Title

Genome biology and evolution of mating type loci in four cereal rust fungi

Author

Zhenyan Luo[1], Alistair McTaggart[2], and Benjamin Schwessinger[1,#]

Other information

[1] Research Biology School, Australian National University, Canberra, ACT, Australia

[2] Centre for Horticultural Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Ecosciences Precinct, Dutton Park, Queensland, Australia

[#]Corresponding author; benjamin.schwessinger@anu.edu.au

Abstract

Sex defining loci or chromosomes often contain large non-recombining regions. In Basidiomycota, two distinct gene loci confer mate compatibility. These loci encode for homeodomain (HD) transcription factors and pheromone receptor (PR)-ligand pairs. In some fungi these two loci are physically linked and give rise to large evolutionary strata with extensive recombination suppression. To date genome level mating type (*MAT*) loci analysis is lacking for obligate biotrophic basidiomycetes that depend on their host plants for survival, this includes the order *Pucciniales*. These economically important plant pathogens have long-term dikaryotic life-stages in which matched haploid nuclei must carry compatible *MAT* loci.

Here we focus on four *Puccinia* cereal rust species, including *Puccinia graminis* f.sp *tritici*, *P. striiformis* f.sp *tritici*, *P. triticina* and *P. coronata* f.sp *avenae,* which infect major cereals including wheat and oat. We identified *MAT* locations on two separate chromosomes that supports previous hypotheses of a tetrapolar mating system in the *Pucciniales*. *HD* loci displayed largely conserved macrosynteny among species and were not suppressed for recombination. Consistently, *HD* transcription factor genes showed no trans-specific polymorphisms between species. Regions, surrounding the *PR* loci however, displayed extensive genomic degeneration, signs of recombination suppression and were found to be linked to centromeres in some species. Expression analysis suggests that both *MAT* loci are involved in nuclear pairing during spore formation in the asexual cycle. Together, our study provides insights into the evolution of *MAT* loci since speciation of key pathogenic *Puccinia* species. This detailed understanding is important to predict the possible combinations of nuclear pairs that can arise via sexual reproduction, somatic recombination, or di-om mating to enable the evolution of newly virulent isolates of these agriculturally important plant pathogens.

Introduction

The evolution of ‘sex’ determining loci and chromosomes is a fundamental question in biology. It is widely accepted that X and Y chromosomes were originally homologous but recombination suppression gradually resulted in genetic degeneration of the Y chromosome (Bachtrog 2006). Fungi are excellent models for studying evolution of sex because they carry mating type (*MAT*)

chromosomes or *MAT* loci that share some similar characteristics with sex chromosomes. This includes extended regions of recombination suppression around *MAT* determined genomic regions. In some model fungal species, regions of recombination suppression range from 100s of kbp (kilo base pairs) to several mbp (mega base pairs) around *MAT* loci (Idnurm, et al. 2015; Coelho, et al. 2017), such as in *Neurospora tetrasperma* (Menkis, et al. 2008), *Podospora anserina* (Hartmann, Ament-Velásquez, et al. 2021) and *Microbotryum lagerheimii* (Branco, et al. 2017).

Basidiomycota have evolved unique mating systems to govern nuclear compatibility of haplotypes and inbreeding within populations. Two *MAT* loci occur in each haplotype, the pheromone-receptor (*PR*) locus and the homeodomain (*HD*) locus (Fraser, et al. 2004). The *PR* locus contains a pheromone receptor gene and at least one mating pheromone gene involved in determining pre-mating compatibility. The *HD* locus contains two tightly linked homeodomain transcription factors (*bW-HD* and *bE-HD*) which are transcribed in opposite directions to control post-mating compatibility (Coelho, et al. 2017). Mating is either bipolar, in which the *HD* and *PR* genes are linked at one locus, or only one *MAT* locus controls mating compatibility, or tetrapolar, in which the *HD* and *PR* loci occur in different parts of the genome and both control mate compatibility. Multiallelic *MAT* loci provide more outcrossing possibility due to molecular nonself-recognition (Hartmann, Duhamel, et al. 2021).

Tetrapolar mating is plesiomorphic in Basidiomycota and decreases inbreeding among siblings compared to bipolar mating, as haploids must differ at two unlinked loci rather than at one locus (James 2007). For example, species of *Microbotryum* are highly selfing, and multiple rearrangements have linked the *HD* locus and the *PR* locus onto the same chromosome (Branco, et al. 2017). Species of *Microbotryum* that evolved bipolar mating systems have increased their opportunities for sexual reproduction in closely related populations (Branco, et al. 2017), and are one example of many Basidiomycota that have changed from a tetrapolar ancestor (James, et al. 2006; Coelho, et al. 2017).

Culturable fungi are commonly used to study mating as compatibility can be tested *in vitro*. We know little about how mating is governed in obligate fungi, including rust fungi (*Pucciniales*) (Cuomo, et al. 2017; Li, et al. 2019; McTaggart, et al. 2022) that cause rust diseases with economic impact for various crops, such as wheat, soybean, and coffee (Aime and McTaggart 2021). Rust fungi have multiple spore stages to complete their asexual and sexual life cycles. Many species of rust fungi require alternate hosts to complete their life cycles (heteroecious), such as species of *Berberis* and *Triticum* for *Puccinia striiformis* f.sp. *tritici* (Wang and Chen 2013). Asexual spore stages dominate life cycles of heteroecious species of *Puccinia* in agricultural systems, as sexual reproduction is prevented by management strategies that remove alternate hosts. Somatic hybridisation, in which nuclei are exchanged between asexual stages of rust fungi, can increase genetic/genotypic diversity when life cycles are interrupted (Pontecorvo 1956; Li, et al. 2019; Sperschneider, et al. 2022). Nuclear compatibility in somatic hybridisation is hypothesised to be governed by *MAT* loci (Wu, et al. 2021), however this is untested *in planta*, both nuclear fidelity in somatic hybridization and in mitosis is unexplored in rust fungi, which lack clamp connections commonly found in other basidomycetes. Sexual recombination and somatic hybridisation may increase fitness of new isolates via pairing of novel haplotype combinations which can also lead to the avoidance of triggering resistance response

of host plants. Understanding how rust fungi change is fundamental to disease management and to predict emerging genotypes (McTaggart, et al. 2022).

Rust fungi likely have tetrapolar mating systems, as hypothesised for *Austropuccinia psidii* (Ferrarezi, et al. 2022)*, Puccinia graminis* f.sp. *tritici* (Li, et al. 2019) and *Puccinia tritici* (Cuomo, et al. 2017; Wu, et al. 2021)*.* However, genome assemblies of rust fungi have long been fragmented and the structure of the *MAT* loci and function of *PR* homologs is not yet understood. Clustering of repetitive elements around *MAT* alleles (Fraser, et al. 2004; Hood, et al. 2004) is a challenge to studying *MAT* loci in species with repeat-rich genomes, including rust fungi, as repetitive sequences cause gaps, fragmentation, and phase-error in the assembled genome.

Here we overcome this challenge and make use of several recently published nuclear assigned and chromosome scale assemblies available for cereal rust fungi. We analysed *MAT* loci in four related species of *Puccinia*, *Puccinia graminis* f.sp *tritici* (*Pgt*) (Li, et al. 2019), *Puccinia striiformis* f.sp *tritici* (*Pst*) (Schwessinger, et al. 2018; Schwessinger, et al. 2020; Schwessinger, et al. 2022), *Puccinia triticina* (*Pt*) (Duan, et al. 2021) and *Puccinia coronata* f.sp. *avenae* (*Pca*) (Miller, et al. 2018; Henningsen, et al. 2022). Based on studies of *Microbotryum* spp. and other rust fungi, we hypothesized that: 1) Species of *Puccinia* have tetrapolar mating systems and should have syntenic *MAT* loci in related species. 2) Recombination suppression causes increased heterozygosity in *MAT* alleles and surrounding genes. 3) Recombination suppression causes genomic degeneration in *MAT* loci. 4) Species of *Puccinia* have multiallelic copies of *HD* and *PR* loci. 5) *MAT* genes maintain nuclear fidelity in both sexual and asexual life cycles.

Results

Genomic organisation and inheritance of mating type genes suggest cereal rust fungi have tetrapolar mating systems

We set out to confirm previous preliminary observations that suggested rust fungi have a tetrapolar mating system (Cuomo, et al. 2017; Li, et al. 2019; Wu, et al. 2021). We determined whether *MAT* loci were linked or not by their location in chromosome-level assemblies of cereal rust fungi including *Pca 203, Pst 134E*, *Pt 76* and *Pgt 210* (Table 1). We included the poplar rust fungus *Melampsora larici-populina* as an outgroup for our study.

We used previously characterised *MAT* genes from the *Pt* isolate *BBBD* (Cuomo, et al. 2017) as queries to identify orthologs in examined genomes and proteomes (Table 1 & Table 2). We identified *HD* (*bW-HD* and *bE-HD*) and *PR* (*STE3.2* and *STE3.3*) loci on chromosome 4 and chromosome 9 respectively, whereas *STE3.1* genes are located on chromosome 1 (Supplementary Fig. S1). This is consistent with previous preliminary analysis for *Pgt 210,* which suggested *HD* and *PR* loci are located on chromosome 4 and 9 respectively for *Pgt* (Li, et al. 2019). We also made use of additional genome resources for cereal rust fungi that are only partially phased and non-chromosome scale. This includes assemblies for three out of four studied cereal rust fungi such as *Pca 12NC29*, *Pca 12SD80*, *Pgt Ug99*, *Pst DK0911* and *Pst 104E* (Miller, et al. 2018; Schwessinger, et al. 2018; Li, et al. 2019). We identified two haplotypes of *bW-HD* and *bE-HD* and one haplotype each of *STE3.2* and

STE3.3 in all genome assemblies analysed with the exception for *STE3.2* in *Pst 104E* and *Pca 12SD80* (Miller, et al. 2018; Schwessinger, et al. 2018). In the case of *Pst 104E, STE3.2* was absent from the genome assembly but we recovered its haplotype by mapping raw sequencing reads of *Pst 104E* to a new phased chromosome scale genome assembly of *Pst 134E* (Schwessinger, et al. 2022). In the case of *Pca 12SD80,* we identified two copies of *STE3.2* on contig 000183F and 000183F_004, respectively. This miss-characterisation of *STE3.2* in *Pst 104E* and *Pca 12SD80* is likely caused by errors in the genome assembly process of early long-read genomes using low accuracy PacBio long-reads and suboptimal genome assembly algorithms that struggle with highly repetitive regions of a genome.

We used a phylogenetic hypothesis to illustrate the inheritance of *STE3* and *HD* genes in each haplotype (Figure 1 and 2). Compatible haplotypes in a dikaryotic genotype have either a copy of *STE3.2* or *STE3.3* and always have different alleles at the *HD* genes, as expected in tetrapolar mating.

We also identified an additional *PR*-like gene with *STE3.1* on chromosome 1 in all analysed species, however copies of this gene are invariant between the two nuclear genomes and not associated with any pheromone peptide precursor genes (see below). Hence *STE3.1* is likely not to be involved mating compatibility in rust fungi.

In summary, *HD* and *PR* loci are unlinked, located on two distinct chromosomes, and heterozygous for all analysed species and isolates.

MFA pheromone peptide precursor genes are closely linked to *STE3.2* but not *STE3.3*

After identifying *STE3.2* and *STE3.3*, we searched the examined genomes for putative pheromone peptide precursor genes, which encode mating factor a (MFA) peptides that bind to the compatible opposing pheromone receptors. We used the previously identified *mfa*1, *mfa2,* and *mfa3* genes of *Pst*, *Pgt* and *Pt* as queries (Cuomo, et al. 2017). We identified a single *mfa2* pheromone peptide precursor gene linked to *STE3.2* in all species. The derived peptides are all 34 amino acids long with a characteristic CAAX motif at the C-terminus, where C is cysteine, A is an aliphatic amino acid, and X is any amino acid (Supplementary Fig. S2) (Raudaskoski and Kothe 2010). *Mfa2* is fully conserved within each species when analysing multiple isolates per species (Supplementary Fig. S2). *STE3.2* and *mfa2* are closely linked, located within 500–1100 bp from each other, and encoded on the same DNA strand in all cereal rust fungal species (Supplementary Fig. S3).

In contrast, the composition, length, number, and location of *mfa* genes associated with *STE3.3* was variable between species. *Mfa1* of *Pca 203* encodes a 66 amino acid peptide and is located 0.54Mb upstream of *STE3.3*. In *Pst*, *mfa1* encodes a 64 amino acid peptide and is located 10kb and 13kb away from *STE3.3* in *Pst 134E* and *Pst DK0911*, respectively. In contrast, *Pt* appears to have two identical copies of *mfa1* associated with *STE3.3* at a distance of 0.24Mbp and 0.27Mbp (Supplementary Fig. S3). These *mfa1* genes encode for 61 amino acid peptides. In the case of *Pgt*, we were able to identify the two distinct pheromone peptide precursor genes *mfa1* and *mfa3* in association with *STE3.3. Mfa1* and *mfa3* encode for 59 and 57 amino acid peptides, respectively.

These two genes are located upstream and downstream of *STE3.3* at a distance 30kb and 96kb, respectively (Supplementary Fig. S3). Except for *Pst*, multiple genes were located between *STE3.3* and *mfa* with relative low gene density in *Pgt*, *Pt* and *Pca*. We could not identify any *mfa* genes associated with *STE3.1*.

Genealogies of *MAT* haplotypes elucidate distinct evolutionary histories for *HD* genes versus *PR* and *mfa* genes

We investigated genealogies of the different *MAT* genes to describe evolutionary relationships between haplotypes, compared their relationship to the species trees to test for phylogenetic concordance, and to compare our observations between *HD* and *PR/mfa* genes. *HD* genes *bE-HD* and *bW-HD* were aligned separately to construct independent gene trees. We used one haplotype of *bE-HD* and *bW-HD* from *Melampsora larici-populina 98AG31* (XP_007413298 and XP_007413233) (Duplessis, et al. 2011) as outgroup in the respective gene trees (Figure 1). We also included *bE-HD* and *bW-HD* from *Pt BBBD* as references (Cuomo, et al. 2017). The gene trees of *bE-HD* and *bW-HD* grouped haplotypes found in the same species closely together (Figure 1). We did not find any evidence for trans-specific polymorphisms for either gene between species and identified at least one shared haplotype between isolates in most species. The tree topology of *bE-HD* and *bW-HD* haplotypes within each species was mostly but not always consistent with our initial assumptions that physically linked *bE-HD/bW-HD* are grouped consistently when comparing the different gene trees. For example, in *Pst* not all *bE-HD1* haplotypes grouped together while all *bW-HD1* loci did (Figure 1). Also *bE-HD* haplotypes within one *Pst* (*Pst104E*) isolate were more similar to each other than between isolates. Similar complex evolutionary histories of *bE-HD* and *bW-HD* haplotypes within each species could be identified for other cereal rust fungi. Furthermore, both gene trees for *bE-HD* and *bW-HD* were incongruent with the widely accepted species tree (Aime, et al. 2018) similar to the gene tree of *STE3.3*.

Next, we explored the evolutionary relationship of the *STE3* pheromone receptor genes. In theory *STE3.2* and *STE3.3* are two haplotypes of the same gene (Coelho, et al. 2017). We used previously described annotations of *STE3.2* and *STE3.4* from *Melampsora larici-populina 98AG31* (XP_007418099 and XP_007419438) as outgroups for generating rooted gene trees for *STE3.2* and *STE3.3* separately, because they are highly diverged within each isolate/species and do not allow for the generation of one combined gene tree of high quality (data not shown). This indicates strong trans-polymorphisms and a long-term suppression of recombination between these haplotypes that predates ancestral species split. The gene tree for *STE3.*2 followed the species tree (Aime, et al. 2018) with *Pca STE3.2* being the outgroup to the wheat rust fungal clade of *STE3.2* containing *Pst*, and *Pt-Pgt* (Figure 2). In contrast, the *STE3.3* gene tree was incongruent with the species tree (Aime, et al. 2018) and the gene tree of *STE3.2*. For *STE3.3 Pst* was the outgroup to the clade that contained *STE3.3* from the three other cereal rust fungi *Pca* and *Pt-Pgt* (Figure 2). These incongruent topologies support long-term recombination suppression around the *PR* loci.

We next investigated if the putative pheromone peptide precursor *mfa* genes followed a similar evolutionary pattern to their associated *STE3* gene. We used all previously identified and above mentioned *mfa* genes (Cuomo, et al. 2017) and haplotypes in addition to two putative pheromone precursors from *Melampsora larici-populina 98AG31* (XP_007419440 and XP_007418113) as outgroups. The gene tree of *mfa* haplotypes also revealed strong trans-specific polymorphisms and followed a similar topology to the gene trees of *STE3.*2 and *STE3.3*. *Mfa* haplotypes clustered according to their physical association with a specific *STE3* haplotype and not by species identity, for example *mfa1* haplotypes from all species formed one specific clade defined by gene identity and not by species of origin (Supplementary Fig. S4). The species specific *mfa3* in *Pgt*, which is physically associated with *STE3.3*, is more closely related to *STE3.2* linked *mfa2* than *mfa1*.

We also investigated the evolutionary relationships of the *PR*-like *STE3.1* haplotypes. One haplotype of *STE3.1* in *M. larici-populina 98AG31* (XP_007405301) was used as outgroup for rooting the gene tree. The two *STE3.1* haplotypes identified in each isolate were highly conserved (almost completely identical) and *STE3.1* showed very little variation within each species (Supplementary Fig. S5). Overall, this suggests that *STE3.1* is highly conserved in each nuclear genome and further supports our hypothesis that *STE3.1* does not function in defining mating compatibility.

Genomic regions containing *HD* transcription factor genes display limited signs of genomic deterioration and their synteny is conserved within each species

Recombination suppression has been observed around *MAT* loci in multiple basidiomycetes (Branco, et al. 2017; Sun, et al. 2019), hence we investigated if we can identify signals for recombination suppression, breaks in synteny, and general genomic deterioration around the *MAT* loci in rust fungi. First, we analysed whether there was a reduction in synteny on chromosome 4 surrounding the *HD* locus as a signature for long-term recombination suppression. We used mummer (Marçais, et al. 2018) to align the haplotypes of chromosome 4 from *Pca, Pst, Pgt* and *Pt* against each other (Supplementary Fig. S6). Overall macro-synteny of chromosome 4 haplotypes was conserved in all cereal rust fungi with *Pt 76* being the most syntenic (Supplementary Fig. S6). Micro-synteny in a 40kb window around the *HD* loci was mostly conserved in *Pt 76*, *Pca 203*, and *Pst 134E* while this initial analysis suggested that this locus was less syntenic in *Pgt 210* (Supplementary Fig. S6).

We investigated the individual gene haplotype divergence along chromosome 4 using synonymous divergence as an additional measure of synteny (Branco, et al. 2017). This analysis tested if the *HD* and associated proximal genes show an increase in synonymous divergence rate, which could be caused by long-term suppression of recombination and accumulation of independent mutations seen in other Basidiomycota (Carpentier, et al. 2019). We calculated the pairwise synonymous divergence rate for all genes across all sister chromosomes in all four species of *Puccinia*. This whole genome analysis of all nuclear gene pairs on sister chromosomes suggested that chromosome 4 does not have an increased global synonymous divergence rate when compared to all other chromosomes (Supplementary Fig. S7). Synonymous divergence rate (dS) of haplotypes for both *HD* genes fell within the upper 95% quantile of dS values of all genes on chromosome 4. Yet dS values of directly adjacent genes are not elevated beyond the background level visible when we plotted dS values for

all gene pairs along chromosome 4 (Figure 3). Consistently, regions around *HD* loci did not show obviously aberrant patterns of gene or transposon density (Figure 3). We tested whether *HD* loci are potentially linked to centromeres as reported in other basidiomycete species (Hood, et al. 2015; Carpentier, et al. 2019). Only *Pst 134E* (Schwessinger, et al. 2022) and *Pgt 210* (Li, et al. 2019; Sperschneider, et al. 2021) have information of centromere locations and their location suggests that the *HD* loci are not directly linked to centromeres in either species.

Lastly, we combined fine scale nucleotide synteny analysis with gene conservation analysis at the protein level, including 40 genes either side of the *HD* locus, to investigate nucleotide vs encoding protein conservation. We specifically tested if cereal rust fungi contain syntenic blocks with conserved gene structures surrounding *MAT* loci as was found for species of *Trichosporonales* (Sun, et al. 2019). We used blastn to identify conserved nucleotide sequences within and between species. Protein sequences were clustered using OrthoFinder (Emms and Kelly 2019). We plotted genes, transposons, nucleotide and protein conservation for the surrounding regions of the *HD* loci within and between cereal rust fungal species (Figure 4A). Protein coding genes and their order are mostly conserved between nuclear genomes within the same species with 58/80 genes in *Pca203*, 62/80 in *Pst134E*, 40/80 in *Pgt 210* and 57/80 in *Pt76* being conserved between the two haplotypes within each species. Similarly, we observed considerable nucleotide conservation and synteny within each species except for *Pgt 210*, which is consistent with our initial mummer based analysis (Supplementary Fig. S6). Overall, synteny and protein gene conservation was very limited between species for *HD* loci surrounding regions. We could identify three conserved genes across all four cereal rust fungi. These three genes code for an integral membrane protein (PTHR12459), D-2-hydroxyglutarate dehydrogenase (PTHR43716) and Glucose-6-phosphate 1-sepimerase (IPR025532) with all three clustering downstream of the *HD* locus (Figure 4A). Composition of transposons vary across species, DNA transposons dominated the *HD* locus of *Pst134E* but *HD* loci of other rust fungi had higher coverage of RNA transposons (Supplementary Fig. S8).

Genomic regions containing *PR* genes display strong signs of loss of synteny, genomic deterioration, recombination suppression, and accumulation of transposable elements

To investigate structure and synteny of *PR* loci we used identical approaches as applied to the *HD* loci and their surrounding regions. Initial mummer based synteny analysis revealed clear macro- and micro-synteny breaks around the *PR* loci in all cereal rust fungi (Supplementary Fig. S9). It also revealed accumulation of highly repetitive regions closely associated with *PR* loci, especially in *Pca 203* and *Pst 134E*. This increase in TEs goes hand in glove with a reduction of gene density in the *PR* loci surrounding regions (Figure 4B and 5). The loss in synteny was specific to *PR* loci neighbouring regions as the overall macro-synteny of chromosome 9 was conserved in all species (Supplementary Fig. S9).

Consistently, the dS values for the *STE3.2- STE3.3* gene pair belongs to the 99.9% quantile of dS values of all gene pairs found on chromosome 9 in all species (Figure 5 and Supplementary Fig. S7). Gene density dropped significantly around *PR* loci with *STE3.2*/*STE3.3* located within or adjacent to 'gene deserts' that extended over 1 mbps in the case of *Pst 134E*. The slight exception here was *Pgt*

210, which is likely caused by assembly errors introduced by the extensive manual curation and assembly processes of this genome that relied on gene synteny between contigs and haplotypes (Li, et al. 2019).

Concomitant with gene depletion, *PR* loci adjacent regions are enriched in transposable elements. Consistently, the *PR* loci of *Pst 134E* was found to be ~164kbp from the centromere of chromosome 9A and ~473kbp from that of chromosome 9B, likely physically-linked to their centromeres. We do not have the information of centromeric regions of *Pca 203* and *Pt 76*, but based on the TE density pattern, it is likely that *PR* loci of *Pca 203* and *Pt 76* are also closely associated to centromeres (Figure 5). Detailed nucleotide synteny and protein coding gene conservation analysis confirmed signals of extended recombination suppression, loss of synteny and accumulation of transposable elements in *PR* loci adjacent regions (Figure 4B). There is very little conservation in these regions between haplotypes of each species and between species. We could not identify a conserved protein coding gene in this region between species. *Pst 210* and *Pt 76* showed strong signals of accumulation of independent transposable element families in each nuclear haplotype, which indicates invasion by TEs since cereal rusts shared a most recent common ancestor. In general, distinct transposable element families contributed to the extensions of the *PR* loci adjacent regions in the different species.

Genomic regions surrounding *STE3.1* are highly conserved within and among cereal rust species consistent with *STE3.1* not being involved in mating compatibility

We analysed *STE3.1* as above for the *MAT* loci. The three independent analyses revealed that *STE3.1* is highly conserved within each species and imbedded in a region of high synteny (Supplementary Fig. S10-12). In addition, genes in *STE3.1* adjacent regions are mostly conserved between species. This clearly sets apart the *STE3.1* locus from the other two *MAT* loci which further supports our hypothesis that this locus is not involved in mating compatibility in rust fungi.

***HD* loci are multiallelic in cereal rust fungi while *PR* loci show little variation at the species level**

We aimed to investigate the species level variation of *MAT* loci encoding genes using publicly available genomic data for *Pca*, *Pst*, *Pgt* and *Pt*. We identified 14 representative isolates per species that were assigned to as many distinct genomic lineages as possible based on previous phylogenetic studies (Table S2). Intraspecific diversity was considered as cereal rust fungi are clonal in most regions of the world and we aimed to avoid sampling from the same clonal lineage. We quality controlled whole genome short read datasets and mapped them against the respective reference genomes to estimate the genetic variation. We identified high levels of polymorphisms at the *HD* loci for all analysed cereal rust fungi. We attempted multiple approaches to phase the entire *HD* loci based on publicly available short read data but failed in all cases (data not shown). We identified multiple SNPs and variations in *bE-HD* across different *Pgt* isolates and this includes a six-nucleotide length deletion at the position between 1083 in *Pgt 126-6711*, *TTTSK*, *UVPgt55* and *UVPgt60*. *Pca* showed the highest level of variation of *HD* loci at the species level (Supplementary Fig. S13). This is likely caused by the fact that isolates selected were mostly linked to sexual populations of *Pca* found

in the USA (Miller, et al. 2020). Hence we also used the most closely related available reference of each sample for short read mapping purposes including references from *Pca 203* (Henningsen, et al. 2022), *Pca 12NC29* and *Pca 12SD80* (Miller, et al. 2020). This showed that the *HD* loci in *Pca* display high levels of polymorphisms including heterozygous SNPs, multiple unaligned regions and gaps, which suggests the *HD* loci present in our selected *Pca* isolates are not well reflected in the available references (data not shown).

Compared to the *HD* locus, the *PR* locus is reported to be far less polymorphic in many basidiomycetes (Urban, et al. 1996; Coelho, et al. 2010; Coelho, et al. 2017). Hence, we next focused our analysis on the *PR* loci using the same whole genome short read mapping datasets as mentioned above (Table S2, Supplementary Fig. S14). Overall, the *PR* loci were far less polymorphic within each species when compared with *HD* loci. *STE3.2* and *STE3.3* of all *Pst* isolates had identical coding regions, suggesting that the *PR* locus is biallelic with one haplotype each in the currently available samples of *Pst* globally. In contrast to the highly polymorphic *HD* loci, we were able to only identify two SNPs in the coding regions of *STE3.2* and *STE3.3* in *Pca* and with a single SNP being non-synonymous. We identified slightly more SNPs in *STE3.2* and *STE3.3* of *Pt* which we estimate to give rise to a maximum of four distinct haplotypes. Yet each predicted haplotype has only a very limited number of variable sites at the protein level. *Pgt* was the most polymorphic at the *STE3.2* and *STE3.3* loci. We identified three haplotypes of *STE3.2* with several amino acid changes. *STE3.3* was the most polymorphic in *Pgt* including two isolates, *TTTSK* and *UVPgt60,* which contain potential nonsense mutations leading to pre-mature stop codons.

Lastly, we also investigated the status of all *mfa* genes at the species level. *Mfa1* and *Mfa2* haplotypes were fully conserved in *Pca* and *Pst* in all isolates tested, while *Pt* had only a single non-synonymous change in *mfa1* (data not shown). In contrast, *mfa1* and *mfa3* in *Pgt* had several non-synonymous variations (Supplemental Figure S2, data not shown). These species level variation in *Pgt* at the *mfa* loci is consistent with significant variation observed in *STE3.2* and *ST3.3*.

MAT genes are expressed late in the asexual infection cycle during urediniospore production

Functions of *MAT* loci in sexual reproduction have been well established for *Pt* (Cuomo, et al. 2017). In addition, they are known to confer vegetative incompatibility during asexual growth in Ascomycota (Glass and Kuldau 1992). Hence, we investigated the expression patterns of *MAT* genes in the four cereal rust fungi. There is a single RNAseq dataset for the sexual infection process of cereal rust fungi in the public domain, which is a *Pst* infection of barberry at two time points (Zhao, et al. 2021) for which opposing monokaryotic single mating type infection structures were not separated. We analyzed the expression of *HD* and *PR* genes in this dataset. We found that both *HD* and *PR* genes are expressed in this dataset at the given timepoints (Figure 6A). We did not pursue more in-depth analysis of this dataset due to its intrinsic limitations.

We performed more in-depth expression analysis from extensive public RNAseq datasets that describe the infection of the cereal hosts during the asexual reproduction cycle (Dobon, et al. 2016; Yadav, et al. 2016; Chen, et al. 2017; Miller, et al. 2018; Schwessinger, et al. 2018; Zhao, et al. 2021).

STE3.2 and *STE3.3* were strongly upregulated during the asexual infection process and always displayed highest expression at the latest time point available, which coincides with production of urediniospores in all cases (Figure 6C). Similarly, *HD* genes are upregulated during asexual infection of the cereal host (Figure 6B).

Lastly, we were also interested to see if we could detect RNA of *MAT* genes in cereal rust spores before infection and in specialized infection structures called haustoria. For this we compared the expression of *MAT* genes in ungerminated spores, germinated spores, 5 days post infection, 9 days post infection and haustoria enriched samples (Figure 6D) (Schwessinger, et al. 2018). *HD* genes displayed some expression in spores albeit lower than observed in later stages of infection during spore maturation in the asexual cycle. In contrast, we could not detect expression of *PR* genes in ungerminated and germinated spores. Yet *PR* genes were upregulated at later stages of the asexual infection (Figure 6D) consistent with other publicly available datasets (Figure 6B and C).

Discussion

The genes behind sexual reproduction and nuclear exchange are poorly understood in rust fungi. We confirmed that cereal rusts use tetrapolar mating and determined the structure of the two loci that control mate compatibility. The *HD* locus had little evidence of recombination suppression and was syntenic within and among cereal rust fungi. Within each species *HD* loci are highly multiallelic with all cereal rust fungi carrying many clearly distinguishable haplotypes. The *PR* locus had strong evidence of recombination suppression, with little synteny within and among species, and is flanked by large stretches of transposable elements. At the species level *PR loci* are likely biallelic for *Pca, Pst*, and *Pt* while *Pgt* might be multiallelic at this second locus too. Both *HD* and *PR* genes are expressed in asexual stages of the life cycle, which may support a hypothesis that *MAT* genes maintain fidelity of nuclei in dikaryotic rust fungi during hyphal development, somatic hybridisation, and potential parasexual reproduction.

Basidiomycota have evolved bipolar and tetrapolar mating systems to control mate compatibility. Tetrapolar mating is thought to be ancestral in Basidiomycota, and we hypothesized that most rust fungi use tetrapolar mating. We identified several lines of evidence for tetrapolar mating in cereal rust fungi by analysing complete chromosome-scale genome assemblies of four species. *HD* and *PR* loci are located on two separate chromosomes in all examined species of cereal rust fungi, and haplotypes segregated with different haplotypes of *HD*, *PR*, and *mfa* genes. We failed to identify evolutionary strata in *Pca*, *Pst*, *Pgt* and *Pt* that link these genes together into a super locus like in other tetrapolar or bipolar species, such as *Trichosporonales* (Sun, et al. 2019), and *Microbotryum* (Branco, et al. 2017). In addition, the *PR* homolog *STE3.1* haplotypes were highly conserved in all cereal rust fungi and we did not find *mfa* genes closely associated with *STE3.1*. Interestingly, in *Pca*, *Pst* and *Pgt*, *STE3.1* is flanked by anaphase promoting complex *APC10/Doc1* which regulates important events in mitosis (Au, et al. 2002) and Elongation factor *EF1B* (Andersen, et al. 2003). Together, these findings suggest that *STE3.1* might play an essential function, possibly related to mitosis activity. At the same time *STE3.1* appears not to regulate mating compatibility and the four cereal rust fungi are truly tetrapolar.

Genomes of cereal rusts are highly heterozygous between haplotypes but chromosomes that contain mating type loci do not have overall higher heterozygosity than other chromosomes, which supports that recombination is not suppressed across entire chromosomes. We found strong evidence for localised recombination suppression surrounding the *PR* loci in conjunction with genomic degradation in all four species. This includes accumulation of TEs, significant reduction in gene density, and complete loss of synteny within and among species surrounding the *PR* loci. The TE families surrounding *PR* loci are not shared within or among species. The specific alleles of *PR* genes, *STE3.2* and *STE3.3*, and their linked *mfa* genes show strong signals of trans-specific polymorphisms. This suggests that recombination suppression around the *PR* loci is ancestral to these Pucciniomycotina and the overall structure of this locus supports accumulation of TEs. Yet the specific TE families that contribute to this phenomenon are distinct for each haplotype in each species and TE accumulation and/or replacement has continued after speciation. Similarly, in *Microbotryum* species the *PR* locus exhibits the oldest trans-specific polymorphism (Petit, et al. 2012; Branco, et al. 2017) , which supports that these loci are under strong selection for mate recognition.

In contrast, we did not observe strong recombination suppression at the *HD* loci, because these loci were syntenic among species. Further, signals of trans-specific polymorphisms and genomic degeneration were not found in *HD* loci. Our analysis might even indicate the potential recombination between haplotypes of *bE-HD* and *bW-HD* in *Pst*. This has been reported for several other Basidiomycota and might be mediated via the highly conserved intergenic promoter regions of *HD* loci haplotypes (Petit, et al. 2012; Branco, et al. 2017; Coelho, et al. 2017). Interestingly, gene trees based on *HD* loci and *STE3.3* were incongruent with the species tree. For these genes, *Pca* was sister to *Pgt* and *Pt* rather than *Pst*. This is in contrast to the accepted phylogenetic relationship among the four cereal rust fungi where *Pgt* and *Pt* are sister species, *Pst* shared a most recent common ancestor with *Pgt* and *Pt*, and *Pca* shared a most recent common ancestor with all three (Aime, et al. 2018). This incongruency between gene and species trees could be caused by incomplete lineage sorting, in which faster genetic change at the *HD* locus increases opportunities for outbreeding. Alternatively, there might have been more recent introgressions between *Pca* and *Pgt* or *Pt* populations. More extensive genome wide analyses are required to resolves these two alternate hypotheses.

We identified many more variations and likely haplotypes of *HD* genes than *PR* genes in all four species. This is consistent with many other Basidiomycota, which have multiallelic *HD* loci but biallelic *PR* loci including species in Agaricomycotina, Ustilaginomycotina, and Pucciniomyctoina (Urban, et al. 1996), (Coelho, et al. 2010). This diversification of *HD* loci haplotypes is preferred under tetrapolar mating that leads to outcrossing because it increases the likelihood of mate compatibility (Coelho, et al. 2017). Syngamy can occur between haploid cells carrying different *PR* haplotypes but the dikaryon aborts rapidly and fails to grow further when *HD* haplotypes are identical (Kües, et al. 2011; Nieuwenhuis, et al. 2013). Failure in passing this post-mating checkpoint is costly to both mating partners. High population level diversity at *HD* loci increases the likelihood of compatibility

downstream of *PR* matching and preserves resources at the whole population and species level (Hartmann et al., 2021, Nieuwenhuis et al., 2013).

The pheromone receptors encoded at biallelic *PR* locus respond to the *mfa* genes located at the opposite haplotype. For example, STE3.3 responds to peptides derived from *mfa2* which is linked to *STE3.2* (Coelho, et al. 2017). We identified one haplotype each of *STE3.*2, *STE3.3, mfa1* and *mfa2* among *Pst* isolates, suggesting the *PR* locus is biallelic in the populations sampled to date. We found a similar scenario for *Pca* and *Pt* where only a very limited number of non-synonymous variations occurred in *STE3.2, STE3.3*, *mfa1* and *mfa2*. We hypothesize that this variation at the amino acid level is functionally insignificant and that *Pca* and *Pt* are also truly biallelic at the *PR* locus. In contrast, we detected multiple haplotypes of *STE3.3* among *Pgt* isolates, with high genetic divergence between several isolates. In addition, *STE3.3* linked *mfa1* and *mfa3* also appear to be truly multiallelic. Also *STE3.2* displayed some variation at the species level, yet it is unclear if this variation is functionally important. A potential explanation for these observations is that the *PR* locus in *Pgt* is transitioning from a biallelic to multiallelic state with multiple compatible haplotypes that allow for more than one compatible haplotype pairing at this locus. Future studies are essential to test this hypothesis which require comprehensive sampling of both asexual and sexual populations of *Pgt* combined with long-read based chromosome scale and nuclear haplotype resolved assemblies.

One main characteristic of asexual spore production, e.g. urediniospores, is that Pucciniomycotina always combine nuclei of opposing mating types. This is despite the fact that nuclei of the same mating type share the same cytoplasm during colonisation of the asexual host. How this proper nuclear pairing and transmission via asexual spores is achieved was unknown. Our detailed transcriptomic analysis of *MAT* genes now gives an initial indication that these genes are important for nuclear pairing in the asexual stage, which is consistent with earlier preliminary reports (Cuomo, et al. 2017). Our analysis suggests that *MAT* genes are expressed in both sexual and asexual life cycle stages of cereal rust fungi. In our detailed transcriptomic analysis, we found both *HD* and *PR* genes were upregulated during the infection of the asexual host with a peak at final infection stages that align with spore formation and dispersal. *HD* genes are known to control post-mating growth, such as control development of fruiting-body in some Ascomycota and Basidiomycota species (Vonk and Ohm 2018). We hypothesized that *HD* gene function in nuclear pairing and post-plasmogamy in general cell proliferation during germination of spores, growing hyphae and formation of haustoria. In contrast, *PR* genes likely have a more restricted role in proper nuclear pairing and spore formation only. *PR* genes were not expressed in urediniospores, during germination and early on in infection.

Taken together, the knowledge that *MAT* loci govern compatibility of nuclei in somatic hybridisation and our detailed study of these loci in cereal rust fungi has an direct outcome for disease management through breeding. If somatic hybridisation occurs in the absence of parasexual reproduction, nuclear haplotypes of rust fungi are entire linkage groups, with effector haplotypes being tied to specific *MAT* haplotypes. Hence, we can predict the possible combinations of effector haplotypes that arise by somatic hybridisation based on the knowledge to which *MAT* haplotypes they are linked. For example, if we identify that the unrecognized haplotypes of an effector are only found

in linkage with *STE3.3* but not *STE3.2,* it is not possible that a new isolate, which can overcome the specific resistance gene, arises via somatic hybridization. Yet when unrecognized effector haplotypes are randomly distributed across nuclear types carrying either *STE3.3* or *STE3.2* and compatible *HD* loci, there is a strong possibility that a somatic hybridisation event can give rise to an isolate that overcomes the specific resistance gene. This observations requires us to generate chromosome scale and nuclear haplotype resolved genome assemblies of all cereal rust fungi at the population and species level. Such resources will improve the prediction of wheat rust resistance genes durability globally and lead to smarter rust resistance breeding programs.

Conclusion

In conclusion, our study provides novel insight into the evolution of *MAT* loci in key species of *Puccinia* on cereals. Cereal rust fungi likely have tetrapolar mating systems. There is no evidence of recombination suppression at *HD* loci, as there is no trans-specific polymorphisms between species and support for macro-synteny among species. The *PR* locus has extensive recombination suppression, which led to extensive genomic degeneration. Both *MAT* loci are likely involved in nuclear pairing during the dikaryotic asexual cycle. This detailed understanding is important to predict the possible novel lineages that may arise via somatic hybridisation that combine advantageous effector alleles linked to specific nuclear mating types. Future studies comparing *MAT* loci across taxa in the Pucciniales will be able to unmask the role of TEs in *MAT* evolution, and essential genes for sexual or asexual reproduction.

Acknowledgments

We thank Dr. P. Tobias for critical comments on manuscript, Dr. M. Bui for advice in iqtree2 usage, Prof. C. Linde for useful suggestions and R. Tam for providing a snakemake pipeline for trimming raw sequencing data and centromere information of *Pst 134E*. And finally, we thank J. Lin for suggestions in data visualization.

Material and methods

Identify *MAT* genes

Source of all species used in this study can be found in supplementary file (Table 1). *MAT* genes including the *HD* and *PR* genes previously identified in *P. triticina* (Cuomo, et al. 2017) were used as blastp and tblastn query to identify orthologs in *P. graminis* f.sp *tritici* (Ug99 and 21-0 isolates (Li, et al. 2019)), *P. striiformis* f.sp *tritici 134E* (Schwessinger, et al. 2022), *104E* (Schwessinger, et al. 2018) and *DK0911*), *P. triticina 76* (Duan, et al. 2021), *P. coronate* f.sp *avenae 203* (Henningsen, et al. 2022), *12NC29* and *12SD80* (Miller, et al. 2018) isolates. Missed *STE3*-like gene in *Pst 104E* was fixed by mapping short reads to reference genome *Pst 134E* with bwa-mem2 (Li and Durbin 2009). *Mfa* was identified by using *mfa* of *Pt BBBD*, *Pgt SCCL* and *Pst 78* (Cuomo, et al. 2017) as motifs to search orthologs in all studied species with Geneious Prime v. 2022.1.1 (2022.0.2).

Phylogenetic analysis of the *HD* and *PR* genes

Both nucleotide sequences and coding sequences of identified *HD* and *PR* genes were aligned by MAFFT (Katoh, et al. 2002) in plug-in Geneious Prime v. 2022.1.1 (2022.0.2).The phylogenetic tree were constructed by using IQ-TREE 2 (version 2.1.4-beta) (Minh, et al. 2020) with 1000 replicates and visualized via Dendroscope 3.7.6 (Huson and Scornavacca 2012).

Identification of variants in *MAT* loci

Sequence data of *Pt* used in this study have been published earlier by (Wu, et al. 2017; Wu, et al. 2020; Fellers, et al. 2021) and under PRJNA39803 and PRJNA39801, data of *Pst* were published by (Hubbard, et al. 2015; Ding, et al. 2021; van Schalkwyk, et al. 2022) and under PRJNA60743, data of *Pgt* were published by (Upadhyaya, et al. 2015; Chen, et al. 2017; Lewis, et al. 2018; Li, et al. 2019) and under PRJNA39437, data of *Pca* were published by (Miller, et al. 2020), all above data were downloaded from NCBI SRA database (URL: https://www.ncbi.nlm.nih.gov/sra) (Table S2). *Pgt 210* (Li, et al. 2019), *Pst 134E* (Schwessinger, et al. 2022), *Pca 203* (Henningsen, et al. 2022) and *Pt* 76 (Duan, et al. 2021) three isolates which have published chromosome-scale haplotype-phased genome assembly were used as reference for mapping. Trimmomatic (Bolger, et al. 2014) was used to remove adapters from raw reads whereas FastQC (Andrews 2010) was applied for assessing the quality before and after trimming. Bwa-mem2 (Li and Durbin 2009) was used to map trimmed reads to reference genome, MarkDuplicates (Picard) (2019) was applied to remove PCR-generated duplicates. Qualimap 2 (Okonechnikov, et al. 2016) was used to check mapping coverage and quality. Freebayes-parallel (Garrison and Marth 2012; Tange 2018) was processed to detect variants, samtools (Li, et al. 2009) was used to filter out regions of interest. The Ensembl Variant Effect Predictor (VEP) (McLaren, et al. 2016) was used to annotate sequence variants. Bcftools (Li, et al. 2009) was used to generate consensus sequence, nucleotide and protein alignments were generated and visualized by Geneious Prime v. 2022.1.1 (2022.0.2).

Investigation of recombination suppression in *MAT* loci

Inspired by the approach used by Brance et al. (Branco, et al. 2017) for identifying evolutionary strata on mating-type chromosomes of Microbotryum species, proteinortho (Lechner, et al. 2011) was used to pair orthologs one-to-one. PAML (Yang 2007) was used to calculate synonymous divergence (dS) values for identified gene pairs. Genomes was split in 10kb per window, then density of genes and repeats were calculated by bedtools v 2.29.2 (Quinlan and Hall 2010).

Synteny analysis of *MAT* loci and flanking regions

Self-alignment analyses between two haplotypes of each species were done with NUCmer (version 3.1) (Marçais, et al. 2018) and visualize via ggplot2 (Villanueva and Chen 2019).

Nucmer (Marçais, et al. 2018) was used to align nucleotide sequences of *MAT* regions, REPET pipeline (Quesneville, et al. 2005; Flutre, et al. 2011) was applied to annotate repeats of all genomes used in this study. Blastp (Johnson, et al. 2008) was used to identified conserved genes (identity

higher than 70%) across and within species. Results were visualized with gggenomes (Thomas Hackl 2022).

Comparison of expression of *MAT* genes in sexual and asexual reproduction

Reads of rust fungi on infected plants were obtained from NCBI SRA database (URL: https://www.ncbi.nlm.nih.gov/sra) including RNA-seq data of *Pst CYR32* on primary and alternate host under PRJNA637808 (Zhao, et al. 2021), *Pst* 87/66 on infected wheat under PRJEB12497 (Dobon, et al. 2016), *Pt* 77-5 on wheat under SRP078210 (Yadav, et al. 2016), *Pca 12NC29* on *Brachypodium distachyon* under PRJNA398546 (Miller, et al. 2018), *Pgt* 21-0 on infected wheat under PRJNA415866 (Chen, et al. 2017), *Pst 104E* under PRJNA396589 (Schwessinger, et al. 2018). Raw reads were checked by using FastQC v0.11.8 (Andrews 2010), trimmomatic 0.39 was used to trim low quality reads with parameters LEADING:3 TRAILING:3 MINLEN:36 (Bolger, et al. 2014). Kallisto 0.44.0 (Bray, et al. 2016) was then applied to align and count transcripts, cDNA data of *Pgt 21-0*, *Pst 134E*, *Pt 76* and *Pca 203* were used as reference. DESeq2 (Love, et al. 2014) was used to normalize and ggplot2 (Wickham 2016) was used to visualize the result.

Data Availability

Analysis code used in this study is available at github repository: https://github.com/ZhenyanLuo/MAT_scripts.git

References

Aime M, McTaggart A. 2021. A higher-rank classification for rust fungi, with notes on genera. Fungal systematics and evolution 7:21-47.

Aime MC, Bell CD, Wilson AW. 2018. Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. Studies in Mycology 89:143-152.

Andersen GR, Nissen P, Nyborg J. 2003. Elongation factors in protein biosynthesis. Trends in biochemical sciences 28:434-441.

FastQC: a quality control tool for high throughput sequence data [Internet]. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Au SW, Leng X, Harper JW, Barford D. 2002. Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit. Journal of molecular biology 316:955-968.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. Current opinion in genetics & development 16:578-585.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Branco S, Badouin H, de la Vega RCR, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg J-T, Coelho MA, Hood ME. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proceedings of the National Academy of Sciences 114:7067-7072.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology 34:525-527.

Carpentier F, de la Vega RCR, Branco S, Snirc A, Coelho MA, Hood ME, Giraud T. 2019. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. Genome research 29:944-953.

Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, Breen S, Dong C, Xu B, Zhang X. 2017. Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. Science 358:1607-1610.

Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the Basidiomycota. The fungal kingdom:147-175.

Coelho MA, Sampaio JP, Goncalves P. 2010. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. PLoS genetics 6:e1001052.

Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, Sakthikumar S, Song X, Adiconis X, Fan L. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. G3: Genes, Genomes, Genetics 7:361-376.

Ding Y, Cuddy WS, Wellings CR, Zhang P, Thach T, Hovmøller MS, Qutob D, Brar GS, Kutcher HR, Park RF. 2021. Incursions of divergent genotypes, evolution of virulence and host jumps shape a continental clonal population of the stripe rust pathogen *Puccinia striiformis*. Molecular Ecology 30:6566-6584.

Dobon A, Bunting DC, Cabrera-Quio LE, Uauy C, Saunders DG. 2016. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC genomics 17:1-14.

Duan H, Jones A, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R, Upadhyaya N, Rathjen J. 2021. Identification and correction of phase switches with Hi-C data in the Nanopore and HiFi chromosome-scale assemblies of the dikaryotic leaf rust fungus *Puccinia triticina*. bioRxiv.

Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proceedings of the National Academy of Sciences 108:9166-9171.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology 20:1-14.

Fellers JP, Sakthikumar S, He F, McRell K, Bakkeren G, Cuomo CA, Kolmer JA. 2021. Whole-genome sequencing of multiple isolates of *Puccinia triticina* reveals asexual lineages evolving by recurrent mutations. G3 11:jkab219.

Ferrarezi J, McTaggart A, Tobias P, Hayashibara C, Degnan R, Shuey L, Franceschini L, Lopes M, Quecine M. 2022. *Austropuccinia psidii* uses tetrapolar mating and produces meiotic spores in older infections on Eucalyptus grandis. Fungal Genetics and Biology:103692.

Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in *de novo* annotation approaches. PloS one 6:e16526.

Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS biology 2:e384.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Geneious Prime 2022.1.1. Available from: https://www.geneious.com

Glass NL, Kuldau GA. 1992. Mating type and vegetative incompatibility in filamentous ascomycetes. Annual review of phytopathology 30:201-224.

Hartmann FE, Ament-Velásquez SL, Vogan AA, Gautier V, Le Prieur S, Berramdane M, Snirc A, Johannesson H, Grognet P, Malagnac F. 2021. Size variation of the nonrecombining region on the mating-type chromosomes in the fungal Podospora anserina species complex. Molecular biology and evolution 38:2475-2492.

Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F, Grognet P, Giraud T. 2021. Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. New Phytologist 229:2470-2491.

Henningsen EC, Hewitt T, Dugyala S, Nazareno E, Gilbert E, Li F, Kianian S, Steffenson BJ, Dodds PN, Sperschneider J. 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. bioRxiv.

Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: Microbotryum violaceum and other model organisms. Genetics 168:141-146.

Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility factors: Tetrapolarity in Microbotryum. Evolution 69:2561-2572.

Hubbard A, Lewis CM, Yoshida K, Ramirez-Gonzalez RH, de Vallavieille-Pope C, Thomas J, Kamoun S, Bayles R, Uauy C, Saunders DG. 2015. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. Genome biology 16:1-15.

Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Systematic biology 61:1061-1067.

Idnurm A, Hood ME, Johannesson H, Giraud T. 2015. Contrasted patterns in mating-type chromosomes in fungi: hotspots versus coldspots of recombination. Fungal Biology Reviews 29:220-229.

James TY. 2007. Analysis of mating-type locus organization and synteny in mushroom fungi: Beyond model species. Sex in fungi: molecular determination and evolutionary implications:317-331.

James TY, Srivilai P, Kues U, Vilgalys R. 2006. Evolution of the bipolar mating system of the mushroom Coprinellus disseminatus from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. Genetics 172:1877-1891.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic acids research 36:W5-W9.

Katoh K, Misawa K, Kuma Ki, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research 30:3059-3066.

Kües U, James TY, Heitman J. 2011. 6 Mating type in basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. In. Evolution of fungi and fungal-like organisms: Springer. p. 97-160.

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC bioinformatics 12:1-9.

Lewis CM, Persoons A, Bebber DP, Kigathi RN, Maintz J, Findlay K, Bueno-Sancho V, Corredor-Moreno P, Harrington SA, Kangara N. 2018. Potential for re-emergence of wheat stem rust in the United Kingdom. Communications biology 1:1-9.

Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C, Miller ME, Silverstein KA, Henningsen E. 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. Nature communications 10:1-15.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15:1-21.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. PLoS computational biology 14:e1005944.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The ensembl variant effect predictor. Genome biology 17:1-14.

McTaggart AR, James TY, Idnurm A, Park RF, Shuey LS, Demers MN, Aime MC. 2022. Sexual reproduction is the null hypothesis for life cycles of rust fungi. PLoS Pathogens 18:e1010439.

Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. 2008. The mating-type chromosome in the filamentous ascomycete Neurospora tetrasperma represents a model for early evolution of sex chromosomes. PLoS genetics 4:e1000030.

Miller ME, Nazareno ES, Rottschaefer SM, Riddle J, Dos Santos Pereira D, Li F, Nguyen-Phuc H, Henningsen EC, Persoons A, Saunders DG. 2020. Increased virulence of Puccinia coronata f. sp. avenae populations through allele frequency changes at multiple putative Avr loci. PLoS genetics 16:e1009291.

Miller ME, Zhang Y, Omidvar V, Sperschneider J, Schwessinger B, Raley C, Palmer JM, Garnica D, Upadhyaya N, Rathjen J. 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:e01650-01617.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution 37:1530-1534.

National Center for Biotechnology Information (NCBI)[Internet]. Available from: https://www.ncbi.nlm.nih.gov/

Nieuwenhuis BP, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013. Evolution of uni-and bifactorial sexual compatibility systems in fungi. Heredity 111:445-455.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292-294.

Petit E, Giraud T, de Vienne DM, Coelho MA, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory F, Wincker P. 2012. Linkage to the mating-type locus across the genus microbotryum: insights into nonrecombining chromosomes. Evolution 66:3519-3533.

Picard toolkit [Internet]. Broad Institute; 2019. Available from: https://github.com/broadinstitute/picard

Pontecorvo G. 1956. The parasexual cycle in fungi. Annual review of microbiology 10:393-400.

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol 1:e22.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Raudaskoski M, Kothe E. 2010. Basidiomycete mating type genes and pheromone signaling. Eukaryotic cell 9:847.

Schwessinger B, Chen Y-J, Tien R, Vogt JK, Sperschneider J, Nagar R, McMullan M, Sicheritz-Ponten T, Sørensen CK, Hovmøller MS. 2020. Distinct life histories impact dikaryotic genome evolution in the rust fungus Puccinia striiformis causing stripe rust in wheat. Genome biology and evolution 12:597-617.

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. Molecular Plant-Microbe Interactions 35:293-296.

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:e02275-02217.

Sperschneider J, Hewitt T, Lewis D, Periyannan S, Milgate A, Hickey L, Mago R, Dodds P, Figueroa M. 2022. Extensive somatic nuclear exchanges shape global populations of the wheat leaf rust pathogen *Puccinia triticina*. bioRxiv:2022.2011. 2028.518271.

Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong C, Upadhyaya NM, Mago R, Hu Y, Figueroa M. 2021. The stem rust fungus *Puccinia graminis* f. sp. *tritici* induces centromeric small RNAs during late infection that are associated with genome-wide DNA methylation. BMC biology 19:1-25.

Sun S, Coelho MA, Heitman J, Nowrousian M. 2019. Convergent evolution of linked mating-type loci in basidiomycete fungi. PLoS genetics 15:e1008365.

Tange O. 2018. GNU parallel 2018: Lulu. com.

gggenomes: A Grammar of Graphics for Comparative Genomics [Internet]. 2022. Available from: https://github.com/thackl/gggenomes

Upadhyaya NM, Garnica DP, Karaoglu H, Sperschneider J, Nemri A, Xu B, Mago R, Cuomo CA, Rathjen JP, Park RF. 2015. Comparative genomics of Australian isolates of the wheat stem rust

pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. Frontiers in plant science 5:759.

Urban M, Kahmann R, Bölker M. 1996. The biallelica mating type locus of *Ustilago maydis*: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. Molecular and General Genetics MGG 250:414-420.

van Schalkwyk HJ, Adams T, Persoons A, Boshoff WH, Wanyera R, Hovmøller MS, Uauy C, Boyd L, Pretorius ZA, Prins R. 2022. Pathogenomic analyses of *Puccinia striiformis* f. sp. *tritici* supports a close genetic relationship between South and East Africa. Plant Pathology 71:279-288.

Villanueva RAM, Chen ZJ. 2019. ggplot2: elegant graphics for data analysis. In: Taylor & Francis.

Vonk PJ, Ohm RA. 2018. The role of homeodomain transcription factors in fungal development. Fungal Biology Reviews 32:219-230.

Wang MN, Chen X. 2013. First report of Oregon grape (*Mahonia aquifolium*) as an alternate host for the wheat stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*) under artificial inoculation. Plant Disease 97:839-839.

Wickham H. 2016. ggplot2: elegant graphics for data analysis: springer.

Wu JQ, Dong C, Song L, Park RF. 2020. Long-Read–Based *de novo* Genome Assembly and Comparative genomics of the wheat leaf rust pathogen *Puccinia triticina* identifies candidates for three avirulence genes. Frontiers in genetics 11:521.

Wu JQ, Sakthikumar S, Dong C, Zhang P, Cuomo CA, Park RF. 2017. Comparative genomics integrated with association analysis identifies candidate effector genes corresponding to Lr20 in phenotype-paired *Puccinia triticina* isolates from Australia. Frontiers in plant science 8:148.

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. Frontiers in Microbiology:2180.

Yadav IS, Sharma A, Kaur S, Nahar N, Bhardwaj SC, Sharma TR, Chhuneja P. 2016. Comparative temporal transcriptome profiling of wheat near isogenic line carrying Lr57 under compatible and incompatible interactions. Frontiers in plant science 7:1943.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution 24:1586-1591.

Zhao J, Duan W, Xu Y, Zhang C, Wang L, Wang J, Tian S, Pei G, Zhan G, Zhuang H. 2021. Distinct transcriptomic reprogramming in the wheat stripe rust fungus during the initial infection of wheat and barberry. Molecular Plant-Microbe Interactions 34:198-209.

Tables

Table 1. Genomes used to search for homologs of *MAT* genes

Genome Source				
Species	**Strain**	**Source**	**Download date**	**Citation**
Puccinia graminis f. sp. *tritici*	*Ug99*	NCBI	26/05/2021	Li et al. 2019
Puccinia graminis f. sp. *tritici*	*21-0*	NCBI	26/05/2021	Li et al. 2019
Puccinia striiformis f. sp. *tritici*	*DK0911*	JGI	1/06/2021	Schwessinger et al. 2020

Puccinia striiformis f. sp. *tritici*	*104E*	JGI	26/05/2021	Schwessinger et al. 2018
Puccinia striiformis f. sp. *tritici*	*134E 16*	From Schwessinger B.	21/06/2021	Schwessinger et al. 2022
Puccinia coronata f. sp. *avenae*	*12NC29*	NCBI	26/05/2021	Miller et al. 2018
Puccinia coronata f. sp. *avenae*	*12SD80*	NCBI	26/05/2021	Miller et al. 2018
Puccinia coronata f. sp. *avenae*	*203*	CSIRO data access portal	20/06/2022	Henningsen et al. 2022
Puccinia triticina	*76*	Cloudstor	28/05/2021	Duan et al. 2022

Table 2. Reference genes used as query and outgroups

Reference Source					
Species	**Strain**	**Source**	**Gene name**	**Gene ID**	**Citation**
Puccinia triticina	*BBBD*	NCBI & Cuomo et al., 2017	*bE-HD*	PTTG_10928 PTTG_03697 PTTG_09683	Cuomo et al., 2017
			bW-HD	PTTG_27730	
			STE3.2	PTTG_09751	
			STE3.1	PTTG_28830	
			STE3.3	PTTG_09693	
			mfa1=3	-	
			mfa2	-	
Melampsora larici-populina	98AG31	NCBI	*bE-HD*	XP_007413298	https://www.ncbi.nlm.nih.gov/
			bW-HD	XP_007413233	
			STE3.4	XP_007419438	
			STE3.2	XP_007418099	
			STE3.1	XP_007405301	
			mfa1	XP_007419440	
			mfa2	XP_007418113	

Figure legends

Figure 1: Gene trees of *HD* haplotypes in rust fungi suggests distinct evolutionary history of *bW-HD* and *bE-HD*.

(A) The phylogenetic tree of *bW-HD* hapolotypes in *Puccinia* species. (B) The phylogenetic tree of *bE-HD* haplotypes in *Puccinia* species. This shows a maximum likelihood gene tree based on the nucleotide alignment of *HD* genes. Branch support was assessed by 1000 replicates, *bW-HD* and *bE-HD* (XP_007413298 and XP_007413298) of *Melampsora larici-populina 98AG31* were used as outgroup for building rooted trees. The scale bar represents 0.1 substitutions per nucleotide site.

Figure 2: Phylogenetic trees of *STE3.2* and *STE3.3* show strong transpolymorphism signal.

(A) Phylogenetic tree of *STE3.2* in *Puccinia* species (B) Phylogenetic tree of *STE3.3* in *Puccinia* species. This shows a maximum likelihood gene tree based on the nucleotide alignment of *PR* genes. *STE3.2* and *STE3.4* (XP_007418099 and XP_007419438) of *Melampsora larici-populina 98AG31*

were used as outgroup for rooting. Branch support was assessed by 1000 replicates. The scale bar represents 0.1 substitutions per nucleotide site.

Figure 3: Synonymous divergence (dS) values of *HD* genes but not their immediate neighbours are slightly elevated on chromosome 4.

In each panel the top plot shows the dS values of gene pairs along chromosome 4. Each dot represents the dS value for a single gene pair. The middle and bottom plot shows the average gene density and TE density along chromosome 4 in 10k bp sized windows respectively. The *HD* gene pairs were labelled in red (*bW-HD*) and green (*bE-HD*) respectively. Centromeric regions of *Pst 134E* and *Pgt 210* are indicated with red line.

Synonymous divergence for all gene pairs are plotted along chromosome 4A for (A) *P. coronara* f.sp. *avenae,* (B) *P. gramminis* f.sp. *tritici,* (C) *P. striiformis* f.sp. *tritici,* and (D) *P. triticina*.

Figure 4: *HD* loci are partially conserved within species whereas *PR* loci are highly heterozygous and display strong signals of genomic degeneration.

(A) Synteny of *HD* locus and relative flanking regions in *Pca 203*, *Pgt 210, Pst 134E* and *Pt 76*. (B) Synteny of *PR* loci and their flanking regions in *Pca 203*, *Pst 134E*, *Pgt 210* and *Pt 76*. The accumulation of transposons disrupts synteny around *PR* loci in *Pgt 210* and *Pt 76*. The low similarity rate between haplotypes suggests inserted TEs are not identical around *PR* loci. Red lines between chromosome sections represent gene pairs with identity higher than 70%, grey shades represent conserved nucleotide sequences and green shades on chromosome sections represent transposons.

Figure 5: Synonymous divergence (dS) rate of *PR* loci are significantly elevated on chromosome 9.

In each panel the top plot shows the dS values of gene pairs along chromosome 9. Each dot represents the dS value for a single gene pair. The middle and bottom plot shows the average gene density and TE density along chromosome 9 in 10k bp sized windows respectively. The *STE3.2*/*STE3.3* gene pair were labelled in red. Centromeric regions of *Pst 134E* and *Pgt 210* are indicated with red line.

Synonymous divergence for all gene pairs are plotted along chromosome 9A for (A) *P. coronara* f.sp. *avenae,* (B) *P. gramminis* f.sp. *tritici,* (C) *P. striiformis* f.sp. *tritici,* and (D) *P. triticina*.

Figure 6: *MAT* genes are expressed during the sexual and asexual infection cycle.

(A) Differential gene expression of *HD* (left) and *PR* (right) genes on primary host and alternate host of *Pst*. (B) Expression of *HD* genes on primary hosts of *Pt, Pst, Pca* and *Pgt*. (C) Expression of *STE3.2* and *STE3.3* on primary hosts of *Pt, Pst, Pca* and *Pgt*. (D) Expression of *MAT* genes in asexual life cycle of *Pst* suggested *MAT* genes upregulated during infection but little or not expressed in mature or germinating spores. (UG: ungerminated urediniospores, GS: germinated urediniospores, H: haustoria). Dots with different colours represent the mean of normalized transcript count of individual genes, dots with light colour represent individual replicates.

Supplemental material legends

Table S1 Counts of shared TE families across haplotypes

Table S2 List of SRA data

Figure S1 Unlinked *MAT* loci suggest tetrapolar mating system in four *Puccinia spp*.

Karyograms of *Pca 203, Pst 134E, Pt 76* and *Pgt 210* with positions of *HD, PR* and *STE3.1* loci marked by black arrow heads. *HD*, *PR* and *STE3.1* all located on chromosome 4, chromosome 9 and chromosome 1 respectively, suggesting tetrapolar mating system for these four species.

Figure S2 Protein alignments indicate mating factor a (*mfa*) are conserved within species.

(A) Alignments of MFA1 cross species. The MFA1*,* which is linked to *STE3.3,* is conserved within species, two identical MFA1 can be identified in *Pt 76*. Compared to MFA1 in *Pgt SCCL* (Cuomo, et al. 2017), MFA1 in *Pgt 210* has four amino acid substitutions. (B) Alignments of MFA2 identified in *Pst*, *Pt*, *Pgt* and *Pca* isolates. MFA2 is highly conserved within species. (C) Alignments of MFA3 identified in *Pgt* isolates. MFA3 identified was used as reference and only can be detected close to *STE3.3* of both *Pgt* isolates, a single amino acid substitution was identified.

Figure S3 Genetic distance between pheromone precursor genes (*mfa*) and pheromone receptor (*PR*) genes vary from hundred to hundred thousand base pairs.

The diagrams display the location of *STE3.2* and *STE3.3* and there linked *mfa* genes on chromosome 9A and 9B. The grey bar represents chromosome subsections containing the genes of interest with the numbers indicating the absolute location in mega base pair on chromosome 9A and 9B.

The genetic distance between *mfa1/mfa3* and *STE3.3* are highly variable between species whereas *STE3.2* and *mfa2* are tightly linked in all species.

Figure S4 Phylogenetic analysis indicates pheromone peptide mating factor a *(mfa*) group according to their linkage to *PR* haplotypes and reveals a strong trans-specific polymorphism signal.

Pca and *Pst* only have one *mfa* in both PR loci, *Pt* has two identical *mfa* around *STE3.3* and *Pgt* has two different *mfa* around *STE3.3*. The *mfa2s* linked to *STE3.2* were grouped in the same clade, and *mfa1/3* linked to *STE3.3* were grouped into one clade. This second clade was split into subgroups where one contains except *mfa3* which are unique to *Pgt*. This shows a maximum likelihood gene tree based on the nucleotide alignment of *mfa*. *Mfa-1* (XP_007419440) and *mfa-2* (XP_007418113) of *M. larici-populina 98AG31* were used as outgroup for rooting. Branch support was assessed by 1000 replicates. The scale bar represents 0.1 substitutions per nucleotide site.

Figure S5 Gene tree of *STE3.1* in rust fungi indicates sequence of *STE3.1* are conserved within species.

This shows a maximum likelihood gene tree based on the nucleotide alignment of *STE3.1*. *STE3.1* (XP_007405301) of *M. larici-populina 98AG31* were used as outgroup for rooting. Branch support was assessed by 1000 replicates. The scale bar represents 0.1 substitutions per nucleotide site.

Two haplotypes of *STE3.1* can be identified in all species and the nucleotide sequences are highly conserved within species.

Figure S6 Whole genome alignments of *HD* loci containing chromosomes between two haplotypes of *Pca 203*, *Pst 134E, Pgt 210* and *Pt 76* exhibit slightly reduced synteny around *HD* locus.

These graphs show whole genome alignments with mummer of chromosome 4A to chromosome 4B. Larger boxes represent chromosome-scale comparison, smaller boxes represent comparison of 20 kbp upstream and downstream of *HD* locus, yellow boxes indicate *HD* loci.

(A) Comparison of nucleotide sequence of chromosome 4A and chromosome 4B of *Pca 203*. (B) Comparison of nucleotide sequence of chromosome 4A and 4B of *Pgt 210*. (C) Comparison of nucleotide sequence of chromosome 4A and 4B of *Pst 134E*. (D) Comparison of nucleotide sequence of chromosome 4A and 4B of *Pt 76*.

Figure S7. Genes pairs on chromosomes 4 and 9 are not overall more diverged when compared to other chromosomes.

Bar plot show threshold of ds value of 99% of genes, red lines represent threshold of ds value of 95% of genes and blue lines represent that of 90% of genes, individual red points represent *HD* and *STE* genes.

The plots show the distribution of ds value capturing synonymous changes of gene pairs located on sister chromosomes in (A) *Pca 203*, (B) *Pst 134E*, (C) *Pgt 210* and (D) *Pt 76*. The ds value of *bW-HD* and *bE-HD* is higher than 95% but smaller than 99% of that of other genes pairs located on chromosome 4. The ds value between *STE3.2* and *STE3.3* is higher than 99% of other genes on same chromosome. The ds value of *STE3.1* is close to 0 in all four species.

Figure S8 Compositions of TE classes are different in *HD*, *PR* and *STE3.1* across species.

This plot shows coverage of different classes of transposons around 40 genes flanking (A) *HD* locus (B) *PR* locus (C) *STE3.1* in *Pca 203*, *Pst 134E*, *Pgt 210* and *Pt 76*.

(A) Coverage of different classes of transposons around 40 genes flanking *HD* locus suggesting TE content are various between haplotypes and species. Coverage of DNA transposons are different across species, RNA transposons occupy larger proportion in *Pca* and *Pgt* whereas DNA transposons have much higher coverage rate than RNA transposons in *Pst*. (B) Coverage of different classes of transposons around 40 genes flanking *PR* locus suggesting TE content are various between haplotypes and species. (C) Coverage of different classes of transposons around 40 genes flanking *STE3.1* locus suggesting TE content are various between haplotypes and species.

Figure S9 Alignment of *PR* locus containing chromosomes between two haplotypes of *Pca 203*, *Pst 134E, Pgt 210* and *Pt 76* exhibits a long synteny gap around *PR* locus.

These graphs show whole genome alignments with mummer of chromosome 9A to chromosome 9B. Larger boxes represent chromosome-scale comparison, smaller boxes represent comparison of 20 kbp upstream and downstream of *PR* locus, yellow boxes indicate *PR* loci.

Long synteny gaps *around* PR locus can be identified in all four species. (A) Comparison of nucleotide sequence of chromosome 9A and 9B of *Pca 203*. (B) Comparison of nucleotide sequence of chromosome 9A and 9B of *Pgt 210*. (C) Comparison of nucleotide sequence of chromosome 9A and 9B of *Pst 134E*. (D) Comparison of nucleotide sequence of chromosome 9A and 9B of *Pt 76*.

Figure S10 Alignment of *STE3.1* containing contigs/chromosomes suggesting conserved synteny around *STE3.1* loci.

These graphs show whole genome alignments with mummer of chromosome 1A to chromosome 1B. Larger boxes represent chromosome-scale comparison, smaller boxes represent comparison of 20 kbp upstream and downstream of *STE3.1*, yellow boxes indicate *HD* loci.

(A) Comparison of nucleotide sequence of chromosome 1A and 1B of *Pca 203*. (B) Comparison of nucleotide sequence of chromosome 1A and 1B of *Pgt 210*. (C) Comparison of nucleotide sequence of chromosome 1A and 1B of *Pst 134E*. (D) Comparison of nucleotide sequence of chromosome 1A and 1B of *Pt 76*.

Figure S11 Synteny analysis of *STE3.1* and their flanking regions reveal no evidence of genetic degeneration.

This plot shows syntenic relationship of *STE3.1* and their flanking regions among *Pca 203*, *Pst 134E*, *Pgt 210* and *Pt 76*. Red lines between contigs represent gene pairs with identity higher than 70%, grey shades represent conserved nucleotide and green shades on contigs represent transposons.

Flanking regions lack of synteny between species, but more conserved genes can be identified across species than *HD* locus and *PR* loci. Compared to other three species, gene density around *STE3.1* in both *Pt 76* haplotypes are relatively lower with higher coverage of transposons.

Figure S12 Synonymous divergence (dS) rate between alleles along the chromosome 1A suggesting absence of evolutionary strata.

In each panel the top plot shows the dS values of gene pairs along chromosome 1. Each dot represents the dS value for a single gene pair. The middle and bottom plot shows the average gene density and TE density along chromosome 1 in 10k bp sized windows respectively. The *STE3.1* gene pairs were labelled in red. Centromeric regions of *Pst 134E* and *Pgt 210* are indicated with red line.

STE3.1 has low ds value in all four species, almost close to 0. Only in *Pt 76*, gene density drops around *STE3.1* loci. (A) Synonymous divergence for all gene pair of *P. coronara* f.sp. *avenae* plotted according to the gene location along chromosome 1A. (B) Synonymous divergence for all gene pair of

P. gramminis f.sp. *tritici* plotted according to the gene location along chromosome 1A. (C) Synonymous divergence for all gene pair of *P. striiformis* f.sp. *tritici* plotted according to the gene location along chromosome 1A. (D) Synonymous divergence for all gene pair of *P. triticina* plotted according to the gene location along chromosome 1A.

Figure S13. Nucleotide alignments of *HD* locus indicate *bW-HD* and *bE-HD* are highly allelic at the population level.

This plot shows nucleotide alignments of *HD* locus from different isolates in *Pca*, *Pst*, *Pgt* and *Pt* constructed by Geneious Primes v2022.0.1, the green blocks represent identical nucleotide sequence, yellow lines represent occurrence of SNPs. Genes are labelled in blue.

SNPs can be identified around and within *HD* locus, suggesting *HD* locus are highly allelic at the population level in all for species.

Figure S14. Nucleotide alignments of *PR* locus suggesting potential biallelic *PR* locus in *Pst*.

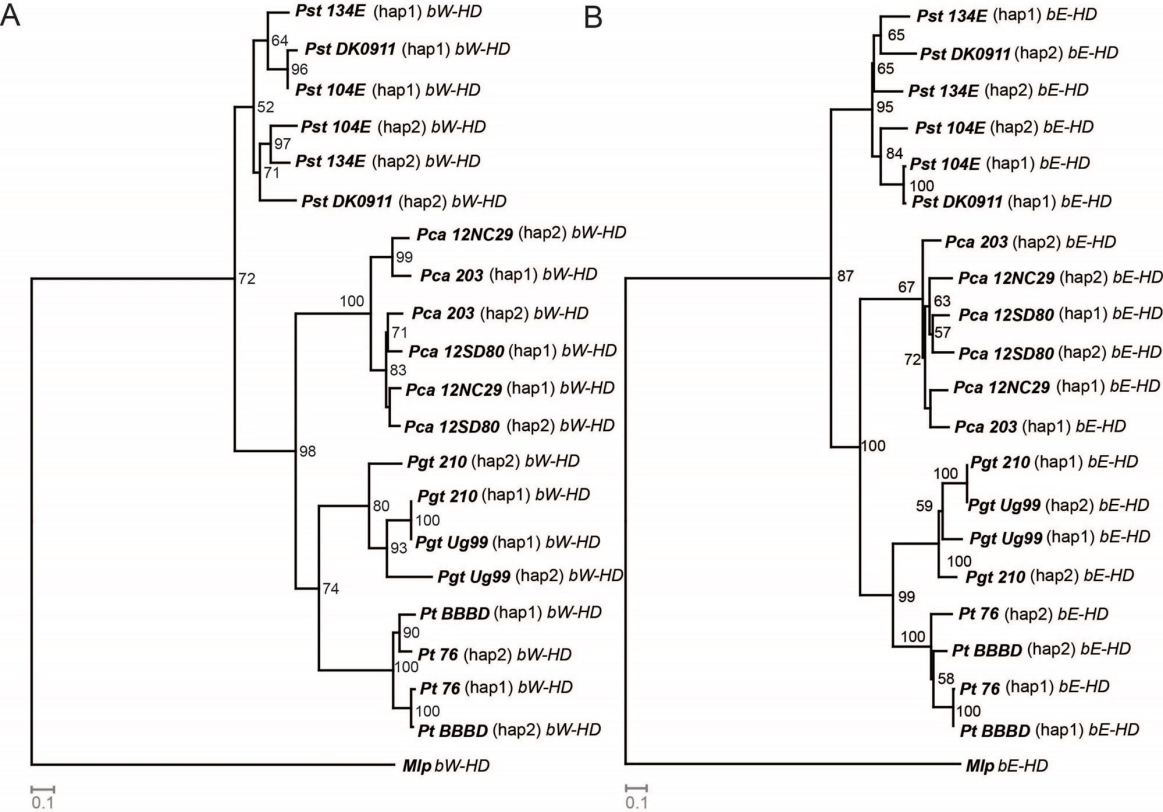
This plot shows nucleotide alignments of *PR* locus from different isolates in *Pca*, *Pst*, *Pgt* and *Pt* constructed by Geneious Primes v2022.0.1, the green blocks represent identical nucleotide sequence, yellow lines represent occurrence of SNPs. Genes are labelled in blue.

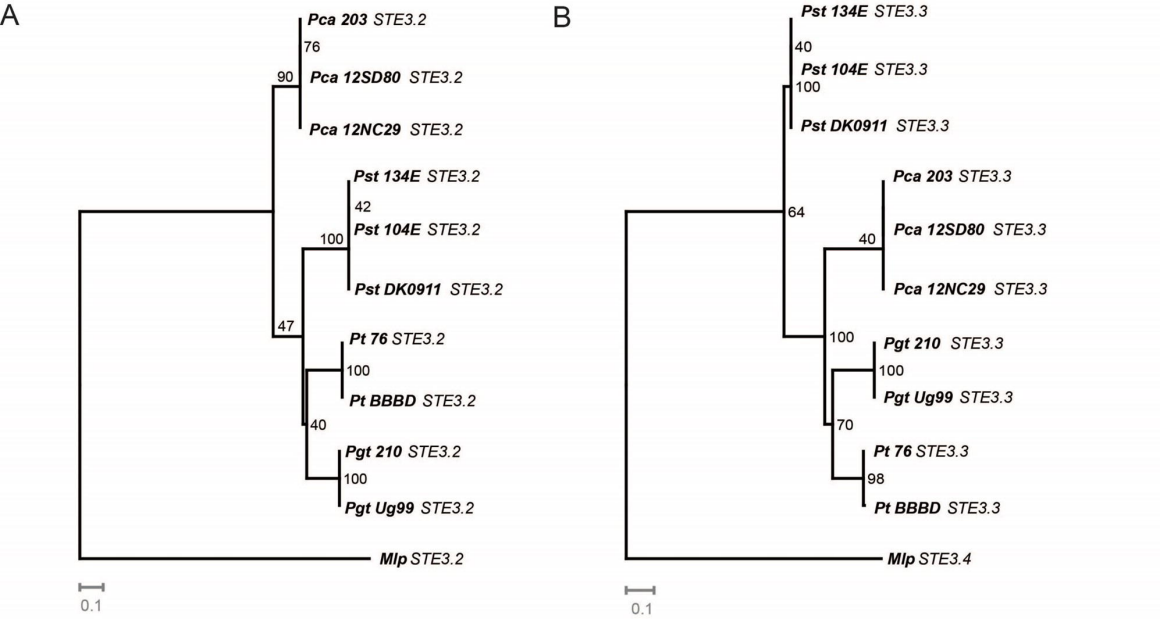
Multiple variations can be identified in *STE3.2* and *STE3.3* of *Pgt* and *Pt*, whereas cds of all *STE3.2* and *STE3.3* are identical in *Pst*. Only two SNPs could be identified in both *STE3.2* and *STE3.3* of *Pca* with only one leads to non-synonymous mutation.

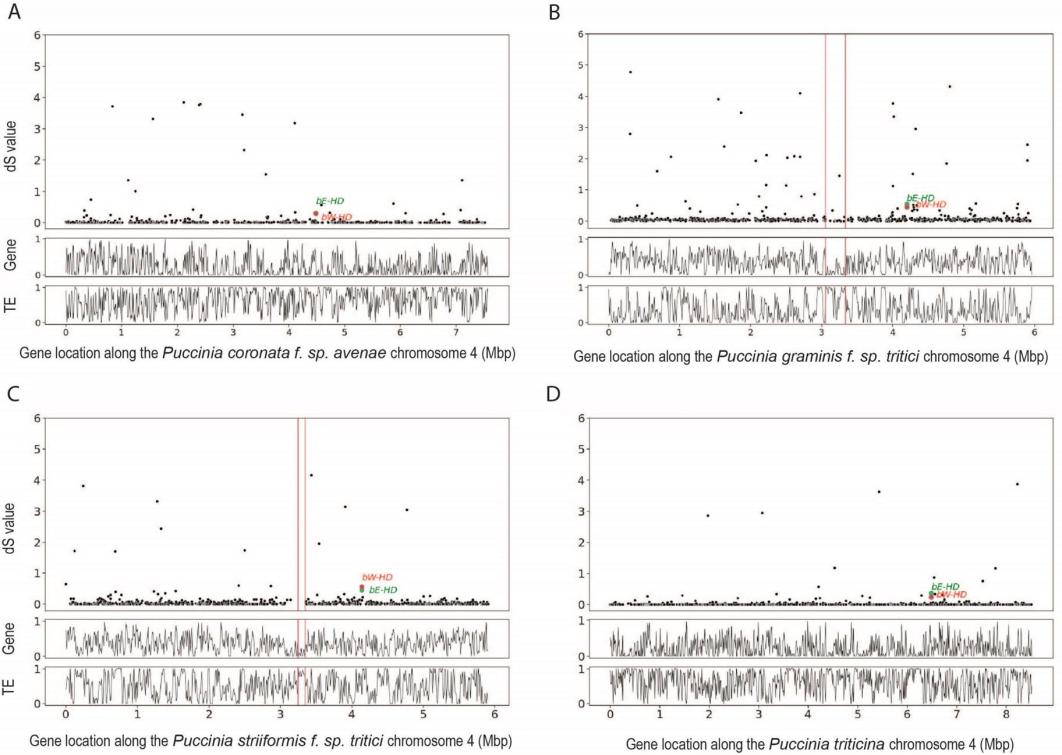
Figure S15 Nucleotide alignments of *STE3.1* locus suggesting *STE3.1* are conserved within species.

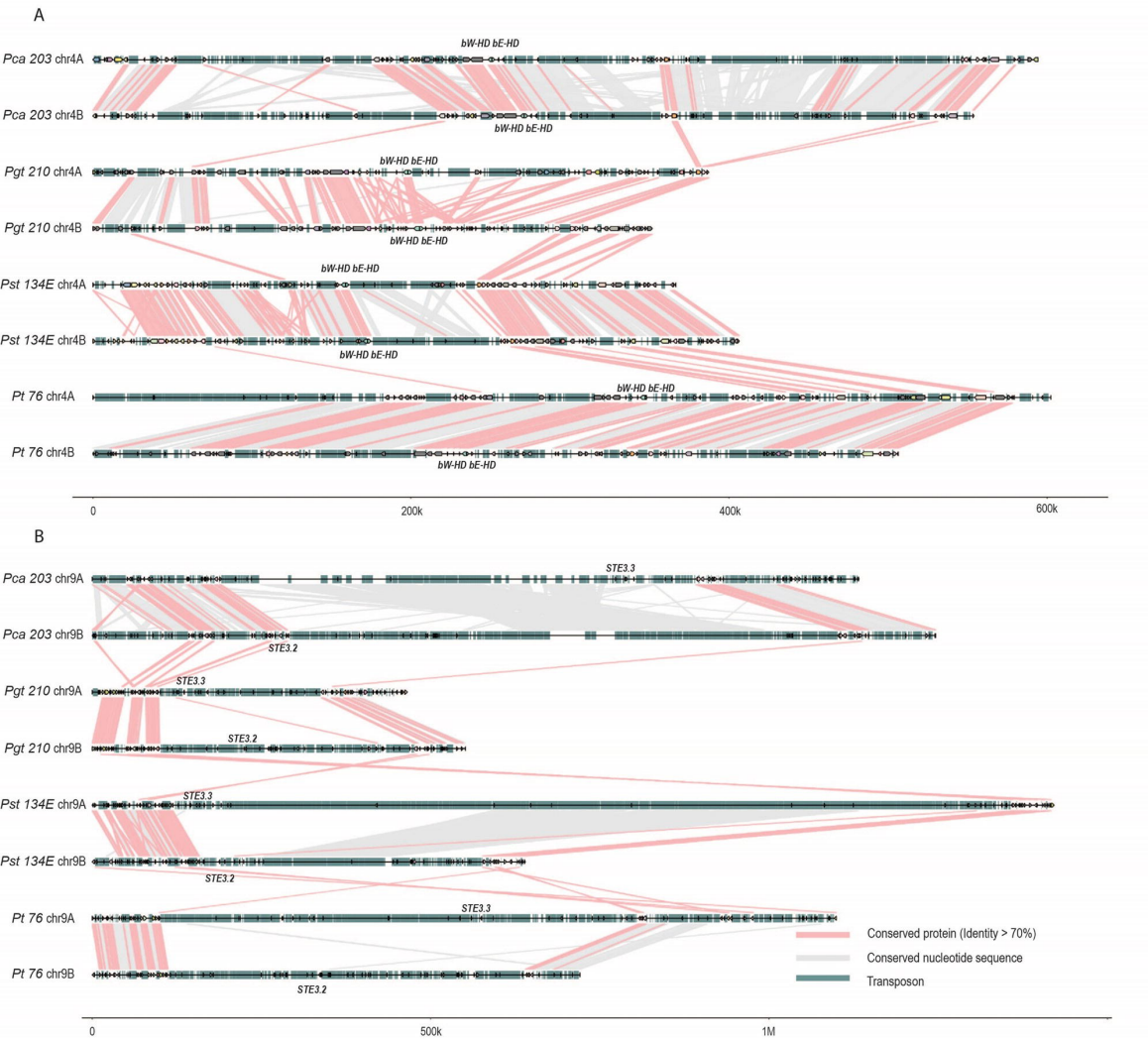
This plot shows nucleotide alignments of *STE3.1* locus from different isolates in *Pca*, *Pst*, *Pgt* and *Pt* constructed by Geneious Primes v2022.0.1, the green blocks represent identical nucleotide sequence, yellow lines represent occurrence of SNPs. Genes are labelled in blue.

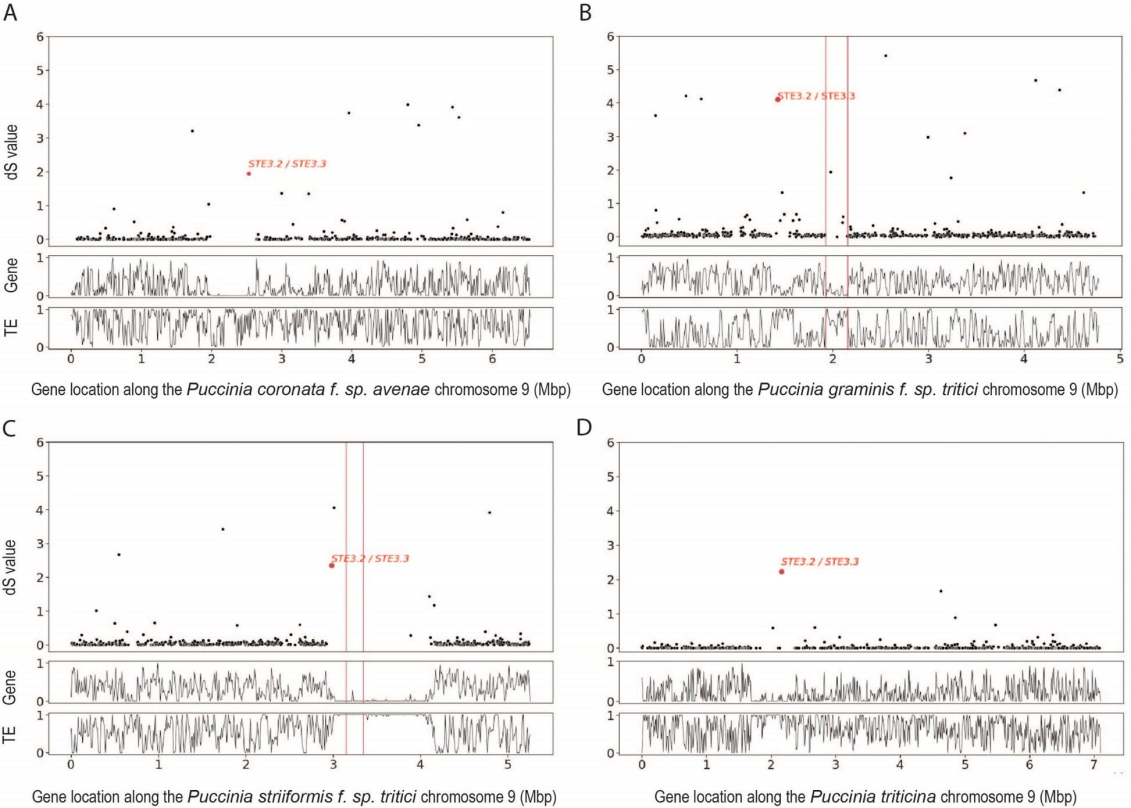
Only a few numbers of SNPs can be identified across isolates. Distance between *STE3.1* and adjacent genes in *Pt 76* are larger than other three specie.

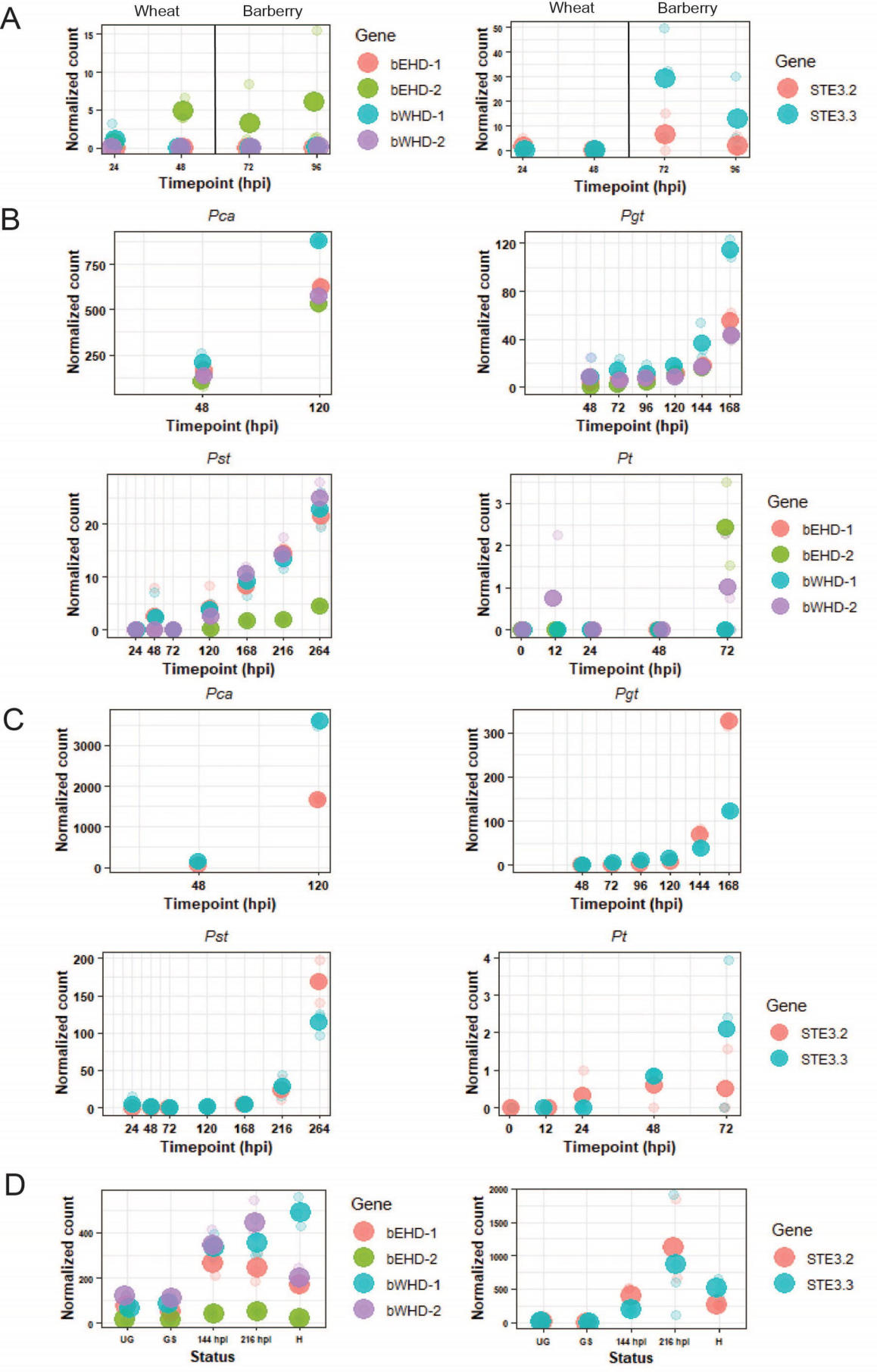

A
Pst 134E (hap1) bW-HD
Pst DK0911 (hap1) bW-HD
Pst 104E (hap1) bW-HD
Pst 104E (hap2) bW-HD
Pst 134E (hap2) bW-HD
Pst DK0911 (hap2) bW-HD
Pca 12NC29 (hap2) bW-HD
Pca 203 (hap1) bW-HD
Pca 203 (hap2) bW-HD
Pca 12SD80 (hap1) bW-HD
Pca 12NC29 (hap1) bW-HD
Pca 12SD80 (hap2) bW-HD
Pgt 210 (hap2) bW-HD
Pgt 210 (hap1) bW-HD
Pgt Ug99 (hap1) bW-HD
Pgt Ug99 (hap2) bW-HD
Pt BBBD (hap1) bW-HD
Pt 76 (hap2) bW-HD
Pt 76 (hap1) bW-HD
Pt BBBD (hap2) bW-HD
Mlp bW-HD
64
96
52
97
71
72
99
100
71
83
98
80
100
93
74
90
100
100
0.1
B
Pst 134E (hap1) bE-HD
Pst DK0911 (hap2) bE-HD
Pst 134E (hap2) bE-HD
Pst 104E (hap2) bE-HD
Pst 104E (hap1) bE-HD
Pst DK0911 (hap1) bE-HD
Pca 203 (hap2) bE-HD
Pca 12NC29 (hap2) bE-HD
Pca 12SD80 (hap1) bE-HD
Pca 12SD80 (hap2) bE-HD
Pca 12NC29 (hap1) bE-HD
Pca 203 (hap1) bE-HD
Pgt 210 (hap1) bE-HD
Pgt Ug99 (hap2) bE-HD
Pgt Ug99 (hap1) bE-HD
Pgt 210 (hap2) bE-HD
Pt 76 (hap2) bE-HD
Pt BBBD (hap2) bE-HD
Pt 76 (hap1) bE-HD
Pt BBBD (hap1) bE-HD
Mlp bE-HD
65
65
95
84
100
87
67
63
57
72
100
100
59
100
99
100
58
100
0.1


A
Pca 203 STE3.2
76
90
Pca 12SD80 STE3.2
Pca 12NC29 STE3.2
Pst 134E STE3.2
42
100
Pst 104E STE3.2
Pst DK0911 STE3.2
47
Pt 76 STE3.2
100
Pt BBBD STE3.2
40
Pgt 210 STE3.2
100
Pgt Ug99 STE3.2
Mlp STE3.2
0.1
B
Pst 134E STE3.3
40
Pst 104E STE3.3
100
Pst DK0911 STE3.3
Pca 203 STE3.3
64
40
Pca 12SD80 STE3.3
Pca 12NC29 STE3.3
100
Pgt 210 STE3.3
100
Pgt Ug99 STE3.3
70
Pt 76 STE3.3
98
Pt BBBD STE3.3
Mlp STE3.4
0.1


A
dS value
bE-HD
bW-HD
Gene
TE
Gene location along the Puccinia coronata f. sp. avenae chromosome 4 (Mbp)
B
bE-HD
bW-HD
Gene location along the Puccinia graminis f. sp. tritici chromosome 4 (Mbp)
C
bW-HD
bE-HD
Gene location along the Puccinia striiformis f. sp. tritici chromosome 4 (Mbp)
D
bE-HD
bW-HD
Gene location along the Puccinia triticina chromosome 4 (Mbp)


A
Pca 203 chr4A
Pca 203 chr4B
Pgt 210 chr4A
Pgt 210 chr4B
Pst 134E chr4A
Pst 134E chr4B
Pt 76 chr4A
Pt 76 chr4B
bW-HD bE-HD
0
200k
400k
600k
B
Pca 203 chr9A
Pca 203 chr9B
Pgt 210 chr9A
Pgt 210 chr9B
Pst 134E chr9A
Pst 134E chr9B
Pt 76 chr9A
Pt 76 chr9B
STE3.3
STE3.2
0
500k
1M
Conserved protein (Identity > 70%)
Conserved nucleotide sequence
Transposon


A
dS value
STE3.2 / STE3.3
Gene
TE
Gene location along the Puccinia coronata f. sp. avenae chromosome 9 (Mbp)
B
STE3.2 / STE3.3
Gene location along the Puccinia graminis f. sp. tritici chromosome 9 (Mbp)
C
dS value
STE3.2 / STE3.3
Gene
TE
Gene location along the Puccinia striiformis f. sp. tritici chromosome 9 (Mbp)
D
STE3.2 / STE3.3
Gene location along the Puccinia triticina chromosome 9 (Mbp)


A
Wheat
Barberry
Normalized count
Timepoint (hpi)
Gene
bEHD-1
bEHD-2
bWHD-1
bWHD-2
STE3.2
STE3.3
B
Pca
Pgt
Pst
Pt
C
D
UG
GS
144 hpi
216 hpi
H
Status