scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of *Fusarium tricinctum*, a fungal pathogen associated with root rot in alfalfa

Shiyang Xu[1], Xizhe Sun[2], Xueping Li[3], Lisong Ma[4] & Xiangling Fang[1]✉

Fusarium tricinctum **is a globally devastating fungal pathogen associated with root rot disease in alfalfa. The absence of a complete genome sequence has hindered molecular investigations into its pathogenic mechanisms. This study reports the first chromosome-level genome assembly of the virulent *F. tricinctum* strain MsR-QD66, generated using a combined sequencing strategy integrating Illumina, PacBio HiFi, and Hi-C technologies. The 46.2 Mb assembly (scaffold N50 = 5.1 Mb) comprises 10 chromosomes with Merqury-validated abase accuracy (QV = 34.8, error rate <0.01%). The genome achieved 95.6% completeness according to CEGMA and complete BUSCO scores. This assembly comprises 13,594 protein-coding genes, 336 non-coding RNAs, and a comprehensive repertoire of repetitive elements. Functional annotation identified 1,871 transcription factors and 1,241 secreted proteins. Comparative genomics and virulence factor analyses revealed 527 genes linked to 36 key diseases across diverse hosts. This chromosome-level genome assembly provides a valuable resource for studying molecular mechanisms of *F. tricinctum* pathogenicity to alfalfa and will facilitate the development of sustainable strategies for managing root rot in alfalfa.**

Background & Summary

Fusarium tricinctum is an economically important, toxin-producing fungal pathogen associated with root rot in a wide range of crops[1,2]. Root rot, often referred to as the "cancer" of alfalfa, causes annual yield losses of up to 20% in global alfalfa production, with plant mortality rates in infected fields ranging from 61% to 73%, and *F. tricinctum* has been identified as one of the primary pathogens responsible for this disease[2]. Root rot caused by *F. tricinctum* exhibits distinct symptoms and greater virulence than other *Fusarium* species, with host susceptibility up to 100% and a disease index exceeding 90[3]. Additionally, *F. tricinctum* produces a broader spectrum of mycotoxins than most *Fusarium* species, including enniatins, beauvericin, moniliformin, T-2 toxin and fusarielins, which pose a substantial threat to agriculture and pasture ecosystems reducing yield losses and product quality[1]. However, the lack of a complete genomic blueprint has impeded molecular investigations of its pathogenicity mechanisms and the development of effective management strategies.

Whole-genome sequencing has advanced fungal genetics by offering high-resolution insights into genomic architecture[4]. Recent improvements in sequencing technologies and assembly algorithms have facilitated chromosome-level assemblies for multiple *Fusarium* species, including *Fusarium oxysporum*, *F. foetens*, and *F. verticillioides*, and these high-quality genomes substantially deepened our understanding of their pathogenic mechanisms[5–8]. However, assemblies of *F. tricinctum* remain at the scaffold or contig level, characterized by a fragmented genomic landscape with numerous gaps and incomplete or erroneous gene models[9]. The lack of comprehensive whole-genome sequences and accurate annotations for *F. tricinctum* limits research into the molecular basis of its pathogenicity, as well as the biosynthetic and regulatory pathways governing its mycotoxin production[4,8].

[1]Center for Grassland Microbiome, State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, 730020, China. [2]Hebei Key Laboratory of Plant Physiology and Molecular Pathology, College of Life Science, Hebei Agricultural University, Baoding, 071001, China. [3]Institute of Plant Protection, Gansu Academy of Agricultural Sciences, Lanzhou, 730070, China. [4]College of Horticulture, Hebei Agricultural University, Baoding, 071001, China. ✉e-mail: xlf@lzu.edu.cn

Statistics	Illumina PE150	PacBio HiFi	Hi-C
Library size (bp)	350	15,000	350
Raw data (Gb)	3.2	14.9	4.4
N50 (bp)	150	16,665	150
Longest reads (bp)	150	49,623	150
Mean read length (bp)	150	16,359	150
Coverage (X)	54.5	322.2	76

Table 1. A summary of the sequencing statistics output of *Fusarium tricinctum* MsR-QD66.

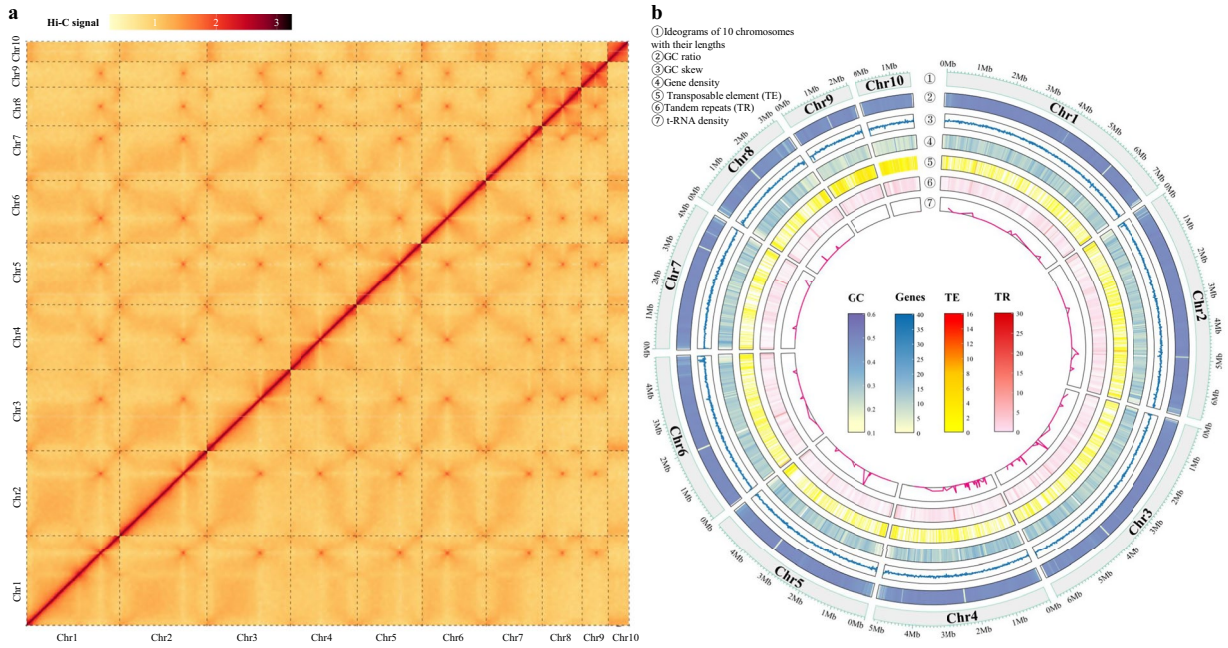

Fig. 1 Overview of the gap-free genome assembly and annotation of *Fusarium tricinctum* MsR-QD66. (**a**) Genome wide Hi-C contact map indicating the interaction matrices among ten chromosomes. (**b**) Circos plot illustrating the genomic features. Chromosomes (Chr) are represented around the plot circumference, with bands indicating genomic windows of 10 kb.

This study implemented a multi-omics approach to resolve the genomic architecture of *F. tricinctum* MsR-QD66. Initial Illumina NovaSeq PE150 sequencing generated 3.2 Gb raw data, which underwent stringent quality filtering to yield 2.5 Gb high-fidelity reads (Table 1). A 15-mer frequency spectrum analysis at 51.3× sequencing depth estimated the haploid genome size at 45.9 Mb, with 0.06% heterozygosity, 18.9% repetitive content, and a GC-content of 48.3%. To achieve chromosomal continuity, PacBio SMRT sequencing produced 14.9 Gb circular consensus sequencing (CCS) HiFi reads (N50: 16,665 kb; mean length: 16,359 bp), providing 322.2× genome coverage (Table 1). Hi-C scaffolding generated 4.4 Gb raw data (76× coverage), processed via Juicer v2.0 to obtain 3.5 Gb valid chromatin interaction pairs (Table 1). The hybrid assembly pipeline combining HiFi data and Hi-C data resolved the genome into 10 gap-free chromosomes (1.6~7.1 Mb) spanning 46.2 Mb, with exceptional continuity metrics (scaffold N50: 5.1 Mb; N90: 3.1 Mb) (Fig. 1a,b; Table 2). Compared to the gap-free assemblies of *F. oxysporum* (GCA_030345115.1) and *F. verticillioides* (GCA_027571605.1), the *F. tricinctum* MsR-QD66 assembly exhibits larger assembly size than GCA_027571605.1 while smaller than that of GCA_030345115.1 (Table 2). This assembly encodes 13,594 protein-coding genes, spanning a total length of 21,276,861 bp with an average gene length of 1,565 bp, representing 46% of the genome (Table 2). Interspersed repeat contains 1,465 LTR elements (265,002 bp), 1,262 DNA transposons (609,476 bp), 637 LINEs (161,941 bp), 59 SINEs (3,794 bp), and 29 rolling-circles (1,994 bp), with the lengths of LTR elements and LINEs are greater than those in GCA_027571605.1 but less than in GCA_030345115.1 (Tables 2, 3). A total of 336 non-coding RNAs were identified, including sRNAs (1,041 bp), tRNAs (27,260 bp), and snRNAs (snRNAs, 2,916 bp), representing approximately 0.07% of the genome (Tables 2, 3). Additionally, 4,880 tandem repeats spanned 295,255 bp (0.64% of the genome), with unit lengths ranging from 1 to 1,102 bp. Minisatellite DNA and microsatellite DNA occupied 0.31% (143,185 bp) and 0.06% (29,759 bp) of the assembly, respectively (Tables 2, 4).

For the genome annotation of protein coding genes in NR, KEGG, Pfam, GO, Swiss-Prot, KOG, CAZy and TCDB databases, individually ~13,436, ~11,931, ~8,381, ~8,381, ~4,040, ~2,250, ~1,018 and ~790 protein coding genes were annotated (Fig. 2). Among them, KEGG database exhibited the highest annotation yield for protein coding genes, whereas TCDB demonstrated the lowest coverage. Besides, 54 coding genes were

Statistics		This study	GCA_030345115.1	GCA_027571605.1
Assembly	Assembly Size (bp)	46,236,445	69,556,834	41,994,356
	Number of contigs	10	18	11
	Contig N50 (bp)	5,140,998	4,366,112	4,275,051
	Contig N90 (bp)	3,081,585	2,829,210	2,453,640
	Number of Scaffolds	10	18	11
	Scaffold N50 (bp)	5,140,998	4,366,112	4,275,051
	Scaffold N90 (bp)	3,081,585	2,829,210	2,453,640
	Longest scaffold (bp)	7,063,146	6,427,461	6,252,867
	Gap bases (bp)	0	0	0
Annotation	Number of genes	13,594	21,460	15,230
	GC content (%)	47.5	47.9	48.3
	LTR elements (bp)	265,002	1,659,789	120,266
	DNA transposons (bp)	609,476	5,489,141	102,640
	LINEs (bp)	161,941	1,227,612	54,265
	SINEs (bp)	3,794	0	10,343
	Rolling-circles (bp)	1,994	1,980,856	Unknown
	TR	295,255	Unknown	Unknown
	Minisatellite DNA	143,185	Unknown	Unknown
	Microsatellite DNA	29,759	Unknown	Unknown
	sRNA (bp)	1,041	8,155	Unknown
	tRNA (bp)	27,260	Unknown	Unknown
	snRNA	2,916	Unknown	Unknown
Quality Assessment	BUSCO (%)	100	98.9	99.9
	Quality value	34.7	Unknown	88.8
	CEGMA (%)	95.6	99.6	99.6
	NGS mapping ratio (%)	96.1	Unknown	98.3

Table 2. Statistics for genome assembly and annotation of *Fusarium tricinctum* MsR-QD66.

Statistics		Number	Average length (bp)	In genome (%)
Interspersed repeats	LTR elements	1,465	183	0.5731
	DNA transposons	1,262	486	1.3182
	LINEs	637	261	0.3502
	SINEs	59	64	0.0082
	Rolling-circles	29	69	0.0043
Non-coding RNAs	Small RNAs	4	260	0.0023
	tRNA	315	86	0.0590
	snRNA	17	171	0.0063

Table 3. Statistics of interspersed repeat and ncRNAs in genome assembly of *Fusarium tricinctum* MsR-QD66.

Tandem repeats	Number	Repeat size (bp)	In genome (%)
Total	4,880	1~1,102	0.6386
Minisatellite DNA	3,167	10~60	0.3097
Microsatellite DNA	700	2~6	0.0644

Table 4. Statistics of tandem repeats in genome assembly of *Fusarium tricinctum* MsR-QD66.

inserted across seven databases, and NR, KEGG, Pfam, and GO databases performed the largest intersection size. Notably, 588 protein-coding genes were uniquely mapped to NR database, while a minimal subset (n = 2) showed exclusive annotation within KEGG, and no database-specific annotations were observed in remaining repositories. Transcriptional regulation analysis identified 1,871 transcription factors (TFs) across 51 families, dominated by ZnCys (n = 649) (Fig. 3). Secretome prediction revealed 1,498 secreted proteins, 1,241 lacking transmembrane domains, suggesting extracellular effector candidates. For the annotation of BCGs, 680 genes afflicted to 56 BCGs were detected in the genome (Fig. 4). Based on the annotation information against DFVF databases, 527 genes were predicted to be the virulence factors for causing 36 kinds of key-diseases in vertebrata, herb, plants, animals, xyloid, and invertebrate (Fig. 5). A total of 11 genes were predicted to be related with

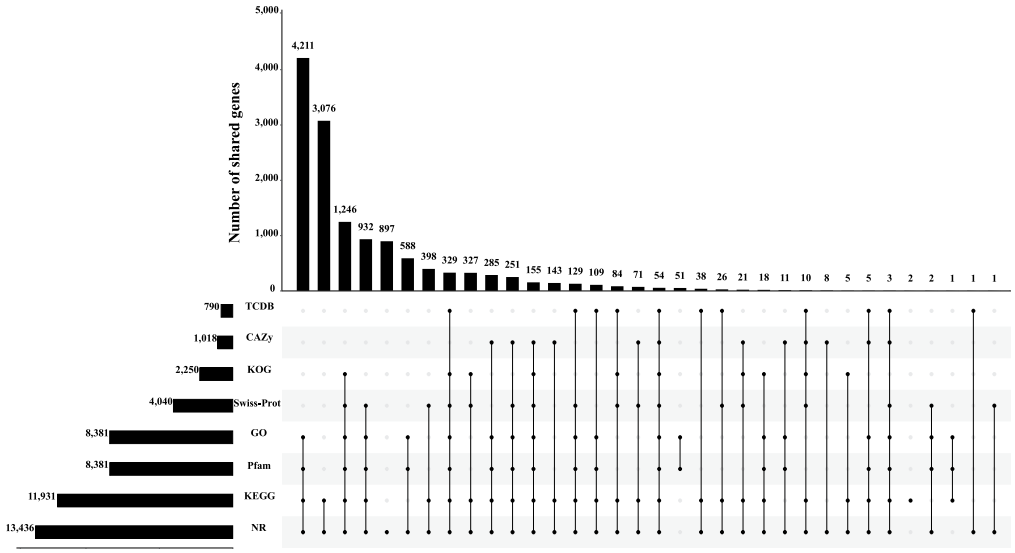

Fig. 2 Distribution of protein-coding genes annotated across multiple databases in the genome assembly of *Fusarium tricinctum* MsR-QD66. The number of protein-coding genes was annotated in the NR, KEGG, Pfam, GO, Swiss-Prot, KOG, CAZy, and TCDB databases. Black dots connected by vertical lines indicate the intersections of genes annotated by different combinations of these databases, illustrating the overlap and unique contributions of each annotation resource.

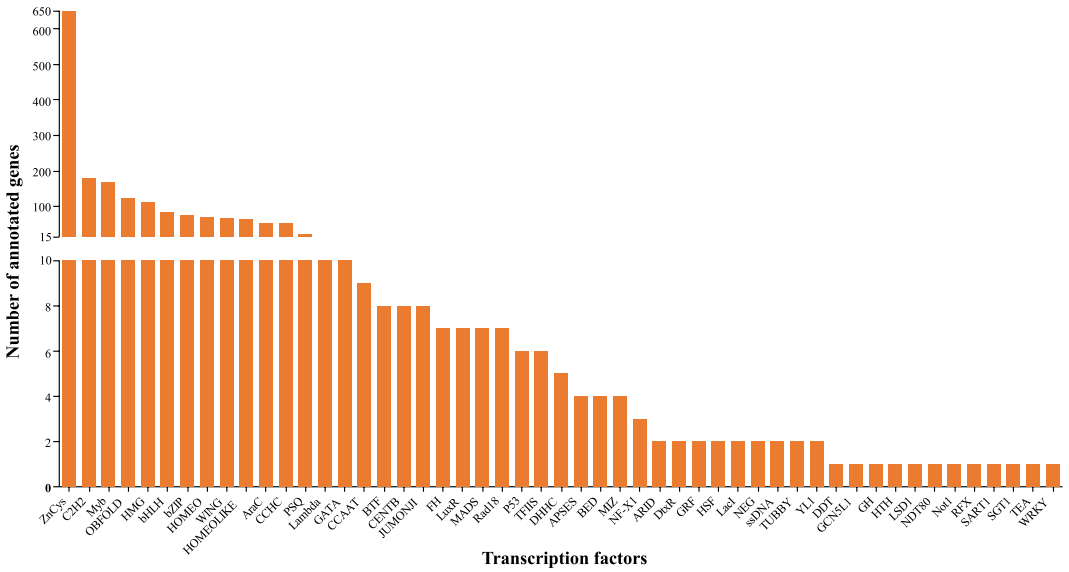

Fig. 3 Statistics of transcription factors annotated in the genome assembly of *Fusarium tricinctum* MsR-QD66.

plant root rot disease, while the functions of 4 candidate genes were still unknown, including *A11162*, *A11163*, *A13129* and *A00249*, with an identity of 71.8 to *PEP5*, 81.7 to *PEP2*, 69 to *PEP1*, 45.8 to *PEP1*, individually in PHI database (Table 5).

This study presents the first chromosome-level genome assembly of *F. tricinctum*. This high-quality genome assembly provides a valuable platform for elucidating the pathogenicity mechanisms of *F. tricinctum* to alfalfa, thereby contributing to the development of effective strategies for the management of root rot disease.

Methods

Fungal collection and genomic DNA extraction. One single-spored isolate *F. tricinctum* MsR-QD66, causal agent of alfalfa (*Medicago sativa*) root rot was conducted for this study (Fig. 6). The isolate was active-cultured on potato dextrose agar (PDA) medium from the −20 °C storage one, and then transferred into potato dextrose broth (PDB) for shanking culture at 150 rpm, 25 °C. The hyphae of MsR-QD66 harvested from the 5-day-old PDB culture set was used for its genomic DNA extraction with the CTAB method[10]. The concentration, integrity and purity of therefore gained DNA was detected by the agarose gel electrophoresis and pulsed-field gel electrophoresis, and quantified by Qubit Fluorometer (Thermo Scientific).

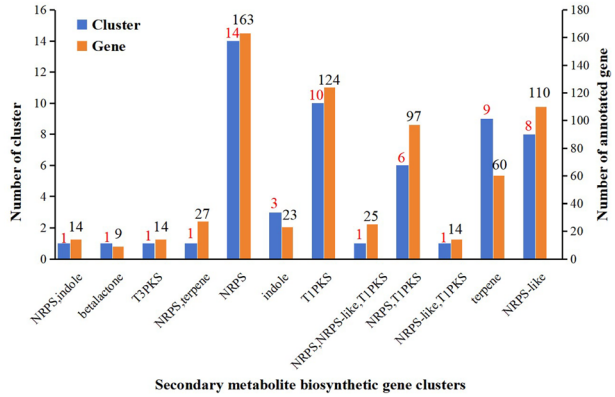

Fig. 4 Statistics of secondary metabolite biosynthetic gene clusters (BGCs) and related genes annotated in the genome assembly of *Fusarium tricinctum* MsR-QD66. The bar chart displays the distribution of different types of secondary metabolite BGCs alongside the number of associated genes. Numbers above each bar indicate the cluster count (in red) and the corresponding annotated gene count (in black).

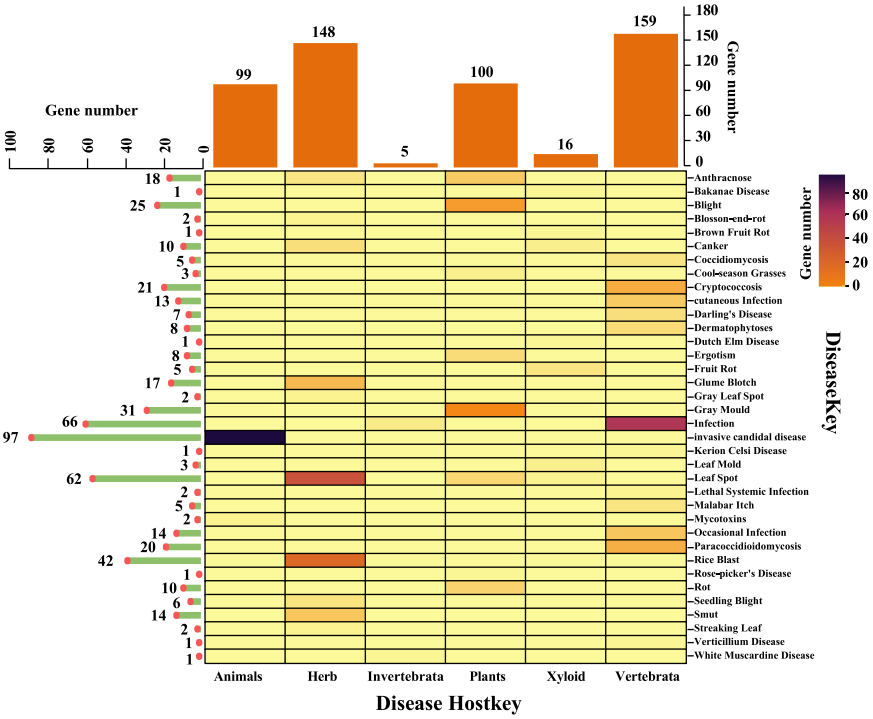

Fig. 5 Statistics of genes annotated against the Database of Fungal Virulence Factors (DFVF) in the genome assembly of *Fusarium tricinctum* MsR-QD66.

Genome survey and chromosome-level assembly. A total amount of 1 μg DNA of MsR-QD66 was used as input material, and sequencing library was generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to it. Briefly, the DNA sample was fragmented by g-TUBE (Covaris, Woburn, MA, USA) to a size of 350 bp, then DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. PCR products were purified (AMPure XP system) and libraries were analysed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. The raw data of MsR-QD66 obtained by Illumina NovaSeq PE150 sequencing was processed to clean data with Fastp v0.20.0[11] through removing the reads containing low-quality bases (mass value ≤ 5) over 40%, N bases reached 10% and adapters. The genomic survey was finally performed with GCE v1.0.0[12] K-mer analysis, to estimate genome size, heterozygous rate and repeat rate.

A total of 2 μg DNA of MsR-QD66 with a main band >30 kb was sheared into 15~18 kb fragments using g-TUBEs. The fragmented DNA was damage repaired, end repaired, adapter ligated and link-failed fragments were removed, by PacBio SMRTbell Express Template Prep Kit 3.0. Size-verified libraries (Agilent 2100

Gene id	Identity	PHI-base accession	Protein ID	Gene Name	Gene Function
A00557	48.4	PHI:179	Q04701	*PELA*	pectate lyase
A02013	71	PHI:179	Q04701	*PELA*	pectate lyase
A08011	84.1	PHI:180	Q00845	*PELD*	pectate lyase
A08366	44.6	PHI:179	Q04701	*PELA*	pectate lyase
A09831	80.7	PHI:180	Q00845	*PELD*	pectate lyase
A11156	52.3	PHI:112	Q01446	*MAK1*	maackiain detoxification
A11162	71.8	PHI:225	Q9C0P7	*PEP5*	unknown
A11163	81.7	PHI:224	Q9C1K0	*PEP2*	unknown
A12587	50	PHI:179	Q04701	*PELA*	pectate lyase
A13129	69	PHI:223	Q9C441	*PEP1*	unknown
A00249	45.8	PHI:223	Q9C441	*PEP1*	unknown

Table 5. PHI-annotated candidate virulence factors implicated in pathogenicity in genome assembly of *Fusarium tricinctum* MsR-QD66.

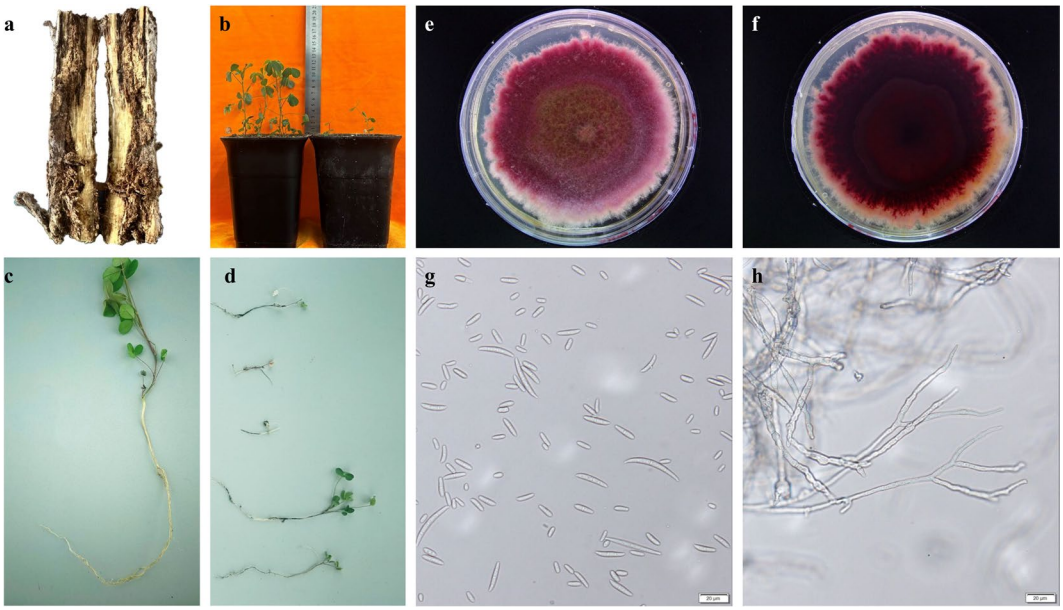

Fig. 6 Morphological characteristics and pathogenicity of *Fusarium tricinctum* MsR-QD66. (**a**) Root rot symptoms observed on alfalfa plants collected from the field. (**b–d**) Pathogenicity assay of MsR-QD66 on alfalfa plants: the left panels in (**b**) and (**c**) show control plants without MsR-QD66 inoculation while the right panels in (**b**) and (**d**) display plants inoculated with MsR-QD66. (**e, f**) Colony morphology of MsR-QD66 cultured on potato dextrose agar (PDA) medium after 7 days. (**g, h**) Microscopic features of MsR-QD66 spores and hyphae observed on carnation leaf agar (CLA) medium. Scale bar = 20 μm.

Bioanalyzer) were subjected to single-molecule real-time (SMRT) sequencing on PacBio Sequel II system (Pacific Biosciences, USA) in Circular Consensus Sequencing (CCS) mode. The obtained polymerase reads were adapter-trimmed and filtered with the minimum length threshold of 50 bp, and then processed through ccs software (https://github.com/PacificBiosciences/ccs) with parameters of 'min-passes = 3' and 'min-rq = 0.99', to generate the HiFi reads for downstream assembly. PacBio HiFi reads were assembled for the sketch genome assembly in the hifiasm v0.14.2-r315[13].

High-throughput resolution chromosome conformation capture (Hi-C) technology was facilitated for the chromosome-level assembly, by utilizing the genome-wide spatial contacts between different regions of chromatin filaments. The Hi-C library was generated according to the standard protocol and sequencing on Illumina NovaSeq 6000 platform. The gained clean reads were compared with the sketch genome to obtain the paired reads, and then quality controlled using HiCUP v0.9.1[14]. Genome clustering, ordering and orientation were performed using LACHESIS v2.0 (http://shendurelab.github.io/LACHESIS/), and chromosomal-level scaffolding was refined through iterative Juicebox v2.13.07 (https://github.com/aidenlab/Juicebox)[15–17].

Genome component prediction. Both *de novo* prediction and homologous analysis from the reference genome of *F. tricinctum* (GCA_020744515.1) and *F. acuminatum* (GCA_038181435.1) annotated resources were integrated for the genome component prediction of *F. tricinctum* MsR-QD66, including the prediction of

the protein coding genes, repetitive sequences and non-coding RNAs(ncRNAs). For the protein coding gene, Augustus v2.7[18] was employed for *de novo* prediction, while the homologous sequences was firstly aligned by TblastN v2.11[19] set at E-value ≤ 1e-5 and then conducted for predicting gene structure with Exonerate v2.2 and Liftoff v1.6.3 in homology-based prediction[20]. The gene models from both sets were consorted using EVidenceModeler v1.1.1 with more evidence weight in Homology-set[21]. In terms of repetitive sequences, the repeat library was firstly constructed by RepeatModeler v1.0.11 (parameters: -database -engine ncbi -pa)[22], and the repetitive sequences were extracted and softmasked by RepeatMasker v4.1.2.p1 ((http://www.repeatmasker.org). Then LTR_FINDER v1.2[23] was deployed for *de novo* searching of long terminal repeat-retrotransposons (LTR-RTs), Tandem Repeat Finder v4.07b[24] was conducted for finding tandem repeats, and the other repeat elements were determined by RepeatMasker in *de novo* prediction. For homologous analysis, transposable elements (TEs) were detected by RepeatMasker and then searching against TE protein database by RepeatProteinMask v4.1.2. The repeat sequences were final gained by integrating the results of both methods. For ncRNAs, transfer RNA (tRNA) genes and ribosome RNA (rRNA) were respectively analyzed by tRNAScan-SE v2.0.12[25] and rRNAmmer v1.2[26], while sRNA, snRNA and miRNA were predicted by BLAST against the Rfam v14.8[27].

Functional annotation and analysis of protein coding genes. To comprehensively decode the functional and metabolic landscape of the genome, we implemented a multi-tiered annotation pipeline. Predicted protein sequences were subjected to homology-based annotation against seven databases: the NR Database (with an E-value cutoff of < 1e-5)[28], GO[29], KEGG[30], KOG[31], TCDB[32], cytochrome P450 family database (CYP450)[33], and Swiss-Prot (UniProtKB)[34]. Protein domains were further annotated via hmmscan (HMMER3.3.2) against Pfam v35.0[35]. Transcription factors (TFs) were systematically identified by Fungal Transcription Factor Database[36]. Secretory proteins were predicted using SignalP v4.1 and filtered for transmembrane domains via TMHMM v2.0c[37]. Secondary metabolite biosynthetic gene clusters (BGCs) were delineated using antiSMASH v7.0[38]. Pathogenicity determinants were annotated by cross-referencing PHI-base v4.12[39] and DFVF v2.2 databases[40]. Carbohydrate-active enzymes (CAZy) were classified through dbCAN2 meta-server predictions against the CAZy database[41].

Data Records

The genome assembly and associated multi-omics datasets of *F. tricinctum* strain MsR-QD66 have been deposited in public repositories to ensure accessibility and reproducibility. The Illumina, PacBio HiFi, and Hi-C sequencing data are archived in NCBI under BioProject number PRJNA1264057, with SRA accession number SRP586064[42]. The final gap-free genome assembly is deposited in the GenBank databases under accession GCA_050859235.1[43] and Genome Warehouse of National Genomics Data Center (https://ngdc.cncb.ac.cn/) at China National Center for Bioinformation with the accession number of GWHGDJR00000000[44]. The annotation resources of genome component prediction and functional annotation of protein coding genes are publicly available on Figshare[45].

Technical Validation

The genome assembly of *F. tricinctum* MsR-QD66 underwent rigorous multidimensional validation to ensure accuracy and biological fidelity. MerQury v1.3[46] quantified assembly precision through K-mer frequency analysis of Illumina sequencing data, achieving an exceptional quality value (QV) of 34.7 ($QV = -10 \times \log E$), corresponding to an error rate of <0.01%, thereby confirming near-perfect base-level accuracy. Evolutionary conservation was assessed via CEGMA v2.5[47], with 95.6% of 248 core eukaryotic genes (CEGs) fully represented and 98% either complete or partially retained, validating structural completeness across conserved genomic regions. Assembly integrity was further corroborated by a 96.1% NGS mapping ratio using BWA v0.7.17-r1188 (http://bio-bwa.sourceforge.net/) aligned short-insert reads[48], demonstrating negligible structural anomalies of 96.1%. BUSCO[49] analysis targeting 290 fungal-specific single-copy orthologs revealed flawless completeness (C:100%, S:290; D/F/M:0), confirming the absence of redundancy, fragmentation, or missing loci (Table 2). These metrics spanning nucleotide-level precision, evolutionary conservation, read alignment fidelity, and ortholog completeness, collectively establish this assembly as a gold-standard genomic resource, meeting the exacting reproducibility criteria required for high-impact research and downstream comparative analyses.

Data availability

The genome assembly of *F. tricinctum* strain MsR-QD66 has been deposited in both NCBI GenBank at https://identifiers.org/ncbi/insdc.gca:GCA_050859235.1, and China National Center for Bioinformation at https://ngdc.cncb.ac.cn/gwh/Assembly/95130/show. The Illumina, PacBio HiFi, and Hi-C sequencing data that support the genome assembly are respectively available in NCBI with run number as SRR33612022 (Illumina sequencing data - SRA - NCBI), SRR33612568 (PacBio HiFi data of Fusarium tricinctum MsR-QD66 - SRA - NCBI), SRR33613311 (HIC data of Fusarium tricinctum MsR-QD66 - SRA - NCBI), under BioProject number PRJNA1264057 (https://www.ncbi.nlm.nih.gov/sra/SRP586064). The genome annotation files have been deposited in Figshare at https://figshare.com/projects/The_annotated_files_for_Fusarium_tricinctum_strain_MsR-QD66/253634. From which, the annotation information for genome component is deposited under Genome_Component (https://figshare.com/articles/dataset/Genome_Component/29375291). The other folders, including AnnoSummary (https://figshare.com/articles/dataset/AnnoSummary/29375333?file=55543073), General_Gene_Annotation (https://figshare.com/articles/dataset/General_Gene_Annotation/29375309), Effector (https://figshare.com/articles/dataset/Effector/29375303), Pathogenicity (https://figshare.com/articles/dataset/Pathogenicity/29375318) are responsible for the annotation results described in the "Functional annotation and analysis of protein coding genes" section of this study.

Code availability

All computational pipelines and command-line operations utilized in this study were executed according to the protocols and manuals of the respective bioinformatic software and performed in the Methods section. If no specified parameters were mentioned for the software, default parameters were employed as recommended by the software developers.

Received: 24 June 2025; Accepted: 10 September 2025;
Published online: 23 October 2025

References

1. Wang, Y., Wang, R. & Sha, Y. Distribution, pathogenicity and disease control of *Fusarium tricinctum*. *Front. Microbiol.* **13**, 2022, https://doi.org/10.3389/fmicb.2022.939927 (2022).
2. Jiang, D. *et al*. Identification of fungal pathogens and analysis of genetic diversity of *Fusarium tricinctum* causing root rots of alfalfa in north-east China. *Plant Pathol.* **70**, 804–814, https://doi.org/10.1111/ppa.13333 (2021).
3. Li, X. P. *et al*. Pathogens identification and pathogenicity analysis of Fusarium root rot on naked barley. *J. Triticeae Crops* **42**, 1149–1161, https://doi.org/10.7606/j.issn.1009-1041.2022.09.12 (2022).
4. Fuentes-Pardo, A. P. & Ruzzante, D. E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* **26**, 5369–5406, https://doi.org/10.1111/mec.14264 (2017).
5. Moolhuijzen, P. M., Manners, J. M., Wilcox, S. A., Bellgard, M. I. & Gardiner, D. M. Genome sequences of six wheat-infecting *Fusarium* species isolates. *Genome Announc.* **1**, e00670-13, https://doi.org/10.1128/genomeA.00670-13 (2013).
6. Fayyaz, G., Robinson, G., Chang, P. L. & Cook, D. R. Hiding in plain sight: Genome-wide recombination and a dynamic accessory genome drive diversity in *Fusarium oxysporum* f.sp. *Ciceris*. *Proc. Natl. Acad. Sci. USA* **120**, e2220570120, https://doi.org/10.1073/pnas.2220570120 (2023).
7. Yao, G. *et al*. Gapless genome assembly of *Fusarium verticillioides*, a filamentous fungus threatening plant and human health. *Sci. Data* **10**, 229, https://doi.org/10.1038/s41597-023-02145-8 (2023).
8. Shang, M. A chromosome-level genome assembly of *Fusarium foetens* that causes sweet potato root rot facilitates the identification of a key Fusarium-specific pathogenicity factor. *Plant commun*. 101421, https://doi.org/10.1016/j.xplc.2025.101421 (2025).
9. https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=61284.
10. Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S. & Thompson, W. F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1**, 2320–2325, https://doi.org/10.1038/nprot.2006.384 (2006).
11. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
12. Liu, B. *et al*. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv.org arXiv:1308.2012*, https://doi.org/10.48550/arXiv.1308.2012 (2013)
13. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
14. Wingett, S. *et al*. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310, https://doi.org/10.12688/f1000research.7334.1 (2015).
15. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845, https://doi.org/10.1038/s41477-019-0487-8 (2019).
16. Burton, J. N. *et al*. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol* **31**, 1119–1125, https://doi.org/10.1038/nbt.2727 (2013).
17. Rao, S. S. *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680, https://doi.org/10.1016/j.cell.2014.11.021 (2014).
18. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637–644, https://doi.org/10.1093/bioinformatics/btn013 (2008).
19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2 (1990).
20. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**, 189, https://doi.org/10.1186/s12859-018-2203-5 (2018).
21. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
22. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
23. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
24. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).
25. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
26. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108, https://doi.org/10.1093/nar/gkm160 (2007).
27. Kalvari, I. *et al*. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–d200, https://doi.org/10.1093/nar/gkaa1047 (2021).
28. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**, W20–W25 (2004).
29. Ashburner, M. *et al*. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29, https://doi.org/10.1038/75556 (2000).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, https://doi.org/10.1093/nar/28.1.27 (2000).
31. Koonin, E. V. *et al*. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7, https://doi.org/10.1186/gb-2004-5-2-r7 (2004).
32. Saier, M. H. Jr. *et al*. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.* **44**, D372–379, https://doi.org/10.1093/nar/gkv1103 (2016).
33. Lisitsa, A. V., Gusev, S. A., Karuzina, I. I., Archakov, A. I. & Koymans, L. Cytochrome P450 Database. *SAR QSAR Environ. Res.* **12**, 359–366, https://doi.org/10.1080/10629360108033244 (2001).
34. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol. Biol.* **406**, 89–112, https://doi.org/10.1007/978-1-59745-535-0_4 (2007).
35. Mistry, J. *et al*. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419, https://doi.org/10.1093/nar/gkaa913 (2021).

36. Park, J. *et al*. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* **24**, 1024–1025, https://doi.org/10.1093/bioinformatics/btn058 (2008).
37. Zeng, R., Gao, S., Xu, L., Liu, X. & Dai, F. Prediction of pathogenesis-related secreted proteins from *Stemphylium lycopersici*. *BMC Microbio.* **18**, 191, https://doi.org/10.1186/s12866-018-1329-y (2018).
38. Blin, K. *et al*. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50, https://doi.org/10.1093/nar/gkad344 (2023).
39. Urban, M. *et al*. PHI-base: the pathogen-host interactions database. *Nucleic Acids Res.* **48**, D613–D620, https://doi.org/10.1093/nar/gkz904 (2020).
40. Lu, T., Yao, B. & Zhang, C. DFVF: database of fungal virulence factors. *Database (Oxford)* **2012**, bas032, https://doi.org/10.1093/database/bas032 (2012).
41. Cantarel, B. L. *et al*. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–238, https://doi.org/10.1093/nar/gkn663 (2009).
42. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP586064 (2025).
43. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_050859235.1 (2025).
44. *NGDC Genome Warehouse* https://ngdc.cncb.ac.cn/gwh/Assembly/95130/show (2025).
45. Xu, S. The annotated files for *Fusarium tricinctum* strain MsR-QD66. *Figshare* https://figshare.com/projects/The_annotated_files_for_Fusarium_tricinctum_strain_MsR-QD66/253634 (2025).
46. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
47. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067, https://doi.org/10.1093/bioinformatics/btm071 (2007).
48. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
49. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245, https://doi.org/10.1007/978-1-4939-9173-0_14 (2019).

Acknowledgements

This project was supported by the National Key Research and Development Program of China (2022YFD1300802), the National Natural Science Foundation of China (32171678), and the Fundamental Research Funds for the Central Universities (lzujbky-2024-it80). We thank Beijing Novogene Bioinformatics Technology Co., Ltd. for their assistance with genome sequencing and analysis.

Author contributions

X.F. conceived and supervised this study. S.X. conducted the experiments and wrote the manuscript; S.X. and X.S. conducted the bioinformatic analysis. X.L. and L.M. revised the manuscript. All authors read and prove the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.F.