The wheat immune receptor Yr28 recognises a highly conserved effector from stripe rust

Peter Dodds
Peter.Dodds@csiro.au
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0003-0620-5923
Camilla Langlands-Perry
Commonwealth Scientific and Industrial Research Organisation
Jana Sperschneider
CSIRO https://orcid.org/0000-0002-9385-8588
Eva Henningsen
Commonwealth Scientific and Industrial Research Organisation
Jian Chen
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0001-8670-6984
David Lewis
Commonwealth Scientific and Industrial Research Organisation
Parastoo Karimian
Commonwealth Scientific and Industrial Research Organisation
Mumta Chhetri
University of Sydney, Plant Breeding Institute
Li Chen Cheah
Commonwealth Scientific and Industrial Research Organisation
Jibril Lubega
Commonwealth Scientific and Industrial Research Organisation
Davinder Singh
The University of Sydney, Faculty of Science, Plant Breeding Institute, Cobbitty, NSW 2570, Australia. https://orcid.org/0000-0003-1411-9291
Michael Ayliffe
CSIRO Agriculture & Food
Rohit Mago
CISRO https://orcid.org/0000-0001-5286-073X
Kostya Kanyuka
NIAB https://orcid.org/0000-0001-6324-4123
Evans Lagudah
CSIRO Agriculture and Food, Canberra, ACT https://orcid.org/0000-0002-6234-1789
Melania Figueroa
CSIRO https://orcid.org/0000-0003-2636-661X

Letter

Keywords:

Posted Date: July 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9855271/v1

Additional Declarations: Yes there is potential Competing Interest. P.N.D., M.F., and J.S. are inventors on patent application WO2024103117 filed by CSIRO and relating to the identification of protein-protein interactions by protoplast screening. J.S. is an inventor on Australian Provisional Patent Application No.2025201797 filed by CSIRO and relating to gene annotation. The remaining authors declare that they have no competing interests.

Abstract

Stripe (yellow) rust is a devastating disease of wheat caused by the globally distributed plant pathogenic fungus Puccinia striiformis f. sp. tritici (Pst), which results in substantial annual yield losses 1, 2. The Yr28 gene from Aegilops tauschii encodes a nucleotide-binding leucine-rich repeat immune receptor (NLR) that provides broad spectrum resistance to Pst 3, 4. Here we identified the corresponding avirulence protein, AvrYr28-01, by screening a library of effector candidates from Pst. A paralogous gene adjacent to AvrYr28-01 at this locus encodes closely related effector variants that are not recognised by Yr28 due to variation in an N-terminal domain region prior to a predicted alpha-helical bundle. Suppression of Yr28-mediated seedling resistance in some wheat genetic backgrounds is alleviated by between 3 and 6 weeks, with older plants showing effective resistance. The AvrYr28-01 gene is homozygous and invariant in major global lineages of stripe rust, which explains the broad-spectrum resistance of Yr28 and suggests it could provide durable resistance in the field.

Main

Wheat (*Triticum aestivum*) is one of the most important grain crops around the world, but production is impacted by several fungal diseases, including three devastating rust diseases (stem, leaf and stripe rust) caused by species of the genus *Puccinia* [5]. Of these, stripe or yellow rust caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) is the most widespread [2] and is estimated to result in global losses averaging 5.5 million tonnes per annum [1, 6]. Although fungicide application and cultural practices can contribute to rust disease management, the preferred method of disease prevention remains genetic control through the deployment of resistance (*R*) genes in wheat cultivars [5, 7]. Most rust *R* genes in wheat are race-specific, with the majority of cloned genes encoding nucleotide-binding leucine-rich repeat immune receptors (NLRs) [8, 9]. A substantial subset of wheat *R* genes encode kinase fusion proteins including tandem kinases [10], at least some of which rely on NLR helper proteins to trigger immunity [11, 12]. These immune receptors detect specific rust effector proteins known as avirulence (Avr) proteins in a gene-for-gene manner [11, 12]. However, rust strains can evolve to overcome specific *R* genes by alteration or disruption of the corresponding *Avr* genes to escape recognition. Because rust fungi are dikaryotic (containing two haploid nuclei) and avirulence is generally a dominant trait, virulence to specific *R* genes requires homozygosity for non-recognised alleles [12]. Multiple *Avr* genes have been identified in the wheat stem rust fungus, *P. graminis* f. sp. *tritici* (*Pgt*), and assessment of their genetic variation provides important information to prioritise deployment of specific stem rust resistance (*Sr*) genes [13]. For instance, major lineages of *Pgt* are homozygous for *AvrSr22*, suggesting the *Sr22* resistance gene may provide more durable resistance than other *Sr* genes where the corresponding *Avr* loci are heterozygous in these strains. Similarly, *Sr26* was shown to recognise multiple independent *AvrSr26* gene loci in *Pgt* [14], again suggesting enhanced durability for this resistance gene due to the requirement for multiple mutations to escape recognition. However, no *Avr* genes corresponding to effective known stripe rust resistance (*Yr*) genes have been identified in *Pst* to date, making it difficult to assess the potential durability of available *Yr* gene sources. The recent emergence of virulence against the previously broadly effective *Yr15* gene in European *Pst* populations [15] highlights the need to identify additional sources of durable resistance to stripe rust.

The *Yr28* resistance gene was introduced from the diploid wheat progenitor *Ae. tauschii* (D genome donor) into the synthetic hexaploid line W7984 [16] and encodes a canonical NLR protein [3, 4]. *Yr28* provides resistance to many *Pst* isolates worldwide, including 16 out of 16 North American races tested and at least six field races in China [4], as well as four *Pst* race groups in Australia [3]. Wheat lines carrying *Yr28* have also shown effective resistance to stripe rust in the field in the United Kingdom [17] and Canada [18]. To understand the broad specificity of *Yr28* we set out to identify the corresponding *Avr* gene from *Pst* using an effector library screening approach [19]. In order to design a stripe rust effector library we first generated a nuclear-phased reference genome assembly and accurate annotation for the rust isolate Pst104 (pathotype 104 E137 A-) [20], which represents the internationally defined lineage PstS0 [21]. PacBio HiFi sequence data was generated and assembled along with Hi-C data to generate two nuclear haplotype genomes (hap01 and hap02) each with 18 chromosomes of total size 75.4 and 77.4 Mbp, respectively. This assembly has 0.55 Mbp of additional sequence and slightly higher BUSCO completeness compared to a recent Nanopore-based assembly of the same isolate [22] (Supplementary Table 1). The Pst104 genome was annotated using both Funannotate [23] and EffectorGeneP [14], a machine learning tool designed to optimise detection of fungal effector genes, which are difficult to annotate using standard pipelines. To support annotation we generated PacBio IsoSeq as well as stranded Illumina RNAseq data from susceptible plants infected with Pst104 (8 days post infection) and combined this with published RNAseq data for this isolate from various life stages including isolated haustoria [24]. We merged the EffectorGeneP and Funannotate annotations into a non-redundant set and identified 7,919 secreted proteins (including isoforms encoded by 6,685 genes), a substantial increase compared to the previous annotation [22] which included 5,346 secreted proteins (Supplementary Table 1).

We selected 1000 effector candidate genes (Supplementary Table 2) based on expression patterns and effector coding potential as predicted by EffectorP 3.0 [25]. These were synthesised in two libraries: one consisting of 694 candidates encoding mature proteins of up to 142 amino acids; and a second including 306 candidates encoding mature proteins >142 amino acids. Both libraries included the wheat stem rust effector AvrSr50-01 [26] as a positive control. Libraries were screened by co-transformation into wheat protoplasts along with the *Sr50* or *Yr28* resistance genes or an empty vector, and the expression of individual effector candidate genes compared between treatments (Figure 1A and Supplementary Figure S1A). The *AvrSr50* control was correctly identified from each library after screening against *Sr50*. In addition, a single candidate *Pst* effector gene (ID #718) showed reduced expression specifically in the presence of *Yr28*. This gene encodes a mature protein of 156 amino acids (after removal of the signal peptide) and represents a candidate for *AvrYr28*.

We further tested this gene (designated *AvrYr28-01*) by individual co-expression with *Yr28* in wheat protoplasts, which confirmed its specific recognition leading to cell death and loss of expression of the yellow fluorescent protein (YFP) reporter gene (Figure 1B and Supplementary Figure S1B). Co-expression of *AvrYr28-01* with *Yr28* in oat (*Avena sativa*) protoplasts (Supplementary Figure S1C) or in *Nicotiana benthamiana* leaves also led to specific recognition and induced cell death (Figure 1C). Interestingly, fusion of YFP to the AvrYr28-01 protein at the N-terminus prevented recognition in *N. benthamiana*, while the untagged version was recognised (Supplementary Figure S1D). These data confirmed the function of *AvrYr28-01* in triggering *Yr28* mediated resistance.

To assess genetic variation at the *AvrYr28* locus in *Pst*, we generated nuclear-phased genome assemblies for additional isolates representing different clonal lineages. We first generated Illumina whole genome sequencing (WGS) data for 14 isolates of *Pst* collected from around Australia (Supplementary Table 3). This was combined with published Illumina sequence data from 92 additional global *Pst* isolates (Supplementary Table 4) and used to generate a phylogenetic tree based on single nucleotide polymorphisms (SNPs) (Figure 2A and Supplementary Figure S2). The fourteen Australian isolates represent four clonal lineages. Three isolates (19VIC05, 20VICV74 and 24ACT159) form a clonal group with the PstS1 reference isolate [27], representing the 134E16 pathotype group known in Australia since 2002, as well as twelve Canadian isolates also assigned to the PstS1 lineage [28]. Similarly, two isolates (22ACT160 and 20ACT99) form a clonal group with another Australian isolate, s674, which represents the 239E237 pathotype group and was previously assigned to the PstS10 lineage [21, 29], as well as 24 European isolates of PstS10 [30]. Seven isolates group with another Australian isolate, s687, from the 198E16 pathotype group that was reported to represent the PstS13 lineage based on SSR marker analysis [21]. Finally, three isolates form an additional clonal group but are not related to other sequenced isolates of known lineage and represent the 238E191 pathotype group (PstSX).

We generated haplotype-phased genomes for four isolates (24ACT159, 22ACT160, 19NSW01, 23NSW60) representing the PstS1, PstS10, PstS13 and PstSX lineages. Each haplotype contains 18 chromosomes with a total size of 75-77 Mbp (Supplementary Table 5). Comparison of nuclear haplotype sequences showed that these isolates and Pst104 (PstS0 lineage) share different combinations of six nuclear haplotypes, designated hap01 to hap06 (Figure 2B). The hap01 nuclear haplotype is shared in both Pst104 (PstS0) and 22ACT160 (PstS10) with only 2,815 SNPs between the two versions, while hap05 is shared by 22ACT160 (PstS10), 19NSW01 (PstS13) and 23NSW60 (PstSX) with just 509 to 769 SNPs between these lines. Similarly, hap03 is shared in 24ACT159 and 23NSW60 with just 996 SNPs. In contrast, the six divergent haplotypes fall into two groups by sequence similarity (Figure 2C), with haplotypes within each group differentiated by about 120,000 to 230,000 SNPs, while haplotypes from the two different groups are differentiated by about 700,000 SNPs. This is consistent with emerging evidence for two divergent sets of haplotypes (A and B types) in *Pst* populations, with sexual populations in China containing exclusively A nuclear types while sexual populations in South Asia contain B types and most globally prevalent clonal lineages have an AB combination and likely derive from hybridisation [31-33].

We also compared these genomes with recently published nuclear-phased assemblies for nine other *Pst* isolates [32, 34] (Figure S3). Hap06 from 19NSW01 is shared with isolates GSY16-9 and GSY17-10, which represent two different clonal lineages that are widespread in China. Hap05 from PstS10, PstS13 and PstSX was shared by an AA type isolate of the PstS7 lineage, which was first detected in Europe in 2011 (Warrior group) [2] but may be derived from the sexual population in China [32]. K-mer containment analysis on WGS data from global *Pst* isolates (Supplementary Table 6) supports these haplotype assignments for all isolates in these defined lineages and also identifies other potential nuclear exchange hybrids. For instance, a clonal lineage of four Canadian isolates (W004, W008, W013, W019 and W024) representing a clonal lineage all contain hap03, while another lineage (US isolates pst21 and pst11-281) contains hap04. The barley stripe rust (*P. striiformis* f.sp. *hordei*) isolate Psh93TX-2 contained hap02. These observations and other recent evidence for nuclear exchange between *Pst* lineages in North America and globally [29, 31] suggests that this process is an important generator of genetic diversity in global stripe rust populations as previously observed in wheat stem rust [29, 31], wheat leaf rust [35], oat crown rust [36] and barley leaf rust [37].

Interestingly, Hovmoller *et al.* [29] proposed based on a limited set of simple sequence repeat (SSR) markers that the PstS14 lineage detected in Africa, Europe and the Middle East since 2016 contains the same nuclear haplotype composition as the PstSX lineage detected here. While this needs to be confirmed by WGS data for PstS14 isolates, this suggests that the emergence of the PstSX lineage in Australia since 2021 may be the result of long-distance dispersal of PstS14. Alternatively, PstSX may result from an independent hybridisation event that occurred in Australia, given that the PstS1 and PstS10 lineages have been prevalent in Australia since 2002 and 2017 respectively [21].

The *AvrYr28-01* gene is invariant in sequence in all six haplotypes derived from the lineages PstS0, PstS1, PstS10, PstS13 and PstSX, so these lineages are all homozygous for this avirulence gene (Figure 3A). This is consistent with the virulence phenotypes, as all five sequenced Australian isolates were avirulent when inoculated onto wheat line W7984 carrying *Yr28* (Supplementary Figure S4, Supplementary Table 7). Other isolates with published genome sequences [32, 34] are also homozygous for this gene (Supplementary Figure S5), including AA and BB types from sexual populations. Mapping of WGS read data from other diverse global isolates indicates that they all contain identical gene sequences to *AvrYr28-01*. There is also an additional closely related gene ~1.5 kbp upstream of *AvrYr28-01* (Figure 3A and Supplementary Figure S5). Three sequence variants of this gene, designated *AvrYr28-02* to *-04*, occur in A-type haplotypes while B-type haplotypes contain a truncated variant (*AvrYr28-trunc1*) that is disrupted by an internal deletion of 189 nucleotides beginning in the fifth codon and ending within the first intron that results in a frameshift and premature stop. The other three variants (*-02* to *-04*) encode proteins of 185 amino acids that differ from each other by single amino acid changes and from AvrYr28-01 by 24-25 amino acid substitutions and the insertion of seven additional amino acids in the N-terminal region (Supplementary Figure S6). These three variants were not recognised by *Yr28* when co-expressed in *N. benthamiana* leaves (Supplementary Figure S7a). In protoplast assays AvrYr28-02 and -03 showed only potentially weak recognition that was not reproducible between experimental repeats (Supplementary Figure S7b), suggesting that these variants do not contribute to the avirulence phenotype of *Pst* on *Yr28* wheat lines.

Protein structures predicted by Alphafold3 for the AvrYr28-01 mature protein and the non-recognised AvrYr28-02 to -04 variants consist of a helical bundle of four a-helices, with non-structured N and C-terminal regions (Supplementary Figure S8A). Structural modelling of the Yr28/AvrYr28-01 protein complex with (Supplementary Figure S8B) showed the AvrYr28 protein interacting with the LRR domain of Yr28 with the a1-helix contacting residues in LRR repeat units 16-18. There were also extensive contacts between the N-terminal region of AvrYr28-01 and the inner surface of the LRR, although this region of AvrYr28-01 showed low prediction confidence. The N-terminal region is highly divergent between AvrYr28-01 and the non-recognised variants. Ten amino acid substitutions occur in the helical bundle region but none of these are in contact with the Yr28 LRR in the structural model. To investigate which regions contribute to the difference in recognition between these proteins, we generated a series of domain swaps between AvrYr28-01 and AvrYr28-04 exchanging the N-terminal unstructured region, the a-1/a-2 helices, or the a-3/a-4 helices and C-terminus (Figure 3B). Recognition assays of these recombinant proteins in wheat protoplasts showed that the N-terminal region was responsible for the difference in recognition specificity, while swaps within the a-helix bundle

region did not affect recognition (Figure 3C). Furthermore, deletion of 4, 8, 12 or 16 amino acids from the N-terminus of AvrYr28-01 also abolished Yr28-specific cell death induction by this protein (Supplementary Figure S9), indicating that this region is necessary for recognition. These results are consistent with the observation that an N-terminal YFP fusion suppressed recognition of AvrYr28-01, likely by steric interference with the interaction complex.

In some hexaploid backgrounds suppression of the *Yr28* resistance phenotype was observed in seedling assays, such as when *Ae. tauschii* carriers were crossed to the tetraploid durum variety Langdon [3, 4]. However, there is emerging evidence that this gene may nevertheless provide effective resistance in the field in adult plants. For instance, Brar *et al.* (2018) found that an Avocet+Yr28 line showed near immunity to stripe rust under field conditions in Canada, despite all tested Canadian stripe rust isolates showing virulence on this line at the seedling stage [38]. Similarly, *Yr28* provides effective adult plant resistance in China despite low levels of seedling resistance in these lines and is additive with the quantitative resistance genes *Yr18* and *Yr36* [39, 40]. Synthetic hexaploid wheat derived lines from a nested association mapping population carrying *Yr28* showed strong resistance in the field to UK isolates of stripe rust [17]. We tested derivatives of two of these lines at the seedling stage with a UK isolate of *Pst* and found only low levels of resistance (Supplementary Table S7), again indicating seedling suppression. To further characterise the age-dependence of *Yr28* resistance in a suppressive genetic background, we examined resistance expression in the LangdonX672 synthetic hexaploid containing *Yr28* [3]. Four *Pst* isolates representing the PstS1, S10, S1 and SX lineages were virulent on 10-day old seedlings of LangdonX672, while Pst104 (S0) was avirulent on both LangdonX672 and the Langdon parent (Supplementary Table 7). Isolate 24ACT159 was used to infect AvocetS, Langdon, LangdonX672 and W7984 plants at ten days, three weeks, six weeks and nine weeks of age (Figure S10). At ten days, W7984 was fully resistant while LangdonX672 again showed similar infection to the susceptible parent Langdon. At three weeks, both Langdon and LangdonX672 showed similar pustule formation, but there was substantial background necrosis in LangdonX672. However, at six weeks and nine weeks a clear difference emerged, with Langdon infection giving rise to extensive pustule formation and prolific spore production while pustules on LangdonX672 were smaller, dark brown in colour and did not release spores. Microscopic analyses using fluorescently labelled wheat germ agglutinin (WGA) to stain fungal chitin (Figure S10B) showed extensive fungal proliferation in Langdon and Avocet with brightly green fluorescing pustules. However, in LangdonX672 pustules lacked internal fluorescence. In W7984, only very limited hyphal spread with no pustule development was observed in the infected leaves. Altogether, these observations suggest that *Yr28* could provide effective adult plant resistance in the field despite low levels of seedling resistance in some backgrounds. Since *AvrYr28-01* is invariant and homozygous in all available genome sequences of *Pst* strains, evolution of virulence to *Yr28* in these clonal lineages would require two independent mutations, one in each nuclear haplotype. This suggests that *Yr28* would be a valuable gene to provide broad spectrum and durable resistance to stripe rust.

In wheat stem rust, the identification of *Avr* genes and assessment of their genetic variation across wheat stem rust lineages has allowed the establishment of a *Avr* gene atlas providing important information to assess the value of specific resistance genes for use in breeding [13, 14]. For instance, *AvrSr22* is homozygous for avirulence in all the sequenced strains of stem rust, while multiple independent *AvrSr26* loci occur in a single isolate. This reduces the risk of virulence evolution and suggests that these genes could provide more durable resistance than other resistance genes for which the corresponding *Avr* loci are heterozygous in many strains. Recently, a stripe rust effector associated with virulence to an unknown resistance gene in the wheat line Kalmar was identified [41], while another effector triggers defense responses in the susceptible wheat line Avocet, possibly by an unknown suppressed resistance gene [42]. Continued characterisation of interacting wheat resistance and *Pst* effector genes and genetic variation in pathogen haplotypes will allow development of a similar comprehensive *Avr* gene atlas for *Pst* to support prioritisation of resistance loci in breeding programs. This will also provide a valuable resource to enable virulence prediction from amplicon sequence-based surveillance tools such as the MARPLE platform [43, 44].

Methods

Rust isolates and plant material

The stripe rust isolate Pst104 (pathotype 104E137A-) was described previously [20]. Additional *Pst* isolates were collected in the field from various locations around Australia from 2019 to 2024, while WYR25-025 was collected in the UK in 2025 (Supplementary Table 3). Single pustules of each isolate were selected after infection of the susceptible wheat cultivars and increased several times to obtain pure cultures. W7984 is the original synthetic hexaploid source of *Yr28* [16]. LangdonX672 is a synthetic hexaploid generated by crossing *Ae. tauschii* line CPI110672 carrying *Yr28* (*672* haplotype) and durum wheat cultivar Langdon [3]. Microscopic visualization of rust infection in plant tissue using wheat germ agglutinin (WGA) at 20 µg/ml conjugated to the fluorophore alexa 488 (Invitrogen, San Diego, CA, U.S.A.) was conducted as previously described [45]. Wheat lines NIAB.SHW.BC.348.1.5.1.1 and NIAB.SHW.BC.222.3.5.1.1 carrying *Yr28* (*AS2388* or *Yr28* haplotype respectively) are derived from primary synthetic wheat lines SHW071 and SHW082 [17] by back crossing to the winter wheat cultivar Robigus. Lines were genotyped using the KASPYr28 marker [3].

Genome sequencing and assembly

DNA for short read sequencing was extracted from 30 mg of urediniospores using the Omniprep™ Genomic DNA isolation kit (#786-136, G-Biosciences), according to the manufacturer's instructions. Samples were sequenced by the Australian Genome Research Facility (AGRF) (Melbourne, Australia) using 150 bp paired-end Illumina Prep (M) and the Illumina NovaSeq 6000 or NovaSeq X platforms (Supplementary Table 3). High molecular weight DNA for long-read sequencing was extracted from ~450 mg of urediniospores as described [46] and PacBio HiFi libraries were sequenced on the PacBio Revio platform at AGRF. For Hi-C sequencing of 19NSW01 and 23NSW60, 200 mg of urediniospores were cross-linked in 1% formaldehyde (20 min at ~22 °C) and then glycine was added to a 125 mM final concentration. Cross-linked samples were processed at the Biomolecular Resource Facility at the Australian National University (BRF-ANU) and sequenced using the Arima High Coverage HiC kit on the MiSeq i100 100M 300 platform. PacBio HiFi reads were assembled using hifiasm v0.23.0 [47] in Hi-C-integration mode using corresponding Hi-C reads (Pst104 SRR31647314; 24ACT159 SRR15357737; 22ACT160 SRR33364113) with default parameters. Initial assemblies were cleaned of contaminant and mitochondrial contigs as described previously [35, 36]. HiFi reads were aligned to the assembly

using minimap2 v2.28 [48] and coverage was assessed using the pileup.sh tool from bbmap v39.06 (https://sourceforge.net/projects/bbmap/). Low coverage contigs were removed from assemblies after determining cut-offs (15x for Pst104, 30x for the other isolates) with a custom R script (https://github.com/JanaSperschneider/GenomeAssemblyTools/tree/master/CollapsedGenomicRegions). The NuclearPhaser v1.1 pipeline (https://github.com/JanaSperschneider/NuclearPhaser) [49] was used to assess phasing along with Hi-C contact maps generated with Hi-C-Prov3.1.0 and Hicexplorer [50]. BUSCO v3.0.2 [51] was used to assess completeness of the assemblies (-l basidiomycota_odb9). Chromosomes were numbered according to the Pst134 [27] and AZ2 [34] assemblies. Haplotypes were numbered 01 to 06 rather than using letters to allow for expansion of the naming scheme to >26 haplotypes as now observed in *Pca* [52].

RNA sequencing and gene annotation

RNA was extracted from infected plants (8 days post infection) using the Maxwell® RSC Plant RNA Kit (Promega, Catalog #AS1500) with the Maxwell® RSC 48 Instrument (Cat. #AS8500). Stranded RNA sequence data was generated using the PacBio Revio IsoSeq long-read platform as well as Illumina RNA sequencing (150 bp paired-end reads) by AGRF (Brisbane, Australia). This data was combined with publicly available unstranded Illumina RNA-seq data for this isolate from isolated haustoria, infected plants (6 days post infection) and germinated and ungerminated spores [24]. RNA-seq reads were cleaned with fastp 0.23.2 [53] using default settings, aligned with STAR 2.7.9a [54] in 2-pass mode (-alignIntronMin 5 --alignIntronMax 3000 --alignMatesGapMax 3000 --outFilterMultimapNmax 100). Transcripts were assembled with Stringtie 2.2.1 [55] (-s 1 -m 200) with the appropriate strand setting. Iso-Seq reads were processed with isoseq refine to remove polyA tails and artificial concatemers (https://github.com/PacificBiosciences/pbbioconda) and then cleaned using fastp (--trim_poly_x) and an adapter file containing Iso-Seq primers. Iso-Seq reads were aligned to the genome with minimap2 [48] (-ax splice:hq -uf -G 3000). Hybrid RNA-seq and Iso-Seq transcripts were assembled with Stringtie 2.2.1 (-s 1.5 -m 200 --mix) and merged with the RNA-seq transcripts (stringtie –merge) into an input set for EffectorGeneP 1.0.0 [14] and CodingQuarry 2.0 in pathogen mode [56]. We also ran Funannotate 1.8.5 (train with --jaccard_clip and all RNA-seq/Iso-seq reads; predict with --optimize_augustus, --ploidy 2 and providing CodingQuarry output as --other_gff). Secreted proteins containing a signal peptide predicted by SignalP 4.0 [57] (-u 0.34 -U 0.34) and no additional transmembrane domains predicted by TMHMM 2.0 [58] were selected from the EffectorGeneP and Funannotate annotations and combined to generate a non-redundant set with agat_sp_merge_annotations.pl [59].

Library Screen

One thousand secreted effector candidates were selected from the new Pst104 genome annotation based on haustorial expression and prediction as cytoplasmic effectors by EffectorP 3.0 [60]. Coding sequences of the mature proteins were codon-shuffled using Geneious to mimic wheat codon usage (according to this table: https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4565) and to maximise sequence differences between related gene sequences. Codon-modified effector coding sequences were synthesised by Twist Bioscience by one of two pipelines. Effectors of less than 143 amino acids (694 candidates plus *AvrSr50*, *AvrSr22* and *AvrSr13* as controls) were synthesised as a pool of multiplexed gene fragments (MGF) and cloned into pTA22 [19]. For this MGF library all genes were synthesised as 450 bp fragments with a primer binding sequence (PBS; CGTACTGTCGAGATCTAGCA) after the stop codon and unique random sequence fillers added to the 3' end of each to make up the total length. Random filler sequences were generated with emboss shuffleseq by shuffling the following filler sequence:ACGCGATCGGAGGCGCTCATTATACGCAGATTCTTTATCGAAGCTGAGGGTGTGCCCGCTGTAACCCGCAAAGCCGTCAATATACAATCCTGACCAAATAGG. Constructs encoding effectors of 143 to 365 amino acids were synthesised as a separate batch of 306 individual clones in pTA22 and then pooled in equimolar amounts using an Eppendorf epMotion 5075 automated liquid handling system (CSIRO robotic systems), with pTA22-AvrSr50 [19] included as a positive control.

Pooled libraries were propagated in ElectroMAX™ Stbl4™ competent cells and DNA was isolated using the QIAGEN EndoFree Plasmid Giga kit (306 candidate library) or the ZYMO Research ZymoPURE II Plasmid Midiprep Kit (MGF library). The empty vector (pTA22), pTA22-Yr28 and pTA22-Sr50plasmids were prepared from *E. coli* DH5α using the Macherey Nagel NucleoBond Xtra Maxi Plus EF kit. For library screening a total of 300,000 protoplast cells of wheat cv Fielder (306 candidates library) or cv Avocet (MGF library) were transformed with an effector library at a multiplicity of transfection (MOT) of 0.14 million molecules per cell for each construct along with either pTA22, pTA22-Yr28 or pTA22-Sr50(MOT = 36 million molecules per cell) in triplicate as described [19]. A control protoplast sample (300,000 cells) was transformed with pTA22-YFP(MOT = 36 million molecules per cell) and pTA22-AvrSr50 (MOT = 100 million molecules per cell) to assess transfection efficiency by fluorescence as described below.

Total RNA was extracted from transformed protoplasts after 20 hours using the Maxwell® RSC Plant RNA Kit (Cat. #AS1500) with the Maxwell® RSC 48 Instrument (Cat. #AS8500) as described [13]. Library-specific cDNA synthesis and PCR was carried out on 1µg RNA using the Invitrogen SuperScript IV One-Step RT-PCR System with ezDNase kit (Cat. # 12595100) with the forward and reverse primers ZmUbi1_5UTR_F3b (5'-GCACACACACACAACCAG-3') and FS_cDNA_R (5'-TGCTAGATCTCGACAGTACG-3') as previously described [19]. For the MGF library, the cDNA was sent to AGRF (Melbourne, Australia) for Illumina library preparation and sequencing on the NovaSeq X on one lane of a 10B flow cell with 150 bp paired-end reads. For the 306 candidate library, the cDNA was sent to BRF-ANU (Canberra, Australia) for Illumina library preparation and sequencing on the NovaSeq X Plus with a 1.5B flow cell with 100bp single-end reads. Sequencing reads were cleaned using fastp 0.23.2 [53] and aligned to the coding sequences of the effector candidates plus the vector-derived UTRs with STAR 2.7.9a [54] (--alignIntronMax 1). Read counts were extracted with samtools idxstats [61] and differential expression analysis was carried out using DESeq2 [62] with default parameters followed by lfcShrink (type="apeglm"). Volcano plots were produced with EnhancedVolcano (https://github.com/kevinblighe/EnhancedVolcano).

Transient expression in protoplastsand *N. benthamiana*

The coding sequences (excluding signal peptides) of *AvrYr28* gene variants and recombinants were codon-modified as aboveand synthesised directly in pTA22 (Twist Bioscience). Plasmids used in this study are listed in Supplementary Table 8. For protoplast assays, plasmids were purified using the NucleoBond Xtra Midi EF kit (Macherey-Nagel). Seedlings of wheat cultivars Fielder or Avocet or oat (*Avena sativa*) cultivar Swan were grown in Martins Seed Raising and Cutting Mix at 24°C under a 12 h light (~130 µmol m^{-2} s^{-1}) and 12 h dark cycle. At seven days, the first leaf was harvested for protoplast isolation and cells transformed as described [19]. Following a 20h incubation, protoplasts were transferred to 2 mL tubes and left to settle for 30 min. Half of the volume of supernatant was removed, and the protoplasts resuspended in the remainder. 130 µL of each sample was transferred in duplicate to a Black 96-Well Immuno Plate (Thermofisher, Cat. #9502867). YFP fluorescence was measured using a CLARIOstar Plus plate reader (BMG Labtech; the top optic was used, optic settings: presetname = YFP, excitation = 497-15, dichroic filter = 517.2, emission = 540-20, focal height = 5.7mm, scan settings: number of flashes per scan = 4, scan mode = matrix scan, scan matrix dimension = 30x30, scan width = 7mm). The gain was set based on fluorescence scanning of YFP control samples (scan settings as above except: number of flashes per scan = 100, scan mode = spiral average). Fluorescence was normalized against the mean of the YFP control samples for each experiment.

N. benthamiana plants were grown in a growth chamber at 23 °C with a 16-hour light period and used for agroinfiltration at the 4-week-old stage. *AvrYr28* gene variants were cloned into pAM-35S-based vectors via Gateway LR reactions (Invitrogen) for *N. benthamiana* expression. *Agrobacterium tumefaciens* (strain GV3101 pMP90RK) cultures were grown at 28°C overnight with shaking at 200 RPM in Luria-Bertani liquid medium with appropriate antibiotic selections. Cells were pelleted by centrifugation and resuspended to a final concentration of OD_{600} = 0.2 for the *Agrobacterium* strains carrying *R* genes and OD_{600} = 0.5 for those carrying *Avr* genes in 10 mM MES pH 5.6, 10 mM $MgCl_2$, 150 µM acetosyringone and incubated at 20-23.5°C for 2 h before infiltration. Leaves were photographed or scanned 3 to 5 days after infiltration.

Phylogenetic analysis

Illumina data generated for fourteen Australian isolates (Supplementary Table 3) was combined with publicly available genomic and RNAseq sequence data from 96 *Pst* isolates [21, 24, 27, 28, 30, 43, 63-69] (Supplementary Table 4). Genomic data for four isolates were excluded because SNP allele frequency and *k*-mer containment analysis showed signs of sample contamination (Supplementary Table 4). Illumina reads were processed using fastp v.0.23.2 [53]. Genomic reads were aligned to the haplotype-phased reference genomes of isolate Pst104E137 using bwa-mem2v.2.2.1 [70]. RNAseq reads were aligned to the same reference with STAR v.2.7.9a [54]. Samtools [61] was used to prepare alignment files for variant calling with Freebayes v1.3.6 (--use-best-n-alleles 6) [66, 71]. Variants were filtered with vcflib v1.0.1 [72] (-f "QUAL > 20 & QUAL / AO > 10 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1 & AC > 0") and vcftools v0.1.16 [61] (--min-alleles 2 --max-alleles 2 --max-missing 0.9 --maf 0.05). The vcf file was converted to phylip format using vcf2phylip (https://github.com/edgardomortiz/vcf2phylip/blob/master/vcf2phylip.py). Maximum likelihood tree search and bootstrapping were conducted with RAxMLv8.2.12 [73] (options -f -a -m GTRCAT -p 12345 -x 12345 -#500 –no-bfgs).

Whole genome comparison and *k*-mer containment analysis

Whole genome alignment was performed with MUMmer v4.0.0rc1 [74] using default nucmer parameters, and sequence divergence was assessed with dnadiff. Recombination analysis was performed using the pangenome-graph approach described by Henningsen et al. [36] using cactus v2.9.0 [75]. Mashv2.0 [76] was used to measure *k*-mer containment of sketched reference haplotype-phased sequences (-s 100000) in genomic Illumina reads from *Pst* isolates. Isolates with *k*-mer identity of ≥ 99.85% and shared *k*-mers ≥ 97.50% were considered as likely containing the screened haplotype. Haplotype ML trees were generated as described in Henningsen et al (2026); briefly, variants on hap01-Pst104 in the other *Pst* haplotypes from cactus-pangenome were filtered with VCFtools v.0.1.16 (Danecek et al. 2011) (--min-alleles 2 –max alleles 2 –max-missing 0.9 –maf 0.05), the filtered vcf was converted to a phylip alignment with vcf2phylip (Ortiz 2019), and this alignment was input into RAxML v8.2.12 for ML tree estimation and bootstrapping (-f a -m GTRCAT -# 500 –no-bfgs) (Stamatakis 2014). The resulting tree was visualized with iTOL **(**Letunic and Bork 2021).

Variant analysis and structure modelling

Multiple sequence alignment of the AvrYr28 protein and its variants was computed with MUSCLE 5.0.1428 [77] and visualized in Geneious Prime 2026.0.2 (https://www.geneious.com). Protein structures and complexes were predicted using Alphafold 3.0.0 [78] and default parameters. Structures were visualized with PyMOL (https://www.pymol.org).

Declarations

Data availability

All datasets used in this paper are publicly accessible. All raw sequence data are available under NCBI BioProject PRJNA1466082. Assembly files are deposited in the CSIRO data access portal (https://doi.org/10.25919/bqck-s546).

Acknowledgements

This project was supported by the Grains Research and Development Corporation projects CSP2403-014RTX and UOS2207-002RTX, BBSRC grant BB/W018403/1 as part of the NSF/BBSRC Lead Agency Agreement, the UK Agriculture and Horticulture Development Board UKCPVS project 21120034a, the CSIRO Research Office and the Catalysing Australia's Biosecurity Initiative (CAB009_2025). We acknowledge Andrew Milgate (NSW Department of Primary Industries), Grant Hollaway and Jennifer Cutajar (Agriculture Victoria) as well as David Eksteen (Agricultural Consulting) for providing field collected rust isolates from Australia; Alicia Farmer for isolating, pathotyping WYR25-025; Megan Outram (CSIRO) for early discussions on the project; Zhouyang Su

(CSIRO) for technical assistance in protoplast isolation; and Tansy Chia (NIAB) for genotyping SHW-derived lines. We thank the CSIRO High Performance Computing Services for computational resources. We also thank Kerith-Rae Dias and Daina Elliot (Decode) as well as Natalie Habel and Bernd Willens (Twist Bioscience) for their technical support for library synthesis.

Author contributions

Conceptualization: PND, MF

Data curation: JS, ECH

Formal analysis: CLP, JS, ECH, JC

Investigation: CLP, JS, ECH, JC, JL, DL, MC, DS, PK, LCC, RM, MA

Project Administration: PND, MF

Software: JS

Methodology: JS, PND, MF

Visualization: CLP, JS, ECH, JC,

Validation: CLP, JC, PK, DL

Writing – original draft: CLP, JS, PND

Writing – review & editing: CLP, JS, ECH, JC, JL, DL, MC, DS, PK, LCC, RM, EL, KK, MA, MF, PND

Resources: MC, EL, KK, MF, DS

Supervision: CLP, JS, KK, MF, PND,

Funding Acquisition: JS, KK, PND, MF

Competing interests

P.N.D., M.F., and J.S. are inventors on patent application WO2024103117 filed by CSIRO and relating to the identification of protein-protein interactions by protoplast screening. J.S. is an inventor on Australian Provisional Patent Application No.2025201797 filed by CSIRO and relating to gene annotation. The remaining authors declare that they have no competing interests.

References

1. Beddow, J.M. et al. Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants* **1**, 15132 (2015).
2. Hovmoller, M.S., Thach, T. & Justesen, A.F. Global dispersal and diversity of rust fungi in the context of plant health. *Curr. Opin. Microbiol.* **71** (2023).
3. Athiyannan, N. et al. Haplotype variants of the stripe rust resistance gene Yr28 in Aegilops tauschii. *Theoretical and Applied Genetics* **135**, 4327-4336 (2022).
4. Zhang, C.Z. et al. An ancestral NB-LRR with duplicated 3'UTRs confers stripe rust resistance in wheat and barley. *Nature Communications* **10**, 4023 (2019).
5. Figueroa, M., Hammond-Kosack, K.E. & Solomon, P.S. A review of wheat diseases—a field perspective. *Molecular Plant Pathology* **19**, 1523-1536 (2018).
6. Savary, S. et al. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* **3**, 430-439 (2019).
7. Carmona, M., Sautua, F., Pérez-Hérnandez, O. & Reis, E.M. Role of Fungicide Applications on the Integrated Management of Wheat Stripe Rust. *Frontiers in Plant Science* **Volume 11 - 2020** (2020).
8. Chen, R., Gajendiran, K. & Wulff, B.B.H. R we there yet? Advances in cloning resistance genes for engineering immunity in crop plants. *Current Opinion in Plant Biology* **77**, 102489 (2024).
9. Dracatos, P.M., Lu, J., Sanchez-Martin, J. & Wulff, B.B.H. Resistance that stacks up: engineering rust and mildew disease control in the cereal crops wheat and barley. *Plant Biotechnology Journal* **21**, 1938-1951 (2023).
10. Fahima, T. & Coaker, G. Pathogen perception and deception in plant immunity by kinase fusion proteins. *Nature Genetics* **55**, 908-909 (2023).
11. Dodds, P.N. From Gene-for-Gene to Resistosomes: Flor's Enduring Legacy. *Molecular Plant-Microbe Interactions* **36**, 461-467 (2023).
12. Henningsen, E.C. et al. Genome Biology of Rust Fungi. *Annual Review of Phytopathology* **63**, 529-552 (2025).
13. Spanner, R.E. et al. Allelic variation of Avr genes in highly virulent strains explains severe wheat stem rust epidemics. *Nature Communications* **17**, 2718 (2026).
14. Sperschneider, J. et al. EffectorGeneP: accurate effector gene annotation in pathogen genomes from infection transcriptomes. *Nature Microbiology* **in press** (2026).

15. Davis, H.R. et al. First Large-Scale Breakdown of Yr15 Resistance to Wheat Yellow Rust (Puccinia striiformis f. sp. tritici) Recorded Globally. *New Disease Reports* **52**, e70067 (2025).
16. Singh, R.P., Nelson, J.C. & Sorrells, M.E. Mapping Yr28 and Other Genes for Resistance to Stripe Rust in Wheat. *Crop Sci* **40**, 1148-1155 (2000).
17. Wright, T.I.C. et al. A new winter wheat genetic resource harbors untapped diversity from synthetic hexaploid wheat. *Theor Appl Genet* **137**, 73 (2024).
18. Brar, G.S., Dhariwal, R. & Randhawa, H.S. Resistance evaluation of differentials and commercial wheat cultivars to stripe rust (*Puccinia striiformis*) infection in hot spot regions of Canada. *Eur J Plant Pathol* **152**, 493-502 (2018).
19. Arndell, T. et al. Pooled effector library screening in protoplasts rapidly identifies novel Avr genes. *Nature Plants* **10**, 572-580 (2024).
20. Garnica, D.P., Upadhyaya, N.M., Dodds, P.N. & Rathjen, J.P. Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. *PLoS ONE* **8**, e67150 (2013).
21. Ding, Y. 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* **30**, 6566-6584 (2021).
22. Tam, R. et al. Long-read genomics reveal extensive nuclear-specific evolution and allele-specific expression in a dikaryotic fungus. *Genome Research* **35**, 1364-1376 (2025).
23. Palmer, J.M. & Stajich, J.S. Funannotate v1.8.1: Eukaryotic genome annotation (v1.8). *Zenodo*, https://doi.org/10.5281/zenodo.1134477 (2020).
24. Schwessinger, B. 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* **9**, e02275-02217 (2018).
25. Sperschneider, J. & Dodds, P.N. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interactions* **35**, 146-156 (2022).
26. Chen, J. et al. Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. *Science* **358**, 1607-1610 (2017).
27. Schwessinger, B. et al. A Chromosome Scale Assembly of an Australian Puccinia striiformis f. sp. tritici Isolate of the PstS1 Lineage. *Molecular Plant-Microbe Interactions®* **35**, 293-296 (2022).
28. Holden, S. et al. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. *BMC Biology* **21**, 233 (2023).
29. Hovmøller, M.S. et al. Long-term surveillance reveals hybridization by nuclear reassortment and intercontinental spread as major evolutionary drivers in wheat yellow rust. *bioRxiv*, 2025.2005.2022.655633 (2025).
30. Breitkreuz, C. et al. Differentiation of European yellow rust subraces within the 'Warrior(-)' genetic group. *PLOS ONE* **20**, e0323046 (2025).
31. Holden, S. et al. Wheat and barley stripe rust pan-genome facilitates discovery of the predominant North American lineage's origin in somatic hybridization. *bioRxiv*, 2025.2010.2024.684071 (2025).
32. Wang, J. et al. Two ancient nuclear lineages and divergent reproductive strategies drive global success of the wheat stripe rust pathogen. *bioRxiv*, 2026.2003.2018.712333 (2026).
33. Wang, Y., Yin, M. & He, F. Two divergent haploid nuclei shaped the landscape of population diversity in wheat stripe rust. *bioRxiv*, 2024.2012.2010.627719 (2024).
34. Wang, J. et al. A fully haplotype-resolved and nearly gap-free genome assembly of wheat stripe rust fungus. *Scientific Data* **11**, 508 (2024).
35. Sperschneider, J. et al. Nuclear exchange generates population diversity in the wheat leaf rust pathogen Puccinia triticina. *Nature Microbiology* **8**, 2130–2141 (2023).
36. Henningsen, E.C. et al. A high-resolution haplotype collection uncovers somatic hybridization, recombination and intercontinental movement in oat crown rust. *Plos Genetics* **20**, e1011493 (2024).
37. Spanner, R.E. et al. Haplotype-phased genomes of the barley leaf rust pathogen (Puccinia hordei) reveal evidence of mobile element expansion and somatic hybridization. . *BioRXiv* (2026).
38. Brar, G.S. & Kutcher, H.R. Race Characterization of *Puccinia striiformis* f. sp *tritici*, the Cause of Wheat Stripe Rust, in Saskatchewan and Southern Alberta, Canada and Virulence Comparison with Races from the United States. *Plant Disease* **100**, 1744-1753 (2016).
39. Wang, F. et al. Pyramiding of Adult-Plant Resistance Genes Enhances All-Stage Resistance to Wheat Stripe Rust. *Plant Disease* **107**, 879-885 (2022).
40. Yang, X. et al. Development of Wheat Lines Pyramiding the Fusarium Head Blight Resistance Gene Fhb1 with the Stripe Rust Resistance Genes Yr18, Yr28, and Yr36. *Plants* **15**, 790 (2026).
41. Domingues Carvalho, R. et al. Virulence gains in the Puccinia striiformis f. sp. tritici PstS10 lineage correlate with expression polymorphism in a candidate Avr effector. *Communications Biology* (2026).
42. Pereira, E.C. et al. Defence Recognition of a Stripe Rust Fungal Effector Is Uncoupled from Disease Outcomes in Wheat. *Molecular Plant-Microbe Interactions®* (2026).
43. Radhakrishnan, G.V. et al. MARPLE, a point-of-care, strain-level disease diagnostics and surveillance tool for complex fungal pathogens. *BMC Biology* **17**, 65 (2019).
44. Savva, L. et al. A portable, nanopore-based genotyping platform for near real-time detection of *Puccinia graminis* f. sp. *tritici* lineages and fungicide sensitivity. *Bmc Genomics* **26** (2025).
45. Ayliffe, M. et al. Nonhost Resistance of Rice to Rust Pathogens. *Molecular Plant-Microbe Interactions®* **24**, 1143-1155 (2011).
46. Li, F. et al. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nat Commun* **10**, 5068 (2019).

47. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170-175 (2021).
48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
49. Duan, H. et al. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biology* **23**, 84 (2022).
50. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16** (2015).
51. Waterhouse, R.M. et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* **35**, 543-548 (2018).
52. Henningsen, E.C. et al. Haplotype-resolved genomes of diverse oat crown rust isolates reveal both global dispersion of long-lived clonal haplotypes and limited recombination between haplotypes. *Molecular Plant-Microbe Iinteractions* **online** (2026).
53. Chen, S.F., Zhou, Y.Q., Chen, Y.R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884-890 (2018).
54. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
55. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290 (2015).
56. Testa, A.C., Hane, J.K., Ellwood, S.R. & Oliver, R.P. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* **16**, 170 (2015).
57. Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786 (2011).
58. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580 (2001).
59. Dainat, J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. *Zenodo* **https://www.doi.org/10.5281/zenodo.3552717** (2020).
60. Sperschneider, J. & Dodds, P. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Molecular Plant-Microbe Interactions®* (2021).
61. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
62. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
63. Cantu, D. et al. 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* **14**, 270 (2013).
64. Hubbard, A. et al. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biology* **16**, 23 (2015).
65. Bueno-Sancho, V. et al. Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biol Evol* **9**, 3282-3296 (2017).
66. Kiran, K. et al. Dissection of genomic features and variations of three pathotypes of Puccinia striiformis through whole genome sequencing. *Scientific Reports* **7**, 42419 (2017).
67. Xia, C. et al. Genomic insights into host adaptation between the wheat stripe rust pathogen (Puccinia striiformis f. sp. tritici) and the barley stripe rust pathogen (Puccinia striiformis f. sp. hordei). *BMC Genomics* **19**, 664 (2018).
68. Schwessinger, B. et al. Distinct Life Histories Impact Dikaryotic Genome Evolution in the Rust Fungus Causing Stripe Rust in Wheat. *Genome Biol Evol* **12**, 597-617 (2020).
69. Xia, C. et al. Folding Features and Dynamics of 3D Genome Architecture in Plant Fungal Pathogens. *Microbiology Spectrum* **10**, e02608-02622 (2022).
70. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *Int Parall Distrib P*, 314-324 (2019).
71. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.3907 (2012).
72. Garrison, E. Vcflib, a simple C++ library for parsing and manipulating VCF files. . *https://github.com/vcflib/vcflib.* (2016).
73. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
74. Marçais, G. et al. MUMmer4: A fast and versatile genome alignment system. *Plos Computational Biology* **14** (2018).
75. Hickey, G. et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology* **42**, 663-673 (2024).
76. Ondov, B.D. et al. Mash Screen: high-throughput sequence containment estimation for genome discovery. *Genome Biology* **20** (2019).
77. Edgar, R.C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications* **13**, 6968 (2022).
78. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature***630**, 493-500 (2024).

Figures

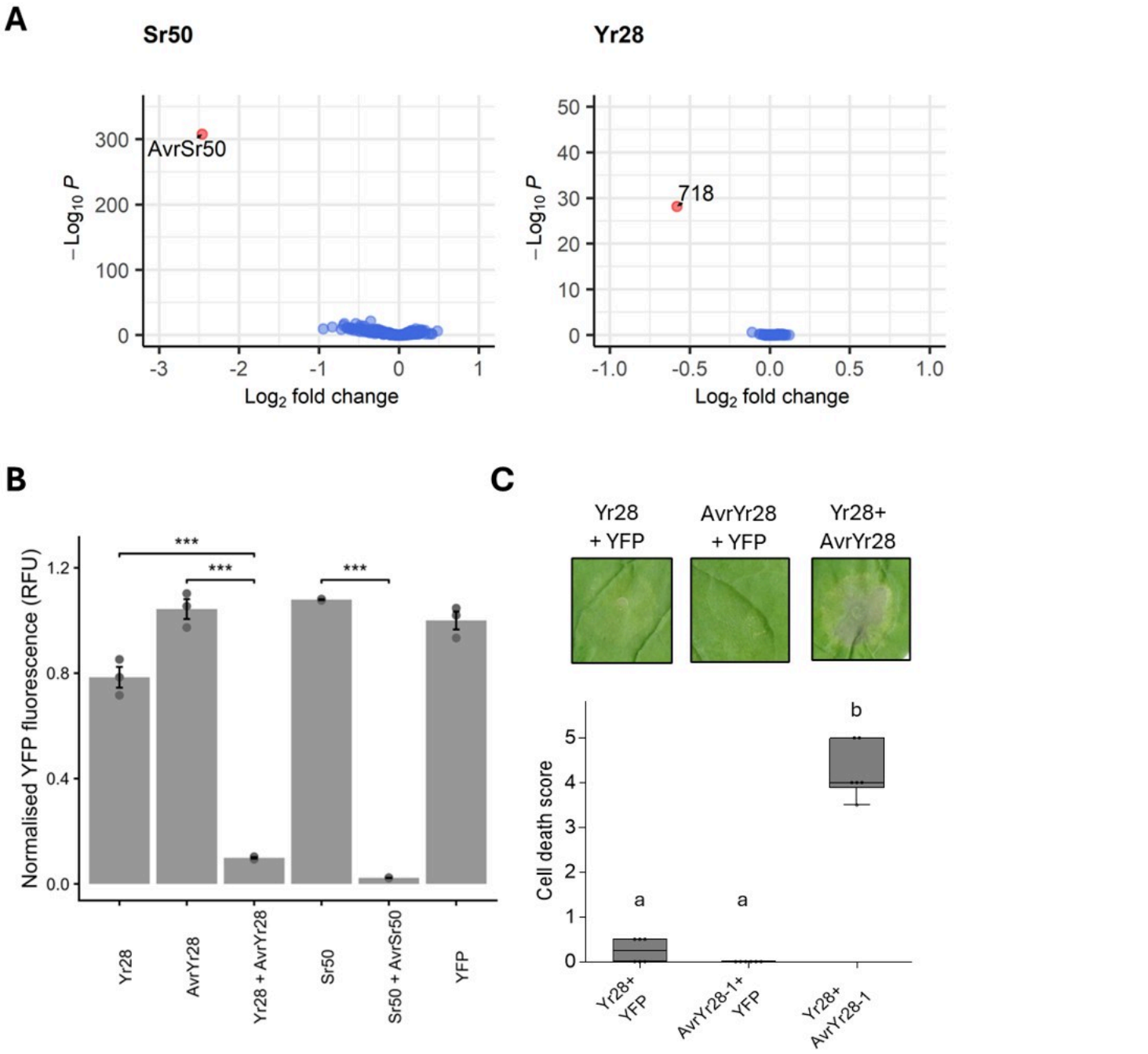

Figure 1

AvrYr28 identification. (**A**) A library of 307 candidate effector genes (>142 aa) was screened against the stem rust resistance gene *Sr50* (left) and stripe rust resistance gene *Yr28* (right) by co-expression in wheat protoplasts (cv Fielder). Graphs show differential gene expression (x-axis) versus adjusted P-value (y-axis) for each effector construct (blue dots). The positive candidate effector (clone #718) detected in the *Yr28* screen is indicated in red as is the *AvrSr50* control. (**B**) *Yr28* and *AvrYr28*genes were expressed singly or in combination in wheat protoplasts (cv Fielder) and fluorescence of a co-transformed YFP reporter gene measured (y-axis). *Sr50*and *AvrSr50* were included as a positive control for cell death and YFP alone as a control for transformation efficiency. Reduced YFP fluorescence when corresponding *R*and *Avr* genes are co-expressed is indicative of immune recognition resulting in cell death. *** indicates P<0.001 (t-test). (**C**) *Yr28*and *AvrYr28* were expressed singly or in combination in *N. benthamiana*leaves by agro-infiltration. Images were recorded after 3 days to observe cell death induced by *R-Avr* recognition (upper panel). Lower panel shows box plot of cell death scores from six infiltration sites per treatment. Different letters on columns represent statistically significant differences (ANOVA, Tukey's test, P < 0.05).

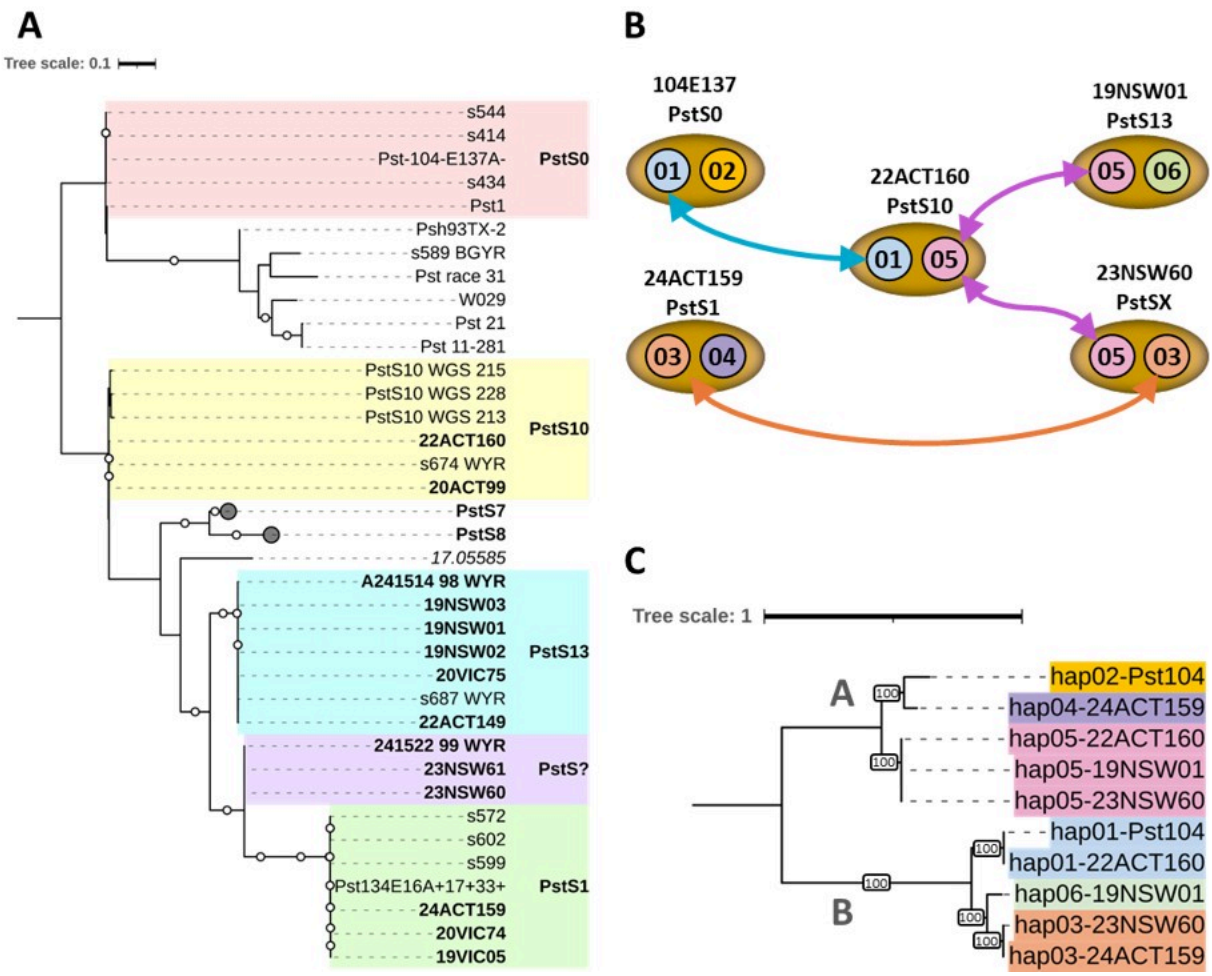

Figure 2

Genomic relationships between *Pst* strains. **A.** Maximum Likelihood (ML) tree of *Pst* isolates based on whole genome SNPs (147,683) called from Illumina sequence data from 106 isolates against the Pst104 genome. Branches with bootstrap support (500 cycles) ≥ 80% are indicated by white circles at branch midpoints. Collapsed clades are indicated with a circle at the branch tip and the full tree is shown in Supplementary Figure S2. Isolates belonging to named clonal lineages PstS0, PstS1 etc are labelled in coloured boxes. Isolates sequenced in this study are labelled with bold text and an RNAseq sample is labelled with italicized text. The tree is mid-point rooted. Tree scale is average nucleotide substitutions per site. **B.** Schematic showing nuclear genome content of five *Pst* isolates representing different lineages (PstS0, PstS1, PstS10, PastS13 and PstSX). Individual haploid nuclei in each strain are numbered (01 to 06) and colour coded according to identity. **C.** Maximum Likelihood (ML) phylogenetic tree based on whole genome alignment of haplotypes hap01 to hap06 from five *Pst* isolates generated by RAxML. The tree is mid-point rooted and scale represents mean substitutions per site. Bootstrap support values (500 cycles) for each branch are indicated.

Supplementary Files

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

- AvrYr28manuscriptsupplementaltables.xlsx
- SupplementaryFigures.docx