

Avirulence genes identified through linkage mapping and region-specific association studies in the wheat leaf rust pathogen *Puccinia triticina*

Lu Liu

Agriculture and Agri-Food Canada

Sean Formby

University of British Columbia

Sang Hu Kim

Agriculture and Agri-Food Canada

David L. Joly

Agriculture and Agri-Food Canada

Barbara Mulock

Morden Research and Development Center, Agriculture and Agri-Food Canada

Mark Lubberts

Agriculture and Agri-Food Canada

Barry Saville

Trent University

Brent McCallum

Morden Research and Development Center, Agriculture and Agri-Food Canada

Guus Bakkeren

`guus.bakkeren@agr.gc.ca`

Agriculture and Agri-Food Canada

Research Article

Keywords: *Puccinia triticina*, Avirulence, Genetic linkage map, Association mapping, Wheat leaf rust, Whole-genome sequencing

Posted Date: August 22nd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7264203/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on February 4th, 2026. See the published version at <https://doi.org/10.1186/s12864-026-12579-0>.

Abstract

Background

Wheat rust fungi can cause significant damage to wheat crops, leading to reduced yields and economic losses. To combat disease, certain plant varieties can trigger defense responses upon recognition of specific pathogen effector proteins, causing avirulence. Identifying such avirulence (*Avr*) genes is crucial for developing strategies to protect crops from devastating losses, from identifying matching resistance genes to designing diagnostic assays for monitoring pathogen populations. *Puccinia triticina* (*Pt*) causes wheat leaf rust and is an obligate biotrophic fungus, and because of its life cycle and mode of reproduction, it is difficult to study genetically.

Results

To identify *Avr* genes in *Pt*, a F_2 population of fifty-seven progeny was generated from a sexual cross of race 9 (SBDG) and race 161 (FBDJ) on the alternate host *Thalictrum speciosissimum* under controlled conditions. The population segregated for avirulent/virulent traits screened at the seedling stage on thirteen single gene resistance lines in the wheat host cultivar Thatcher background. The genomes of the parents, F_1 , and progeny were sequenced and mapped onto an assembled parental race 9 phased haplotype B genome, resulting in the generation of 21,154 high-quality SNP markers suitable for genetic mapping of the F_2 population. A genetic map composed of 61 linkage groups was obtained, containing a total of 10,923 markers, and spanning 10,730.5 centimorgans. Effector loci correlating with avirulence to specific leaf rust resistance (*Lr*) genes, *PtAvrLr14a*, *PtAvrLr11* and *PtAvrLr2a*, were mapped to chromosome 1, chromosome 3 and chromosome 4, respectively. To strengthen the identification of candidate *Avr* genes, a region-specific association study was done on a natural population of fifty-nine *Pt* isolates that were collected in Canada and whose genomes were sequenced using Illumina.

Conclusion

Significant markers and corresponding candidate effector genes were identified for these mapped *Avr* loci. The identification of these candidate genes is an essential step towards cloning *Avr* and subsequently their matching host resistance genes, and for studying the molecular mechanisms underlying pathogen-host interactions and host defense.

Introduction

Wheat rust diseases, including stem, leaf and stripe rust, pose a major threat to global wheat production. Among these, wheat leaf rust caused by the fungal pathogen *Puccinia triticina* Eriks. (*Pt*) is the most common and widely distributed [1, 2]. In North America, yield losses ranging from 5–10% are common depending on the wheat variety and growing conditions [2, 3]. Although rust diseases can be controlled by fungicides, genetic resistance is the most effective, economical and environmentally sustainable method [4]. Significant progress has been made in developing resistant wheat cultivars. However, rust fungi evolve rapidly and can often evade plant immune systems, leading to ineffective host resistance [5]. Therefore, there is great interest in understanding the molecular mechanisms underlying host-pathogen interactions, in particular as they pertain to pathogens evading recognition by host resistance genes.

The genetics of plant-pathogen interactions were initially studied by Harold Flor with the flax-rust system and described in the 'gene-for-gene' theory [6]. In this genetic concept, the interaction of specific avirulence (*Avr*) genes in the pathogen with corresponding resistance (*R*) genes in the host activates host defense responses. Various molecular mechanisms underpinning these interactions have been revealed over the last decade. *R* genes can code for plant receptors on plant cell surfaces responding to *Avr* effectors secreted by the pathogen in the apoplast, but many *R* genes encode intracellular immune receptor proteins of the nucleotide binding leucine rich repeat (NB-LRR) class [7–10] that recognize small effector proteins delivered by the pathogen into host cells to suppress host basal defenses and to assist in the establishment of

pathogen infection [11, 12]. R proteins can interact directly with effector proteins or guard other cellular targets of effectors and indirectly sense their actions [13]. Since the outcome is pathogen arrest and 'avirulence', such effectors correspond to the *Avr* genes in the genetic concept. In wheat rust fungi, three stem rust pathogen, *P. graminis* f.sp. *tritici* (*Pgt*) *Avr* genes have been successfully cloned, namely *AvrSr27* [14], *AvrSr35* [15] and *AvrSr50* [16]. Several *Avr* effector candidates for *Pt* have been reported with two, *PtAvrLr15* and *PtAvrLr21* identified [17–22]. Functional characterization of such effectors will further our understanding of host-pathogen interactions and allow for more rational deployment of resistance genes in crop resistance breeding.

One approach to identify and clone *Avr* genes is through genetic mapping. This approach has been successfully employed to clone *Avr* genes of the flax rust fungal pathogen, *Melampsora lini* [23, 24], which is an autoecious fungus. However, wheat rust fungi are heteroecious, meaning they require different host species to complete their life cycle. This makes sexual crosses more difficult to achieve. *Pt* cycles asexually on wheat but to complete its sexual cycle, requires alternate hosts, *Thalictrum* species including *Thalictrum speciosissimum* L. (common meadow rue), which is the most susceptible alternate host of *Pt* and found in southern Europe and west Asia [1, 25]. In North America, *Pt* reproduces by the continuous production of asexually produced urediniospores and the sexual cycle does not occur naturally since the native North American *Thalictrum* species have limited susceptibility to infection by *Pt* basidiospores [18]. Starting in 1968, Samborski and Dyck conducted a series of studies on the genetics of pathogenicity in *Pt*. They produced populations of progeny that segregated for avirulence on various wheat cultivars by sexually crossing *Pt* races. These investigations formed the basis for identifying the genetic background in the generation of the standard differential wheat varieties and led to the early development of several near-isogenic wheat lines specific to leaf rust [26–28]. Since then, Dr. Peter Dyck produced additional near-isogenic wheat lines, allowing for continued and expanded research on the pathogenicity of the wheat leaf rust fungus.

The rapid advancement of genome sequencing technologies and data analysis tools have accelerated *Avr* gene identification and have promoted our understanding of various aspects of rust biology [29, 30]. However, compared with most other plant pathogens, rust fungi have larger genome sizes, with assembled genomes ranging from 53 Mbp to 1.25 Gbp [31–34]. In addition, rust fungi have large numbers of up to 93% repetitive sequences and transposable elements (TEs) in their genomes, among the largest known for plant pathogens [29, 34–36]. Moreover, the biological materials easiest to increase and collect for genomics study are the urediniospores which are dikaryotic with two highly heterozygous nuclei in one cell [35], making genome assembly particularly challenging. In *Pt*, a draft genome sequence of race 1 (BBBD) first became available in 2017 [37]. This genome contained 135 Mb of DNA, with 51% TEs and repetitive elements and an estimated 14,800 genes, which has been useful as a reference genome. In recent years, the combination of long-read and Hi-C sequencing data has helped achieve chromosome-level and haplotype-phased assemblies for some rust species including *Pt* [38–43]. The generation of complete and phased genomes has proved to be important not only for identifying intrinsic differences between two nuclei, but also for elucidating the structure of rust genomes and allowing for a deeper understanding of the role of TEs in rust virulence and evolutionary processes [44].

In this study, we made use of a sexual cross between *Pt* race 9 (SBDG) and race 161 (FBDJ), an F_1 and 57 selected progeny, and recent genome sequencing and assembly techniques. Phased genomes were constructed for the parental isolates whereby race 9 haplotype B was used as a reference for read mapping. The F_1 , progeny, and a natural population of 59 isolates collected from common wheat in Canada were sequenced and compared to this *Pt* race 9 haplotype B reference genome. The objective of this study was to generate a genetic map and identify *Pt Avr* genes and lay the basis for their cloning and characterization.

Material and methods

Fungal material

For the bi-parental mapping population, the parental isolates, race 9 (SBDG) and race 161 (FBDS), were originally selected by Samborski and Dyck [45]. Race 9 was a prevalent race throughout North America prior to the 1940's, and race 161 was a

predominant race in the Pacific region in the 1960's [54]. For the natural population used for association studies, 59 isolates representing different race phenotypes were collected in Canada from different years (1998 to 2012). To diversify the population, race 9 and race 161 were also included in association studies. Isolates were increased via single pustule purification and stored at -80 °C in glass tubes sealed under vacuum.

Developing a sexual population

Samborski and Dyck purified urediniospores for race 9 and race 161 via single pustule increases. Teliospores of each race were produced and subsequently used to infect the alternate host, *T. speciosissimum*. Aeciospores, generated from the reciprocal transfer of nectar between isolated pycnia from each parental isolate, were used to infect the rust susceptible wheat cultivar 'Little Club'. F₁ hybrids produced in this manner were similarly self-fertilized to produce the four F₂ populations (43 F₂1, 63 F₂2, 39 F₂3, 76 F₂4). Among the four F₂ populations, F₂1 and F₂3 were derived from one same F₁, while F₂2 and F₂4 were derived from another shared F₁. Urediniospores from parents, F₁ and the progeny were stored at -80 °C in glass tubes sealed under vacuum.

To initiate this study, the F₂1 and F₂3 populations generated by Samborski and Dyck, F₁ and both parents were revived and viability was assessed. Glass tubes were opened and urediniospores allowed to rehydrate for 24 hr prior to heat shock treatment (suspension of the tube in a 40 °C water bath for 5–10 min). Urediniospores were resuspended in a light mineral oil (Bayol, Esso Canada) and inoculated to individual 8 day-old seedlings of rust-susceptible wheat cultivar Thatcher. Inoculated plants were incubated in dew chambers (100% RH; 20 °C) overnight prior to transfer to a greenhouse (20 ± 4 °C). At 14 days post inoculation, urediniospores from all viable isolates were collected from individual plants. Parents, F₁ and 57 isolates of the F₂ progeny resulted in infection and the production of fresh urediniospores. Depending on the amount of urediniospores produced in the initial increase of viable cultures, successive inoculations were carried out to provide sufficient urediniospores for the completion of the study.

Virulence phenotyping and genetic analysis

All of the 59 isolates collected in Canada and stored at -80 °C in glass tubes sealed under vacuum were revived and increased as described above. For virulence profiling, urediniospores from each of these 59 isolates, race 9, race 161, the F₁ and 57 F₂ progeny were inoculated to 8 day-old germplings of a set of twenty near- isogenic lines (NILs) in the Thatcher background, each harboring a single leaf rust resistance gene: *Lr1*, *Lr2a*, *Lr2c*, *Lr3*, *Lr9*, *Lr16*, *Lr24*, *Lr26*, *Lr3ka*, *Lr11*, *Lr17*, *Lr30*, *LrB*, *Lr10*, *Lr14a*, *Lr18*, *Lr3bg*, *Lr14b*, *Lr20* and *Lr28*. Twelve to 14 days after inoculation, plants were examined. Infection types were rated based on the appearance and size of the rust pustules. Hypersensitive flecks (;), small pustules with necrosis (1), and small to medium sized pustules with chlorosis (2) were considered avirulent infection types. Virulent infection types were medium and large sized pustules without chlorosis or necrosis (rating of 3 and 4, respectively). Rust isolates found to produce more than one infection type on a single isogenic line were considered to be contaminated and were purified by single pustule increases. To achieve this, a small amount of urediniospores was inoculated to the susceptible wheat cultivar 'Little Club'. Seven days after inoculation, leaves were trimmed to isolate single uredinia. Urediniospores were collected from separately spaced pustules and subsequently tested for purity on the differential NILs. Virulence ratings were repeated for each isolate at least twice. The rust infection type on each isogenic line was converted to a binary code where an avirulent rating (IT < 3) was equal to 0 and a virulent rating (IT ≥ 3) was equal to 1. The virulence profile was then converted to a 4 or 5-letter code as per convention [55].

Segregation ratios of avirulence to virulence for each host *Lr* gene were determined for the F₂ population. Genotypes were determined by comparing observed ratios with ratios expected based on Mendelian genetics. Chi-square values were calculated to determine goodness of fit for segregation ratios; $p < 0.05$ was the level for rejecting the null hypothesis that data did not fit the expected values.

DNA extraction and whole-genome sequencing

Approximately 200 mg of urediniospores from each isolate were dispensed over individual Petri plates containing 10 ml H₂O treated with Gramacidin D (0.006 mg/ml). To enhance germination, filter paper treated with the plant volatile, Nonanol, (250 µl 0.06%) (Sigma, MO) was suspended over each plate. The following day, viable urediniospores had germinated to form mats over the surface of the plates. Mats were removed, air-dried on filter paper, and lyophilized for 24 hour. Freeze-dried mats were finely ground with glass beads and incubated for 3 h in buffer (1.0 M Tris/HCl, 0.5 M EDTA, 5.0 M NaCl, 10% cetyltrimethylammonium bromide (CTAB)), containing protease K (10 mg/ml), and SDS (2%). DNA was extracted by the phenol/coloroform/isoamyl alcohol method as described by Anikster et al. [56] and precipitated. The resulting pellet was resuspended in TE buffer, treated with RNase and re-extracted with phenol/coloroform/isoamyl alcohol. The precipitated gDNA was resuspended in 50 µl sterile H₂O.

Sequencing services were performed in 2012 at the Michael Smith Genome Sciences Centre in Vancouver, BC, Canada. DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay and Quant-iT dsDNA HS Assay Kit using a Qubit fluorometer (Invitrogen). Plate-based Illumina PET adapter ligation followed. Products were purified using Ampure XP SPRI beads and then PCR-amplified with Phusion DNA Polymerase (Thermo Fisher Scientific Inc. USA) using Illumina's PE indexed primer sets, with cycle condition 30 sec at 98°C, followed by 10 cycles of 98°C for 10 sec, 65°C for 30 sec and 72°C for 30 sec, and 72°C for 5 min. The PCR products were purified using Ampure XP SPRI beads, and checked with Caliper LabChip GX for DNA samples, using the High Sensitivity Assay (PerkinElmer, Inc. USA). PCR products of desired size range were purified on a 8% PAGE and DNA quality was assessed and quantified as described above, then diluted to 8 nM. The final concentration was double checked and determined by Quant-iT dsDNA HS Assay again. Clusters were generated on a cBot and equimolar amounts from 6–7 isolates were pooled to be sequenced per lane on an Illumina HiSeq 2000 to 76 bases in paired-end reads following standard Illumina protocols.

Genomic variation analysis

Short-read sequencing data were quality-filtered and trimmed to remove adapter sequences using fastp v0.23.4 [57]. Reads were then aligned to the phased *Pt* race 9 haplotype B reference genome (S. Formby, G. Bakkeren et al., manuscript in preparation; see Data Availability) using BWA-MEM2 [58], a performance-optimized implementation of the original BWA-MEM algorithm [59]. Alignments were filtered with SAMtools v1.14 [60] to exclude unmapped, secondary, and low-quality alignments (-q 10 -F 2316), followed by coordinate sorting and indexing.

Variant calling was performed using bcftools v1.21 [61]. bcftools mpileup was run in GVCF mode (--gvcf 1) with a minimum mapping quality of 20, and annotated with allele depth (AD), strand-aware read depth (ADF, ADR), total depth (DP), and strand bias (SP) from both FORMAT and INFO fields. Variant calls were then generated using bcftools call with multiallelic calling (-m) and genotype quality output (-f GQ).

VCFtools (version 0.1.16) was used to filter samples and markers in vcf format. Isolates with more than 10% missing data were filtered out and only biallelic markers with a minimum quality of 1,000, a minimum mean depth of 10, a maximum mean depth of 60, less than 10% missing data and minor allele frequency greater than 1% were selected. For the bi-parental mapping population, GATK SelectVariants was used to further select SNPs of different types (AA/BB, AA/AB, AB/AB, AB/BB). SNP effects were annotated by SnpEff (version 4.3) [62].

Genetic mapping

After filtering samples and SNPs, 53 progeny and 310,544 SNPs were retained in the bi-parental mapping population for the genetic map construction. Prior to map construction, a Chi-squared test was performed. SNPs that did not fit the expected 1 (homozygous AA):2 (heterozygous AB):1 (homozygous BB) segregation ratio at $P > 0.05$ were excluded. In addition, only one of the identical SNPs was retained. In the end, 11,138 SNPs were used for genetic map construction by ASMap (version 1.0–4) [63] in the R program. *Avr* genes showing an A/V segregation ratio with $P > 0.05$, including *PtAvrLr1*, *PtAvrLrB*, *PtAvrLr14a*, *PtAvrLr3bg*, *PtAvrLr14b*, *PtAvrLr11*, *PtAvrLr2c*, and *PtAvrLr28*, were included in the linkage map construction. The minimum spanning tree algorithm was used to assign linkage groups, and the Kosambi mapping function was used to calculate

genetic distances in CentiMorgans (cM). Linkage groups were initially determined using a p -value = $1e-10$ and further selected by a genetic distance of 30 cM to ensure consistency with different chromosomes. Linkage groups with less than three markers were excluded. The number of observed crossovers per individual was estimated using the ASMap package. Linkages without *Avr* genes and the A/V segregation phenotypes of *Avr* genes in the F_2 population were also used for the QTL analysis. This analysis was performed in R/qtl using the EM algorithm (method = "em") and a binary model (model = "binary"), which is appropriate for binary phenotypic data. A LOD threshold of 3 was applied for all traits.

Region-specific association studies

To strengthen identification of candidate *Avr* genes, a complementary approach using region-specific association studies was performed for *PtAvrLr14a*, *PtAvrLr11*, *PtAvrLr1* and *PtAvrLr2a*, which were identified by the linkage mapping and QTL analysis. Genomic DNA-derived Illumina reads were generated and treated for each of the 59 *Pt* Canadian isolates as described above for the progeny. Sequence reads were also aligned to the *Pt* race 9 haplotype B assembly as described above. Sample TBJJK_98_MB5-2 was eliminated from the study due to the high percentage of missing data (Table S1) and 58 isolates were retained in this natural population to be used for association studies. After filtering samples, variants were called as above yielding 408,473 high-quality markers and 58 isolates were retained in this natural population to be used for GWAS. Markers on contig16, contig3 and contig13 were used for *PtAvrLr14a*, *PtAvrLr1*, and *PtAvrLr3bg* association analysis, respectively. Missing data were imputed with BEAGLE 5.1 [64]. The CMLM model [65] implemented in GAPIT (version 3) [66] was used to perform association analysis with three principal components (PCs) as covariates. The associations between SNPs and virulence phenotypes were considered significant if the p value was below the 5% significance threshold after Bonferroni correction [67]. Manhattan plots were drawn using the CMplot package (version 4.2.0) [68] in the R program. FIMO (version 5.5.8) was used to search for regulatory motifs within 50 bp upstream and downstream of significant SNPs, using the JASPAR CORE Fungi database [69].

Results

Virulence phenotyping

Samborski and Dyck originally established four F_2 populations (consisting of 43 progeny in F_{21} , 63 in F_{22} , 39 in F_{23} , 76 in F_{24}). Among the four F_2 populations, F_{21} and F_{23} were derived from one same F_1 , while F_{22} and F_{24} were derived from another shared F_1 . Segregation data of avirulence/virulence (A/V) in the four F_2 populations on different near-isogenic lines (NILs) is presented in Table S2. For this study, 57 F_2 progeny were revived from the F_{21} and F_{23} . Parental *Pt* isolates race 9, race 161, F_1 and the 57 F_2 progeny were tested on 20 NILs in the Thatcher background. All isolates were avirulent to the NILs harboring *Lr9*, *Lr16*, *Lr24*, *Lr26*, *Lr3ka* and *Lr30*, indicating that race 9 and race 161 were homozygous for avirulence on these *Lr* genes. When tested on *Lr20*, all isolates showed virulence, suggesting that the parental isolates were homozygous for virulence towards *Lr20*. As a result, it was not possible using this cross to map these loci responsible for avirulence or virulence using linkage mapping.

Race 161 was avirulent and race 9 was virulent towards *Lr1*, while the opposite was observed towards *Lr14a* (Table S2). The F_2 progeny segregated for A/V on both *Lr1* and *Lr14a*, fitting a 3:1 ratio, suggesting that avirulence towards *Lr1* and *Lr14a* was each controlled by a single dominant gene. Race 9 and race 161 were both avirulent towards *Lr11*, and a 3:1 A/V segregation ratio was observed in the F_2 progeny, indicating that avirulence towards *Lr11* was also controlled by a single dominant gene, with parental isolates each being heterozygous avirulent. Even though parental isolates had different phenotypes towards *Lr2c*, *LrB*, *Lr14b* and *Lr28* (Table S2), segregation ratios of A/V in the F_2 progeny fitted a 1:3 ratio, indicating that avirulence to these genes was each controlled by a single recessive gene.

On the NIL containing *Lr2a*, the observed A/V ratio of 7:9 in the F_2 progeny indicated that two independent recessive genes controlled avirulence. On *Lr3*, *Lr10*, *Lr17* and *Lr18*, only a small number of isolates exhibited avirulence/virulence traits different from most isolates in the population. It is very likely that these F_2 progeny did not segregate on these NILs and that

some of the isolates are contaminants (see below), or their responses to these NILs were misinterpreted. The segregation of virulence towards *Lr3bg* more closely fits an A/V ratio of 13:3 ($P = 0.21$), rather than a ratio of 3:1 ($P = 0.03$). However, considering the entire isolate population ($F_21 + F_23$), which yields more precise outcomes, the segregation better conformed to a 3:1 ratio ($P = 0.90$) (Table S2). Consequently, it can be inferred that avirulence towards *Lr3bg* is under the control of a single dominant gene. Detailed phenotypes of progeny isolates are provided in Table S3.

The fifty-nine *Pt* isolates in the natural population collected in Canada were also screened on the same 20 wheat differential NILs. Their virulence phenotypes are provided in Table S1. On the 20 isogenic lines tested, all *Pt* isolates were virulent to *Lr1* and *Lr3*. Therefore, in the association study, we were not able to identify SNP markers linked to these avirulence or virulence loci corresponding to *Lr1* and *Lr3*.

Genotyping by whole-genome sequencing

Four complete phased haplotype genomes were generated for the two parental isolates *Pt* race 9 and 161 through the combination of Hi-C and long-reads. The genome of *Pt* race 9 haplotype B selected for read mapping is of high quality and comprises 18 main contigs that correspond to the 18 predicted *Pt* chromosomes. It contains 121.3 Mb of DNA and has 24,520 annotated genes (S. Formby, G. Bakkeren et al., manuscript in preparation). To produce genome-wide molecular markers, Illumina HiSeq 2000 technology was used to generate whole genomic DNA-derived sequences of the F_1 and 57 progeny in the bi-parental mapping population from the race 9 x race 161 sexual cross, as well as of the 59 isolates in the natural population collected in Canada (see Material and methods).

For *Pt* isolates in the bi-parental mapping population, an average of 31.1 million pairs of filtered sequence reads were generated for each isolate, of which 90.2% were mapped to the parental race 9 phased 121.3 Mbp haplotype B reference genome. The mapping coverage ranged from 13.0 to 47.5x with a mean coverage of 34.1x (Table S4). Using the sequence data, 838,419 SNPs were identified, and 310,544 high-quality SNPs were kept after filtering (Fig. 1). Only SNPs that were heterozygous in the F_1 were selected, and among them, 5,866 were homozygous in each parental isolate but heterozygous between the two parents (AA/BB type), 19,880 were heterozygous in each parental isolate (AB/AB type), and 46,099 were heterozygous in only one parental isolate (AA/AB or AB/BB type).

To identify potentially contaminated isolates within the bi-parental mapping population, a Principal Component Analysis (PCA) was performed using a subset of 90K SNP markers. The results revealed that four progeny (19A, 19B, S22, S47) were clearly separated from the main group (Fig. 2). These were judged to be contaminants and eliminated from the study (Table S3).

For the 59 isolates in the natural population, an average of 28.9 million pairs of filtered sequence reads were generated for each isolate, with 93.5% of them mapped to the reference genome. The mapping coverage ranged from 22.1 to 49.3x, with a mean coverage of 35.9x (Table S4). To diversify the population, Illumina reads from race 9 and race 161 were also included in the population and a total of 1,210,015 SNPs were identified. After filtering, 408,473 high-quality SNPs were kept and used for the association studies.

Genetic linkage mapping

After removing the four contaminated isolates, a total of 21,135 filtered and robust SNPs with a segregation ratio of $P > 0.05$ were selected for constructing a genetic map. Only one of redundant SNPs (SNPs that show the same segregation) was kept and inadequate linkage groups with less than three SNPs were removed. A final genetic map was created revealing 61 linkage groups, comprising a total of 10,923 SNPs and spanning 10,730.5 cM (Fig. 1). The individual linkage groups ranged in size from 5.7 to 490.5 cM, and the average genetic distance between SNPs was 1.0 cM throughout the genome (Table 2). Comparison between linkage groups and the reference haplotype B genome showed one chromosome could be represented by multiple linkage groups (Fig. 1, Table 2). Detailed genetic map information is listed in Table S5.

Table 2
Linkage groups in the *Puccinia triticina* genetic map

Linkage Group (LG)	Number of Markers	Genetic Distance (cM)	Mean distance between markers (cM)	Mean recombination rate (cM/10 kb)	Chromosome
LG1	413	490.5	1.2	1.3	chr1
LG2	477	455.1	1.0	1.0	chr10
LG3	559	428.2	0.8	1.0	chr12
LG4	465	418.0	0.9	1.1	chr3
LG5	482	380.3	0.8	1.4	chr6
LG6	289	360.9	1.2	1.3	chr7
LG7	379	354.1	0.9	0.5	chr9
LG8	475	345.8	0.7	1.1	chr4
LG9	320	342.7	1.1	0.8	chr13
LG10	471	329.8	0.7	0.7	chr4
LG11	210	324.8	1.5	0.7	chr11
LG12	218	312.6	1.4	0.6	chr5
LG13	291	309.4	1.1	0.8	chr8
LG14	445	309.3	0.7	0.8	chr2
LG15	364	289.3	0.8	1.5	chr10
LG16	173	276.3	1.6	0.6	chr17
LG17	317	266.3	0.8	0.4	chr1
LG18	118	245.5	2.1	0.8	chr1
LG19	186	228.9	1.2	0.5	chr9
LG20	271	224.5	0.8	1.1	chr7
LG21	159	217.9	1.4	0.9	chr5
LG22	186	212.1	1.1	0.7	chr15
LG23	196	203.2	1.0	0.6	chr13
LG24	274	188.9	0.7	0.9	chr14
LG25	253	186.1	0.7	1.1	chr15
LG26	108	176.7	1.6	0.6	chr6
LG27	263	175.0	0.7	1.0	chr4
LG28	142	166.7	1.2	0.6	chr15
LG29	228	159.7	0.7	0.9	chr18
LG30	179	144.5	0.8	1.4	chr6
LG31	95	144.2	1.5	0.6	chr16
LG32	73	143.0	2.0	1.0	chr3

Linkage Group (LG)	Number of Markers	Genetic Distance (cM)	Mean distance between markers (cM)	Mean recombination rate (cM/10 kb)	Chromosome
LG33	147	132.1	0.9	0.7	chr7
LG34	130	127.9	1.0	2.3	chr8
LG35	71	118.8	1.7	0.6	chr4
LG36	57	115.7	2.0	0.2	chr7
LG37	99	114.7	1.2	1.0	chr18
LG38	53	111.1	2.1	1.1	chr9
LG39	178	109.1	0.6	1.8	chr10
LG40	26	100.9	3.9	0.5	chr16
LG41	24	90.1	3.8	0.5	chr8
LG42	52	87.2	1.7	0.3	chr16
LG43	54	83.5	1.5	0.5	chr14
LG44	144	72.7	0.5	0.6	chr3
LG45	72	62.9	0.9	0.5	chr6
LG46	59	62.2	1.1	0.5	chr12
LG47	81	58.1	0.7	0.5	chr13
LG48	24	57.6	2.4	0.3	chr1
LG49	56	54.5	1.0	0.6	chr2
LG50	31	51.7	1.7	0.5	chr9
LG51	49	42.8	0.9	1.0	chr2
LG52	52	41.1	0.8	0.4	chr3
LG53	42	39.7	0.9	0.7	chr10
LG54	45	38.1	0.8	0.6	chr11
LG55	49	35.5	0.7	0.5	chr3
LG56	105	27.9	0.3	0.3	chr2
LG57	26	23.4	0.9	1.5	chr1
LG58	29	22.8	0.8	0.3	chr2
LG59	27	19.0	0.7	0.8	chr5
LG60	40	13.5	0.3	0.3	chr18
LG61	22	5.7	0.3	0.1	chr8
Total	10923	10730.5	1.0	0.8	

Markers were used to anchor the race 9 haplotype B reference genome onto the genetic map. All 18 chromosomes were tagged by SNPs and a total of 107.2 Mb (81.3% of the total genome assembly length) were covered by SNPs. A total of 10,831 single-crossover and 765 double-crossover events were detected in the 53 members of the F₂ population (Fig. 3) and the average recombination rate across these linkage groups was estimated to be 0.8 cM/10 kb.

Out of the 20 wheat NILs that were tested, segregation of virulence phenotype among the F₂ progeny was observed in only 13 of them. Among these 13 NILs, it was found that both race 9 and race 161 displayed the same virulence phenotype on *Lr2c*, *Lr11*, *Lr17*, *Lr10*, *Lr18*, and *Lr28* (Table 1). To map all possible *Avr* genes corresponding to these *Lr* genes, we attempted both linkage phases, testing whether the *Avr* allele was from race 9 or race 161. This way, *PtAvrLr11* was successfully mapped to chromosome 3, with the closest marker being chr3_6726200 (Fig. 4), under the assumption that the *Avr* allele derived from race 161. Among the remaining segregating *Avr* genes, *PtAvrLr14a* was mapped to the end of chromosome 1, with the closest marker being chr1_8705360 (Fig. 4). To map additional *Avr* genes, the segregation of A/V in the 53 F₂ progeny was treated as a binary phenotype and analyzed using QTL mapping. For *PtAvrLr14a*, one QTL was identified on chromosome 1, with the closely linked marker chr1_8705360, consistent with results from linkage mapping (Fig. 5A). For *PtAvrLr11*, three QTL were detected on chromosomes 3, 4, and 18, with closely linked SNPs chr3_6726200 (consistent with linkage mapping), chr4_1557250, and chr18_3996934, respectively (Figs. 5B-5D). Additional QTLs were also detected for *PtAvrLr1* and *PtAvrLr2a*. A single QTL for *PtAvrLr1* was mapped to chromosome 1, with marker chr1_6500516 (Fig. 5E). Two QTLs were detected for *PtAvrLr2a*, located on LG8 and LG10, which both reside on chromosome 4 (Fig. 5F and 5G). The closely linked SNPs for these QTLs were chr4_4153557 and chr4_4105933, respectively. These two loci most likely represent the same QTL, indicating that *PtAvrLr2a* maps to chromosome 4. Candidate genes predicted to be secreted and hence are potential (avirulence) effectors, and located within the QTL regions for *PtAvrLr14a*, *PtAvrLr11* (on chromosomes 4 and 18), and *PtAvrLr2a* (on chromosome 4) were identified and are listed in Table S6. In contrast, no secreted genes were found within the QTL regions for *PtAvrLr1* or *PtAvrLr11* on chromosome 3.

Table 1
Phenotypes of parental and F₁ isolates and segregation of avirulent/virulent in the F₂ progeny generated from the sexual cross of *Pt* race 9 x race161 on differential wheat NILs

	Phenotypes			No. of progeny isolates		exp. ratio	
<i>Lr</i> gene	race9	race161	F ₁	Avirulent	Virulent	(A/V)	<i>P</i> ^b
<i>Lr1</i>	1 ^a	0	0	41	16	3:1	0.59
<i>Lr2a</i>	1	0	0	21	36	7:9	0.29
<i>Lr2c</i>	1	1	1	18	39	1:3	0.25
<i>Lr3</i>	0	1	0	52	5	HA ^c	
<i>Lr11</i>	0	0	0	46	11	3:1	0.32
<i>Lr17</i>	1	1	1	1	56	HV ^d	
<i>LrB</i>	0	1	1	17	40	1:3	0.40
<i>Lr10</i>	1	1	1	1	56	HV	
<i>Lr14a</i>	0	1	0	41	16	3:1	0.59
<i>Lr18</i>	0	0	0	54	3	HA	
<i>Lr3bg</i>	0	1	0	50	7	3:1	0.03
<i>Lr14b</i>	1	0	1	16	41	1:3	0.59
<i>Lr28</i>	1	1	1	10	47	1:3	0.19
^a 1 = virulence (V); 0 = avirulence (A)							
^b <i>P</i> , probability of goodness of fit by a Chi-square test.							
^c HA, homozygous avirulent							
^d HV, homozygous virulent							

Region-specific association study

Although four *Pt Avr* loci were identified by linkage mapping and QTL analysis, the corresponding genomic regions are relatively large. To refine the position of these potential avirulence effectors at these genomic loci, a region-specific association study was performed using a natural population of 58 isolates collected in Canada (Table S1). In this approach, SNPs from corresponding genomic regions were assessed for correlation with the specific A/V on the four respective *Lr* gene-containing NILs. After the generation of Illumina genomic DNA reads, filtering and aligning to the same *Pt* race 9 haplotype B assembly, variants were called (see Materials and methods). A total of 408,473 high-quality SNPs were generated for the whole genome. Specific SNPs on chromosomes 1, 3, 4, and 18 were selected to perform the region-specific association studies for *PtAvrLr14a*, *PtAvrLr11* and *PtAvrLr2a*; the association study could not be performed for *PtAvrLr1*, since all of the 58 isolates were virulent on the *Lr1* line.

Three significant SNPs—chr1_8762635, chr1_8776624, and chr1_9064782—were identified for *PtAvrLr14a* (Fig. 4A; Table 3). Among these, chr1_8762635 is an upstream_gene_variant of PTTG_R9hB_001601, which is predicted to encode a secreted hypothetical protein and chr1_8776624 is in two regulatory motifs (Table 3). Similarly, three significant SNPs, chr3_6565157, chr3_6569082, and chr3_6616613, were detected for *PtAvrLr11* on chromosome 3 (Fig. 4B; Table 3). While none of these

were associated with genes encoding secreted proteins, chr3_6565157 overlaps with a C6 zinc cluster transcription factor (Table 3). Additionally, two significant SNPs were identified for *PtAvrLr2a* on chromosome 4 (Fig. 6, Table 3), but neither was associated with secreted protein-coding genes. However, both were located within regulatory motifs, including a CCAAT-binding factor and a C6 zinc cluster factor site. No significant SNPs were identified for *PtAvrLr11* on chromosome 4 and 18.

Table 3
SNPs significantly associated with PtAvrLr14a, PtAvrLr11 and PtAvrLr2a

<i>Avr gene</i>	SNP	Annotation	Associated Genes	Gene Function	Predicted DNA motifs overlapping the SNP
<i>PtAvrLr14a</i>	chr1_8762635	3_prime_UTR_variant	PTR9hB_001599	hypothetical protein	None
		upstream_gene_variant	PTR9hB_001600	hypothetical protein	
		upstream_gene_variant	PTR9hB_001601	hypothetical protein; SECRETED: SignalP(1-31)	
		upstream_gene_variant	PTR9hB_001602	hypothetical protein	
		downstream_gene_variant	PTR9hB_001597	Arf GTPase	
		downstream_gene_variant	PTR9hB_001598	hypothetical protein	
	chr1_8776624	upstream_gene_variant	PTR9hB_001603	Aminomethyl-transferase	IXR1: High-mobility group (HMG) domain factors; MATALPHA2: Homeo domain factors
		upstream_gene_variant	PTR9hB_001605	hypothetical protein	
		upstream_gene_variant	PTR9hB_001607	hypothetical protein	
		downstream_gene_variant	PTR9hB_001604	54S ribosomal protein L7	
		downstream_gene_variant	PTR9hB_001606	tRNA-Pro	
	chr1_9064782	downstream_gene_variant	PTR9hB_001661	hypothetical protein	None
<i>PtAvrLr11</i>	chr3_6565157	missense_variant	PTR9hB_003527	hypothetical protein	UME6: C6 zinc cluster factors
		upstream_gene_variant	PTR9hB_003529	hypothetical protein	
		upstream_gene_variant	PTR9hB_003526	hypothetical protein	
		downstream_gene_variant	PTR9hB_003528	hypothetical protein	
	chr3_6569082	downstream_gene_variant	PTR9hB_003528	hypothetical protein	None
		downstream_gene_variant	PTR9hB_003527	hypothetical protein	
		downstream_gene_variant	PTR9hB_003526	hypothetical protein	
		downstream_gene_variant	PTR9hB_003525	hypothetical protein	
		downstream_gene_variant	PTR9hB_003524	hypothetical protein	
	chr3_6616613	upstream_gene_variant	PTR9hB_003514	metalloendopeptidase	None
		upstream_gene_variant	PTR9hB_003511	hypothetical protein	
		upstream_gene_variant	PTR9hB_003510	hypothetical protein	
		upstream_gene_variant	PTR9hB_003508	hypothetical protein	
		upstream_gene_variant	PTR9hB_003506	hypothetical protein	
<i>PtAvrLr2a</i>	chr4_4288311	missense_variant	PTR9hB_005386	hypothetical protein	HAP5 and HAP3: Heteromeric CCAAT-binding factors
		upstream_gene_variant	PTR9hB_005387	hypothetical protein	
		downstream_gene_variant	PTR9hB_005385	hypothetical protein	

Avr gene	SNP	Annotation	Associated Genes	Gene Function	Predicted DNA motifs overlapping the SNP
	chr4_4248321	upstream_gene_variant	PTR9hB_005373	hypothetical protein	YPR196W: C6 zinc cluster factors
		upstream_gene_variant	PTR9hB_005375	hypothetical protein	
		downstream_gene_variant	PTR9hB_005374	hypothetical protein	
		downstream_gene_variant	PTR9hB_005376	34-kDa subunit of RNA polymerase III	

Discussion

In our cross, segregation for avirulence was observed for *Lr1*, *Lr2a*, *Lr2c*, *Lr11*, *LrB*, *Lr14a*, *Lr3bg*, *Lr14b* and *Lr28*. In terms of inheritance patterns, avirulence in the pathogen seemed to be controlled by single dominant genes towards *Lr1*, *Lr11*, *Lr14a* and *Lr3bg*. Conversely, avirulence appeared to be controlled by single recessive genes towards *Lr2c*, *LrB*, *Lr14b*, and *Lr28*. The inheritance of avirulence towards *Lr2a* did not fit single gene models, suggesting that multiple genes may be involved in this case.

Previous studies [28, 45, 46] have reported similar inheritance patterns for avirulence to *Lr1*, *Lr11*, *Lr14a*, and *Lr14b*. However, there is notable variability in the inheritance patterns of the remaining *Avr* genes between our study and previous studies. In our study, segregation of F₂ progeny on *Lr2c* fit a 1:3 ratio, while isolates from the entire population (F₂1 + F₂3) followed a 7:9 segregation ratio (Table S2), suggesting two independent recessive genes for avirulence. Samborski and Dyck [45], as well as Statler [47] found that avirulence to *Lr2c* and *Lr2a* was governed by single dominant genes in different sexual crosses of *Pt*, which differs from the findings of our study. Samborski and Dyck [45] reported that two genes seemed to condition virulence on *LrB*, while in our study the segregation ratio suggested a single recessive gene for avirulence towards *LrB*. Similarly, Statler [47] in a different sexual cross, suggested avirulence on *Lr28* was determined by a single dominant gene, whereas our study revealed that avirulence on *Lr28* was controlled by a recessive gene in the F₂ population.

Haggag *et al.* [27] found that segregation of virulence towards *Lr3* fit a 3:1 and 13:3 ratio in two F₂ populations, suggesting that avirulence was governed by a dominant gene and an additional gene in the pathogen, possibly *PtAvr3bg*. More recently, Statler [47], based on his own cross, identified a digenic ratio of virulence towards *Lr3*. In our study, F₂ progeny did not segregate on *Lr3*. However, when examining isolates from the entire population (F₂1 + F₂3), their segregation followed a 13:3 ratio (Table S2), consistent with the findings by Haggag *et al.* [27]. The segregation fit a 3:1 ratio on *Lr3bg*, indicating the involvement of a single dominant gene.

The inconsistent inheritance patterns of avirulence in *Pt* observed across various studies can be attributed to the isolate-dependent nature of avirulence inheritance, a phenomenon that has also been reported in other rust fungi [48–51]. Avirulence to a host resistance gene can be controlled by a single gene in one isolate but by two or more genes in another isolate, or exhibit a switch from dominant to recessive inheritance. Additionally, it is possible that some of the wheat cultivars used to test *Pt* populations were not pure *Lr* single-gene lines, which would compound the interactions between such host lines and *Pt* races. Lastly, it should be noted that when the size of the test population is small, misclassification of segregation is more likely to occur. For instance, the misplacement of intermediate interactions can have a larger impact on segregation ratios in a small population.

In general, the markers belonging to each chromosome exhibited proximity in the genetic map, aligning well with the physical map (Table S5). However, certain inconsistencies were observed. For example, although markers from the same chromosome were often positioned closely in the genetic map, their locations did not always show a linear correlation. Furthermore, one chromosome is represented by multiple linkage groups. These inconsistencies could potentially be

attributed to two factors: firstly, the highly repetitive nature of the *Pt* genome can complicate the mapping process, and secondly, the presence of structural variations between the two parental isolates used in the study.

In this study, the generated genetic map of *Pt* spanned a total genetic distance of 10,730.5 cM, which is comparable to other rust fungi. For instance, in the pine fusiform rust fungus *Cronartium quercuum* f. sp. *fusiforme*, a genetic map spanning 3006 cM was constructed for its 76.6 Mb genome [52]. Similarly, Anderson *et al.* [24] generated a genetic map spanning 5,860 cM for the 189.5 Mb genome of the flax rust fungus *M. lini*. Another example is from the wheat stripe rust fungus *P. striiformis* f. sp. *tritici*, for which Xia *et al.* [46] generated a genetic map spanning 7,715 cM for its 84.5 Mb genome. The construction of the linkage map in this study involved selecting markers without segregation distortion. This selection criterion resulted in a substantial reduction in the number of markers, leading to some parts of the genome not being tagged by markers. The resulting 61 linkage groups representing the 18 chromosomes of the *Pt* haploid genome can be attributed to this factor. It is likely that regions lacking developed markers caused the fragmentation of linkage groups. Additionally, the presence of genetic gaps exceeding the 30 cM threshold used in map construction might have occurred in regions characterized by repetitive DNA sequences. These factors contribute to the understanding of the challenges faced in achieving complete and contiguous linkage groups for *Pt*.

The recombination rate in *Pt* exhibited significant variations among different linkage groups as well as within regions of the same linkage group (Table 2 and Fig. 3). In the F₂ population, an average of 204.4 single-crossover and 14.4 double-crossover events per individual were detected, which was higher compared to the 115 crossover events in *M. lini* [24] and 133.2 events in *P. striiformis* f. sp. *tritici* [49]. Previous studies have suggested that the recombination rate is influenced by factors such as repetitive DNA content and the percentage of CpG island [24, 49]. However, further research is needed to gain a deeper understanding of the regulatory mechanisms governing recombination in filamentous fungi.

In this study, a total of four *Avr* loci were found using multiple methods. *PtAvrLr14a* was mapped to the telomeric region of chromosome 1 by both linkage mapping and QTL analysis. Region-specific association studies using a natural *Pt* population also identified significant SNPs at the telomere of chromosome 1, further supporting this finding. *PtAvrLr11* on chromosome 3 was similarly mapped by both linkage mapping and QTL analysis. However, the QTL associated with *PtAvrLr11* on chromosome 3 had the lowest LOD score among the three QTLs identified on chromosomes 3, 4, and 18 by QTL analysis. In addition, segregation of A/V phenotypes indicated that avirulence towards *Lr11* is controlled by a single dominant gene, suggesting that only one of these three QTLs represents the true *PtAvrLr11* locus. Significant SNPs were identified for *PtAvrLr14a*, *PtAvrLr2a* and *PtAvrLr11* (on chromosome 3). Only one SNP (chr1_8762635) associated with *PtAvrLr14a*, was linked to a gene encoding a predicted secreted protein, indicating its potential role as an (*Avr*) effector. Interestingly, a majority of these significant SNPs were found in intergenic regions, in between candidate predicted genes/ORFs. Studies have shown that non-coding SNPs associated with diseases are often located in or near regulatory regions, suggesting their potential to disrupt or modify the binding sites of transcription factors and other regulatory proteins [53]. Consequently, these SNPs may influence gene expression patterns to impact phenotypic traits. Thus, the significant SNPs identified in the intergenic regions of this study and located within regulatory motifs, such as chr1_8776624, chr3_6565157, chr4_4248321 and chr4_4288311, could potentially interfere with the normal functioning of regulatory elements, leading to altered gene expression and the observed phenotypic variations.

Conclusions

A comprehensive genetic map was successfully constructed for *Pt*, providing a valuable resource for genomics research in this pathogen. By utilizing this genetic map in conjunction with the complete phased *Pt* genome, genomic features such as recombination rate were estimated. By integrating linkage mapping and region-specific association approaches, significant SNPs associated with *PtAvrLr14a* were identified with high confidence. These findings represent a significant advancement in our quest to clone and characterize *Avr* genes and shed light on the molecular mechanisms underlying the complex interactions between this pathogen and its host.

Declarations

Data availability

Sets of sequence read data from this study were deposited in the National Center for Biotechnology Information's Sequence Read Archive and are accessible through the BioProject Accession IDs PRJNA1111826 (Canadian isolate reads, Table S1; confidential for now and for reviewers only: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1111826?reviewer=4qrfvpevd52i466ej7jega1g9m>) and PRJNA1259536 (progeny reads, Table S3; confidential for now and for reviewers only: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1259536?reviewer=4cogjl39o7c4h7jls2iingt6dv>). The *Pt* race9 haplotype B genome and its annotation can be found at NCBI under project number PRJNA1298919 (provisional and confidential link for now and for reviewers only: https://drive.google.com/drive/folders/1sMUGtqqWc_DbJLHgOIJTk7j2rT15XeMU?usp=sharing).

[Reviewers, data will be released at the NCBI upon publication of this study]

Contributions

This work was designed by BMc and GB, and co-funded by BMc, BS and GB. BM and BMc were responsible for isolate collections and performed phenotyping, and BM and DLJ initial mapping. BM performed gDNA isolations. SF assembled and haplophased the rust genomes, and S-HK, ML and SF performed genome annotations. LL performed all genetic map constructions and GWAS. GB and LL wrote the manuscript, and all authors read and approved the final version of the manuscript.

Acknowledgements

We thank Curt McCartney for support with the initial map constructions, and Rob Linning for his outstanding technical assistance. The Michael Smith Genome Sciences Centre in Vancouver, BC, Canada, is acknowledged for their expert advice and sequencing of all samples, under supervision of Dr. Richard Moore.

Funding

This project was funded partly through Canada's Digital Technology Supercluster Program, project "Computational Biochemistry Platform for Crop Health," the Genome BC Strategic Opportunities Fund, SOF131 project "Poplar and cereal rust comparative genomics: identification of pathogen determinants to prevent and predict epidemics", the Natural Resources Canada's Genomics Research and Development Initiative (both, joint R. Hamelin and G. Bakkeren), and an Ontario Ministry of Research and Innovation grant #RE-03-056, "Genomics Approaches to Mitigate Fungal Threats to Crops" (joint B. Saville and G. Bakkeren). We also gratefully acknowledge the support of the Agriculture and Agri-Food Canada Genomics Initiative.

References

1. Bolton MD, Kolmer JA, Garvin DF. Wheat leaf rust caused by *Puccinia triticina*. *Mol Plant Pathol*. 2008;9(5):563–75.
2. Huerta-Espino J, Singh RP, Germán S, McCallum BD, Park RF, Chen WQ, Bhardwaj SC, Goyeau H. Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica*. 2011;179(1):143–60.
3. Kolmer JA, Jin Y, Long DL. Wheat leaf and stem rust in the United States. *Aust J Agric Res*. 2007;58(6):631–8.
4. McCallum BD, Hiebert CW, Cloutier S, Bakkeren G, Rosa SB, Humphreys DG, Marais GF, McCartney CA, Panwar V, Rampitsch C, et al. A review of wheat leaf rust research and the development of resistant cultivars in Canada. *Can J Plant Pathol*. 2016;38(1):1–18.
5. Kolmer JA. Tracking wheat rust on a continental scale. *Curr Opin Plant Biol*. 2005;8(4):441–9.
6. Flor HH. Host-parasite interaction in flax rust-its genetics and other implications. *Phytopathology*. 1955;45:680–5.

7. Bernoux M, Ellis JG, Dodds PN. New insights in plant immunity signaling activation. *Curr Opin Plant Biol.* 2011;14(5):512–8.
8. Ngou BPM, Jones JDG, Ding P. Plant immune networks. *Trends Plant Sci.* 2022;27(3):255–73.
9. Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, Baggs E, Witek AI, Yadav I, Krasileva KV, et al. The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol.* 2020;183(2):468–82.
10. Krattinger SG, Keller B. Molecular genetics and evolution of disease resistance in cereals. *New Phytol.* 2016;212(2):320–32.
11. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet.* 2010;11(8):539–48.
12. Koeck M, Hardham AR, Dodds PN. The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cell Microbiol.* 2011;13(12):1849–57.
13. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. *Nat Rev Microbiol.* 2013;11(11):800–14.
14. Upadhyaya NM, Mago R, Panwar V, Hewitt T, Luo M, Chen J, Sperschneider J, Nguyen-Phuc H, Wang A, Ortiz D, et al. Genomics accelerated isolation of a new stem rust avirulence gene-wheat resistance gene pair. *Nat Plants.* 2021;7(9):1220–8.
15. Salcedo A, Rutter W, Wang S, Akhunova A, Bolus S, Chao S, Anderson N, De Soto MF, Rouse M, Szabo L, et al. Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99. *Science.* 2017;358(6370):1604–6.
16. Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, Breen S, Dong C, Xu B, Zhang X, 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.
17. 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. *Front Plant Sci.* 2017;8:148.
18. Kolmer JA, Herman A, Fellers JP. Genotype groups of the wheat leaf rust fungus *Puccinia triticina* in the United States as determined by genotyping by sequencing. *Phytopathology.* 2022;112(3):653–62.
19. Song L, Wu JQ, Dong CM, Park RF. Integrated analysis of gene expression, SNP, InDel, and CNV identifies candidate avirulence genes in Australian isolates of the wheat leaf rust pathogen *Puccinia triticina*. *Genes.* 2020;11(9):1107.
20. Shen S, Wang F, Cui Z, Yuan S, Meng L, Liu D, Ma L, Wang H. *Puccinia triticina* avirulence protein AvrLr21 directly interacts with wheat resistance protein Lr21 to activate wheat immune response. *Commun Biol.* 2024;7:1170.
21. Cui Z, Wu W, Fan F, Wang F, Liu D, Di D, Wang H. Transcriptome analysis of *Lr19*-virulent mutants provides clues for the *AvrLr19* of *Puccinia triticina*. *Front Microbiol.* 2023;14:1062548.
22. Cui Z, Shen S, Meng L, Sun X, Jin Y, Liu Y, Liu D, Ma L, Wang H. Evasion of wheat resistance gene *Lr15* recognition by the leaf rust fungus is attributed to the coincidence of natural mutations and deletion in *AvrLr15* gene. *Mol Plant Pathol.* 2024;25:e13490.
23. Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell.* 2004;16(3):755–68.
24. Anderson C, Khan MA, Catanzariti A-M, Jack CA, Nemri A, Lawrence GJ, Upadhyaya NM, Hardham AR, Ellis JG, Dodds PN, et al. Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini*. *BMC Genomics.* 2016;17(1):667.
25. Saari EE, Young HC, Kernkamp MF. Infection of North American *Thalictrum ssp.* with *Puccinia recondita* f.sp. *tritici*. *Phytopathology.* 1968;58:939–43.
26. Bartos P, Dyck PL, Samborski DJ. Adult-plant leaf rust resistance in Thatcher and Marquis wheat: a genetic analysis of the host-parasite interaction. *Can J Bot.* 1969;47(2):267–9.
27. Haggag MEA, Samborski DJ, Dyck PL. Genetics of pathogenicity in three races of leaf rust on four wheat varieties. *Can J Genet Cytol.* 1973;15(1):73–82.

28. Samborski DJ, Dyck PL. Inheritance of virulence in *Puccinia recondita* on six backcross lines of wheat with single genes for resistance to leaf rust. *Can J Bot.* 1976;54(14):1666–71.
29. Xia C, Qiu A, Wang M, Liu T, Chen W, Chen X. Current status and future perspectives of genomics research in the rust fungi. *Int J Mol Sci.* 2022;23(17):9629.
30. Bakkeren G, Szabo LJ. Progress on molecular genetics and manipulation of rust fungi. *Phytopathology.* 2020;110(3):532–43.
31. Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, Abranches R, Silva MC, Voegelé RT, Loureiro J, et al. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Front Plant Sci.* 2014;5:422.
32. Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders DGO, Uauy C. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics.* 2013;14(1):270.
33. Tobias PA, Schwessinger B, Deng CH, Wu C, Dong C, Sperschneider J, Jones A, Luo Z, Zhang P, Sandhu K et al. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. *G3.* 2021;11(3):jkaa015.
34. Gupta YK, Marcelino-Guimarães FC, Lorrain C, Farmer A, Haridas S, Ferreira EGC, Lopes-Caitar VS, Oliveira LS, Morin E, Widdison S, et al. Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. *Nat Commun.* 2023;14(1):1835.
35. Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, Guo J, Zhuang H, Qiu C, Liu J, et al. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat Commun.* 2013;4(1):2673.
36. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci.* 2011;108(22):9166–71.
37. Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, Sakthikumar S, Song X, Adiconis X, Fan L et al. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3.* 2017;7(2):361–76.
38. Li C, Qiao L, Lu Y, Xing G, Wang X, Zhang G, Qian H, Shen Y, Zhang Y, Yao W, et al. Gapless genome assembly of *Puccinia triticina* provides insights into chromosome evolution in Pucciniales. *Mircobiology Spectr.* 2023;11(1):e02828–22.
39. Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R, Upadhyaya NM, Rathjen JP, et al. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biol.* 2022;23(1):84.
40. Edwards RJ, Dong C, Park RF, Tobias PA. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, *Austropuccinia psidii*. *bioRxiv.* 2022; <https://doi.org/10.1101/2022.04.22.489119>
41. Schwessinger B, Jones A, Albekaa M, Hu Y, Mackenzie A, Tam R, Nagar R, Milgate A, Rathjen JP, Periyannan S. A chromosome scale assembly of an australian *Puccinia striiformis* f. sp. *tritici* isolate of the *PstS1* Lineage. *Mol Plant Microbe Interact.* 2022;35(3):293–6.
42. Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, Kianian SF, Steffenson BJ, Dodds PN, Sperschneider J, 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.
43. Liang J, Li Y, Dodds PN, Figueroa M, Sperschneider J, Han S, Tsui CKM, Zhang K, Li L, Ma Z, et al. Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust. *Mol Ecol Resour.* 2023;23(3):601–20.
44. Lorrain C, Gonçalves Dos Santos KC, Germain H, Hecker A, Duplessis S. Advances in understanding obligate biotrophy in rust fungi. *New Phytol.* 2019;222(3):1190–206.
45. Samborski DJ, Dyck PL. Inheritance of virulence in wheat leaf rust on the standard differential wheat varieties. *Can J Genet Cytol.* 1968;10(1):24–32.

46. Statler GD. Inheritance of virulence of *Puccinia recondita* f. sp. *tritici* on durum and spring wheat cultivars. *Phytopathology*. 1982;72(2):210–3.
47. Statler GD. Inheritance of virulence of *Puccinia triticina* culture X47, the F1 of the cross 71–112 x 70 – 1. *Can J Plant Pathol*. 2000;22(3):276–9.
48. Yuan C, Wang M, Skinner DZ, See DR, Xia C, Guo X, Chen X. Inheritance of virulence, construction of a linkage map, and mapping dominant virulence genes in *Puccinia striiformis* f. sp. *tritici* through characterization of a sexual population with genotyping-by-sequencing. *Phytopathology*. 2018;108(1):133–41.
49. Xia C, Lei Y, Wang M, Chen W, Chen X. An avirulence gene cluster in the wheat stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*) identified through genetic mapping and whole-genome sequencing of a sexual population. *mSphere*. 2020;5(3):e00128–20.
50. Webb CA, Fellers JP. Cereal rust fungi genomics and the pursuit of virulence and avirulence factors. *FEMS Microbiol Lett*. 2006;264(1):1–7.
51. Zambino PJ, Kubelik AR, Szabo LJ. Gene action and linkage of avirulence genes to DNA markers in the rust fungus *Puccinia graminis*. *Phytopathology*. 2000;90(8):819–26.
52. Kubisiak TL, Anderson CL, Amerson HV, Smith JA, Davis JM, Nelson CD. A genomic map enriched for markers linked to *Avr1* in *Cronartium quercuum* f.sp. *fusiforme*. *Fungal Genet Biol*. 2011;48(3):266–74.
53. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res*. 2016;44(18):8641–54.
54. Kolmer JA. Virulence and race dynamics of *Puccinia recondita* f. sp. *tritici* in Canada during 1956–1987. *Phytopathology*. 1989;79(3):349–56.
55. Long DL, Kolmer JA. A North American system of nomenclature for *Puccinia recondita* f. sp. *tritici*. *Phytopathology*. 1989;79:525–9.
56. Anikster Y, Szabo LJ, Eilam T, Manisterski J, Koike ST, Bushnell WR. Morphology, life cycle biology, and DNA sequence analysis of rust fungi on garlic and chives from california. *Phytopathology*. 2004;94(6):569–77.
57. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
58. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA–MEM for multicore systems. In: *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 2019;2019:314–324.
59. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
60. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008.
61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93.
62. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso–2; iso–3. *Fly*. 2012;6(2):80–92.
63. Wu Y, Bhat PR, Close TJ, Lonardi S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet*. 2008;4(10):e1000212.
64. Browning BL, Zhou Y, Browning SR. A One-Penny imputed genome from Next–Generation reference panels. *Am J Hum Genet*. 2018;103(3):338–48.
65. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. 2010;42(4):355–60.
66. Wang J, Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom Proteom Bioinform*. 2021;19(4):629–40.

67. Bland JM, Altman DG. Multiple significance tests: The Bonferroni method. *Br Med J (Clin Res Ed)*. 1995;310:170.
68. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, et al. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinform*. 2021;19(4):619–28.
69. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, Lemma RB, Lucas J, Chèneby J, Baranasic D, et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2024;52(D1):D174–82.

Figures



Figure 1

Chromosomes of the race 9 haplotype B genome with aligned linkage groups (LGs, shown as green bars). Marker densities are calculated using 1 Mb sliding windows across the genomic regions.

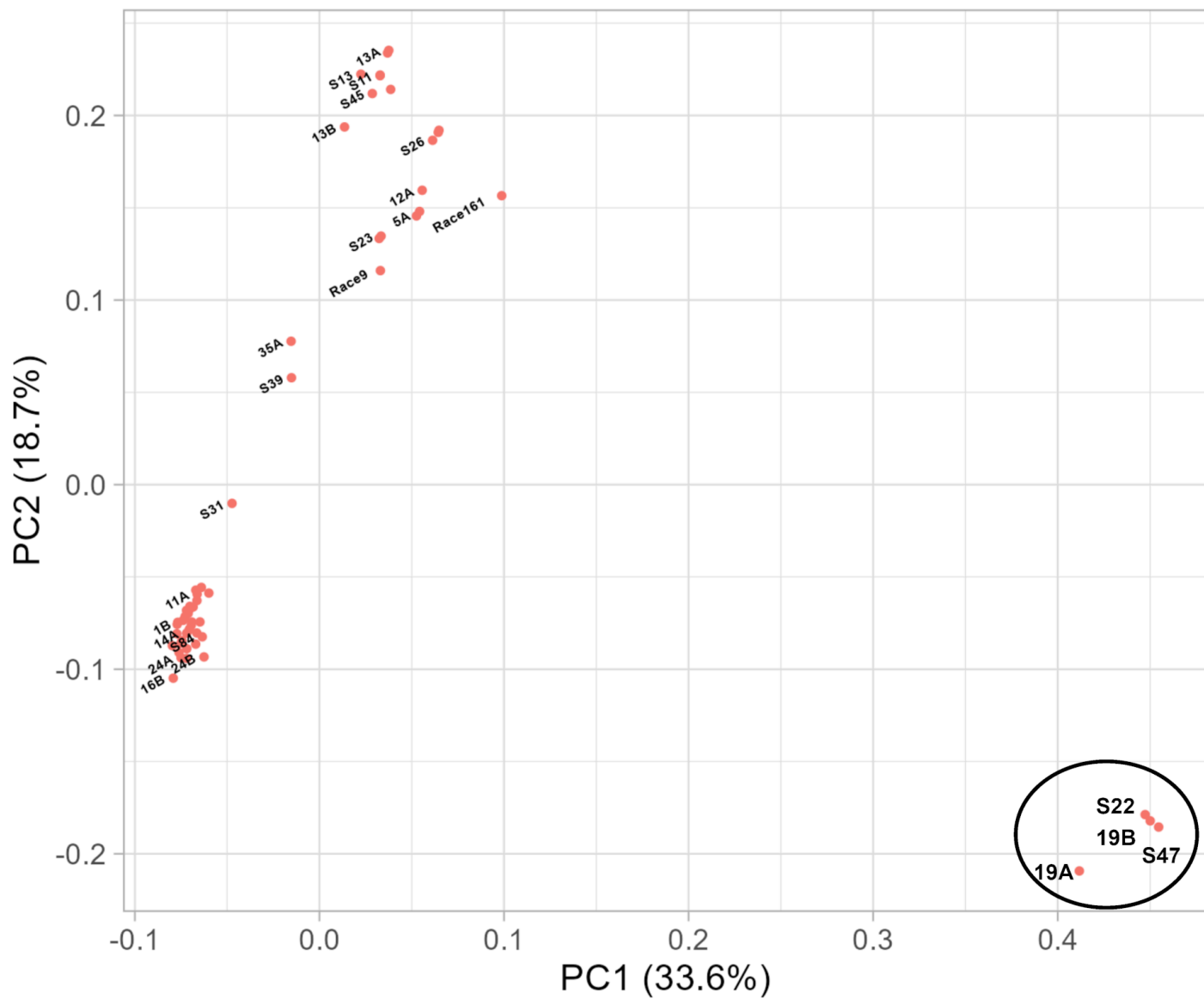


Figure 2

Principal component analysis (PCA) for parental isolates of the sexual cross race 9 x race 161 and the resulting F₁ and F₂ progeny. The circle indicates four contaminated isolates.

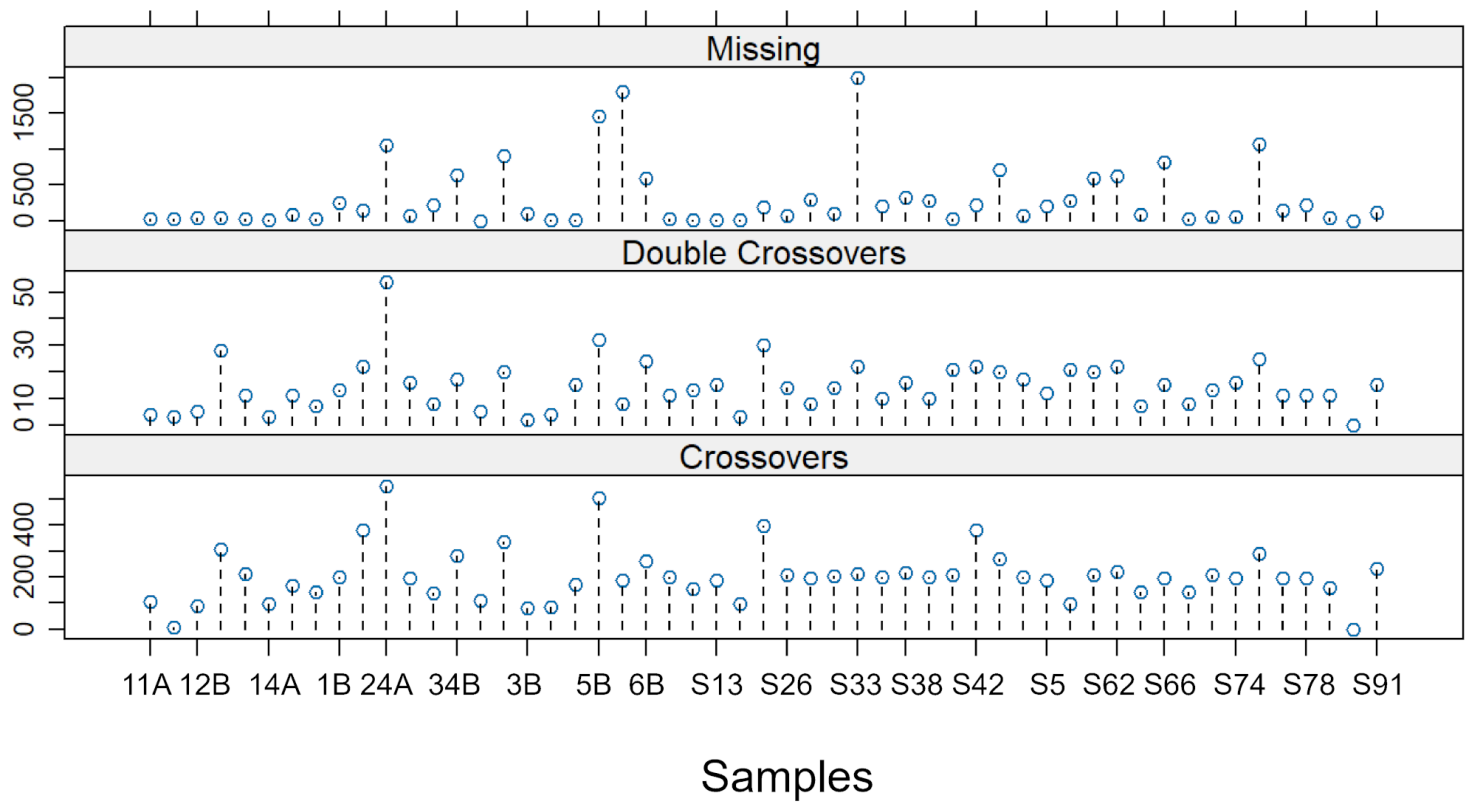


Figure 3

Genotype profiles of missing values, single and double crossover events during recombination for the F₂ progeny isolates generated from the sexual cross of *Pt* race 9 x race161.

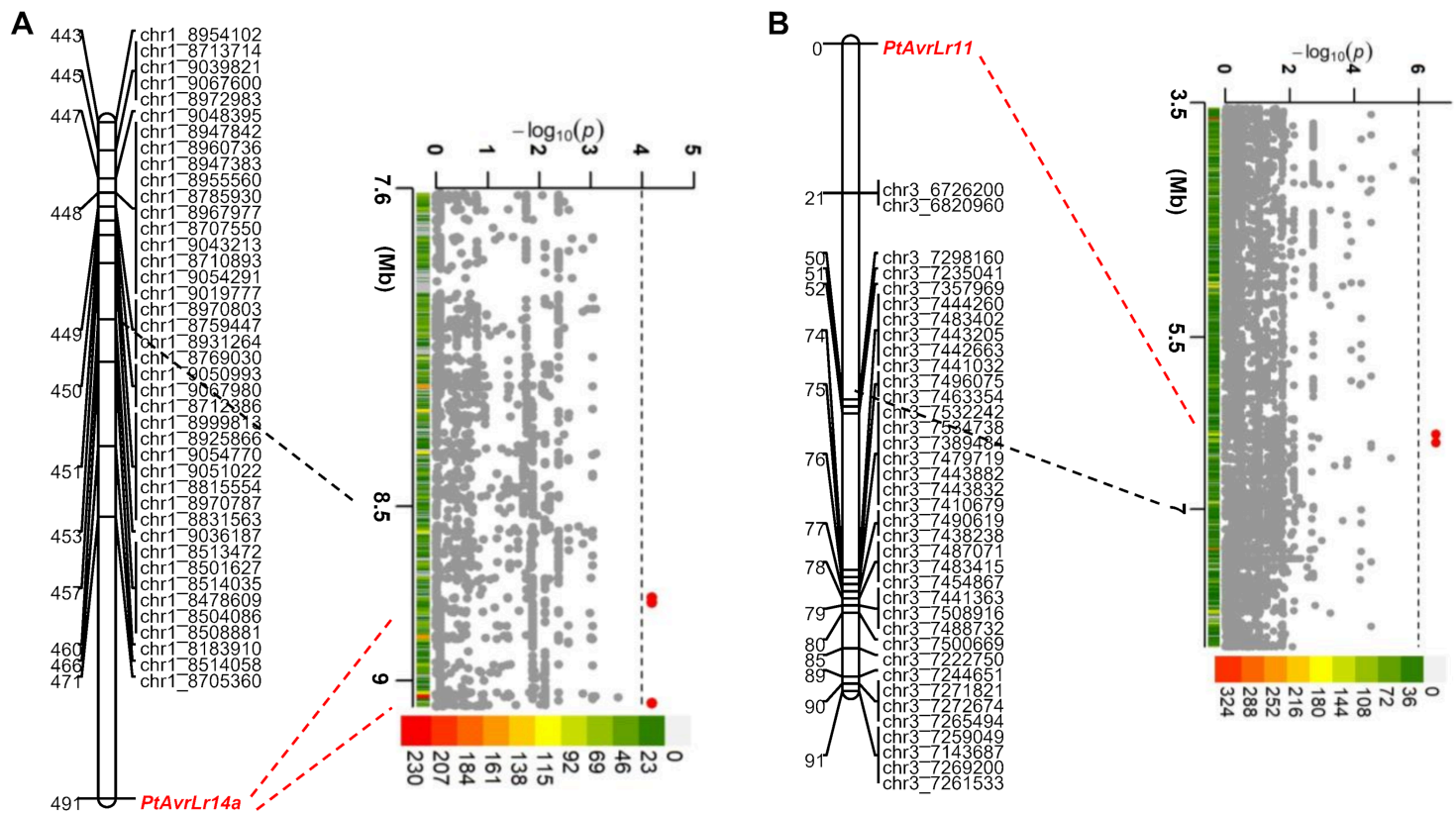


Figure 4

QTL mapping identified loci associated with *PtAvrLr14a* (A), *PtAvrLr11* (B-D), *PtAvrLr1* (E), and *PtAvrLr2a* (F,G).

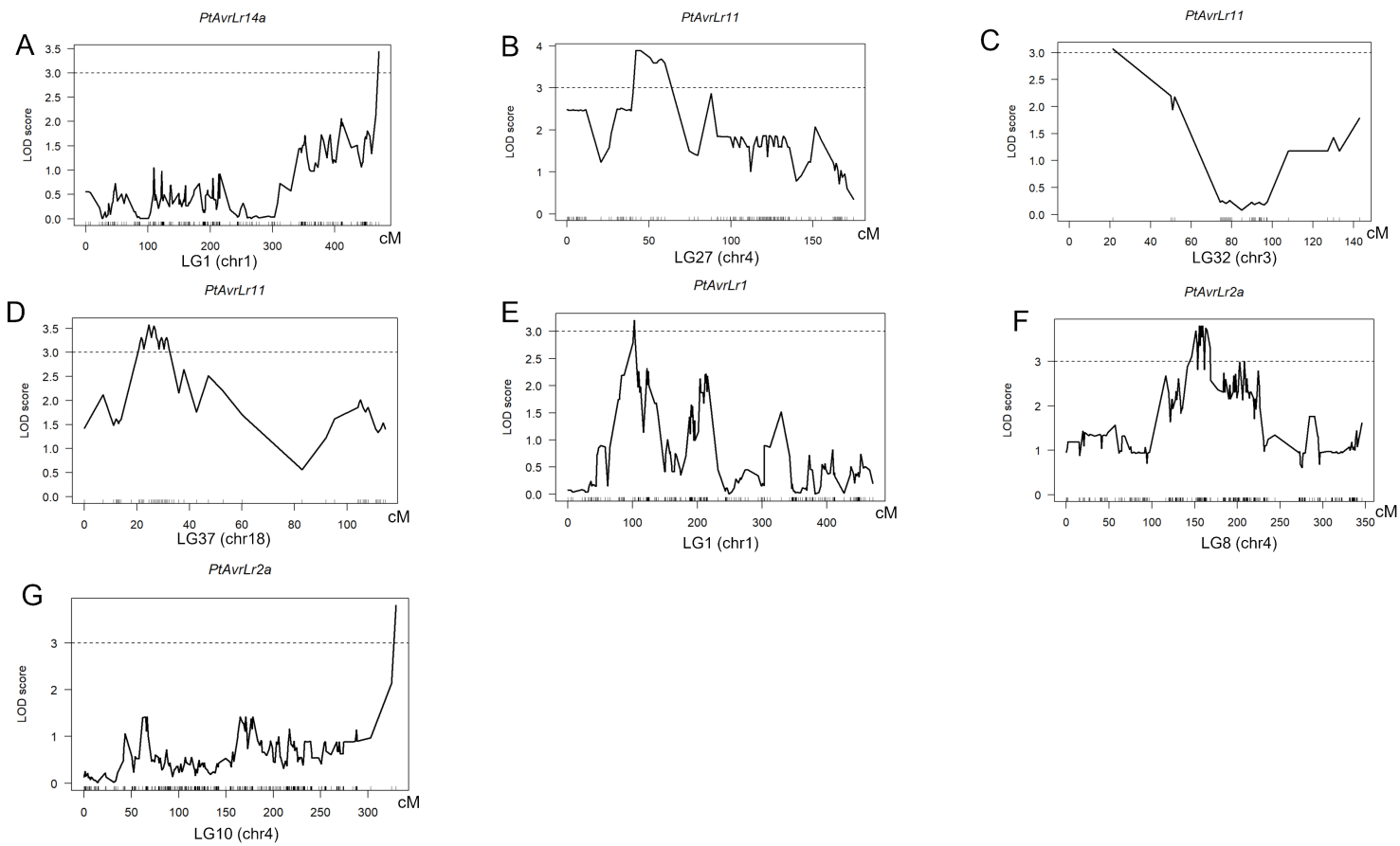


Figure 5

Linkage maps and Manhattan plots for *PtAvrLr14a* (A) and *PtAvrLr11* (B). Only part of the linkage groups or chromosomes are shown. Significant SNPs are indicated with red dots. The bar plots beside the Manhattan plots represent SNP densities within 1Mb window size across the regions.

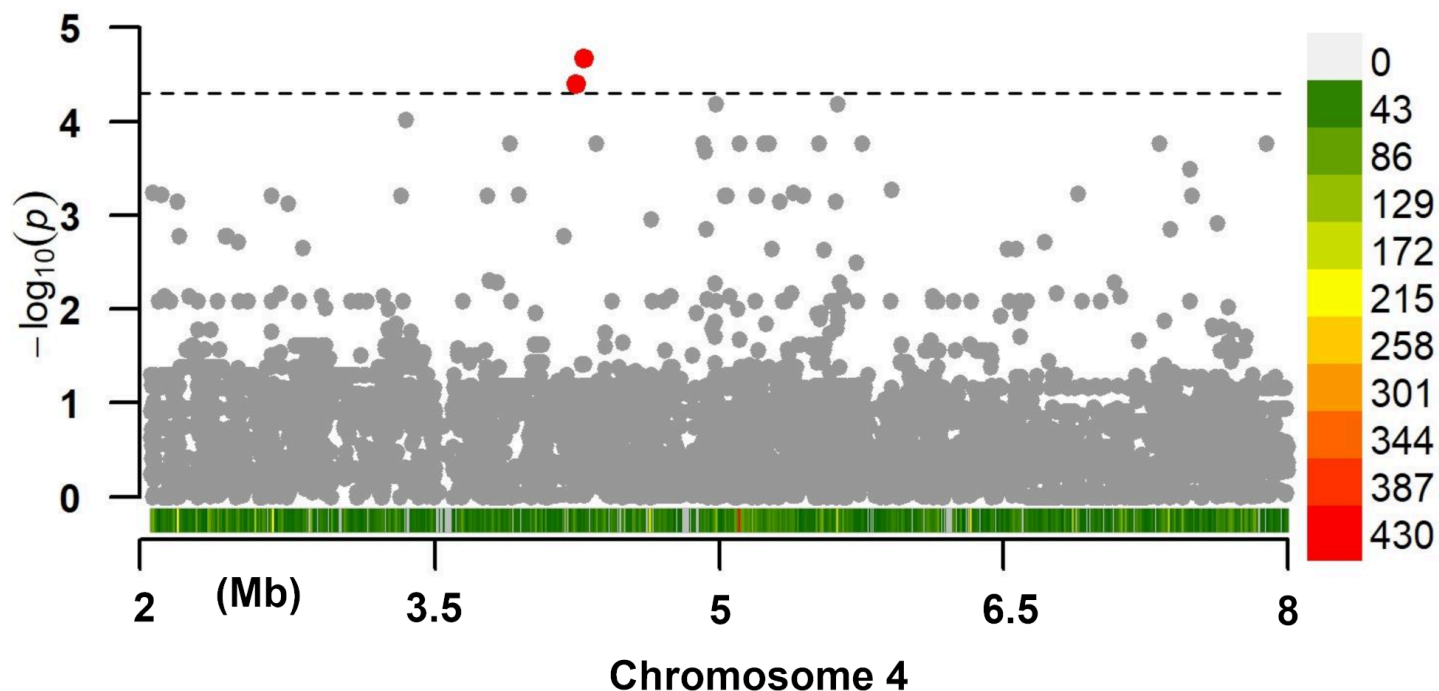


Figure 6

Manhattan plot showing significant SNPs on chromosome 4 for *PtAvrLr2a*. Only a portion of chromosome 4 is displayed. Significant SNPs are highlighted in red. The bar plot beside the Manhattan plot represents SNP density within 1Mb window size across the region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.docx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)