AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Host-Microbial Interactions | Announcement

Genome annotations for the ascomycete fungi *Trichoderma harzianum*, *Trichoderma aggressivum*, and *Purpureocillium lilacinum*

Erik P. W. Beijen,[1] **Robin A. Ohm**[1]

AUTHOR AFFILIATION See affiliation list on p. 3.

ABSTRACT We sequenced and annotated the genomes of the ascomycete fungi *Trichoderma harzianum*, *Trichoderma aggressivum f. europaeum*, and *Purpureocillium lilacinum*. Moreover, we developed a website to allow users to interactively analyze the assemblies, gene predictions, and functional annotations of these species and 70+ previously sequenced fungi.

KEYWORDS fungal genomics, ascomycetes, mycoparasites

Trichoderma harzianum and *Trichoderma aggressivum* are mycoparasites that can cause green mold disease on mushroom-forming fungi, including the white button mushroom *Agaricus bisporus* (1). Phylogenetically, they are members of the Harzianum/Virens clade of the *Trichoderma* species complex (2). The species *T. aggressivum* encompasses the biotypes *aggressivum* and *europaeum*, which originate from North America and Europe, respectively (3). The ascomycete *P. lilacinum* is a ubiquitous saprotroph, found in soil, dead organic matter, insects, and air (4). We serendipitously discovered that it can inhibit growth of the mushroom-forming fungus *Schizophyllum commune*.

T. harzianum CBS 354.33, originally isolated from soil, and *T. aggressivum* CBS 100526, originally isolated from mushroom compost, were obtained from the Westerdijk Institute. *P. lilacinum* was isolated from the air and deposited in the Westerdijk collection with accession number CBS 150709. A spore suspension was incubated for 3 days at 30°C in liquid culture in *Schizophyllum commune* minimal medium (5). Genomic DNA was extracted and purified from lysed cells using the MagMAX Plant DNA Isolation Kit (Thermo Scientific, USA) and sonicated using a Bioruptor (Diagenode, Belgium) for 10 minutes, with a 30-/90-second on/off cycle on low power and the temperature in the water bath at 4°C, resulting in fragments of approximately 700 bp. The NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) was used to generate Illumina libraries. These libraries were sequenced by the Utrecht Sequencing Facility (Utrecht; The Netherlands; https://useq.nl) using an Illumina NextSeq2000 sequencer with a 2 × 100 bp protocol. The reads were trimmed on both ends to a quality score of 15 using BBMap version 37.88 (https://sourceforge.net/projects/bbmap/) and assembled with SPAdes version 3.15.5 (6) with kmer lengths 21, 33, 55, 77, and 99 and the parameter "--careful." Scaffolds shorter than 1,000 bp were removed from the assembly. Genes were predicted with AUGUSTUS version 3.3.2 (7) using the provided parameter set of the related ascomycete *Aspergillus nidulans* and the settings "--genemodel=complete" and "--noInframeStop=true." The predicted proteins were functionally annotated as previously described (8). Repetitive sequences were identified with RepeatModeler version 2.0.3 (9).

Editor Jennifer Geddes-McAlister, University of Guelph, Canada

Address correspondence to Robin A. Ohm, r.a.ohm@uu.nl.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 22 November 2023
Accepted 11 February 2024
Published 22 February 2024

TABLE 1 Statistics of the three genome annotations

	Trichoderma harzianum	*Trichoderma aggressivum f. europaeum*	*Purpureocillium lilacinum*
Strain	CBS 354.33	CBS 100526	CBS 150709
Sequencing read pairs	23,531,919	18,241,525	19,347,672
Version of assembly and annotations	Trihar35433_1	Triagg1	Purlil1
Assembly			
Sequences in assembly	171	318	928
Coverage	122×	93×	100×
Total assembly length (Mbp)	38.51	38.95	40.62
Assembly GC[a] content (%)	49.3	49.06	58.33
Assembly gaps (%)	0	0.01	0.02
L50 number (#)	6	20	25
N50 length (bp)	2,296,050	579,860	470,782
Repetitive content (%)	4.23	1.9	3.99
Predicted genes			
Genes	11,326	11,015	14,434
Gene length (median)	1,422	1,431	1,489
Transcript length (median)	1,281	1,281	1,296
Exon length (median)	297	298	288
CDS length (median)	1,278	1,278	1,293
Protein length (median)	426	426	431
Spliced genes (total, %)	8,366 (73.87%)	8,160 (74.08%)	10,853 (75.19%)
Exons per gene (median)	2	2	3
Intron length (median)	65	66	65
Introns per spliced gene (median)	2	2	2
Gene density (genes/Mbp)	294.09	282.79	355.36
Coding content of assembly (bp, %)	17,783,979 (46.18%)	17,306,514 (44.43%)	23,597,817 (58.1%)
Functional annotations			
Unique PFAM domains	4,125	4,118	4,129
Genes with PFAM domain (total, %)	8,661 (76.47%)	8,226 (74.68%)	8,386 (58.1%)
Genes with GO annotation (total, %)	5,414 (47.8%)	5,142 (46.68%)	5,360 (37.13%)
Genes with secretion signal (total, %)	1,053 (9.3%)	970 (8.81%)	1,259 (8.72%)
Genes annotated as small secreted protein (total, %)	911 (8.04%)	991 (9.00%)	1,146 (7.94%)
Genes with transmembrane domain (total, %)	2,141 (18.9%)	2,063 (18.73%)	2,543 (17.62%)
Genes annotated as transcription factor (total, %)	487 (4.3%)	437 (3.97%)	497 (3.44%)
Genes annotated as protease (total, %)	477 (4.21%)	438 (3.98%)	474 (3.28%)
Genes annotated as CAZyme (total, %)	456 (4.03%)	430 (3.9%)	454 (3.15%)
Secondary metabolism gene cluster	54 (0.48%)	58 (0.53%)	32 (0.22%)
BUSCO2 completeness (%)	99.66% (single copy: 99.31%, duplicated: 0.34%)	99.66% (single copy: 99.66%, duplicated: 0.0%)	98.62% (single copy: 97.93%, duplicated: 0.69%)

[a]guanine-cytosine.

A website was created with a portal for each genome and for all 70+ genomes that were previously sequenced and published by us (https://fungalgenomics.science.uu.nl/). The genome portals allow users to interactively analyze the assembly, gene predictions, and functional annotations. This includes a JBrowse genome browser (10) with Apollo plugin to allow the manual curation of gene models (11), a Blast function to search for sequence similarity in the assembly and gene predictions (12), and full access to the functional annotations of the proteins.

The sizes of the genome assembly were 38.51, 38.95, and 40.62 Mbp for *T. harzianum*, *T. aggressivum*, and *P. lilacinum*, respectively, and BUSCO completeness scores were 99.66% for both *Trichoderma* spp. and 98.62% for *P. lilacinum*, indicating a high quality of assembly and gene prediction for all three species (Table 1). These genome annotations

and the genome portals will be a valuable resource to study these three competitors of mushroom-forming fungi.

ACKNOWLEDGMENTS

This publication is part of project number VI.Vidi.192.065 of the VIDI research program which is (partly) financed by the Dutch Research Council (NWO). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement number 716132). We acknowledge the Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidized by the University Medical Center Utrecht and The Netherlands X-omics Initiative (NWO project 184.034.019).

AUTHOR AFFILIATION

[1]Department of Biology, Microbiology, Faculty of Science, Utrecht University, Utrecht, the Netherlands

AUTHOR ORCIDs

Erik P. W. Beijen http://orcid.org/0000-0003-2548-0074
Robin A. Ohm http://orcid.org/0000-0003-2548-0074

FUNDING

Funder	Grant(s)	Author(s)
Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)	VI.Vidi.192.065	Robin A. Ohm
EC | European Research Council (ERC)	716132	Robin A. Ohm

AUTHOR CONTRIBUTIONS

Erik P. W. Beijen, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft | Robin A. Ohm, Conceptualization, Funding acquisition, Project administration, Resources, Software, Supervision, Writing – review and editing

DATA AVAILABILITY

All data are available under NCBI BioProject PRJNA1034581. The raw sequence reads of *T. aggressivum*, *T. harzianum*, and *P. lilacinum* are deposited in NCBI Sequence Read Archive (SRA) under accession numbers SRR26638389, SRR26638390, and SRR26638391, respectively. The genome assemblies and gene predictions are deposited in NCBI GenBank under accession numbers JAWRVG000000000, JAWRVH000000000, and JAWRVI000000000, respectively. The genome annotations can also be accessed at https://fungalgenomics.science.uu.nl.

REFERENCES

1. Miyazaki K, Tsuchiya Y, Okuda T. 1987. Infection of mushroom compost by *Trichoderma* species. Mushroom J 179:355–361.
2. Kubicek CP, Steindorff AS, Chenthamara K, Manganiello G, Henrissat B, Zhang J, Cai F, Kopchinskiy AG, Kubicek EM, Kuo A, Baroncelli R, Sarrocco S, Noronha EF, Vannacci G, Shen Q, Grigoriev IV, Druzhinina IS. 2019. Evolution and comparative genomics of the most common *Trichoderma* species. BMC Genom 20:485. https://doi.org/10.1186/s12864-019-5680-7
3. Samuels GJ, Dodd SL, Gams W, Castlebury LA, Petrini O. 2002. *Trichoderma* species associated with the green mold epidemic of commercially grown *Agaricus bisporus*. Mycologia 94:146–170.
4. Luangsa-Ard J, Houbraken J, van Doorn T, Hong S-B, Borman AM, Hywel-Jones NL, Samson RA. 2011. *Purpureocillium*, a new genus for the medically important *Paecilomyces lilacinus*. FEMS Microbiol Lett 321:141–149. https://doi.org/10.1111/j.1574-6968.2011.02322.x
5. Dons JJM, de Vries OMH, Wessels JGH. 1979. Characterization of the genome of the basidiomycete *Schizophyllum commune*. Biochim Biophys Acta 563:100–112. https://doi.org/10.1016/0005-2787(79)90011-x
6. 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
7. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that

uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62
8. Marian IM, Vonk PJ, Valdes ID, Barry K, Bostock B, Carver A, Daum C, Lerner H, Lipzen A, Park H, Schuller MBP, Tegelaar M, Tritt A, Schmutz J, Grimwood J, Lugones LG, Choi I-G, Wösten HAB, Grigoriev IV, Ohm RA. 2022. The transcription factor Roc1 is a key regulator of cellulose degradation in the wood-decaying mushroom *Schizophyllum commune*. mBio 13:e0062822. https://doi.org/10.1128/mbio.00628-22
9. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117:9451–9457. https://doi.org/10.1073/pnas.1921046117
10. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, Hogue G, Duncan A, Morgan M, Flores T, Bimber BN, Haw R, Cain S, Buels RM, Stein LD, Holmes IH. 2023. Jbrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol 24:1–21. https://doi.org/10.1186/s13059-023-02914-z
11. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. 2019. Apollo: democratizing genome annotation. PLoS Comput Biol 15:e1006790. https://doi.org/10.1371/journal.pcbi.1006790
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2