Near-complete genomes of two *Trichoderma* species: a resource for biological control of plant pathogens

Introduction

Trichoderma species are widely used to control fungal and nematode diseases of crops. To date, only one complete *Trichoderma* genome has been sequenced, *T. reesei* stain QM6a, a model fungus for industrial enzyme production, while the species or strains used for biological control of plant diseases are only available as draft genomes. Previously, we demonstrated that two *Trichoderma* strains (*T. harzianum* and *T. cyanodichotomus*) provide effective control of nematode and fungal plant pathogens. Based on deep sequencing using Illumina and Pacbio platforms, we have assembled high-quality genomes of the above two strains, with contig N50 reaching 4.2 Mbp and 1.7 Mbp, respectively, higher than published draft genomes. The genome data will provide a resources to assist research on the biological control mechanisms of *Trichoderma* spp.

Data

The raw sequence and assembled genome have been submitted to the NCBI database: BioProject: PRJNA596042, BioSample: SAMN13611475, accession number: WUWT00000000, for *Trichoderma harzianum* strain T11-W; and BioProject: PRJNA598077, BioSample: SAMN13698093, accession number: WXUD00000000, for *Trichoderma cyanodichotomus* strain TW21990-1.

Here we provide the sequences of each gene identified from the genomes (the .fasta files) and the detailed genome annotation (.xlsx files) including non-coding RNA (sRNA, rRNA, tRNA, snRNA and miRNA), transposon sequences, and the annotation of the genes from the National Center for Biotechnology Information (NCBI) non-redundant (NR) database, Gene Ontology (GO), Cluster of Orthologous Groups of proteins (COG) and CAZy.

Methods

Non-coding RNA: rRNAs was identified by comparing with rRNA database or predicting with RNAmmer software (version 1.2 www.cbs.dtu.dk/services/RNAmmer Parameters:–s Species –m Type –gff *. rRNA.gff –f *.rRNA.fq); tRNA was predicted by tRNAscan-SE (version:1.3.1, https://gtrnadb.ucsc.edu Parameters:–Spec_tag(BAOG) –o *. tRNA –f *); sRNAs was identified by comparing with database Rfam (version:9.1, https://rfam.sanger.ac.uk Parameters:–p blastn –W 7 –e 1 –v 10000 –b 10000 –m 8 –i subfile –o *.blast.m8).

The transposon sequence was searched by aligning the assembled genome with the known transposon sequence database and the *De novo* method, using Repeatmasker (version:4-0-6, www.repeatmasker.org Parameters:–nolow –no_is –norna –engine wublast –parallel 1 –lib lib file) and Tandem Repeat Finder (version:4.04, https://tandem.bu.edu/trf/trf.html Parameters:2 7 7 80 10 50 Period_size –d –h seqfile).

Function annotation is completed by blasting genes with different databases (https://blast.ncbi.nlm.nih.gov/).