

## RESEARCH ARTICLE

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

Zhenyan Luo<sup>1</sup>, Alistair McTaggart<sup>2</sup>, Benjamin Schwessinger<sup>1\*</sup>

**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

\* [benjamin.schwessinger@anu.edu.au](mailto:benjamin.schwessinger@anu.edu.au)

## Abstract

Permanent heterozygous loci, such as sex- or mating-compatibility regions, often display suppression of recombination and signals of genomic degeneration. In Basidiomycota, two distinct loci confer mating compatibility. These loci encode homeodomain (*HD*) transcription factors and pheromone receptor (*Pra*)-ligand allele pairs. To date, an analysis of genome level mating-type (*MAT*) loci is lacking for obligate biotrophic basidiomycetes in the *Pucciniales*, an order containing serious agricultural plant pathogens. Here, we focus on four species of *Puccinia* that infect oat and wheat, including *P. coronata* f. sp. *avenae*, *P. graminis* f. sp. *tritici*, *P. triticina* and *P. striiformis* f. sp. *tritici*. *MAT* loci are located on two separate chromosomes supporting previous hypotheses of a tetrapolar mating compatibility system in the *Pucciniales*. The *HD* genes are multiallelic in all four species while the *PR* locus appears biallelic, except for *P. graminis* f. sp. *tritici*, which potentially has multiple alleles. *HD* loci are largely conserved in their macrosynteny, both within and between species, without strong signals of recombination suppression. Regions proximal to the *PR* locus, however, displayed signs of recombination suppression and genomic degeneration in the three species with a biallelic *PR* locus. Our observations support a link between recombination suppression, genomic degeneration, and allele diversity of *MAT* loci that is consistent with recent mathematical modelling and simulations. Finally, we confirm that *MAT* genes are expressed during the asexual infection cycle, and we propose that this may support regulating nuclear maintenance and pairing during infection and spore formation. Our study provides insights into the evolution of *MAT* loci of key pathogenic *Puccinia* species. Understanding mating compatibility can help predict possible combinations of nuclear pairs, generated by sexual reproduction or somatic recombination, and the potential evolution of new virulent isolates of these important plant pathogens.

## OPEN ACCESS

**Citation:** Luo Z, McTaggart A, Schwessinger B (2024) Genome biology and evolution of mating-type loci in four cereal rust fungi. *PLoS Genet* 20(3): e1011207. <https://doi.org/10.1371/journal.pgen.1011207>

**Editor:** Tatiana Giraud, CNRS, FRANCE

**Received:** August 28, 2023

**Accepted:** March 4, 2024

**Published:** March 18, 2024

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1011207>

**Copyright:** © 2024 Luo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Analysis code used in this study is available at github repository: <https://github.com/ZhenyanLuo/codes-used-for-mating-type>. Alignments used in genealogical studies, exported RDP5 project files, ds value of gene pairs in studied cereal rust species, presumed CDS of

## Author summary

Sexes in animals and some plants are determined by sex chromosomes. In fungi, mate compatibility is determined by mating-type (*MAT*) loci, which share some features with sex chromosomes including recombination suppression around heterozygous loci. Here,

reconstructed HD alleles, *Pra* alleles, TE annotation and classification files, normalized gene expression matrix files are available at Dryad (Luo Z, Schwessinger B. [Supporting information of genome biology and evolution of mating type loci in four cereal rust fungi \[Dataset\] 2024](#). Available from: <https://doi.org/10.5061/dryad.w0vt4b8zm>).

**Funding:** This work was supported by an Australian Research Council Future Fellowship FT180100024 to B. S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

we study the MAT loci in fungal pathogens from the order *Pucciniales*, which cause rust diseases on many economically important plants, including wheat and oats. We show that one of the MAT loci is multiallelic, while the other is biallelic in most cases. The biallelic locus shows strong signs of recombination suppression and genetic deterioration with an increase in number of transposable elements and gene deserts surrounding the locus. Our findings on the genome biology of MAT loci in four economically important pathogens will improve predictions on potential novel virulent isolates that can lead to large scale pandemics in agriculture.

## Introduction

The evolutionary origin of sex and mating-type determining chromosomes or loci is a fundamental question in biology. It is widely accepted that X and Y chromosomes were originally homologous, but recombination suppression caused gradual genetic degeneration of the Y chromosome [1]. Footprints of genetic degeneration include increased rates of (non-)synonymous substitutions ( $d_N$ ,  $d_S$ ), accumulation of transposable elements (TEs), accumulation of inversions, reduced gene expression and/or reduced gene numbers, all of which are consequences of recombination cessation [2–5].

In contrast to many animals and plants, antagonistic selection is unlikely to be the evolutionary mechanism that causes recombination cessation in fungal mating-type (MAT) loci [6–8]. Instead, mathematical modelling and stochastic simulations suggest that non-recombining DNA fragments, caused for example by inversions, can be fixed solely due to the presence of deleterious mutations in genomes [2]. Non-recombining fragments are beneficial if they carry fewer deleterious mutations than average and can increase in frequency. However, when becoming frequent, their recessive deleterious mutations will be exposed, selected against and thereby prevent fixation. Yet, they can be fix at permanent heterozygous loci due to their sheltering effect for deleterious mutations [2].

Initial fixation of recombination suppressors, such as inversions, can lead to progressive extension of non-recombining DNA fragments by an accumulation of additional inversions over longer evolutionary timeframes [2]. Importantly, this model predicts that non-recombining DNA fragments around MAT loci are larger at biallelic loci and in species with smaller effective population sizes, short haploid phases, outcrossing mating systems, high mutation rates, and extended dikaryotic life stages [2]. Indeed, several fungal species carry extensive non-recombining DNA fragments around MAT loci. Recombination suppression can range from hundreds of kbp (kilo base pairs) to several Mbp (mega base pairs) and lead to genetic degeneration [5,9], such as in *Neurospora tetrasperma* [10], *Podospora anserina* [11], *Schizothecium tetrasporum* [12], *Agaricus bisporus* [13], *Ustilago hordei* [14], several species of *Microbotryum* [8,15] and *Cryptococcus* spp. [16,17].

Basidiomycota have evolved unique mating systems to govern nuclear compatibility, mate selection, and life cycles [9]. From a genetic perspective, two MAT loci determine mating-type identity and non-self-recognition in most Basidiomycota. The pheromone receptor (PR) locus contains a pheromone receptor gene (*Pra*) and at least one pheromone peptide precursor gene (*mfa*) [18]. The PRA protein is a transmembrane localized G protein-coupled receptor, which recognizes the processed and post-translationally modified mature pheromone peptide MFA encoded by compatible PR allele [9]. PR locus defines pre-mating compatibility and gamete fusion [3]. Downstream of initial gamete fusion, the homeodomain (HD) locus determines success in post-mating development [9]. The HD locus contains two tightly linked

homeodomain transcription factor genes (*bW-HD1* and *bE-HD2* in *Pucciniales* [19, 20], originally designated as *bE-HD1* and *bW-HD2* in *Ustilago maydis* [21]), which are linked by a short DNA fragment (~1 kbp) and are outwardly transcribed in opposite directions. Their protein products must be of different allelic specificity to form heterocomplexes and to activate a transcriptional cascade downstream of PRA-MFA [9]. The HD transcription factors regulate cellular development during mating, maintenance of the dikaryotic state, and control pathogenicity in some plant pathogens such as smuts [15].

The HD and PR loci can be either physically linked or unlinked, which means that they segregate together or independently, respectively. The mating compatibility system in Basidiomycota influences genomic organization and resulting segregation patterns of MAT loci. Most basidiomycete species display either bipolar or tetrapolar mating compatibility systems and sexual gametes (haploids) must have different alleles at MAT loci to be compatible [9]. In bipolar species the HD and PR loci are physically linked or alternatively, one locus has lost its function in mating compatibility. In bipolar species, only one MAT locus determines mate compatibility. In tetrapolar species, the HD and PR loci are unlinked, and both determine mate compatibility. A bipolar mating compatibility system is favorable in inbreeding populations that are primarily selfing, as the likelihood of compatibility amongst gametes derived from the same diploid individual is 50 percent [15]. In comparison, tetrapolar mating has 25 percent compatibility of gametes under selfing conditions. Outcrossing mating behavior leads to multiallelism at MAT loci because syngamy of haploid gametes derived from different individuals increases compatibility odds [6,9,15]. In many tetrapolar basidiomycete species, the HD locus is highly polymorphic within the population, with tens to hundreds of known or estimated alleles that are under negative frequency-dependent selection [15,22,23]. A tetrapolar mating compatibility system is thought to be the ancestral state in Basidiomycota, yet several species have evolved bipolarity independently. For example, *Microbotryum* spp. are highly selfing, and multiple independent evolutionary events have linked the HD and the PR locus into DNA fragments of different sizes and ages within this genus [3,7,8,24,25]. These independent events at different timescales show clear signatures of recombination suppression and genetic degeneration. Similarly, in the Ustilaginomycotina, several species evolved bipolarity by linking both MAT loci, such as *U. hordei*, *U. bromivora* or *Sporisorium scitamineum* [9]. In contrast, a recent study showed that tetrapolarity has evolved multiple times in the human skin fungus, *Malassezia* spp., where the ancestral behavior is pseudobipolar [26].

Detailed analyses of MAT loci and their genome biology is lacking for rust fungi (order *Pucciniales*, division Basidiomycota) [9]. This lack of knowledge on mating-type is despite *Pucciniales* being the largest order of fungal plant pathogens, causing diseases with significant environmental and economic impact for trees and crops such as poplar, paperbarks, wheat, oat, soybean, and coffee [27–29]. Studying mating systems in *Pucciniales* has been difficult as they are obligate biotrophs with complex life cycles. They cannot be cultured in vitro and many species require multiple hosts to complete their sexual life cycle [9]. For example, macrocyclic and heteroecious rust fungi have five spore stages and require two hosts to complete their sexual life cycles. During the asexual infection cycle on the “primary” host, for example wheat and oat for rust on cereal crops, fungi are dikaryotic and produce re-infective urediniospores that harbor two distinct nuclei carrying compatible MAT loci. The sexual cycle is initiated with the production of teliospores, which represent the diploid phase of rust fungi and the site of meiosis. Resulting haploid basidiospores infect the “alternate” host, for example *Berberis* spp. for rust fungi of cereal crops, and form haploid infection structures with distinct MAT loci. Fusion of cells with compatible MAT loci generates dikaryotic intercellular mycelia within the plant tissue that gives rise to dikaryotic aeciospores that infect the “primary” host [7] to complete the full life cycle [9, 30]. However, it is important to note that recent cytogenomic

and cytogenetic work questions this clear delineation of nuclear state and ploidy in *Pucciniales* at different life stages, with evidence for the occurrence of diploid nuclei throughout the life cycle in many species [31].

The comprehensive role of *MAT* genes during the life cycle of rust fungi is not yet understood. It is hypothesized that *MAT* genes regulate nuclear pairing during dikaryotic spore production and mediate compatibility of haploid cell fusions [9,30]. Original research suggested that most species of rust fungi are bipolar [30], yet more recent studies on *Melampsora lini* and *P. coronata* var. *coronata* revealed a tetrapolar mating compatibility system [32,33]. In the absence of direct experimental evidence such as experimental mating or gene knockout studies, genomic insights into *MAT* loci from genome assemblies of dikaryons can provide hypotheses on the mating compatibility system. Independent initial genome analyses of several rust fungi revealed two unlinked *MAT* loci with PR and HD loci located on different contigs, scaffolds or chromosomes including *Austropuccinia psidii* [34], *P. graminis* f. sp. *tritici* [35] and *P. tritici* [19,20]. This genome organization supports hypotheses of a tetrapolar mating compatibility system in these rust fungi. These rust fungi encode *Pra* receptor genes belonging to the *STE3* gene family. The likely *bona fide* receptors are encoded by *STE3.2-2* and *STE3.2-3* that display clear signatures of ancient trans-specific polymorphisms, which are proposed to be ancestral to Basidiomycota [36]. Here the sequences of a given mating-type are more similar to pheromone receptor alleles of the same mating-type in distantly related fungi than the alternate allele in the same species. In most cases, at least one *mfa* allele has been found in proximity to the *STE3.2-2/3* locus. To date, the PR locus is supported as biallelic in available population level datasets for *P. tritici* and *P. striiformis* f. sp. *tritici* [19,23,37]. *Melampsora larici-populina* may be multiallelic at the PR locus [9] and population level analyses are missing for other rust fungi. The HD locus is multiallelic in *P. triticina* and *P. striiformis* f. sp. *tritici* [19,23,37] encoding variants of *bW-HD1* and *bE-HD2* that are highly dissimilar only in the variable N-terminal domain, shown to be essential for functional heterodimerization in other Basidiomycota [38,39].

It is noteworthy that direct biochemical or genetic evidence is lacking to show that any of these *MAT* genes govern mating compatibility in rust fungi. Yet, complementation assays in *U. maydis* and host induced gene silencing on wheat demonstrated the importance of *Pra* and *HD* genes of *P. triticina* for mating in a heterologous system and for spore production during asexual infection [19]. In addition, *MAT* genes were expressed during the asexual and/or sexual infection cycle in several rust fungal species [19,40]. These initial studies brought insights and left several outstanding knowledge gaps for *MAT* loci in rust fungi: 1) What is the chromosomal organization of PR and HD loci and how does it compare between closely related species? 2) What is the allelic diversity of *MAT* genes within rust fungal species? 3) What is the composition and organization of *MAT* loci proximal regions, how does it compare between PR and HD loci, and how does it vary between closely related species? 4) What is the compositional and organizational variation of *MAT* loci within rust fungal species? 5) Is the expression of *MAT* genes during the asexual infection cycle common in rust fungi?

Addressing these questions has been challenging because initial genome assemblies of rust fungi were fragmented and the structure of *MAT* loci could not be assessed consistently. For example, clustering of repetitive elements around *MAT* loci [18,41,42] is a challenge for studying rust fungi with repeat-rich genomes because repetitive sequences cause gaps, fragmentation, and phase-errors in assembled genomes using short-read or noisy long-read sequencing technologies. Here, we overcome this challenge and address important knowledge gaps around *MAT* loci in four cereal rust fungi by providing a detailed comparative analysis of publicly available chromosome-scale phased genome assemblies and complementary Illumina short-read datasets. The study species include the oat crown rust fungus, *P. coronata* f. sp. *avenae*

[43–45], and three wheat rust fungi, *P. graminis* f. sp. *tritici* [35,46,47], *P. triticina* [48–51] and *P. striiformis* f. sp. *tritici* [52–58], which combined are the biggest threat to wheat production globally causing losses of several billion dollars every year [29]. Our detailed comparative analyses set out to address the following specific objectives. 1) To determine the chromosomal organization of MAT loci. 2) To assess the allelic diversity of these loci in the four cereal rust fungi. 3) To evaluate synteny, recombination suppression, and genomic degeneration around PR and HD loci with the prediction that multiallelic loci are more syntenic than biallelic loci and that the latter show stronger footprints of recombination suppression and genomic degeneration. 4) To gauge the conservation and/or plasticity of MAT loci within a cereal rust fungus species using *P. triticina* as model system. And 5) to determine the expression of *Pra* and *HD* genes during the asexual infection cycle of the cereal hosts.

We show that *HD* genes are multiallelic in four species of cereal rust fungi and *Pra* genes are most likely biallelic in three out of four species. Only biallelic MAT loci show strong signs of recombination suppression and genomic degeneration which supports recent mathematical modelling. Our results provide novel insights into the genome biology of MAT loci in cereal rust fungi and will benefit predictions about mating and hybridization events that may be linked to the evolution of novel pathogenicity traits.

## Results

### Genomic organisation and inheritance of mating-type genes suggest cereal rust fungi have a tetrapolar mating compatibility system

We set out to test previous observations that suggest cereal rust fungi have a tetrapolar mating compatibility system with unlinked PR and HD loci [19,20]. We made use of seven available chromosome-level genome assemblies of four cereal rust fungi, namely *P. coronata* f. sp. *avenae* (*Pca*) [43–45], *P. graminis* f. sp. *tritici* (*Pgt*) [35,46,47,58], *P. triticina* (*Pt*) [37,48–51,59] and *P. striiformis* f. sp. *tritici* (*Pst*) [52,54–57,60], in addition to five partially phased assemblies (Table 1). Our initial analysis focused on one reference genome per species with *Pca* 203, *Pgt* 21–0, *Pt* 76, and *Pst* 134E.

We used previously characterized *MAT* genes from the *Pt* isolate *BBBD* (Table 2) [19] as queries to identify orthologs in available genomes and proteomes (Table 1). We identified *HD* (*bW-HD1* and *bE-HD2*) and *Pra* (*STE3.2–2* and *STE3.2–3*) alleles on chromosome 4 and chromosome 9 respectively, whereas *STE3.2–1* alleles were located on chromosome 1 (S1 Fig). This is consistent with previous preliminary analyses for *Pgt* and *Pt* that suggested HD and PR loci are located on two distinct chromosomes [35,37]. We confirmed these results using five genome resources for cereal rust fungi that are only partially phased and non-chromosome scale (Table 1). We identified two alleles of *bW-HD1*, *bE-HD2* and *Pra* (*STE3.2–2* or *STE3.2–3*) in all five genome assemblies with the exception for *STE3.2–2* in *Pst* 104E and *Pca* 12SD80 [43,53]. In the case of *Pst* 104E, *STE3.2–2* was absent from the genome assembly but we confirmed its presence in the raw sequencing data by mapping sequencing reads against the *Pst* 134E phased chromosome scale genome assembly [52]. *STE3.2–2* displayed average genome-wide read coverage without any nucleotide variation confirming that *Pst* 104E was also heterozygous for *Pra*. In the case of *Pca* 12SD80, we identified two copies of *STE3.2–2* on contig 000183F and 000183F\_004. This mis-characterization of *STE3.2–2* in *Pst* 104E and *Pca* 12SD80 is likely caused by early long-read genome assembly errors using noisy PacBio long-reads and suboptimal genome assembly algorithms that struggle with highly repetitive regions. Similar observations have been made for missing *Pra* genes reported for earlier genome versions of *Melampsora larici-populina* [61].

Table 1. Cereal Rust Fungi Genomes used in the present study.

Species	Strain	Genome size	BUSCO (%)	Source	Download date	Citation
<i>Puccinia graminis</i> f. sp. <i>tritici</i>	Ug99	176.2 Mb	93.4 [S:7.7, D:85.7]	NCBI	26/05/2021	[35]
	21-0*	176.9 Mb	94.2 [S:8.3, D:85.9]	NCBI	26/05/2021	
<i>Puccinia striiformis</i> f. sp. <i>tritici</i>	DK0911	157.8 Mb	91.1 [S:38.4, D:52.7]	JGI	1/06/2021	[54]
	104E	126.5 Mb	92.1 [S:7.9, D:84.2]	JGI	26/05/2021	[53]
	134E*	167.7 Mb	92.6 [S:6.8, D:85.8]	Provided by the authors	21/06/2021	[52]
<i>Puccinia coronata</i> f. sp. <i>avenae</i>	12NC29	105.2 Mb	89.9 [S:18.1, D:71.8]	NCBI	26/05/2021	[43]
	12SD80	99.2 Mb	90.2 [S:33.1, D:57.1]	NCBI	26/05/2021	[43]
	203*	208.1 Mb	92.1 [S:6.9, D:85.2]	CSIRO data access portal	20/06/2022	[44]
<i>Puccinia triticina</i>	76*	253.5 Mb	91.0 [S:1.7, D:89.3]	Provided by the authors	28/05/2021	[48]
	15*	243.9 Mb	91.1 [S:1.6, D:89.5]	NCBI	02/11/2023	[59]
	19NSW04*	253 Mb	91.2[S:0.9, D:90.3]	NCBI	10/10/2023	[37]
	20QLD87*	248.1 Mb	91.3[S:1.3, D:90.0]	NCBI	10/10/2023	[37]
<i>Puccinia polysora</i> f. sp. <i>zea</i>	GD1913*	1.7 Gb	91.1 [S:1.4, D:89.7]	NCBI	07/14/2023	[81]

The table provides information on all genome assemblies used in this study, their total dikaryotic genome assembly size (“Genome size”), and completeness as assessed with BUSCO (“BUSCO (%)”). For the “BUSCO (%)” column the first number presents the percentage of complete single BUSCO genes identified. The numbers in the brackets provide the percentage of single copy and of duplicated BUSCO genes in the dikaryotic genome assembly. Additional presented metadata includes species name (“Species”), strain name (“Strain”), source (“Source”), download date in DD/MM/YYYY (“Download date”) and the initial reference (“Citation”). The \* genome denotes dikaryotic genome assemblies which are chromosome scale phased genome. NCBI—National Center for Biotechnology information, JGI—Joint Genome Institute, CSIRO—Commonwealth Scientific and Industrial Research Organization.

<https://doi.org/10.1371/journal.pgen.1011207.t001>

Overall, the two MAT loci, HD and PR, are unlinked and heterozygous in all twelve dikaryotic genome assemblies, which is consistent with a genome informed hypothesis that cereal rust fungi are tetrapolar.

### ***Mfa* pheromone peptide precursor genes are closely linked to *STE3.2–2* but not *STE3.2–3***

We searched the available dikaryotic genomes for putative pheromone peptide precursor genes, which encode mating-factor-a (MFA) peptides. *Mfas* are often linked to *Pra* alleles and are predicted to bind to the compatible pheromone receptors encoded at the allelic PR locus. We used the previously identified *mfa1*, *mfa2*, and *mfa3* genes of *Pst*, *Pgt*, and *Pt* as queries [19]. We identified a single *mfa2* pheromone peptide precursor gene in all species. In all cases, *mfa2* and *STE3.2–2* were closely linked, located within 500–1100 bp from each other, and encoded on the same DNA strand (S2 Fig). The *mfa2* derived precursor peptides were 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 [62] (S3B Fig). The MFA2 amino acid sequences were 100 percent identical at species rank (S3B Fig).

In contrast, the sequence, length, number, and location of *mfa* genes associated with *STE3.2–3* varied between species (S2, S3A, and S3C Figs). *Pca* and *Pst* encoded a single *mfa1* allele. *Pca-mfa1* of *Pca* 203 encoded a 76 amino acid peptide and was located 0.54 Mbp upstream of *STE3.2–3*. In *Pst*, *Pst-mfa1* encoded a 74 amino acid peptide and was located 10 kbp and 13 kbp away from *STE3.2–3* in *Pst* 134E and *Pst* DK0911, respectively. In contrast, *Pt* carried two identical copies of *Pt-mfa1/3* encoding 61 amino acid peptides. In the *Pt* 76 reference, *Pt-mfa1/3s* were associated with *STE3.2–3* at a distance of 0.24 mbp and 0.27 mbp, respectively (S2 Fig). In addition, *Pgt* encoded for two *mfa* genes in close proximity to *STE3.2–3* that were located upstream and downstream at a distance of 30 kbp and 96 kbp, respectively

Table 2. MAT reference genes used as query and outgroups.

Species	Strain	Source	Gene name	Gene ID	Citation
<i>Puccinia graminis</i> f. sp. <i>tritici</i>	isolate CDL 75–36–700–3 (race SCCL)	NCBI & Cuomo et al., 2017	<i>bW-HD1</i>	PGTG_05143	[19]
			<i>bE-HD2</i>	PGTG_05144	
			<i>STE3.2-1</i>	PGTG_00333	
			<i>STE3.2-2</i>	PGTG_19559	
			<i>STE3.2-3</i>	PGTG_01392	
			<i>Pgtmfa1</i>	-	
			<i>Pgtmfa2</i>	-	
<i>Puccinia triticina</i>	isolate 1-1 / race 1 (BBBD)	NCBI & Cuomo et al., 2017	<i>bW-HD1</i>	PTTG_09683 PTTG_27730	[19]
			<i>bE-HD2</i>	PTTG_10928 PTTG_03697	
			<i>STE3.2-1</i>	PTTG_09751	
			<i>STE3.2-2</i>	PTTG_28830	
			<i>STE3.2-3</i>	PTTG_09693	
			<i>Ptmfa1/3</i>	-	
			<i>Ptmfa2</i>	-	
<i>Puccinia striiformis</i> f. sp. <i>tritici</i>	78	NCBI & Cuomo et al., 2017	<i>bW-HD1</i>	PSTG_05919 PSTG_18670	[19]
			<i>bE-HD2</i>	PSTG_05918 PSTG_19315	
			<i>STE3.2-1</i>	PSTG_02613	
			<i>STE3.2-2</i>	PSTG_15127	
			<i>STE3.2-3</i>	PSTG_15070	
			<i>Pstmfa1</i>	-	
			<i>Pstmfa2</i>	-	
<i>Puccinia polysora</i> f. sp. <i>zea</i>	GD1913	NCBI	<i>bW-HD1</i>	FUNB_006397	[81]
			<i>bE-HD2</i>	FUNA_006226	
			<i>STE3.2-2</i>	FUNA_023003	
			<i>STE3.2-1</i>	FUNA_001167	
			<i>STE3.2-3</i>	FUNA_013409	

The table provides gene names (“Gene name”) and gene identifier (“Gene ID”) of MAT genes from *Puccinia triticina* initially used as query sequences and from *Puccinia polysora* f. sp. *zea* used as outgroups. Additional presented metadata includes species name (“Species”), strain name (“Strain”), source (“Source”), and the initial reference (“Citation”). NCBI–National Center for Biotechnology information.

<https://doi.org/10.1371/journal.pgen.1011207.t002>

(S2 Fig). Yet in contrast to *Pt*, *Pgt mfa1* and *mfa3* encoded for highly distinct pheromone peptide precursor that were only 23.10% identical at the amino acid level and varied in length (68 vs 57 amino acids, respectively).

Further analysis of all three MFA precursor peptide alleles suggests a distinct maturation pathway for each. MFA1 precursor peptides appear to contain three tandem pheromone peptide repeats of similar sequence (S3A Fig), which has been reported for *Microbotryum* spp. [22,63] and Ascomycete fungi [64] previously. The predicted mature peptide encoded by the first repeat is sequence conserved between all four cereal rust species while the second repeat are species-specific, and the third repeat is sequenced conserved between *Pgt*, *Pst* and *Pca*, and has a single amino acid variation in *Pt* (S3A Fig). In contrast, MFA2 precursor peptides lack detectable pheromone peptide repeats and are likely processed into a single mature MFA peptide which is sequence-specific in each species (S3B Fig). Similarly, MFA3 of *Pgt* did not contain amino acid repeat sequences, however we identified the “QWNGNGSHYC” amino acid sequence at the C-terminus of the *Pgt*-MFA3 precursor peptide (S3C Fig). This predicted mature peptide sequence of *Pgt*-MFA3 is highly similar to the predicted mature pheromone

peptides of Pgt-MFA1 “QWGNGSHMC” with a single amino acid variation. We cannot exclude the possibility that MFA3 is processed into additional mature peptides and future studies are needed to define if the observed amino acid variations lead to different receptor specificities.

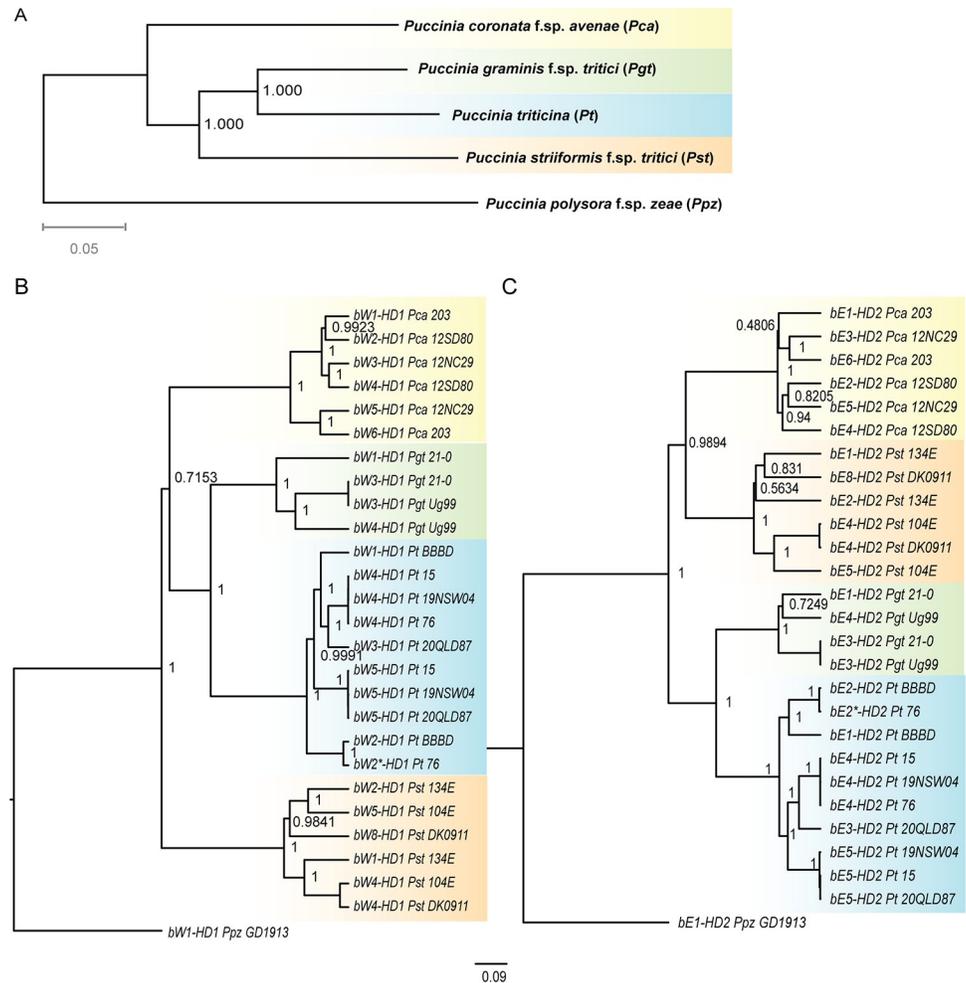
Taken together, in all four cereal rust fungi, *Pra* alleles were associated with species-specific *mfa* genes. Yet, the number, distance, organization, sequence of *mfa* genes and their predicted mature pheromone peptides varied between species.

## Genealogies elucidate distinct evolutionary histories for the different *MAT* genes

Having identified the *MAT* genes, we investigated the genealogical relationships of each individual gene and compared them within and between species. This tested for shared or distinct evolutionary histories, which can provide indications for recombination within and between genes. For example, trans-specific polymorphisms indicate that recombination cessation is older than speciation [36,65]. In the absence of a robust phylogeny including all four study species [66], we first generated a multilocus species tree based on a multiple sequence alignment of 2,284 single ortholog protein sequences [67] including *P. polysora* f. sp. *zeae* as an outgroup (Fig 1A). The species tree showed that the three wheat rust fungi shared a most recent common ancestor when compared to the oat rust fungus *Pca* (Fig 1A). In the wheat rust clade, *Pgt* was sister to *Pt*, and *Pst* shared a common ancestor with them both. We aligned *HD* genes *bW-HD1* and *bE-HD2* separately to construct independent gene trees using *P. polysora* f. sp. *zeae* orthologs as outgroups (Table 2 and Fig 1B and 1C). The gene trees of *bW-HD1* and *bE-HD2* grouped alleles from each species into species-specific clades (Fig 1B and 1C) and we therefore did not find any evidence for trans-specific polymorphisms for either *HD* gene.

We identified multiple *bW-HD1* and *bE-HD2* alleles for the four cereal rust fungi suggesting that the *HD* locus is multiallelic (see also below). We observed several shared *HD* alleles between isolates in the case of *Pt* while *HD* alleles in *Pca* and *Pst* appeared to be more diverse. In the case of *Pgt*, we were limited to only four alleles, making any meaningful conclusion difficult. We next tested our null hypothesis that *bW-HD1* and *bE-HD2* have similar evolutionary histories because they are closely linked by a short DNA fragment. We performed approximately unbiased (AU) tests [68] within each species to investigate if the tree topologies of *bW-HD1* and *bE-HD2* are congruent with each other. In *Pca*, *Pst*, and *Pt* the  $p_{AU}$ -value was less than 0.05, which rejected the null hypothesis of congruent tree topologies. This suggests distinct evolutionary histories for the *bW-HD1* and *bE-HD2* genes in these species, which might be caused by recombination within the *HD* locus. In contrast, the AU-test for *Pgt* suggested similar tree topologies for *bW-HD1* and *bE-HD2*, which likely is influenced by the low sample number of four alleles. We also applied RDP5 [69] to detect potential signals of recombination at the nucleotide level. We detected potential signals for recombination in the *HD* locus for all species, which further supports the results of the AU test [70] and is similar to recombination events between *b* alleles reported in *U. marydis* [71].

Next, we explored the evolutionary relationship between *Pra* alleles at the PR locus by building a single gene tree using *P. polysora* f.sp. *zeae* *STE3.2-2* and *STE3.2-3* as outgroups. The *Pra* gene tree formed two obvious clades based on allele identity grouping the *STE3.2-2* alleles of all species into one clade and all *STE3.2-3* alleles into another clade including respective alleles of *P. polysora* f.sp. *zeae* (Fig 2). In each clade, there was little to no intra-species variation for *STE3.2-2* and *STE3.2-3* based on the minimal branch lengths that separated allele copies in each species sub-clade. This indicates that *Pra* might be biallelic in cereal rust fungi (see also below). The clear grouping by allele identity rather than by species of rust fungi

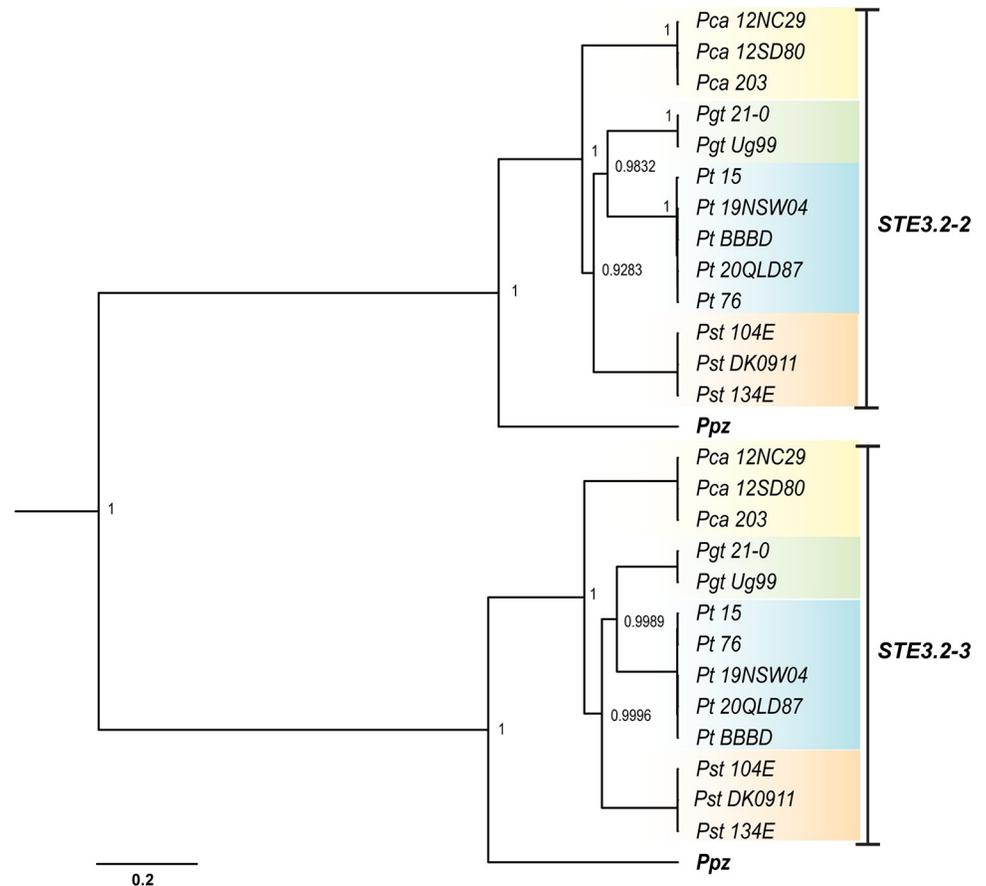


**Fig 1. HD genealogies suggest distinct evolutionary histories for *bW-HD1* and *bE-HD2* in four cereal rust fungi.** (A) Species tree of *Puccinia polysora* f. sp. *zeae* (*Ppz*), *P. coronata* f. sp. *avenae* (*Pca*), *P. graminis* f. sp. *tritici* (*Pgt*), *P. triticina* (*Pt*) and *P. striiformis* f. sp. *tritici* (*Pst*) inferred from 2284 single-copy orthogroups. Numbers on nodes represent local support value computed by Fasttree. *Puccinia polysora* f. sp. *zeae* (*Ppz*) has been used for rooting the species tree. (B) and (C) Bayesian rooted gene tree built from *bW-HD1* or *bE-HD2* coding-based sequence alignment, respectively. Trees are based on a HKY + G model of molecular evolution. In either case, alleles from the same species are grouped into the same clade. Each node is labelled with its values of posterior probability (PP). PP values above 0.95 are considered to have strong evidence for monophyly of a clade and PP values of identical alleles are not displayed. The scale bar represents the number of nucleotide substitutions per site. \* marks an allele with minor variations only outside the variable domain, which means this allele is predicted to be functionally equivalent to its closest neighbor. Alleles of the same species are colored with identical background: *Pca* (yellow), *Pgt* (green), *Pt* (blue), *Pst* (orange).

<https://doi.org/10.1371/journal.pgen.1011207.g001>

indicates strong trans-polymorphisms and a long-term suppression of recombination at the PR locus that predates ancestral speciation.

We next investigated if the putative pheromone peptide precursor *mfa* genes at the PR locus followed similar evolutionary histories compared to their co-located *Pra* genes. Like the *Pra* genealogy, the gene tree of *mfa* alleles revealed strong trans-specific polymorphisms because *mfa* copies clustered by allele and not species identity (S4 Fig). For example, all *mfa2* alleles grouped into one clade according to their physical association with the specific *Pra* allele *STE3.2-2*. Similarly, all *mfa1* alleles grouped together with similar topology as their linked *STE3.2-3* copies. The species-specific *Pgt-mfa3s*, which are also physically linked to *STE3.2-3*,



**Fig 2. *Pra* genealogy displays ancient trans-specific polymorphism in cereal rust fungi.** Bayesian gene tree built from *Pra* (*STE3.2-2* and *STE3.2-3*) coding-based sequence alignment. The tree is based on a TN93 + I model of molecular evolution. The *Pra* alleles, *STE3.2-2* and *STE3.2-3*, were grouped into two clades by allele identity and not species identity. Each node is labelled with its values of posterior probability (PP). PP values above 0.95 are considered as strong evidence for monophyly of a clade and PP values of identical alleles are not displayed. The scale bar represents the number of nucleotide substitutions per site. Species-specific background coloring is the same as for Fig 1.

<https://doi.org/10.1371/journal.pgen.1011207.g002>

were placed as a sister group to the *mfa1* clade, being more closely related to *mfa1* than to *mfa2*.

### HD loci display limited signs of genomic deterioration and their synteny is conserved within species

Different effects on recombination have been reported around MAT loci in basidiomycetes [8,17]. Hence, we investigated signals of altered recombination, synteny, and genomic deterioration around the MAT loci in rust fungi. We defined the HD locus as the DNA fragment which includes *bW-HD1*, *bE-HD2*, and the DNA sequence in between [22]. First, we analyzed 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 software [72] to align the haplotypes of chromosome 4 from *Pca*, *Pgt*, *Pt*, and *Pst*, against each other (S5 Fig). Overall macro-synteny of chromosome 4 haplotypes was conserved in all cereal rust fungi with *Pt* 76 the most syntenic (S5C Fig). Macro-synteny in a 40 kb-sized window around the HD locus was mostly conserved in *Pt* 76, *Pca* 203, and *Pst* 134E while this initial analysis suggested that this locus was less syntenic in *Pgt* 21-0 (S5 Fig).

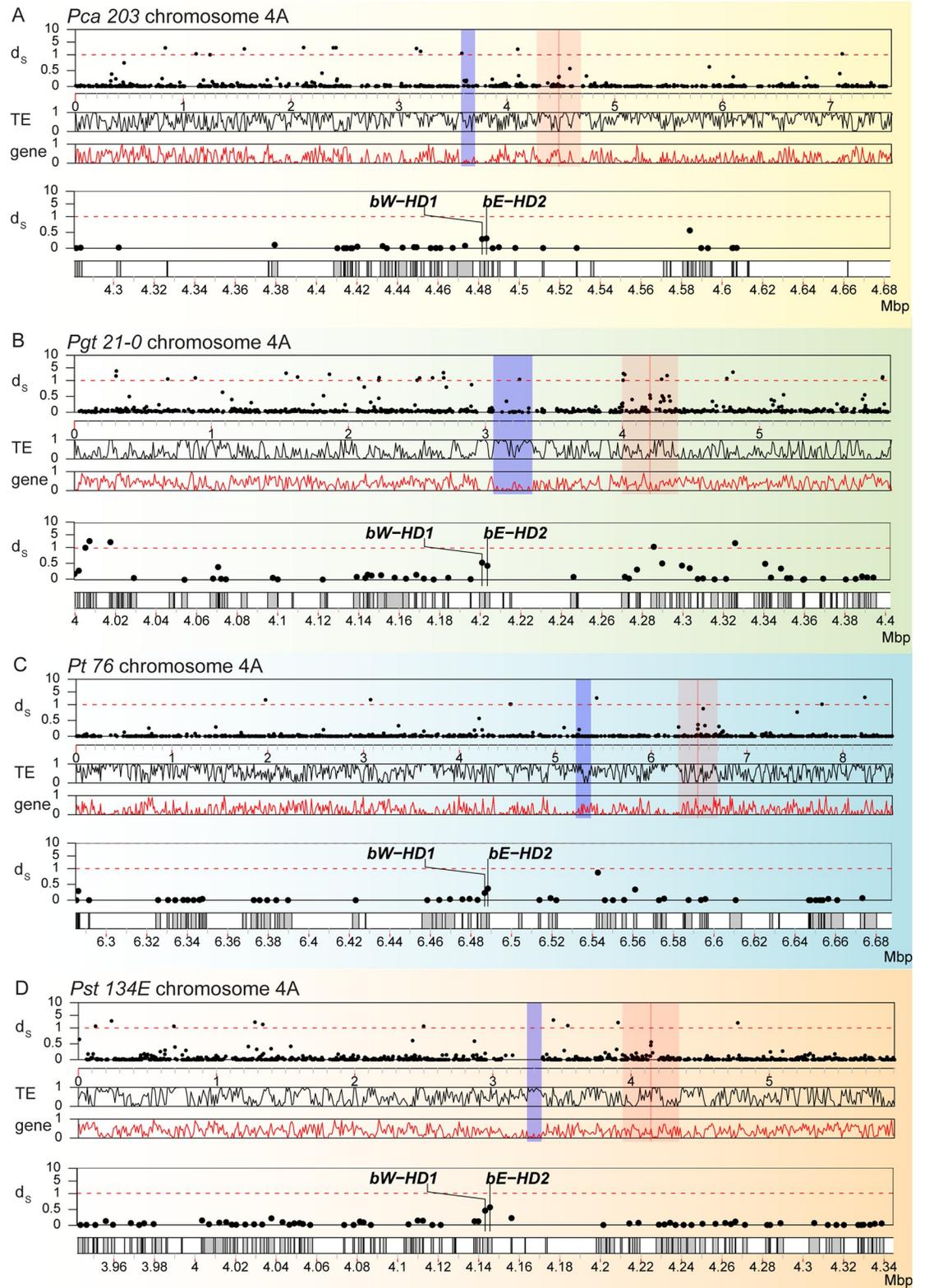
We investigated the individual allelic divergence along chromosome 4 using synonymous divergence ( $d_S$ ) value as an additional measure for footprints of ancient recombination suppression [8]. This analysis tested if the HD locus and proximal genes showed an increase in  $d_S$  values, which could be caused by long-term suppression of recombination and accumulation of independent mutations, as seen in other Basidiomycota [15]. We calculated the pairwise  $d_S$  values for all genes across all sister chromosomes in all four species of *Puccinia*. This whole-genome analysis of all allele pairs on sister chromosomes suggested that chromosome 4 did not have increased  $d_S$  values when compared to all other chromosomes (S6 Fig). The  $d_S$  value of alleles for both HD genes fell within the upper 95% quantile of  $d_S$  values on chromosome 4. Yet,  $d_S$  values of directly adjacent genes were not elevated beyond the background level when we plotted  $d_S$  values for allele pairs along chromosome 4 (Fig 3). Consistently, regions around the HD locus did not show any obviously particular patterns of gene or transposon density, which are common features of genetic deterioration (Fig 3). We tested whether the HD locus was potentially linked to centromeres as reported in other Basidiomycota [15,73]. We used previously identified centromere locations for *Pgt* and *Pst* [52,74] and identified centromere locations for *Pca* and *Pt* using genome-wide Hi-C heatmaps based on the bowtie-like Hi-C interaction features known for fungal centromeres (S7–10 Figs) [75]. This analysis revealed that the HD locus is likely not directly linked to centromeres in any of the four cereal rust fungi.

Lastly, we investigated overall conservation of nucleotide and gene-coding regions for the HD locus including proximal regions containing 40 neighboring genes on either side of the locus. We combined fine scale nucleotide synteny with gene conservation analysis. We specifically tested if cereal rust fungi contain syntenic blocks with conserved protein-coding genes surrounding MAT loci as was found for species e. g. *Trichosporonales* spp. [17]. We used blastn to identify conserved nucleotide sequences within dikaryotic genomes and between species. We plotted genes, transposons, nucleotide and protein-coding gene conservation for proximal regions to the HD locus for all four cereal rust fungi (Fig 4A). Protein-coding genes and their order are mostly conserved within dikaryotic genomes of the same species with 58/80 genes in *Pca* 203, 40/80 in *Pgt* 21–0, 57/80 in *Pt* 76, and 62/80 in *Pst* 134E being conserved. Similarly, we observed considerable nucleotide conservation and synteny within dikaryotic genomes with the exception of *Pgt* 21–0, which is consistent with our initial MUMmer-based analysis (S5 Fig). Overall, synteny and protein-coding gene conservation was very limited between species for regions proximal to the HD locus. We could only 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 (Fig 4A). Our TE analysis revealed that their coverage was overall consistent at the order classification level [76] within species but varied across species (S11 Fig). For example, terminal inverted repeats (TIRs) and other undetermined Class II DNA transposons dominated the HD locus of *Pst* 134E, while the HD locus of other rust fungi had higher coverage of Class I RNA transposons including long terminal repeats (LTRs).

Overall, our analysis of the HD locus suggests that it is conserved within each dikaryotic genome at species rank, while conservation has been eroded between species over longer evolutionary timescales.

### The PR locus displays strong signs of genomic degeneration

We investigated organization and synteny of the PR locus which we defined as the DNA fragments including *Pra*, *mfa* genes, and DNA sequences in between. Initial MUMmer based



**Fig 3. Synonymous divergence ( $d_s$ ) values of HD genes but not their immediate neighbors are slightly elevated on chromosome 4.** Synonymous divergence values ( $d_s$ ) for all allele pairs are plotted along chromosome 4A for (A) *P. coronata* f. sp. *avenae* ("Pca 203"), (B) *P. graminis* f. sp. *tritici* ("Pgt 21-0"), (C) *P. triticina* ("Pt 76"), and (D) *P. striiformis* f. sp. *Tritici* ("Pst 134E").

In each panel, the top track shows the  $d_s$  values (“ $d_s$ ”) of allele pairs along chromosome 4. Each dot corresponds to the  $d_s$  value of a single allele pair. The second and third track shows the averaged TE (“TE”) and gene (“gene”) density along chromosome 4 in 10 kbp-sized windows, respectively. The HD genes (*bW-HD1* and *bE-HD2*) are highlighted with a red line and red shading indicates a 0.4 mbp-sized window around the HD locus. Predicted centromeric regions are marked with blue shading. The two lower tracks ( $d_s$  values and gene locations) provide a detailed zoomed in view of red shaded area around the HD locus. Species-specific background coloring is the same as for Fig 1.

<https://doi.org/10.1371/journal.pgen.1011207.g003>

synteny analysis revealed clear macro- and micro-synteny breaks around the PR locus in all cereal rust fungi (S12 Fig). It also revealed accumulation of highly repetitive sequences closely associated with the PR locus, especially in *Pca 203* and *Pst 134E*. This increase in TEs correlates with a reduction in gene density around the PR locus (Figs 4B and 5). The loss in synteny was specific to the PR locus, as overall macro-synteny of chromosome 9 was conserved in all species (S12 Fig).

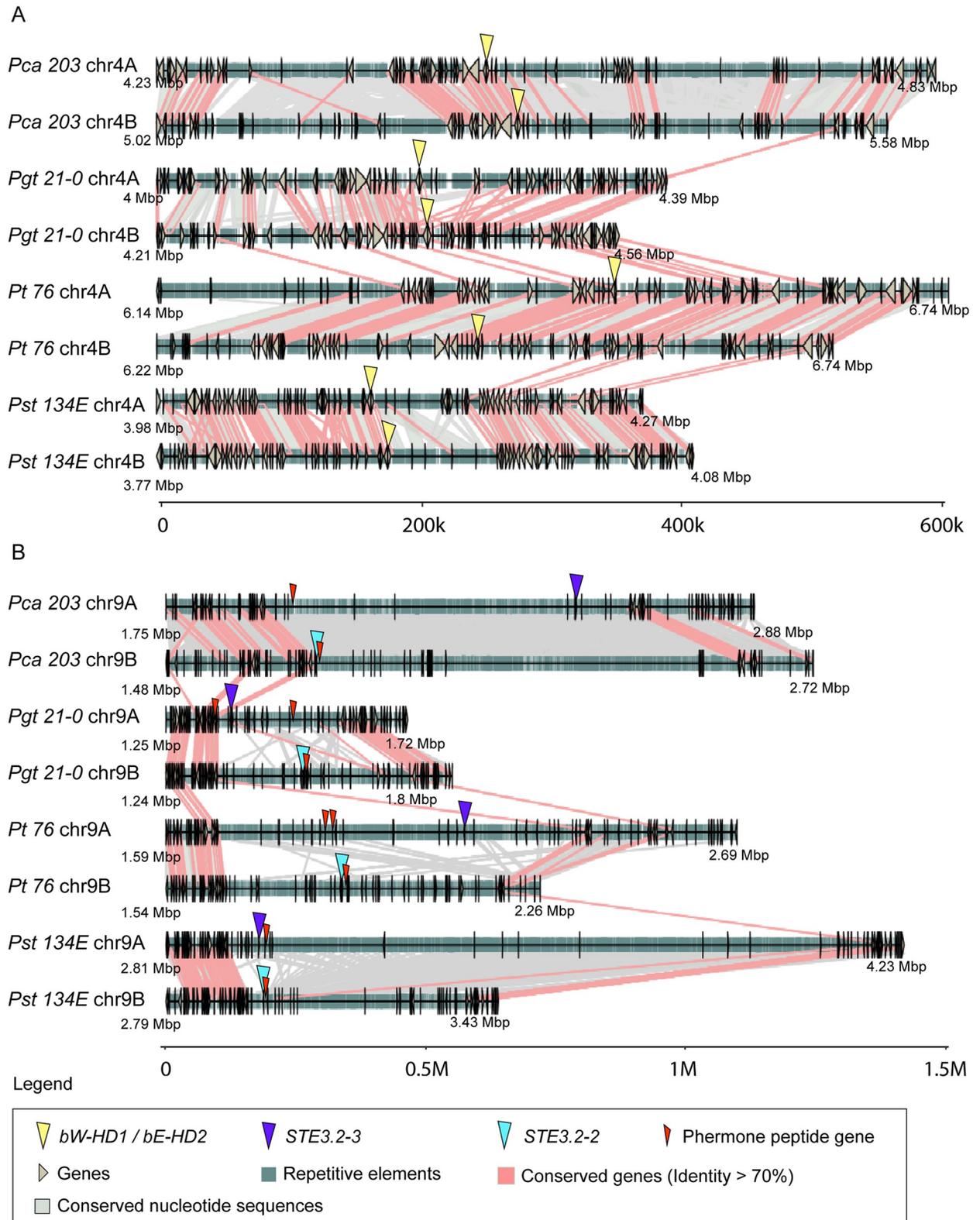
Consistent with the macro-synteny analysis,  $d_s$  values for the *Pra* alleles *STE3.2-2-STE3.2-3* belonged to the 99.9% quantile of  $d_s$  values of all allele pairs found on chromosome 9 in all species (Figs 5 and S6). Gene density dropped significantly around *Pra* genes with *STE3.2-2/STE3.2-3* located within or adjacent to ‘gene deserts’ that extended over 1 mbp in the case of *Pst 134E*. The PR locus and its proximal regions also appear to be enriched in TEs. One exception was observed around the PR locus of *Pgt 21-0* which exhibited a higher gene density and lower amounts of TEs when compared to the other three species, which might be a biological feature or a technical artefact (see Discussion). None of the *Pra* alleles was physically closely linked to centromeres except for *Pst 134E STE3.2-3*, which was found 164 kbp from the centromere, embedded in a > 1 Mb long stretch of TEs depleted of genes (Fig 5).

Detailed analyses of nucleotide synteny and protein-coding gene conservation confirmed signals of extended recombination suppression, loss of synteny and accumulation of transposable elements around the PR locus (Fig 4B). We observed little conservation of nucleotide sequences and protein-coding genes in regions proximal to *Pra* within dikaryotic genomes or between species. Indeed, we could not identify any conserved protein-coding gene proximal to *Pra* across the four cereal rust fungi. The TE composition in regions proximal to *Pra* varied within dikaryotic genomes and between species (S11B Fig). This indicates the accumulation of distinct transposable element families around the PR locus in different cereal rust species (see more detail for *Pt* below).

Our results show that the PR locus appears highly plastic with little conservation within and between species.

### **HD genes are multiallelic in cereal rust fungi while *Pra* genes display little variation at the species rank**

We extended our species rank analysis of variation in *MAT* genes (Figs 1 and 2) using publicly available genomic data for all four species (S1 Table). We identified between 12 or 15 representative isolates per species that were assigned to as many distinct genomic lineages as possible based on previous population genetic studies [35,45–47,49–51,55–58] (S1 Table). Our reasoning for this approach was to avoid repeated sampling from clonal lineages that are predicted to show no genetic variation at *MAT* genes. We quality controlled whole-genome short-read datasets and mapped them against the respective reference genome to estimate the genetic variation in *MAT* genes (S13–S16 and S18–S21 Figs). We identified high levels of polymorphisms at the HD locus for all analyzed cereal rust fungi. In our dataset, *Pca* showed the highest level of variation of the HD locus for all species



**Fig 4. HD loci are partially conserved within species whereas PR loci are highly heterozygous and display strong signals of genomic degeneration.** Synteny graphs of HD locus (A) and (B) PR locus including proximal regions in *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f.

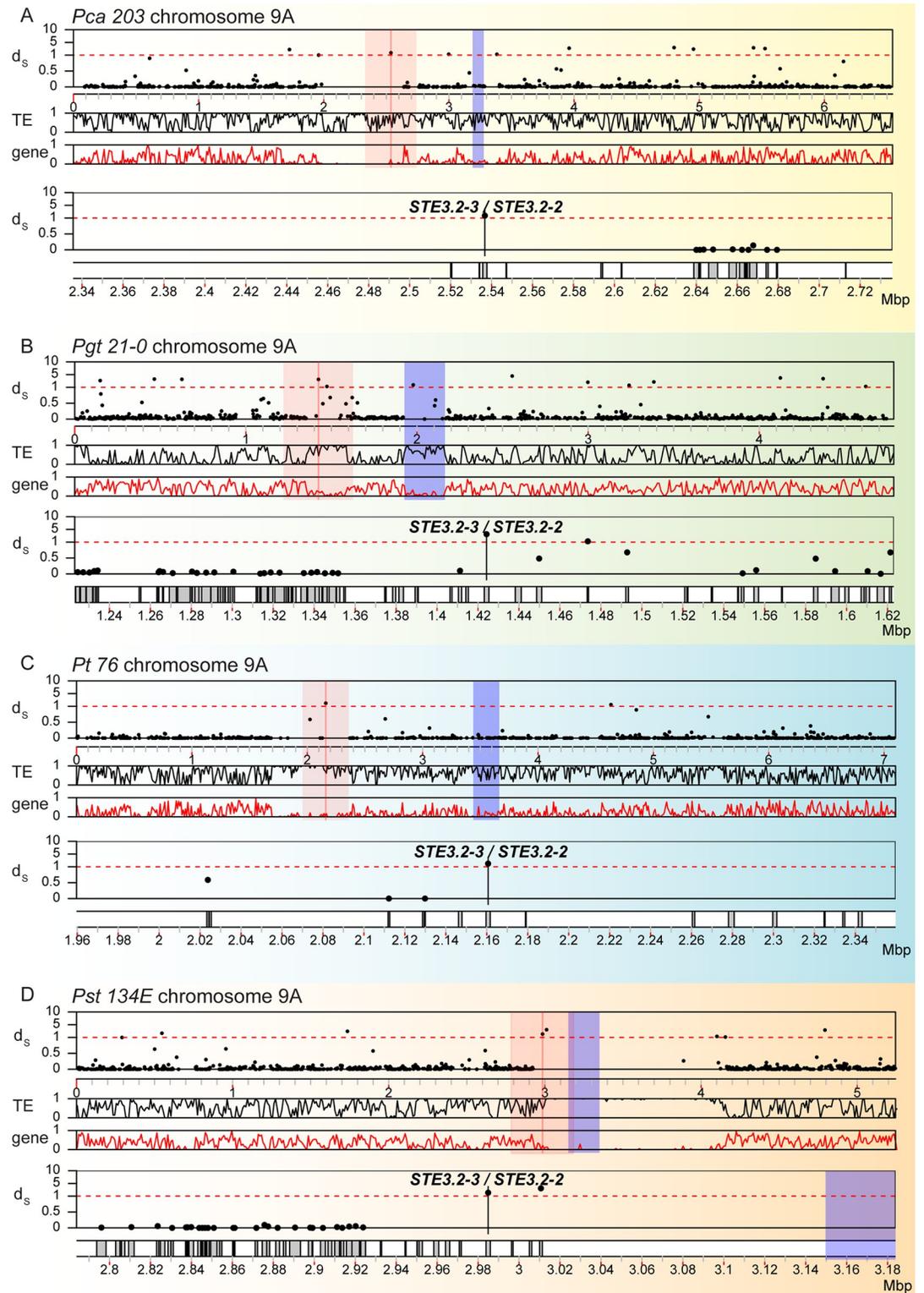
sp. *tritici* (“*Pgt 21-0*”), (C) *P. triticina* (“*Pt 76*”), and (D) *P. striiformis* f. sp. *tritici* (“*Pst 134E*”). Proximal regions are defined as 40 genes downstream and upstream of *HD* and *Pra* genes, respectively. HD loci are relatively syntenic within each dikaryotic genome but less conserved between species. PR loci display little synteny in *Pgt 21-0*, *Pt 76* and *Pst 134E* and show strong signs of transposon accumulation (see also S13 Fig). There is very little conservation of the PR loci across species. Red lines between chromosome sections represent gene pairs with identity higher than 70% and grey shades represent conserved nucleotide sequences (> = 1000 bp and identity > = 90%). For additional annotations please refer to the included legend (“Legend”).

<https://doi.org/10.1371/journal.pgen.1011207.g004>

(S13 Fig). When mapping Illumina short-read data of the *Pca* isolates against the *Pca 203* reference, the HD locus was highly polymorphic including heterozygous single nucleotide polymorphisms (SNPs) and multiple unaligned regions and gaps. This suggested that the HD locus present in our selected *Pca* isolates were not well reflected in the *Pca 203* reference assembly. This is likely due to sampling from diverse sexual populations of *Pca* in the USA [45]. We therefore implemented a novel *de novo* reconstruction approach for the HD locus its encoded *HD* genes directly from whole-genome Illumina short-read datasets (see Methods for detail). We confirmed our approach by reconstructing the *Pca 203* HD locus from publicly available Illumina short-read datasets. We confirmed that the two reconstructed *HD* alleles were identical to the ones derived from the dikaryotic reference genome assembly (Fig 1) via a nucleotide alignment of the two reference *bW-HD1* and *bE-HD2* alleles to the *de novo* reconstructed *HD* alleles (S17A Fig). In total, we reconstructed eleven *bW-HD1* and twelve *bE-HD2* alleles for *Pca* (S17B Fig). Using this approach, we identified extensive genetic variation for *bW-HD1* and *bE-HD2* in *Pgt* and *Pt* (S17C–S17D Figs) while making use of recently published *bW-HD1* and *bE-HD2* alleles in case of *Pst* [23] (S17E Fig). We identified six, nine, and ten *bW-HD1* and six, eight, and ten *bE-HD2* alleles for *Pgt*, *Pt*, and *Pst*, respectively. Overall, this suggests that the *HD* genes are multiallelic in all four cereal rust fungi, which is consistent with other recent reports [23,37].

Compared to the *HD* genes, the *Pra* genes are reported to be far less polymorphic in many basidiomycetes [9,77,78]. Hence, we next focused our analysis on *Pra* genes using the same whole-genome Illumina short-read datasets (S1 Table). As expected, the *Pra* genes were far less polymorphic within each species compared to *HD* genes (S18–S22 Figs), which is consistent with our initial analysis of *Pra* genes extracted from genome assemblies (Figs 1 and 2). In contrast to the highly polymorphic *HD* genes in *Pca*, we identified only two SNPs in the coding regions of *STE3.2-2* and *STE3.2-3* with a single SNP being non-synonymous (S22 Fig). Similarly, *STE3.2-2* and *STE3.2-3* copies of all *Pst* isolates had identical coding sequences. Our short-read mapping analysis of *Pra* in *Pt* identified four SNPs in coding regions of *STE3.2-2* and *STE3.2-3*. These SNPs gave rise to a single additional *STE3.2-2* variant with two amino acid substitutions close to the C-terminus (S22 Fig). *Pgt* was the most polymorphic for *Pra*. We identified multiple distinct copies of *STE3.2-2* and *STE3.2-3* in the global *Pgt* population. We identified one *STE3.2-2* variant with several amino acid changes (S22B Fig). *STE3.2-3* was the most polymorphic in *Pgt* including two isolates, *TTTSK* and *UVPgt60*, which contain potential non-sense mutations leading to pre-mature stop codons (S22B Fig). Overall, our analysis suggests that *Pra* is most likely biallelic in *Pca*, *Pt*, and *Pst*, while *Pgt* may have more than two functional *Pra* alleles.

Lastly, we investigated the nucleotide sequences of all *mfa* alleles at the species rank. *mfa1* and *mfa2* alleles were fully conserved in *Pca* and *Pst*, while *Pt* had only a single non-synonymous change in *mfa1/3* (S3 Fig). In contrast, *mfa1* and *mfa3* in *Pgt* had several non-synonymous variations, yet all were located outside the predicted mature pheromone peptide sequences (S23 Fig). This species level variation in *Pgt* at *mfa* is consistent with the variation observed in *STE3.2-2* and *STE3.2-3*.



**Fig 5. PR loci display elevated synonymous divergence ( $d_s$ ) values, accumulation of transposable elements and depletion of genes.** Synonymous divergence values ( $d_s$ ) for all allele pairs are plotted along chromosome 9A for (A) *P. coronata* f. sp. *avenae* (“Pca 203”), (B) *P. graminis* f. sp. *tritici* (“Pgt 21-0”), (C) *P. triticina* (“Pt 76”), and (D) *P. striiformis* f. sp. *tritici* (“Pst 134E”). In each panel, the top track shows the  $d_s$  values (“ $d_s$ ”) of allele pairs along chromosome 9. Each dot corresponds to the  $d_s$  value of a single allele pair. The second and third track shows the averaged TE (“TE”) and gene (“gene”) density along chromosome 9 in 10

kbp-sized windows, respectively. The *Pra* alleles (*STE3.2-2* and *STE3.2-3*) are highlighted with a red line and red shading indicates a 0.4 mbp-sized window around the PR locus. Predicted centromeric regions are marked with blue shading. The two lower tracks ( $d_s$  values and gene locations) provide a detailed zoomed in view of red shaded area around the PR locus. Species-specific background coloring is the same as for Fig 1.

<https://doi.org/10.1371/journal.pgen.1011207.g005>

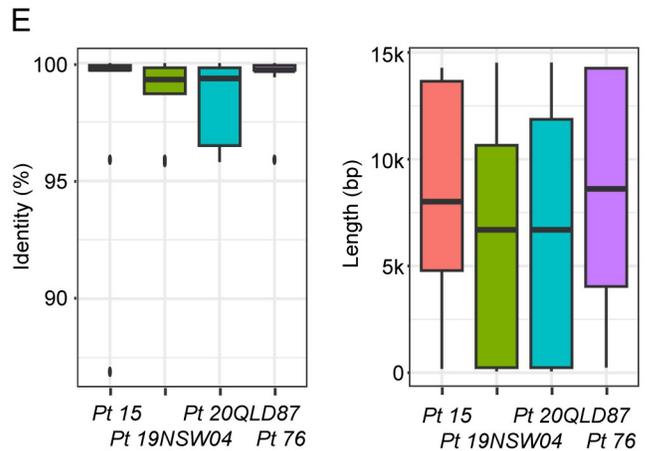
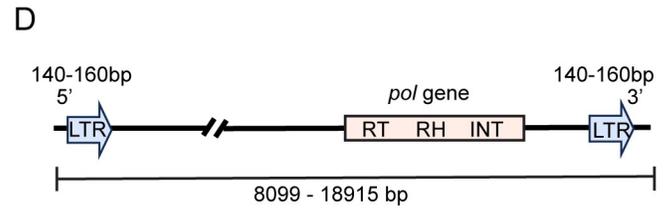
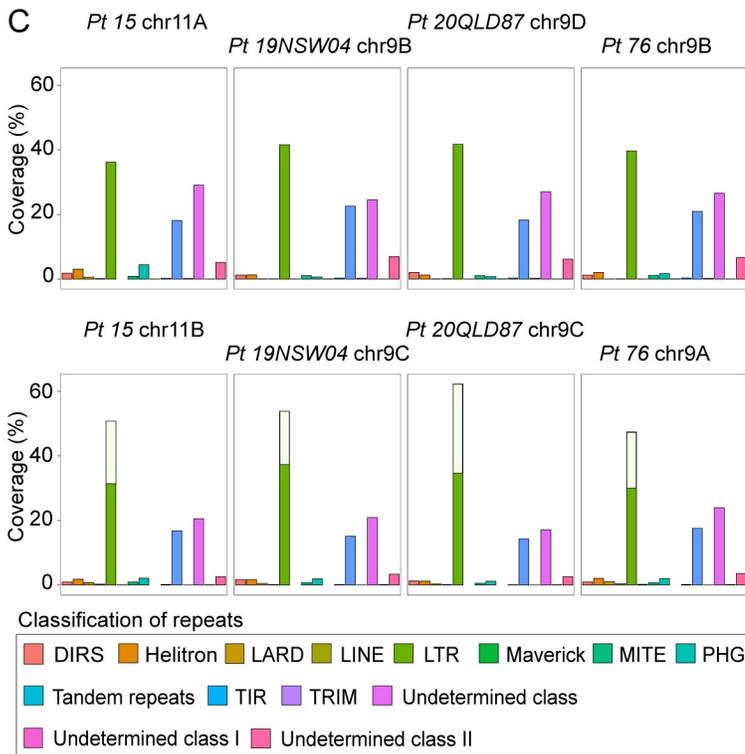
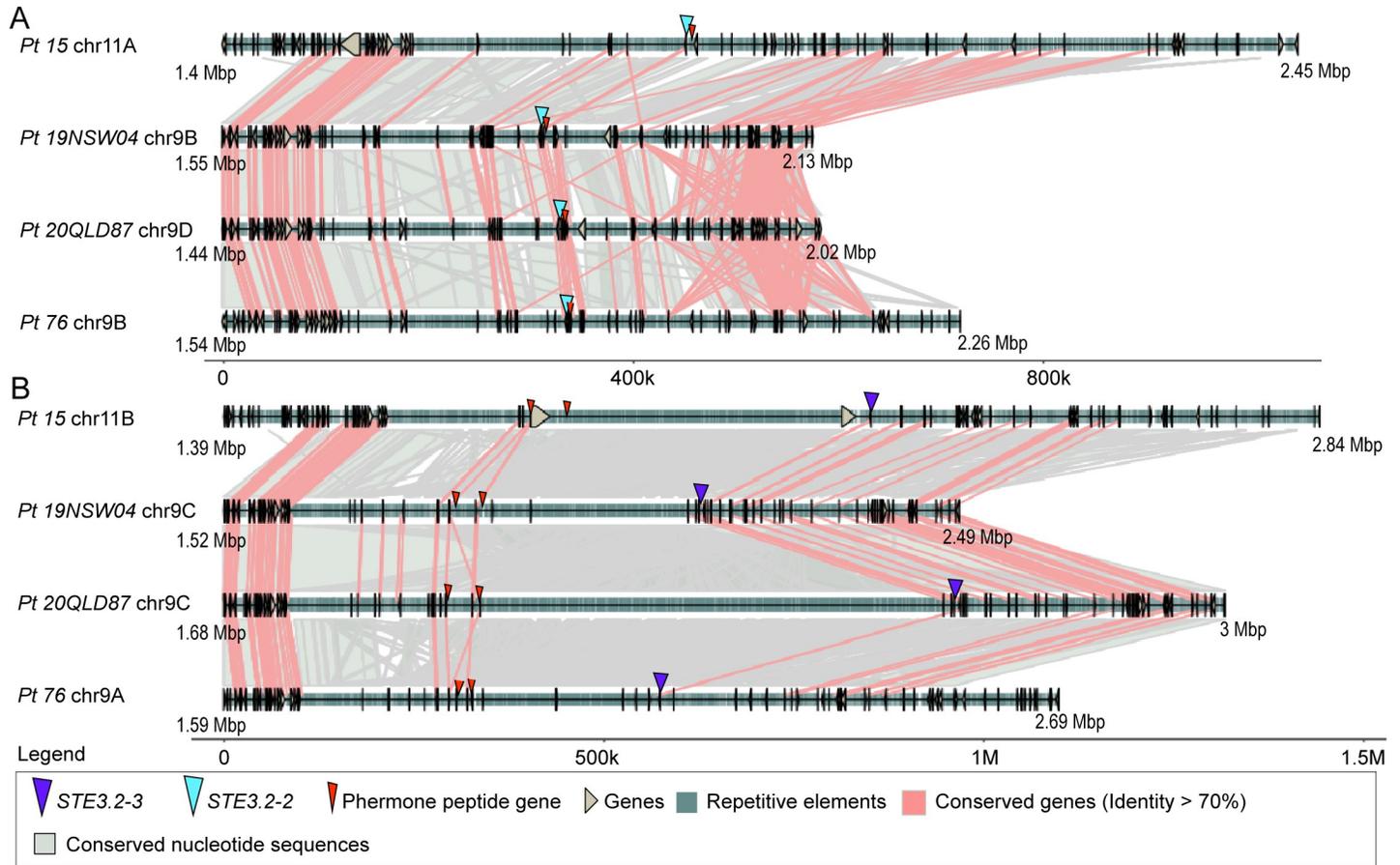
## MAT loci are conserved at the species level in *Puccinia triticina* while showing transposable element expansion specific to the *STE3.2-3* allele

We made use of four chromosome-scale and fully-phased genome assemblies for *Pt* (Table 1) to explore structural conservation or plasticity of the MAT loci within a species of cereal rust fungi. We analyzed macro-synteny, detailed nucleotide synteny, protein-coding gene conservation and TE composition for the HD and PR locus within and between *Pt* 76 [48], *Pt* 15 [59], *Pt* 19NSW04 and *Pt* 20QLD87 [37]. A recent study suggested that *Pt* 19NSW04 arose via somatic hybridization between *Pt* 76 and *Pt* 20QLD87 or close relatives in Australia mid-2010 [37]. Hence, *Pt* 19NSW04 and *Pt* 76 share one complete nuclear haplotype B, which contains *STE3.2-2*, and *Pt* 19NSW04 and *Pt* 20QLD87 share one complete nuclear haplotype C, which contains *STE3.2-3*. The *Pt* 15 isolate is unrelated and was sampled in China in 2015.

Macro-synteny plots of the HD locus containing chromosomes suggested that shared nuclear haplotypes e.g. *Pt* 19NSW04 and *Pt* 76 hapB or *Pt* 19NSW04 and *Pt* 20QLD87 hapC, are more conserved when compared to the HD locus contained within each dikaryotic genome (S5 and S24 Figs). Detailed nucleotide synteny and protein-coding gene conservation analysis revealed that gene order and function is highly conserved in regions proximal to the HD locus between all tested nuclear haplotypes (S25 Fig). To investigate whether any specific TE families cluster around the HD locus, we built a TE database based on the *Pt* 76 reference genome and used it to annotate TEs in all four isolates. We identified similar coverage of TEs, at the order classification level, around the HD locus in all dikaryotic genomes (S26 Fig).

We used identical analyses for the PR locus to understand its evolution at species level in *Pt*. We compared the overall structural relationship between chromosomes containing *STE3.2-2* or *STE3.2-3* using macro-synteny plots. Like the HD locus, chromosomes of shared nuclear haplotypes, e.g. *Pt* 19NSW04 vs *Pt* 76 hapB with *STE3.2-2*, were more similar to each other than non-shared haplotypes, e.g. *Pt* 76 hapB vs *Pt* 15 hapA with *STE3.2-3* (S12 and S27A Figs). The same applied for *STE3.2-2* (S12 and S27B Figs). We clearly observed clustering of repetitive sequences around the *STE3.2-2* and *STE3.2-3* locus with the latter more pronounced (S27 Fig). These repetitive sequences appeared to be allele specific as they were not visible in macro-synteny plots comparing *STE3.2-2* and *STE3.2-3* containing chromosomes (S12 Fig).

Detailed analyses of nucleotide synteny and protein-coding genes showed that genes proximal to either *Pra* allele were mostly conserved while some haplotypes appeared to have inversions downstream of *STE3.2-2* relative to the *Pt* 76 reference (Fig 6A and 6B). Gene synteny was more conserved for genes surrounding *STE3.2-3* yet gene distance appeared to be especially variable between the two *mfa1/3s* and *STE3.2-3* (Fig 6B). We explored the TE composition around each *Pra* allele and tested if TE expansions could explain the difference in intergenic distances around the *STE3.2-3* allele. TE composition around *STE3.2-2* and *STE3.2-3* were markedly different with the *STE3.2-3* allele displaying an increase of LTR retrotransposons (Fig 6C). TE coverage and composition around *STE3.2-2* was overall consistent between the different haplotypes (Fig 6C). In the case of *STE3.2-3* we found a single LTR Ty3 (also known as Gypsy [79]) TE family (Ty3\_Pt\_STE3.2-3, Fig 6D) highly expanded at the locus with varying coverage between haplotypes ranging from 14.7% to 25.4% with *Pt* 20QLD87 having the highest coverage (Fig 6C). The percentage identity relative to the consensus sequence was > 99% and most copies were around 8 kb (Fig 6E), which indicates active transposition of



**Fig 6. The PR locus shows an STE3.2–3 allele specific accumulation of a unique Ty3-transposon family in *Puccinia triticina*.** Synteny graphs of (A) STE3.2–2 and (B) STE3.2–3 including proximal regions in four different *P. triticina* isolates *Pt* 15, *Pt* 19NSW04, *Pt* 20QLD87, and *Pt* 76. Both loci are overall syntenic while the STE3.2–3 locus displays an extension of the intergenic distance especially between the pheromone receptor and the two *mfa* genes. Red lines between chromosome sections represent gene pairs with identity higher than 70% and grey shades represent conserved nucleotide sequences (> = 1000 bp and identity > = 90%). For additional annotations please refer to the included legend (“Legend”). (C) Transposable element coverage at the order classification level at the PR locus in the four different *P. triticina* isolates, empty white bars represent the coverage of Ty3\_Pt\_STE3.2–3. (D) Structure of a representative Ty3\_Pt\_STE3.2–3 copy. LTR–Long Terminal Repeat, RT–Reverse transcriptase, RH–RNase H, and INT–Integrase. Distributions of (E) percentage identity relative to the consensus sequence (left) and lengths (right) of individual Ty3\_Pt\_STE3.2–3 insertions at the STE3.2–3 gene and proximal regions in four different *P. triticina* isolates.

<https://doi.org/10.1371/journal.pgen.1011207.g006>

Ty3\_Pt\_STE3.2–3 in recent history. To investigate whether this TE family was specific to the STE3.2–3 allele, we compared its abundance at the PR locus with its abundance across the whole-genome. We found that Ty3\_Pt\_STE3.2–3 only clustered on STE3.2–3 containing chromosomes in all four strains with only a very limited number of single copies found on other chromosomes (S28 Fig). Indeed, Ty3\_Pt\_STE3.2–3 had a strong preference to insert between *Ptmfa1/3s* and STE3.2–3 (S29 Fig). To further investigate the structure of Ty3\_Pt\_STE3.2–3, we used LTR\_FINDER [80] to identify 5' and 3' LTR regions and other features related to retrotransposons. Ty3\_Pt\_STE3.2–3 was found to have a short LTR, around 150 bp in all four strains, with target site repeat (TSR) region of ‘AAGT’, *pol* genes containing reverse transcriptase (RT), ribonuclease H (RH), integrase (INT). However, we failed to identify any group-specific antigen (*gag*) in these TEs (Fig 6D).

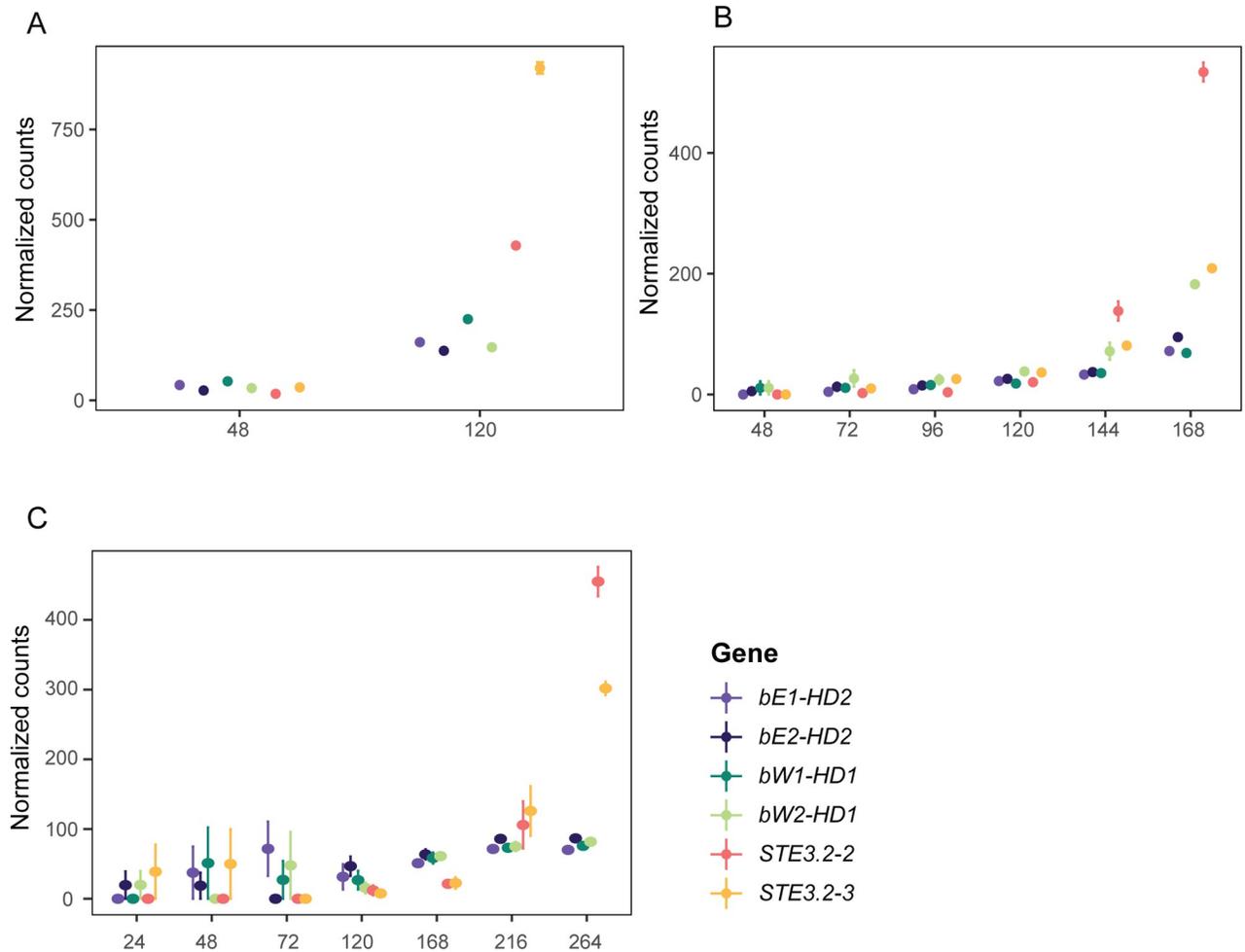
### STE3.2–1 genes are highly conserved within and among cereal rust fungi

Similar to previous studies [19,34], we identified an additional *Pra*-like gene, STE3.2–1, on chromosome 1. However, its two alleles are nearly identical in all dikaryotic genome assemblies and are not associated with any pheromone peptide precursor genes [70]. We investigated the genealogy of the *Pra*-like STE3.2–1 with using *P. polysora* f. sp. *zeae* STE3.2–1 as an outgroup [81]. The two STE3.2–1 alleles identified in each isolate were almost identical and grouped by species identity. Overall, STE3.2–1 showed very little variation within each species (S30 Fig).

We also analyzed STE3.2–1 and its proximal regions as above for the MAT loci. The three independent analyses revealed that STE3.2–1 and its proximal regions are highly conserved at the nucleotide and protein-coding gene level within dikaryotic genomes and are syntenic between species (S31–S33 Figs). This shows a clear difference between the STE3.2–1 gene from the two MAT genes. Hence, STE3.2–1 is likely not involved in mating compatibility in the examined cereal rust fungi as previously suggested.

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

Importance of MAT genes in mate compatibility is well established for *Pt* and *Pst* [19,82], yet it is unclear if they are expressed and functional during asexual reproduction on cereal hosts as found in *Pt* [19]. Hence, we investigated the expression patterns of MAT genes in *Pca*, *Pgt* and *Pst* using publicly available RNA-seq infection time series of their cereal hosts [43,46,53,83] (S2 Table). We applied the trimmed mean of M-values (TMM) normalization to read counts and assessed quality of the RNA-seq datasets by multidimensional scaling (MDS) plots (S34 Fig). The MDS plots confirmed the suitability of the datasets for detailed expression analysis based on technical replicates clustering closely together and one dimension separating samples according to their infection progress. The stable expression of two house-keeping genes for each species across all samples further confirmed the suitability of the datasets for detailed MAT gene expression analysis (S35 Fig).



**Fig 7. MAT gene expression is upregulated in the late stage of asexual infection cycle of the cereal host.** (A) Trimmed mean of M-values (TMM)-normalized values of MAT genes in *P. coronata* f. sp. *avenae* (“Pca 12NC29”) at 48 and 120 hour post infection (hpi). (B) TMM-normalized values of MAT genes in *P. graminis* f. sp. *tritici* (“Pgt 21-0”) at 48, 72, 96, 120, 144 and 168 hpi. (C) TMM-normalized values of MAT genes in *P. striiformis* f. sp. *tritici* (“Pst 87/66”) at 24, 48, 72, 120, 168, 216 and 264 hpi.

<https://doi.org/10.1371/journal.pgen.1011207.g007>

*STE3.2-2* and *STE3.2-3* were upregulated during the asexual infection process and always displayed highest expression at the latest time point available, which coincides with sporulation and production of urediniospores (Fig 7). We confirmed the differential expression at later infection timepoints with a likelihood ratio test with a p-value cut-off of < 0.05 (S36 Fig). Similarly, *HD* genes are upregulated during asexual infection of the cereal host (Fig 7). We did not observe any expression of *STE3.2-1* in any of the samples. Lastly, we were also interested to see if we could detect transcripts of MAT genes in cereal rust spores before infection and in specialized infection structures called haustoria. We used an additional *Pst* RNA-seq dataset and compared the expression of MAT genes in ungerminated spores, germinated spores, 5-days post infection, 9-days post infection and haustoria enriched samples (S37 Fig) [53]. *HD* genes displayed some expression in spores albeit lower than observed in later stages of infection during spore production in the asexual cycle. In contrast, we did not detect any expression of *Pra* genes in ungerminated and germinated spores. Yet, *Pra* genes were expressed and upregulated at later stages of the asexual infection consistent with other publicly available datasets (Figs 7C and S37).

## Discussion

Our comparative analysis of predicted *MAT* genes and *MAT* loci in four cereal rust fungi provides novel insight into the evolution of these genes and their proximal regions within and between species.

First, we confirmed that HD and PR loci are unlinked and located on two distinct chromosomes in all four cereal rust fungi, which supports previous analyses on more fragmented genome assemblies or for individual rust fungal species [19,23,35]. Further, both loci were heterozygous in all twelve dikaryotic genome assemblies. Taken together, these observations strongly support the hypothesis of rust fungi displaying a tetrapolar mating compatibility system. Though this genome biology informed hypothesis requires experimental validation via crosses and genetic analysis of resulting offspring.

As previously observed for *Pt* [19], the *Pra* homologs *STE3.2-1* were highly conserved in all cereal rust fungi and had no associated *mfa* genes. This suggests that *STE3.2-1* does not regulate mate compatibility. In *P. coronata* f. sp. *avenae*, *P. striiformis* f. sp. *graminis* and *P. striiformis* f. sp. *tritici*, *STE3.2-1* is flanked by anaphase promoting complex *APC10/Doc1*, which regulates mitosis [84], and Elongation factor *EF1B* [85]. Together, these findings suggest that *STE3.2-1* might play a role during the cell cycle or different developmental programs as reported for other basidiomycete fungi [86,87]. However, as previously reported for *P. triticina* [19], expression for *STE3.2-1* was absent or very low during the infection cycle of the cereal host.

Fungi with a tetrapolar mating compatibility system are often multiallelic for at least one of the two *MAT* loci. For example, *U. maydis* has many distinct transcription factor alleles at the HD locus [88,89]. We performed extensive analysis of the HD locus extracted from dikaryotic genome assemblies and developed a workflow to *de novo* reconstruct loci directly from Illumina short-read sequencing data. We investigated the allelic diversity of *bW-HD1* and *bE-HD2* in 12–15 additional isolates for each cereal rust fungal species. For each species these were sampled in a way to provide the best available representation of unique genotypes from global populations. This analysis demonstrated that the HD locus is multiallelic in the four species of cereal rust fungi. The four species had between 6–12 alleles of *bW-HD1* and *bE-HD2*. This supports previous analyses of the allelic diversity of these genes in *P. striiformis* f. sp. *tritici*, which identified nine alleles of *bW-HD1* and *bE-HD2* [23]. Our reported number of *bW-HD1* and *bE-HD2* alleles is likely an underestimate given the current limited sampling publicly available. Most analyzed isolates were retrieved from cereal hosts during the asexual reproduction cycle and from regions that are known to be deprived of sexually recombining populations. The only exception was *P. coronata* f. sp. *avenae* isolates, which are reported to be derived from sexual recombining populations in the United States of America [45]. We identified the highest allelic variation of HD genes (eleven *bW-HD1* and twelve *bE-HD2*) in *P. coronata* f. sp. *avenae*. This is similar to previous studies of isolates from likely sexual populations of *P. striiformis* f. sp. *tritici* in China, India, and Pakistan, which showed the highest diversity in HD alleles compared to other clonal lineages in Western wheat growing regions [23].

In contrast to multiallelic HD genes, the *Pra* and associated *mfa* genes were biallelic in *P. coronata* f. sp. *avenae*, *P. triticina* and *P. striiformis* f. sp. *tritici*. In all cases, one *Pra* allele was linked to one specific invariant *mfa* allele whereby *P. triticina* carried two *mfa1/3* genes that were nearly identical in the sequence of the predicted mature pheromone peptide. This biallelism of *Pra/mfa* is similar to other tetrapolar dimorphic basidiomycetes like *Microbotryum* sp., *U. maydis* and *Cryptococcus* sp. [63,90,91]. The notable exception was *P. graminis* f. sp. *tritici*, which encodes two *mfa* genes (*mfa1* and *mfa3*) in proximity to *STE3.2-3*. The two encoded precursor peptides are highly variable and are likely processed into two distinct mature

pheromone peptides with at least one amino acid difference. In addition, we cannot exclude that MFA3 encodes for additional mature pheromone peptides, which might lead to distinct receptor activation. Interestingly, *Pgt STE3.2–3* was also the most variable *Pra* allele in all species. We identified three *STE3.2–3* variants that have at least four or more non-synonymous nucleotide changes relative to the most common *Pgt STE3.2–3* variant. Moreover, the *STE3.2–3* variants in *Pgt 126–6711*, *Pgt ME-02*, and *Pgt PK-01* shared four distinctive amino acid changes in the C-terminus while being linked to an *mfa1* allele that gives rise to MFA1 with four amino acid substitutions. All analyzed *P. graminis* f. sp. *tritici* isolates carried an invariant *STE3.2–2* allele with only one *mfa* gene in proximity and we did not identify a clearly distinct additional *Pra* allele. Hence the functional significance of the observed variation at the *STE3.2–3* locus is currently difficult to assess without more extensive sampling, more phased chromosome scale genome assemblies and in the absence of direct functional crossing studies in *P. graminis* f. sp. *tritici*. Yet, the organization of the *STE3.2–3* gene is reminiscent of the PR locus of *Sporisorium reilianum* in the order Ustilaginales. In *S. reilianum*, the PR locus is at least triallelic with each allele encoding one pheromone receptor and two distinct pheromone peptides. Each of the mature pheromones derived from one allele specifically activates only one of the receptors encoded by the two other alleles [92]. More studies are needed to assess if the PR locus in *P. graminis* f. sp. *tritici* is bi- or multiallelic.

We initially predicted that biallelic MAT loci would show stronger signatures of recombination suppression and genomic degeneration than multiallelic loci based on a published mathematical model [2] and observations in other basidiomycete species [9,93]. Indeed, the multiallelic HD locus was mostly syntenic within dikaryotic genomes of the four cereal rust species. This initial analysis with a limited set of genomes per species is likely to hold true at the population level. In the case of *P. triticina*, the eight HD alleles of four isolates were all highly syntenic with most proximal genes conserved between the different alleles. Consistently, the composition of TEs at order classification level was highly similar between HD alleles of the same species.

In contrast, the PR locus showed strong signs of recombination suppression and genomic degeneration, however, the exact patterns were not identical in all four species. PR proximal regions showed the strongest signs of recombination suppression and were the most extended in *P. coronata* f. sp. *avenae*, *P. triticina* and *P. striiformis* f. sp. *tritici* with *Pra* genes located at edges or within large gene deserts and TE islands of 0.7–1.2 Mbp. At least some of these repeats were shared between the two PR alleles within each dikaryotic genome of *P. coronata* f. sp. *avenae* and *P. striiformis* f. sp. *tritici* based on whole chromosome alignments. This was not the case for PR alleles in *P. triticina*. In all three cases, there appeared to be differences in composition and coverage of TE at the order, superfamily, and family classification level between the two PR alleles. This was most obvious for *P. triticina*, where we analyzed four dikaryotic genome assemblies and identified a TE family specific to *STE3.2–3*, namely Ty3\_Pt\_*STE3.2–3*. Ty3\_Pt\_*STE3.2–3* displayed preferential insertion for the *STE3.2–3* allele between the pheromone receptor and the two *mfa* genes. Ty3\_Pt\_*STE3.2–3* is likely a currently active TE based on its sequence conservation, length of individual TE copies, and the suggested nuclear exchange between the studied Australian isolates within the last decade [37]. These initial observations of partial allele-specific TE composition and coverage in *P. coronata* f. sp. *avenae*, *P. triticina* and *P. striiformis* f. sp. *tritici* contrast with observations of TE analysis of biallelic MAT locus in 15 species of *Microbotryum*. The MAT locus in *Microbotryum* spp. did not display allele specific patterns of TE composition, yet were likely reservoirs for TE families that drove TE expansion at a genome scale at discrete time points associated with the extension of evolutionary strata around the MAT locus [42].

Compared to these observations at biallelic PR loci, the PR locus in *P. graminis* f. sp. *tritici* contained fewer TEs, more genes, and was much shorter at ~300 kbp. In addition, the TE composition and coverage was very similar between the two alleles within the dikaryotic genome of *Pgt 21-0*. These observations could be biological or a technical artifact of the assembly process of this specific genome. The *Pgt 21-0* assembly is based on older, noisy PacBio long-read technologies, included extensive manual curation and relied on gene synteny between contigs and haplotypes for scaffolding. This gene-based scaffolding might have been broken by long-stretches of TEs around the PR locus, as observed for the three other cereal rusts, and thereby introduced scaffolding errors in *Pgt 21-0* [35]. Several other genomes generated with older, error-prone PacBio long-read technologies (e. g. *Pst 104E* and *Pca 12SD80*) had issues assembling the PR locus correctly. Alternatively, these marked differences of the *P. graminis* f. sp. *tritici* PR locus might reflect biological reality, as models predict that recombination suppression and genomic degeneration is more pronounced at biallelic loci [2]. If the PR locus in *P. graminis* f. sp. *tritici* is not truly biallelic, as potentially indicated by the gene-based analysis of *Pra* and *mfa* alleles, the observed organization and level of genomic degeneration at the PR locus could support this prediction. However, more sampling and high-quality, phased chromosome-scale genome assemblies are necessary to differentiate these alternative hypotheses concerning the PR locus in *P. graminis* f. sp. *tritici*.

The functions of *MAT* genes in rust fungi are currently unknown, beyond their likely role in mediating mate compatibility. However, in *P. triticina*, *P. striiformis* f. sp. *tritici* and *M. larici-populina* [19,40,82,94], *MAT* genes are expressed during the sexual and asexual infection cycle. Suppression of *MAT* gene expression in *P. triticina* during the asexual infection cycle on wheat reduces infection severity and spore production [19]. Here, we show that *MAT* genes are also upregulated during the late stage of the asexual infection cycle of the cereal host in *P. coronata* f. sp. *avenae*, *P. graminis* f. sp. *tritici*, and *P. striiformis* f. sp. *tritici*. This suggests a broader conservation of *MAT* gene expression during the late stages of the asexual infection cycle of rust fungi. Cells of rust fungi carry more than two nuclei in the same cytosol during the asexual infection cycle [95]. Urediniospores always carry two nuclei of compatible mating-types and never two nuclei of identical mating-types. Studies from the early 20<sup>th</sup> century show that the two nuclei of urediniospore mother cells undergo synchronous nuclear division within the same cytoplasm giving rise to four nuclei within the same urediniospore precursor cells [95]. This is followed by nuclear movement to restore appropriate nuclear pairing and planar cell division that enables urediniospore maturation and reconstitutes the dikaryotic state [95]. The observed upregulation of *MAT* genes aligns with the development of urediniospore primordia and spore production. Hence, we hypothesize that *MAT* genes regulate nuclear pairing during production of dikaryotic urediniospores from primordia.

Understanding mating-types in cereal rust fungi has significant implications for agriculture and prediction of new virulent isolates. Recent reports suggest that nuclear exchange between isolates during the asexual infection cycle (also known as somatic hybridization [96]), in addition to sexual reproduction, can lead to nuclear assortment [96]. Nuclear exchange and the viability of the resulting offspring is likely regulated by *MAT* genes because all proposed events have given rise to isolates with dikaryotic genomes that carry opposing *MAT* gene pairs. We confirmed these observations with our *MAT* gene analysis in the case of *Pgt Ug99* (*Pgt 21-0* and unknown nuclear donor) and *Pt 19NSW04* (*Pt 76* and *Pt 20QLD87* as proposed nuclear donors). Conversely, our analysis showed that nuclear exchange has not occurred for the analyzed isolates of *P. coronata* f. sp. *avenae*, and *P. striiformis* f. sp. *tritici*, as none of the isolates share identical *HD* and *Pra* alleles that could be carried in the same nucleus.

In cases of nuclear exchange, these nuclei behave like entire linkage groups with virulence (also known as effectors) allele complements being tied to specific *MAT* alleles, if somatic hybridization occurs in the absence of parasexual reproduction as currently proposed. This implies that the nuclear *MAT* alleles define the possible virulence allele combinations that can arise via nuclear exchange, from either sexual or somatic reproduction. We hence can predict the most likely novel virulence allele combinations that might arise given circulating nuclear genomes in existing cereal rust fungal populations on their cereal hosts. This observation requires us to generate phased and chromosome scale genome assemblies of all cereal rust fungi at the population level. These will improve disease management strategies and the durability of characterized resistance loci in the global crop germplasm leading to smarter rust resistance breeding programs.

## Material and methods

### Assessment of used genomic resources

Detailed information on isolates and genomes used in this study can be found in [Table 1](#). Completeness of genome assemblies were assessed by BUSCO v5.5.0 [97] in the genome mode with the basidiomycota\_odb10 gene set.

### Identification of *MAT* genes

Annotation of *Pt 19NSW04* and *Pt 20QLD87* were downloaded from CSIRO Data Access Portal (URL: <https://data.csiro.au/>), CDS and protein data of *Pt 19NSW04* and *Pt 20QLD87* were generated with Gffread [98]. *MAT* genes including the *HD* and *Pra* genes previously identified in *P. triticina* ([Table 2](#)) [19] were used as blastp and blastn [99] query to identify orthologs in *P. graminis* f. sp. *tritici* *Ug99* and *21-0* [35], *P. striiformis* f. sp. *tritici* *134E* [52], *104E* [53] and *DK0911* [54], *P. triticina* *76* [48], *19NSW04*, *20QLD87* [37] and *15* [59], *P. polysora* *GD1913* [81], *P. coronate* f. sp. *avenae* *203* [44], *12NC29* and *12SD80* [43] isolates. Missed *STE3.2-3* in *Pst 104E* was fixed by mapping short-read to reference genome *Pst 134E* with bwa-mem2 [100]. *Mfa* genes were identified by using *mfas* of *Pt BBBB*, *Pgt SCCL* and *Pst 78* [19] as a custom motif database to search similar pattern in all studied species with Geneious Prime v. 2022.1.1 [101], using a function built from fuzznuc [102]. Karyograms of *Pca 203*, *Pgt 21-0*, *Pt 76* and *Pst 134E* were made with RIdeogram 0.2.2 [103].

### Genealogical analysis of the *HD* and *Pra* genes

To construct a species tree for cereal rust fungi, protein sequences encoded in one nuclear genome from each cereal rust fungus species under study was used as input data. The input data was processed by Orthofinder 2.5.5 [67,104] to infer maximum likelihood trees from multiple sequence alignments (MSA). fasttree v2.1 [105] was used to calculate likelihood-based local support values. Coding regions of identified *HD* and *Pra* genes were aligned with MACSE v2.07 [106] with default setting. Coding-sequence based alignments were trimmed by using TrimAl v1.2 [107] with option (-gt 0.9) to remove sequences which have gaps in more than 10% of samples, trimmed alignments were realigned with MACSE again. Nucleotide alignments were imported into BEAUti v2.7.6 [108] to generate.xml files, bModelTest 1.3.3 [109] was used to test the best nucleotide substitution model separately with a Markov chain length of 15 million. HKY+G was chosen for inferring two *HD* genes, whereas TN93+I was chosen for *Pra* genes and *STE3.2-1* individually. Then BEAST 2 MCMC was run under a strict molecular clock with the Yule Model tree. Fifteen million length Markov chains were used to provide a sufficient effective sample size (ESS) of independent samples. Tracer v.1.7.2 [110]

was used to check if ESS values of all parameters are above 200 for each run. TreeAnnotator [108] was used to generate a maximum clade credibility tree with 10% burn in. Figtree v.1.4.4 [111] was used to visualize the final output.

Nucleotide sequences of identified *mfa* genes were aligned with MACSE v2.07 [106] with default settings, trimAl v1.2 [107] with option (-gt 0.9) to remove gaps, trimmed alignment was used to construct maximum likelihood tree via IQ-TREE 2 (version 2.1.4-beta) [112], the best model was estimated by ModelFinder (-m MFP) [113] based on Akaike Information Corrected Criterion (AICc) score [114], 10000 replicates were set for ultrafast bootstrap [115]. K2P+G4 substitution model was chosen by ModelFinder [113].

To compare topologies of *bW-HD1* and *bE-HD2* gene trees, IQ-TREE 2 (version 2.1.4-beta) [112] was used to build unrooted maximum likelihood trees for each species and both genes separately, the best model for each alignment estimated by ModelFinder (-m MFP) [113] based on AICc score [114], 10000 replicates were set for ultrafast bootstrap [115]. Top 100 optimal gene trees of *bW-HD1* and *bE-HD2* were used as alternate topologies after removing branch length values to perform AU tests [68] with IQ-TREE2 (-keep-ident-zb 10000 -n 0 -au),  $p_{AU} < 0.05$  suggesting topological incongruence between the two *HD* genes.

In order to detect potential recombination events, the HD locus plus 1kb of proximal DNA sequence of each species were aligned with MAFFT v7.490 [116,117] in Geneious Prime v. 2022.1.1 [101], both sides of each alignment were trimmed manually. The RDP [118], GENECONV [119], BOOTSCAN [120], MAXCHI [121], CHIMAERA [122], SISCAN [123], and 3SEQ [124] methods in the program RDP5 [69] were processed to detect unique recombination events for the HD locus of each species using default settings. Recombination events which were detected by at least five of the above methods were accepted. RDP5 projects of each species are available in Dryad Database [70].

### Estimation of centromeric regions

Hi-C reads of *Pca 203* and *Pt 76* were mapped back to their corresponding haploid genome assembly with Juicer 1.6 [75] and processed with the 3d-dna [125] pipeline. Juicebox 1.8.8 [75] was used to visualize Hi-C maps. The approximate positions of centromeric regions were estimated by zooming in on the Hi-C heatmap in Juicebox and selecting the region corresponding to strong bowtie like interaction signals, which have been validated and widely used in previous studies [126, 127].

### $d_s$ value estimation for allele pairs and plotting

Inspired by the approach used by Branco et al. [8] for identifying evolutionary strata on mating-type chromosomes of *Microbotryum* species, proteinortho 6.0.22 [128] was used to pair orthologs one-to-one with -synteny and -singles tag. Muscle 3.8.1551 [129] was applied to generate coding-sequence based protein alignments with default settings. PAML 4.9 [130] was used to calculate synonymous divergence values from alignments generated with Muscle. Genomes were split in 10 kbp per window, then density of genes and repeats were calculated with bedtools v 2.29.2 [131]. Results were visualized with karyoploteR v3.17 [132].

### Synteny analysis of *MAT* genes and proximal regions

For investigating macrosyntenic relationship between haplotypes, self-alignment analyses between two haplotypes of each species were done with NUCmer (version 3.23) [72] by using-maxmatch option, alignments with a size of less than 1000 bp or less than 90% identity were removed, filtered data were visualized via Matplotlib [133].

To investigate synteny conservation, NUCmer (version 3.23) [72] was used to align nucleotide sequences of *MAT* genes and proximal regions. Blastn 2.10.0 [99] was used to identify conserved genes (identity higher than 70%, e-value < 0.05) between and within species. Results of 40 genes flanking the *MAT* genes were visualized with gggenomes 0.9.9.9000 [134].

### General repeat annotations

REPET v3.0 pipeline [135,136] was applied to annotate repeats in all genomes used in this study, together with Repbase v22.05 [137], config files were based on default settings, Wicker classification system was used to classify TE families into Class-Order-Superfamily-Family [76]. TEdenovo was used to predict novel repetitive elements and to construct specific custom databases from each genome individually as described in the manual instructions. Constructed custom databases were used for repeat annotation with TEannot. The reference TE database built for *Pt 76* from TEdenovo with the REPET v3.0 pipeline was used as reference for annotating repeats in *Pt 15* [59], *Pt 19NSW04* and *Pt 20QLD87* [37] with RepeatMasker 4.1.5 [138]. To make analyses consistent, the *Pt 76* genome was reannotated in the same way as the other three isolates.

To calculate the coverage of each TE family, classification information from TEannot was used. For TEs with more than one potential classification, hits from Repbase which have more than 70% of similarity were used as evidence for reclassification.

LTR\_FINDER v1.07 [80] with ps\_scan [139,140] was used to identify detailed structures of LTRs of Ty3\_Pt\_STE3.2–3 in default setting. Reverse transcriptase (A0A0C4EV49\_PUCT1), Retrotransposon gag domain (A0A0C4EKM8\_PUCT1), Integrase, catalytic domain (A0A0C4EMG6\_PUCT1) and Retropepsins (A0A0C4F2V2\_PUCT1) [141], were additionally used as query to search for motifs in identified LTR retrotransposons using tblastn 2.12.0+ [99].

### Identification of variants in *MAT* loci

Sequence data of *Pca* were published by [45], data of *Pgt* were published by [35,46,47,58] and under PRJNA39437, data of *Pt* used in this study have been published earlier by [49–51] and under PRJNA39803 and PRJNA39801, data of *Pst* were published by [55–57] and under PRJNA60743, all above data were downloaded from NCBI SRA database (URL: <https://www.ncbi.nlm.nih.gov/sra>) (S1 Table). Four *Pca* isolates, *Pca 203* [44], *Pgt 21–0* [35], *Pt 76* [48] and *Pst 134E* [52], with published chromosome-scale haplotype-phased genome assemblies were used as reference for mapping. Trimmomatic 0.39.2 [142] was used to remove adapters from raw reads whereas FastQC v0.11.8 [143] was applied for assessing the quality before and after trimming. Bwa-mem2 [100] was used to map trimmed reads to reference genomes. MarkDuplicates (Picard) 1.70 [144] was applied to remove PCR-generated duplicates. Samtools 1.12 [145] was used to remove mapped reads with mapping quality lower than 30. IGV 2.16.2 [146] and Qualimap 2 [147] were both used to check mapping quality.

For reconstruction of *Pra* alleles, freebayes-parallel 1.3.6 [148,149] was applied to detect variants, bcftools 1.12 [145] was used to exclude low quality variants with `-e 'QUAL<40'`. Samtools 1.12 [145] was used to filter out regions of interest. The Ensembl Variant Effect Predictor (VEP) 88.9 [150] was used to annotate sequence variants. Bcftools 1.12 [145] was used to generate consensus sequences, nucleotide and protein alignments were generated and visualized with Geneious Prime v. 2022.1.1 [101].

For *de novo* reconstruction of *HD* alleles, Spades [151], a *de novo* assembler, was used for building draft genome assemblies from available whole-genome Illumina short-read datasets (S1 Table). For isolates with average read lengths > 101 bp, k-mer size of 101 was chosen for building draft genomes. For isolates with average read lengths < 101 bp, k-mer size of 51 bp was chosen.

*HD* genes used in genealogical analysis were used as reference genes, blastn 2.12.0+ [99] was used to search and identify *HD* allele containing contigs of the draft genome assemblies. KAT [152] was used to compute a k-mer matrix from *de novo* assembled *HD* contigs and reference *HD* loci sequences with default setting (k-mer size = 27). KAT filter was used to obtain all paired-end reads which contained *HD* k-mers based on the created *HD* k-mer matrix (-T 0.2). Spades was used to reassemble all *HD* related paired-end reads of the specific isolate. Final outputs were visualized in Geneious, CDS regions of reconstructed *HD* alleles were predicted by mapping the CDS of reference *HD* alleles to reconstructed contigs, ORFs were presumed with Geneious.

### Expression analysis of *MAT* genes

RNA-seq reads of rust fungi on infected plants were obtained from NCBI SRA database (URL: <https://www.ncbi.nlm.nih.gov/sra>). The datasets included: RNA-seq data of *Pst* 87/66 on wheat (PRJEB12497 [83]), *Pca* 12NC29 on *Brachypodium distachyon* (PRJNA398546 [43]), *Pgt* 21-0 on wheat (PRJNA415866 [46]), *Pst* 104E on wheat (PRJNA396589 [53]). Raw reads were quality checked with FastQC v0.11.8 [143], trimmomatic 0.50 [142] was used to trim low quality reads with parameters: ILLUMINACLIP:adapter.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Kallisto 0.44.0 [153] was then applied to align and count transcripts with -b 100 parameter for all paired-end reads and -b 50 -l 200 -s 20—single—single-overhang for single-end reads. *Pst* 104E cDNA data was used as reference for *Pst* 104E infected tissues. Since no reference genome of *Pst* 87/66 is available, cDNA data of *Pst* 134E was used as reference instead with *HD* genes of *Pst* 87/66 supplemented to obtain better quantification of the specific *HD* alleles. *Pgt* 21-0 and *Pca* 12NC29 were used as reference for mapping reads from *Pgt* 21-0 and *Pca* 12NC29 infected tissues, respectively. Gene names of *Pca* 12NC29 and *Pst* 104E were added with funannotate annotate v 1.8.5 [154,155] with default parameters. Kallisto outputs were imported to EdgeR 3.17 [156] with tximportdata [157] following user instruction, followed by normalizing with the TMM method, significant upregulation of *MAT* genes was assessed with the LRT method and ggplot2 [158] was used to visualize the result. Housekeeping genes were identified in each dataset based on the following conditions: *p-value* > 0.1 and log(FC) < 0.5, functional annotations of candidate housekeeping genes were retrieved from interproscan [155].

### Dryad DOI

<https://doi.org/10.5061/dryad.w0vt4b8zm> [70].

### Supporting information

**S1 Table. List of whole-genome Sequence Read Archive data used in the present study.** The table provides information on all whole-genome Sequence Read Archive (SRA) data used in the present study. Metadata includes species abbreviation (“Species”), isolate name (“Isolate”), SRA identifier (“SRA ID”), and the initial reference (“Citation”). *Pca*—*P. coronata* f. sp. *avenae*, *Pgt*—*Puccinia graminis* f. sp. *tritici*, *Pt*—*P. triticina* and *Pst*—*P. striiformis* f. sp. *tritici*. (XLSX)

**S2 Table. List of RNAseq Sequence Read Archive data used in the present study.** The table provides information on all RNAseq Sequence Read Archive (SRA) data used in the present study. Metadata includes species abbreviation (“Species”), isolate name (“Isolate”), timepoint of infection (hpi, hours post infection) or sample type (“Sample Type”), timepoint/status of spores (“Time point/Status”), replicate number (“Replicate”), SRA identifier (“SRA ID”), and the initial reference (“Citation”). *Pca*—*P. coronata* f. sp. *avenae*, *Pgt*—*Puccinia graminis* f. sp.

*tritici*, and *Pst*—*P. striiformis* f. sp. *tritici*.  
(XLSX)

**S1 Fig. Two unlinked MAT loci suggest tetrapolar mating types in four *Puccinia* spp.** Karyograms of *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f. sp. *tritici* (“*Pgt* 21-0”), *P. triticina* (“*Pt* 76”) and *P. striiformis* f. sp. *tritici* (“*Pst* 134E”) with the positions of HD, PR and STE3.2-1 loci marked by black arrow heads. HD, PR and STE3.2-1, located on chromosome 4, chromosome 9 and chromosome 1, respectively, suggest tetrapolar mating types in these four species.

(TIF)

**S2 Fig. Genetic distance between pheromone precursor genes (*mfa*) and pheromone receptor (*Pra*) genes varies from 100 bp to 100 kb.** The diagrams display the location of STE3.2-2 and STE3.2-3 and their linked *mfa* genes on chromosome 9A and 9B in *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f. sp. *tritici* (“*Pgt* 21-0”), *P. triticina* (“*Pt* 76”) and *P. striiformis* f. sp. *tritici* (“*Pst* 134E”). The grey bar represents chromosome subsections containing the genes of interest with the numbers indicating the absolute location in mega base pairs on chromosome 9A and 9B. The genetic distance between *mfa1/mfa3* and STE3.2-3 are highly variable between species, whereas STE3.2-2 and *mfa2* are tightly linked in all species.

(TIF)

**S3 Fig. MFA protein alignments indicate mating factor a proteins are overall conserved within species.** (A) Alignment of MFA1 protein sequences from four different rust fungal species. The MFA1, which is linked to STE3.2-3, is mostly conserved within species. Two near identical MFA1/3 copies can be identified in four *P. triticina* isolates. In *P. graminis* f. sp. *tritici* MFA1 has four amino acid substitutions in *Pgt* SCCL versus *Pgt* 21-0. (B) Alignment of MFA2 protein sequences from four different rust fungal species. MFA2 is fully conserved within each species. (C) Alignment of MFA3 protein sequences from three *P. graminis* f. sp. *tritici* isolates. *Mfa3* is only present *P. graminis* f. sp. *tritici* in close proximity to STE3.2-3. Predicted mature pheromone sequences are outlined by boxes.

(TIF)

**S4 Fig. Pheromone precursor genes group according to their proximate *Pra* alleles and display strong signals of trans-specific polymorphisms.** Maximum likelihood tree of *mfas* identified in four cereal rust fungi including multiple isolates per species. Tips are labelled with the species abbreviation, gene names, and isolate names are provided in parentheses. Branch support was assessed by 10000 replicates. The scale bar represents 0.5 substitutions per site. *Pca*—*P. coronata* f. sp. *avenae*, *Pgt*—*Puccinia graminis* f. sp. *tritici*, *Pt*—*P. triticina* and *Pst*—*P. striiformis* f. sp. *tritici*.

(TIF)

**S5 Fig. Whole chromosome alignments of HD loci containing chromosomes between the two haplotypes of each dikaryotic genome exhibit slightly reduced synteny around HD locus.** The figure shows dot plots of whole chromosome alignments between the two HD loci containing chromosomes from dikaryotic genome assemblies. Each panel consists of dot plots of the whole chromosome and subset dot plot zooming into the HD locus. The HD locus is labelled and line colors show the nucleotide percentage identity and nucleotide orientation as indicated in the figure legend. Subfigure A to D show *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f. sp. *tritici* (“*Pgt* 21-0”), *P. triticina* (“*Pt* 76”) and *P. striiformis* f. sp. *tritici* (“*Pst* 134E”), respectively.

(TIF)

**S6 Fig. Allele pairs on chromosomes 4 and 9 are not overall more diverged when compared to other chromosomes.** The plots show the distribution of  $d_S$  values of allele pairs on sister chromosomes in four cereal rust fungal species. Bar plots show the distribution of  $d_S$  values up to the 99% quantile. Red lines represent threshold of  $d_S$  values of 95% of all alleles. Blue lines represent threshold of  $d_S$  values of 90% of all alleles. Black points show  $d_S$  values of individual allele pairs. *HD* and *STE3* allele pairs are highlighted as red points. Subfigures A to D show *P. coronata* f. sp. *avenae* (“*Pca 203*”), *P. graminis* f. sp. *tritici* (“*Pgt 21-0*”), *P. triticina* (“*Pt 76*”) and *P. striiformis* f. sp. *tritici* (“*Pst 134E*”), respectively.  
(TIF)

**S7 Fig. Hi-C heatmap of *P. coronata* f. sp. *avenae* (“*Pca 203*”) hapA.**  
(TIF)

**S8 Fig. Hi-C heatmap of *P. coronata* f. sp. *avenae* (“*Pca 203*”) hapB.**  
(TIF)

**S9 Fig. Hi-C heatmap of *P. triticina* (“*Pt 76*”) hapA.**  
(TIF)

**S10 Fig. Hi-C heatmap of *P. triticina* (“*Pt 76*”) hapB.**  
(TIF)

**S11 Fig. Coverage of different transposable element orders at the HD, PR, and STE3.2–1 locus.** The plots show the percentage of nucleotides covered by different transposable element orders at (A) HD locus (B) PR locus (C) STE3.2–1 locus. Each subfigure A to C shows the coverage in each haplotype of the dikaryotic genomes of *P. coronata* f. sp. *avenae* (“*Pca 203*”), *P. graminis* f. sp. *tritici* (“*Pgt 21-0*”), *P. triticina* (“*Pt 76*”) and *P. striiformis* f. sp. *tritici* (“*Pst 134E*”). Different TE orders are color coded as shown in the legend. TEs with no assigned class are labelled “Undetermined”. TEs with no assigned order but belonging to Class I (RNA retrotransposons) or Class II (DNA transposons) are labelled “Undetermined Class I” or “Undetermined Class II”, respectively.  
(TIF)

**S12 Fig. Whole chromosome alignments of PR loci containing chromosomes between two haplotypes of each dikaryotic genome exhibit strong signs of synteny loss.** The figure shows dot plots of whole chromosome alignments between the two PR loci containing chromosomes from the dikaryotic genome assemblies. Each panel consists of dot plots of the whole chromosome and subset dot plots zooming into the PR locus. The PR locus is labelled and line colors show the nucleotide percentage identity and nucleotide orientation as indicated in the figure legend. Subfigures A to D show *P. coronata* f. sp. *avenae* (“*Pca 203*”), *P. graminis* f. sp. *tritici* (“*Pgt 21-0*”), *P. triticina* (“*Pt 76*”) and *P. striiformis* f. sp. *tritici* (“*Pst 134E*”), respectively.  
(TIF)

**S13 Fig. IGV screen shots of short-read Illumina mapping of various *P. coronata* f. sp. *avenae* isolates against the HD locus in the *Pca 203* reference.** (A) shows mapping against the HD locus on chromosome 4A and (B) chromosome 4B, respectively.  
(TIF)

**S14 Fig. IGV screen shot of short-read Illumina mapping of various *P. graminis* f. sp. *tritici* isolates against the HD locus in the *Pgt 21-0* reference.** (A) shows mapping against the HD locus on chromosome 4A and (B) chromosome 4B, respectively.  
(TIF)

**S15 Fig. IGV screen shots of short-read Illumina mapping of various *P. triticina* against the HD isolates locus in the *Pt 76* reference.** (A) shows mapping against the HD locus on chromosome 4A and (B) chromosome 4B, respectively.

(TIF)

**S16 Fig. IGV screen shots of short-read Illumina mapping of various *P. striiformis* f. sp. *tritici* isolates against the HD locus in the *Pst 134E* reference.** (A) shows mapping against the HD locus on chromosome 4A and (B) chromosome 4B, respectively.

(TIF)

**S17 Fig. Nucleotide alignments of the HD gene coding sequence in four cereal rust fungi indicate *bW-HD1* and *bE-HD2* are multiallelic in each species.** (A) Nucleotide alignment of the *de novo* reconstructed HD locus from *Pca 203* Illumina short-read data with the coding regions of *bW-HD1* and *bE-HD2* alleles from *Pca 203* dikaryotic reference genome. (B) to (D) multiple sequence alignments of *de novo* reconstructed HD coding regions, (E) multiple sequence alignments of *Pst bW-HD1* and *bE-HD2* alleles [159]. In each subfigure B to E, the top two track shows the consensus sequence length and relative sequence identity, respectively. Subfigure B to E show *P. coronata* f. sp. *avenae* (“*Pca*”), *P. graminis* f. sp. *tritici* (“*Pgt*”), *P. triticina* (“*Pt*”) and *P. striiformis* f. sp. *tritici* (“*Pst*”), respectively. The *bW-HD1* and *bE-HD2* are numbered in accordance with Fig 1.

(TIF)

**S18 Fig. IGV screen shots of short-read Illumina mapping of various *P. coronata* f. sp. *avenae* isolates against the PR locus in the *Pca 203* reference.** (A) shows mapping against the PR locus on chromosome 9A and (B) chromosome 9B, respectively.

(TIF)

**S19 Fig. IGV screen shot of short-read Illumina mapping of various *P. graminis* f. sp. *tritici* isolates against the PR locus in the *Pgt 21-0* reference.** (A) shows mapping against the PR locus on chromosome 9A and (B) chromosome 9B, respectively.

(TIF)

**S20 Fig. IGV screen shots of short-read Illumina mapping of various *P. triticina* isolates against the PR locus in the *Pt 76* reference.** (A) shows mapping against the PR locus on chromosome 9A and (B) chromosome 9B, respectively.

(TIF)

**S21 Fig. IGV screen shots of short-read Illumina mapping of various *P. striiformis* f. sp. *tritici* isolates against the PR locus in the *Pst 134E* reference.** (A) shows mapping against the PR locus on chromosome 9A and (B) chromosome 9B, respectively.

(TIF)

**S22 Fig. Amino acid alignments of STE3.2–2 and STE3.2–3 alleles of four cereal rust fungi.** Multiple sequence alignment of *de novo* reconstructed STE3.2–2 and STE3.2–3 protein sequences. Subfigure contains MSAs; one for STE3.2–2 and one for STE3.2–3. Subfigure A to D show *P. coronata* f. sp. *avenae* (“*Pca*”), *P. graminis* f. sp. *tritici* (“*Pgt*”), *P. triticina* (“*Pt*”) and *P. striiformis* f. sp. *tritici* (“*Pst*”), respectively.

(TIF)

**S23 Fig. Amino acid alignments of the three MFA alleles from various *P. graminis* f. sp. *tritici* isolates.** Amino acid substitutions are highlighted by color, whereas predicted mature pheromone sequences are outlined by boxes.

(TIF)

**S24 Fig. Whole chromosome alignments of HD genes containing chromosomes between haplotypes of four different *P. triticina* isolates.** The figure shows dot plots of whole chromosome alignments of HD loci containing chromosomes derived from distinct dikaryotic genomes of four different *P. triticina* isolates including *Pt 15*, *Pt 19NSW04*, *Pt 20QLD87* against *Pt 76*. Each panel consists of a dot plot of the whole chromosome and a subset dot plot zooming into the HD locus. The HD locus is labelled and line colors show the nucleotide percentage identity and nucleotide orientation as indicated in the figure legend. (A) Comparison of nucleotide sequence of chromosome 4s of *Pt 15* and *Pt 76*. (B) Comparison of nucleotide sequence of chromosome 4s of *Pt 19NSW04* and *Pt 76*. (C) Comparison of nucleotide sequence of chromosome 4s of *Pt 20QLD87* and *Pt 76*. (D) Comparison of nucleotide sequence of chromosome 4s of *Pt 19NSW04* and *Pt 20QLD87*.

(TIF)

**S25 Fig. The HD locus is highly conserved in four *P. triticina* isolates.** Synteny graphs of HD loci including proximal regions in the four *P. triticina* isolates *Pt 15*, *Pt 19NSW04*, *Pt 20QLD87* and *Pt 76*. Red lines between chromosome sections represent gene pairs with nucleotide sequence identity higher than 70% and grey shades between conserved nucleotide sequences ( $> = 1000$  bp and identity  $> = 90\%$ ). For additional annotations please refer to the provide legend (“Legend”).

(TIF)

**S26 Fig. Coverage of different transposable element orders at the HD locus in four *P. triticina* isolates.** The plots show the percentage of nucleotides covered by different transposable element orders at the HD locus. Each subfigure shows the coverage in each haplotype of the dikaryotic genomes of *P. triticina* isolates *Pt 15*, *Pt 19NSW04*, *Pt 20QLD87* and *Pt 76*. Different TE orders are color coded as shown in the legend. TEs with no assigned class are labelled “Undetermined”. TEs with no assigned order but belonging to Class I (RNA retrotransposons) or Class II (DNA transposons) are labelled “Undetermined Class I” or “Undetermined Class II”, respectively.

(TIF)

**S27 Fig. Whole chromosome alignments of *Pra* gene containing chromosomes between haplotypes of four different *P. triticina* isolates.** The figure shows dot plots of whole chromosome alignments of *STE3.2-2* or *STE3.2-3* containing chromosomes derived from distinct dikaryotic genomes of four different *P. triticina* isolates including *Pt 15*, *Pt 19NSW04*, *Pt 20QLD87* against *Pt 76*. Each panel consists of a dot plot of the whole chromosome and a subset dot plot zooming into the proximal region of *STE3.2-2* or *STE3.2-3*. The *STE3.2-2* or *STE3.2-* are labelled and line colors show the nucleotide percentage identity and nucleotide orientation as indicated in the figure legend. (A) Comparison of nucleotide sequence of chromosome 9s containing *STE3.2-2* gene (B) Comparison of nucleotide sequence of chromosome 9s containing *STE3.2-3* gene.

(TIF)

**S28 Fig. Nucleotide coverage distribution of Ty3\_Pt\_Ste3.2-3 on the chromosomes of four *P. triticina* isolates.** The plots show the percentage of nucleotides covered by the transposable element family Ty3\_Pt\_Ste3.2-3 on each chromosome of four *P. triticina* isolates *Pt 15* (A), *Pt 19NSW04* (B), *Pt 20QLD87* (C) and *Pt 76* (D). Chromosomes carrying *STE3.2-3* are highlighted in red.

(TIF)

**S29 Fig. Schematic illustration of the location of Ty3\_Pt\_STE3.2–3 copies at the STE3.2–3 locus in four *P. triticina* isolates.**

(TIF)

**S30 Fig. Genealogy of STE3.2–1 in rust fungi indicates sequence of STE3.2–1 are conserved within species.** Bayesian rooted gene tree built from STE3.2–1 coding-based sequence alignment from four cereal rust fungi: *P. coronata* f. sp. *avenae* (*Pca*), *P. graminis* f. sp. *tritici* (*Pgt*), *P. triticina* (*Pt*) and *P. striiformis* f. sp. *tritici* (*Pst*) and *P. polysora* f. sp. *zeae* (*Ppz*) GD1913 was included as outgroup. Trees are based on a TN93+I model of molecular evolution. Each node is labelled with its values of posterior probability (PP). PP values above 0.95 are considered to have strong evidence for monophyly of a clade and PP values of identical alleles are not displayed. The scale bar represents the number of nucleotide substitutions per site. Alleles of the same species are colored with identical background: *Pca* (yellow), *Pgt* (green), *Pt* (blue), *Pst* (orange).

(TIF)

**S31 Fig. Whole chromosome alignments of STE3.2–1 gene containing chromosomes between two haplotypes of each dikaryotic genome suggest high conservation of STE3.2–1.** The figure shows dot plots of whole chromosome alignments between the two STE3.2–1 gene containing chromosomes from dikaryotic genome assemblies. Each panel consists of dot plots of the whole chromosome and subset dot plots zooming into the STE3.2–1 proximal region. The position of STE3.2–1 gene is labelled, and line colors show the nucleotide percentage identity and nucleotide orientation as indicated in the figure legend. Subfigures A to D show *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f. sp. *tritici* (“*Pgt* 21–0”), *P. triticina* (“*Pt* 76”) and *P. striiformis* f. sp. *tritici* (“*Pst* 134E”), respectively.

(TIF)

**S32 Fig. Synteny analysis of STE3.2–1 genes and their flanking regions reveal no evidence of genetic degeneration.** Synteny graphs of the STE3.2–1 locus including proximal regions in *P. coronata* f. sp. *avenae* (“*Pca* 203”), *P. graminis* f. sp. *tritici* (“*Pgt* 21–0”), *P. triticina* (“*Pt* 76”), and *P. striiformis* f. sp. *tritici* (“*Pst* 134E”). Proximal regions are defined as 40 genes downstream and upstream of the STE3.2–1 alleles, respectively. STE3.2–1 proximal regions are highly syntenic within each dikaryotic genome and conserved between species. Red lines between chromosome sections represent gene pairs with sequence identity higher than 70% and grey shades represent conserved nucleotide sequences (> = 1000 bp and identity > = 90%). For additional annotations please refer to the provided legend (“Legend”).

(TIF)

**S33 Fig. Synonymous divergence ( $d_s$ ) values between STE3.2–1 alleles along chromosome 1A suggest conservation of this locus.** Synonymous divergence values ( $d_s$ ) for all allele pairs are plotted along chromosome 1A for (A) *P. coronata* f. sp. *avenae* (“*Pca* 203”), (B) *P. graminis* f. sp. *tritici* (“*Pgt* 21–0”), (C) *P. triticina* (“*Pt* 76”), and (D) *P. striiformis* f. sp. *tritici* (“*Pst* 134E”). In each panel, the top track shows the  $d_s$  values (“ $d_s$ ”) of allele pairs along chromosome 1. Each dot corresponds to the  $d_s$  value of a single allele pair. The second and third track show the averaged TE (“TE”) and gene (“gene”) density along chromosome 1 in 10 kbp-sized windows, respectively. The STE3.2–1 alleles are highlighted with a red line and red shading indicates a 0.4 mbp-sized window around the STE3.2–1 genes. Predicted centromeric regions are marked with blue shading. The two lower tracks ( $d_s$  values and gene locations) provide a detailed zoomed in view of red shaded area around the STE3.2–1 alleles. Species-specific background coloring is the same as for Fig 1.

(TIF)

**S34 Fig. Multidimensional scaling (MDS) plot of RNA dataset used in this study.** MDS plots were made with TMM normalized counts for quality control, each dot represents a single sample, and replicates were color coded as indicated in each subfigure legend. (A) MDS plot of TMM-normalized value of *Pca* at 48 hour post infection (hpi) and 120 hpi. (B) MDS plot of TMM-normalized value of *Pgt* at 48, 72, 96, 120, 148 and 168 hpi. (C) MDS plot of TMM-normalized value of *Pst* at 24, 48, 72, 120, 168, 216 and 264 hpi. (D) MDS plot of TMM-normalized value of ungerminated spore (US), germinated spores (GS) stages, 144 hpi, 216 hpi and haustoria enriched samples (HE) of *Pst*. (TIF)

**S35 Fig. Expression of housekeeping genes in four RNAseq datasets.** (A) TMM-normalized value of housekeeping genes in *P. coronata* f. sp. *avenae* (“*Pca 12NC29*”), *P. graminis* f. sp. *tritici* (“*Pgt 21-0*”), and *P. striiformis* f. sp. *tritici* (“*Pst 87/66*” and “*Pst 104E*”). (B) Likelihood ratio test (LRT) method was applied to test significant upregulation of housekeeping genes between timepoints, none of the housekeeping genes show significant upregulation between timepoints. red dashed lines indicate  $\log_{2}FC = 0.5$  and  $\log_{2}FC = -0.5$  respectively, stars above or below bars indicate statistically significant differences between the two adjacent time points: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Genes are labelled with different colors. Transcription elongation factor TFIIS (*TFIIS*), Actin-related protein 3 (*ARP3*), Actin/actin-like protein 2 (*ACTIN2*), Ubiquitin carboxyl-terminal hydrolase 6 (*UBP6*), Conserved oligomeric Golgi complex subunit 3 (*COG3*), Ctr copper transporter 2 (*CTR2*). (TIF)

**S36 Fig. Upregulation of MAT genes in late stages of the asexual life cycle.** Likelihood ratio test (LRT) method was applied to test significant upregulation of *MAT* genes between timepoints, red dashed lines indicate  $\log_{2}FC = 0.5$  and  $\log_{2}FC = -0.5$  respectively, stars above or below bars indicate statistically significant differences between the two adjacent time points: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . *MAT* genes were labelled with different colors. (A) The expression levels of *MAT* genes in *P. coronata* f. sp. *avenae* (“*Pca 12NC29*”) were compared between 120 hours post infection (hpi) and 48 hpi. (B) The expression levels of *MAT* genes in *P. graminis* f. sp. *tritici* (“*Pgt 21-0*”) were compared between 72 hpi and 48 hpi, 96 hpi and 72 hpi, 120 hpi and 96 hpi, 144 hpi and 120 hpi, 168 hpi and 144 hpi. (C) The expression levels of *MAT* genes in *P. striiformis* f. sp. *tritici* (“*Pst 87/66*”) were compared between 48 hpi and 24 hpi, 72 hpi and 48 hpi, 120 hpi and 72 hpi, 168 hpi and 120 hpi, 216 hpi and 168 hpi, 264 hpi and 216 hpi. (D) The expression levels of *MAT* genes in *P. striiformis* f. sp. *tritici* (“*Pst 104E*”) were compared between germinated spores (GS) and ungerminated spores (US), 144 hpi and GS, 216 hpi and 144 hpi, haustoria enriched samples (HE) and 216 hpi. (TIF)

**S37 Fig. MAT genes are upregulated in the late asexual infection stage of *P. striiformis* f. sp. *tritici* (“*Pst 104E*”).** TMM-normalized value of *MAT* genes in ungerminated spores (US), germinated spores (GS) stages, 144 hpi, 216 hpi and in haustoria enriched samples (HE) of *Pst*. Genes are labelled with different colours. (TIF)

## Acknowledgments

We thank Dr. P. Tobias and Dr. M. Moeller kindly for critical reading and comments on this manuscript. We thank Dr. M. Bui for advice in iqtrees2 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*. We thank J. Lin for suggestions on data visualization. This work was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the ANU Merit Allocation Scheme.

## Author Contributions

**Conceptualization:** Benjamin Schwessinger.

**Data curation:** Zhenyan Luo.

**Formal analysis:** Zhenyan Luo, Alistair McTaggart, Benjamin Schwessinger.

**Funding acquisition:** Benjamin Schwessinger.

**Investigation:** Zhenyan Luo.

**Methodology:** Zhenyan Luo, Benjamin Schwessinger.

**Project administration:** Benjamin Schwessinger.

**Resources:** Benjamin Schwessinger.

**Software:** Zhenyan Luo, Benjamin Schwessinger.

**Supervision:** Benjamin Schwessinger.

**Validation:** Zhenyan Luo.

**Visualization:** Zhenyan Luo, Alistair McTaggart.

**Writing – original draft:** Zhenyan Luo, Alistair McTaggart, Benjamin Schwessinger.

**Writing – review & editing:** Zhenyan Luo, Alistair McTaggart, Benjamin Schwessinger.

## References

1. Bachtrog D. A dynamic view of sex chromosome evolution. *Current opinion in genetics & development*. 2006; 16(6):578–85. <https://doi.org/10.1016/j.gde.2006.10.007> PMID: 17055249
2. Jay P, Tezenas E, Véber A, Giraud T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLoS Biology*. 2022; 20(7):e3001698. <https://doi.org/10.1371/journal.pbio.3001698> PMID: 35853091
3. Duhamel M, Carpentier F, Begerow D, Hood ME, Rodríguez de la Vega RC, Giraud T. Onset and stepwise extensions of recombination suppression are common in mating-type chromosomes of *Microbotryum anther-smut* fungi. *Journal of Evolutionary Biology*. 2022.
4. Carpentier F, Rodríguez de la Vega RC, Jay P, Duhamel M, Shykoff JA, Perlin MH, et al. Tempo of degeneration across independently evolved nonrecombining regions. *Molecular Biology and Evolution*. 2022; 39(4):msac060. <https://doi.org/10.1093/molbev/msac060> PMID: 35325190
5. Idnurm A, Hood ME, Johannesson H, Giraud T. Contrasted patterns in mating-type chromosomes in fungi: hotspots versus coldspots of recombination. *Fungal Biology Reviews*. 2015; 29(3–4):220–9. <https://doi.org/10.1016/j.fbr.2015.06.001> PMID: 26688691
6. Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, et al. Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. *New Phytologist*. 2021; 229(5):2470–91. <https://doi.org/10.1111/nph.17039> PMID: 33113229
7. Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). *G3: Genes, Genomes, Genetics*. 2019; 9(6):1987–98.
8. Branco S, Badouin H, de la Vega RCR, Gouzy J, Carpentier F, Aguilera G, et al. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proceedings of the National Academy of Sciences*. 2017; 114(27):7067–72.

9. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. Fungal sex: the Basidiomycota. The fungal kingdom. 2017;147–75. <https://doi.org/10.1128/microbiolspec.FUNK-0046-2016> PMID: 28597825
10. Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. PLoS Genetics. 2008; 4(3):e1000030.
11. Hartmann FE, Ament-Velásquez SL, Vogan AA, Gautier V, Le Prieur S, Berramdane M, et al. Size variation of the nonrecombining region on the mating-type chromosomes in the fungal *Podospora anserina* species complex. Molecular biology and evolution. 2021; 38(6):2475–92.
12. Vittorelli N, Rodríguez de la Vega RC, Snirc A, Levert E, Gautier V, Lalanne C, et al. Stepwise recombination suppression around the mating-type locus in an ascomycete fungus with self-fertile spores. PLoS Genetics. 2023; 19(2):e1010347. <https://doi.org/10.1371/journal.pgen.1010347> PMID: 36763677
13. Jalalzadeh B, Barroso G, Savoie J-M, Callac P. Experimental Outcrossing in *Agaricus bisporus* Revealed a Major and Unexpected Involvement of Airborne Mycelium Fragments. Journal of Fungi. 2022; 8(12):1278. <https://doi.org/10.3390/jof8121278> PMID: 36547611
14. Liang S-W, Huang Y-H, Chiu J-Y, Tseng H-W, Huang J-H, Shen W-C. The smut fungus *Ustilago esculenta* has a bipolar mating system with three idiomorphs larger than 500 kb. Fungal Genetics and Biology. 2019; 126:61–74.
15. Carpentier F, de la Vega RCR, Branco S, Snirc A, Coelho MA, Hood ME, et al. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. Genome research. 2019; 29(6):944–53. <https://doi.org/10.1101/gr.242578.118> PMID: 31043437
16. Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, et al. Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. Eukaryotic Cell. 2002; 1(5):704–18.
17. Sun S, Coelho MA, Heitman J, Nowrousian M. Convergent evolution of linked mating-type loci in basidiomycete fungi. PLoS genetics. 2019; 15(9):e1008365. <https://doi.org/10.1371/journal.pgen.1008365> PMID: 31490920
18. Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, et al. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS biology. 2004; 2(12):e384. <https://doi.org/10.1371/journal.pbio.0020384> PMID: 15538538
19. Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, et al. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. G3: Genes, Genomes, Genetics. 2017; 7(2):361–76.
20. Wu JQ, Song L, Ding Y, Dong C, Hasan M, Park RF. 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. 2021:2180.
21. Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R. *Ustilago maydis* as a pathogen. Annual review of phytopathology. 2009; 47:423–45.
22. Maia TM, Lopes ST, Almeida JMGC, Rosa LH, Sampaio JP, Gonçalves P, et al. Evolution of Mating Systems in Basidiomycetes and the Genetic Architecture Underlying Mating-Type Determination in the Yeast *Leucosporidium scottii*. Genetics. 2015; 201(1):75–89.
23. Holden S, Bakkeren G, Hubensky J, Bamrah R, Abbasi M, Qutob D, et al. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. BMC Biology. 2023; 21(1):233. <https://doi.org/10.1186/s12915-023-01717-9> PMID: 37880702
24. Badouin H, Hood ME, Gouzy J, Aguilera G, Siguenza S, Perlin MH, et al. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. Genetics. 2015; 200(4):1275–84.
25. Petit E, Giraud T, de Vienne DM, Coelho MA, Aguilera G, Amselem J, et al. LINKAGE TO THE MATING-TYPE LOCUS ACROSS THE GENUS MICROBOTRYUM: INSIGHTS INTO NONRECOMBINATING CHROMOSOMES. Evolution. 2012; 66(11):3519–33. <https://doi.org/10.1111/j.1558-5646.2012.01703.x> PMID: 23106715
26. Coelho MA, Ianiri G, David-Palma M, Theelen B, Goyal R, Narayanan A, et al. Frequent transitions in mating-type locus chromosomal organization in *Malassezia* and early steps in sexual reproduction. Proceedings of the National Academy of Sciences. 2023; 120(32):e2305094120.
27. Aime M, McTaggart A. A higher-rank classification for rust fungi, with notes on genera. Fungal systematics and evolution. 2021; 7(1):21–47. <https://doi.org/10.3114/fuse.2021.07.02> PMID: 34124616
28. Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. Molecular plant pathology. 2012; 13(4):414–30. <https://doi.org/10.1111/j.1364-3703.2011.00783.x> PMID: 22471698

29. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. *Nature ecology & evolution*. 2019; 3(3):430–9. <https://doi.org/10.1038/s41559-018-0793-y> PMID: 30718852
30. Buller AHR. *Researches on Fungi, Vol. VII: The Sexual Process in the Uredinales*: University of Toronto Press; 1950.
31. Talhinhos P, Carvalho R, Tavares S, Ribeiro T, Azinheira H, Ramos AP, et al. Diploid Nuclei Occur throughout the Life Cycles of Pucciniales Fungi. *Microbiology Spectrum*. 2023:e01532–23. <https://doi.org/10.1128/spectrum.01532-23> PMID: 37289058
32. Narisawa K, Yamaoka Y, Katsuya K. Mating type of isolates derived from the spermogonial state of *Puccinia coronata* var. *coronata*. *Mycoscience*. 1994; 35(2):131–5.
33. Nemri A, Saunders DG, Anderson C, Upadhyaya NM, Win J, Lawrence G, et al. The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Frontiers in plant science*. 2014; 5:98.
34. Ferrarezi J, McTaggart A, Tobias P, Hayashibara C, Degnan R, Shuey L, et al. *Austropuccinia psidii* uses tetrapolar mating and produces meiotic spores in older infections on *Eucalyptus grandis*. *Fungal Genetics and Biology*. 2022:103692.
35. Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, et al. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nature communications*. 2019; 10(1):1–15.
36. Devier B, Aguileta G, Hood ME, Giraud T. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics*. 2009; 181(1):209–23. <https://doi.org/10.1534/genetics.108.093708> PMID: 19001292
37. Sperschneider J, Hewitt T, Lewis DC, Periyannan S, Milgate AW, Hickey LT, et al. Nuclear exchange generates population diversity in the wheat leaf rust pathogen *Puccinia triticina*. *Nature Microbiology*. 2023:1–12.
38. Kronstad J, Leong S. The b mating-type locus of *Ustilago maydis* contains variable and constant regions. *Genes & development*. 1990; 4(8):1384–95.
39. Casselton LA, Olesnicky NS. Molecular genetics of mating recognition in basidiomycete fungi. *Microbiology and molecular biology reviews*. 1998; 62(1):55–70. <https://doi.org/10.1128/MMBR.62.1.55-70.1998> PMID: 9529887
40. Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, Martin F, et al. *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular plant-microbe interactions*. 2011; 24(7):808–18.
41. Hood ME, Antonovics J, Koskella B. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics*. 2004; 168(1):141–6.
42. Duhamel M, Hood ME, Rodríguez de la Vega RC, Giraud T. Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi. *Nature Communications*. 2023; 14(1):5692.
43. Miller ME, Zhang Y, Omidvar V, Sperschneider J, Schwessinger B, Raley C, et al. 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*. 2018; 9(1):e01650–17.
44. Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, et al. A chromosome-level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp. *avenae*: a resource to enable comparative genomics in the cereal rusts. *G3*. 2022; 12(8):jkac149.
45. Miller ME, Nazareno ES, Rottschaefer SM, Riddle J, Dos Santos Pereira D, Li F, et al. Increased virulence of *Puccinia coronata* f. sp. *avenae* populations through allele frequency changes at multiple putative Avr loci. *PLoS genetics*. 2020; 16(12):e1009291.
46. Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, et al. Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50* resistance in wheat. *Science*. 2017; 358(6370):1607–10.
47. Lewis CM, Persoons A, Bebbler DP, Kigathi RN, Maintz J, Findlay K, et al. Potential for re-emergence of wheat stem rust in the United Kingdom. *Communications biology*. 2018; 1(1):1–9. <https://doi.org/10.1038/s42003-018-0013-y> PMID: 30271900
48. Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, et al. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome biology*. 2022; 23(1):84. <https://doi.org/10.1186/s13059-022-02658-2> PMID: 35337367

49. Wu JQ, Dong C, Song L, Park RF. 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*. 2020; 11:521.
50. Wu JQ, Sakthikumar S, Dong C, Zhang P, Cuomo CA, Park RF. 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*. 2017; 8:148.
51. Fellers JP, Sakthikumar S, He F, McRell K, Bakkeren G, Cuomo CA, et al. Whole-genome sequencing of multiple isolates of *Puccinia triticina* reveals asexual lineages evolving by recurrent mutations. *G3*. 2021; 11(9):jkab219. <https://doi.org/10.1093/g3journal/jkab219> PMID: 34544127
52. Schwessinger B, Jones A, Albekaa M, Hu Y, Mackenzie A, Tam R, et al. A Chromosome Scale Assembly of an Australian *Puccinia striiformis* f. sp. tritici Isolate of the PstS1 Lineage. *Molecular Plant-Microbe Interactions*. 2022; 35(3):293–6.
53. Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM, et al. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. tritici reveals high interhaplotype diversity. *MBio*. 2018; 9(1):e02275–17.
54. Schwessinger B, Chen Y-J, Tien R, Vogt JK, Sperschneider J, Nagar R, et al. Distinct life histories impact dikaryotic genome evolution in the rust fungus *Puccinia striiformis* causing stripe rust in wheat. *Genome biology and evolution*. 2020; 12(5):597–617.
55. van Schalkwyk HJ, Adams T, Persoons A, Boshoff WH, Wanyera R, Hovmøller MS, et al. Pathogenic analyses of *Puccinia striiformis* f. sp. tritici supports a close genetic relationship between South and East Africa. *Plant Pathology*. 2022; 71(2):279–88.
56. Hubbard A, Lewis CM, Yoshida K, Ramirez-Gonzalez RH, de Vallavieille-Pope C, Thomas J, et al. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome biology*. 2015; 16(1):1–15.
57. Ding Y, Cuddy WS, Wellings CR, Zhang P, Thach T, Hovmøller MS, et al. Incursions of divergent genotypes, evolution of virulence and host jumps shape a continental clonal population of the stripe rust pathogen *Puccinia striiformis*. *Molecular Ecology*. 2021; 30(24):6566–84.
58. Upadhyaya NM, Garnica DP, Karaoglu H, Sperschneider J, Nemri A, Xu B, et al. 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*. 2015; 5:759.
59. Li C, Qiao L, Lu Y, Xing G, Wang X, Zhang G, et al. Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome Evolution in Pucciniales. *Microbiology Spectrum*. 2023; 11(1):e02828–22.
60. Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM, et al. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. tritici reveals high interhaplotype diversity. *MBio*. 2018; 9(1).
61. Persoons A, Morin E, Delaruelle C, Payen T, Halkett F, Frey P, et al. Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in plant science*. 2014; 5:450.
62. Raudaskoski M, Kothe E. Basidiomycete mating type genes and pheromone signaling. *Eukaryotic cell*. 2010; 9(6):847. <https://doi.org/10.1128/EC.00319-09> PMID: 20190072
63. Xu L, Petit E, Hood M. Variation in mate-recognition pheromones of the fungal genus *Microbotryum*. *Heredity*. 2016; 116(1):44–51.
64. Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET. Causes and consequences of variability in peptide mating pheromones of ascomycete fungi. *Molecular biology and evolution*. 2011; 28(7):1987–2003. <https://doi.org/10.1093/molbev/msr022> PMID: 21252281
65. van Diepen LTA, Olson A, Ihrmark K, Stenlid J, James TY. Extensive Trans-Specific Polymorphism at the Mating Type Locus of the Root Decay Fungus *Heterobasidion*. *Molecular Biology and Evolution*. 2013; 30(10):2286–301. <https://doi.org/10.1093/molbev/mst126> PMID: 23864721
66. Aime MC, Bell CD, Wilson AW. Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology*. 2018; 89(1):143–52. <https://doi.org/10.1016/j.simyco.2018.02.002> PMID: 29910520
67. Emms DM, Kelly S. STRIDE: species tree root inference from gene duplication events. *Molecular biology and evolution*. 2017; 34(12):3267–78. <https://doi.org/10.1093/molbev/msx259> PMID: 29029342
68. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Systematic biology*. 2002; 51(3):492–508. <https://doi.org/10.1080/10635150290069913> PMID: 12079646
69. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide

- sequence datasets. *Virus Evolution*. 2021; 7(1):veaa087. <https://doi.org/10.1093/ve/veaa087> PMID: 33936774
70. Luo Z, Schwessinger B. Supporting information of genome biology and evolution of mating type loci in four cereal rust fungi. Dryad, Dataset, <https://doi.org/10.5061/dryad.w0vt4b8zm>.
  71. Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*. 2006; 444(7115):97–101. <https://doi.org/10.1038/nature05248> PMID: 17080091
  72. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS computational biology*. 2018; 14(1):e1005944. <https://doi.org/10.1371/journal.pcbi.1005944> PMID: 29373581
  73. Hood ME, Scott M, Hwang M. Breaking linkage between mating compatibility factors: Tetrapolarity in *Microbotryum*. *Evolution*. 2015; 69(10):2561–72. <https://doi.org/10.1111/evo.12765> PMID: 26339889
  74. Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong C, et al. 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*. 2021; 19(1):1–25.
  75. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems*. 2016; 3(1):99–101. <https://doi.org/10.1016/j.cels.2015.07.012> PMID: 27467250
  76. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*. 2007; 8(12):973–82. <https://doi.org/10.1038/nrg2165> PMID: 17984973
  77. Coelho MA, Sampaio JP, Goncalves P. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. *PLoS genetics*. 2010; 6(8):e1001052. <https://doi.org/10.1371/journal.pgen.1001052> PMID: 20700437
  78. Urban M, Kahmann R, Bölker M. The biallelic mating type locus of *Ustilago maydis*: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. *Molecular and General Genetics MGG*. 1996; 250(4):414–20.
  79. Wei K, Aldaimalani R, Mai D, Zinshteyn D, Satyaki P, Blumenstiel JP, et al. Rethinking the “gypsy” retrotransposon: A roadmap for community-driven reconsideration of problematic gene names. 2022.
  80. Xu Z, Wang H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research*. 2007; 35(suppl\_2):W265–W8.
  81. Liang J, Li Y, Dodds PN, Figueroa M, Sperschneider J, Han S, et al. Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust. *Molecular Ecology Resources*. 2023; 23(3):601–20.
  82. Zhan G, Guo J, Tian Y, Ji F, Bai X, Zhao J, et al. High-throughput RNA sequencing reveals differences between the transcriptomes of the five spore forms of *Puccinia striiformis* f. sp. *tritici*, the wheat stripe rust pathogen. *Stress Biology*. 2023; 3(1):29.
  83. Dobon A, Bunting DC, Cabrera-Quio LE, Uauy C, Saunders DG. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC genomics*. 2016; 17(1):1–14. <https://doi.org/10.1186/s12864-016-2684-4> PMID: 27207100
  84. Au SW, Leng X, Harper JW, Barford D. Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the *Doc1/Apc10* subunit. *Journal of molecular biology*. 2002; 316(4):955–68.
  85. Andersen GR, Nissen P, Nyborg J. Elongation factors in protein biosynthesis. *Trends in biochemical sciences*. 2003; 28(8):434–41. [https://doi.org/10.1016/S0968-0004\(03\)00162-2](https://doi.org/10.1016/S0968-0004(03)00162-2) PMID: 12932732
  86. Wirth S, Freihorst D, Krause K, Kothe E. What role might non-mating receptors play in *Schizophyllum commune*? *Journal of Fungi*. 2021; 7(5):399.
  87. Hsueh YP, Xue C, Heitman J. A constitutively active GPCR governs morphogenic transitions in *Cryptococcus neoformans*. *The EMBO journal*. 2009; 28(9):1220–33.
  88. Zambino P, Groth JV, Lukens L, Garton JR, May G. Variation at the b mating type locus of *Ustilago maydis*. *Phytopathology*. 1997; 87(12):1233–9.
  89. Silva J. Alleles at the b incompatibility locus in Polish and North American populations of *Ustilago maydis* (DC) Corda. *Physiological Plant Pathology*. 1972; 2(4):333–7.
  90. Metin B, Findley K, Heitman J. The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*: insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS Genetics*. 2010; 6(5):e1000961.
  91. Bölker M, Urban M, Kahmann R. The a mating type locus of *U. maydis* specifies cell signaling components. *Cell*. 1992; 68(3):441–50.

92. Schirawski J, Heinze B, Wagenknecht M, Kahmann R. Mating type loci of *Sporisorium reilianum*: novel pattern with three a and multiple b specificities. *Eukaryotic cell*. 2005; 4(8):1317–27.
93. Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, et al. Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. *Molecular biology and evolution*. 2015; 32(4):928–43. <https://doi.org/10.1093/molbev/msu396> PMID: 25534033
94. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*. 2011; 108(22):9166–71. <https://doi.org/10.1073/pnas.1019315108> PMID: 21536894
95. Savile DBO. Nuclear structure and behavior in species of the Uredinales. *American Journal of Botany*. 1939:585–609.
96. Park RF, Wellings CR. Somatic hybridization in the Uredinales. *Annual review of Phytopathology*. 2012; 50:219–39. <https://doi.org/10.1146/annurev-phyto-072910-095405> PMID: 22920559
97. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution*. 2021; 38(10):4647–54. <https://doi.org/10.1093/molbev/msab199> PMID: 34320186
98. Perteza G, Perteza M. GFF utilities: GffRead and GffCompare. *F1000Research*. 2020; 9. <https://doi.org/10.12688/f1000research.23297.2> PMID: 32489650
99. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic acids research*. 2008; 36(suppl\_2):W5–W9.
100. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
101. Geneious Prime 2022.0.2 [Available from: <https://www.geneious.com>].
102. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Elsevier current trends*; 2000.
103. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, et al. Rldeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science*. 2020; 6:e251. <https://doi.org/10.7717/peerj-cs.251> PMID: 33816903
104. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*. 2019; 20(1):1–14.
105. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one*. 2010; 5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490> PMID: 20224823
106. Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular biology and evolution*. 2018; 35(10):2582–4. <https://doi.org/10.1093/molbev/msy159> PMID: 30165589
107. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
108. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*. 2014; 10(4):e1003537. <https://doi.org/10.1371/journal.pcbi.1003537> PMID: 24722319
109. Bouckaert RR, Drummond AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC evolutionary biology*. 2017; 17:1–11.
110. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*. 2018; 67(5):901–4. <https://doi.org/10.1093/sysbio/syy032> PMID: 29718447
111. FigTree 2016 [Available from: <http://tree.bio.ed.ac.uk/software/figtree/>].
112. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*. 2020; 37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015> PMID: 32011700
113. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. 2017; 14(6):587–9. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
114. HURVICH CM TSAI C-L. Regression and time series model selection in small samples. *Biometrika*. 1989; 76(2):297–307.
115. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution*. 2013; 30(5):1188–95. <https://doi.org/10.1093/molbev/mst024> PMID: 23418397

116. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
117. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*. 2002; 30(14):3059–66. <https://doi.org/10.1093/nar/gkf436> PMID: 12136088
118. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*. 2000; 16(6):562–3. <https://doi.org/10.1093/bioinformatics/16.6.562> PMID: 10980155
119. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology*. 1999; 265(2):218–25. <https://doi.org/10.1006/viro.1999.0056> PMID: 10600594
120. Martin D, Posada D, Crandall K, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *Virologica Sinica*. 2005; 37(2). <https://doi.org/10.1089/aid.2005.21.98> PMID: 15665649
121. Smith JM. Analyzing the mosaic structure of genes. *Journal of molecular evolution*. 1992; 34:126–9. <https://doi.org/10.1007/BF00182389> PMID: 1556748
122. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences*. 2001; 98(24):13757–62. <https://doi.org/10.1073/pnas.241370698> PMID: 11717435
123. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 2000; 16(7):573–82. <https://doi.org/10.1093/bioinformatics/16.7.573> PMID: 11038328
124. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*. 2007; 176(2):1035–47. <https://doi.org/10.1534/genetics.106.068874> PMID: 17409078
125. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017; 356(6333):92–5. <https://doi.org/10.1126/science.aal3327> PMID: 28336562
126. Yow AG, Zhang Y, Bansal K, Eacker SM, Sullivan S, Liachko I, et al. Genome sequence of *Monilinia vaccinii-corymbosi* sheds light on mummy berry disease infection of blueberry and mating type. *G3*. 2021; 11(2):jkaa052.
127. Schultz DT, Haddock SH, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. *Nature*. 2023:1–8. <https://doi.org/10.1038/s41586-023-05936-6> PMID: 37198475
128. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics*. 2011; 12(1):1–9. <https://doi.org/10.1186/1471-2105-12-124> PMID: 21526987
129. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004; 32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
130. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 2007; 24(8):1586–91. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113
131. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
132. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 2017; 33(19):3088–90. <https://doi.org/10.1093/bioinformatics/btx346> PMID: 28575171
133. Hunter JD. Matplotlib: A 2D graphics environment. *Computing in science & engineering*. 2007; 9(03):90–5.
134. Thomas Hackl MJA. gggenomes: A Grammar of Graphics for Comparative Genomics 2022 [Available from: <https://github.com/thackl/gggenomes>.
135. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. *PLoS one*. 2011; 6(1):e16526. <https://doi.org/10.1371/journal.pone.0016526> PMID: 21304975
136. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol*. 2005; 1(2):e22. <https://doi.org/10.1371/journal.pcbi.0010022>
137. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna*. 2015; 6(1):11. <https://doi.org/10.1186/s13100-015-0041-9> PMID: 26045719
138. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. 2015.

139. Gattiker A, Gasteiger E, Bairoch A. ScanProsite: a reference implementation of a PROSITE scanning tool. *Applied bioinformatics*. 2002; 1(2):107–8. PMID: [15130850](#)
140. Sigrist CJ, De Castro E, Cerutti L, Cucho BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic acids research*. 2012; 41(D1):D344–D7. <https://doi.org/10.1093/nar/gks1067>
141. Consortium U. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*. 2019; 47(D1):D506–D15. <https://doi.org/10.1093/nar/gky1049> PMID: [30395287](#)
142. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: [24695404](#)
143. Andrews FK, Anne Segonds-Pichon, Laura Biggins, Christel Krueger, Jo Montgomery. FastQC: a quality control tool for high throughput sequence data 2010 [Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>].
144. Picard toolkit: Broad Institute; 2019 [Available from: <https://github.com/broadinstitute/picard>].
145. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](#)
146. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. 2013; 14(2):178–92. <https://doi.org/10.1093/bib/bbs017> PMID: [22517427](#)
147. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016; 32(2):292–4. <https://doi.org/10.1093/bioinformatics/btv566> PMID: [26428292](#)
148. Tange O. GNU parallel 2018: Lulu. com; 2018.
149. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907. 2012.
150. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome biology*. 2016; 17(1):1–14. <https://doi.org/10.1186/s13059-016-0974-4> PMID: [27268795](#)
151. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021> PMID: [22506599](#)
152. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*. 2017; 33(4):574–6. <https://doi.org/10.1093/bioinformatics/btw663> PMID: [27797770](#)
153. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*. 2016; 34(5):525–7. <https://doi.org/10.1038/nbt.3519> PMID: [27043002](#)
154. Jon Palmer JS. nextgenusfs/funcannotate: funcannotate v1.5.3 Zenodo2019 [Available from: <https://doi.org/10.5281/zenodo.2604804>].
155. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014; 30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031> PMID: [24451626](#)
156. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: [19910308](#)
157. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015; 4. <https://doi.org/10.12688/f1000research.7563.2> PMID: [26925227](#)
158. Wickham H. ggplot2: elegant graphics for data analysis: springer; 2016.
159. Holden S, Bakkeren G, Hubensky J, Bamrah R, Abbasi M, Qutob D, et al. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating-type alleles. *bioRxiv*. 2023:2023.03. 30.534825. <https://doi.org/10.1101/2023.03.30.534825> PMID: [37880702](#)