A *Cladosporium halotolerans* strain 550W1 isolated from Lushan Russet Potato confers broad-spectrum plant growth-promoting activity

Libing Liao[1,2,a], Yihong Hu[1,2,a], Yuhua Lu[1,2,a], Nan Wu[1,2,3], Shifeng Liu[4], Min Zhong[5] and Deshui Yu[1,2,3*]

[1] Lushan Botanical Garden, Jiangxi province and Chinese Academy of Sciences, Jiujiang, 332900, Jiangxi, China

[2] Jiangxi Provincial Key Laboratory of Wetland Plant Resources Conservation and Utilization, Jiujiang, 332900, Jiangxi, China

[3] School of Life Sciences, Nanchang University, Nanchang, 330031, Jiangxi, China

[4] Panxi Crop Improvement Key Laboratory of Sichuan Province, College of Agricultural Science, Xichang University, Liangshan 615300, China

[5]Jiangxi Cash Crop Research Institute, Nanchang 330006, Jiangxi, China

[a] These authors contribute equally to this work.

[*] Corresponding author Deshui Yu, E-mail address: yuds@lsbg.cn

Supplementary results

Consistency and completeness evaluation of assembled genome

BWA v0.7.13 was used to map Illumina clean reads back to assembled contigs to evaluate genome consistency. The mapping rate reached 99.73%, and average sequencing depth was 202.22×, with genome-wide coverage of 99.45%. The nearly complete read mapping and uniform coverage indicated high consistency between sequencing data and assembled contigs.

BUSCO v4.1.2 with fungi_odb10 dataset was adopted to assess assembly completeness. Among 758 universal fungal single-copy orthologs, 741 (97.8%) were complete, including 720 single-copy BUSCOs (95.0%) and 21 duplicated BUSCOs (2.8%). Only 2 fragmented (0.3%) and 15 missing BUSCOs (1.9%) were detected, demonstrating high integrity of the assembled genome (Supplementary Table5).

Genome annotation

The MAKER v2.31.10 pipeline integrated three independent evidence sets: (1) ab initio prediction by GlimmerHMM v3.0.4 and Augustus v3.3.2; (2) homology-based prediction via Exonerate v2.4.0 using five Metarhizium reference protein sequences (M. album, M. anisopliae, M. brunneum, M. acridum, M. robertsii); (3) transcript evidence from StringTie-assembled RNA-seq transcripts processed by TransDecoder v5.1.0.

After integration and filtering, a total of 7,593 non-redundant protein-coding genes were obtained. Statistical structural features: average mRNA length = 2,985.67 bp,

average CDS length = 1,503.13 bp, average 2.39 exons per gene, mean exon length = 1,178.94 bp, mean intron length = 118.32 bp. Length distribution curves of gene, CDS, exon and intron were plotted for interspecific comparison with related Metarhizium species (Supplementary Table6).

Three classes of non-coding RNAs were annotated separately. tRNAscan-SE v1.23 predicted 184 tRNA copies with average length of 94 bp. Homology search identified 76 rRNA fragments (12 × 18S, 24 × 5.8S, 40 × 5S). INFERNAL v1.1.2 combined with Rfam database annotated 35 snRNA genes. The total length of all ncRNAs was 51,075 bp, occupying only 0.18% of the full genome (Supplementary Table7).

BUSCO v5.2.2 (fungi_odb10) was performed on predicted protein sequences to evaluate gene set integrity. A total of 721 complete BUSCOs (95.1%) were recovered, including 701 single-copy and 20 duplicated orthologs; 2 fragmented and 35 missing BUSCOs were identified. The BUSCO completeness higher than 90% represented reliable and high-quality genome annotation. The slightly lower completeness relative to raw genome BUSCO score was attributed to incomplete capture of low-expression genes in transcriptome data.

Gene functional annotation

A total of 7,593 protein-coding genes predicted from the genome of strain 550W-1 were subjected to systematic multi-database functional annotation, with detailed annotation statistics summarized in the functional classification table (Supplementary Table10 and Supplementary Table11).

Homology searches were conducted using Diamond blastp (version 2.0.11.149) with a threshold E-value of 1e-5 against three databases: the NCBI Nr database annotated 7,219 genes, accounting for 95.07% of the total gene set; the UniProt database matched 7,186 genes, occupying 94.64%; and the KEGG metabolic pathway database provided orthology annotations for 5,566 genes, corresponding to 73.30% of all genes (Supplementary Fig. S12). KEGG annotation was further processed using the KOBAS v3.0 pipeline to associate genes with KEGG ORTHOLOGY and PATHWAY entries, resulting in 2,899 genes (38.18%) being assigned to specific metabolic pathways. In addition, the RefSeq database provided homologous functional descriptions for 7,150 genes, corresponding to 94.17% of all genes.

Conserved protein domain detection was carried out via InterProScan v5.52-86.0 and hmmscan (version 3.3.2, E-value 0.01) based on hidden Markov models. A total of 7,265 genes (95.68%) were identified with conserved domains recorded in the InterPro integrated database, while 5,913 genes (77.87%) were assigned to specific Pfam domain families; supplementary annotation against the Tigerfam database further identified 2,253 annotated genes, making up 29.67% of the full gene collection.

Orthologous functional grouping was performed using the eggNOG v5.0 database, and 1,674 genes (22.05%) were successfully classified into eukaryotic orthologous groups (KOGs) (Supplementary Fig. S13).

In addition, Gene Ontology (GO) classification extracted functional terms covering biological process, cellular component and molecular function from InterPro and UniProt matching results, and 5,968 genes (78.60%) were assigned clear GO classification information (Supplementary Fig. S14).

Overall, 7,446 out of the total 7,593 genes received reliable functional annotation evidence from at least one universal database, reaching an overall annotation rate of 98.06%, which reflected the high integrity of gene functional information of strain 550W1(Supplementary Table10).

Supplementary Figure Legends

Supplementary Fig. S1. Colony growth of strain 550W1 on five different culture media (PDA, MEA, SDA, CDA, MS). Optimal growth was observed on PDA. Bar=2 cm.

Supplementary Fig. S2. Microscopic observation of septate hyphae and spherical conidia of 550W1, bar=40 μm.

Supplementary Fig. S3. Salt tolerance assay of strain 550W1 cultured on PDA medium amended with NaCl at concentrations ranging from 0 M (CK) to 4.0 M. The strain maintained normal colony growth when NaCl concentration was no more than 2.0 M, whereas its growth was completely suppressed at 4.0 M NaCl. Bar=2 cm.

Supplementary Fig. S4. Colonial growth characteristics of strain 550W1 cultured on PDA medium under a pH gradient from 4.0 to 11.0. Bar=2 cm.

Supplementary Fig. S5. Mycelial dry weight growth curve of 550W1 during 0–8 days of liquid incubation.

Supplementary Fig. S6. Maximum rosette leaf length and leaf SPAD value of Arabidopsis plants with or without strain 550W1 inoculation.

A: maximum rosette leaf length; **B:** leaf SPAD value. ($n \geqslant 7$). Statistical significance was determined using Student's *t*-test (** $P < 0.01$).

Supplementary Fig. S7. Morphology of field-cultivated soybean cv. Huadou 16.

A: Representative soybean plants of non-inoculated control (CK, left) and 550W1 inoculation treatment (right). Scale bar=10 cm. **B:** Boxplot illustrating shoot height for CK and 550W1 groups, n=20, statistical significance was determined using Student's *t*-test (** $P < 0.01$).

Supplementary Fig. S8. Divided plate volatile organic compound (VOC) assay. Left half without fungal inoculation (CK), right half co-cultivated with 550W1 to detect *A. thaliana* growth promotion triggered by fungal VOCs, bar=2 cm.

Supplementary Fig. S9. MLSA-based phylogeny and sequence similarity of *Cladosporium* isolates.

A: Maximum-likelihood tree inferred from concatenated ITS-*tef1*-*act*-*tub2*-*gapdh*-*rpb2* sequences. Strain 550W1 is highlighted in red, with *Cercospora berteroae* CBS538 as outgroup. **B:** Heatmap of pairwise concatenated-sequence identity (%) for all strains. Color scale denotes identity. Strain 550W1 is indicated in red.

Supplementary Fig. S10. Phylogenetic analysis of strain 550W1 and related

Cladosporium species using whole-genome sequences.

Maximum-likelihood phylogenetic tree constructed from whole-genome sequences of strain 550W1 and related *Cladosporium* strains. Numbers on branches represent genetic distances, and bootstrap support values are indicated at the nodes. *Cercospora berteroae* CBS538 was used as the outgroup.

Supplementary Fig. S11. Whole-genome sequences average nucleotide identity (ANI) of strain 550W1 and related *Cladosporium* species.

Heatmap of pairwise average nucleotide identity (ANI) values between strain 550W1 and the selected *Cladosporium* strains. The color gradient from blue to red indicates increasing ANI values.

Supplementary Fig. S12. Classification statistics of annotated genes of strain 550W1 mapped to KEGG metabolic pathways.

Supplementary Fig. S13. KOG functional classification of all protein-coding genes of strain 550W1.

Supplementary Fig. S14. Gene Ontology (GO) functional classification of 550W1.

The bar charts show the top enriched GO terms categorized into three main ontologies: biological process **A**, cellular component **B**, and molecular function **C**. The horizontal axis represents the number of annotated genes assigned to each GO term, and the vertical axis lists the enriched functional terms. GO annotation was performed on all predicted protein-coding genes of strain 550W1 genome. BP, biological process; CC, cellular component; MF, molecular function.

Supplementary Fig. S15. Genomic locations of genes involved in IAA, GA and CTK biosynthesis across all contigs of strain 550W1.

Horizontal grey bars represent each assembled contig with horizontal axis showing length in kilobases. Red vertical marks denote IAA biosynthetic genes, blue for GA-related genes, green for CTK biosynthetic genes.

Supplementary Table Legends

Supplementary Table1. Reference genome information of Cladosporium isolates used in this study.

Supplementary Table2. Statistics of Illumina NovaSeq 6000 sequencing data of strain 550W1 genome.

Supplementary Table3. Statistics of Oxford Nanopore PromethION long-read sequencing data of strain 550W1 genome.

Rank represents read length thresholds (in bp), with >0 covering all reads. Flag indicates data categories: all for total raw reads, pass for high-quality valid reads retained after filtering, and fail for low-quality discarded reads. TotalBase and TotalReads denote total base count and total read number respectively. MaxLen and AvgLen refer to maximum and average read length. N50/N90 are defined as the length of the last read when reads sorted from longest to shortest are accumulated to reach 50%/90% of total bases; L50/L90 correspond to the minimum number of reads required to reach 50%/90% of total bases. meanQ represents the average Phred quality score of all reads.

Supplementary Table4. Summary statistics of the de novo assembled genome of strain 550W1.

Total length represents the total base pair length of the final assembled genome; Total number refers to the total number of assembled contigs. GC content indicates the genomic GC percentage. Contig N50 and N90 are calculated by sorting all contigs from longest to shortest and accumulating their lengths until reaching 50% and 90% of the total genome length, respectively, with the length of the last accumulated contig defined as N50/N90. Average, Median, Min and Max separately stand for the average length, median length, minimum length and maximum length of all contigs.

Supplementary Table5. BUSCO completeness assessment of the assembled genome of strain 550W1.

Complete BUSCOs (C) represent full-length conserved orthologs; Complete and single-copy BUSCOs (S) are intact orthologs with only one copy in the genome; Complete and duplicated BUSCOs (D) refer to full-length orthologs with multiple copies. Fragmented BUSCOs (F) stand for partially recovered incomplete orthologs, while Missing BUSCOs (M) are orthologs absent from the assembly. Total BUSCO groups searched indicates the total number of universal fungal single-copy orthologs contained in the fungi_odb10 database.

Supplementary Table6. Protein-coding genes prediction of strain 550W1 genome.

Supplementary Table7. Statistical results of non-coding RNA of strain 550W1 genome.

Supplementary Table8. Reference primers for retrieving housekeeping gene sequences from genome assemblies.

Supplementary Table9. Pairwise ANI values of 550W1 and related Cladosporium isolates.

Supplementary Table10. Statistical summary of functional annotation for protein-coding genes of strain 550W1

Supplementary Table11. Functional annotation for protein-coding genes of strain 550W1.

Supplementary Table12. Candidate functional genes annotated in the genome of 550W1, including genes involved in phytohormone biosynthesis, antioxidant enzymes (SOD, CAT) and NRPS secondary metabolism.