



Long-read genomics reveal extensive nuclear-specific evolution and allele-specific expression in a dikaryotic fungus

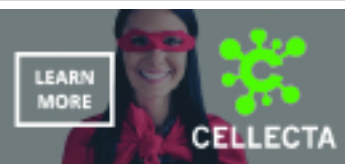
Rita Tam, Mareike Möller, Runpeng Luo, et al.

Genome Res. published online April 11, 2025

Access the most recent version at doi:[10.1101/gr.280359.124](https://doi.org/10.1101/gr.280359.124)

P<P	Published online April 11, 2025 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

Long-read genomics reveal extensive nuclear-specific evolution and allele-specific expression in a dikaryotic fungus

Rita Tam¹, Mareike Möller¹, Runpeng Luo¹, Zhenyan Luo¹, Ashley Jones¹, Sambasivam Periyannan^{1,2,3}, John P. Rathjen^{1†}, Benjamin Schwessinger^{1†}

¹Research School of Biology, Australian National University, Canberra ACT 2601, Australia

²Commonwealth Scientific and Industrial Research Organisation Agriculture and Food, Canberra ACT 2601, Australia

³School of Agriculture and Environmental Science, Centre for Crop Health, University of Southern Queensland, Toowoomba QLD 4350, Australia

†Corresponding authors: John P. Rathjen (john.rathjen@anu.edu.au) and Benjamin Schwessinger (benjamin.schwessinger@anu.edu.au)

Abstract

Phased telomere-to-telomere (T2T) genome assemblies are revolutionising our understanding of long-hidden genome biology “dark matter” such as centromeres, rDNA repeats, inter-haplotype variation, and allele-specific expression (ASE). Yet insights into dikaryotic fungi that separate their haploid genomes into distinct nuclei are limited. Here we explore the impact of dikaryotism on the genome biology of a long-term asexual clone of the wheat pathogenic fungus *Puccinia striiformis* f. sp. *tritici*. We use Oxford Nanopore (ONT) duplex sequencing combined with Hi-C to generate a T2T nuclear-phased assembly with >99.999% consensus accuracy. We show that this fungus has large regional centromeres enriched in LTR retrotransposons, with a single centromeric dip in methylation that suggests one kinetochore attachment site per chromosome. The centromeres of homologous chromosomes are most often highly diverse in sequence and kinetochore attachment sites are not always positionally conserved. Each nucleus carries a unique array of rDNAs with

>200 copies that harbour nucleus-specific sequence variations. The inter-haplotype diversity between the two nuclear genomes is shaped by large-scale structural variations linked to transposable elements. ONT long-read cDNA analysis across dormancy and distinct host infection conditions revealed pervasive ASE for nearly 20% of the heterozygous genes. Genes encoding secreted proteins, including putative virulence effectors, are significantly enriched in ASE genes which appear to be linked to elevated CpG gene body methylation of the lower-expressed allele. This suggests that epigenetically regulated ASE is likely a previously overlooked mechanism facilitating plant infection. Overall, our study reveals how dikaryotism uniquely shapes key eukaryotic genome features.

Introduction

Telomere-to-telomere (T2T) and haplotype-resolved genome assemblies have become the norm in eukaryotic genomics with advances in long-read sequencing technologies. Complete genome assemblies are fundamental for addressing key questions in genome biology that were previously hidden in the “dark matter” of genomes. Key breakthroughs have revolved around centromeres and the embedded kinetochore attachment sites pivotal for understanding karyotype diversity and evolution (Logsdon et al. 2024; Mastorosa et al. 2024), as well as the notoriously repetitive ribosomal DNA (rDNA) arrays whose sequences have only been recently completed in humans (Nurk et al. 2022) and *Arabidopsis* (Fultz et al. 2023). Full haplotype resolution enables precise characterisation of inter-haplotype variations predominantly shaped by heterozygosity, structural rearrangements and transposable element movements (Ferguson et al. 2024; Gluck-Thaler et al. 2022; Hartmann 2022). Further, the complete picture of allelic information facilitates robust assessment of allele-specific expression (ASE) uncovering the underlying variations in *cis*-regulatory elements and epigenetic regulation, which can have important implications on phenotypic variability, as well-established in mammals and plants (e.g. Shi et al. 2024; Cleary and Seoighe 2021; Tian et al. 2022; St. Pierre et al. 2022; Shao et al. 2019).

While there have been important novel insights into diploid and polyploid genome organisation, especially in plants (e.g. Belser et al. 2021; Sun et al. 2022; Hu et al. 2021), less is known about the fungi-specific dikaryotic state where two haploid genomes are contained in separate nuclei in the same cytoplasm and propagated in a coordinated manner during cell division (Anderson and Kohn 2014; Li et al. 2019; Sperschneider et al. 2023a, 2023b). Dikaryotism is highly successful with an estimated 400,000 taxa in the fungal subphylum Basidiomycota relying on it for significant time periods, for example during fruiting body formation in mushrooms and infection processes in rusts and smuts (James et al. 2020; Schmidt-Dannert 2016). Early partially-phased assemblies of rust fungi genomes indicated high levels of heterozygosity and presence/absence polymorphisms between the nuclear genomes, however they lacked the resolution to identify individual haplotypes (Schwessinger et al. 2020, 2018; Vasquez-Gross et al. 2020; Cuomo et al. 2017). The latest fully haplotype-phased and nuclear-assigned T2T genomes have provided the first insights into organisation of genes involved in mating behaviour and contributed to our understanding of reproductive mechanisms generating novel genetic diversity (Luo et al. 2024; Li et al. 2023a; Schwessinger et al. 2022). This includes the somatic exchanges of nuclei between asexual rust lineages that are adapted to the infection of cereals (Henningsson et al. 2024; Sperschneider et al. 2023a), and sexual recombination in permissive hosts (Wang et al. 2022; Rodriguez-Algaba et al. 2022; Du et al. 2023). Such genetic reshuffling produces novel allele combinations of so-called avirulence (*Avr*) genes that encode secreted effector proteins essential for host infection. Rust effectors are under strong selection pressure to diversify into non-recognised virulence alleles because they can be recognised by cognate immune receptors encoded by wheat resistance genes (Salcedo et al. 2017; Ortiz et al. 2022; Chen et al. 2017). This is necessary to escape the plant immune system and to confer the ability to infect new host varieties.

Recent progress in heterokaryotic fungi genomics, such as in arbuscular mycorrhizal fungi, has revealed extensive nuclear genome variations in structural and gene content (Sperschneider et al. 2023b; Li et al. 2019), as well as nuclear-level transcriptomic and epigenetic differences similar to findings in button mushroom *Agaricus bisporus* (Gehrmann et al. 2018). Though key questions

remain around the impact of dikaryotic genome organisation on the evolution of centromeres, rDNA repeats, detailed inter-haplotype variations, ASE and methylation differences at locus resolution. Here we use a long-term asexual clone of the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*) dating back at least 80 years to address these questions in a globally important wheat pathogen (Schwessinger et al. 2018; Wellings 2007; Thach et al. 2016). We leveraged high-accuracy Oxford Nanopore Technology (ONT) duplex long reads to generate the first ONT-based T2T, fully nuclear-phased genome assembly for a dikaryotic fungus. We combined this with comprehensive ONT long-read cDNA datasets sampled during dormancy and host pathogenesis for high-quality gene annotations and differential expression analysis at gene- and allele-specific levels. Our study sheds new light onto the genome biology and adaptive evolutionary potential of rust fungi, with important implications for managing their agricultural impacts.

Results

A T2T nuclear-phased genome assembly of *Puccinia striiformis* f. sp. *tritici* 104E based on high-quality ONT long-read sequencing

We generated a dikaryotic *Puccinia striiformis* f. sp. *tritici* genome assembly for a representative isolate of the Australian founder pathotype 104E137A- (abbreviated as *Pst*104E), which belongs to the long-term asexual *Pst*S0 lineage (Schwessinger et al. 2018, 2020; Wellings 2007). We assembled ONT duplex reads combined with ultralong simplex reads and Hi-C (Supplemental Table S1), followed by scaffolding and manual curation. This resulted in 36 chromosome assemblies corresponding to the 18 homologous pairs, which we sorted by average length of both haplotypes (Chr1 to Chr18) (Fig. 1A; Table 1; Supplemental Fig. S1). Of these, 35 were assembled T2T and each telomere had ~43 repeats on average, consistent with other basidiomycetes (Ramírez Nasto et al. 2011; Schwessinger et al. 2020; Sperschneider et al. 2021). The Hi-C contact heatmap clearly grouped the homologous chromosomes into two nuclear complements as expected for dikaryons (Fig. 1B). Five remaining gaps were found at repetitive regions; for example, a ~650 kbp TE-rich

region on Chr6A near the previously identified mating type *PR* locus (Luo et al. 2024) (Supplemental Fig. S2), and two rDNA arrays on Chr13A and 13B (Supplemental Fig. S3).

We evaluated the quality of the final curated assembly. The full assembly yielded 92.6% of complete BUSCOs (Manni et al. 2021). We used the long terminal repeat (LTR) assembly index (LAI) (Ou et al. 2018) to assess contiguity at repetitive LTR retrotransposons. Haplotype A and B assemblies each received LAI score of 27.2 and 24.8, classifying them to the highest rank based on well-assembled repeats. The per-base accuracy Phred score was 57.4 corresponding to >99.999% consensus accuracy, in line with recent PacBio HiFi assemblies (Wang et al. 2024; Li et al. 2023a; Sperschneider et al. 2023a). CRAQ (Li et al. 2023b) was used to detect clipped alignments indicative of regional and structural errors that could be computed into R/S- Assembly Quality Indices (AQIs). Our assembly achieved R/S-AQI of 94.7 and 97.4, both meeting reference quality. Rare residual errors were reflected by the low density of SNPs detected as homozygous (1.1/Mbp, 161 total) and heterozygous (6.8/Mbp, 1038 total).

Analysis of the mapping read depth supported full haplotype phasing with a single peak corresponding to the expected haploid 1× depth (Supplemental Fig. S4). To detect putative phase switch errors, we quantified Hi-C paired alignments within- and between-haplotypes. About 99.2% of the Hi-C mappings were between chromosomes contained within one nucleus (within-haplotype links), with 0.8% linking chromosomes contained in separate nuclei (cross-haplotype links) (Table 2; Supplemental Fig. S5). These results demonstrate a highly complete and accurate nuclear-phased genome assembly of *Pst104E*.

ONT cDNA sequencing enables high-quality evidence-guided gene annotations

We aimed to improve on current fungal gene annotations by incorporating extensive long-read cDNA sequencing datasets (Pardo-Palacios et al. 2024). We generated a detailed time course of ONT direct cDNA datasets for *Pst104E* gene annotation. The transcripts were sampled from six conditions with four replicates each (Supplemental Table S2). These included ungerminated (UG)

urediniospores as the dormancy control, and infected wheat leaf tissues at 4-, 6-, 8-, 10- and 12-days post infection (dpi). Principal component analysis (PCA) demonstrated clear clustering of the technical replicates of each sample, with separation of samples representing dormancy (UG), macroscopically asymptomatic (4, 6 and 8 dpi) and symptomatic (10 and 12 dpi) stages (Supplemental Fig. S6). We also complemented gene annotation with multiple publicly available Illumina RNA-seq datasets (Supplemental Table S3) (Dobon et al. 2016; Schwessinger et al. 2018; Zhao et al. 2021). In total, we annotated 15,142 protein-coding genes on haplotype A and 14,938 on haplotype B, improving the complete BUSCO score to 94.8%. Functional annotation identified ~15% of the gene models that encode secreted proteins without predicted transmembrane domain.

ONT direct cDNA sequencing has demonstrated quantitative power consistent with short-read RNA-seq, making it suitable for differential expression analysis (Pardo-Palacios et al. 2024; Grünberger et al. 2022; Sessegolo et al. 2019). We performed differential expression analysis with our ONT cDNA datasets to identify candidate *Avr* effector genes. We searched for secretome genes that are upregulated early in wheat infection (4, 6, and 8 dpi) relative to UG, as their functions likely correlate with pathogenesis. A total of 1,318 secretome genes were found to be upregulated early during infection. These were shortlisted to hemizygous genes (single copy) for reduced functional redundancy, resulting in 97 high-priority candidates for future functional validation (Supplemental Table S4).

Transposable element annotations

Transposable elements (TEs) are major components of genomes of many basidiomycetes (Castanera et al. 2017; Corre et al. 2025). TE annotations for our *Pst104E* assembly revealed that both haploid genomes shared similar TE content and composition, covering 44.51% of the genome space (Table 1). This was consistent with findings in other published *Pst* genomes (Schwessinger et al. 2018; Zheng et al. 2013; Schwessinger et al. 2022). Class I (retrotransposons) and Class II (DNA transposons) accounted for 15% and 18.7% of the genome respectively (Supplemental Table S5).

LTRs comprised the most abundant retrotransposons, predominated by the Ty3/Gypsy. Terminal inverted repeats (TIRs) represented the majority of DNA transposons.

***Pst* centromeres are highly diverse and enriched with LTR retrotransposons**

We set out to identify *Pst104E*'s centromeric regions and analysed their sequence composition. Basidiomycete centromeres are often characterised by hypermethylated TE-rich and gene-poor regions (Guin et al. 2020). We estimated their positions from the Hi-C heatmap based on the strong inter-chromosomal contact peaks ("bowtie" shapes) caused by *Rab1* centromere clustering (Fig. 1B) (Varoquaux et al. 2015; Muller et al. 2019). These interaction sites were enriched for TEs and overlapped with gene-sparse regions spanning hundreds of kilobases (Fig. 1A, 2A). Using ONT-derived methylation data, we found substantially more methylated CpGs within all 36 inferred centromeric regions, *Cen1A* to *Cen18B* (94.9%–98.3%), than in non-centromeric regions (32.9%–45.3%) (Fig. 2B).

Pst104E centromere sizes varied ~2.5-fold between 210 and 538 kbp (mean 304 kbp; Supplemental Table S6), categorising them as "large regional" centromeres which are known to support multiple spindle microtubule attachments (Yadav et al. 2018a; Bodor et al. 2014). Most homologous chromosomes shared similar centromere lengths with differences under 1.4-fold, while *Cen5A* stood out being double the length of *Cen5B* (Fig. 2C). Pairwise alignments of centromeric regions revealed varying levels of macro-collinearity between haplotypes, ranging from almost complete (e.g., *Cen3*, 12, 13 and 18) to negligible synteny (e.g., *Cen2*, 5, 15 and 17) (Fig. 2D). Given the high TE density, such a diverse range in centromeric sequence conservation prompted us to examine their TE composition in search of elements possibly linked to centromere function.

We conducted permutation tests to determine the enrichment of abundant TE superfamilies within each *Pst* centromere compared to non-centromeric regions (Fig. 2E; Supplemental Table S7). Most centromeres (25 out of 36) were found to be significantly enriched for one to two TE superfamilies that belonged to retro- and/or DNA transposons (Fig. 2E). Of these, 19 centromeres were

significantly enriched for Ty3/Gypsy LTRs, occasionally co-colonised by TIRs. Given the balanced representation of retro- and DNA transposons throughout the genome, this finding suggests that retrotransposons might have a more prominent role in *Pst* centromere formation than DNA transposons.

Centromeres contain a single putative kinetochore attachment site

It is unclear if rust fungi have one or multiple kinetochore attachment sites given they have “large regional” centromeres (Fig. 2) (Yadav et al. 2018a; Sperschneider et al. 2021). Kinetochores typically assemble at a hypomethylated stretch of DNA embedded within the centromere termed the “centromere dip region” (CDR) which is marked by the centromere-specific histone variant CENP-A (Logsdon et al. 2024; Sundararajan and Straight 2022; Akiyoshi 2019). We examined the CpG methylation pattern along the *Pst*104E centromeres to locate CDRs as a qualitative proxy for potential kinetochore sites. Throughout all centromeres, we consistently observed a single methylation depletion “valley” spanning 24.8 kbp on average, which is characteristic of CDR (Fig. 3A; Supplemental Fig. S7; Supplemental Table S6). The CDR lengths were similar between haplotypes (Fig. 3B). The positioning of CDRs were generally conserved (Fig. 3C, 2D). The two notable exceptions were *Cen3* and *Cen13*, whose CDR haplotypes were placed further than 20% of their centromere lengths apart (Fig. 3C).

We next investigated whether the CDR signals were linked to increased AT content as reported for other fungi (Narayanan et al. 2024; Yadav et al. 2018a; Sankaranarayanan et al. 2020). No differences in AT content were detected when comparing CDRs with centromeric or non-centromeric regions (Supplemental Fig. S8). The only exception was *Cen3B* CDR, which has an elevated AT content of 59.8% when compared to the rest of the centromere and the overall genome average of 55.6%. Close investigation revealed two nearly identical AT-rich copies of a Ty3/Gypsy retrotransposon family (hapB-B-G1437-Map9), together covering 91% of *Cen3B* CDR. Sequence alignment revealed that the corresponding syntenic region on *Cen3A* contained only one copy of

this TE family which was not involved in CDR formation (Fig. 3D). Consistently, this single copy on *Cen3A* was highly methylated on its CpG sites (93.8%) while the two TE copies on *Cen3B* CDR were lowly methylated (34.7% and 49.8%) (Fig. 3E). This was despite over 99.5% identity shared by the three TE copies. These results highlight that centromere and kinetochore attachment site formation are not solely driven by primary DNA sequence composition.

Nucleus-specific variations in the rDNA arrays

We investigated the rDNA composition in *Pst104E* to better understand its dynamics in the context of a dikaryotic genome with physical separation of the haploid genomes. *Pst104E* has a single rDNA cluster per nuclear genome on the q-arm of Chr13 (Fig. 1A). Both haplotypes were incompletely assembled with a gap at the rDNA cluster indicating an underrepresentation of this complex locus (Supplemental Fig. S3). The assembled rDNA copies were oriented with transcription directed away from the centromere and towards the telomere (Fig. 4A). We reconstructed the canonical rDNA repeat (see Methods) which included the 45S transcription unit (18S, 5.8S and 25S rRNA genes) separated by two internal transcribed spacers (ITS1 and ITS2), followed by two intergenic spacers (IGS1 and IGS2) and a 5S rRNA gene that is barely transcribed in between (Fig. 4B; Supplemental Fig. S9). We defined 18S and IGS2 as the start and end of an rDNA unit for the subsequent analysis.

Detailed analyses of sequencing reads revealed two major rDNA subtypes (#1 and #2) supported by substantial read coverage at near-equal frequencies. The estimated copy number ranged from 206 to 212 copies for rDNA subtype #1, and 223 to 228 copies for subtype #2, totalling ~434 copies (Fig. 4C). The two subtypes were 9,343 bp and 9,711 bp in length, with polymorphisms contained in the repeats nested within IGS1 and IGS2 (96.1% sequence identity; Supplemental Fig. S10). We also identified twelve variants of the two subtypes (#1.1-1.9 and #2.1-2.3) based on low-frequency SNP analysis (Supplemental Fig. S11; Supplemental Table S8). Most of these SNPs occurred in the

18S gene. The subtype variants #1.1-1.9 and #2.1-2.3 collectively accounted for ~23% of the total copy number, suggesting incomplete rDNA homogenisation (Supplemental Table S8).

Given the long clonal history of *Pst104E* (Schwessinger et al. 2020), the individual haplotypes are expected to have been stably inherited in separate nuclei without karyogamy or meiotic crossovers. We therefore hypothesised that the sequence variations in the two dominant rDNA subtypes might be nucleus-specific. To test this, we took advantage of Hi-C read pairs that contained rDNA subtype-specific *k*-mers and asked where the alternate read mate pair mapped (Fig. 4D). About 94.3% of mates associated with rDNA subtype #1 mapped to haplotype B, whereas 90.2% of those associated with subtype #2 mapped to haplotype A (Fig. 4E); the remaining cross-haplotype links were likely technical noises due to some less informative subtype-specific *k*-mers (such as those derived from the IGS2 minisatellite-like repeats) and sequencing artefacts. The control procedure, which involved aligning subtype-nonspecific rDNA Hi-C read pairs, showed a near-equal proportion of Hi-C mates mapped against each haplotype. Together, this implies that each rDNA subtype was associated with a different haplotype and that each nuclear genome contains its own major rDNA subtype array.

Large-scale structural variations driven by transposable elements shape inter-haplotype diversity

The phased *Pst104E* assembly allowed us to assess its inter-haplotype structural variations (SVs). The haplotypes were 78.8% syntenic (Fig. 5A), with 60.2 Mbp of haplotype A (78.0%) and 60.1 Mbp of haplotype B (79.6%) identified as highly continuous syntenic blocks (Fig. 5B). All identified SV types (duplications, large indels, translocations and inversions) collectively occupied about 11% and 9% of the total lengths of haplotypes A and B, respectively (Fig. 5B; Supplemental Table S9). About 10% of each haplotype's sequence was not alignable and was therefore hemizygous. Syntenic regions had the largest average length with most ranging 10–100 kb. The different SV types had various length distributions mostly centred around 1–10 kb (Fig. 5B). We used permutations to test if

the SVs and their 2 kbp flanking regions were enriched for specific genomic features. This revealed a significant enrichment of TEs across all analysed SV types, especially duplications, possibly due to the replicative mechanism of retrotransposons (Fig. 5C; Supplemental Table S10). In contrast, protein-coding genes were depleted at SVs.

Next, we assessed the impact of inter-haplotype variation on genes by analysing their sequence conservation and synteny (Fig. 5D). We robustly identified 3,391 and 3,214 hemizygous genes unique to either haplotype A or B, respectively. A total of 21,744 genes shared at least one homolog in the alternative haplotype. We analysed one-to-one gene pairs and performed codon-aware alignment to compute their divergence values at synonymous and nonsynonymous sites (Supplemental Table S11), identifying 2,730 homozygous pairs. The remaining 8,122 displayed divergence greater than zero and were therefore defined as heterozygous biallelic pairs.

Allele-specific expression is correlated with gene body methylation and enriched for genes encoding effector/secreted proteins

We next investigated if any of the heterozygous biallelic genes displayed allele-specific expression (ASE) in any condition using our ONT cDNA datasets.

We first tested if there was an overall gene expression bias at the individual nuclear haplotype level. Both haplotypes displayed balanced expression without evidence of nuclear dominance (Supplemental Fig. S12); however, an expression bias for haplotype A was detected when considering only the heterozygous biallelic genes (Fig. 6B).

We compared the transcript abundance for each heterozygous allele pair to determine its ASE across all six tested conditions (Fig. 6A; Supplemental Table S12). We classified the ASE status using the following criteria (Shi et al. 2024): (1) no observed expression or no unambiguous transcript mapping at both alleles (NA); (2) no significant difference between alleles with false discovery rate (FDR) adjusted p-value > 0.05 (NS); and (3) significant difference between alleles with adjusted p-value < 0.05, indicating differential ASE. The differential ASE pairs were further

classified based on $\log_2$ fold change (LFC): weak ASE, $|\text{LFC}| < 2$ (Diff0); moderate ASE, $2 \leq |\text{LFC}| < 4$ (Diff2); and strong ASE, $|\text{LFC}| \geq 4$ (Diff4). While most heterozygous genes showed no evidence of ASE, a substantial proportion exhibited ASE in each condition. The UG stage had most ASE pairs (18.7%), followed by 8 to 12 dpi which maintained stable ASE rates (11.1–13.2%). Early infection at 4 dpi had the fewest (1.2%) ASE pairs, potentially due to the low fungal biomass sampled. To reduce noise in the subsequent ASE analyses, we applied a $|\text{LFC}|$ threshold of two to define the ASE set (Diff2 and Diff4), while all other categories were defined as non-ASE, for the remainder of this study. Intersection analysis identified 678 allele pairs (8.3%) that were assigned to Diff2 or Diff4 ASE in at least one condition (Supplemental Fig. S13).

Given that pathogenicity-related effectors typically undergo diversifying selection (Sperschneider et al. 2014; Stukenbrock and McDonald 2009), we hypothesised that ASE might be an additional pathway to drive virulence dynamics. We therefore asked whether secretome genes, including putative effectors, were overrepresented in each condition-specific ASE versus non-ASE set compared to those of evolutionarily conserved BUSCOs. A Fisher's exact test ($\text{FDR} < 5\%$) for every condition consistently showed significant ASE enrichment for secretome genes relative to BUSCOs with odd ratios greater than one (Fig. 6C; contingency tables in Supplemental Table S13).

We explored several epigenetic and genetic factors that potentially underlie the ASE of secretome genes. No association was found between ASE and SVs (Fig. 5C), TE occupancy and allelic sequence divergence (Supplemental Fig. S14, S15; Supplemental Note). We compared the CpG methylation density between secretome alleles along their gene bodies (start to stop codon) and 2 kbp flanking sequences to include proximal *cis*-regulatory elements such as promoters. Significant methylation differences were observed around the start codon between ASE alleles across all *in planta* infection conditions when compared to the non-ASE alleles (Fig. 6D; Supplemental Fig. S16). Higher-expressed alleles were found to be hypomethylated at the start codon windows (1.1–2.5% CpGs methylated), whereas those of their lower-expressed counterparts which were frequently more heavily methylated (19.2–36.1% CpGs methylated). Non-ASE secretome genes showed no

methylation difference between alleles, regardless of their expression levels. These data suggest that DNA methylation imbalance may be involved in secretome ASE.

Discussion

Recent advances in long-read sequencing technologies have made T2T haplotype-phased genome assemblies the new gold standard for eukaryotes, including di- and heterokaryotic fungi such as rust fungi. Here we report a fully nuclear-phased T2T genome assembly for *Pst*, the first reconstructed using high-accuracy ONT duplex sequencing. We show that our ONT-only T2T genome assembly is of comparable or superior quality to recently published PacBio HiFi-based haplotype-resolved T2T assemblies of other di- and heterokaryotic fungi (e.g. Wang et al. 2024; Sperschneider et al. 2023a; Henningsen et al. 2024). With the complete resolution of both nuclear haplotypes in our *Pst* assembly, we were able to uncover novel insights into the impact of dikaryotism on the genome biology of a long-term asexual clonal isolate of this fungus.

Our detailed analyses of *Pst* centromeres show that they adopt the classic *Rab1* configuration with clustering of heterochromatic centromeres (Torres et al. 2023; Xia et al. 2022). This arrangement allowed us to identify large regional centromeres from Hi-C contact hotspots that coincide with hypermethylated, gene-poor genomic signatures. Each centromere has a single potential kinetochore attachment site marked by a hypomethylated pocket known as the CDR (Logsdon and Eichler 2022; Logsdon et al. 2024; Mastroiosa et al. 2024), within otherwise fully CpG-methylated centromeres without signs of elevated AT content. Direct experimental evidence from CENP-A (CenH3) chromatin immunoprecipitation sequencing will be required to validate centromere localisation. Our comparative inter-haplotype analysis revealed that *Pst* centromeres are highly variable in length and sequence, lacking characteristic motifs that define all centromeres. In most cases, the inferred kinetochore site appears to be consistently positioned in homologous chromosomes even considering highly divergent homologous centromeres; and where shifts occur,

we did not detect an associated sequence presence/absence pattern. These findings point to the conclusion that formation of *Pst* centromeres is a sequence-independent process, consistent with observations in many other fungal regional centromeres (Cissé et al. 2024; Schotanus et al. 2015; Roy and Sanyal 2011; Smith et al. 2012; Sperschneider et al. 2021). Most but not all *Pst* centromeres are enriched for LTR retrotransposons, especially the Ty3/Gypsy superfamily. LTR-rich centromeres have also been reported for other pathogenic fungi with large regional type centromeres, such as the closely related stem rust fungus *P. graminis* f. sp. *tritici* (Sperschneider et al. 2021), a human pathogenic yeast *Cryptococcus neoformans* (Yadav et al. 2018b) and an ascomycetous phytopathogen *Verticillium dahliae* (Seidl et al. 2020), but direct roles for LTR retrotransposons in centromere establishment remain unclear. In *C. neoformans*, the RNAi machinery and DNA methylation have been proposed as key epigenetic drivers for centromere identity via suppressing transposition and deleterious recombination among centromeric LTRs (Yadav et al. 2018b). However, the fact that the correlation with LTR enrichment does not hold in some *Pst* centromeres supports the idea that the LTR sequence itself does not define centromeres (Lynch et al. 2010; Guin et al. 2020). Rather, LTRs might be preferentially inserted due to their high proliferative potential which may, in turn, promote RNAi-directed silencing to reinforce centromere formation (Balzano and Giunta 2020). Further studies will be required to elucidate such links in *Pst*.

The dikaryotic configuration of *Pst* prompted us to explore its effect on intraspecific rDNA dynamics under strict clonality. Generally, rDNA tandem repeats are thought to undergo concerted evolution towards sequence homogenisation via repeated homologous recombination, namely unequal crossovers and gene conversion (Garcia et al. 2024; Symonová 2019; Mullis et al. 2020). Because *Pst*104E has genetically distinct nuclei, we hypothesised that its rDNA variants may persist or emerge within individual nuclei in the absence of meiotic exchange. Our results show that each nucleus of *Pst*104E carries a unique array with over 200 repeats, predominated by an rDNA sequence subtype that harbours nucleus-specific variations within both IGS regions. A low proportion of these repeats have diversified through accumulating point mutations, and some appeared to have become fixed. We speculate that such nucleus-specific rDNA subtype

homogeneity might be the consequence of compartmentalised concerted evolution due to the individual inheritance of each nucleus over prolonged clonal history. The intra-array diversification, however, signifies a relaxation of concerted evolution leading to incomplete homogenisation within each nucleus (Wang et al. 2023; Xu et al. 2017). A possible explanation for this could be the reliance on limited non-meiotic homologous recombination (e.g. intrachromosomal or between sister chromatids) which might less efficiently purge the newly spreading variants (Paloi et al. 2022). This parallels previous observations from homokaryotic (i.e. has genetically uniform nuclei) arbuscular mycorrhizal fungi strains (Serghi et al. 2021) which displayed extensive rDNA heterogeneity within each nucleus, consistent with their ancient clonality (Lin et al. 2014; Pawlowska and Taylor 2004). In future, it will be interesting to survey intraspecific rDNA variations in rust isolates that have arisen via recent sexual recombination (Wang et al. 2022; Heitman et al. 2013; Wallen and Perlin 2018), where we expect that divergent rDNA subtypes will be more evenly distributed among nuclei. Dikaryotism therefore presents an excellent opportunity for understanding the rates and dynamics of concerted evolution in fungi.

A fundamental question in genome biology is whether ASE could have functional consequences that lead to phenotypic variability (Cleary and Seoighe 2021; St. Pierre et al. 2022). Here we show that ASE is pervasive in *Pst* and appears to be inversely related to gene body methylation, where lower-expressed alleles display higher levels of CpG methylation. In *Pst*104E, ASE is overrepresented in secretome genes including putative effectors that could be involved in host pathogenesis. Therefore, ASE might present a novel transcriptomic regulatory mechanism to generate effector diversity beyond protein sequence variations (Chen et al. 2017; Salcedo et al. 2017; Ortiz et al. 2022). Expression level polymorphisms between recognised and non-recognised effector alleles may also explain virulence switching. For example, in *P. graminis* f. sp. *tritici*, a virulence allele of *AvrSr27* was expressed at a much lower level than its avirulence allele counterpart; yet when over-expressed *in planta*, recognition took place (Upadhyaya et al. 2021). Similar observations have been recently made in a common rust disease of maize caused by *P. sorghi*. A lowly expressed virulence allele of *AvrRp1-D*, differing by only one amino acid from its avirulence counterpart,

became recognised by the cognate resistance gene *Rp1-D* after co-overexpression *in planta* (Kim et al. 2024). Non-recognition can be therefore due to low expression rather than an inability of the host and pathogen proteins to interact. Future studies will inform whether the ASE of these *P. graminis* f. sp. *tritici* and *P. sorghi* effector genes is linked to changes in gene body methylation. The role of epigenetic regulation of *Avr* gene expression was previously highlighted in the soybean pathogen *Phytophthora sojae*, where the natural silencing of the avirulence gene *Avr3a* resulted in gain-of-virulence (Hale et al. 2023; Qutob et al. 2013). Such epigenetics-mediated ASE may offer a reversible means to selectively silence or “archive” avirulence alleles to escape immune recognition while retaining the unmutated gene.

One important limitation of our study is that our DNA methylation data only reflects dormancy. A previous study in *Zymoseptoria tritici* (Meile et al. 2020) demonstrated that chromatin remodelling via histone modifications, which are highly correlated with DNA methylation in fungi (Rose and Klose 2014; He et al. 2020), can occur during host pathogenesis to de-repress effector expression. To extend our observations, we will need DNA methylation and histone modification data sampled from *Pst* undergoing pathogenesis to determine if chromatin dynamism underpins ASE *in planta*. Flor’s classic “gene-for-gene” hypothesis (Flor 1942) that has shaped our understanding of plant resistance to pathogens since the 1940’s might be an oversimplification, and we may have to consider that differences in effector gene expression underlie disease outcomes in the field.

Methods

Methods for *Pst*104E DNA and RNA extraction, ONT long-read genome and transcriptome sequencing, and Hi-C library sequencing are detailed in Supplemental Methods.

Genome assembly and quality evaluation

Filtered duplex (>10 kbp; Q30; 32×/haplotype) and simplex (>40 kbp; Q10; 117×/haplotype) reads were assembled using Verkko v1.3.1 (Rautiainen et al. 2023). The assembly was scaffolded with Hi-C using Juicer v2.0 (Durand et al. 2016b) and 3D-DNA v180114 (Dudchenko et al. 2017), followed by manual curation (Supplemental Methods). Each haplotype was quality-assessed using assembly statistics, BUSCO v5.5.0 (basidiomycota_odb10) (Manni et al. 2021), Merqury v1.3 (Rhie et al. 2020) and LAI (Ou et al. 2018; Ou and Jiang 2018). CRAQ v1.0.9 (Li et al. 2023b) was launched to compute R/S-AQI inferred from regional and structural errors from clipped alignments that may indicate misjoins. To evaluate phasing quality, HiC-Pro v3.1.0 (Servant et al. 2015) was used to generate contact matrices from Hi-C alignments (MAPQ ≥ 20), which were analysed using scripts from <https://github.com/RunpengLuo/HiC-Analysis> to quantify *cis*- and *trans*-chromosome contacts.

TE and gene annotations

TE and gene annotations were performed for each haplotype separately. We predicted and annotated TEs using the REPET pipeline v3.0 (Flutre et al. 2011; Quesneville et al. 2005). Prior to gene annotation, we filtered host RNA from all transcriptomic datasets by mapping them against the *Pst104E* assembly with minimap2 v2.26 (Li 2018) and retaining only the mappable reads. We then independently processed and assembled the Illumina RNA-seq and the ONT cDNA datasets to generate transcript evidence. For ONT cDNA, reads were trimmed with Porechop_ABI v0.5.0 (Bonenfant et al. 2023), aligned to the dikaryotic assembly with minimap2 in splice-aware mode (-ax splice -ub -G 3000 --secondary=no), and partitioned into haplotype sets. To identify transcript structures from noisy long reads reference-guided and annotation-free, we employed two long-read-tailored tools: StringTie2 (-L -s2 -m50) (Kovaka et al. 2019) for better single-exon transcript discovery, and ESPRESSO v1.4.0 (ESPRESSO_S.pl -Q0) (Gao et al. 2023) for improved splice site detection. Transcript annotations were merged across all samples using StringTie2 (--merge). Gene predictions and functional annotations were performed using funannotate v1.8.15 (Jonathan and Jason 2023). See Supplemental Methods for details.

Differential expression analysis

Transcript abundance was quantified from ONT cDNA spliced alignments using Bambu v3.4.1 (Chen et al. 2023) with isoform discovery mode disabled. The resulting count matrices were imported to DESeq2 v1.38.3 (Love et al. 2014) for analysis. PCA was performed on variance-stabilising transformed read counts to visualise sample clustering. DESeq2 was executed on default settings to identify genes differentially expressed at host infection conditions relative to UG. We considered genes with an FDR-adjusted p-value < 0.05 and $\text{LFC} \geq 2$ to be upregulated. Secretome genes upregulated at 4, 6 or 8 dpi were identified as our preliminary *Avr* effector candidates.

Centromere inference and analysis

Centromere locations were determined based on strong inter-chromosomal interaction signals on Juicebox v2.17.00 Hi-C heatmap (Durand et al. 2016a) and confirmed by analysing ONT-derived DNA methylation data (Supplemental Methods).

CDRs were located via visually selecting the largest hypomethylated region as this single pattern consistently appeared throughout all the inferred centromeres. Relative CDR position was calculated by dividing its midpoint coordinate by centromere length; then compared between haplotypes to detect CDR shifts. High-resolution alignment dotplots were generated in Gepard v2.1 to investigate sequence composition at shifted CDRs (Krumsiek et al. 2007).

Centromeric TE enrichment was analysed with permutation tests using our custom Python script. The rationale behind was to randomly reshuffle locations of target features (e.g. TEs) throughout a given chromosome or genome (5,000 times) to remove their potential biological association with region of interest (e.g. centromere), creating a null distribution of test statistics (e.g. TE coverage difference between centromere versus non-centromere) which enabled a two-tailed hypothesis test for feature enrichment or depletion. P-values were defined as the proportion of permuted values equal to or more extreme than the observed value ($< 5\%$ FDR).

Ribosomal DNA analysis

The canonical rDNA unit was defined by aligning reference ITS, 18S and 5S of *Puccinia* species retrieved from databases including Gold Standard (Eenjes et al. 2022), EukRibo (Berney et al. 2022) and 5SrRNAdb (Szymanski et al. 2016) to distinguish conserved and variable elements. Long rRNA reads were aligned to confirm the transcribed units. Analyses of rDNA subtype variations and copy numbers were conducted on duplex and Illumina read alignments (Supplemental Methods).

The nuclear association of the two dominant rDNA subtypes was tested by analysing rDNA Hi-C reads with a *k*-mer approach. Unique 31-mers of each subtype were identified using UniqueKMER (Chen et al. 2021). Subtype-specific 31-mers, along with their reverse complements, were used to tag rDNA Hi-C reads, whose paired Hi-C mates were fetched from the corresponding R1/R2 read file. Mates were mapped to the dikaryotic assembly with bwa-mem2 (Vasimuddin et al. 2019). Mapping locations and MAPQ scores were extracted with BEDTools *bamtobed* (Quinlan and Hall 2010) for plotting. This was repeated on subtype-nonspecific 31-mers as control. Note if a read is tagged by both subtype-specific and unspecific *k*-mers, it is defined as subtype-specific.

Synteny and SV detection

Whole-genome alignment between haplotypes A and B was conducted with NUCmer (--maxmatch -l 200 -b 500 -c 500). Only alignment blocks with >90% identity were retained. SyRI (Goel et al. 2019) was used to annotate SVs, syntenic and unaligned regions, then visualised with plotsr (Goel and Schneeberger 2022). To analyse features within and nearby SVs, BEDTools *slop* was used to extend SV coordinates by 2 kbp in both directions. Enrichment or depletion for genomic features, including TEs, genes, and subsets such as secretome and ASE-only genes, was statistically assessed through two-tailed permutation tests, as described above.

Identification of hemizygous and heterozygous genes

Protein sequences of all annotated genes were analysed using Proteinortho v6.3.1 (-synteny) (Lechner et al. 2011) to detect homologs between haplotypes. Genes lacking a hit on the alternative haplotype were considered hemizygous candidates. These were further filtered via reciprocal BLASTP (>70% identity; >70% query/subject coverage) to ensure the absence of alleles. One-to-one gene pairs with divergence values greater than zero (computed from codon-aware alignments generated using script “dN_dS_Pst134E.ipynb” (<https://github.com/ZhenyanLuo/codes-used-for-mating-type>) (Luo et al. 2024)) were defined to be heterozygous biallelic.

Allele-specific expression analysis

We reformatted the Bambu gene-level read count matrices to test differential allele expression at heterozygous biallelic genes per condition using DESeq2 (Supplemental Methods). Allele pairs were grouped into different ASE status, as detailed in Results. To assess nuclear dominance, allele read counts were normalised using DESeq2’s median-of-ratio method and transformed into expression levels as $\log_{10}(\text{median of ratio} + 1)$, then averaged across replicates for haplotype comparisons. Overrepresentation of secretome genes in the ASE set (Diff2 and Diff4) relative to BUSCOs was evaluated via a two-sided Fisher’s exact test on 2×2 contingency table constructed per condition (<5% FDR). Methylation differences between ASE and non-ASE secretome alleles at gene-body and flanking regions were analysed using a custom Python script (Supplemental Methods).

Data access

All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1195871. All custom scripts for analyses and figures are available on <https://github.com/ritatam/Pst104EGenomeAnalysis> and in Supplemental Scripts. Genome

assembly, rDNA sequences, gene and TE annotations are available on Zenodo (doi:
[10.5281/zenodo.14885411](https://doi.org/10.5281/zenodo.14885411)).

Competing interest statement

The authors declare no competing interests.

Acknowledgements

R.T. was supported by a Grains Research and Development Corporation Graduate Research Scholarship. This work was supported by an Australian Research Council Future Fellowship to B.S. (FT180100024) and a Discovery Project Grant (DP230100941) to J.P.R. and B.S. This work was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the ANU Merit Allocation Scheme. We would like to acknowledge the contribution of the Plant Pathogen 'Omics Initiative consortium in the generation of data used in this publication. The Initiative is supported by funding from Bioplatforms Australia, enabled by the Commonwealth Government National Collaborative Research Infrastructure Strategy (NCRIS).

Author contributions: R.T., B.S. and J.P.R. designed the project. R.T. led the project and performed most of the formal analyses and visualisation. B.S. and J.P.R. supervised the project. R.T., B.S., J.P.R., and M.M. acquired funding. M.M. conducted RNA extraction and ONT cDNA sequencing. R.L. analysed Hi-C data for evaluating assembly phasing quality. Z.L. identified mating type loci and shared scripts for allele pairing. B.S., A.J. and S.P. contributed to experiments. R.T., M.M., J.P.R., and B.S. wrote the manuscript. All authors have reviewed the final manuscript.

538 References

- 539 Akiyoshi B. 2019. Evolution: A Mosaic-type Centromere in an Early-Diverging Fungus. *Curr Biol* **29**:
540 R1184–R1186.
- 541 Anderson JB, Kohn LM. 2014. Dikaryons, Diploids, and Evolution. In *Sex in Fungi* (eds. J. Heitman,
542 J.W. Kronstad, J.W. Taylor, and L.A. Casselton), pp. 333–348, ASM Press, Washington, DC,
543 USA <http://doi.wiley.com/10.1128/9781555815837.ch20> (Accessed January 29, 2021).
- 544 Balzano E, Giunta S. 2020. Centromeres under Pressure: Evolutionary Innovation in Conflict with
545 Conserved Function. *Genes* **11**: 912.
- 546 Belser C, Baurens F-C, Noel B, Martin G, Cruaud C, Istace B, Yahiaoui N, Labadie K, Hřibová E,
547 Doležel J, et al. 2021. Telomere-to-telomere gapless chromosomes of banana using
548 nanopore sequencing. *Commun Biol* **4**: 1047.
- 549 Berney C, Henry N, Mahé F, Richter DJ, Vargas C de. 2022. EukRibo: a manually curated
550 eukaryotic 18S rDNA reference database to facilitate identification of new diversity.
551 2022.11.03.515105. <https://www.biorxiv.org/content/10.1101/2022.11.03.515105v1>
552 (Accessed November 25, 2024).
- 553 Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, Cleveland DW, Black BE,
554 Shah JV, Jansen LE. 2014. The quantitative architecture of centromeric chromatin ed. J.
555 Pines. *eLife* **3**: e02137.
- 556 Castanera R, Borgognone A, Pisabarro AG, Ramírez L. 2017. Biology, dynamics, and applications
557 of transposable elements in basidiomycete fungi. *Appl Microbiol Biotechnol* **101**: 1337–1350.
- 558 Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, Breen S, Dong C, Xu B, Zhang
559 X, et al. 2017. Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50
560 resistance in wheat. *Science* **358**: 1607–1610.
- 561 Chen S, He C, Li Y, Li Z, Melançon CE III. 2021. A computational toolset for rapid identification of
562 SARS-CoV-2, other viruses and microorganisms from sequencing data. *Brief Bioinform* **22**:
563 924–935.
- 564 Chen Y, Sim A, Wan YK, Yeo K, Lee JJX, Ling MH, Love MI, Göke J. 2023. Context-aware transcript
565 quantification from long-read RNA-seq data with Bambu. *Nat Methods* **20**: 1187–1195.
- 566 Cissé OH, Curran SJ, Folco HD, Liu Y, Bishop L, Wang H, Fischer ER, Davis AS, Combs C, Thapar
567 S, et al. 2024. Regional centromere configuration in the fungal pathogens of the
568 *Pneumocystis* genus ed. L.M. Weiss. *mBio* **15**: e03185-23.
- 569 Cleary S, Seoighe C. 2021. Perspectives on Allele-Specific Expression. *Annu Rev Biomed Data Sci*
570 **4**: 101–122.
- 571 Corre E, Morin E, Duplessis S, Lorrain C. 2025. Ancestral and recent bursts of transposition shaped
572 the massive genomes of plant pathogenic rust fungi. 2025.01.10.632365.
573 <https://www.biorxiv.org/content/10.1101/2025.01.10.632365v1> (Accessed January 23, 2025).
- 574 Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, Sakthikumar S, Song X, Adiconis
575 X, Fan L, et al. 2017. Comparative Analysis Highlights Variable Genome Content of Wheat
576 Rusts and Divergence of the Mating Loci. *G3 GenesGenomesGenetics* **7**: 361–376.

- 577 Dobon A, Bunting DCE, Cabrera-Quio LE, Uauy C, Saunders DGO. 2016. The host-pathogen
578 interaction between wheat and yellow rust induces temporally coordinated waves of gene
579 expression. *BMC Genomics* **17**: 380.
- 580 Du Z, Peng Y, Zhang G, Chen L, Jiang S, Kang Z, Zhao J. 2023. Direct Evidence Demonstrates that
581 *Puccinia striiformis* f. sp. *tritici* Infects Susceptible Barberry to Complete Sexual Cycle in
582 Autumn. *Plant Dis* **107**: 771–783.
- 583 Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I,
584 Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-
585 C yields chromosome-length scaffolds. *Science* **356**: 92–95.
- 586 Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a.
587 Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell*
588 *Syst* **3**: 99–101.
- 589 Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer
590 Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**:
591 95–98.
- 592 Eenjes T, Hu Y, Irinyi L, Hoang MTV, Smith LM, Linde CC, Milgate AW, Meyer W, Stone EA, Rathjen
593 JP, et al. 2022. Linked machine learning classifiers improve species classification of fungi
594 when using error-prone long-reads on extended metabarcodes. 2021.05.01.442223.
595 <https://www.biorxiv.org/content/10.1101/2021.05.01.442223v4> (Accessed November 25,
596 2024).
- 597 Ferguson S, Jones A, Murray K, Andrew R, Schwessinger B, Borevitz J. 2024. Plant genome
598 evolution in the genus *Eucalyptus* is driven by structural rearrangements that promote
599 sequence divergence. *Genome Res* **34**: 606–619.
- 600 Flor HH. 1942. Inheritance of pathogenicity in *Melampsora lini*. *Phytopathol* **32**: 653-669.
- 601 Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering Transposable Element
602 Diversification in De Novo Annotation Approaches ed. Y. Xu. *PLoS ONE* **6**: e16526.
- 603 Fultz D, McKinlay A, Enganti R, Pikaard CS. 2023. Sequence and epigenetic landscapes of active
604 and silent nucleolus organizer regions in *Arabidopsis*. *Sci Adv* **9**: eadj4509.
- 605 Garcia S, Kovarik A, Maiwald S, Mann L, Schmidt N, Pascual-Díaz JP, Vitales D, Weber B, Heitkam
606 T. 2024. The Dynamic Interplay Between Ribosomal DNA and Transposable Elements: A
607 Perspective From Genomics and Cytogenetics ed. M. Barlow. *Mol Biol Evol* **41**: msae025.
- 608 Gehrman T, Pelkmans JF, Ohm RA, Vos AM, Sonnenberg ASM, Baars JJP, Wösten HAB, Reinders
609 MJT, Abeel T. 2018. Nucleus-specific expression in the multinuclear mushroom-forming
610 fungus *Agaricus bisporus* reveals different nuclear regulatory programs. *Proc Natl Acad Sci*
611 **115**: 4429–4434.
- 612 Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood
613 CW, Slot JC, Lopez-Nicora HD, et al. 2022. Giant *Starship* Elements Mobilize Accessory
614 Genes in Fungal Genomes ed. A. Larracuenta. *Mol Biol Evol* **39**: msac109.
- 615 Goel M, Schneeberger K. 2022. plotsr: visualizing structural similarities and rearrangements
616 between multiple genomes. *Bioinformatics* **38**: 2922–2926.

- 617 Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local
618 sequence differences from whole-genome assemblies. *Genome Biol* **20**: 277.
- 619 Grünberger F, Ferreira-Cerca S, Grohmann D. 2022. Nanopore sequencing of RNA and cDNA
620 molecules in *Escherichia coli*. *RNA* **28**: 400–417.
- 621 Guin K, Sreekumar L, Sanyal K. 2020. Implications of the Evolutionary Trajectory of Centromeres in
622 the Fungal Kingdom. *Annu Rev Microbiol* **74**: 835–853.
- 623 Hale B, Brown E, Wijeratne A. 2023. An updated assessment of the soybean–*Phytophthora sojae*
624 pathosystem. *Plant Pathol* **72**: 843–860.
- 625 Hartmann FE. 2022. Using structural variants to understand the ecological and evolutionary
626 dynamics of fungal plant pathogens. *New Phytol* **234**: 43–49.
- 627 He C, Zhang Z, Li B, Tian S. 2020. The Pattern and Function of DNA Methylation in Fungal Plant
628 Pathogens. *Microorganisms* **8**: 227.
- 629 Heitman J, Sun S, James TY. 2013. Evolution of fungal sexual reproduction. *Mycologia* **105**: 1–27.
- 630 Henningsen EC, Lewis D, Nazareno E, Huang Y-F, Steffenson BJ, Boesen B, Kianian SF, Stone E,
631 Dodds PN, Sperschneider J, et al. 2024. A high-resolution haplotype pangenome uncovers
632 somatic hybridization, recombination and intercontinental migration in oat crown rust.
633 2024.03.27.583983. <https://www.biorxiv.org/content/10.1101/2024.03.27.583983v2>
634 (Accessed July 17, 2024).
- 635 Hu G, Grover CE, Arick MA, Liu M, Peterson DG, Wendel JF. 2021. Homoeologous gene
636 expression and co-expression network analyses and evolutionary inference in allopolyploids.
637 *Brief Bioinform* **22**: 1819–1835.
- 638 James TY, Stajich JE, Hittinger CT, Rokas A. 2020. Toward a Fully Resolved Fungal Tree of Life.
639 *Annu Rev Microbiol* **74**: 291–313.
- 640 Kim S-B, Kim K-T, In S, Jaiswal N, Lee G-W, Jung S, Rogers A, Gómez-Trejo LF, Gautam S, Helm
641 M, et al. 2024. Use of the *Puccinia sorghi* haustorial transcriptome to identify and
642 characterize AvrRp1-D recognized by the maize Rp1-D resistance protein. *PLOS Pathog* **20**:
643 e1012662.
- 644 Kovaka S, Zimin AV, Perteza GM, Razaghi R, Salzberg SL, Perteza M. 2019. Transcriptome assembly
645 from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**: 278.
- 646 Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on
647 genome scale. *Bioinformatics* **23**: 1026–1028.
- 648 Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: Detection of
649 (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**: 124.
- 650 Li C, Qiao L, Lu Y, Xing G, Wang X, Zhang G, Qian H, Shen Y, Zhang Y, Yao W, et al. 2023a.
651 Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome
652 Evolution in Pucciniales ed. L.P. Burbank. *Microbiol Spectr* **11**: e02828-22.
- 653 Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C, Miller ME,
654 Silverstein KAT, Henningsen E, et al. 2019. Emergence of the Ug99 lineage of the wheat
655 stem rust pathogen through somatic hybridisation. *Nat Commun* **10**: 5068.

- 656 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.
- 657 Li K, Xu P, Wang J, Yi X, Jiao Y. 2023b. Identification of errors in draft genome assemblies at single-
658 nucleotide resolution for quality assessment and improvement. *Nat Commun* **14**: 6556.
- 659 Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, et al.
660 2014. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an
661 Endomycorrhizal Fungus. *PLOS Genet* **10**: e1004078.
- 662 Logsdon GA, Eichler EE. 2022. The Dynamic Structure and Rapid Evolution of Human Centromeric
663 Satellite DNA. *Genes* **14**: 92.
- 664 Logsdon GA, Rozanski AN, Ryabov F, Potapova T, Shepelev VA, Catacchio CR, Porubsky D, Mao
665 Y, Yoo D, Rautiainen M, et al. 2024. The variation and evolution of complete human
666 centromeres. *Nature*. <https://www.nature.com/articles/s41586-024-07278-3> (Accessed April
667 23, 2024).
- 668 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-
669 seq data with DESeq2. *Genome Biol* **15**: 550.
- 670 Luo Z, McTaggart A, Schwessinger B. 2024. Genome biology and evolution of mating-type loci in
671 four cereal rust fungi ed. T. Giraud. *PLOS Genet* **20**: e1011207.
- 672 Lynch DB, Logue ME, Butler G, Wolfe KH. 2010. Chromosomal G + C Content Evolution in Yeasts:
673 Systematic Interspecies Differences, and GC-Poor Troughs at Centromeres. *Genome Biol
674 Evol* **2**: 572–583.
- 675 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and
676 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring
677 of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**: 4647–4654.
- 678 Mastroi Rosa FK, Rozanski AN, Harvey WT, Knuth J, Garcia G, Munson KM, Hoekzema K, Logsdon
679 GA, Eichler EE. 2024. Complete chromosome 21 centromere sequences from a Down
680 syndrome family reveal size asymmetry and differences in kinetochore attachment. *bioRxiv*
681 2024.02.25.581464.
- 682 Meile L, Peter J, Puccetti G, Alassimone J, McDonald BA, Sánchez-Vallet A. 2020. Chromatin
683 Dynamics Contribute to the Spatiotemporal Expression Pattern of Virulence Genes in a
684 Fungal Plant Pathogen. *mBio* **11**: e02343-20.
- 685 Muller H, Gil J, Drinnenberg IA. 2019. The Impact of Centromeres on Spatial Genome Architecture.
686 *Trends Genet* **35**: 565–578.
- 687 Mullis A, Lu Z, Zhan Y, Wang T-Y, Rodriguez J, Rajeh A, Chatrath A, Lin Z. 2020. Parallel Concerted
688 Evolution of Ribosomal Protein Genes in Fungi and Its Adaptive Significance. *Mol Biol Evol*
689 **37**: 455–468.
- 690 Narayanan A, Reza MdH, Sanyal K. 2024. Behind the scenes: Centromere-driven genomic
691 innovations in fungal pathogens. *PLOS Pathog* **20**: e1012080.
- 692 Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N,
693 Uralsky L, Gershman A, et al. 2022. The complete sequence of a human genome. *Science*
694 **376**: 44–53.

- 695 Ortiz D, Chen J, Outram MA, Saur IML, Upadhyaya NM, Mago R, Ericsson DJ, Cesari S, Chen C,
696 Williams SJ, et al. 2022. The stem rust effector protein AvrSr50 escapes Sr50 recognition by
697 a substitution in a single surface-exposed residue. *New Phytol* **234**: 592–606.
- 698 Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index
699 (LAI). *Nucleic Acids Res.* [https://academic.oup.com/nar/advance-](https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky730/5068908)
700 [article/doi/10.1093/nar/gky730/5068908](https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky730/5068908) (Accessed April 5, 2024).
- 701 Ou S, Jiang N. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of
702 Long Terminal Repeat Retrotransposons. *Plant Physiol* **176**: 1410–1422.
- 703 Paloi S, Luangsa-ard JJ, Mhuantong W, Stadler M, Kobmoo N. 2022. Intragenomic variation in
704 nuclear ribosomal markers and its implication in species delimitation, identification and
705 barcoding in fungi. *Fungal Biol Rev* **42**: 1–33.
- 706 Pardo-Palacios FJ, Wang D, Reese F, Diekhans M, Carbonell-Sala S, Williams B, Loveland JE, De
707 María M, Adams MS, Balderrama-Gutierrez G, et al. 2024. Systematic assessment of long-
708 read RNA-seq methods for transcript identification and quantification. *Nat Methods* **21**:
709 1349–1363.
- 710 Pawlowska TE, Taylor JW. 2004. Organization of genetic variation in individuals of arbuscular
711 mycorrhizal fungi. *Nature* **427**: 733–737.
- 712 Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005.
713 Combined Evidence Annotation of Transposable Elements in Genome Sequences. *PLoS*
714 *Comput Biol* **1**: e22.
- 715 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
716 *Bioinformatics* **26**: 841–842.
- 717 Qutob D, Patrick Chapman B, Gijzen M. 2013. Transgenerational gene silencing causes gain of
718 virulence in a plant pathogen. *Nat Commun* **4**: 1349.
- 719 Ramírez Nasto L, Pérez Garrido MG, Castanera Andrés R, Santoyo Santos F, Pisabarro de Lucas
720 G. 2011. Basidiomycetes telomeres: a bioinformatics approach. [https://academic-](https://academic.e.unavarra.es/handle/2454/35292)
721 [e.unavarra.es/handle/2454/35292](https://academic.e.unavarra.es/handle/2454/35292) (Accessed September 13, 2024).
- 722 Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, Eichler EE, Phillippy AM,
723 Koren S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat*
724 *Biotechnol* 1–9.
- 725 Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness,
726 and phasing assessment for genome assemblies. *Genome Biol* **21**: 245.
- 727 Rodriguez-Algaba J, Hovmøller MS, Schulz P, Hansen JG, Lezáun JA, Joaquim J, Randazzo B,
728 Czembor P, Zemeca L, Slikova S, et al. 2022. Stem rust on barberry species in Europe: Host
729 specificities and genetic diversity. *Front Genet* **13**.
730 <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.988031/full>
731 (Accessed December 12, 2024).
- 732 Rose NR, Klose RJ. 2014. Understanding the relationship between DNA methylation and histone
733 lysine methylation. *Biochim Biophys Acta Gene Regul Mech* **1839**: 1362–1372.
- 734 Roy B, Sanyal K. 2011. Diversity in Requirement of Genetic and Epigenetic Factors for Centromere
735 Function in Fungi. *Eukaryot Cell* **10**: 1384–1395.

- 736 Salcedo A, Rutter W, Wang S, Akhunova A, Bolus S, Chao S, Anderson N, De Soto MF, Rouse M,
737 Szabo L, et al. 2017. Variation in the *AvrSr35* gene determines *Sr35* resistance against
738 wheat stem rust race Ug99. *Science* **358**: 1604–1606.
- 739 Sankaranarayanan SR, Ianiri G, Coelho MA, Reza MH, Thimmappa BC, Ganguly P, Vadnala RN,
740 Sun S, Siddharthan R, Tellgren-Roth C, et al. 2020. Loss of centromere function drives
741 karyotype evolution in closely related *Malassezia* species eds. K. Struhl, W.-D. Heyer, W.-D.
742 Heyer, and P.B. Talbert. *eLife* **9**: e53944.
- 743 Schmidt-Dannert C. 2016. Biocatalytic Portfolio of Basidiomycota. *Curr Opin Chem Biol* **31**: 40–49.
- 744 Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock
745 EH. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria*
746 *tritici* distinguish core and accessory chromosomes. *Epigenetics Chromatin* **8**: 41.
- 747 Schwessinger B, Chen Y-J, Tien R, Vogt JK, Sperschneider J, Nagar R, McMullan M, Sicheritz-
748 Ponten T, Sørensen CK, Hovmøller MS, et al. 2020. Distinct Life Histories Impact Dikaryotic
749 Genome Evolution in the Rust Fungus *Puccinia striiformis* Causing Stripe Rust in Wheat ed.
750 J. Stajich. *Genome Biol Evol* **12**: 597–617.
- 751 Schwessinger B, Jones A, Albekaa M, Hu Y, Mackenzie A, Tam R, Nagar R, Milgate A, Rathjen JP,
752 Periyannan S. 2022. A Chromosome Scale Assembly of an Australian *Puccinia striiformis* f.
753 sp. *tritici* Isolate of the PstS1 Lineage. *Mol Plant-Microbe Interactions*® **35**: 293–296.
- 754 Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM, Dodds PN,
755 Figueroa M, Park RF, Rathjen JP. 2018. A Near-Complete Haplotype-Phased Genome of the
756 Dikaryotic Wheat Stripe Rust Fungus *Puccinia striiformis* f. sp. *tritici* Reveals High
757 Interhaplotype Diversity ed. A. Di Pietro. *mBio* **9**: e02275-17.
- 758 Seidl MF, Kramer HM, Cook DE, Fiorin GL, Van Den Berg GCM, Faino L, Thomma BPHJ. 2020.
759 Repetitive Elements Contribute to the Diversity and Evolution of Centromeres in the Fungal
760 Genus *Verticillium* ed. J. Heitman. *mBio* **11**: e01714-20.
- 761 Serghi EU, Kokkoris V, Cornell C, Dettman J, Stefani F, Corradi N. 2021. Homo- and Dikaryons of
762 the Arbuscular Mycorrhizal Fungus *Rhizophagus irregularis* Differ in Life History Strategy.
763 *Front Plant Sci* **12**. [https://www.frontiersin.org/journals/plant-](https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.715377/full)
764 [science/articles/10.3389/fpls.2021.715377/full](https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.715377/full) (Accessed December 4, 2024).
- 765 Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E.
766 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**:
767 259.
- 768 Sessegolo C, Cruaud C, Da Silva C, Cologne A, Dubarry M, Derrien T, Lacroix V, Aury J-M. 2019.
769 Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA
770 molecules. *Sci Rep* **9**: 14908.
- 771 Shao L, Xing F, Xu C, Zhang Q, Che J, Wang X, Song J, Li X, Xiao J, Chen L-L, et al. 2019.
772 Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the
773 genetic basis of heterosis. *Proc Natl Acad Sci* **116**: 5653–5658.
- 774 Shi T-L, Jia K-H, Bao Y-T, Nie S, Tian X-C, Yan X-M, Chen Z-Y, Li Z-C, Zhao S-W, Ma H-Y, et al.
775 2024. High-quality genome assembly enables prediction of allele-specific gene expression in
776 hybrid poplar. *Plant Physiol* **195**: 652–670.

- 777 Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. 2012. Centromeres of filamentous
778 fungi. *Chromosome Res* **20**: 635–656.
- 779 Sperschneider J, Hewitt T, Lewis DC, Periyannan S, Milgate AW, Hickey LT, Mago R, Dodds PN,
780 Figueroa M. 2023a. Nuclear exchange generates population diversity in the wheat leaf rust
781 pathogen *Puccinia triticina*. *Nat Microbiol* **8**: 2130–2141.
- 782 Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong C, Upadhyaya NM, Mago R, Hu Y,
783 Figueroa M, et al. 2021. The stem rust fungus *Puccinia graminis* f. sp. *tritici* induces
784 centromeric small RNAs during late infection that are associated with genome-wide DNA
785 methylation. *BMC Biol* **19**: 203.
- 786 Sperschneider J, Yildirim G, Rizzi YS, Malar C M, Mayrand Nicol A, Sorwar E, Villeneuve-Laroche M,
787 Chen ECH, Iwasaki W, Brauer EK, et al. 2023b. Arbuscular mycorrhizal fungi heterokaryons
788 have two nuclear populations with distinct roles in host–plant interactions. *Nat Microbiol* **8**:
789 2142–2153.
- 790 Sperschneider J, Ying H, Dodds PN, Gardiner DM, Upadhyaya NM, Singh KB, Manners JM, Taylor
791 JM. 2014. Diversifying selection in the wheat stem rust fungus acts predominantly on
792 pathogen-associated gene families and reveals candidate effectors. *Front Plant Sci* **5**.
793 <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2014.00372/full>
794 (Accessed November 12, 2024).
- 795 St. Pierre CL, Macias-Velasco JF, Wayhart JP, Yin L, Semenkovich CF, Lawson HA. 2022. Genetic,
796 epigenetic, and environmental mechanisms govern allele-specific gene expression. *Genome*
797 *Res* **32**: 1042–1057.
- 798 Stukenbrock EH, McDonald BA. 2009. Population Genetics of Fungal and Oomycete Effectors
799 Involved in Gene-for-Gene Interactions. *Mol Plant-Microbe Interactions*® **22**: 371–380.
- 800 Sun H, Jiao W-B, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukat C, Huettel B,
801 Schneeberger K. 2022. Chromosome-scale and haplotype-resolved genome assembly of a
802 tetraploid potato cultivar. *Nat Genet* **54**: 342–348.
- 803 Sundararajan K, Straight AF. 2022. Centromere Identity and the Regulation of Chromosome
804 Segregation. *Front Cell Dev Biol* **10**: 914249.
- 805 Symonová R. 2019. Integrative rDNAomics—Importance of the Oldest Repetitive Fraction of the
806 Eukaryote Genome. *Genes* **10**: 345.
- 807 Szymanski M, Zielezinski A, Barciszewski J, Erdmann VA, Karlowski WM. 2016. 5SRNadb: an
808 information resource for 5S ribosomal RNAs. *Nucleic Acids Res* **44**: D180–D183.
- 809 Thach T, Ali S, de Vallavieille-Pope C, Justesen AF, Hovmøller MS. 2016. Worldwide population
810 structure of the wheat rust fungus *Puccinia striiformis* in the past. *Fungal Genet Biol FG B*
811 **87**: 1–8.
- 812 Tian Y, Thrimawithana A, Ding T, Guo J, Gleave A, Chagné D, Ampomah-Dwamena C, Ireland HS,
813 Schaffer RJ, Luo Z, et al. 2022. Transposon insertions regulate genome-wide allele-specific
814 expression and underpin flower colour variations in apple (*Malus* spp.). *Plant Biotechnol J*
815 **20**: 1285–1297.
- 816 Torres DE, Reckard AT, Klocko AD, Seidl MF. 2023. Nuclear genome organization in fungi: from
817 gene folding to Rabl chromosomes. *FEMS Microbiol Rev* **47**: fuad021.

- 818 Upadhyaya NM, Mago R, Panwar V, Hewitt T, Luo M, Chen J, Sperschneider J, Nguyen-Phuc H,
819 Wang A, Ortiz D, et al. 2021. Genomics accelerated isolation of a new stem rust avirulence
820 gene–wheat resistance gene pair. *Nat Plants* **7**: 1220–1228.
- 821 Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert J-P, Noble WS. 2015.
822 Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids*
823 *Res* **43**: 5331–5339.
- 824 Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient Architecture-Aware Acceleration of BWA-
825 MEM for Multicore Systems. In *2019 IEEE International Parallel and Distributed Processing*
826 *Symposium (IPDPS)*, pp. 314–324 <https://ieeexplore.ieee.org/document/8820962> (Accessed
827 November 25, 2024).
- 828 Vasquez-Gross H, Kaur S, Epstein L, Dubcovsky J. 2020. A haplotype-phased genome of wheat
829 stripe rust pathogen *Puccinia striiformis* f. sp. *tritici*, race PST-130 from the Western USA ed.
830 D. Perovic. *PLOS ONE* **15**: e0238611.
- 831 Wallen RM, Perlin MH. 2018. An Overview of the Function and Maintenance of Sexual Reproduction
832 in Dikaryotic Fungi. *Front Microbiol* **9**.
833 <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00503/full> (Accessed February 25,
834 2021).
- 835 Wang J, Xu Y, Peng Y, Wang Y, Kang Z, Zhao J. 2024. A fully haplotype-resolved and nearly gap-
836 free genome assembly of wheat stripe rust fungus. *Sci Data* **11**: 508.
- 837 Wang J, Zhan G, Tian Y, Zhang Y, Xu Y, Kang Z, Zhao J. 2022. Role of Sexual Reproduction in the
838 Evolution of the Wheat Stripe Rust Fungus Races in China. *Phytopathology*® **112**: 1063–
839 1071.
- 840 Wang W, Zhang X, Garcia S, Leitch AR, Kovařík A. 2023. Intragenomic rDNA variation - the product
841 of concerted evolution, mutation, or something in between? *Heredity* **131**: 179–188.
- 842 Wellings CR. 2007. *Puccinia striiformis* in Australia: a review of the incursion, evolution, and
843 adaptation of stripe rust in the period 1979–2006. *Aust J Agric Res* **58**: 567–575.
- 844 Xia C, Huang L, Huang J, Zhang H, Huang Y, Benhamed M, Wang M, Chen X, Zhang M, Liu T, et al.
845 2022. Folding Features and Dynamics of 3D Genome Architecture in Plant Fungal
846 Pathogens ed. J. Han. *Microbiol Spectr* **10**: e02608-22.
- 847 Xu B, Zeng X-M, Gao X-F, Jin D-P, Zhang L-B. 2017. ITS non-concerted evolution and rampant
848 hybridization in the legume genus *Lespedeza* (Fabaceae). *Sci Rep* **7**: 40057.
- 849 Yadav V, Sreekumar L, Guin K, Sanyal K. 2018a. Five pillars of centromeric chromatin in fungal
850 pathogens ed. D.C. Sheppard. *PLOS Pathog* **14**: e1007150.
- 851 Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman
852 J, Sanyal K. 2018b. RNAi is a critical determinant of centromere evolution in closely related
853 fungi. *Proc Natl Acad Sci* **115**: 3108–3113.
- 854 Zhao J, Duan W, Xu Y, Zhang C, Wang L, Wang J, Tian S, Pei G, Zhan G, Zhuang H, et al. 2021.
855 Distinct Transcriptomic Reprogramming in the Wheat Stripe Rust Fungus During the Initial
856 Infection of Wheat and Barberry. *Mol Plant-Microbe Interact MPMI* **34**: 198–209.

857 Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, Guo J, Zhuang H, Qiu C, Liu J, et al. 2013.
858 High genome heterozygosity and endemic genetic recombination in the wheat stripe rust
859 fungus. *Nat Commun* **4**: 2673.

860 nextgenusfs/funannotate: Eukaryotic Genome Annotation Pipeline.
861 <https://github.com/nextgenusfs/funannotate> (Accessed June 11, 2024).

862

863 **Tables**

864 **Table 1. Summary of assembly statistics and quality metrics of the ONT duplex genome**
865 **assembly of *Pst104E*.**

Statistic	Full dikaryotic	Haplotype A	Haplotype B
Assembly size (bp)	152,328,638	77,128,783	75,199,855
# T2T chromosome	35/36	17/18	18/18
N50 (bp)	4,614,849	4,713,886	4,606,998
%GC content	44.41	44.41	44.42
%TE content	44.51 (68 Mbp)	45.16 (35 Mbp)	43.87 (33 Mbp)
# Gaps	5	4	1
%Complete BUSCOs	92.6	91.9	91.4
# Heterozygous SNPs/Mbp	6.8	6.9	6.7
# Homozygous SNPs/Mbp	1.1	1.3	0.8
LAI	18.3	27.2	24.8
CRAQ R-AQI	94.7	94.3	95.1
CRAQ S-AQI	97.4	97.4	97.4

866

867 **Table 2. Hi-C contact statistics confirming phasing correctness of the *Pst104E* genome**
868 **assembly.**

Types of Hi-C contacts (MAPQ ≥ 20)	Counts
Total Hi-C links	1,126,566
Within-haplotype links	1,117,107 (99.2%)
<i>cis</i> -chromosome	
Haplotype A	479,583 (42.6%)
Haplotype B	467,937 (41.5%)
<i>trans</i> -chromosome	
Haplotype A	85,667 (7.6%)
Haplotype B	83,920 (7.5%)
Cross-haplotype links	9,459 (0.8%)

869

Figure legends

Figure 1. A nuclear-phased, chromosome-scale genome assembly of the dikaryotic fungus *Puccinia striiformis* f. sp. *tritici* isolate *Pst104E*.

(A) Karyoplot of the 18 chromosomes of haplotype A, showing CpG methylation and transposable elements (TEs) density as peaks, and gene density as heatmaps within chromosome ideograms (10 kbp sliding windows). Locations of centromeres, telomeres, secretome genes, the ribosomal DNA array, and assembly gaps are annotated as per legend. **(B)** Hi-C contact heatmap of the full dikaryotic genome assembly consisting of 36 chromosomes. The two nuclear haplotypes A and B display a clear signal of spatial separation.

Figure 2. *Pst104E* centromeres are highly diverse with haplotype-specific sequences and are enriched in retrotransposons.

(A) Percentage of TE coverage in centromere and non-centromere regions of *Pst104E*. Each dot represents one of the 36 chromosomes. Student's *t*-test (***, $p < 0.001$). **(B)** Percentage of coverage of methylated CpG sites in centromere and non-centromere regions of *Pst104E*. Each dot represents one of the 36 *Pst104E* chromosomes. Student's *t*-test (***, $p < 0.001$). **(C)** Size ratio of haplotype A centromeres compared to haplotype B centromeres. Outlier ratios exceeding 1.5 times the interquartile range are highlighted in red. **(D)** Pairwise alignment dotplot between centromeres of haplotypes A and B. Only alignment blocks longer than 100 bp with minimum 90% sequence identity are shown. Each colour denotes an alignment type: blue, unique forward alignments; green, unique reverse alignments; orange, repetitive alignments. **(E)** Enrichment of TE superfamilies within centromeres compared with non-centromere regions. Statistical significance was assessed using permutation tests on each chromosome. Only abundant TE superfamilies with >1% total genome coverage are shown. The colour scalebar denotes the test statistic, defined as the observed difference in % TE coverage within and outside centromeres. P-values represent the proportion of permuted values equal to or more extreme than observed. FDR < 5% was applied to correct for

multiple testing (*, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$).

898

Figure 3. Analysis of centromere dip regions (CDRs) within *Pst104E* centromeres suggest a single putative kinetochore attachment site per chromosome.

(A) CpG methylation profiles (black histograms) and % AT content (red lines) across *Pst104E* centromeres *Cen1A* to *Cen9A* as examples. A consistent methylation depletion valley with a mean size of ~24.8 kbp was observed throughout all centromeres, indicating CDR signals (marked by grey dotted line). **(B)** Estimated sizes of the CDRs for each chromosome in haplotype A and B. Student's *t*-test (NS, $p > 0.05$). **(C)** Differences in the relative positions of CDRs between haplotypes for each homologous chromosome pair. Differences that exceed 0.2 are highlighted in red. **(D)** Sequence alignment dotplot between *Cen3A* and *Cen3B* and their respective CpG methylation profiles (black histograms). The dotted red box highlights sequence divergence between *Cen3A* and *Cen3B* corresponding to the *Cen3B* CDR. **(E)** Detailed synteny analysis of the region highlighted in (D). Long orange arrows indicate copies of hapB-B-G1437-Map9 belonging to the Ty3/Gypsy LTR retrotransposon superfamily. Grey shading indicates homologous sequences with % identity shown in boxes. Numbers above and below the orange arrows indicate the respective % methylated CpGs.

Figure 4. The two dikaryotic nuclear haplotypes of *Pst104E* harbour distinct rDNA subtypes.

(A) Schematic view of the rDNA tandem repeat array located on *Pst104E* Chr13. **(B)** Diagram of a single canonical rDNA unit of *Pst104E* containing the transcription start site in the 5' external transcribed spacer (5' ETS; not shown), the catalytic rRNA genes (18S, 5.8S, 25S and 5S), two internal transcribed spacers (ITS1 and ITS2) and two intergenic spacers (IGS1 and IGS2). The alignment shows two dominant rDNA subtypes (#1 and #2) which differ by SNPs and structural variations in both IGS. **(C)** Estimated copy number of rDNA subtypes #1 and #2 based on ONT duplex and Illumina read datasets. **(D)** Workflow of our rDNA Hi-C analysis designed to determine the physical location of the two dominant rDNA subtypes in each nuclear haplotype. **(E)** Mapping quality (MAPQ) distributions of Hi-C reads whose paired mates contained rDNA subtype-specific or

non-subtype-specific 31-mers, categorised by the nuclear haplotype they mapped against.

Figure 5. *Pst*104E's inter-haplotype variation is shaped by large-scale structural variation and transposable elements.

(A) Synteny and structural rearrangements between chromosome pairs of the two nuclear haplotypes, with A as reference and B as query. **(B)** Length distribution (left) of different structural variation (SV) types, including syntenic and unaligned regions. Counts of each SV type are shown. The bar chart (right) represents the total genome length covered by each sequence category. **(C)** Enrichment (red) and depletion (blue) of different genomic features (TEs and genes) within SVs including +/- 2 kbp flanking regions tested using permutations. The colour scale denotes the observed difference in % coverage of a genomic feature located within and outside of a given set of SVs. P-values indicate the proportion of permuted coverage values less than, equal to, or greater than the observed (*, $p < 0.05$; **, < 0.01 ; ***, < 0.001 ; FDR-corrected). **(D)** Flowchart summarising the classification of hemizygous, heterozygous and homozygous protein-coding genes, and their further categorisation. Hemizygous genes were intersected with secretome genes upregulated during infection timepoints (4, 6 or 8 dpi) to shortlist *Avr* candidates. Heterozygous biallelic genes were used for allele-specific expression analysis.

Figure 6. Allele-specific expression (ASE) is prevalent among secretome genes and correlates with gene methylation patterns.

(A) The five ASE categories detected across six different transcript sampling conditions. Allele pairs displaying absolute $\log_2$ fold changes greater than two (Diff2 and Diff4) in at least one condition were defined as ASE in subsequent analysis. **(B)** Expression levels of alleles belonging to the two nuclear haplotypes A and B. Mann-Whitney *U* test (***, $p < 0.001$). **(C)** Odds ratios and $\log_{10}$ p-values from Fisher's exact tests comparing the proportion of ASE genes among secretome genes and evolutionarily conserved BUSCOs. Each red dot represents a transcript sampling conditions as labelled. Dotted red and grey lines highlight cut-offs matching the null hypothesis of no difference

951 between the two gene groups. **(D)** Distribution of % CpG methylation (sampled at UG) across ASE
952 or non-ASE secretome gene bodies (here defined as start to stop codon) including +/- 2 kbp
953 flanking regions. Solid lines represent mean % CpG methylation; shaded areas represent 95%
954 bootstrapping confidence intervals. Yellow inset shows the number of allele pairs included.

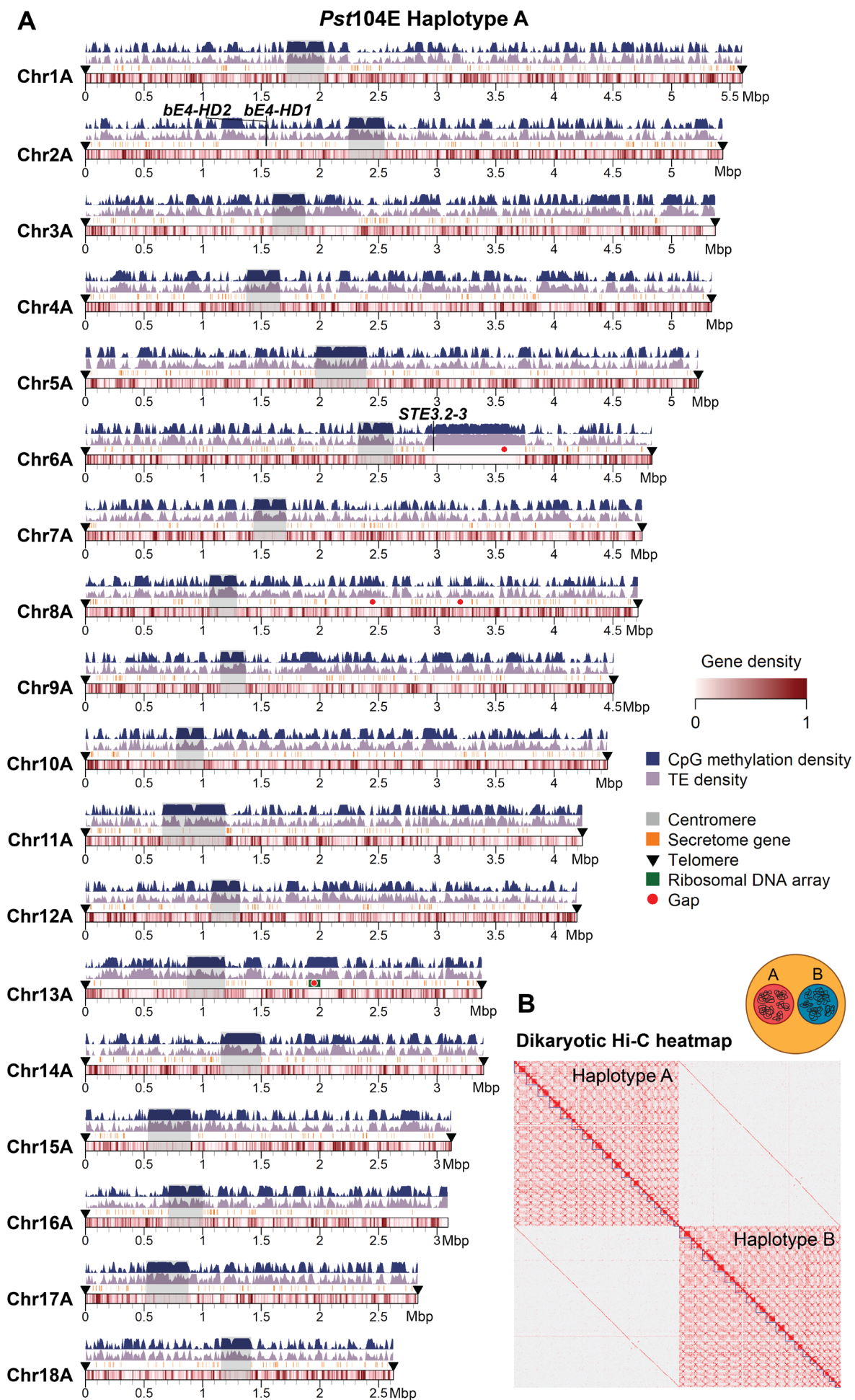
Figure 1

Figure 2

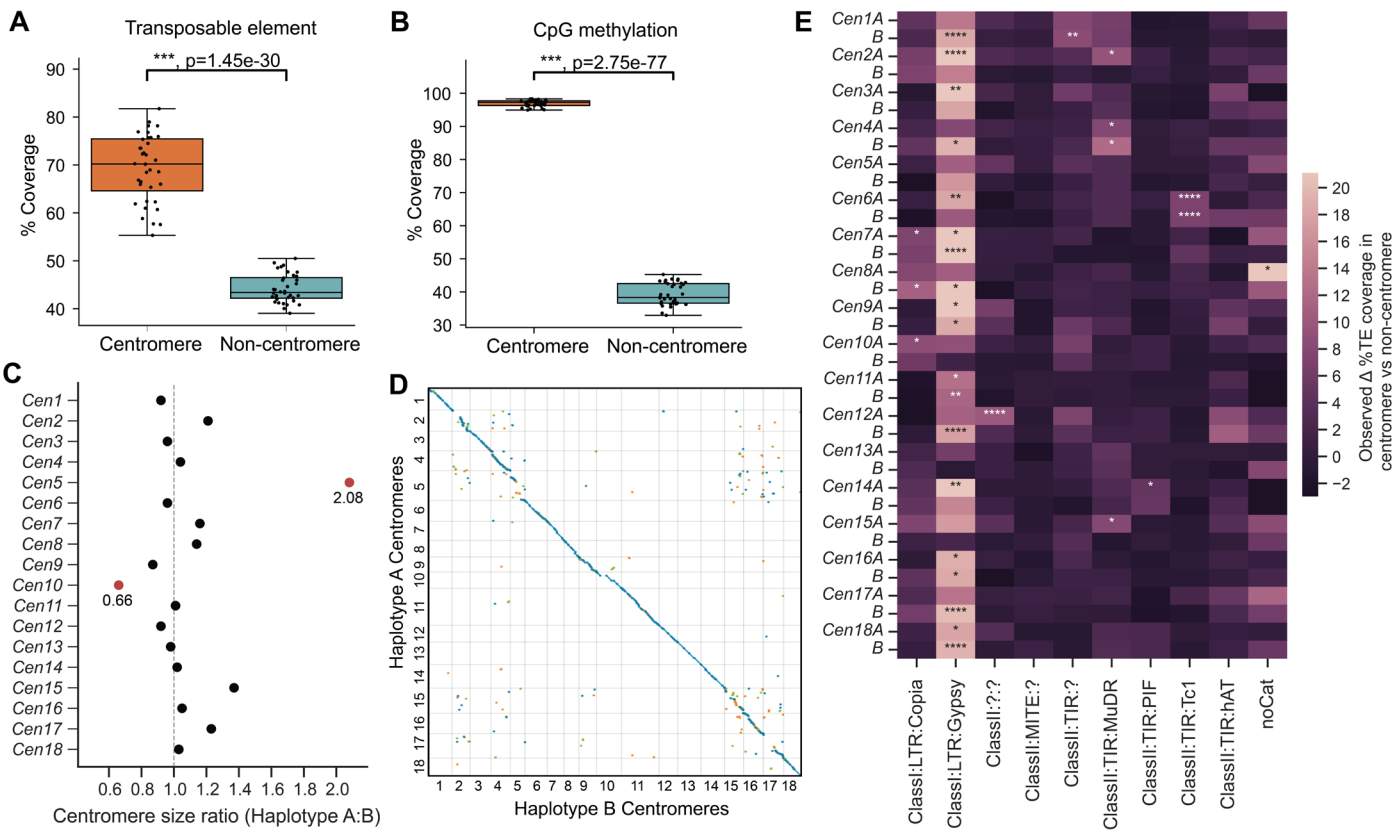
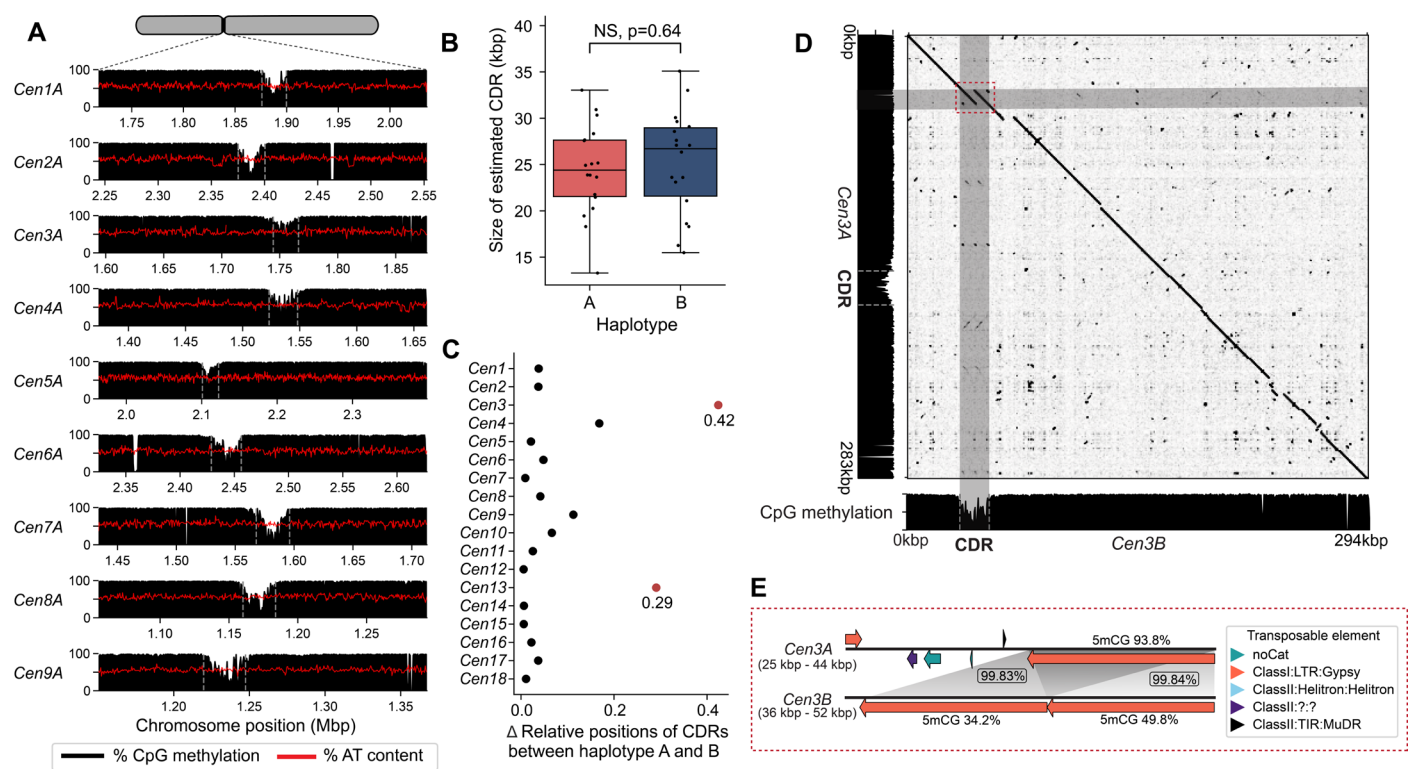


Figure 3



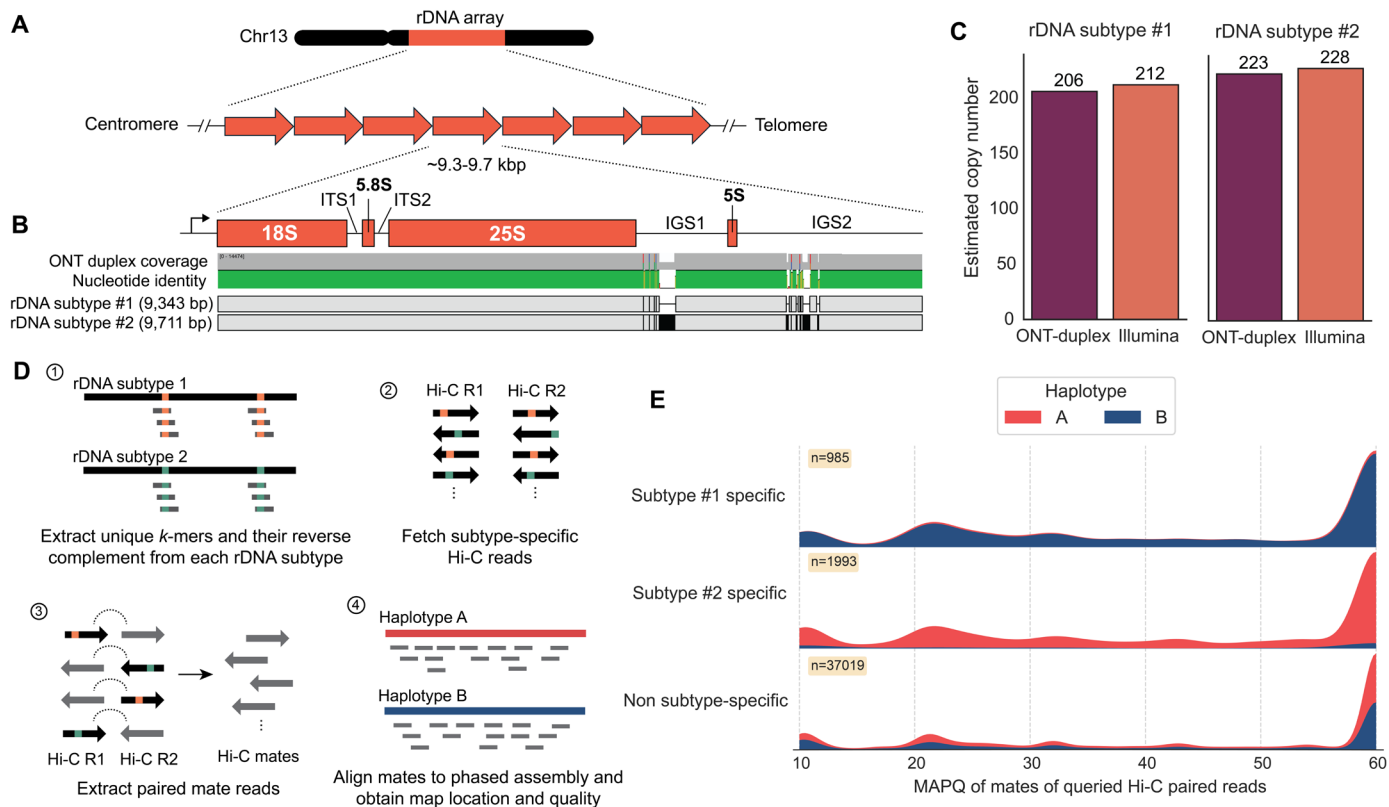


Figure 5

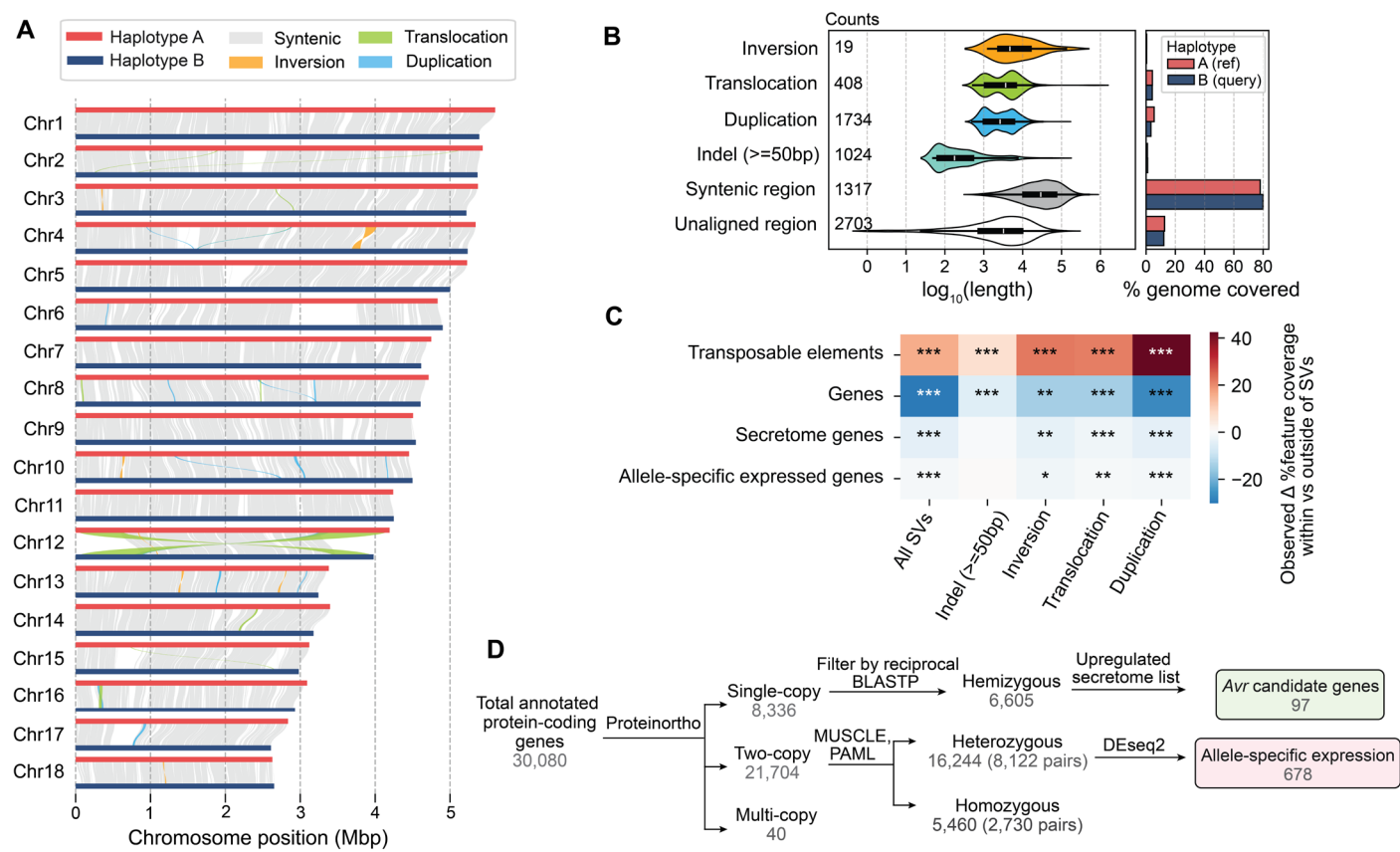


Figure 6

