scientific data

OPEN

DATA DESCRIPTOR

Chromosome-level genome assemblies of the pink snow mold pathogens *Microdochium majus* and *Microdochium nivale*

Meixin Yang[1,2], Micong Xu[2], Wanquan Chen[2], Hao Zhang[2,3] ✉, Xu Cheng[1] ✉ & Taiguo Liu[2,3] ✉

Wheat Snow Mold (WSM), caused by *Microdochium* species, poses a serious threat to global wheat production. Despite its importance, the genetic and molecular mechanisms of *Microdochium* remain poorly understood. In particular, genome-wide differences between *M. majus* and *M. nivale*, which were previously considered a single species, have not been fully elucidated. Here, we present the first high-quality telomere-to-telomere genome assemblies of *M. majus* (231095) and *M. nivale* (231047), based on Nanopore and Illumina sequencing, with genome sizes of 36.50 Mb and 37.27 Mb. Each assembly was anchored to 13 chromosomes and one circular mitochondrial genome. We identified 11,432 and 11,904 protein-coding genes, with BUSCO completeness scores of 98.5% and 99.3%; of these, 11,094 and 11,504 genes were functionally annotated. Comparative genomics revealed a high degree of collinearity between the two strains, along with segment relocations and gene presence/absence variations. This study enhances our understanding of the genetic foundations of *M. majus* and *M. nivale*, laying the groundwork for future research on genetic evolution and disease management.

Background & Summary

Wheat is one of the most important cereal crops worldwide and constitutes a cornerstone of global food security. Following its domestication in the Fertile Crescent approximately 10,000 years ago, wheat cultivation has fundamentally shaped agricultural systems and driven the development of human societies[1]. The early domestication of wheat contributed to the transition from nomadic subsistence strategies to settled agricultural systems, providing a basis for the development of complex societies[2]. Currently, wheat is cultivated on a global scale and constitutes a major contributor to human dietary energy across diverse populations.

A diversity of phytopathological agents continues to represent major challenges to wheat production. WSM was first discovered in 1825 by Swedish mycologist E.M. Fries[3]. This disease occurs in cold, humid, high-latitude regions and is a devastating condition caused by a complex of soil-borne fungi and oomycetes, among which *M. nivale* and *M. majus* are the dominant pathogens[4]. WSM commonly occurs in the Northern Hemisphere and is prevalent in regions such as North America, Europe, and Russia[4–9]. In China, it is widely distributed in the northwest and southwest regions, as well as in the middle and lower reaches of the Yellow River and Yangtze River[10–15]. Since its first detection in Latvia in 2005, the average incidence rate reached 40% within three years[16]. In 2019, it affected up to 850,000 hectares in Russia[17]. In Xinjiang, China, the disease has caused yield losses exceeding 30% in severely affected fields, with more than 15% of fields rendered unsuitable for cultivation[14]. The pathogens of WSM can infect a variety of hosts, including wheat, barley, oats, rye, and triticale, with wheat being the most severely affected[18–20]. Although *Microdochium* species responsible for WSM do not produce secondary metabolites toxic to mammals or humans, they can infect wheat plants at all growth stages, from germination to maturity, leading to symptoms such as seedling blight, sheath rot,

[1]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518120, P. R. China. [2]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, P. R. China. [3]National Agricultural Experimental Station for Plant Protection, Gangu, Ministry of Agriculture and Rural Affairs, Tianshui, 741200, P. R. China. ✉e-mail: zhanghao@caas.cn; chengxu@caas.cn; liutaiguo@caas.cn

Sample	Read Bases (bp)	Number of Reads	Mean Length (bp)	N50 Length (bp)
231095	14,111,087,441	1,369,253	10,305.7	23,119
231047	10,295,646,642	666,100	15,456.6	23,286

Table 1. Summary of Nanopore sequencing data for strains 231095 and 231047.

Sample	Stat	Bases(bp)	Reads	Q20(%)	Q30(%)	GC(%)
231095	Before filter	6.46 G	43.07 M	96.92	91.49	49.56
	After filter	6.34 G	42.47 M	97.39	92.30	49.47
231047	Before filter	9.43 G	62.87 M	97.41	92.43	53.91
	After filter	9.25 G	62.27 M	97.74	93.01	53.89

Table 2. Summary of Illumina sequencing data for strains 231095 and 231047.

leaf blight, and head blight[14]. Among these, leaf blight and sheath rot are the most significant. When wheat is infected and leaf blight develops, mature leaves display characteristic large oval or semicircular lesions that appear dark green and water-soaked or dark brown with concentric rings, often covered with sparse moldy growth on the surface. Initially, lesions on the leaves appear water-soaked and gradually expand into nearly round or oval blotches; at the leaf margins, they appear semicircular. The pathogen's conidiophores emerge from the stomata, producing abundant conidia, resulting in brick-red mold-like growth covering the lesion surface. Under high humidity, spores at lesion edges germinate to form a thin layer of white mycelium. In the late stages of infection, affected leaves typically die off entirely, and small black dots (perithecia) appear scattered in the center of necrotic lesions[21,22]. The pathogen can survive in soil for several years, making snow mold leaf blight particularly difficult to control.

Due to the difficulty in distinguishing the hyphal morphology of *M. majus* and *M. nivale*, they were initially regarded as two varieties of the same species—*Microdochium* var. *majus* and *Microdochium* var. *nivale*[23]. With the advancement of molecular biology, Lees *et al.*[24] used molecular genotyping methods to identify them as two distinct species[24]. However, until now, chromosome-level genome assemblies for *M. majus* and *M. nivale* have not been reported. To address these limitations and deepen our understanding of WSM pathogen biology, we employed both long-read and short-read sequencing approaches to sequence the whole genomes of strains 231095 (*M. majus*) and 231047 (*M. nivale*). We evaluated the genome assemblies and repetitive sequences, annotated genes using multiple databases, predicted effectors, and compared the whole genomes of *M. majus* and *M. nivale*. This study provides insights into the genome structure and functional characteristics of *M. majus* and *M. nivale*, enhancing our understanding of their biology and pathogenic mechanisms.

Methods

Sample preparation and DNA extraction. Strains 231095 (*M. majus*) and 231047 (*M. nivale*), stored at −80°C, were cultured on potato dextrose agar (PDA) plates at 20°C for 5 days. Three mycelial plugs were taken from the peripheral region of the colonies using a sterile punch and placed into yeast extract peptone dextrose (YEPD) liquid medium. The cultures were incubated with shaking at 20°C for 7 days. Mycelia were collected by moving the cultures into sterile 50 mL centrifuge tubes and centrifuging at 4000 x g for 5–7 minutes. The pellet was washed twice with sterile water to obtain fresh mycelia. Collected mycelia were put into 2 mL centrifuge tubes, freeze-dried under vacuum for 24 hours, and then mixed with glass beads and an appropriate amount of quartz sand. The tubes were frozen in liquid nitrogen with a grinding module and then immediately homogenized for 90 seconds using a MiniBeadbeater-96 homogenizer. Genomic DNA was extracted using the OMEGA fungal DNA extraction kit (D3390-02) following the manufacturer's instructions.

Whole genome sequencing and quality control. Whole-genome sequencing of 231095 and 231047 was performed by Sinobiocore Biotechnology Co., Ltd. (Beijing, China) using both Oxford Nanopore (Oxford Nanopore Technologies, Oxford, UK) and Illumina NovaSeq 6000 (Illumina, CA, USA) platforms. A sequencing library was constructed following the protocol of the Nanopore Native Barcoding Kit (SQK-NBD114.24). Long-read sequencing was performed using the PromethION platform employing R10.4.1 flow cells. Nanopore reads were quality filtered and trimmed using NanoFilt v2.8.0[25] from the NanoPack toolkit. In parallel, Illumina raw reads were processed with fastp v0.23.4[26] to remove adapter sequences and low-quality bases. These preprocessing steps ensured high-quality input data for subsequent genome assembly.

After filtering, a total of 14.11 Gb and 10.30 Gb of Nanopore long-read data (Table 1), and 6.34 Gb and 9.25 Gb of Illumina short-read data (Table 2) were generated for strains 231095 and 231047, respectively. Furthermore, 92.30% and 93.01% of Illumina bases for strains 231095 and 231047, respectively, had Phred scores ≥ Q30.

Genome assembly of the telomere-to-telomere (T2T) genomes. The genome sizes of strains 231095 and 231047 were estimated using clean Illumina sequencing data and the GenomeScope tool[27], based on k-mer frequency distributions generated by Jellyfish v2.3.0[28] (k-mer size = 17). Independent primary genome assemblies of strains 231095 and 231047 were generated using NextDenovo v2.5.2[29] with the

	231095	231047
Genome Size (bp)	36,504,744	37,267,004
Number of Chromosome	13	13
Mitochondrial genome	Present (circular)	Present (circular)
GC Content (%)	55.48	55.10
Repeat content (%)	4.73	4.85
N50 Length (bp)	3,418,793	3,509,414
N90 Length (bp)	1,742,796	1,501,069
L50 number	5	5
L90 number	11	11
Gaps in chromosomes	0	0

Table 3. Summary of the T2T genome assemblies of strains 231095 and 231047.

parameters 'read_type = ont, read_cutoff = 15k, genome_size = 38000000', and NECAT (v0.0.1, accessed on 3 August 2020)[30] with the parameters 'GENOME_SIZE = 38000000, MIN_READ_LENGTH = 15000'. The initial assembly generated by NECAT was used as the backbone genome. The assembly produced by NextDenovo was then employed to identify and supplement the missing telomeric sequences. Telomeric regions were identified in the assembled genomes using trf v. 4.09.1[31] with the parameters '2 7 7 80 10 90 2000 -d -m -l 2', revealing canonical telomeric repeat units $(TTAGGG/CCCTAA)_n$ at both termini of the chromosomes. Finally, T2T genome assemblies were generated. Two rounds of polishing were performed on the assemblies using Pilon v1.24[32] with default parameters, incorporating corrected ONT long reads and cleaned short reads to produce the final assemblies. Mummer v4.0.0[33] was used to construct nucleotide alignments between the assembled genomes of 231095 and 231047. JCVI v1.3.9[34], the Python implementation of MCScan, was used to identify synteny blocks between the gene sets of the two strains. Orthologous gene pairs were identified using the jcvi.compara.catalog ortholog module. The depth of synteny blocks was then calculated with jcvi.compara.synteny depth to assess collinearity. Redundant or low-confidence anchors were filtered using jcvi.compara.synteny screen with the–minspan = 1 and–simple options. Finally, the synteny relationships were visualized using jcvi.graphics.karyotype.

The final assemblies of the 231095 and 231047 are summarized in Table 3. The assembled genome of 231095 is 36,504,744 bp with a GC content of 55.48%, and 4.73% of the genome consists of repetitive elements. For strain 231047, the genome is 37,267,004 bp with 55.10% GC content, and 4.85% repetitive content (Table 3). Both strains were assembled at the chromosome level with telomeric completeness (Fig. 1b,c), each consisting of 13 nuclear chromosomes and one circular mitochondrial genome. The nucleotide-based alignment between strain 231095 and 231047 is shown in Fig. 1a. In both strains, the rDNA sequences were assembled on Chromosome 5, and the telomeric sequences were located distal to the rDNA regions (Fig. 1b,c).

Genome structural annotation. Before predicting gene structures and annotating the genome, repetitive elements within the assemblies were first detected and masked. RepeatModeler v2.0.1 and RepeatMasker v4.1.1[35] were used to perform de novo repeat annotation on the *M. nivale* and *M. majus* genomes.

The annotation of protein-coding genes was performed by integrating ab initio predictions, transcriptome evidence, and protein predictions using EVidenceModeler v2.1.0[36]. We used four tools for ab initio gene predictions, including BRAKER3 v3.0.3[37,38], GeneMark-ES v4.71_lic[39], GlimmerHMM v3.0.4[40] and SNAP v2017_03_01[41]. BRAKER pipeline B was employed to train GeneMark-ET and to predict genes using AUGUSTUS, with both steps utilizing RNA-Seq spliced alignment information as supporting evidence. GeneMark-ES was utilized for ab initio gene prediction, operating in fungi mode with default parameters. GlimmerHMM and SNAP were also used for ab initio gene prediction with default parameters. For transcriptome-based gene prediction, three publicly available RNA-seq datasets derived from *M. nivale* strain F00608 and were retrieved from the NCBI Sequence Read Archive (SRR21231516, SRR21231517 and SRR21231518)[42]. The paired-end RNA-seq reads were aligned to the soft-masked genome assemblies of each strain using HISAT2 v2.2.1[43] with the–dta parameter and a maximum intron length of 5000 bp. Genome indices were first built using hisat2-build. The resulting SAM files were converted to sorted BAM format using SAMtools v1.9[44]. Subsequently, transcript assemblies were constructed from the sorted BAM files using StringTie v2.2.1[45] with default parameters. TransDecoder v5.7.1[46] was used to identify candidate coding regions from the assembled transcript sequences. To predict gene structure based on proteins, sequences from the Sordariomycetes OrthoDB reference were mapped to the soft-masked genome using Miniprot v0.12-r237[47]. In total, we identified 11,432 protein-coding genes for 231095 and 11,904 for 231047, as shown in Table 4.

Genome functional annotation. Functional annotation was also performed on the nucleotide sequences of both strains using InterProScan v5.56-89.0[48], which integrates multiple annotation systems including PFAM, InterPro, Gene Ontology (GO), and SignalP. Next, we utilized EffectorP v3.0.0[49] to enhance effector prediction with an EffectorP score (effector probability) greater than 0.5. Subsequently, proteins were identified as effectors if they simultaneously met the following three criteria: predicted to contain a signal peptide by SignalP, to

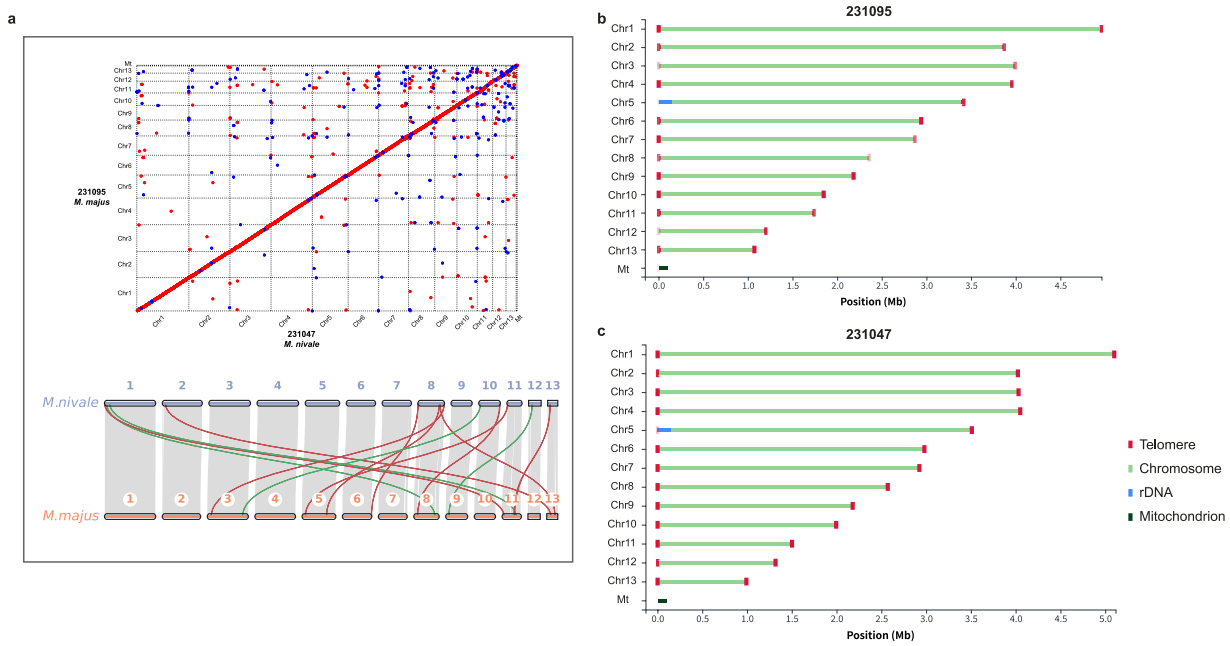

Fig. 1 Comparative genomic synteny and chromosomal organization of strains 231095 and 231047. (**a**) Nucleotide-based alignment with Nucmer between the strain 231095 and strain 231047. (**b,c**) Chromosomal organization of 231095 and 231047. Green bars indicate the chromosomal sequences, red bars mark the positions of telomeres (darker colors indicate a higher number of TTAGGG/CCCTAA repeats, whereas lighter colors indicate fewer repeats), blue bars show the rDNA on Chr5, and dark green bar represents the mitochondrial genome.

		231095	231047
Protein-coding gene number		11,432	11,904
BUSCO	Complete (%)	98.5	99.3
	Single-copy (%)	98.3	99.1
	Duplicated (%)	0.2	0.2
	Fragmented (%)	0.6	0.2
	Missing (%)	0.9	0.5
Functional gene number	PFAM	7,963	8,075
	InterPro	8,705	8,843
	GO	6,265	6,358
	SignalP with transmembrane	29	27
	SignalP without transmembrane	1,112	1,173
	EffectorP	2,696	2,872
	BGCs	28	31

Table 4. Summary of the T2T genome annotations of strains 231095 and 231047.

lack transmembrane domains by TMHMM, and to be predicted as secreted effectors by EffectorP. AntiSMASH v6.1.1[50] was performed to analyze the whole genome sequences to identify biosynthetic gene clusters (BGCs).

Functional annotation revealed that at least 97.04% (11,094) and 96.64% (11,504) of gene models in strains 231095 and 231047, respectively, were annotated in at least one functional database. A total of 28 and 31 BGCs were predicted and identified in strains 231095 and 231047, respectively (Table 4). This underscores the reliability of the gene prediction results and offers a sturdy foundation for future genomic research on these species.

Data overview. Genomic features, GC content was calculated using bedtools v2.30.0[51] nuc with a sliding window approach applied to the soft-masked genome assemblies of strains 231095 and 231047. In parallel, gene density was estimated using bedtools intersect with the -c option, which counts the number of gene annotations overlapping each genomic window based on the GFF3 files. Subsequently, custom Python scripts were employed to calculate the number of BGCs and effector gene density within each genomic window for both strains. Finally, these results were visualized using Circos v0.69-8[52] (Fig. 2).

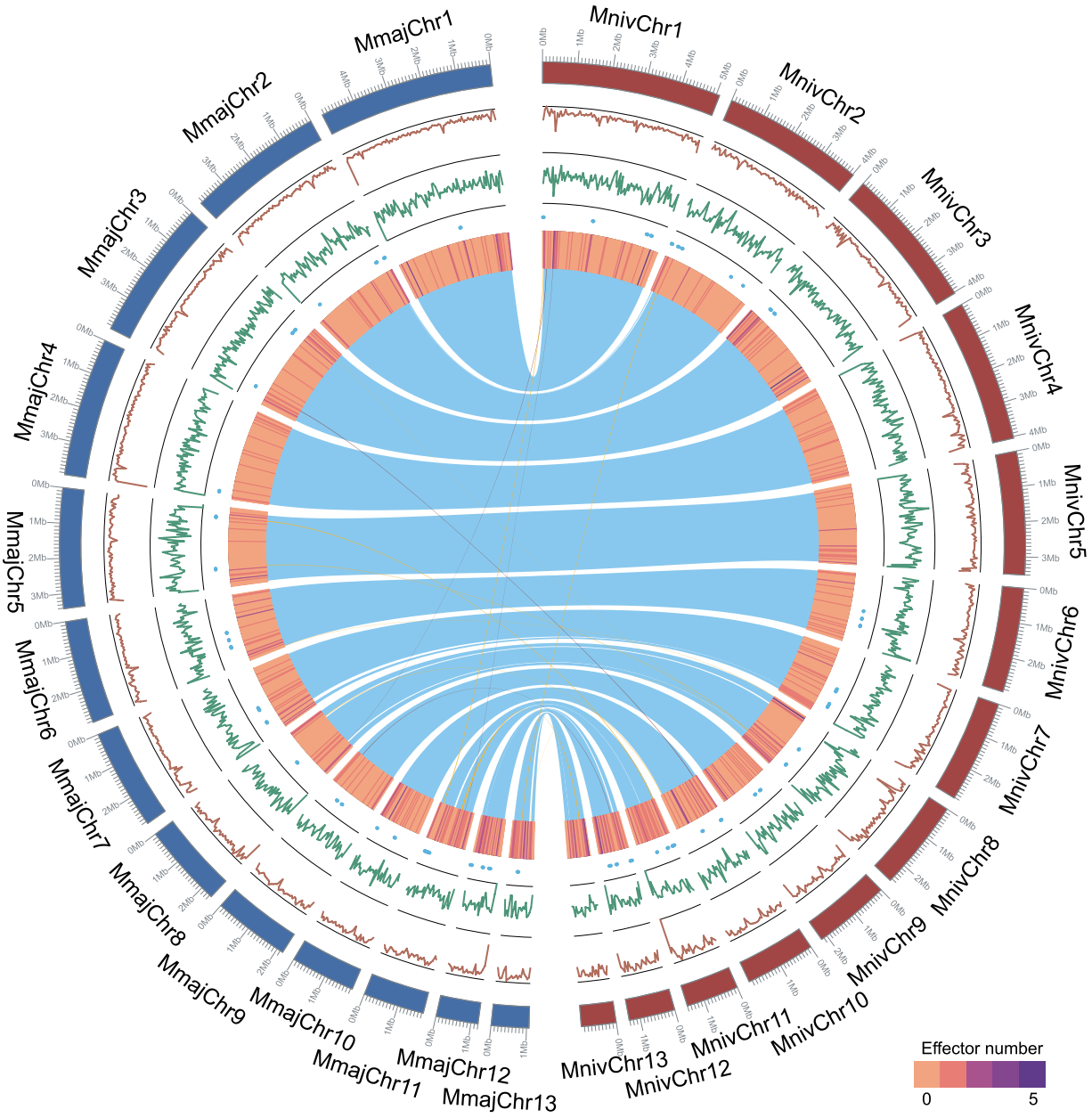

Fig. 2 Genome features of the nuclear genomes of strains 231095 and 231047. In the circular map, Track a on the outer ring illustrates the assembled T2T genomes of 231095 (left, blue) and 231047 (right, red). Track b-f represent the GC ratio, the protein-coding genes density, BGCs positions, the effector heatmap, and collinearity regions between 231095 and 231047 chromosomes, respectively. A 50 kb sliding window was used to calculate the values of the parameters shown in Tracks b-e.

Data Records

All data generated in this study are available at the ENA and NCBI databases. Raw sequencing reads for strains 231095 and 231047 were generated using both Illumina short-read sequencing and Oxford Nanopore long-read sequencing platforms. The raw reads are available in FASTQ format in the NCBI SRA under BioProject accession PRJNA1298527 and SRA accession SRP604710[53]. Individual sequencing libraries are accessible under the following run accessions: Nanopore long reads (SRR34788127 and SRR34788126) and paired-end Illumina reads (SRR34788129 and SRR34788128). The genome assemblies and annotations were originally submitted to the ENA and are available through the INSDC databases under accession numbers GCA_977970195.1 (strain 231095)[54] and GCA_977970205.1 (231047)[55]. Each assembly record includes: (i) a genome assembly file in FASTA format; (ii) a genome annotation file in EMBL format.

Technical Validation

Quality control of DNA extraction. DNA concentration and purity were assessed using a NanoDrop spectrophotometer. Purity was evaluated by measuring the 260/280 nm ratio (1.8-1.9) and 260/230 nm ratio (2.0-2.2), ensuring DNA quality was suitable for downstream library preparation and sequencing.

Sequencing quality control. Quality control of Illumina raw reads was performed using fastp. Adapter sequences were automatically detected and trimmed. Low-quality bases at both ends of the reads were removed using sliding window trimming with a default quality threshold (Phred score < 20). Reads containing more than 5% ambiguous bases (N) or more than 40% low-quality bases (Phred score < 15) were discarded. Reads shorter than 15 bp after trimming were also removed. For Nanopore sequencing data, low-quality reads with an average Phred score below Q7 were filtered out. These stringent quality control steps ensured the accuracy and reliability of the input data, providing a solid foundation for high-quality genome assembly.

Quality assessments of the assemblies. Initial assemblies were generated using Nanopore sequencing data and subsequently polished twice with Illumina short reads to improve base-level accuracy. Telomere completeness was subsequently validated using tapestry v1.0.1[56] (weave -t CCTAAC GTTAGG -c 100), which confirmed the presence of telomeric repeats at both ends of all 13 chromosomes in strains 231047 and 231095 (Fig. 1b,c). In strain 231047, 25 telomeres contained more than 10 repeat units, with only one telomere exhibiting a shorter array of five repeats. In strain 231095, 22 telomeres harbored more than 10 repeat units, while the remaining four telomeres contained 6, 6, 5, and 1 repeat units, respectively. Detection of rDNA regions was performed using barrnap v0.9[57] (parameters:–kingdom euk) for both strains. The rDNA loci were identified at the 5' terminus of the chromosomes in strains 231047 and 231095 (Fig. 1b,c).

The completeness of protein-coding genes in the genome assemblies was assessed with BUSCO v5.5.0[58] using the sordariomycetes_odb10 dataset and default settings. The result showed that 231095 and 231047 contained 98.5% and 99.3% complete single-copy genes, respectively, with low proportions of missing (0.9% and 0.5%) genes (Table 4). These findings emphasize the high quality of the genome assemblies and annotations.

Data availability

Raw sequencing reads generated using both Illumina and Nanopore platforms for strains 231095 and 231047 are available in the NCBI SRA under BioProject PRJNA1298527 and SRA accession SRP604710[53]. Genome assemblies and annotations are available through the INSDC databases GCA_977970195.1 (strain 231095)[54] and GCA_977970205.1 (strain 231047)[55].

Code availability

The software utilized in this study is detailed in the Methods section. Default parameters were applied if specific ones were not provided. The genome assembly pipelines and custom scripts are available at https://github.com/ymxNancy/Assembly_Microdochium_genomes.

Received: 4 August 2025; Accepted: 2 March 2026;
Published online: 10 March 2026

References

1. Callaway, E. Farming invented twice in middle east, genomes study reveals. *Nature*, https://doi.org/10.1038/nature.2016.20119 (2016).
2. Peleg, Z., Fahima, T., Korol, A. B., Abbo, S. & Saranga, Y. Genetic analysis of wheat domestication and evolution under domestication. *J. Exp. Bot.* **62**, 5051–5061 (2011).
3. Fries E. M. *Systema Orbis Vegetabilis. I. Plantae homonemeae*. (Lund, 1925).
4. Ponomareva, M. L., Gorshkov, V. Y., Ponomarev, S. N., Korzun, V. & Miedaner, T. Snow mold of winter cereals: a complex disease and a challenge for resistance breeding. *Theor. Appl. Genet.* **134**, 419–433 (2021).
5. Tanaka, F., Saito, I., Miyajima, K., Tsuchiya, S. & Tsuboki, K. Occurrence of thiophanate-methyl tolerant isolates of *Fusarium nivale* (= *Gerlachia nivalis*), a causal fungus of snow mold of winter wheat, in Japan. *Jpn. J. Phytopathol.* **49**, 565–566 (1983).
6. Hayashi, Y. *et al*. Population genetic structure of *Microdochium majus* and *Microdochium nivale* associated with Fusarium head blight of wheat in Hokkaido, Japan. *Eur. J. Plant Pathol.* **140**, 787–795 (2014).
7. Walker, A.-S., Auclair, C., Gredt, M. & Leroux, P. First occurrence of resistance to strobilurin fungicides in *Microdochium nivale* and *Microdochium majus* from french naturally infected wheat grains. *Pest Manag. Sci.* **65**, 906–915 (2009).
8. Iriki, N., Nakajima, T. & Kawakami, A. Reaction of winter wheat cultivars to artificially inoculated seed-borne pink snow mold. *Breed. Sci.* **52**, 231–233 (2002).
9. Ioos, R., Belhadj, A. & Menez, M. Occurrence and distribution of *Microdochium nivale* and *Fusarium* species isolated from barley, durum and soft wheat grains in France from 2000 to 2002. *Mycopathologia* **158**, 351–362 (2004).
10. S Hongsheng. Occurrence and control of wheat Typhula leaf blight. *Plant Protection* 3–6 (1980).
11. H Xishu, W Chengfang, C Shinan, & X Mankai. Preliminary study on the occurrence of wheat Typhula leaf blight. *Anhui Agricultural Science* 321–322 (1995).
12. H Haiming. Brief analysis of the causes of severe wheat snow mold (mildew) outbreak in the Ili region in 1999 and future prevention measures. *Rural Science & Technology* **7** (1999).
13. Guoping, W., Xiangjun, Z. & Jiang, H. Preliminary study on seed treatment with different fungicides for controlling wheat snow rot (mildew) in the winter wheat region of Tacheng. *Xinjiang Agricultural Sciences* **44**, 158–160 (2007).
14. Qiu Zhuo, S. B., Yongsheng, Z., Yilin, Z. & Wanquan, C. Analysis of causes and control strategies of wheat snow rot (mildew). *Plant Protection* **36**, 154–156 (2010).
15. Yi, L. *et al*. Analysis of the occurrence and damage of wheat snow rot (mildew) and the application of seed coating control in Xinjiang. *China Plant Protection* **40**, 72–76 (2020).
16. Bankina, B., Ruza, A., Katamadze, M., Kreita, D. & Paura, L. Snow mould development under conditions of central part of Latvia. *Rural Sustain. Res.* **27**, 1–5 (2012).

17. Glynn, N. C., Hare, M. C. & Edwards, S. G. Fungicide seed treatment efficacy against *Microdochium nivale* and *M. majus in vitro* and *in vivo*. *Pest Manag. Sci.* **64**, 793–799 (2008).
18. Simpson, D. R., Thomsett, M. A. & Nicholson, P. Competitive interactions between *Microdochium nivale* var. *majus*, *M. nivale* var. *nivale* and *Fusarium culmorum in planta* and *in vitro*. *Environ. Microbiol.* **6**, 79–87 (2004).
19. Chen, H., Xue, L. H., Li, C. J., Zhao, G. Q. & Qin, Z. M. First Report of Seedling Blight of Oat (*Avena sativa*) Caused by *Microdochium nivale* in China. *Plant Disease* **105**, 704–704 (2021).
20. Nielsen, L. K. *et al*. Fusarium head blight of cereals in denmark: species complex and related mycotoxins. *Phytopathology* **101**, 960–969 (2011).
21. Årsvoll, K. & Smith, J. D. *Typhula ishikariensis* and its varieties, var. *idahoensis* comb.nov. and var. *canadensis* var.nov. *Can. J. Bot.* **56**, 348–364 (1978).
22. Smith, J. D. Snow molds of winter cereals: guide for diagnosis, culture, and pathogenicity. *Can. J. Plant Pathol.* **3**, 15–25 (1981).
23. Samuels, G. J. & Hallett, I. C. *Microdochium stoveri* and *Monographella stoveri*, new combinations for *Fusarium stoveri* and *Micronectriella stoveri*. *Trans. Br. Mycol. Soc.* **81**, 473–483 (1983).
24. Lees, A. K., Nicholson, P., Rezanoor, H. N. & Parry, D. W. Analysis of variation within *Microdochium nivale* from wheat: Evidence for a distinct sub-group. *Mycol. Res.* **99**, 103–109 (1995).
25. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
26. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
27. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
28. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
29. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024).
30. Chen, Y. *et al*. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* **12**, 60 (2021).
31. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
32. Walker, B. J. *et al*. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **9**, e112963 (2014).
33. Marçais, G. *et al*. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
34. Tang, H. *et al*. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* **3**, e211 (2024).
35. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinformatics* **25**, (2009).
36. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
37. 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 (2008).
38. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
39. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2**, lqaa026 (2020).
40. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
41. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
42. https://trace.ncbi.nlm.nih.gov/Traces?study=SRP392084 (2022).
43. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).
44. Li, H. *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
46. Haas, B. J. TransDecoder. https://github.com/TransDecoder/TransDecoder.
47. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
48. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
49. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant Microbe Interact.* **35**, 146–156 (2022).
50. Blin, K. *et al*. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
51. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
52. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
53. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP604710 (2025).
54. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_977970195.1 (2025).
55. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_977970205.1 (2025).
56. Davey, J. W. *et al*. Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*. *G3 Genes Genomes Genet.* **11**, jkaa018 (2021).
57. Seemann, T. Barrnap 0.9: rapid ribosomal RNA prediction (2013).
58. Huang, N. & Li, H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, btad595 (2023).

Acknowledgements

This study was funded by the Key R&D Program of Xinjiang Uygur Autonomous Region (2022B02015-3) and the National Natural Science Foundation of China (32172379).

Author contributions

Hao Zhang designed the research and collected the samples; Wanquan Chen, Taiguo Liu and Xu Cheng supervised the study; Micong Xu isolated the strains and extracted the DNA; Meixin Yang analyzed the data and wrote the manuscript. All the authors were involved in revising the manuscript and approving the version that was submitted.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.Z., X.C. or T.L.