Genome Sequence of the Fungus *Trichoderma asperellum* SM-12F1 Revealing Candidate Functions of Growth Promotion, Biocontrol, and Bioremediation

Lijuan Li,¹ Xibai Zeng,¹ Jieyin Chen,^{2,†} Jian Tian,³ Jinqun Huang,⁴ and Shiming Su^{1,†}

¹ Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences/Key Laboratory of Agro-Environment, Ministry of Agriculture, Beijing, 100081, P.R. China

² Institute of Plant Protection, c/o State Key Laboratory for Biology of Plant Diseases and Insect Pests, Chinese Academy of Agricultural Sciences, Beijing, 100193, P.R. China

³ Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, 100081, P.R. China

⁴ BGI Genomics, BGI-Shenzhen, Shenzhen, 518083, China

Funding

Financial support was provided by the National Foundation of Natural Science of China (grant number 41671328) and the Science Innovation Project of the Chinese Academy of Agricultural Science (CAAS-ASTIP-2016-IEDA).

Keywords

arsenic bioremediation, biocontrol, genomic basis, plant growth promotion, *Trichoderma asperellum*

Abstract

Trichoderma spp. are versatile probiotic fungi that are known to promote plant growth and disease resistance under biotic stress, abiotic stress, or both. They are often used to prevent plant diseases caused by pathogens, act as biofertilizers, and are used in mycoremediation. In our previous study, *Trichoderma asperellum* strain SM-12F1 was isolated from soils contaminated with arsenic (As), adjacent to a realgar mine. SM-12F1 promoted plant growth and was useful for biocontrol and bioremediation. However, the genomic sequence of this strain was not characterized. This study aimed to generate high-quality genome resources for *T. asperellum* SM-12F1, and to determine the genomic basis of mechanisms behind plant growth promotion, biocontrol, and bioremediation of As in soil. Genome data of this fungus will provide perspectives on the molecular basis underlying biocontrol activity and mycoremediation.

Genome Announcement

Trichoderma spp. (teleomorph Hypocreaceae, Hypocreales, Ascomycota, and Dikarya) are recognized as biocontrol agents and have been widely utilized as probiotic fungi in agricultural applications because of their capacity to prevent phytopathogen infection and promote plant growth (Viterbo et al. 2010; Wijesinghe et al. 2010). *Trichoderma harzianum*, *T. atroviride*, and *T. virens* are known to enhance plant resistance against many pathogens (Kubicek and Druzhinina 2013) whereas *T. reesei* is known to promote plant growth by producing enzymes that facilitate the use of renewable nutrients (Kubicek 2013). In addition, *Trichoderma* spp. also can remediate heavy-metal-contaminated environments (Venice et al. 2020; Wang et al. 2015; X. Zhang et al. 2018). For example, *T. asperellum* protects crop health by reducing the bioavailability of heavy metals such as cadmium (Cd) and lead (Pb) through detoxification and sequestration (X. Zhang et al. 2018). *T. asperellum* has been successfully produced as chlamydospores to remediate arsenic (As)-contaminated soils via biovolatilization (Wang et al. 2015).

In 2010, we isolated *T. asperellum* strain SM-12F1 from soils contaminated with As, adjacent to a realgar mine located in Shimen County, Hunan Province, China (Zeng et al. 2010). Based on our previous study and other literature reports, three important functions of this strain are known to provide benefits for agricultural production. The first function is growth promotion; *T. asperellum* SM-12F1 can improve the activities of soil enzymes associated with nutrient

[†]Corresponding authors: S. Su; sushiming@caas.cn; and J. Chen; chenjieyin@caas.cn

The author(s) declare no conflict of interest.

Accepted for publication 3 February 2021.

activation and antioxidant stress (H. Zhang et al. 2018b) and induce plant growth-promoting attributes of 1-aminocyclopropane-1-carboxylate deaminase, auxin, and siderophore production (Qi and Zhao 2013; Viterbo et al. 2010). This strain can also produce induced systemic resistance (ISR) in plants by regulating ethylene pathways and signaling jasmonic acid release (Guo et al. 2020). The second important function is biocontrol, in which *T. asperellum* resists pathogenic fungi and nematodes by competing for nutrients and growing space, destroying the hyphae of pathogenic fungi and mycoparasitic activity, improving antibiotic production, and inducing disease resistance (Guo et al. 2020). The third important function is bioremediation, in which *T. asperellum* SM-12F1 can remediate As-contaminated environments by triggering the transformation of As species by reduction, oxidation, methylation (Su et al. 2012), demethylation (Viterbo et al. 2010; Su et al. 2017), and cellular sequestration (Li et al. 2021). *T. asperellum* SM-12F1 can also remediate environments contaminated by Pb and Cd (X. Zhang et al. 2018).

The genomes of *T. asperellum* strains CBS 433.97 (Druzhinina et al. 2018), B05 (Kwak 2020), Ts93 (GenBank accession number ML142921.1), and IC-1 (Takahashi et al. 2020) have been sequenced by using Illumina technology. The sequences are available at a publicly accessible website (<https://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/18221/>). The genome of the model strain CBS 433.97 consists of 419 scaffolds, with a total assembly length of 37.46 Mb and 12,557 protein-coding genes (H. Zhang et al. 2018a). However, the new sequencing technology allows improvement of the quality of this genome assembly. This is important because it will enable the discovery of biological characteristics based on globally comparative genomic data.

Presently, we generated a high-quality draft genome assembly of *T. asperellum* SM-12F1. The entire genome was systematically sequenced using long-read technology of the PacBio RS II System (Pacific Biosystems, Menlo Park, CA, U.S.A.). The high-quality genome assembly allowed for further analysis of the genomic basis of growth promotion, biocontrol, and the unique characteristics of As bioremediation by *T. asperellum* strain SM-12F1.

To obtain the whole genome of *T. asperellum* strain SM-12F1, mycelial tissue was collected by culturing on potato dextrose agar medium at 25°C in darkness for 7 days. High-quality genomic DNA was extracted for genome sequencing using the Quick-DNA Fungal/Bacterial Midiprep Kit (Zymo Research, Orange, CA, U.S.A.), following the manufacturer's protocol. Genomic DNA was used to construct a library with an insertion size of 20 kb using the SMRTbell Template Prep Kits (Pacific Biosciences). A library with an insert size of 400 bp was also constructed using TruSeq Nano DNA Library Prep Kits (Illumina, San Diego, CA, U.S.A.) to generate pair-end short reads for error correction of de novo assembly. The 20- and 400-bp libraries were sequenced using the PacBio RS II platform and Illumina HiSeq 2000 platform, respectively.

The genome sequence was assembled de novo using the PacBio subreads with the SMARTdenovo program and default parameters (-c 1 -t 8 -k 16). The assembled genome sequences were then extended by the SSPACE program using subreads (parameters: -o 50 -l 3) and Illumina clean reads (parameters: -o 20 -m 32 -k 5 -n 15) (Boetzer and Pirovano 2014). The PBJelly2 program (English et al. 2012) was used to minimize the gap between assembled sequences using default parameters. Error corrections were carried out using subreads by the variantCaller program from the SMRTlink v4 package (Pacific Biosystems). Subsequently, the GATK program (McCormick et al. 2015) was used for further error corrections with custom parameters (-cluster 2 -window 5 -stand_call_conf 50 -stand_emit_conf 10.0 -dcov 200 MQ0 ≥ 4). Finally, the assembled sequences were corrected using the SOAPsnp (parameters: -u -t -z @ -Q i -q) and SOAPindel program (parameters: -c 3 -h 1 -u 2 -m 2) (Li 2013) using Illumina clean reads. Transposable elements were identified by RepeatMasker (open 3.2.8, detailed parameters: -no_is, -norna, -engine, -s, -parallel = 1, using Repbase version 15.08) and Repeat-ProteinMask (-noLowSimple, -P value = 1e-4) (<http://www.repeatmasker.org>). The output files were summarized using a custom Perl script.

The genome assembly of SM-12F1 consisted of 18 contigs (N₅₀ length of 2,903,986 bp and N₉₀ length of 1,183,523 bp) and was further assembled into 14 scaffolds (N₅₀ length of 3,339,580 bp and N₉₀ length of 1,815,149 bp). Finally, a genome size of 36,780,254 bp was obtained. Each of the top seven assembled scaffold sequences exceeded 1 Mb in length and accounted for 28,449,162 bp in total length (77.3% of the genome size) (Table 1), close to the total predicted seven chromosomes for *T. asperellum* (Mäntylä et al. 1992; Srivastava et al. 2014). The analysis revealed that 2.97% of the SM-12F1 genome was composed of repetitive DNA (Table 1). The genome size of SM-12F1 was similar in size to the 37.46 to 38.47 Mb sizes of the CBS 433.97, B05, Ts93, and IC-1 *T. asperellum* genomes. The assembly quality was significantly improved by assembly into 14 scaffolds, in contrast to the more fragmented genomes of the

Table 1. Genome features of *Trichoderma asperellum* strain SM-12F1 and CBS 433.97

Parameters	SM-12F1	CBS 433.97
Sequence		
Total subreads (PacBio)	996,928	—
Mean subread length (bp, PacBio)	13,087	—
Total bases (Mb, PacBio)	12,655	—
Depth of PacBio sequence	344	—
Total reads (Illumina)	34,934,854	—
Total bases of clean reads (Mb, Illumina)	4,674	—
Depth of Illumina sequence	127×	130×
Assembly		
Assembly size (bp)	36,780,254	37,464,610
Number of scaffolds	14	419
N ₅₀ length (bp)	2,903,986	163,984
Size of top seven scaffolds (bp)	28,449,162	19,621,238
GC content (%)	48.13	47.3
Repeat content (%)	2.97	—
Annotation		
Predicted protein-coding genes	10,771	12,557
BUSCO (%) ^a	C = 99.6, S = 99.3, D = 0.3, F = 0.3, M = 0.1	
Annotated proteins with Swiss-Prot	6,673	6,709
Annotated proteins with eggNOG	9,954	10,213
Annotated proteins with KEGG ^b	7,189	4,417
Secreted proteins	721	776
Secreted carbohydrate active enzymes	218	227
Carbohydrate active enzymes	481	476

^a BUSCOs = Benchmarking Universal Single-Copy Orthologs ($n = 290$): C = complete BUSCOs; S = complete and single-copy BUSCOs; D = complete and duplicated BUSCOs; F = fragmented BUSCOs; and M = missing BUSCOs.

^b KEGG = Kyoto Encyclopedia of Genes and Genomes.

CBS 433.97, B05, Ts93 and IC-1 strains with 419, 1,439, 873, and 72 scaffolds, respectively (<https://www.ncbi.nlm.nih.gov/genome/18221/>). These results provide a higher quality and more integrated whole-genome assembly of *T. asperellum*.

For prediction of protein-coding genes, the gene annotation programs Augustus v3.3.2 (Stanke and Morgenstern 2005) and GeneMark-ES v4.46 (Ter-Hovhannisyan et al. 2008) were trained for use with MAKER v2.31.10 (Holt and Yandell 2011). GeneMark-ES was trained on the genomes of SM-12F1. The output results were used to train Augustus. In total, 10,771 protein-coding genes were predicted in the SM-12F1 genome. Analyses showed the presence of 99.6% complete Benchmarking Universal Single-Copy Orthologs, indicating a high-quality assembly. Functional annotation of the predicted genes was performed using the publicly available Swiss-Prot, eggNOG, and Kyoto Encyclopedia of Genes and Genomes databases (similarity > 30%, e-value < 1e-5). The annotation revealed 6,673 genes (61.95%), and 7,189 genes (66.74%), respectively (Table 1). Putative carbohydrate active enzymes (CAZymes) were identified with the HMM-based and BLASTP routine of the CAZymes database. The analysis revealed 481 genes that encoded CAZymes in the SM-12F1 genome. Of these genes, 218 displayed potential secretory characteristics. Putative secretory proteins contain the signal peptide, which is predicted using SignalP v5.0, (<http://www.cbs.dtu.dk/services/SignalP/>) and lack transmembrane domains, as jointly predicted using TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) and Phobius (<http://phobius.sbc.su.se/>). Furthermore, the putative proteins with extracellular location were predicted using WoLF PSORT (https://www.genscript.com/psort/wolf_psort.html). The comparative genomic results indicated that the SM-12F1 genome and the CBS 433.97 reference genomes encoded a similar number of genes encoding CAZymes and secretory CAZymes (Table 1).

Finally, the As transformation, biocontrol, and growth promotion-related genes (Wijesinghe et al. 2010; Zhao et al. 2018) in the SM-12F1 genome were systematically compared with the CBS 433.97 reference genome using functional annotations (Fig. 1). In total, 464 genes from SM-12F1 and 445 genes from CBS 433.97 were predicted to be functional genes (HMMER 3.1b2 [<http://hmmer.org/>]; e-value < 1e-5). The SM-12F1 and CBS 433.97 genomes encode 91 and 88 genes, respectively, associated with six functions of As transformation (Zhang et al. 2015). These genomes encode 175 and 181 genes, respectively, that are potentially involved in biocontrol and enhance the disease resistance of the host plant. These genes are involved

Figure 1. Functional genes of growth promotion, biocontrol, and bioremediation from *Trichoderma asperellum* strains SM-12F1 and CBS 433.97. The histogram represents the total number of functional genes and the pie chart represents the number and the proportion of different genetic types represented in the total number of functional genes.

in mycoparasitism (a total of 12 functional types, involving 28 genes in each genome), competition (hydrophobin and vegetative incompatibility proteins, involving 10 genes in each genome), and secondary metabolite biosynthesis (5 functional types in total, involving 137 and 143 genes, respectively) (Ansari et al. 2004). Moreover, the two genomes also encode genes that potentially participate in plant growth promotion (Li et al. 2019), including some functional genes associated with ISR (11 and 10 genes in the SM-12F1 and CBS 433.97 genome, respectively) and plant growth promotion (187 and 166 genes in the SM-12F1 and CBS 433.97 genome, respectively).

The capabilities of growth promotion, biocontrol, and bioremediation by *T. asperellum* SM-12F1 were determined by the whole-genome analysis. The genome sequences and corresponding annotation information presented here improve the understanding of plant-microbiome interactions in the environment and will aid future explorations of functional genes associated with plant growth and health.

The data from the Whole Genome Shotgun project have been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession numbers BioProject PRJNA647971, BioSample SAMN15666258, and Assembly GCA_014839745.1.

Literature Cited

- Ansari, M. Z., Yadav, G., Gokhale, R. S., and Mohanty, D. 2004. NRPS-PKS: A knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* 32:W405-W413.
- Boetzer, M., and Pirovano, W. 2014. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15:211.
- Druzhinina, I. S., Chenthamara, K., Zhang, J., Atanasova, L., Yang, D., Miao, Y., Rahimi, J. M., Grujic, M., Cai, F., Pourmehdi, S., Salim, K. A., Pretzer, C.,

- Kopchinskiy, A. G., Henrissat, B., Kuo, A., Hundley, H., Wang, M., Aerts, A., Salamov, A., Lipzen, A., LaButti, K., Barry, K., Grigoriev, I. V., and Kubicek, C. P. 2018. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. *PLoS Genet.* 14:e1007322.
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Worley, K. C., and Gibbs, R. A. 2012. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7:e47768.
- Guo, R., Wang, Z., Zhou, C., Liu, Z., Zhang, P., and Fan, H. 2020. Transcriptomic analysis reveals biocontrol mechanisms of *Trichoderma harzianum* ACCC30371 under eight culture conditions. *J. For. Res.* 31:1863-1873.
- Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12:491.
- Kubicek, C. P. 2013. Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *J. Biotechnol.* 163:133-142.
- Kubicek, C. P., and Druzhinina, I. S. 2013. *Trichoderma*: Genomic aspects of mycoparasitism and biomass degradation. Pages 127-156 in: *Genomics of Soil- and Plant-Associated Fungi*. Soil Biology, Vol. 36. B. Horwitz, P. Mukherjee, M. Mukherjee, and C. Kubicek, eds. Springer, Berlin, Heidelberg, Germany.
- Kwak, Y. 2020. Complete mitochondrial genome of the fungal biocontrol agent *Trichoderma atroviride*: Genomic features, comparative analysis and insight into the mitochondrial evolution in *Trichoderma*. *Front. Microbiol.* 11:785.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.
- Li, L., Zeng, X., Williams, P. N., Gao, X., Zhang, L., Zhang, J., Shan, H., and Su, S. 2021. Arsenic resistance in fungi conferred by extracellular bonding and vacuole-septa compartmentalization. *J. Hazard. Mater.* 401:123370.
- Li, X., Jousset, A., de Boer, W., Carrión, V. J., Zhang, T., Wang, X., and Kuramae, E. E. 2019. Legacy of land use history determines reprogramming of plant physiology by soil microbiome. *ISME J.* 13:738-751.
- Mäntylä, A. L., Rossi, K. H., Vanhanen, S. A., Penttilä, M. E., Suominen, P. L., and Nevalainen, K. H. 1992. Electrophoretic karyotyping of wild-type and mutant *Trichoderma longibrachiatum* (reesei) strains. *Curr. Genet.* 21:471-477.
- McCormick, R. F., Truong, S. K., and Mullet, J. E. 2015. RIG: Recalibration and interrelation of genomic sequence data with the GATK. *G3: Genes Genomes Genetics* 5:655-665.
- Qi, W., and Zhao, L. 2013. Study of the siderophore-producing *Trichoderma asperellum* Q1 on cucumber growth promotion under salt stress. *J. Basic Microbiol.* 53:355-364.
- Srivastava, M., Shahid, M., Pandey, S., Singh, A., Kumar, V., Gupta, S., and Maurya, M. 2014. *Trichoderma* genome to genomics: A review. *J. Data Min. Genomics Proteomics.* 5:1000162.
- Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33:W465-W467.
- Su, S., Zeng, X., Bai, L., Wang, Y., Zhang, L., Li, M., and Wu, C. 2017. Concurrent methylation and demethylation of arsenic by fungi and their differential expression in the protoplasm proteome. *Environ. Pollut.* 225:620-627.
- Su, S. M., Zeng, X. B., Li, L. F., Duan, R., Bai, L. Y., Li, A. G., Wang, J., and Jiang, S. 2012. Arsenate reduction and methylation in the cells of *Trichoderma asperellum* SM-12F1, *Penicillium janthinellum* SM-12F4, and *Fusarium oxysporum* CZ-8F1 investigated with X-ray absorption near edge structure. *J. Hazard. Mater.* 243:364-367.
- Takahashi, H., Sakagawa, E., Toyoda, R., Motomura, T., Murase, M., Takahashi, A., Fukuyoshi, S., and Kanamasa, D. 2020. Draft genome sequence of glycoside hydrolase-producing *Trichoderma asperellum* strain IC-1. *Microbiol. Resour. Announce.* 9:e00958-20.
- Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18:1979-1990.
- Venice, F., Davolos, D., Spina, F., Poli, A., Prigione, V. P., Varese, G. C., and Ghignone, S. 2020. Genome sequence of *Trichoderma lixii* MUT3171, a promising strain for mycoremediation of PAH-contaminated sites. *Microorganisms* 8:1258.
- Viterbo, A., Landau, U., Kim, S., Chernin, L., and Chet, I. 2010. Characterization of ACC deaminase from the biocontrol and plant growth-promoting agent *Trichoderma asperellum* T203. *FEMS Microbiol. Lett.* 305:42-48.
- Wang, X. R., Su, S. M., Zeng, X. B., Bai, L. Y., Li, L. F., Duan, R., Wang, Y. N., and Wu, C. X. 2015. Inoculation with chlamydospores of *Trichoderma asperellum* SM-12F1 accelerated arsenic volatilization and influenced arsenic availability in soils. *J. Integr. Agric.* 14:389-397.
- Wijesinghe, C. J., Wijeratnam, R. W., Samarasekera, J. K. R. R., and Wijesundera, R. L. C. 2010. Biological control of *Thielaviopsis paradoxa* on pineapple by an isolate of *Trichoderma asperellum*. *Biol. Control* 53:285-290.
- Zeng, X., Su, S., Jiang, X., Li, L., Bai, L., and Zhang, Y. 2010. Capability of pentavalent arsenic bioaccumulation and biovolatilization of three fungal strains under laboratory conditions. *Clean: Soil, Air, Water* 38:238-241.
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018a. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46:W95-W101.
- Zhang, H., Zeng, X., Bai, L., Shan, H., Wang, Y., Wu, C., Duan, R., and Su, S. 2018b. Reduced arsenic availability and plant uptake and improved soil microbial diversity through combined addition of ferrihydrite and *Trichoderma asperellum* SM-12F1. *Environ. Sci. Pollut. Res.* 25:24125-24134.
- Zhang, S. Y., Zhao, F. J., Sun, G. X., Su, J. Q., Yang, X. R., Li, H., and Zhu, Y. G. 2015. Diversity and abundance of arsenic biotransformation genes in paddy soils from southern China. *Environ. Sci. Technol.* 49:4138-4146.
- Zhang, X., Li, X., Yang, H., and Cui, Z. 2018. Biochemical mechanism of phytoremediation process of lead and cadmium pollution with *Mucor circinelloides* and *Trichoderma asperellum*. *Ecotoxicol. Environ. Saf.* 157:21-28.
- Zhao, Y., Su, J. Q., Ye, J., Rensing, C., Tardif, S., Zhu, Y. G., and Brandt, K. K. 2018. AsChip: A high-throughput qPCR chip for comprehensive profiling of genes linked to microbial cycling of arsenic. *Environ. Sci. Technol.* 53:798-807.