi_odb10" library as the reference dataset. This analysis reported a completeness score of 99.1% (*n* = 758). BUSCO dataset "sordariomycetes_odb10" was also used in the evaluation and 98.1% complement of a total of 3,817 BUSCOs was obtained. All of these indicated the reliability of the genome assembly.

Protein-coding genes were annotated through a combined pipeline. First, BRAKER2 (with `--softmasking --gff3 --fungus --gth2traingenes --prg=gth`) (Brůna et al. 2021) with a combination of protein-based and RNA-based evidence was performed. All proteins of the genus *Trichoderma* in the Uniref90 database (Suzek et al. 2015) were used as protein-based training evidence in BRAKER2, with AUGUSTUS v3.4.0 (Keller et al. 2011) and Genemark-EP+ (Brůna et al. 2020) as ab initio prediction tools. Additionally, RNA-Seq data aligned onto the genome by Hisat2 (with `-t -dta`) (Kim et al. 2019) were used as RNA-based training evidence in BRAKER2. Then, RNA-Seq data were assembled by StringTie2 v2.1.4 (with default parameters) (Kovaka et al. 2019), and the genus *Trichoderma* proteins were aligned to the assembled genome by GenomeThreader v1.7.3 (Gremme et al. 2005). Finally, all of the BRAKER2

predictions, transcript assembly, and protein alignments were combined by EvidenceModeler v1.1.1 and Program to Assemble Spliced Alignments v2.4.1 (Haas et al. 2008). In total, 10,251 protein coding genes were predicted. The comparative analysis between *T. semiorbis* and 12 publicly available *Trichoderma* genomes (Fig. 1B) was performed by OrthFinder v2.5.4 (Emms and Kelly 2019). A species tree was inferred from orthogroups by STAG (Emms and Kelly 2018) and rooted by STRIDE (Emms and Kelly 2017) (Fig. 1B).

Seven databases were used for the functional annotation of protein coding genes: that is, Gene Ontology (Ashburner et al. 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), eukaryotic orthologous groups (KOG) (Koonin et al. 2004), GenBank nonredundant (NR) database, Pfam, carbohydrate-active enzymes (CAZymes) (<http://www.cazy.org>), and AntiSMASH (Blin et al. 2021). In total, 10,098 genes were mapped to the NR database by Blastp (e-value = 1e-6). Among these genes, 6,714 were classified by the KOG database, and 4,238 and 7,829 genes were assigned KEGG and Pfam terms, respectively. We also identified 1,016 genes annotated with secretion signal peptides by SignalP v5.0 (with -format short -org euk) (Almagro Armenteros et al. 2019). The genomic repository of CAZymes predicted the pattern of carbohydrate metabolism. In total, 422 CAZymes were identified, including 212 glycoside hydrolases, 61 auxiliary activities, and 83 glycosyl transferases (Fig. 1C). AntiSMASH v6.0.0 (with -taxon fungi -cb-general -cb-knownclusters -cb-subclusters -asf -pfam2go -smcog-trees) predicted gene clusters involved in the formation of secondary metabolites. We identified 17 polyketide synthases (PKS), 9 nonribosomal peptide synthases (NRPS), 6 terpenes, and 5 NRPS-PKS hybrid gene clusters (Fig. 1D). The highly contiguous genome resource given in this study is the first reported genome of *T. semiorbis* FJ059. It will serve as a fundamental resource for further research of this strain's biocontrol mechanism.

Data Availability

The genome sequence of *T. semiorbis* FJ059 is available in NCBI under accession number JAIMJC000000000, BioProject PRJNA756961, and BioSample SAMN20930937.

Literature Cited

- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37:420-423.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Wezel, G. P., Medema, M. H., and Weber, T. 2021. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49:W29-W35.
- Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
- Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. 2021. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinf.* 3:lqaa108.
- Brůna, T., Lomsadze, A., and Borodovsky, M. 2020. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinf.* 2:lqaa026.
- Chen, K., and Zhuang, W.-Y. 2017. Discovery from a large-scaled survey of *Trichoderma* in soil of China. *Sci. Rep.* 7:9090.
- Choi, Y. W., Hyde, K. D., and Wellcom, W. H. H. 1999. Single spore isolation of fungi. *Fungal Divers.* 3:29-38.
- El-Bondkly, A. M. A. 2014. Chapter 28 – Sequence analysis of industrially important genes from *Trichoderma*. Pages 377-392 in: *Biotechnology and Biology of Trichoderma*. V. K. Gupta, M. Schmoll, A. Herrera-Estrella, R. S. Upadhyay, I. Druzhinina, and M. G. Tuohy, eds. Elsevier, Amsterdam, The Netherlands.
- Emms, D. M., and Kelly, S. 2017. STRIDE: Species tree root inference from gene duplication events. *Mol. Biol. Evol.* 34:3267-3278.
- Emms, D. M., and Kelly, S. 2018. STAG: Species tree inference from all genes. *bioRxiv* 267914.
- Emms, D. M., and Kelly, S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Gremme, G., Brendel, V., Sparks, M. E., and Kurtz, S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* 47:965-978.
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., and Wortman, J. R. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9:R7.
- Hermosa, R., Viterbo, A., Chet, I., and Monte, E. 2012. Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology (Reading)* 158:17-25.
- Jaklitsch, W. M., and Voglmayr, H. 2015. Biodiversity of *Trichoderma* (Hypocreaceae) in Southern Europe and Macaronesia. *Stud. Mycol.* 80:1-87.
- Kanehisa, M., and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27-30.
- Keller, O., Kollmar, M., Stanke, M., and Waack, S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27:757-763.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37:907-915.
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Rogozin, I. B., Smirnov, S., Sorokin, A. V., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:R7.
- Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. 2019. Transcriptome assembly from long-read RNA-seq alignments with String-Tie2. *Genome Biol.* 20:278.
- Marques, E., Martins, I., and Mello, S. C. M. 2018. Antifungal potential of crude extracts of *Trichoderma* spp. *Biota Neotrop.* 18:20170418.
- Mukherjee, P. K., Horwitz, B. A., Herrera-Estrella, A., Schmoll, M., and Kenerley, C. M. 2013. *Trichoderma* research in the genome era. *Annu. Rev. Phytopathol.* 51:105-129.
- Mukhopadhyay, R., and Kumar, D. 2020. *Trichoderma*: A beneficial antifungal agent and insights into its mechanism of biocontrol potential. *Egypt. J. Biol. Pest Control* 30:133.

- Samuels, G. J., Petrini, O., and Manguin, S. 1994. Morphological and macromolecular characterization of *Hypocrea schweinitzii* and its *Trichoderma* anamorph. *Mycologia* 86:421-435.
- Seppely, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962:227-245.
- Shen, W., Le, S., Li, Y., and Hu, F. 2016. SeqKit: A cross-platform and ultrafast tool-kit for FASTA/Q file manipulation. *PLoS One* 11:e0163962.
- Shores, M., Harman, G. E., and Mastouri, F. 2010. Induced systemic resistance and plant responses to fungal biocontrol agents. *Annu. Rev. Phytopathol.* 48:21-43.
- Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., and Wu, C. H. 2015. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31:926-932.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27:737-746.
- Verma, M., Brar, S. K., Tyagi, R. D., Surampalli, R. Y., and Valéro, J. R. 2007. Antagonistic fungi, *Trichoderma* spp.: Panoply of biological control. *Biochem. Eng. J.* 37:1-20.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. K. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Weiland, J. J. 1997. Rapid procedure for the extraction of DNA from fungal spores and mycelia. *Fungal Genet. Rep.* 44:60-63.