



# Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome Evolution in Pucciniales

Chuang Li,<sup>a,b</sup> Liuhui Qiao,<sup>a,b</sup> Yanan Lu,<sup>a</sup> Guozhen Xing,<sup>a</sup> Xiaodong Wang,<sup>c</sup> Gengyun Zhang,<sup>d</sup> Huimin Qian,<sup>a</sup> Yilin Shen,<sup>a</sup> Yibo Zhang,<sup>a</sup> Wen Yao,<sup>a</sup> Kun Cheng,<sup>a</sup> Zhenling Ma,<sup>a</sup> Na Liu,<sup>a</sup> Daowen Wang,<sup>b</sup>  Wenming Zheng<sup>a</sup>

<sup>a</sup>State Key Laboratory of Wheat and Maize Crop Science, College of Life Sciences, Henan Agricultural University, Zhengzhou, China

<sup>b</sup>State Key Laboratory of Wheat and Maize Crop Science, College of Agronomy and Center for Crop Genome Engineering, Henan Agricultural University, Zhengzhou, China

<sup>c</sup>State Key Laboratory of North China Crop Improvement and Regulation, College of Plant Protection, Hebei Agricultural University, Baoding, China

<sup>d</sup>BGI-Shenzhen, Shenzhen, China

**ABSTRACT** Chromosome evolution drives species evolution, speciation, and adaptive radiation. Accurate genome assembly is crucial to understanding chromosome evolution of species, such as dikaryotic fungi. Rust fungi (Pucciniales) in dikaryons represent the largest group of plant pathogens, but the evolutionary process of adaptive radiation in Pucciniales remains poorly understood. Here, we report a gapless genome for the wheat leaf rust fungus *Puccinia triticina* determined using PacBio high-fidelity (HiFi) sequencing. This gapless assembly contains two sets of chromosomes, showing that one contig represents one chromosome. Comparisons of homologous chromosomes between the phased haplotypes revealed that highly frequent small-scale sequence divergence shapes haplotypic variation. Genome analyses of *Puccinia triticina* along with other rusts revealed that recent transposable element bursts and extensive segmental gene duplications synergistically highlight the evolution of chromosome structures. Comparative analysis of chromosomes indicated that frequent chromosomal rearrangements may act as a major contributor to rapid radiation of Pucciniales. This study presents the first gapless, phased assembly for a dikaryotic rust fungus and provides insights into adaptive evolution and species radiation in Pucciniales.

**IMPORTANCE** Rust fungi (Pucciniales) are the largest group of plant pathogens. Adaptive radiation is a predominant feature in Pucciniales evolution. Chromosome evolution plays an important role in adaptive evolution. Accurate chromosome-scale assembly is required to understand the role of chromosome evolution in Pucciniales. We took advantage of HiFi sequencing to construct a gapless, phased genome for *Puccinia triticina*. Further analyses revealed that the evolution of chromosome structures in rust lineage is shaped by the combination of transposable element bursts and segmental gene duplications. Chromosome comparisons of *Puccinia triticina* and other rusts suggested that frequent chromosomal arrangements may make remarkable contributions to high species diversity of rust fungi. Our results present the first gapless genome for Pucciniales and shed light on the feature of chromosome evolution in Pucciniales.

**KEYWORDS** HiFi sequencing, gapless genome, rust fungi, chromosomal rearrangement, evolutionary radiation

Advances in long-read sequencing technologies have accelerated complete assemblies for complex genomes (1, 2). Highly identical repeats are major limitations due to discrimination errors in these genome projects. PacBio high-fidelity (HiFi) long reads with high accuracy of (>99.9%) conduce to the assembly of complex repeat regions in resolving complex genomes (3). Recently, the completion of the first telomere-to-telomere (T2T) human genome (4) and gapless rice genomes (5, 6) has benefited from HiFi

**Editor** Lindsey Price Burbank, USDA—San Joaquin Valley Agricultural Sciences Center

**Copyright** © 2023 Li et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Wenming Zheng, [wzmzheng@henau.edu.cn](mailto:wzmzheng@henau.edu.cn), Daowen Wang, [dwwang@henau.edu.cn](mailto:dwwang@henau.edu.cn), or Na Liu, [naliu@henau.edu.cn](mailto:naliu@henau.edu.cn).

The authors declare no conflict of interest.

**Received** 28 July 2022

**Accepted** 2 December 2022

**Published** 23 January 2023

sequencing. More genome projects with the goal of T2T and gapless assemblies are ongoing. Noticeably, most of animals, plants, and fungi contain more than one chromosome copy, and thus, accurate phasing of haplotypes is essential to shedding new light on genome organization and evolution. The emergence of more scaffold-level and haplotype-resolved chromosome assemblies has been greatly driven by long-read sequencing, Hi-C sequencing, and improved algorithms (7–9).

The ultimate goal of assembly is a high-quality haplotype-resolved genome without gaps. The closure of gaps in phased assemblies, especially for complex genomes, is still not trivial. Given that high accuracy of HiFi reads is in favor of haplotype phasing, the assembly software hifiasm (10), designed for haplotype-resolved *de novo* assembly using graph-binning algorithm and maintaining the contiguity of each haplotype, is a powerful tool to obtain phased assembly with high accuracy and contiguity. Other methods are also available in different projects (11). Overall, coupled with different strategies, HiFi reads are expected to help achieve gapless and phased assembly or highly contiguous chromosome sequences.

Dikaryotic fungi carry two different haploid nuclei in a cell along with the two chromosome sets contributing to adaptive evolution. Rust fungi (Pucciniales) in dikaryons are the most speciose natural group of plant pathogens, and many members of this group are the causal agents of prevalent diseases in agriculture and ecology (12, 13). In the life cycle of rust fungi, the predominant stage is generally the dikaryotic developmental phase of urediniospores that recurrently infect hosts. Rust fungi have to enhance their genetic variation to generate intraspecies diversity to overcome numerous host resistances. Comparative genomic analyses have unraveled that sexual recombination and somatic exchange are important drivers of genetic diversity (14, 15). High interhaplotype diversity has been shown to contribute to adaptive evolution (16, 17). Although intraspecies diversity in rust fungi has been reviewed extensively (12, 13), the genetic mechanism driving evolutionary radiation and interspecies diversity is poorly understood (13, 18).

Chromosome evolution plays a pivotal role during speciation and adaptive radiation. It is unclear how chromosomes have evolved to promote species radiation of Pucciniales. Recently, the combination of long-read sequencing and Hi-C data has helped achieve chromosome-level and haplotype-phased assemblies for some rust species, although too many gaps exist in the assembled chromosomes (9, 15, 19–23). *Puccinia triticina* is the most widely distributed wheat rust fungus (24) and is one of the key species used in rust fungal genomics. To dissect chromosome changes for rust lineage, our study aimed to obtain a highly contiguous, phased chromosome assembly for *P. triticina* and to investigate the role of chromosome evolution in Pucciniales through evolutionary and comparative analyses.

## RESULTS

**Gapless assembly of *Puccinia triticina* isolate.** *Puccinia triticina* isolate Pt15 was from field samples in central China in 2015. Its virulence profiles were tested using a wheat differential set (see Table S1 in the supplemental material). Whole-genome sequencing of Pt15 yielded 19.83-Gb HiFi reads with a mean length of 14.6 kb and 29.42-Gb Illumina reads (Table S2). hifiasm (10) was used for initial phased assembly with  $\geq 10$ -kb HiFi reads. Read correction was conducted through two rounds of hifiasm correction in this step. Then, the redundant sequences from heterozygous regions and mitochondria for each haplotype were identified and filtered out, producing two highly contiguous haplotype assemblies (Table S3). Using all long reads, the HERA pipeline (25) was used to connect the remaining contigs of the two haplotypes. The resulting assembly contained 18 contigs totaling 122.82 Mb for haplotype A and 18 contigs totaling 121.11 Mb for haplotype B (Table 1). Given the haploid chromosome number of *P. triticina* (26) and genome synteny between the two haplotype assemblies, our phased assembly was thought to represent 36 pseudochromosomes in dikaryons. The numbers of chromosomes were assigned according to the length in haplotype A, from the longest, as chromosome 1 (chr1A), to

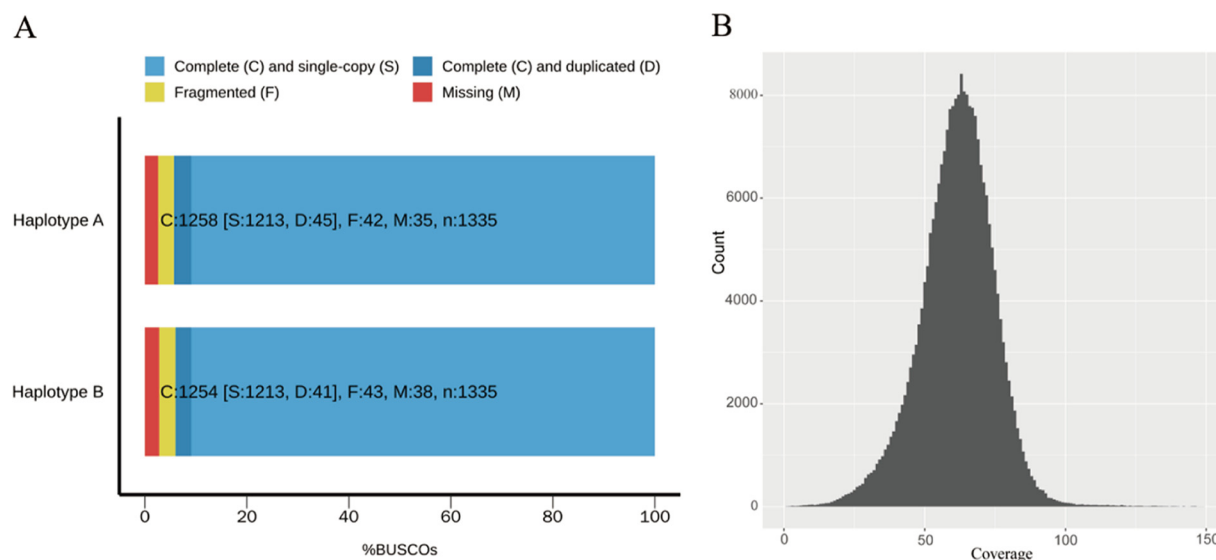
**TABLE 1** Summary of the Pt15 genome assembly and annotation

Parameter	Value for:	
	Haplotype A	Haplotype B
Total length (bp)	122,823,596	121,114,073
Contig no.	18	18
Contig $N_{50}$ (bp)	7,485,535	7,574,744
GC content (%)	46.63	46.63
Telomere no.	32	35
Repetitive sequences (%)	63.64	63.31
No. of protein-coding genes	12,737	12,692
Mean gene length (bp)	2,130.81	2,031.33
No. of secreted proteins	1,135	1,119

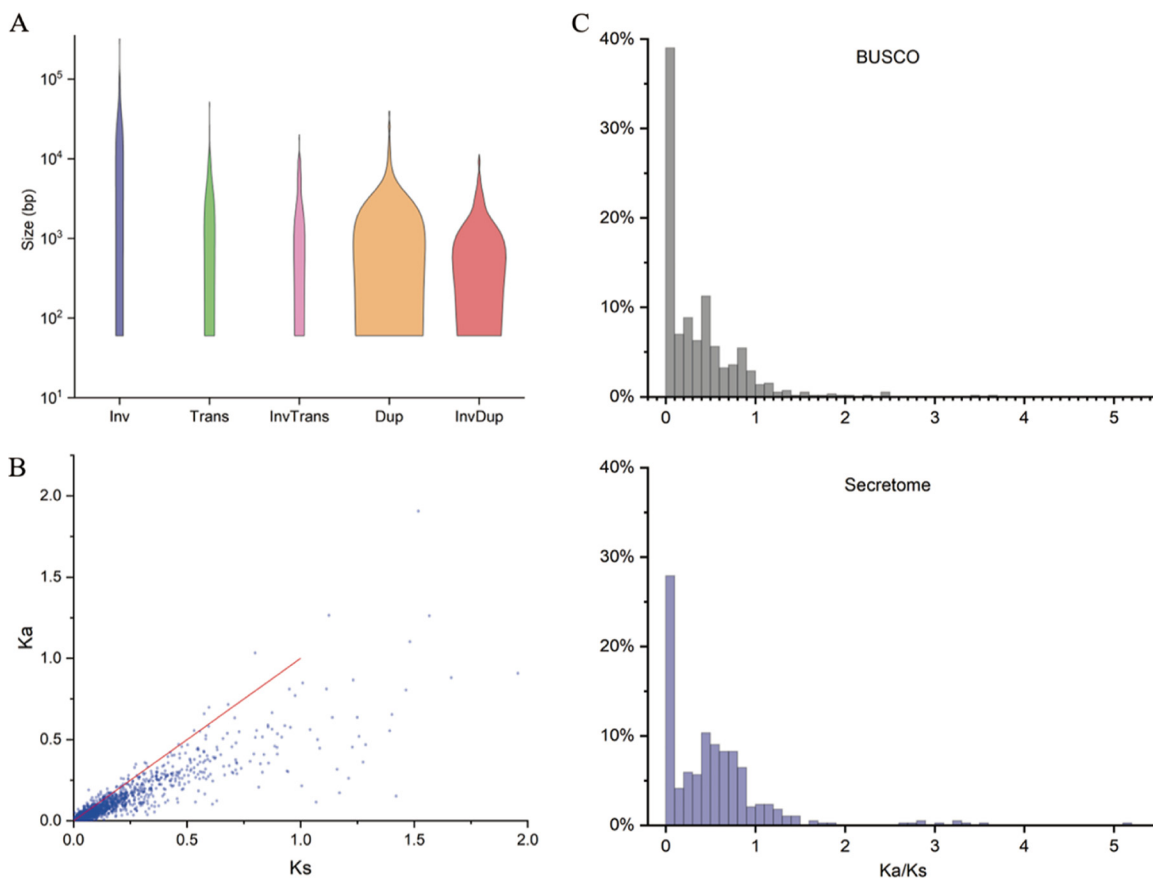
the shortest, as chromosome 18 (chr18A), and homologous chromosomes in haplotype B were identified using haplotype synteny (Fig. S1A). Additionally, our phased genome contained 67 telomeric sequences, and 31 chromosomes contained telomeric sequences at each end (Fig. S1B). Genome synteny between our assembly and Pt76 scaffold-level chromosomes validated the chromosomal structure of the assembled contigs (Fig. S2). The chromosomal structure was also confirmed using Hi-C contact maps by mapping Hi-C sequencing reads from the Pt76 and Pt64 genomes (9, 19) to our assembly (Fig. S3).

We next evaluated the completeness, continuity, accuracy, and phasing quality of our assembly. Benchmarking Universal Single-Copy Orthologs (BUSCO) (27) analysis for evaluating gene space completeness showed that of 1,335 single-copy orthologues in Basidiomycota, 94.3% and 93.7% were complete in the two haplotypes, respectively. The missing rates of BUSCO genes were 2.6% and 2.8% (Fig. 1A). The long terminal repeat (LTR) assembly index (LAI) was used to assess the continuity of repeat regions (28). The LAI scores of the two haplotypes were 20.63 and 18.93 respectively, which was indicative of a well-assembled genome with many repeats. Long-read coverage analysis at the whole-genome level showed that the two haplotypes were fully phased (Fig. 1B). Merqury analysis (29) with highly accurate Illumina reads showed that the estimated completeness was 97.24% based on k-mers of 19, and the base call accuracy of each chromosome was  $>QV50$  (QV: quality value). Collectively, these analyses demonstrated that our *de novo* phased assembly is of high quality.

We predicted a total of 25,429 protein-coding genes for haplotype A and haplotype B (Table 1; Table S4). BUSCO evaluation showed that the gene annotation complete-



**FIG 1** Assessment of genome assembly. (A) Genome completeness of the two haplotypes was assessed using conserved fungal BUSCO genes. (B) Read depth histogram plotted by mapping HiFi reads to the phased assembly, showing full phasing of the assembled genome.

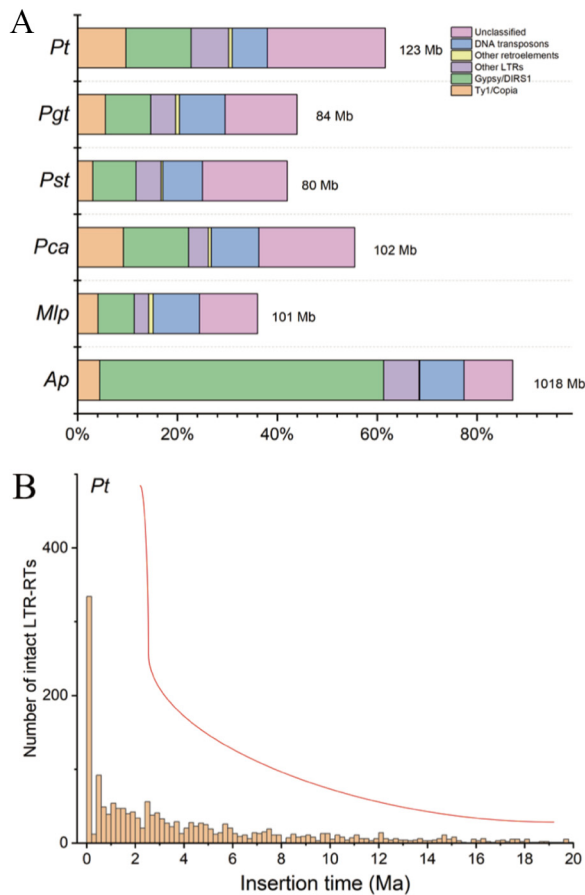


**FIG 2** Haplotype variation. (A) Size distribution of structural variations between homologous chromosomes. Inv, inversions; Trans, translocations; InvTrans, invert translocations; Dup, duplications; InvDup, invert duplications. (B)  $K_a$  and  $K_s$  distribution for allelic genes. (C) Distribution of  $K_a/K_s$  ratios in conserved fungal BUSCO genes and secreted protein genes.

ness was 95.9%, and the missing rate of conserved genes was 1% (Fig. S4). Predicted secreted proteins of the two haplotypes accounted for ~9% of total proteins. Repeat and gene density plots for the chromosomes showed that the distribution of secreted protein genes was not correlated with repeat sequences (Fig. S5).

**Haplotype variations between homologous chromosomes.** The fully phased genome at the chromosomal level allowed us to discover genomic changes between homologous chromosomes. We found 98.79% sequence identity between the two haplotypes by calculating the average nucleotide identity (ANI) (30). We used Syri (31) to identify structural variations (SVs) between homologous chromosomes (Fig. S6). A total of 108.60 Mb (88.42%) in haplotype A and 108.48 Mb (89.57%) in haplotype B were syntenic blocks. We identified 53 inversions (1,349,548 bp), 73 translocations (280,613 bp), 76 inverted translocations (236,774 bp), 495 duplications (1,544,911 bp), and 370 inverted duplications (587,203 bp) (Table S5). These SVs ranged from over 100 bp to hundreds of kilobases, and the predominant SVs were inversions (Fig. 2A). Sequence variations included 370,259 single nucleotide polymorphisms (SNPs), 19,177 insertions (317,638 bp), and 19,834 deletions (364,941 bp). These variations between homologous chromosomes indicated that haplotypic divergence in Pt15 was largely caused by highly frequent small-scale sequence divergence.

We identified a total of 9,956 allelic gene pairs using gene synteny between homologous chromosomes. All allelic pairs were one-one syntenic genes. All allelic genes were colinear except that only one inversion was found on chr4. Most allelic genes underwent purifying selection, and 5.4% underwent positive selection (Fig. 2B). We compared the evolutionary differences between secreted protein genes and conserved fungal BUSCO genes. The ratio of the number of substitutions per nonsynonymous site



**FIG 3** Analysis of TEs in rust species. (A) Comparison of different TEs in *P. triticina* (*Pt*), *P. graminis* f. sp. *tritici* (*Pgt*), *P. coronata* f. sp. *avenae* (*Pca*), *P. striiformis* f. sp. *tritici* (*Pst*), *M. larici-populina* (*Mlp*), and *A. psidii* (*Ap*). (B) Insertion times of LTR-RTs in *P. triticina*. The red curve represents the trend of the accumulation of TEs in selected rust species. Ma, 1 million years.

to the number of substitutions per synonymous site ( $K_a/K_s$  ratio; median, 0.4381; mean, 0.5314) of secreted protein genes was higher than that of BUSCO genes (median, 0.2307; mean, 0.3626). The proportion of positively selected allelic genes encoding secreted proteins was higher than that of BUSCO genes, and the same was true for the proportion of allelic genes that underwent more relaxation of purifying selection (Fig. 2C), suggesting that secreted protein genes evolved more rapidly and were subjected to stronger selective pressure in response to host adaptation.

**Characterization of repetitive sequences.** The two haplotypes in Pt15 contained similar levels of repetitive sequences (haplotype A, 63.64%; haplotype B, 63.31%). The repetitive sequences were almost evenly distributed on each chromosome (Table S3). The predominant transposable element (TE) type was LTR retrotransposons (LTR-RTs), which mainly included Gypsy and Copia elements (Fig. 3A). We also reanalyzed repetitive sequences of the scaffold-level chromosome assembly for *Puccinia graminis* f. sp. *tritici*, *Puccinia striiformis* f. sp. *tritici*, *Puccinia coronata* f. sp. *avenae*, and *Melampsora larici-populina* and the primary assembly for *Austropuccinia psidii* (Table S6). The distribution of TEs in these rust species was similar to that in our assembly (Fig. 3A), indicating that the LTR-RTs, including Gypsy and Copia, and unclassified TEs were the dominant repetitive elements in rust fungi.

We calculated the insertion time of the LTR-RTs in the six genomes (Fig. 3B; Fig. S7), based on the divergence between the LTRs at the ends of each LTR-RT. Although the sequence accuracy of TEs may vary in these genomes and could influence the calculation, the six rust species were found to undergo the significant accumulation of LTR-

RTs for genome expansion millions of years ago. Recent large-scale bursts of LTR-RTs were observed over the last 0.2 million years, which contributed dramatically to genome expansion in *Puccinia* species and *M. larici-populina*. The gigabase-sized genome in *A. psidii* was attributed to the recent burst and the accumulation over a long period. Given that different TE families in *A. psidii* showed the burst at discrete time scales (32), we inferred that *A. psidii* has evolved the complex and efficient mechanism of TE expansion in shaping the superset of the *A. psidii* genome compared to those of many species of the rust lineage. Our data also indicated that the expansion of LTR-RTs may still be ongoing.

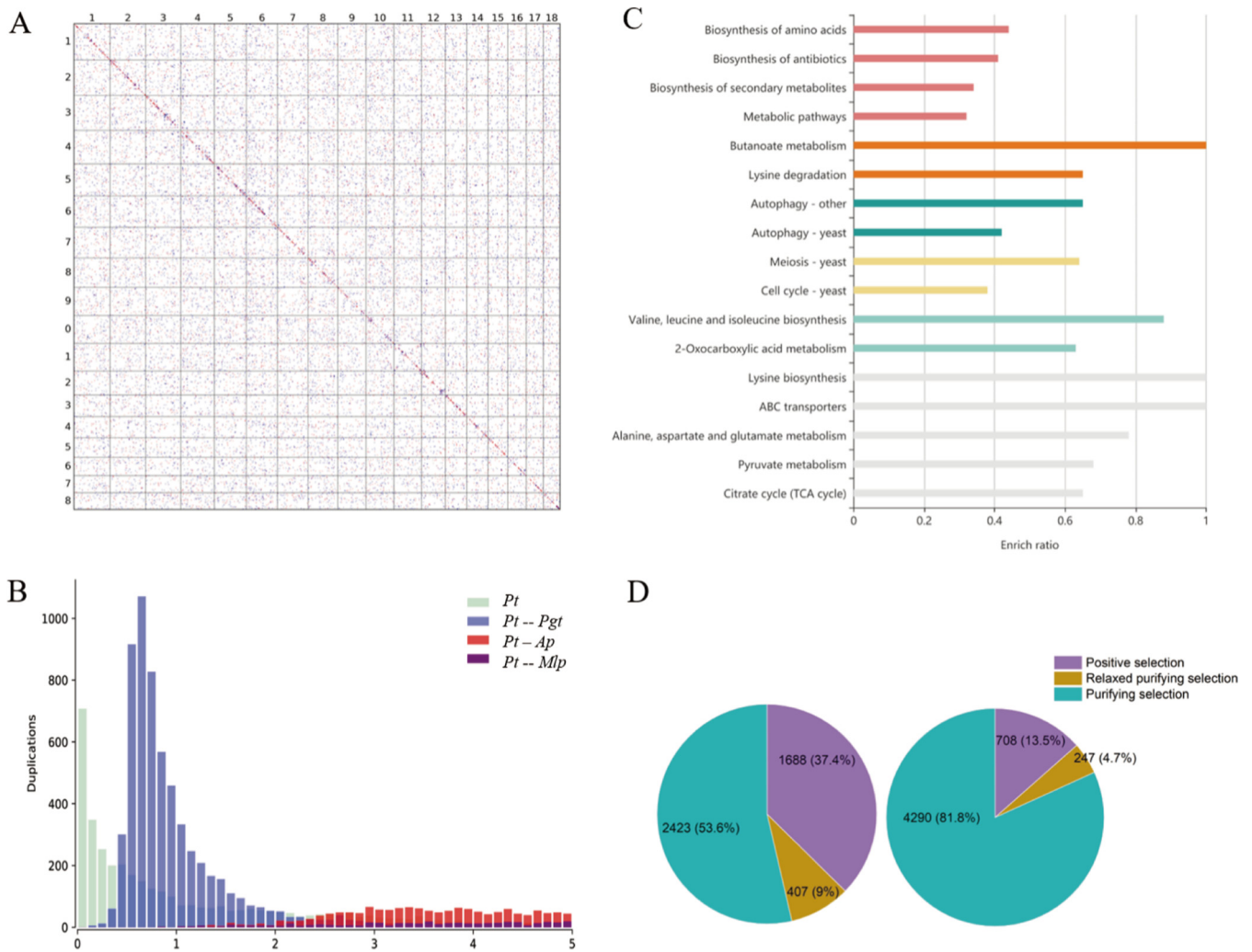
**Gene duplication drives genome evolution.** Gene duplications play a crucial role in species adaptation to diverse environmental changes. Collapsed or incomplete haplotype assembly often leads to errors in the identification of gene duplications in heterozygous diploid genomes. Using whole-genome pairwise alignments of homologous genes in *P. triticina*, we did not observe significant syntenic blocks within the haploid chromosome set (Fig. 4A) and found that gene duplicates were randomly distributed in the chromosomal structures. Gene collinearity supported the randomness of gene duplicates (Fig. S8). We detected the distribution of  $K_s$  values of paralogues in *P. triticina*, and no  $K_s$  peak was observed (Fig. 4B). These findings suggested that *P. triticina* did not experience recent whole-genome duplications (WGDs) or large-scale gene duplication events, consistent with the observations for  $K_s$  distribution of paralogues in *P. graminis*, *P. striiformis* f. sp. *tritici*, *P. coronata* f. sp. *avenae*, *A. psidii*, and *M. larici-populina* (Fig. S9). Most gene duplicates in *P. triticina* and other rust species were dispersed duplicates, which may result from extensive segmental duplications, given that segmental duplication events often occur in eukaryotic genomes with a high TE content (33–35).

We performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of gene duplicates in *P. triticina* after *P. triticina*-*P. graminis* divergence ( $K_s < 0.7$ ). The results showed that gene duplicates were mainly enriched in pathways including metabolism, biosynthesis, and ABC transporters, which contribute to lifestyle adaptation (Fig. 4C). The number and proportion of paralogues that underwent positive selection ( $K_a/K_s$  ratio  $> 1$ ) and relaxed purifying selection ( $K_a/K_s > 0.8$ ) were significantly higher after *P. triticina*-*P. graminis* divergence than before *P. triticina*-*P. graminis* divergence (Fig. 4D). Species-specific genes in *P. triticina* were inferred using protein clusters of *P. triticina* and five other rusts. Species-specific genes shared over one-third of genes that underwent positive selection after *P. triticina*-*P. graminis* divergence (Fig. S10), suggesting that rapid evolution of species-specific genes in *P. triticina* contributed to environmental adaptability.

**Chromosomal rearrangement among rust genomes.** To investigate chromosome structural changes among rust species, we used *P. triticina* chromosomes as the reference genome to perform comparative analysis with high-quality genomes of five other rust species. By mapping the *P. graminis*, *P. striiformis* f. sp. *tritici*, and *P. coronata* f. sp. *avenae* chromosome sequences onto the *P. triticina* genome, we found that overall synteny existed between genomes and genes of these closely related species, and we identified multiple chromosomal rearrangements on different chromosomes, most of which were large inversions (Fig. 5A; Fig. S11). The irregular locations of rearrangements suggested that rearrangement events may occur randomly. In combination with the timescale of TEs of rust genomes in our analysis, the overall genome synteny between *P. triticina* and the other three *Puccinia* species should be attributed to homologous gene synteny rather than to TEs. Each chromosome of *P. graminis*, *P. striiformis* f. sp. *tritici*, and *P. coronata* f. sp. *avenae* had one matched chromosome in *P. triticina* (Fig. 5B), meaning that continual rearrangements occurred on each chromosome. These analyses suggested that frequent rearrangements of ancestral chromosomes probably led to the current karyotypes in these species.

*P. triticina* and *P. graminis* shared 10,637 syntenic genes of 5,326 pairs that accounted for 34.85% of their total genes. *P. triticina* and *P. striiformis* f. sp. *tritici* shared 9,422 syntenic genes of 4,756 pairs that accounted for 33.45% of their total genes. *P. triticina* and *P. coronata* f. sp. *avenae* shared 8,407 genes of 4,218 pairs that accounted for 28.58% of their total genes. A total of 5,317, 4,681, and 4,198 genes in *P. triticina*

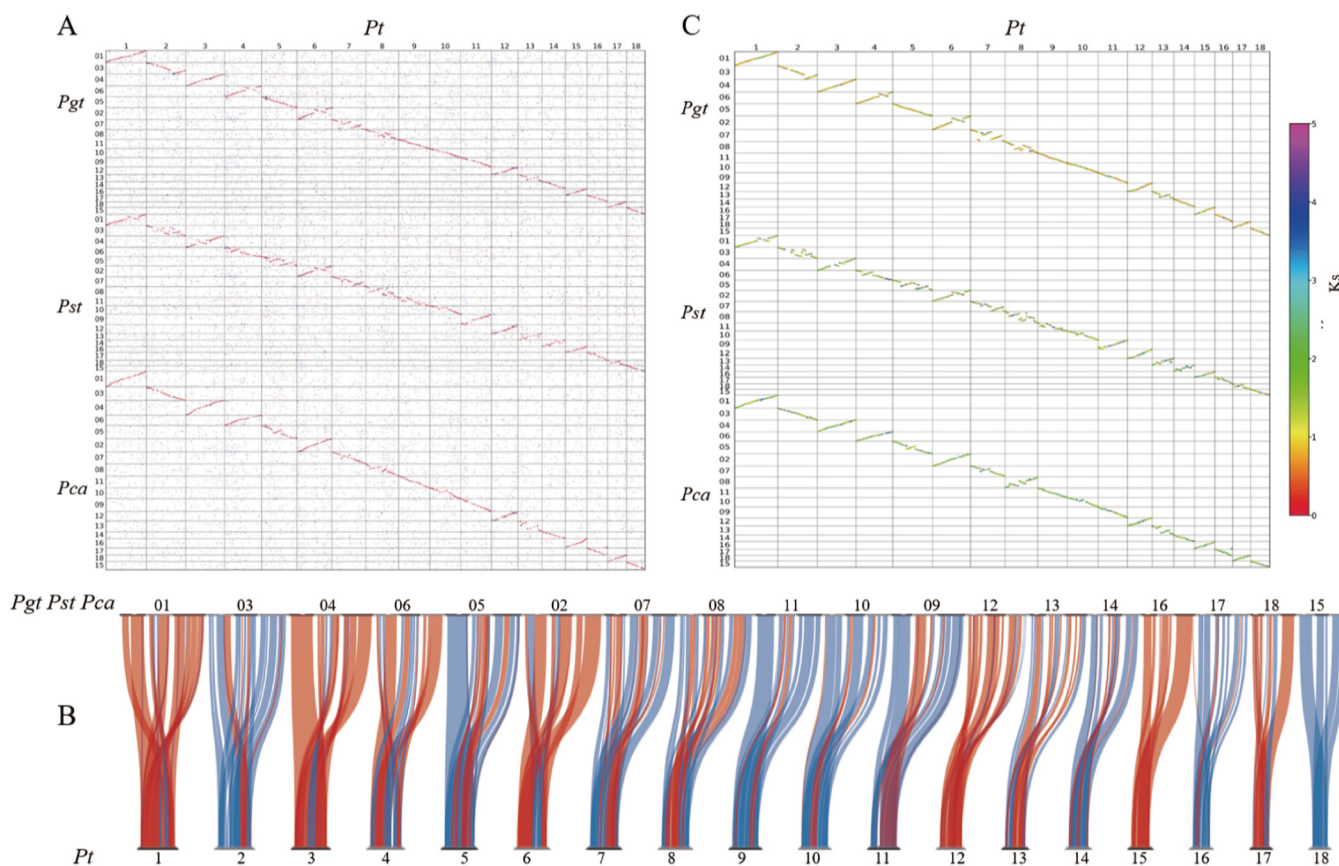




**FIG 4** Gene duplication and evolution. (A) Dot plots showing gene duplicates in *P. triticina*. Chromosome numbers are shown. Best hits, secondary hits, and other hits are shown in red, blue, and gray, respectively. (B)  $K_s$  distribution for paralogues in *P. triticina* and orthologues in *P. triticina* and other rust species. (C) Bar plot of KEGG pathway enrichment for *P. triticina* genes after *P. triticina-P. graminis* divergence. TCA, tricarboxylic acid. (D) Distribution of  $K_a/K_s$  ratios for paralogues in *P. triticina*. (Left) After *P. triticina-P. graminis* divergence; (right) before *P. triticina-P. graminis* divergence. Positive selection, relaxed purifying selection, and purifying selection represent  $K_a/K_s > 1$ ,  $0.8 < K_a/K_s < 1$ , and  $K_a/K_s \leq 0.8$ , respectively.

were syntenic to those of *P. graminis*, *P. striiformis* f. sp. *tritici*, and *P. coronata* f. sp. *avenae*, respectively. We used  $K_s$  values of gene pairs between syntenic blocks to evaluate gene divergence among these species. The mean  $K_s$  values of *P. triticina-P. graminis*, *P. triticina-P. striiformis* f. sp. *tritici*, and *P. triticina-P. coronata* f. sp. *avenae* were 1.22732, 1.88705, and 1.88731, respectively. These results showed the phylogenetic divergences of the four *Puccinia* species. The  $K_s$  comparisons of matched chromosomes of the three species and *P. triticina* showed that rearrangement may promote gene divergence; for example, gene divergence on chr13 was relatively greater (Fig. 5C; Fig. S12). Altogether, the comparisons suggested that rearrangements may lead to genetic innovations responsible for adaptation and/or speciation.

Subsequently, we detected chromosome structural changes of *M. larici-populina* and *A. psidii* from Melampsoraceae and Sphaerophragmiaceae, respectively, compared to *P. triticina*. For the *M. larici-populina* chromosome assembly, homologous genes were disordered compared to *P. triticina* chromosomes but were mostly clustered on each matched chromosome (Fig. 6A). Gene synteny identified retained signals of recurrent chromosome shuffling between *M. larici-populina* and *P. triticina* (Fig. S13A). No significant large rearrangement was observed, largely due to the larger phylogenetic distance between *M. larici-populina* and *P. triticina*. A high degree of gene order divergence may be



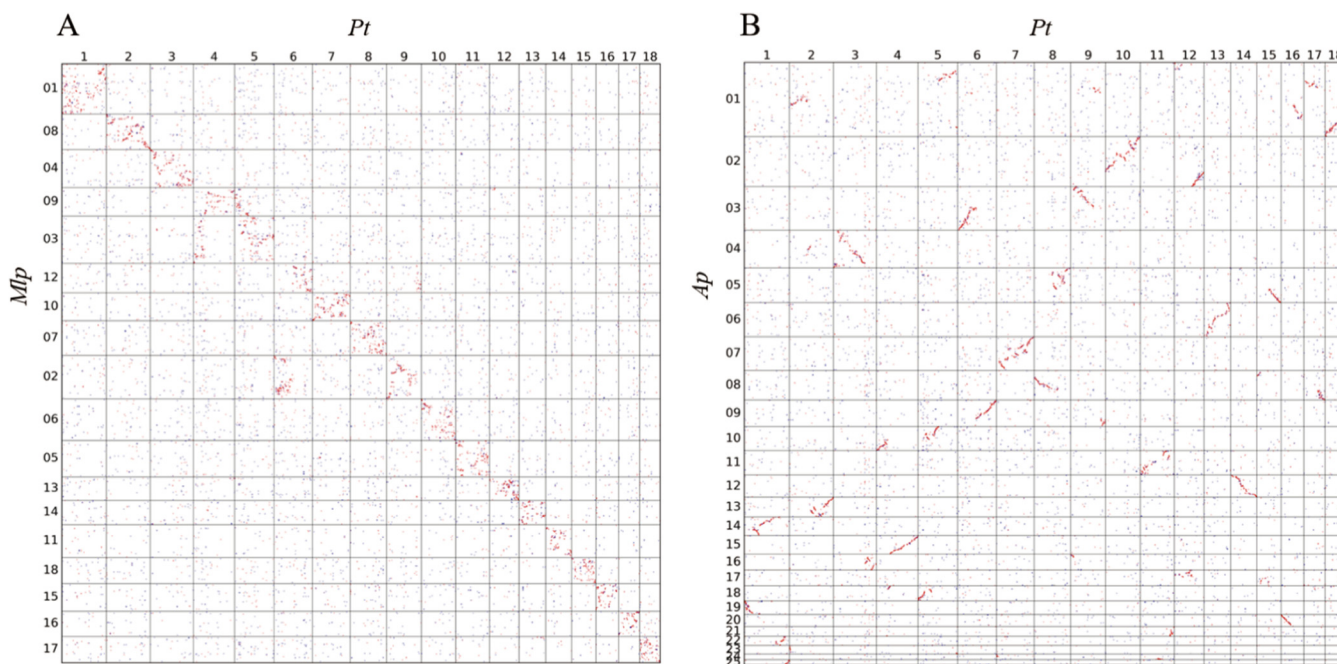
**FIG 5** (A) Dot plots showing homologous genes in *P. triticina-P. graminis*, *P. triticina-P. striiformis f. sp. tritici*, and *P. triticina-P. coronata f. sp. avenae*. Best hits, secondary hits, and other hits are shown in red, blue, and gray, respectively. (B) Gene synteny in *P. triticina-P. graminis*, *P. triticina-P. striiformis f. sp. tritici*, and *P. triticina-P. coronata f. sp. avenae*. Syntenic blocks containing at least 10 genes are shown. Regular chromosomes are in blue and reversed chromosomes are in red. (C)  $K_s$  distribution of gene pairs between syntenic blocks. The median value is shown to represent the  $K_s$  distribution of the blocks containing at least 10 gene pairs. Chromosome numbers are shown for the four species.

attributed to recurrent rearrangements from their ancestral chromosomes. When the supersized assembly in *A. psidii* was compared to the *P. triticina* chromosomes, multiple chromosomal rearrangements were detected regardless of the effect of large numbers of TEs (Fig. 6B; Fig. S13B). This is likely due to the phylogenetic relationship between Sphaerophragmiaceae and Pucciniaceae. Chromosome structural changes in *M. laricipopulina* and *A. psidii* suggested that rearrangements may exist extensively in a broad array of species in the rust lineage.

## DISCUSSION

An accurate, chromosome-level phased genome is necessary to understand the biology and evolution of species, such as dikaryotic fungi. Previous genome projects have shown that Hi-C data appeared to be indispensable for the completion of chromosome-level and phased assembly in dikaryons as well as diploid organisms (15, 19–23). Using higher coverage (80×) of HiFi reads, we obtained a gapless, phased genome of *P. triticina* isolate Pt15 with hifiasm (10) and HERA (25), and its assembly quality was confirmed by high contiguity and accuracy. While Hi-C sequencing was unavailable in the current project, which limited the assessment of phase switches in the two haplotypes (9), the chromosomal structure and its quality in Pt15 were determined by multiple approaches, such as the assessment of genome data of other *P. triticina* isolates (9, 19) and the coverage of long reads in the two sets of chromosomes. Overall, this chromosome assembly presents the first gapless, phased genome for a dikaryotic rust fungus and provides a feasible strategy for resolving phased, chromosome-scale genomes of dikaryotic fungi.





**FIG 6** Dot plots showing homologous genes in *P. triticina*-*M. larici-populina* (A) and *P. triticina*-*A. psidii* (B). Chromosome numbers are shown for *P. triticina* and *M. larici-populina*. Sequence numbers are shown for *A. psidii*.

The gene set identified in the two haplotypes of Pt15 showed high completeness of conserved fungal genes. Nevertheless, the number of identified genes in each haplotype (12,737 and 12,692) was relatively lower than the previous haplotypes of other *P. triticina* isolates and *Puccinia* species (9, 21, 22). This was likely due to the effect of Pt15 haplotypes with high contiguity and lack of reductant heterozygous sequences or was associated with the differences in the transcript data.

The dramatic expansions of TEs and gene families have shaped adaptive evolution of rust fungi (13, 36). We evaluated the role of the expansions at the chromosomal scale. Recent bursts of TEs and extensive segmental gene duplications characterize the evolution of chromosome structures in *P. triticina* and five other rust species, suggesting that TEs and segmental duplications synergistically contribute to genomic adaptation in the rust lineage.

Via comparative analysis of *P. triticina* chromosomes, we uncovered frequent chromosomal rearrangements in species from Pucciniaceae, Sphaerophragmiaceae, and Melampsoraceae. Given that these species have the larger phylogenetic distance across the rust lineage (13, 37) and lack evidence of recent WGDs, we speculate that chromosome structural changes act regularly and chromosome number changes may be rare in the formation and evolution of modern rust species. This leads to our hypothesis that haploid chromosome number ( $n = 18$ ) may be highly conserved in a wide range of species of the rust lineage, although multiple assembled chromosomes (9, 15, 19–23, 38) support a consistent chromosome number and cytological evidence of karyotypes still remains insufficient (26).

It has been proposed that speciation is driven by chromosomal rearrangements in eukaryotes. For instance, chromosomal inversion can suppress recombination, resulting in reproductive isolation to speciation (39–41). DNA double-strand break repair during the proliferation of TEs has been known to mediate rearrangements (41–43). Strikingly, a relatively higher proportion of TEs in the rust lineage than in other groups of plant pathogens (44–47) may confer the benefit of frequent chromosomal rearrangements, facilitating rapid species radiation by triggering a high rate of formation of new species. This inference is consistent with the observation that species radiation is driven by frequent rearrangements across other eukaryotes, such as the fungal phylum Ascomycota (48) and the angiosperm family Cyperaceae (49). It remains unclear

how the occurrence and fixation of chromosomal rearrangements are regulated by TEs or natural selection. Chromosome assembly and comparisons among more diverse taxa are required to reveal the mechanism and trajectory of evolutionary radiation for rust fungi and other fungal lineages.

In summary, this study reports a high-quality gapless genome for *P. triticina* and provides insights into resolving phased genomes for dikaryotic fungi. Further evolutionary and comparative analyses will enhance the understanding of how chromosome evolution facilitates adaptation and species radiation in Pucciniales.

## MATERIALS AND METHODS

***Puccinia triticina* isolate and pathotyping.** *P. triticina* samples were collected from Henan province of China. After single-spore isolation and purification, urediniospores were produced on seedlings of the wheat cultivar Mingxian16. Pathotype identification was tested on the standard wheat differential set for wheat leaf rust.

**DNA extraction and sequencing.** High-molecular-weight DNA was extracted from urediniospores as previously described (50). Qualified DNA was used for constructing a PCR-free SMRT bell library (15 kb) and was sequenced in a SMRT cell on the PacBio Sequel II platform. A PCR-free library for Illumina sequencing was prepared, and 150-bp paired-end reads were produced on the Illumina NovaSeq 6000 platform.

**Genome assembly and evaluation.** Genome assembly was first obtained with HiFi reads using hifiasm 0.15 (10). A BLASTN (51) search against the *P. triticina* mitochondrial sequences and the NCBI NT database was performed to discard contaminated sequences (52). Heterozygous sequences in each haplotype assembly were identified and removed using Purge Haplotigs (53). HERA (25) was used to construct the final assembly. Hi-C reads for Pt76 and Pt64 were downloaded from the NCBI SRA database (Pt76, [SRR14386306](https://www.ncbi.nlm.nih.gov/sra/SRR14386306); Pt64, [SRR14470629](https://www.ncbi.nlm.nih.gov/sra/SRR14470629)). Hi-C contact maps were produced with HiC-Pro 3.10 (54) and HiCExplorer 3.7.1 (55).

Telomeres were identified by searching multiple repeats of TTAGGG or CCCTAA within 200 bp. Genome completeness was assessed using BUSCO 3.02 with genome mode (27) based on the data set basidiomycota\_odb9 (2016-02-13). The LAI program (28) was used to evaluate LTRs with default parameters. Genome completeness and accuracy were measured using Merqury (29) with Illumina reads. A genome-wide coverage plot was generated by mapping long reads to genome assembly with the CollapsedGenomicRegions tool (<https://github.com/JanaSperschneider/GenomeAssemblyTools>).

**Genome annotation.** *De novo* repeats were identified with RepeatModeler 2.0.2a and the option -LTRstruct (56). These repetitive elements and Pucciniomycetes elements from the RepeatMasker library (<http://www.repeatmasker.org>) (57) were merged as the annotation library, and then RepeatMasker 4.1.2 was run to annotate repetitive sequences. Annotation of protein-coding genes was performed on the repeat-masked genome with the Funannotate pipeline (<https://github.com/nextgenusfs/funannotate>). Transcript and protein evidence included full-length cDNA from germinated urediniospores in our genome project and transcript sequences from *Puccinia* genomes (Table S6). *Ab initio* gene prediction tool AUGUSTUS (58) was trained using transcript evidence. GeneMark-ES (59) and GlimmerHMM (60) were used for self-training. Finally, all the above-mentioned gene models were combined to produce the consensus gene sets using EvidenceModeler (61).

Gene annotation was evaluated using BUSCO 3.02 with protein mode (27). Secreted proteins were predicted using a combination of SignalP 5.0 (62) and TMHMM 2.0 (63). Repeat and gene densities for the chromosomes were plotted using karyoploteR (64).

**Analysis of LTR-RTs.** Intact LTR-RTs were identified with LTR\_retriever (65) based on the merged results from LTR\_FINDER (66) and LTRharvest (67). The nucleotide alignments between intact LTR-RTs were performed by MUSCLE (68). The insertion time ( $T$ ) of LTR-RTs was calculated based on the following formula:  $T = K/2r$ , where  $K$  is the divergence rate between the 5' and 3' ends of LTRs and  $r$  is the fungal substitution rate of  $1.05 \times 10^{-9}$  nucleotides per site per year (69).  $K$  was estimated using R package ape (<http://ape-package.ird.fr/>) with the K80 model.

**Structure variation detection.** The ANI was calculated using fastANI with “-fragLen 1 Mb” (30). Whole-genome alignments between the two haplotypes was performed using Nucmer of MUMmer 3.23 with the parameters “-maxmatch -c 100 -b 500 -l 50,” followed by delta-filter and show-coords (70). SyRi (31) was used to identify genomic rearrangements.

**Analysis of gene duplications.**  $K_s$  distributions of paralogues and orthologues were generated using the wgd pipeline (<https://github.com/arzwa/wgd>) (71). KEGG pathway enrichment analysis was conducted using KOBAS-i (72) with a  $P$  value of  $<0.05$ . OrthoFinder 2.3.3 (73) was used to infer species-specific genes using Markov cluster algorithm (MCL) clustering.

**Synten analysis.** A whole-genome pairwise alignment of genome assemblies was generated using d-Genies (74). Homologous genes within and between the genomes were identified with their protein sequences by BLASTP alignments (E value  $< 1e-5$ ). WGDI (75) was used to construct homologous-gene dot plots. Gene synteny was identified using MCScanX (76) and visualized using SynVisio (<https://synvisio.github.io/#/>).

**$K_a/K_s$  calculation.** The protein sequences were aligned using MUSCLE (68). Protein alignments were converted into coding sequence alignments using PAL2NAL (77). The number of substitutions per non-synonymous site ( $K_a$ ), the number of substitutions per synonymous site ( $K_s$ ), and  $K_a/K_s$  values were estimated with the YN00 model in the PAML program (78).  $K_s$  values of  $>5$  in the analysis were excluded due to saturation.

**Data availability.** The genome sequence of Pt15 has been deposited in the NCBI database under BioProject numbers [PRJNA848635](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA848635) and [PRJNA848634](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA848634). All sequencing reads are available in the SRA database under BioProject number [PRJNA848635](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA848635).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 1.5 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

The study was supported by the National Key R&D Program of China (2021YFD1401000), the Henan Provincial Science and Technology Major Project (221100110100), the Natural Science Foundation of Henan Province (202300410202), and the Research Start-up Fund to Topnotch Talents of Henan Agricultural University (30500924).

W.Z., D.W., and N.L. conceived and designed the study. C.L. prepared samples, analyzed all data sets, and wrote the manuscript. L.Q., Y.L., G.X., Y.S., Y.Z., W.Y., K.C., Z.M., and G.Z. helped in preparing the sample and discussions. X.W. performed the pathotyping experiment. All authors read and approved the final manuscript.

## REFERENCES

- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 21:30. <https://doi.org/10.1186/s13059-020-1935-5>.
- Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, Sudmant PH, Graves TA, Alkan C, Dennis MY, Wilson RK, Turner SW, Korlach J, Eichler EE. 2014. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res* 24:688–696. <https://doi.org/10.1101/gr.168450.113>.
- Logsdon GA, Vollger MR, Eichler EE. 2020. Long-read human genome sequencing and its applications. *Nat Rev Genet* 21:597–614. <https://doi.org/10.1038/s41576-020-0236-x>.
- Nurk S, Koren S, Rhie A, Rautiainen M, Bizkadez AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt SJ, Diekhans M, Logsdon GA, Alonge M, Antonarakis SE, Borchers M, Bouffard GG, Brooks SY, Caldas GV, Cheng H, Chin C-S, Chow W, de Lima LG, Dishuck PC, Durbin R, Dvorkina T, Fiddes IT, Formenti G, Fulton RS, Fungtammasan A, Garrison E, Grady PGS, Graves-Lindsay TA, Hall IM, Hansen NF, Hartley GA, Haukness M, Howe K, Hunkapiller MW, Jain C, Jain M, Jarvis ED, Kerpedjiev P, Kirsche M, Kolmogorov M, Korlach J, Kremitzki M, Li H, Maduro VV, et al. 2022. The complete sequence of a human genome. *Science* 376:44–53. <https://doi.org/10.1126/science.abj6987>.
- Song JM, Xie WZ, Wang S, Guo YX, Koo DH, Kudrna D, Gong C, Huang Y, Feng JW, Zhang W, Zhou Y, Zuccolo A, Long E, Lee S, Talag J, Zhou R, Zhu XT, Yuan D, Udall J, Xie W, Wang RA, Zhang Q, Poland J, Zhang J, Chen LL. 2021. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant* 14:1757–1767. <https://doi.org/10.1016/j.molp.2021.06.018>.
- Li K, Jiang W, Hui Y, Kong M, Feng LY, Gao LZ, Li P, Lu S. 2021. Gapless indica rice genome reveals synergistic contributions of active transposable elements and segmental duplications to rice genome evolution. *Mol Plant* 14:1745–1756. <https://doi.org/10.1016/j.molp.2021.06.017>.
- Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, Wai CM, Zheng C, Shi Y, Chen S, Xu X, Yue J, Nelson DR, Huang L, Li Z, Xu H, Zhou D, Wang Y, Hu W, Lin J, Deng Y, Pandey N, Mancini M, Zerpá D, Nguyen JK, Wang L, Yu L, Xin Y, Ge L, Arro J, Han JO, Chakrabarty S, Pushko M, Zhang W, Ma Y, Ma P, Lv M, Chen F, Zheng G, Xu J, Yang Z, Deng F, Chen X, Liao Z, Zhang X, Lin Z, et al. 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat Genet* 50:1565–1573. <https://doi.org/10.1038/s41588-018-0237-2>.
- Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, Deng G, Wang H, Hu F, Zhao S, Fernie AR, Boerno S, Timmermann B, Zhang P, Vingron M. 2017. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat Plants* 3:696–703. <https://doi.org/10.1038/s41477-017-0002-z>.
- Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R, Upadhyaya NM, Rathjen JP, Stone EA, Schwessinger B, Figueroa M, Dodds PN, Periyannan S, Sperschneider J. 2022. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biol* 23:84. <https://doi.org/10.1186/s13059-022-02658-2>.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. 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>.
- Guk JY, Jang MJ, Choi JW, Lee YM, Kim S. 2022. De novo phasing resolves haplotype sequences in complex plant genomes. *Plant Biotechnol J* 20:1031–1041. <https://doi.org/10.1111/pbi.13815>.
- Lorrain C, Goncalves Dos Santos KC, Germain H, Hecker A, Duplessis S. 2019. Advances in understanding obligate biotrophy in rust fungi. *New Phytol* 222:1190–1206. <https://doi.org/10.1111/nph.15641>.
- Aime MC, McTaggart AR, Mondo SJ, Duplessis S. 2017. Phylogenetics and phylogenomics of rust fungi. *Adv Genet* 100:267–307. <https://doi.org/10.1016/bs.adgen.2017.09.011>.
- Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, Guo J, Zhuang H, Qiu C, Liu J, Liu H, Huang X, Pei G, Zhan G, Tang C, Cheng Y, Liu M, Zhang J, Zhao Z, Zhang S, Han Q, Han D, Zhang H, Zhao J, Gao X, Wang J, Ni P, Dong W, Yang L, Yang H, Xu JR, Zhang G, Kang Z. 2013. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat Commun* 4:2673. <https://doi.org/10.1038/ncomms3673>.
- Li F, Upadhyaya NM, Sperschneider J, Matry O, Nguyen-Phuc H, Mago R, Raley C, Miller ME, Silverstein KAT, Henningsen E, Hirsch CD, Visser B, Pretorius ZA, Steffenson BJ, Schwessinger B, Dodds PN, Figueroa M. 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nat Commun* 10:5068. <https://doi.org/10.1038/s41467-019-12927-7>.
- Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM, Dodds PN, Figueroa M, Park RF, Rathjen JP. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. *mBio* 9:e02275-17. <https://doi.org/10.1128/mBio.02275-17>.
- Miller ME, Zhang Y, Omidvar V, Sperschneider J, Schwessinger B, Raley C, Palmer JM, Garnica D, Upadhyaya N, Rathjen J, Taylor JM, Park RF, Dodds PN, Hirsch CD, Kianian SF, Figueroa M. 2018. De novo assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal agent of oat crown rust. *mBio* 9:e01650-17. <https://doi.org/10.1128/mBio.01650-17>.
- Aime MC, Bell CD, Wilson AW. 2018. Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. *Stud Mycol* 89:143–152. <https://doi.org/10.1016/j.simyco.2018.02.002>.
- Wu JQ, Song L, Ding Y, Dong C, Hasan M, Park RF. 2021. A chromosome-scale assembly of the wheat leaf rust pathogen *Puccinia triticina* provides insights into structural variations and genetic relationships with haplotype resolution. *Front Microbiol* 12:704253. <https://doi.org/10.3389/fmicb.2021.704253>.



20. Edwards RJ, Dong C, Park RF, Tobias PA. 2022. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, *Austropuccinia psidii*. bioRxiv. 2022.04.22.489119. <https://doi.org/10.1101/2022.04.22.489119>.
21. Schwessinger B, Jones A, Albekaa M, Hu Y, Mackenzie A, Tam R, Nagar R, Milgate A, Rathjen JP, Periyannan S. 2022. A chromosome scale assembly of an Australian *Puccinia striiformis* f. sp. *tritici* isolate of the PstS1 lineage. Mol Plant Microbe Interact 35:293–296. <https://doi.org/10.1094/MPMI-09-21-0236-A>.
22. Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, Kianian SF, Steffenson BJ, Dodds PN, Sperschnieder J, Figueroa M. 2022. A chromosome-level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp. *avenae*: a resource to enable comparative genomics in the cereal rusts. G3 (Bethesda) 12:jkac149. <https://doi.org/10.1093/g3journal/jkac149>.
23. Liang J, Li Y, Dodds PN, Figueroa M, Sperschnieder J, Han S, Tsui CKM, Zhang K, Li L, Ma Z, Cai L. 2022. The largest haplotype-phased and chromosome-scale genome assembly of the fungal pathogen, *Puccinia polysora*, by HiFi reads and Hi-C data. bioRxiv. <https://doi.org/10.1101/2022.05.18.492400>.
24. Bolton MD, Kolmer JA, Garvin DF. 2008. Wheat leaf rust caused by *Puccinia triticina*. Mol Plant Pathol 9:563–575. <https://doi.org/10.1111/j.1364-3703.2008.00487.x>.
25. Du H, Liang C. 2019. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. Nat Commun 10:5360. <https://doi.org/10.1038/s41467-019-13355-3>.
26. Leonard KJ, Szabo LJ. 2005. Stem rust of small grains and grasses caused by *Puccinia graminis*. Mol Plant Pathol 6:99–111. <https://doi.org/10.1111/j.1364-3703.2005.00273.x>.
27. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35: 543–548. <https://doi.org/10.1093/molbev/msx319>.
28. Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res 46:e126.
29. Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 21:245. <https://doi.org/10.1186/s13059-020-02134-9>.
30. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
31. Goel M, Sun H, Jiao WB, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol 20:277. <https://doi.org/10.1186/s13059-019-1911-0>.
32. Tobias PA, Schwessinger B, Deng CH, Wu C, Dong C, Sperschnieder J, Jones A, Lou Z, Zhang P, Sandhu K, Smith GR, Tibbitts J, Chagne D, Park RF. 2021. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 (Bethesda) 11:jkaa015. <https://doi.org/10.1093/g3journal/jkaa015>.
33. Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. Plant Physiol 171:2294–2316. <https://doi.org/10.1104/pp.16.00523>.
34. Grandaubert J, Balesdent M-H, Rouxel T. 2014. Evolutionary and adaptive role of transposable elements in fungal genomes. Adv Bot Res 70:79–107. <https://doi.org/10.1016/B978-0-12-397940-7.00003-3>.
35. Jiao C, Sorensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun L, Reed R, Jeon S, Kiyonami R, Zhang S, Fernie AR, Brumer H, Domozych DS, Fei Z, Rose JKC. 2020. The *Penium margaritaceum* genome: hallmarks of the origins of land plants. Cell 181:1097–1111.e12. <https://doi.org/10.1016/j.cell.2020.04.019>.
36. Duplessis S, Bakkeren G, Hamelin R. 2014. Advancing knowledge on biology of rust fungi through genomics. Adv Bot Res 70:173–209. <https://doi.org/10.1016/B978-0-12-397940-7.00006-9>.
37. Aime MC, McTaggart AR. 2021. A higher-rank classification for rust fungi, with notes on genera. Fungal Syst Evol 7:21–47. <https://doi.org/10.3114/fuse.2021.07.02>.
38. Persoons A, Maupetit A, Louet C, Andrieux A, Lipzen A, Barry KW, Na H, Adam C, Grigoriev IV, Segura V, Duplessis S, Frey P, Halkett F, De Mita S. 2022. Genomic signatures of a major adaptive event in the pathogenic fungus *Melampsora larici-populina*. Genome Biol Evol 14:evab279. <https://doi.org/10.1093/gbe/evab279>.
39. Wellenreuther M, Bernatchez L. 2018. Eco-evolutionary genomics of chromosomal inversions. Trends Ecol Evol 33:427–440. <https://doi.org/10.1016/j.tree.2018.04.002>.
40. Faria R, Navarro A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol Evol 25:660–669. <https://doi.org/10.1016/j.tree.2010.07.008>.
41. Priest SJ, Yadav V, Heitman J. 2020. Advances in understanding the evolution of fungal genome architecture. F1000Res 9:776. <https://doi.org/10.12688/f1000research.25424.1>.
42. Mehrabi R, Mirzadi Gohari A, Kema GHJ. 2017. Karyotype variability in plant-pathogenic fungi. Annu Rev Phytopathol 55:483–503. <https://doi.org/10.1146/annurev-phyto-080615-095928>.
43. Klein SJ, O'Neill RJ. 2018. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome Res 26: 5–23. <https://doi.org/10.1007/s10577-017-9569-5>.
44. Frantzeskakis L, Németh MZ, Barsoum M, Kusch S, Kiss L, Takamatsu S, Panstruga R. 2019. The *Parauncinula polyspora* draft genome provides insights into patterns of gene erosion and genome expansion in powdery mildew fungi. mBio 10:e01692-19. <https://doi.org/10.1128/mBio.01692-19>.
45. Fouché S, Oggenfuss U, Chanclud E, Croll D. 2022. A devil's bargain with transposable elements in plant pathogens. Trends Genet 38:222–230. <https://doi.org/10.1016/j.tig.2021.08.005>.
46. Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, Donofrio N, Polson SW, Maia TA, Chen C, Adetunji M, Kingham B, Dalio RJD, Resende MLV. 2019. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. PLoS One 14:e0215598. <https://doi.org/10.1371/journal.pone.0215598>.
47. Kijpornyongpan T, Mondo SJ, Barry K, Sandor L, Lee J, Lipzen A, Pangilinan J, LaButti K, Hainaut M, Henrissat B, Grigoriev IV, Spatafora JW, Aime MC. 2018. Broad genomic sampling reveals a smut pathogenic ancestry of the fungal clade Ustilaginomycotina. Mol Biol Evol 35:1840–1854. <https://doi.org/10.1093/molbev/msy072>.
48. Rajeh A, Lv J, Lin Z. 2018. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. BMC Genomics 19:282. <https://doi.org/10.1186/s12864-018-4683-0>.
49. Escudero M, Hipp AL, Waterway MJ, Valente LM. 2012. Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (*Carex*, Cyperaceae). Mol Phylogenet Evol 63: 650–655. <https://doi.org/10.1016/j.ympev.2012.02.005>.
50. Schwessinger B, Rathjen JP. 2017. Extraction of high molecular weight DNA from fungal rust spores for long read sequencing. Methods Mol Biol 1659:49–57. [https://doi.org/10.1007/978-1-4939-7249-4\\_5](https://doi.org/10.1007/978-1-4939-7249-4_5).
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
52. Li C, Lu X, Zhang Y, Liu N, Li C, Zheng W. 2019. The complete mitochondrial genomes of *Puccinia striiformis* f. sp. *tritici* and *Puccinia recondita* f. sp. *tritici*. Mitochondrial DNA B Resour 5:29–30. <https://doi.org/10.1080/23802359.2019.1674744>.
53. Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19:460. <https://doi.org/10.1186/s12859-018-2485-7>.
54. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16:259. <https://doi.org/10.1186/s13059-015-0831-x>.
55. Ramirez F, Bhardwaj V, Arrigoni L, Lam KC, Gruning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 9:189. <https://doi.org/10.1038/s41467-017-02525-w>.
56. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
57. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. <https://doi.org/10.1186/s13100-015-0041-9>.
58. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. <https://doi.org/10.1093/nar/gkl200>.
59. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33: W451–W454. <https://doi.org/10.1093/nar/gki487>.
60. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20: 2878–2879. <https://doi.org/10.1093/bioinformatics/bth315>.

61. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. 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>.
62. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
63. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
64. Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33:3088–3090. <https://doi.org/10.1093/bioinformatics/btx346>.
65. Ou S, Jiang N. 2018. LTR\_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* 176:1410–1422. <https://doi.org/10.1104/pp.17.01310>.
66. Xu Z, Wang H. 2007. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35:W265–W268. <https://doi.org/10.1093/nar/gkm286>.
67. Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18. <https://doi.org/10.1186/1471-2105-9-18>.
68. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
69. Castanera R, Lopez-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Perez G, Pisabarro AG, Grigoriev IV, Stajich JE, Ramirez L. 2016. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genet* 12:e1006108. <https://doi.org/10.1371/journal.pgen.1006108>.
70. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
71. Zwaenepoel A, Van de Peer Y. 2019. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35:2153–2155. <https://doi.org/10.1093/bioinformatics/bty915>.
72. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, Fang S, Cao W, Yi L, Zhao Y, Kong L. 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* 49:W317–W325. <https://doi.org/10.1093/nar/gkab447>.
73. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
74. Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958. <https://doi.org/10.7717/peerj.4958>.
75. Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. 2021. WGDl: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *bioRxiv*. 2021.04.29.441969. <https://doi.org/10.1101/2021.04.29.441969>.
76. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. 2012. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40:e49. <https://doi.org/10.1093/nar/gkr1293>.
77. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612. <https://doi.org/10.1093/nar/gkl315>.
78. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>.