Quantitative pathogenicity and host adaptation in a fungal plant pathogen revealed by whole-genome sequencing

Reda AMEZROU*[1], Aurélie DUCASSE[1], Jérôme COMPAIN[2], Nicolas LAPALU[1,2], Anais PITARCH[1], Laetitia DUPONT[1], Johann CONFAIS[1], Henriette GOYEAU[1], Gert HJ KEMA[3], Daniel CROLL[4], Joëlle AMSELEM[2], Andrea SANCHEZ-VALLET[5], Thierry C MARCEL*[1]

[1] Université Paris-Saclay, INRAE, UR BIOGER, Palaiseau, France
[2] Université Paris-Saclay, INRAE, UR URGI, Versailles, France
[3] Plant Research International B.V., Wageningen, The Netherlands
[4] Université de Neuchâtel, Department of Ecology and Evolution, Neuchâtel, Switzerland
[5] CBGP, INIA, Campus de Montegancedo UPM, Pozuelo de Alarcón, Madrid, Spain

*** Corresponding authors:** thierry.marcel@inrae.fr; reddamez@gmail.com

Abstract

Knowledge of genetic determinism and evolutionary dynamics mediating host-pathogen interactions is essential to manage fungal plant diseases. Studies on the genetic architecture of fungal pathogenicity often focus on large-effect effector genes triggering strong, qualitative resistance. It is not clear how this translates to predominately quantitative interactions. Here, we use the *Zymoseptoria tritici*-wheat model to elucidate the genetic architecture of quantitative pathogenicity and mechanisms mediating host adaptation. With a multi-host genome-wide association study, we identify 19 high-confidence candidate genes associated with quantitative pathogenicity. Analysis of genetic diversity reveals that sequence polymorphism is the main evolutionary process mediating differences in quantitative pathogenicity, a process that is likely facilitated by genetic recombination and transposable element dynamics. Finally, we use functional approaches to confirm the role of an effector-like gene and a methyltransferase in phenotypic variation. This study highlights the complex genetic architecture of quantitative pathogenicity, extensive diversifying selection and plausible mechanisms facilitating pathogen adaptation.

Introduction

Fungal diseases cause major damage to crop production and threaten food security worldwide[1]. Understanding the molecular dialogue between pathogens and their hosts is essential to design durable and effective control strategies. Major molecular factors in the ability of fungal pathogens to cause disease are effectors. Those are proteins delivered into the host apoplast or translocated inside cells to manipulate host physiology or suppress its immunity to favour infection[2]. In turn, plants respond to pathogen invasion by activating defence mechanisms based on recognition of pathogen-associated

38 molecular patterns and effectors by cell surface and intracellular receptors. The discovery of resistance
39 genes (*R*) that encode receptors has provided potential means to control fungal diseases through
40 resistance breeding[3]. However, pathogens have repeatedly overcome host resistance by evading
41 recognition or by suppression of host immunity. These rapid adaptations are mainly driven by genetic
42 variation at pathogenicity genes and genome evolution[4]. Gene diversification, deletions or horizontal
43 acquisitions have generated adaptive variants at loci encoding effectors. Transposable elements (TEs),
44 high mutation and recombination rates are thought to contribute considerably to the extensive genome
45 variation in many fungal species[5-8].
46 In the case of antagonistic host-pathogen interactions, the fate of genetic variation is determined partly
47 by the selection pressure exerted by host *R* proteins. This is especially true when *R* genes and effectors
48 are involved in a gene-for-gene (GFG) relationship[9-12]. Moreover, the diversity of host genotypes in *R*-
49 gene content exerts heterogeneous selection pressure on pathogen populations, giving rise to the co-
50 maintenance of multiple pathogen variants better adapted to specific subsets of hosts[3,13]. Hence, in
51 addition to the identification of effector genes, elucidating the mechanisms contributing to gene
52 diversification provides critical information about their evolvability. This knowledge has tremendous
53 applications for the durability of crop protection strategies[14].
54 The first fungal effectors to be identified were small secreted proteins (SSP). SSPs trigger strong
55 resistance (*i.e.* encoded by avirulence genes) as a result of their recognition by plant immune receptors
56 following a typical GFG relationship[15]. The increasing availability of fungal genomes over the past
57 decade allowed significant advances in investigating molecular plant–fungal interactions and facilitated
58 effector discovery. However, effector biology has focused almost exclusively on SSPs despite the
59 growing evidence that other genetic factors are also involved in host colonization[16]. For instance, small
60 noncoding RNAs and fungal secondary metabolites play similar effector roles in various plant
61 pathogenic fungi[17]. Other fungal proteins were shown to interact with the host immune system or act to
62 promote pathogenicity; these include for example cell wall–degrading enzymes[18], protease inhibitors[19]
63 or disruptors of the hormone-signalling pathway[20].
64
65 Effector discovery also led to the systematic search for host genes conferring resistance[15]. Resistance to
66 many fungal diseases can be divided into two categories: (*i*) qualitative (or race-specific) resistance,
67 which is usually controlled by large-effect genes and is effective against avirulent strains; and (*ii*)
68 quantitative (or non–race-specific) resistance, usually controlled by genes with small-to-moderate
69 effects and is thought to be less specific. While cloning plant immune receptors remains challenging,
70 the discovery of the interacting fungal effectors is considered more straightforward[14]. However, fungal
71 pathogenicity studies often focus on large-effect genes, whose interaction with genetic factors from the
72 host trigger a strong immune response, leading to qualitative resistance. It remains unclear how these
73 findings translate to quantitative models, where the interaction of quantitative components of
74 pathogenicity (also called aggressiveness in the plant pathology literature) and plant resistance, trigger

a partial/incomplete immune response. While the quantitative components of pathogenicity are frequently measured in various plant-pathogen interactions[21], there have been only few reports on the genetic basis of quantitative pathogenicity in pathogens and the process of adaptation to quantitative host resistance[22,23]. Thus, elucidating the biology of quantitative pathogenicity factors and their evolutionary dynamics is timely and may require the adoption of complementary experimental approaches. To address this, we used the *Zymoseptoria tritici* – wheat pathosystem, which is known to be largely quantitative *in natura*.

Zymoseptoria tritici is an ascomycete filamentous fungus that causes the Septoria tritici blotch (STB) disease on wheat. *STB* is a major threat to wheat crops in Europe and in temperate regions around the world[24]. *Z. tritici* is hemibiotrophic and remains exclusively apoplastic through its infection cycle, which is characterized by an asymptomatic phase of 7 to 14 days, followed by an induction of leaf necrosis[25]. This stage is accompanied by the development of pycnidia containing asexual spores that further spread the disease[26,27]. To date, three *Z. tritici* pathogenicity genes that encode avirulence factors have been cloned. These include *AvrStb6* and *AvrStb9* that are recognized by the *R* genes *Stb6* and *Stb9*, respectively, and lead to a strong qualitative resistance response[12,28,29], and *Avr3D1* which triggers quantitative resistance on cultivars harbouring *Stb7* or *Stb12*[10]. In addition to the above-mentioned *R* genes, 18 other race-specific *Stb* resistance genes and numerous quantitative resistance loci have been genetically mapped in wheat[30], but only *Stb6* and *Stb16q* have been cloned to date[31,32]. An ideal approach to identify genetic determinants of complex traits is through Genome-Wide Association Studies (GWAS). GWAS aims to associate sequence polymorphisms to phenotypic variation, and relies on linkage disequilibrium (LD) –the non-random association of alleles at different loci– to identify causal variants. The approach has been used extensively in plants but remains at its infancy on fungal populations[33,34].

The aim of the present work was to characterize the genetic determinants underlying quantitative pathogenicity in *Z. tritici* and to investigate the evolutionary processes that could facilitate pathogen adaptation to host resistances. We hypothesized that quantitative pathogenicity relies mostly on gene-for-gene interactions between host and pathogen, and that the genes involved are not all encoding SSPs but also genetic factors from the non-secretory pathway. To test our hypotheses, we performed an extensive genome wide association study combined with transcriptomics and population genomics. We identified candidate genes putatively mediating quantitative pathogenicity and confirmed the role of two of these genes using functional approaches. Together, our results generate a comprehensive view on the complex genetic architecture of quantitative pathogenicity and pathogen adaptation to host resistances. Beyond this model, our findings may have broad implications for sustainable disease management.

Results

Characterization of the fungal collection. To efficiently perform GWAS and reduce confounding effects, it is important to select a population that is genetically and phenotypically diverse with little

112 stratification. Following these guidelines, we focused on a subset of 103 French *Z. tritici* isolates selected
113 from a larger population for their genetic diversity. We sequenced their genomes and generated a total
114 of 247 Gb of sequence data with an average depth of 61x. After quality assessment, we obtained a total
115 of 1,463,638 Single Nucleotide Polymorphisms (SNPs) that we further filtered for Minor Allele
116 Frequency (MAF) and missing rates, resulting in a final set of 718,810 SNPs. Analysis of functional
117 variant effect revealed that 59,318 SNPs (8.25%) induced missense mutations and 2,493 (0.3%) resulted
118 in a loss of function such as premature stop codons, loss of stop or start codons (Table 1). As expected,
119 principal component analysis (PCA) using full polymorphism data did not show any genetic structure;
120 the first three PCs explained less than 4% of the total variance (Supplementary Fig. 1). We further
121 confirmed this with STUCTURE using a subset of evenly distributed SNPs (1 SNP per 5 kb) and like
122 the PCA analysis, Bayesian clustering failed to infer sub-populations (Fig. 1A). Similarly, individual
123 isolates showed very low genetic relatedness (Supplementary Fig. 2). Genome-wide LD decayed
124 (r^2<0.2) over physical distance at ~0.5 kb (Fig. 1B). Given our high SNP density (1 SNP every 55 bp)
125 we expected marker-trait associations to be resolved to the individual gene level. When pathotyped on
126 wheat differential cultivars, isolates showed a wide virulence spectrum depending on the host cultivar,
127 and strong isolate × cultivar interactions (Supplementary Tables 1, 2). Overall, strains isolated from the
128 cultivar 'Premio' were more aggressive compared to those isolated from the susceptible cultivar
129 'Apache', even when tested on the cultivar 'Apache' (Supplementary Fig. 3).

130 **Table 1 |** Functional variants identified in a population of 103 *Z. tritici* isolates

Functional variant effect	Count	Percent
5' UTR	10431	1.45
Gain of start codon	1084	0.15
Loss of start codon	127	0.02
Synonymous coding	162938	22.67
Non-synonymous coding	59318	8.25
Intron	58217	8.10
Gain of stop codon	1109	0.15
Loss of stop codon	173	0.02
3' UTR	8794	1.22

131

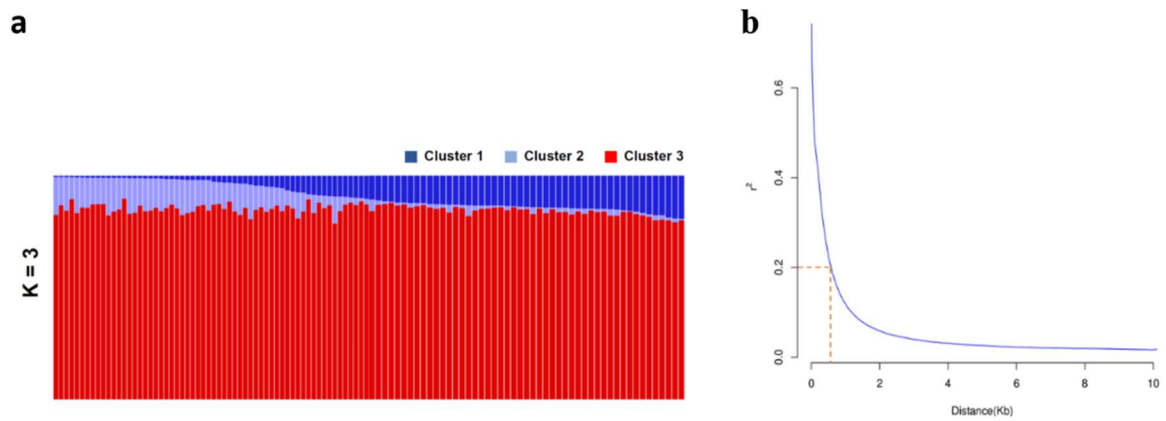


Fig. 1 | Population structure and Linkage Disequilibrium (LD) in the *Z. tritici* population (n=103). **a.** Admixture plot of the fungal population. Each vertical bar represents a single isolate and is colored according to the membership coefficient (Q_i) to the three sub-population (K) clusters identified by STRUCTURE. **b.** LD decay over physical distance. The r^2 values were calculated between pairs of SNPs up to a physical distance of 20 kb and fitted against the physical distance using a non-linear model. The dotted orange line shows the distance (0.5 kb) at which LD decays at the $r^2 < 0.2$ level.

A complex genetic architecture and host-specificity underlies pathogenicity. To elucidate the genetic basis of the observed pathogenicity differences, we investigated two key quantitative phenotypes (PLACN and PLACP) using mixed linear models for marker-trait association tests, performed independently on each of the studied cultivars. Collectively, we identified 94 genome-wide significant SNPs and an additional 53 at the 10% False Discovery Rate (FDR) threshold, defining 58 distinct genes (Fig. 2, Supplementary Data 1). The strongest associations were at the *AvrStb6* and *AvrStb9* loci and were detected in all *Stb6*- and *Stb9*- containing cultivars, respectively. Initially not thought to harbour *Stb9*, PLACN on cultivar 'Cadenza' was highly associated (P=3.57e-13) with polymorphisms in *AvrStb9*, suggesting that the cultivar possesses the corresponding *R* gene. The presence of *Stb9* in 'Cadenza' was recently confirmed by its tightly linked genetic markers. Interestingly, the single-locus model failed to detect *AvrStb9* and demonstrates the usefulness of the multi-locus mixed model in detecting loci with larger effects. However, if multiple SNPs were in strong LD, biases may be created while estimating the genetic variance. Nevertheless, the multi-locus approach in our study generally outperformed the traditional single-locus tests (Supplementary Fig. 4). Candidate pathogenicity loci were identified in the thirteen *Z. tritici* core chromosomes. Despite recent suggestions of accessory chromosomes being enriched in putative effectors[35], we could not map any gene residing on accessory chromosomes in this multi-host GWAS (Fig. 2, Supplementary Data 1). Only eleven genes were shared among PLACP and PLACN (Fig. 2, Supplementary Data 1), which was expected, since necrosis coverage and pycnidia production levels were highly variable (Supplementary Fig. 3). Hence, considering both traits in association analyses could reveal genetic determinants influencing both, the conidial production and host susceptibility. The phenotypic variance explained by the identified loci

associated with pathogenicity traits on each cultivar ranged from 9.1% to 65.7% (Supplementary Data 1). With the exception of *AvrStb6* and *AvrStb9*, all of the detected genes accounted for small-to-moderate effects (R^2<30%) and similarly, analysis of virulent/avirulent alleles showed quantitative differences (Supplementary Data 1, Fig. 3B). Further, the number of associated genes varied greatly between cultivars (three candidates for 'Bulgaria-88' and 'Toisondor' *vs* ten for 'Israel-493') and ~76% were specific to a single host cultivar (Supplementary Data 1, Fig. 3A). This weak overlap observed between genes is a strong indication of host-specificity, in consistence with the different *Stb* genes or *Stb* combinations harboured by the cultivar set used here.

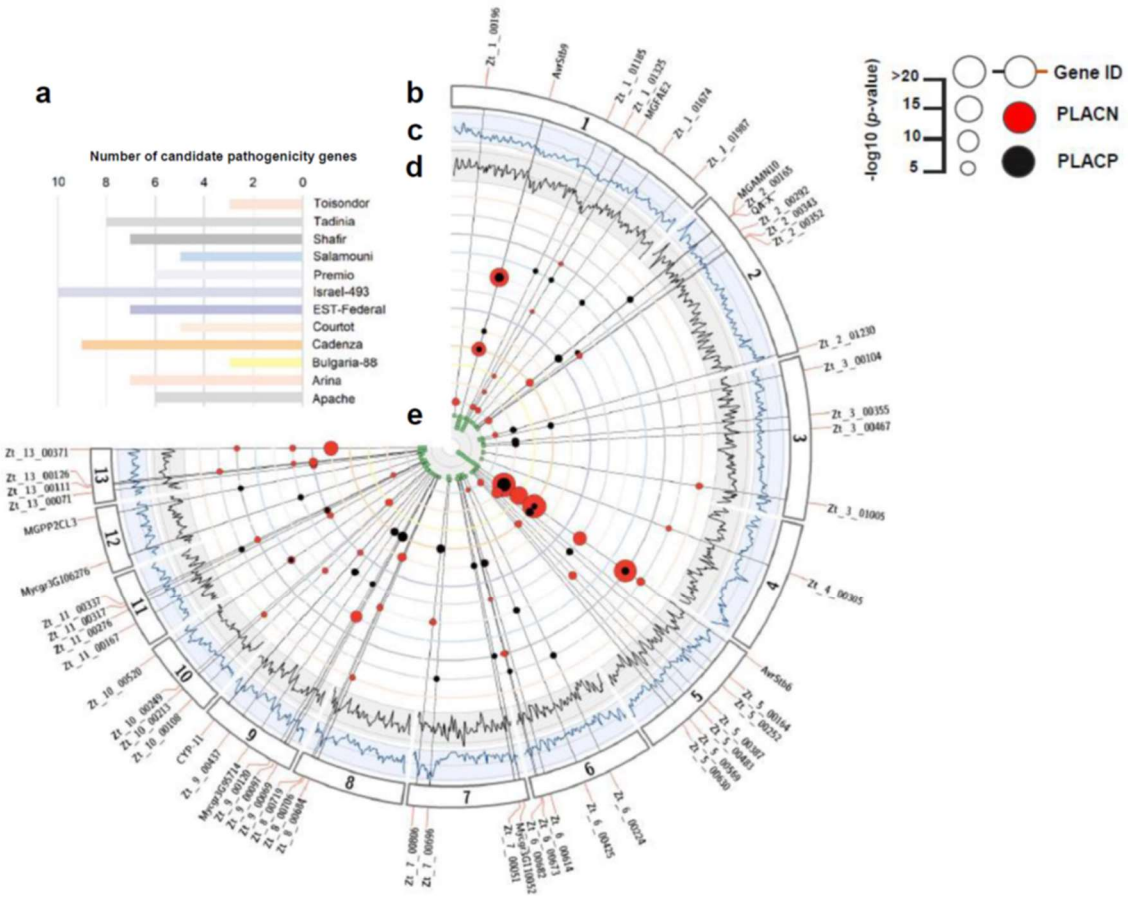
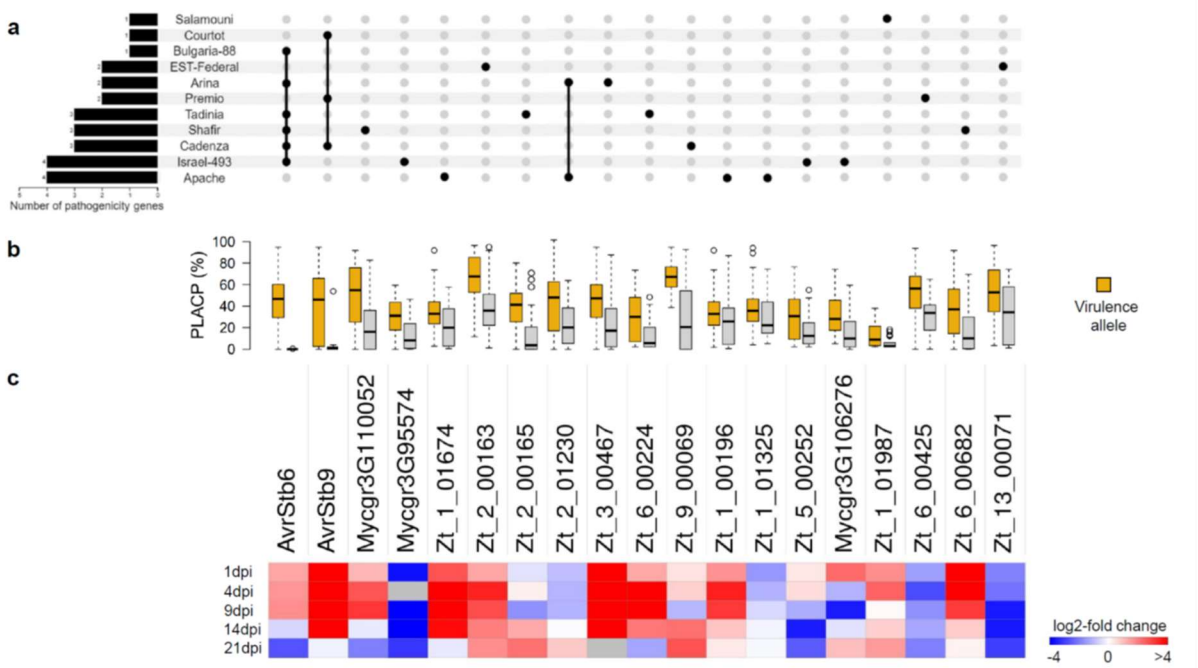


Fig. 2 | Overview of the identified candidate pathogenicity genes from the multi-host GWAS. **a.** Number of identified genes per cultivar. **b.** Gene IDs and their significant associations. Black dots correspond to PLACP (percentage of leaf area covered by pycnidia) and red dots to PLACN (percentage of leaf area covered by necrosis) marker-trait associations. The larger the dot the stronger the association (*P*-value). Radial lines connect dots to the cultivar where it was detected and to the gene ID. **c.** Nucleotide diversity (π) estimates in a 50 kb sliding window with a step size of 25 kb. Genomic positions are displayed on the x-axis and π values are displayed on the y-axis. The y-axis ranges from 0 to 0.045. **d.** Tajima's D estimates in 50 kb sliding windows with a step size of 25 kb. Genomic positions are displayed on the x-axis. Tajima's D values are displayed on the y-axis. The y-axis ranges from -2.5 to 2. **e.** The number of associations for each gene (PLACN/PLACP per cultivar).

180 ***In planta* gene expression and functional annotation.** To reveal the expression profile of the identified
181 candidate pathogenicity genes using GWAS, we calculated the relative abundance of *Z. tritici*
182 transcripts, using RNAseq data generated during different *Z. tritici* infection time courses[25]. Of the 58
183 GWAS candidates, ten genes exhibited no expression (FPKM = 0) or very low expression (FPKM < 5)
184 throughout the infection cycle, while another ten genes were found to be highly expressed (FPKM above
185 the 95th percentile of the mean expression value) either at specific time points or throughout the infection
186 cycle (Fig. 3C, Supplementary Data 2). Interestingly, *AvrStb6*, *Zt_6_00224* and *Zt_3_00467* were
187 among the most highly expressed *Z. tritici* genes at the disease transitional phase (9 dpi; Supplementary
188 Data 2). Differentially expressed genes *in planta* were identified by examining the ratio of expression
189 between the Czapek-Dox broth (CDB) culture medium and the five time points of wheat infection. In
190 our set of candidate genes, sixteen were differentially expressed with a FDR below 0.10 and a log2-fold
191 change greater than 2. Among them, eleven were up-regulated at specific time points of the infection
192 (Fig. 3C; Supplementary Data 2). For subsequent investigations, we considered a set of nineteen genes
193 as high-confidence candidates, whose expression profiles supported their potential role in pathogenicity;
194 this included the sixteen differentially expressed genes and three additional ones that were highly
195 expressed during the infection (Fig. 3; Supplementary Data 2).



196

197 **Fig. 3 |** The quantitative nature and host-specificity of *Z. tritici* pathogenicity. **a.** An upset plot of candidate
198 pathogenicity genes across host cultivars. **b.** Boxplots showing variation in quantitative pathogenicity
199 (PLACP) of the isolates (n=103 isolates) carrying different lead SNP alleles of the candidate pathogenicity
200 genes (centre line at the median, upper bound at 75th percentile, lower bound at 25th percentile, with whiskers
201 chosen to show the 1.5 of the interquartile range). **c.** Differential gene expression shown as log2-fold change
202 compared to the culture medium. Red shades indicate up-regulated genes and blue shades indicate down-
203 regulated genes during the infection cycle.

204 Eleven protein products out of the nineteen retained as best candidates were predicted to be secreted
205 (Supplementary Table 4). These secreted proteins could be effectors promoting disease or interacting
206 with the cognate *Stb* resistance genes analysed in this study. We also found three proteins with predicted
207 transmembrane helices (THMs), a class that is systematically discarded in effector discovery studies.
208 EffectorP predicted seven as putative effectors, with five (excluding *AvrStb6*) lacking homology with
209 other fungal species and might be considered as *Z. tritici*-specific candidate effectors (Supplementary
210 Table 4). The identified conserved protein domains included functions such as hydrolase (*e.g.*, Glycosyl
211 hydrolase), oxidoreductase (*e.g.*, Thioredoxin reductase) or transferase activities (*e.g.*,
212 Methyltransferase; Supplementary Table 4). Using PHIB-Blast, we found that seven candidates shared
213 homology with known pathogenicity factors from other pathogenic fungi and therefore constitute
214 promising candidates mediating quantitative pathogenicity in *Z. tritici* (Supplementary Table 4).

215 **Factors of genetic diversity in pathogenicity genes.** Gain in virulence and/or escape from host
216 recognition is often mediated by polymorphisms in pathogenicity genes. We investigated sequence
217 variation and diversity in the identified genes and detailed results are summarized in Supplementary
218 Table 5. In particular, we compared patterns of nucleotide diversity in candidate pathogenicity genes
219 with the genome-wide levels. We observed higher genetic diversity in pathogenicity genes compared to
220 other protein-coding genes (median $\pi_{Pathogenicity}$=0.0212 vs. $\pi_{Genome\text{-}wide}$=0.0079; $P < 0.01$; Fig. 4A). We
221 found no significant impact of loss of function mutations or presence/absence polymorphisms, except
222 for *Mycgr3G110052*. In addition, we wanted to examine which mechanisms were likely to contribute to
223 the high sequence diversity in our candidates. As TE dynamics have been observed to contribute to
224 fungal genome plasticity and diversity, we calculated the proximity of candidate pathogenicity genes to
225 the nearest TEs. 52.6% of these genes were flanked by or adjacent to TEs (*i.e.* at less than 5 kb distance;
226 Supplementary Table 5). Comparison with genome-wide genes with regards to TE proximity revealed
227 that candidate pathogenicity genes were significantly closer to TEs ($P < 0.05$; Fig. 4B). Gene
228 diversification is also accelerated by meiotic recombination, which influences both natural population
229 dynamics and selection. We used a combination of algorithms in two independent software (RDP4 and
230 GARD) to detect signatures of recombination. We found signatures of recombination supported by both
231 detection methods in a significant proportion (63.2%) of the putative pathogenicity genes
232 (Supplementary Table 6). Interestingly, genes with low or no signatures of recombination were in close
233 proximity to TEs, whereas those with higher recombination events were found at greater distances from
234 TEs (Fig. 4C). We hypothesize that recombination in particular and proximity to TEs promotes allelic
235 diversity in genes that are putatively involved in quantitative pathogenicity.

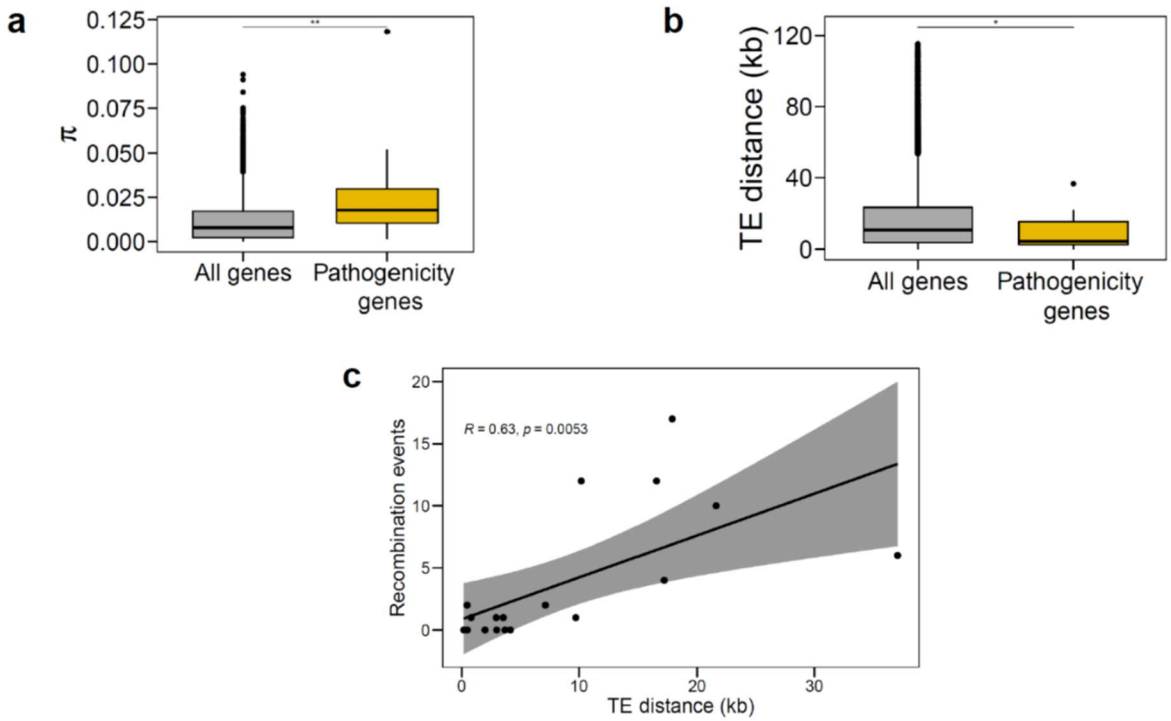


Fig. 4 | Genetic diversity of quantitative pathogenicity genes in *Z. tritici* and factors contributing to gene diversification. **a.** Distribution of nucleotide diversity (π) among candidate pathogenicity genes (n=58) and the rest of the protein-coding genes (n=10,993). Statistical significance was determined using a permutation test with 1000 iterations; *P*=0.001818**. **b.** TE distance compared between pathogenicity genes (n=58) and the rest of the protein-coding genes in *Z. tritici* (n=12,012). Statistical significance was determined using a permutation test with 1000 iterations; *P*=0.04683*. In a and b data are presented as box plots (centre line at the median, upper bound at 75th percentile, lower bound at 25th percentile) with whiskers chosen to show the 1.5 of the interquartile range. **c.** The relationship between TE distance and recombination events in candidate pathogenicity genes. The error band indicates the 95% confidence interval.

Pathogenicity genes are hotspots of positive diversifying selection. We assumed that candidate pathogenicity genes may have evolved under *R*-genes selective pressure on pathogen populations to escape recognition. Therefore, we searched for episodes of positive diversifying selection on protein-coding sequences by using the ω (d_N/d_S) parameter. Codon sites under positive selection are defined as those with elevated nonsynonymous/synonymous substitution ratios (ω > 1) compared to the expectation under neutral evolution, ω = 1. The proportion of positively-selected sites ($\omega > 1$) varied greatly among genes and ranged from 0.42% in the case of *Zt_1_01987* to 10.97% for *AvrStb6* (Supplementary Table 5). We also found four genes with a $d_N/d_S > 1$ (*Zt_3_00467*, ω=7.03; *AvrStb6*, ω=2.84; *Mycgr3G106276*, ω=1.18 and *AvrStb9*, ω=1.04), that could be considered as genes under strong selection pressure. Since ω averaged across all gene-tree branches rarely surpasses 1, and as the selective pressure could affect codon sites differently, we tested different codeML models to detect positively-selected codons. For all model comparisons, we found signatures of positive diversifying selection in 68.4% of candidate genes using both, the more conservative M1a vs. M2a and the M7 vs. M8 model

259 pairs (Supplementary Table 5). Together, this shows that pathogenicity genes are hotspots of positive
260 diversifying selection, likely as a result of a selection pressure exerted by host *R* genes.

261 **Two functionally distinct genes mediate quantitative pathogenicity.** To determine the role of the
262 candidate genes in quantitative pathogenicity, we selected two distinct genes for functional validation:
263 *Zt_3_00467* (Fig. 5A), which encodes a SSP that is highly expressed *in planta*, with a peak at the disease
264 transition phase (Fig. 5B); and *Zt_6_00682* (Fig. 6A), which encodes a S-adenosylmethionine-
265 dependent methyltransferase (SAM-MTase) that is differentially expressed during the disease
266 asymptomatic phase (Fig. 6B). For *Zt_3_00467*, we ectopically expressed the avirulent (IPO10273) and
267 virulent (3D7) alleles in the 3D7 strain background and assessed quantitative pathogenicity traits on the
268 host cultivars 'Arina' and 'Riband'. Overall, Mutants expressing the avirulent allele (*3D7* +
269 $Zt_3_00467_{IPO10273}$) showed a slight reduction of PLACP levels compared to those expressing the
270 virulent allele (*3D7* + $Zt_3_00467_{3D7}$; Fig. 5C). When compared individually, one mutant expressing
271 the avirulent allele showed a phenotype similar to the wild-type, while another expressing the virulent
272 allele had higher PLACP compared to the wild-type (Supplementary Fig. 7). The difference in
273 phenotypic responses between ectopic mutants is a well-known phenomenon that may be caused by
274 distinct allele integration sites in the different mutant lines. Testing transformants expressing the
275 avirulent allele on the cultivar 'Riband' led to lower development of pycnidia compared to the virulent
276 and the wild-type strains (Supplementary Fig. 8). 'Arina' and 'Riband' share a common resistance gene
277 known as *Stb15*. These phenotypic observations suggest that *Zt_3_00467* recognition leads to
278 quantitative resistance which could be attributed to *Stb15*. Comparison of the avirulent and virulent
279 protein sequences showed seven amino acid substitutions, all positively selected (Fig. 5D). Analysis of
280 *Zt_3_00467* in the global *Z. tritici* collection showed that the gene was conserved and that most protein
281 variants were population specific and segregating by geographical origin (Supplementary Fig. 9).

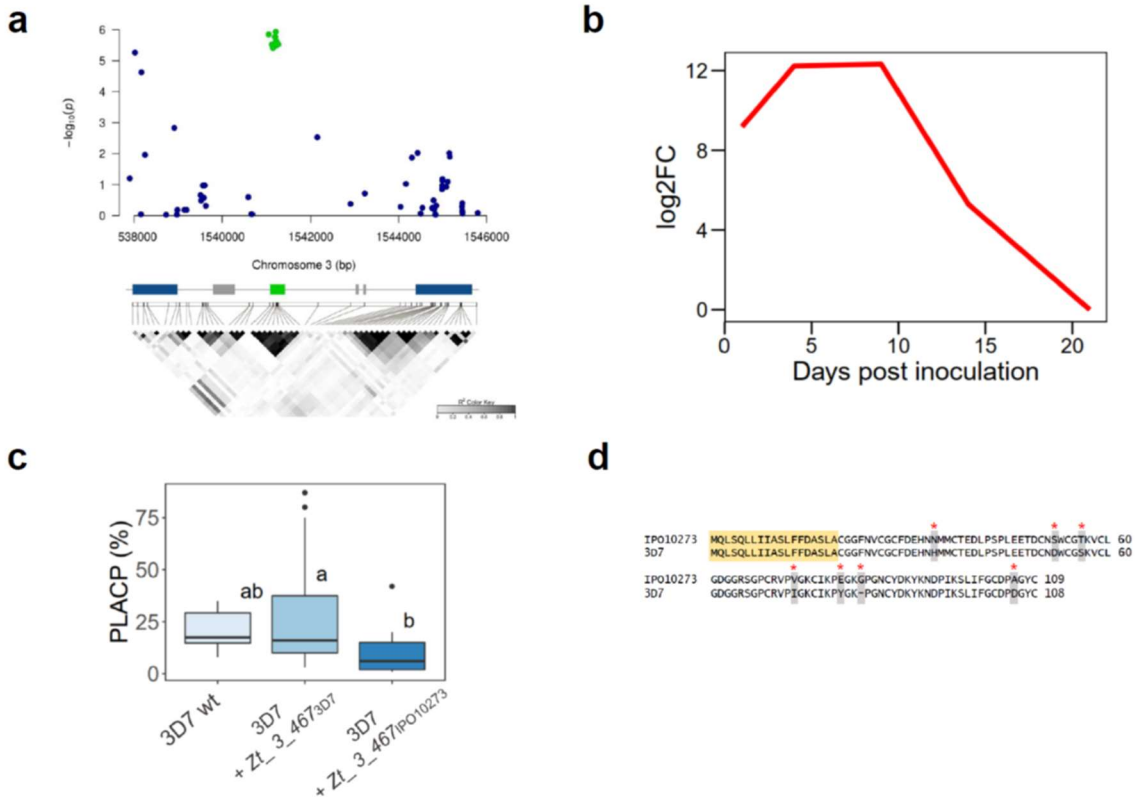


Fig. 5 | An effector gene encoding a positively-selected small secreted protein is involved in quantitative pathogenicity of *Z. tritici*. **a.** Regional association plot of a candidate effector protein detected on cultivar ''Arina''. X and Y axes indicate positions on chromosome 6 and −log10 (*P*-value) for associations with PLACP, respectively. Annotations on the reference genome are shown at the bottom with the candidate gene represented by a green box, other genes by blue boxes and transposable elements by grey boxes. A linkage disequilibrium plot is shown below the annotation bar. **b.** log2-fold expression changes of *Zt_3_00467* based on RNAseq data collected throughout the time course of *Z. tritici* infection on wheat. **c.** Percentage of leaf area covered by pycnidia (PLACP) produced by the wild-type strain 3D7 (n=6 leaves) and the ectopic transformants expressing the virulent (3D7) and avirulent (IPO10273) alleles (n=18 leaves each). Data are presented as box plots (centre line at the median, upper bound at 75th percentile, lower bound at 25th percentile) with whiskers chosen to show the 1.5 of the interquartile range. Significant difference between group means are represented with different letters above the boxplots after a Tukey's HSD test. **d.** Alignment of the virulent and avirulent protein sequences. Residues in yellow constitute the predicted signal peptide, grey residues show protein mutations and red asterisks show positively selected sites with a Bayes Empirical Bayes posterior probability > 0.95.

For *Zt_6_00682*, we generated knockout (KO) mutants of the 3D7 strain by targeted gene disruption and ectopic transformants with the virulent (3D7) and avirulent (3D1) alleles in the KO background. This allowed us to assess allele-dependent effects of *Zt_6_00682.* Three independent KO and ectopic transformants for each allele were tested on the cultivar 'Shafir', where the association was detected. Compared to the wild-type, KO transformants resulted in unexpectedly high PLACP values ranging from 67% to 100% (Fig. 6C). Ectopic transformation of the 3D1 allele in the KO background

(3D7Δ*Zt_ 6_00682*+ *Zt_ 6_00682*$_{3D1}$) resulted in reduced levels of PLACP, while ectopic lines transformed with the 3D7 allele (3D7Δ*Zt_ 6_00682*+ *Zt_ 6_00682*$_{3D7}$) showed a slightly higher PLACP than the wild-type (Fig. 6C). Besides 'Shafir', the deletion mutants were aggressive on another four cultivars, but phenotypic differences between ectopic transformants and wild-type were not significant (Supplementary Fig. 6). Five amino acid differences were found between the avirulent and virulent protein sequences, of which three were under positive selection (Fig. 6D). Like *Zt_3_00467, Zt_ 6_00682* was conserved in the global population but was highly polymorphic (Supplementary Fig. 9).

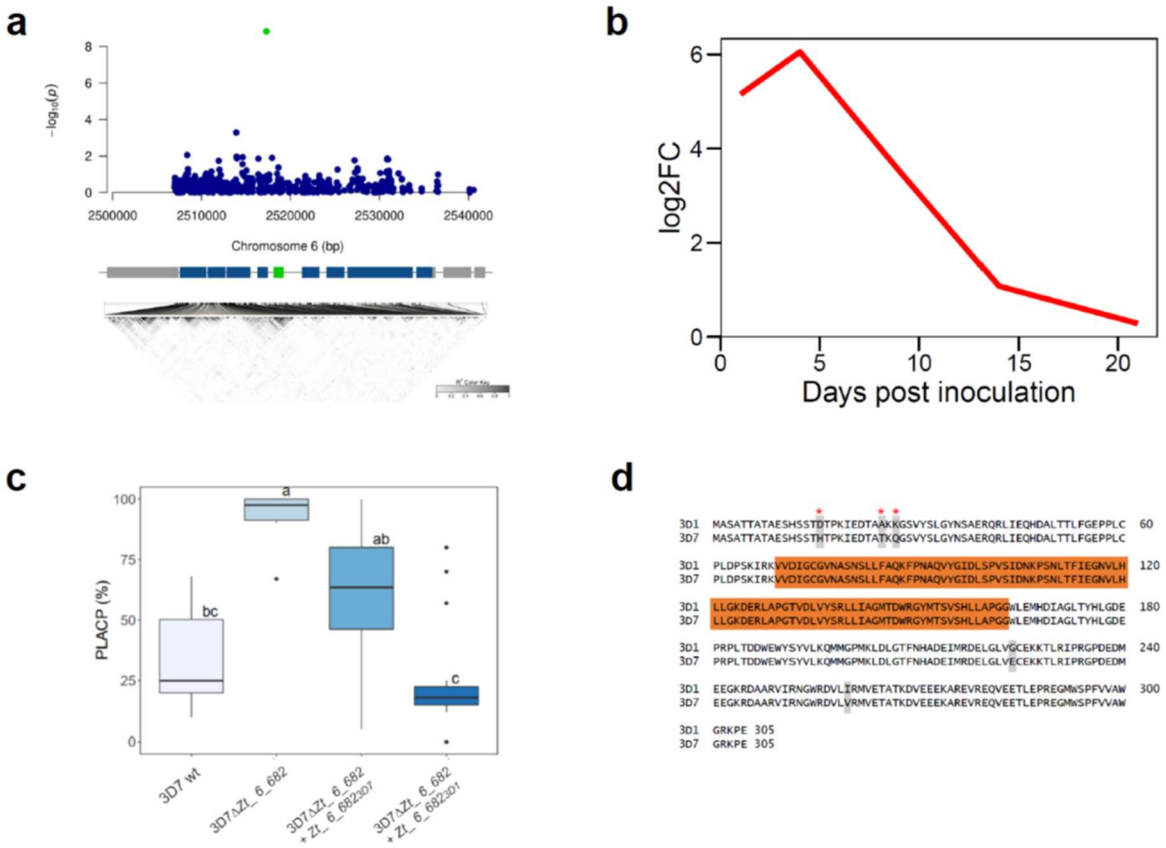


Fig. 6 | A gene encoding a methyltransferase mediates quantitative pathogenicity in *Z. tritici*. **a.** Regional association plot of the pathogenicity gene detected on cultivar ''Shafir''. X and Y axes indicate positions on chromosome 6 and −log10 (*P*-value) for associations with PLACP, respectively. Annotations on the reference genome are shown at the bottom with the candidate gene represented by a green box, other genes by blue boxes and transposable elements by grey boxes. A linkage disequilibrium plot is shown below the annotation bar. **b.** log2-fold expression changes of *Zt_6_00682* based on RNAseq transcription data collected throughout the time course of *Z. tritici* infection on wheat. **c.** Percentage of leaf area covered by pycnidia (PLACP) produced by the wild-type strain 3D7 (n=6 leaves), knockout mutants (n=18 leaves) and the complementation mutants carrying the virulent (3D7) and avirulent (3D1) alleles (n=18 leaves each). Significant difference between group means are represented with different letters above the boxplots after a Tukey's HSD test. Data are presented as box plots (centre line at the median, upper bound at 75th percentile, lower bound at 25th percentile) with whiskers chosen to show the 1.5 of the interquartile

range. **d.** Alignment of the virulent and avirulent protein sequence. Residues in orange constitute the methyltransferase domain, grey residues show protein mutations and red asterisks show positively selected sites with a Bayes Empirical Bayes posterior probability > 0.95.

Discussion

Our results point to a complex genetic architecture of pathogenicity in *Z. tritici*, involving many genes with variable additive effects, resulting in quantitative phenotypes on the host cultivar. These findings are consistent with previous studies[36,37] and provide further evidence that pathogenicity in this pathosystem is complex and results from the contribution of both large- and small- effect genes. We highlighted nineteen differentially and highly expressed genes *in planta* and focused on their features for subsequent investigations. Most of these genes had a peak expression in the phases corresponding to the colonization and the pre-pycnidia formation before the appearance of symptoms. During colonization, hyphae penetrate host tissue *via* stomata and colonize the substomatal cavity but remain exclusively in the intercellular space. This is followed by hyphal invasion of the apoplast leading to pre-pycnidia formation in substomatal cavities[26]. It was shown that these are the key phases of effector gene deployment by plant pathogens[38]. These generally encode small, secreted, cysteine-rich proteins with no sequence homology in other species and play a vital role in successful pathogen invasion, suppression of host immune perception and manipulation of its physiology[39]. Here, we mapped a set of diverse secreted proteins. While effector-like proteins may initiate an effector-triggered immune response, as was shown for the gene *Zt_3_00467*, which triggered quantitative resistance in cultivars 'Riband' and 'Arina', secreted molecules with a known functional domain could also promote disease.

Of the functionally annotated secreted proteins, we identified domains associated with macromolecule degradation, fungal metabolism, or stress response, among others. Some genes encoded proteins that show strong homology with functionally characterized virulence factors from other pathogenic fungi. For instance, *Zt_1_01674* encodes a Thioredoxin reductase (*Trr*), a flavoenzyme that plays a central role in the thioredoxin oxidative stress resistance pathway and is involved in regulating gene transcription, cell growth and apoptosis[40]. Deletion of the *Trr* ortholog resulted in reduced hyphal growth, sexual reproduction, and virulence in *Fusarium graminearum*[41] and profoundly affected pathogenicity of the entomopathogenic fungus *Beauveria bassiana*[42]. We hypothesize that *Zt_1_01674* may play similar roles and could be an important factor in *Z. tritici* development and pathogenicity. We also identified a cell-wall degrading enzyme (*Zt_2_00163*) belonging to the Glycoside Hydrolase protein family. Glycoside hydrolases are known to play important roles in fungal invasion and nutrition acquisition via the degradation or modification of plant cell wall components, such as cellulose or hemicellulose[43].

Genes encoding proteins predicted to be non-secreted were common in our candidate list, highlighting not only the polygenic nature of pathogenicity, but also the potential role of the non-secretory pathway

360 in fungal pathogenicity. This was demonstrated by the functional studies of *Zt_6_00682*, which encodes
361 a SAM-dependent Methyltransferase (SAM-MTase). SAM-MTases catalyse the transfer of methyl
362 groups from SAM to a large variety of substrates, ranging from metabolites to macromolecules,
363 including proteins[44]. They are involved in epigenetic regulation with functions ranging from the control
364 of gene expression, silencing or activation of transposon control[45]. *Zt_6_00682* shares some sequence
365 similarities with the putative SAM-MTase gene *LaeA*, a component of the velvet transcription factor
366 complex that functions as a global regulator of fungal secondary metabolism and development for a
367 handful of fungi[46]. Recent work showed that mutants lacking the velvet gene *velB* in *Z. tritici* have
368 defects in vegetative growth and asexual reproduction[47], while the loss of *LaeA* attenuates pathogenicity
369 in a variety of filamentous fungi[48]. However, the inactivation of *Zt_6_00682* promoted pycnidia
370 formation, suggesting that it functions differently from *LaeA* and may negatively regulate the
371 transcriptional activity of genes or metabolites associated with host colonization. However, how this
372 SAM-MTase affects the regulation of pathogenicity factors needs to be investigated as a key step
373 towards understanding the evolutionary mechanisms associated with the emergence of highly aggressive
374 strains.

375 In addition, we identified three genes with predicted transmembrane domains. For instance, the *DPM2*
376 ortholog (*Zt_1_01325*) was highly expressed throughout the infection cycle, highlighting its potential
377 role in *Z. tritici* pathogenicity. *DPM2* is a regulatory subunit of the dolichol-phosphate mannose synthase
378 complex, key enzymes for GPI anchor biosynthesis[49]. The alteration of the *DPM* complex affected cell
379 wall integrity, fungal growth and increased chitin levels in *Candida albicans*[49]. Traditionally,
380 transmembrane proteins are systematically discarded in effector gene prediction studies. There is
381 however growing evidence that they could play key roles in pathogenicity of filamentous fungi and our
382 results support this hypothesis[50,51].

383 We found substantial evidence for host specificity demonstrated by a highly significant isolate × cultivar
384 interaction and few shared associations between host cultivars. It is also probable that the weak overlap
385 in candidate pathogenicity genes among host cultivars is partly due to a lack of statistical power to detect
386 variants of small effect[52]. However, besides *Stb6*- and *Stb9*-carrying cultivars, there was no meaningful
387 correlation among phenotypes on different hosts, which further supports host specificity. In several well
388 studied host-pathogen systems, specificity is determined mainly by the matching of resistance or
389 susceptibility genes and their corresponding effectors[53-55], following a GFG relationship (or inverse GFG
390 in the case of necrotrophic fungal pathogens).

391 In *Z. tritici*, pathogenicity can be both qualitative[29,56,57] and quantitative[10,58]. Here, we showed that
392 pathogenicity is predominantly quantitative based on multiple genes of small-to-moderate effects. This
393 was demonstrated with the allele analysis of candidate genes and functional studies of *Zt_3_00467* and
394 *Zt_6_00682*. In contrast to what has been shown for most GFG interactions, the recognition of the
395 *Zt_3_00467* and *Zt_6_00682* avirulent alleles does not trigger strong resistance, but rather a reduction

396 at the onset of disease symptoms. Furthermore, the fact that avirulent alleles are only recognized by
397 certain wheat cultivars suggests that the interaction also follows a GFG model and demonstrates that
398 *Zt_3_00467* and *Zt_6_00682* are likely quantitative avirulence factors interacting with specific host
399 resistance genes. This is further supported by recent work in *Z. tritici* as shown for *Avr3D1* and a recently
400 reported effector gene (*G_07189*) that triggers quantitative *Stb20q* resistance[11,59]. This is not uncommon
401 in other filamentous fungi. For instance, the late effector LmSTEE98 from *Leptosphaeria maculans*
402 triggers quantitative resistance following a GFG model[60], similarly, silencing three *Blumeria graminis*
403 f sp. *tritici* effectors lead to quantitative gain of resistance[61]. Since *Zt_3_00467* encodes a typical
404 effector protein and is recognized by hosts harbouring the *Stb15* resistance gene, we suggest that
405 *Zt_3_00467* could be *AvrStb15* and trigger an effector-induced defence. We hypothesise that *Stb15*
406 resistance is not strong enough to block pathogen infection, although we cannot completely rule out the
407 existence of a *Zt_3_00467* haplotype that triggers qualitative resistance. However, additional genetic
408 factors might also contribute to differences in pathogenicity between avirulent and virulent strains. On
409 the other hand, *Zt_6_00682* encodes an atypical avirulence factor and may interact indirectly with the
410 host immune system. The fact that *Zt_6_00682* resembles genes involved in transcriptional regulation
411 and that its deletion mutants are more aggressive on some cultivars, suggests that *Zt_6_00682* negatively
412 regulates pathogenicity factors that are specifically recognized by one or more (unknown) host
413 quantitative resistance gene(s). Complementation mutants further support host specificity, where
414 recognition of the avirulent allele leads to a dramatic reduction in pycnidia formation. Unlike deletion
415 mutants, this was only observed on the cultivar where the gene was detected initially ('Shafir').
416 However, the resistance gene involved and the underlying mechanisms through which the avirulent
417 allele is recognized remain unknown.

418 In addition to *AvrStb6* and *AvrStb9*, we found the SSP *Zt_2_01230* to be significantly associated with
419 variation in quantitative pathogenicity on two cultivars, suggesting a common yet unknown *Stb*
420 resistance gene. We also identified a number of secreted and species-specific proteins. Since canonical
421 fungal effectors are generally thought to be secreted, species-specific and without conserved protein
422 domain[62], we propose that the ones identified here constitute a set of candidate effectors interacting with
423 their cognate host *Stb* resistance products and should be a priority target for functional studies.
424 Unlocking the molecular basis underpinning these interactions and how it could impact host defence
425 would lead to a better exploitation of quantitative resistance, which is thought to be more durable[63].

426 Evolution and diversification of pathogenicity genes enables adaptation to new host cultivars and, in
427 response to host population changes, they undergo changes in rapid allele frequency[64]. We observed
428 high levels of genetic diversity in our candidate pathogenicity genes when compared to the genome-
429 wide gene content. We found a single case of gene deletion and no significant impact of loss of function.
430 Because of the intact and functional coding sequences for nearly all of the identified genes, we suggest
431 that sequence polymorphism is the primary mechanism behind the rapid adaptation of *Z. tritici* to host

432 populations, permitting the pathogen to circumvent host resistance while conserving the presumable
433 function of the pathogenicity protein. However, it is important to note that the utilization of short-read
434 sequencing technology imposes limitations on our ability to identify structural variants that may play a
435 role in pathogenicity. Genes exhibiting presence/absence polymorphism could be underrepresented in
436 our final set of candidate pathogenicity genes. We next investigated what could contribute to this
437 extensive genetic diversity. Recombination is a central source of diversity in sexual fungal pathogens[65].
438 Here, we detected recombination events for a significant number of candidates, which suggests a role in
439 the observed genetic diversity of pathogenicity genes. In *Z. tritici*, it was reported that recombination is
440 a major driver of its evolution and is highly heterogeneous, concentrated in narrow genomic regions that
441 are enriched in putative effector genes[8,66]. TEs are also thought to contribute to gene diversification by
442 creating fast-evolving genomic regions through non-homologous recombination or repeat-induced point
443 mutations[67]. Here, we found that candidate pathogenicity genes are indeed at closer distances to TEs,
444 suggesting their potential role in sequence diversity. We also observed that TE distance was correlated
445 with recombination events in putative pathogenicity genes, and that the levels of genetic diversity were
446 high for genes with the highest recombination events, regardless of their distances to TEs. We
447 hypothesize that recombination is likely the major driver underlying the diversification of pathogenicity
448 genes. A recent study supports this hypothesis where, unlike distance to TEs, the recombination rate
449 was strongly associated with adaptive substitutions in *Z. tritici* genomes[4].

450 A key determinant of genetic diversity at pathogenicity genes is selection pressure. Understanding the
451 evolution of pathogenicity in fungal populations with respect to the selection pressure imposed by host
452 populations is necessary for the sustainable management of *R* genes in the field. This can be achieved
453 by searching for signatures of positive diversifying selection at the risk of recovering some false positive
454 signals. In this study, we used conservative maximum likelihood models and models that have been
455 shown to be robust to recombination and significantly reduce false positives[68]. More than two thirds of
456 candidate genes harbour signatures of positive selection acting on specific codons, suggesting that the
457 diversifying selection is likely the primary evolutionary force maintaining non-synonymous
458 substitutions relevant for *Z. tritici*-wheat quantitative interactions. For instance, the vast majority of
459 different mutations between the avirulent and virulent versions of *Zt_3_00467* and *Zt_6_00682* have
460 been identified as positively selected. Such selection pressure could be the result of host *R* gene
461 recognition, and may have greatly contributed to the accumulation of beneficial mutations. This agrees
462 with a previous study, which demonstrated that rapidly evolving genes play a prominent role in host-
463 specific disease development[69]. Together, these results suggest that host–pathogen coevolution drives
464 the emergence and maintenance of allelic diversity at loci likely involved in direct (or indirect)
465 interactions with host *R* genes. Future surveys of the diversity and structure of *R* genes in wheat
466 populations will provide a deeper insight into reciprocal patterns of selection, coevolution, and the
467 maintenance of diversity in resistance and pathogenicity in natural populations.

468

469 ## Methods

470 **Fungal material.** We analysed 103 *Z. tritici* isolates in this study deriving predominantly from a large
471 collection of French isolates collected in 2009 and 2010. This genetically diverse sample was mainly
472 isolated from the French cultivars 'Apache' (n=44) and 'Premio' (n=42). The remaining were isolated
473 from cultivars 'Soissons' (n=9), 'Caphorn' (n=3), 'Alixan' (n=1), 'Bermude' (n=1), 'Dinosor' (n=1),
474 and 'Garcia' (n=1). One isolate was collected in the United Kingdom in 2009 on cultivar 'Humbert'.
475 All strains are monospore non-clonal isolates were originally isolated either at the Plant Research
476 International in the Netherlands or at INRAE UR BIOGER in France. For the functional validation of
477 candidate pathogenicity genes, we used the two strains 3D1 (*i.e.*, ST99CH3D1) and 3D7 (*i.e.*,
478 ST99CH3D7) collected in 1999 from a Swiss wheat field and previously used for the functional
479 validation of *Avr3D1*[10].

480

481 **Plant material.** The pathogenicity of the 103 isolates was initially evaluated on 29 wheat cultivars,
482 including 15 carrying known *Stb* resistance genes. After the first replicate, only cultivars on which we
483 observed phenotypic variability were kept for second and third replicates. These were the cultivars
484 'Bulgaria-88' (*Stb1* and *Stb6*), 'Israel-493' (*Stb3* and *Stb6*), 'Tadinia' (*Stb4* and *Stb6*), 'Shafir' and
485 'Cadenza' (*Stb6*), 'Estanzuela Federal' (*Stb7*), 'Courtot' (*Stb9*), 'Apache' (*Stb4/5* and *Stb11*),
486 'Salamouni' (*Stb13*, *Stb14* and *StbSm3*) and 'Arina' (*Stb6* and *Stb15*); the resistance background of
487 cultivars 'Premio' and 'Toisondor' was not known. Details on the differential lines are available in
488 Supplementary Table 3. This genetic material is available from the INRAE Small grain cereals
489 Biological Resources Centre in Clermont-Ferrand
490 (<https://www6.clermont.inrae.fr/umr1095_eng/Organisation>).

491

492 **Quantitative pathogenicity assays.** For inoculum preparation, blastospores of *Z. tritici* stored at -80°C
493 were used to inoculate a liquid growth medium (Difco™ YPD Broth Cat. No. 242820) agitated at 140
494 rpm in a 17°C climate cell for 144 h and transferred to agar growth medium (Difco™ YPD Agar Cat.
495 No. 242720) for 96 h. Fresh blastospores were recovered with loops, resuspended in sterile water, and
496 the inoculum was adjusted to a concentration of 10^6 spores.L^{-1}, to which one drop of Tween®20 (Sigma
497 Aldrich® Cat. No. 9005-64-5) was added per 15 mL before inoculation. Five seeds from each cultivar
498 were sown in 90*90*95 mm pots containing Floradur B soil (80% blond peat / 20% black peat, with 1.2
499 kg.m^{-3} fertilizer 18/10/20). One day before inoculation, we kept three homogeneous seedlings on which
500 a 75-mm-length portion was delimited on the first true leaf using a black marker. Sixteen days old
501 seedlings were inoculated by applying the inoculum six times on the marked leaves with a flat tipped
502 paintbrush (synthetic bristles, 12 mm width), depositing approximately 30,000 spores.cm^{-2} on the leaf
503 surface. After inoculation, plants were watered and enclosed in transparent polyethylene high-density
504 bags for 72 h to initiate the infection. Ten days post-inoculation, leaves were removed keeping only the

505 marked leaf to ensure light homogeneity and longevity of the inoculated leaf. From sowing to disease
506 scoring, plants were kept in growth chambers regulated at 80% relative humidity, 16 h of light per 24 h
507 at a photon flux of 300 $\mu mol.m^{-2}.s^{-1}$, 22°C temperature under light and 18°C in the darkness. Quantitative
508 pathogenicity was assessed as percentages of leaf area covered by pycnidia (PLACP) and by necrosis
509 (PLACN) at 21 days post inoculation, averaged over three leaves and three replicates (nine leaves per
510 interaction). Phenotypic characterization of the generated deletion and ectopic transformants was done
511 following the same procedure described above in two replicates.

512 **Whole-genome resequencing.** Blastospores were harvested after seven days growth in liquid cultures
513 on yeast extract-peptone-dextrose Broth (Difco™ YPD Broth), and then lyophilised for 24 h. High
514 molecular weight DNA was extracted using a phenol-chloroform-based procedure[70]. Illumina paired-
515 end libraries were prepared for each sample and DNA fragments were re-sequenced with 2x100 bp reads
516 on a HiSeq-2000 sequencing system at Genewiz, Inc. (formerly Beckman Coulter Genomics). Depth of
517 sequence coverage was variable depending on the isolate, ranging from 16 to 303 genome equivalents
518 with an average of 61 genome equivalents.

519 **SNP calling and quality filtering.** For each isolate, reads were trimmed using trimmomatic v0.32[71] and
520 then mapped to the reference genome IPO323[72] using the mem algorithm from BWA v0.7.7 with default
521 settings[73]. Samtools v0.1.19 and Picard tools v1.106 were used to filter reads[74]
522 (<http://broadinstitute.github.io/picard/>); to remove optical duplicates, secondary alignments, reads
523 with a mapping quality below 30 and to keep only pairs in which both reads met quality checks. SNP
524 calling was done with Freebayes v0.9 using the option --report-monomorphic[75] and all positions
525 detected as low complexity regions or TEs were excluded. Low-complexity regions and TEs were
526 detected using RepeatMasker with default settings (RepeatMasker Open-4.0
527 <http://www.repeatmasker.org>) and REPET v.2.5[76], respectively. Quality filters were applied to VCF
528 files using in-house scripts. Depth of coverage (DP) calculated at each position followed a normal
529 distribution around the depth main peak. Hence, a DP filter was set at twice the standard deviation of
530 the mean. All VCFs were stored as a matrix reduced to keep only positions where at least one isolate
531 carries an alternative allele at a non-filtered position. The SNP matrix was parsed as a .tped format,
532 converted to .ped/.map. The final matrix was filtered for less than 50% missing data and more than 10%
533 Minor Allele Frequency (MAF) using PLINK v.1.9[77] <https://cog-genomics.org/plink/>.

534 **Linkage disequilibrium (LD).** To assess the mapping power and resolution of GWAS, we estimated
535 genome-wide LD decay using PopLDdecay v.3.40[78]. The average squared Pearson's correlation
536 coefficient (r^2) was computed between pairs of SNPs up to 20 kb and plotted over physical distance by
537 fitting a nonlinear regression model.

538 **Population structure analyses.** Two approaches were used to investigate the fungal population
539 structure: (1) we performed a principal component analysis (PCA) using independent SNPs ($r^2 < 0.2$)

540 with the --pca command PLINK v.1.9 and (2) the admixture model in STRUCTURE v.2.2 software[79]
541 with a 10^5 burn-in period, 5×10^5 MCMC steps and 5 iterations for each k value from 1 to 10. The optimal
542 number of sub-populations was determined with the Evanno method[80].

543 **GWAS.** For each quantitative trait, we conducted GWAS using single- and multi-locus linear mixed
544 models based on the EMMA algorithm implemented in GAPIT v.3.0[81-83]. We used a VanRaden derived
545 kinship matrix as a random effect, and set the PCA option to 'true' to determine the optimum number
546 of PCs to be included in the model, following the Bayesian Information Criterion (BIC). The most robust
547 model was chosen by visualizing Quantile-Quantile plots that illustrate the observed vs expected *P*
548 values. A uniform threshold of $P=1/n$ (n being the effective number of independent SNPs) was used as
549 the genome-wide significance threshold. The effective number of independent SNPs was calculated
550 using the Genetic type 1 Error Calculator; <http://grass.cgs.hku.hk/gec/download.php>. The suggestive
551 genome-wide threshold for the 718,810 SNPs was 3.96×10^{-6} and the significant *P* value was 1.98×10^{-7}.
552 We also adjusted SNP *P*-values to account for multiple testing correction using the false discovery rate
553 (FDR) procedure. SNPs with an FDR-adjusted *P* value < 0.10 were deemed significant. For LD analysis,
554 we first calculated the pairwise LD with the lead SNP in a window size of 1 Mb using the --r^2 command
555 in PLINK v.1.9. We set a cut-off of $r^2 > 0.5$ to determine candidate genomic regions and assigned
556 predicted genes using a custom Perl script. In case more than one gene is assigned to the candidate
557 region, we picked the closest one from the lead SNP having pathogenicity-like properties (*e.g.*, secretion
558 signal, differential expression *in planta*). If no gene fell in the LD region, we selected the closest gene
559 to the lead SNP. LD heatmaps were constructed using Haploview v4.2[84].

560 **Functional annotation.** We inferred gene models from the *ab initio Z. tritici* annotation[85] in the
561 candidate genomic regions identified by GWAS. We used RNAseq datasets of the reference isolate
562 IPO323[25] to check for missing genes or manually refine gene models (*e.g.*, splice junctions, missing
563 exons). N-terminal secretion signals were predicted using the Hidden Markov Model (HMM) scoring
564 method implemented in SignalP 3.0[86]. Transmembrane domains were identified with TMHMM 2.0[87].
565 Effector probabilities were predicted through EffectorP 3.0[88]. Conserved protein domains were assigned
566 using HMM scan against the PFAM 35.0 database and blastp was used to search for homologs from
567 other fungi. Finally, we used PHIB-BLAST, a blast server for PHI-base (Pathogen-Host Interactions),
568 to search for homologs against functionally characterized pathogenicity factors.

569 **Gene expression analysis.** RNAseq data from *in vitro* and *in planta* (1-, 4-, 9-, 14- and 21-days post
570 infection; dpi) were obtained from a previous study[25]. We computed gene expression quantified as
571 fragments per kilobase of exon per million fragments mapped (FPKM) using the improved gene models
572 with Cuffdiff v.2.2.1[89]. We inferred pairwise log2-fold expression changes and *P* value adjustment for
573 *in planta* RNAseq samples compared to the Czapek-Dox broth (CDB) culture medium. Genes with an
574 FDR-adjusted *P*-value < 0.10 and a log2|fold change|> 2 were considered as differentially expressed
575 during the infection.

576 ***De novo* assembly and sequence diversity.** We used SPAdes v.3.14.1 to produce draft assemblies for
577 each isolate[90]. We ran the tool with the following settings: -k 21,33,55,75,95 --careful to reduce
578 mismatches and indel errors. We performed blastn[91] searches using ~1000 bp of the target locus as a
579 query to retrieve matching contig segments in each assembly. High-confidence matches were first
580 extracted from the contig set using Samtools v.1.10 and subsequently aligned with MAFFT v7.464 using
581 the --auto option and 1,000 iterative refinement cycles[92]. Alignments were inspected using JalView
582 v.2.11.1[93] and ambiguous regions were removed with trimAL v1.2[94] with less stringent parameters. We
583 used the batch mode of DNAsp v.6[95] to calculate summary statistics of population genetic parameters
584 associated with candidate genes such as nucleotide diversities (π)[96] and Tajima's D[97]. TE locations for
585 *Z. tritici* were retrieved from a previous study[85] and we parsed gene distances to TEs using the command
586 'closest' from bedtools v.2.29.2[98].

587 **Evolutionary analysis.** Before testing for positive diversifying selection, we checked for evidence of
588 recombination, as this can cause false positives. We used the software packages RDP4.101[99] and
589 GARD[100] for recombination detection. We considered genes with recombination events if both detection
590 methods found evidence of recombination. We then constructed non-recombinant phylogenetic trees of
591 gene coding sequences using the RAxML algorithm[101] implemented in RDP4. Statistical support for the
592 nodes was obtained after 100 bootstrap replicates. We used codeml of the PAML v.4.9 package[102] to
593 estimate the evolutionary rates of candidate genes, based on Markov models of codon evolution. Branch
594 lengths of the inferred ML trees were estimated by using the model M0 and then fixed in further
595 analyses. The model M0 assumes constant selective pressure across codon sites and over time. The
596 selective pressure at the coding-sequence level was measured by ω, the ratio of non-synonymous to
597 synonymous rates (d_N/d_S). A d_N/d_S ratio of 1 ($\omega = 1$) indicates neutrality, whereas $\omega < 1$ and $\omega > 1$
598 suggests purifying and diversifying selection, respectively. Models M1a, M2a, M7 and M8 of variable
599 selective pressure across codon sites were used to estimate selective pressure and test for positive
600 selection[103,104]. For each candidate gene, we performed pairwise likelihood ratio tests (LRTs) for positive
601 selection, comparing models that allow sites with $\omega > 1$ (M2a and M8) with models that do not (M1a
602 and M7). We considered genes under positive diversifying selection if both model comparisons in LRT
603 were significant. Finally, positively selected sites were inferred with a Bayes Empirical Bayes (BEB)
604 probability[104] and codons with $\omega > 1$ at a probability > 0.95 were considered positively-selected.

605 **Functional validation of candidate pathogenicity genes.** For *Zt_6_00682*, we generated knockout
606 mutants in the 3D7 isolate and ectopic transformants expressing the virulent and avirulent alleles in the
607 knock-out background. To create plasmid constructs for gene disruption, flanking regions (~1 kb before
608 start and stop codons) of *Zt_6_00682* were amplified from gDNA using primers listed in Supplementary
609 Table 7. The hygromycin resistance gene cassette was amplified from the plasmid
610 pNOV2114_HygR_attP1_attP2 and cloned into a pNOV2214 vector. The vector was then digested with
611 *Hind*III restriction enzyme (New England Biolabs), amplified, and used for the cloning of PCR

612 fragments with a Gibson Assembly cloning kit (New England Biolabs) following the manufacturer's
613 instructions. Constructs were transformed into *Escherichia coli* NEB 5-α using a heat shock
614 transformation, screened by PCR and verified by Sanger sequencing using primers listed in
615 Supplementary Table 7. Confirmed plasmids were transformed into *Agrobacterium tumefaciens* cells
616 by heat shock and validated by PCR. The vector was used to transform *Z. tritici* via *A. tumefaciens*-
617 mediated transformation (ATMT) following a procedure adapted from Zwiers & de Waard[105].
618 Transformed *Z. tritici* colonies growing in the presence of hygromycin were tested by PCR to confirm
619 the successful deletion of the gene. To validate the allele-dependent effect of *Zt_6_00682*, we generated
620 and validated ectopic transformants by placing both the virulent (3D7) and the avirulent (3D1) alleles
621 into the 3D7 KO background, following the same procedure described above with the following
622 modifications: (*i*) we used the vector pNOV_3Gate_SUL with sulfonylurea as a resistance marker and
623 the digestion was achieved with *Hind*III and *Kpn*I restriction enzymes. The resulting vectors,
624 pNOV_sulf_3D7g.7232 and pNOV_sulf_3D1g.7155, were used to transform *Z. tritici* via ATMT
625 following the procedure described above. Primers used to verify gene disruptions and ectopic
626 integrations are listed in Supplementary Table 7. Finally, three independent transformants carrying the
627 different virulence alleles and three knock-out mutants were used for phenotypic characterization *in*
628 *planta*. For *Zt_3_00467*, we ectopically expressed the virulent (3D7) and avirulent (IPO10273) alleles
629 in the 3D7 background. A fragment containing the *Zt_3_00467* gene, including 1800 bp upstream of the
630 start codon and 463 bp downstream of the stop codon was amplified using Phusion DNA polymerase
631 (NEB) and the primers included in Supplementary Table 7. This fragment was cloned into a pCGEN
632 vector[106] previously digested with *Kpn*I (New England Biolabs), using the In-Fusion HD Cloning Kit
633 (Takara Bio) and following the manufacturer's instructions. Vectors were transformed using heat-
634 shocked *E. coli* Stellar cells and verified using Sanger sequencing. Transformation via ATMT was
635 performed as described above and selection for the *Z. tritici* mutants was performed on geneticin
636 (150 $\mu g.mL^{-1}$)-containing plates.
637

638 **Data availability**

639 The sequence and annotation of *Zymoseptoria tritici* IPO323 reference genome are available from DOE
640 Joint Genome Institute website <https://mycocosm.jgi.doe.gov/Zymtr1/Zymtr1.home.html>. The
641 sequence data generated in this study have been deposited in the NCBI Sequence Read Archive under
642 the accession number PRJNA777581 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA777581].
643 Sequence data for the global isolates are available at the NCBI Sequence Read Archive under the
644 accession number PRJNA327615 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA327615]. The
645 processed summary GWAS and gene expression data are provided in the Supplementary Data files 1
646 and 2.

647

648 **References**

649 1. Savary, S., et al. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **3**,
650 430–439 (2019). https://doi.org/10.1038/s41559-018-0793-y

651 2. Franceschetti, M., et al. Effectors of filamentous plant pathogens: commonalities amid diversity.
652 *Microbiol. Mol. Biol. Rev.* **81**, e00066-16 (2022). https://doi.org/10.1128/MMBR.00066-16

653 3. lo Presti, L., et al. Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* **66**, 513–545
654 (2015). https://doi.org/10.1146/annurev-arplant-043014-114623

655 4. Grandaubert, J., Dutheil, J. Y., & Stukenbrock, E. H. The genomic determinants of adaptive
656 evolution in a fungal pathogen. *Evol. Lett.* **3**, 299–312 (2019). https://doi.org/10.1002/evl3.117

657 5. Möller, M., & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens.
658 *Nat. Rev. Microbiol.* **15**, 756–771 (2017). https://doi.org/10.1038/nrmicro.2017.76

659 6. Raffaele, S., & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be
660 better. *Nat. Rev. Microbiol.* **10**, 417–430 (2012). https://doi.org/10.1038/nrmicro2790

661 7. Rouxel, T., et al. Effector diversification within compartments of the *Leptosphaeria maculans*
662 genome affected by repeat-induced point mutations. *Nat. Commun.* **2**, 202 (2011).
663 https://doi.org/10.1038/ncomms1189

664 8. Stukenbrock, E. H., & Dutheil, J. Y. Fine-scale recombination maps of fungal plant pathogens reveal
665 dynamic recombination landscapes and intragenic hotspots. *Genetics* **208**, 1209–1229 (2018).
666 https://doi.org/10.1534/genetics.117.300502

667 9. Singh, P.K., et al. Co-evolutionary interactions between host resistance and pathogen avirulence
668 genes in rice-*Magnaporthe oryzae* pathosystem, *Fungal Genet. Biol.* **115**, 9-19 (2018).
669 https://doi.org/10.1016/j.fgb.2018.04.005

670 10. Meile, L., et al. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial
671 resistance to septoria tritici blotch. *New Phytol.* **219**, 1048–1061 (2018).
672 https://doi.org/10.1111/nph.15180

673 11. Langlands-Perry, C., et al. Quantitative and qualitative plant-pathogen interactions call upon similar
674 pathogenicity genes with a spectrum of effects. *Front. Plant Sci.* **14**, 1128546 (2023).
675 https://doi.org/10.3389/fpls.2023.1128546

676 12. Amezrou, R., et al. A secreted protease-like protein in *Zymoseptoria tritici* is responsible for
677 avirulence on *Stb9* resistance gene in wheat. *PLoS Pathog.* **19**, e1011376 (2023).
678 https://doi.org/10.1371/journal.ppat.1011376

679 13. Stukenbrock, E. H., & McDonald, B. A. Population genetics of fungal and oomycete effectors
680 involved in gene-for-gene interactions. *Mol. Plant-Microbe Interact.* **22**, 371–380 (2009).
681 https://doi.org/10.1094/MPMI-22-4-0371

682 14. van de Wouw, A. P., & Idnurm, A. Biotechnological potential of engineering pathogen effector
683 proteins for use in plant disease management. *Biotechnol. Adv.* **37**, 107387 (2019).
684 https://doi.org/10.1016/j.biotechadv.2019.04.009

685 15. Domazakis, E., et al. Effectoromics-based identification of cell surface receptors in potato. *Methods*
686 *mol. biol.* **1578**, 337–353 (2017). https://doi.org/10.1007/978-1-4939-6859-6_29

687 16. Varden, F. A., de la Concepcion, J. C., Maidment, J. H. R., & Banfield, M. J. Taking the stage:
688 effectors in the spotlight. *Curr. Opin. Plant Biol.* **38**, 25–33 (2017).
689 https://doi.org/10.1016/j.pbi.2017.04.013

690 17. Collemare, J., O'Connell, R., & Lebrun, M.-H. Nonproteinaceous effectors: the terra incognita of
691 plant–fungal interactions. *New Phytol.* **223**, 590–596 (2019). https://doi.org/10.1111/nph.15785

692 18. Kubicek, C. P., Starr, T. L., & Glass, N. L. Plant cell wall–degrading enzymes and their secretion
693 in plant-pathogenic fungi. *Annu. Rev. Phytopathol.* **52**, 427–451 (2014).
694 https://doi.org/10.1146/annurev-phyto-102313-045831

695 19. van der Hoorn, R. A. L. Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev.*
696 *Plant Biol.* **59**, 191–223 (2008). https://doi.org/10.1146/annurev.arplant.59.032607.092835

697 20. Ma, K.-W., & Ma, W. Phytohormone pathways as targets of pathogens to facilitate infection. *Plant*
698 *Molecular Biology* **91**, 713–725 (2016). https://doi.org/10.1007/s11103-016-0452-0

699 21. Lannou, C. Variation and selection of quantitative traits in plant pathogens. *Annu. Rev. Phytopatol.*
700 **50**, 319-338 (2012). https://doi.org/10.1146/annurev-phyto-081211-173031

701 22. Croll, D., & McDonald, B. A. The genetic basis of local adaptation for pathogenic fungi in
702 agricultural ecosystems. *Mol. Ecol.* **26(7)**, 2027-2040 (2017). https://doi.org/10.1111/mec.13870

703 23. Carlier, J., et al. Convergent adaptation to quantitative host resistance in a major plant pathogen.
704 *mBio* **12(1)**, e03129-20 (2021). https://doi.org/10.1128/mbio.03129-20

705 24. O'Driscoll, A., Kildea, S., Doohan, F., Spink, J., & Mullins, E. The wheat Septoria conflict: a new
706 front opening up? *Trends Plant Sci.* **19**, 602–610 (2014).
707 https://doi.org/10.1016/j.tplants.2014.04.011

708 25. Rudd, J. J., et al. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria*
709 *tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen
710 chromosomal contributions and a variation on the hemibiobiotrophic lifestyle definition. *Plant*
711 *Physiol.* **167**, 1158–1185 (2015). https://doi.org/10.1104/pp.114.255927

712 26. Duncan, K. E., & Howard, R. J. Cytological analysis of wheat infection by the leaf blotch pathogen
713 *Mycosphaerella graminicola*. *Mycol. Res.* **104**, 1074–1082 (2000).
714 https://doi.org/10.1017/S0953756299002294

715 27. Kema, G. H. J., Sayoud, R., Annone, J. G., & van Silfhout, C. H. Genetic variation for virulence
716 and resistance in the wheat-*Mycosphaerella graminicola* pathosystem II. Analysis of interactions
717 between pathogen isolates and host cultivars. *Phytopathology* **86**, 213–220 (1996).
718 https://doi.org/10.1094/Phyto-86-213

719 28. Kema, G. H. J., et al. Stress and sexual reproduction affect the dynamics of the wheat pathogen
720 effector AvrStb6 and strobilurin resistance. *Nat. Genet.* **50**, 375–380 (2018).
721 https://doi.org/10.1038/s41588-018-0052-9

722 29. Zhong, Z., et al. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on
723 wheat cultivars carrying the *Stb6* resistance gene. *New Phytol.* **214**, 619–631 (2017).
724 https://doi.org/10.1111/nph.14434

725 30. Brown, J. K. M., Chartrain, L., Lasserre-Zuber, P., & Saintenac, C. Genetics of resistance to
726 *Zymoseptoria tritici* and applications to wheat breeding. *Fungal Genet. Biol.* **79**, 33–41 (2015).
727 https://doi.org/10.1016/j.fgb.2015.04.017

728 31. Saintenac, C., et al. Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to
729 fungal pathogen *Zymoseptoria tritici*. *Nat. Genet.* **50**, 368–374 (2018).
730 https://doi.org/10.1038/s41588-018-0051-x

731 32. Saintenac, C., et al. A wheat cysteine-rich receptor-like kinase confers broad-spectrum resistance
732 against Septoria tritici blotch. *Nat. Commun.* **12**, 433 (2021). https://doi.org/10.1038/s41467-020-
733 20685-0

734 33. Bartoli, C., & Roux, F. Genome-wide association studies in plant pathosystems: toward an
735 ecological genomics approach. *Front. Plant Sci.* **8**, 763 (2017).
736 https://www.frontiersin.org/article/10.3389/fpls.2017.00763

34. Sánchez-Vallet, A., Hartmann, F. E., Marcel, T. C., & Croll, D. Nature's genetic screens: using genome-wide association studies for effector discovery. *Mol. Plant Pathol.* **19**, 3–6 (2018). https://doi.org/10.1111/mpp.12592

35. Habig, M., Quade, J., & Stukenbrock, E. H. Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen *Zymoseptoria tritici*. *mBio* **8**, e01919–17 (2017). https://doi.org/10.1128/mBio.01919-17

36. Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J.* **11**, 1189–1204 (2017). https://doi.org/10.1038/ismej.2016.196

37. Stewart, E. L., et al. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.* **19**, 201–216 (2018). https://doi.org/10.1111/mpp.12515

38. Gohari, A.M., et al. Effector discovery in the fungal wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.* **16**, 931–945 (2015). https://doi.org/10.1111/mpp.12251

39. Toruño, T. Y., Stergiopoulos, I., & Coaker, G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annu. Rev. Phytopathol.* **54**, 419–441 (2016). https://doi.org/10.1146/annurev-phyto-080615-100204

40. Arnér, E. S. J., & Holmgren, A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* **267**, 6102–6109 (2000). https://doi.org/10.1046/j.1432-1327.2000.01701.x

41. Fan, X., et al. Thioredoxin reductase is involved in development and pathogenicity in *Fusarium graminearum*. *Front. Microbiol.* **10**, 393 (2019). https://www.frontiersin.org/article/10.3389/fmicb.2019.00393

42. Zhang, L.-B., Tang, L., Ying, S.-H., & Feng, M.-G. Distinct roles of two cytoplasmic thioredoxin reductases (Trr1/2) in the redox system involving cysteine synthesis and host infection of *Beauveria bassiana*. *Appl. Microbiol. Biotechnol.* **100**, 10363–10374 (2016). https://doi.org/10.1007/s00253-016-7688-0

43. Esquerré-Tugayé, M.-T., Boudart, G., & Dumas, B. Cell wall degrading enzymes, inhibitory proteins, and oligosaccharides participate in the molecular dialogue between plants and pathogens. *Plant Physiology and Biochemistry*, *38*(1), 157–163 (2000). https://doi.org/10.1016/S0981-9428(00)00161-3

44. Zhang, J., & Zheng, Y. G. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chem. Biol.* **11**, 583–597 (2016). https://doi.org/10.1021/acschembio.5b00812

45. Martienssen, R. A., & Vincent, C. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001). https://doi.org/10.1126/science.293.5532.1070

46. Bayram, Ö., & Braus, G. H. Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. *FEMS Microbiol. Rev.* **36**, 1–24 (2012). https://doi.org/10.1111/j.1574-6976.2011.00285.x

47. Tiley, A.M.M., White, H.J., Foster, G.D., & Bailey, A.M. The *ZtvelB* gene is required for vegetative growth and sporulation in the wheat pathogen *Zymoseptoria tritici*. *Front. Microbiol.* **10,** 2210 (2019). https://doi.org/10.3389/fmicb.2019.02210

48. Jain, S., & Keller, N. Insights to fungal biology through LaeA sleuthing. *Fungal Biol. Rev.* **27**, 51–59 (2013). https://doi.org/10.1016/j.fbr.2013.05.004

780 49. Juchimiuk, M., Kruszewska, J., & Palamarczyk, G. Dolichol phosphate mannose synthase from the
781 pathogenic yeast *Candida albicans* is a multimeric enzyme. *Biochim. Biophys. Acta Gen. Subj.*
782 **1850**, 2265–2275 (2015). https://doi.org/10.1016/j.bbagen.2015.08.012
783 50. Weiland, P., & Altegoer, F. Identification and characterization of two transmembrane proteins
784 required for virulence of *Ustilago maydis*. *Front. Plant Sci.* **12**, 669835 (2021).
785 https://doi.org/10.3389/fpls.2021.669835
786 51. Gu, Q., Chen, Y., Liu, Y., Zhang, C., & Ma, Z. The transmembrane protein FgSho1 regulates fungal
787 development and pathogenicity via the MAPK module Ste50-Ste11-Ste7 in *Fusarium*
788 *graminearum*. *New Phytol.* **206**, 315–328 (2015). https://doi.org/10.1111/nph.13158
789 52. Korte, A., & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review.
790 *Plant Methods* **9**, 29 (2013). https://doi.org/10.1186/1746-4811-9-29
791 53. Bourras, S., et al. Multiple avirulence loci and allele-specific effector recognition control the *Pm3*
792 race-specific resistance of wheat to powdery mildew. *The Plant Cell* **27**, 2991–3012 (2015).
793 https://doi.org/10.1105/tpc.15.00171
794 54. Faris, J. D., et al. A unique wheat disease resistance-like gene governs effector-triggered
795 susceptibility to necrotrophic pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13544–13549 (2010).
796 https://doi.org/10.1073/pnas.1004090107
797 55. Liu, W., et al. Recent progress in understanding PAMP- and effector-triggered immunity against
798 the rice blast fungus *Magnaporthe oryzae*. *Mol. Plant* **6**, 605–620 (2013).
799 https://doi.org/10.1093/mp/sst015
800 56. Brading, P. A., Verstappen, E. C. P., Kema, G. H. J., & Brown, J. K. M. A gene-for-gene relationship
801 between wheat and *Mycosphaerella graminicola*, the Septoria tritici blotch pathogen.
802 *Phytopathology* **92**, 439–445 (2012). https://doi.org/10.1094/PHYTO.2002.92.4.439
803 57. Kema, G. H. J., Verstappen, E. C. P., & Waalwijk, C. Avirulence in the wheat Septoria tritici leaf
804 blotch fungus *Mycosphaerella graminicola* is controlled by a single locus. *Mol. Plant-Microbe*
805 *Interact.* **13**, 1375–1379 (2000). https://doi.org/10.1094/MPMI.2000.13.12.1375
806 58. Zhan, F., et al. Linear correlation analysis of *Zymoseptoria tritici* aggressiveness with *in vitro*
807 growth rate. *Phytopathology* **106**, 1255–1261 (2016). https://doi.org/10.1094/PHYTO-12-15-0338-
808 R
809 59. Meile L, et al. Natural variation in Avr3D1 from *Zymoseptoria* sp. contributes to quantitative gene-
810 for-gene resistance and to host specificity. *New Phytol.* **238**, 1562-1577 (2023).
811 https://doi.org/10.1111/nph.18690
812 60. Jiquel, A., et al. A gene-for-gene interaction involving a 'late' effector contributes to quantitative
813 resistance to the stem canker disease in *Brassica napus*. *New Phytol.* **231**, 1510–1524 (2021).
814 https://doi.org/10.1111/nph.17292
815 61. Schaefer, L.K., et al. Cross-kingdom RNAi of pathogen effectors leads to quantitative adult plant
816 resistance in wheat. *Front. Plant Sci.* **11**, 253 (2020). https://doi.org/10.3389/fpls.2020.00253
817 62. Kim, K.-T., et al. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their
818 potential role in host association. *Front. Plant Sci.* **7**, 186 (2016).
819 https://www.frontiersin.org/article/10.3389/fpls.2016.00186
820 63. Kettles, G. J., & Kanyuka, K. Dissecting the molecular interactions between wheat and the fungal
821 pathogen *Zymoseptoria tritici*. *Front. Plant Sci.* **7**, 508 (2016).
822 https://www.frontiersin.org/article/10.3389/fpls.2016.00508

823 64. Barrett, L. G., et al. Diversity and evolution of effector loci in natural populations of the plant
824 pathogen *Melampsora lini*. *Mol. Biol. Evol.* **26**, 2499–2513 (2009).
825 https://doi.org/10.1093/molbev/msp166

826 65. Burdon, J. J., & Silk, J. Sources and patterns of diversity in plant-pathogenic fungi. *Phytopathology*
827 **87**, 664–669 (1997). https://doi.org/10.1094/PHYTO.1997.87.7.664

828 66. Croll, D., Lendenmann, M. H., Stewart, E., & McDonald, B. A. The impact of recombination
829 hotspots on genome evolution of a fungal plant pathogen. *Genetics* **201**, 1213–1228 (2015).
830 https://doi.org/10.1534/genetics.115.180968

831 67. Seidl, M. F., & Thomma, B. P. H. J. Transposable elements direct the coevolution between plants
832 and microbes. *Trends Genet.* **33**, 842–851 (2017). https://doi.org/10.1016/j.tig.2017.07.003

833 68. Anisimova, M., Nielsen, R., & Yang, Z.. Effect of recombination on the accuracy of the likelihood
834 method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229–1236 (2003).
835 https://doi.org/10.1093/genetics/164.3.1229

836 69. Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. Rapidly evolving genes are key players
837 in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici*
838 (*Mycosphaerella graminicola*). *PLoS Pathog.* **11**, e1005055 (2015).
839 https://doi.org/10.1371/journal.ppat.1005055

840 70. Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. Dissecting the biology of the fungal wheat
841 pathogen *Zymoseptoria tritici*: A laboratory workflow. *Curr. Protoc. Microbiol.* **59**, e128 (2020).
842 https://doi:10.1002/cpmc.128

843 71. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
844 data. *Bioinformatics* **30**, 2114–2120 (2014). https://doi.org/10.1093/bioinformatics/btu170

845 72. Goodwin, S. B., et al. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*
846 reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.* **7**,
847 e1002070 (2011). https://doi.org/10.1371/journal.pgen.1002070

848 73. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform.
849 *Bioinformatics* **25**, 1754–1760 (2009). https://doi.org/10.1093/bioinformatics/btp324

850 74. Li, H., et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
851 (2009). https://doi.org/10.1093/bioinformatics/btp352

852 75. Garrison, E., & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv*,
853 1207.3907 (2012). http://arxiv.org/abs/1207.3907

854 76. Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. Considering transposable element
855 diversification in *de novo* annotation approaches. *PLoS ONE* **6**, e16526 (2011).
856 https://doi.org/10.1371/journal.pone.0016526

857 77. Chang, C. C., et al. Second-generation PLINK: rising to the challenge of larger and richer datasets.
858 *GigaScience* **4** (2015). https://doi.org/10.1186/s13742-015-0047-8

859 78. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. PopLDdecay: a fast and effective tool
860 for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**,
861 1786–1788 (2019). https://doi.org/10.1093/bioinformatics/bty875

862 79. Pritchard, J. K., Stephens, M., & Donnelly, P. Inference of population structure using multilocus
863 genotype data. *Genetics* **155**, 945–959 (2000). https://doi.org/10.1093/genetics/155.2.945

864 80. Evanno, G., Regnaut, S., & Gouget, J. Detecting the number of clusters of individuals using the
865 software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
866 https://doi.org/10.1111/j.1365-294X.2005.02553.x

867 81. Kang, H. M., et al. Efficient control of population structure in model organism association mapping.
868 *Genetics* **178**, 1709–1723 (2008). https://doi.org/10.1534/genetics.107.080101
869 82. Lipka, A. E., et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**,
870 2397–2399 (2012). https://doi.org/10.1093/bioinformatics/bts444
871 83. Segura, V., et al. An efficient multi-locus mixed-model approach for genome-wide association
872 studies in structured populations. *Nat. Genet.* **44**, 825–830 (2012). https://doi.org/10.1038/ng.2314
873 84. Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. Haploview: analysis and visualization of LD and
874 haplotype maps. *Bioinformatics* **21**, 263–265 (2005). https://doi.org/10.1093/bioinformatics/bth457
875 85. Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. RNA-seq based gene annotation and
876 comparative genomics of four fungal grass pathogens in the genus Zymoseptoria identify novel
877 orphan genes and species-specific invasions of transposable elements. *G3-Genes Genom. Genet.* **5**,
878 1323-1333 (2015). https://doi.org/10.1534/g3.115.017731
879 86. Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. Improved prediction of signal
880 peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004). https://doi.org/10.1016/j.jmb.2004.05.028
881 87. Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. Predicting transmembrane protein
882 topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–
883 580 (2001). https://doi.org/10.1006/jmbi.2000.4315
884 88. Sperschneider, J., & Dodds, P. N. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors
885 in fungi and oomycetes. *Mol. Plant-Microbe Interact.* **35**, 146–156 (2021).
886 https://doi.org/10.1094/MPMI-08-21-0201-R
887 89. Trapnell, C., et al. Differential gene and transcript expression analysis of RNA-seq experiments with
888 TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012). https://doi.org/10.1038/nprot.2012.016
889 90. Bankevich, A., et al. SPAdes: A new genome assembly algorithm and its applications to single-cell
890 sequencing. *J. Comput. Biol.* **19**, 455–477 (2012). https://doi.org/10.1089/cmb.2012.0021
891 91. Camacho, C., et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
892 https://doi.org/10.1186/1471-2105-10-421
893 92. Katoh, K., & Standley, D. M. MAFFT multiple sequence alignment software version 7:
894 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
895 https://doi.org/10.1093/molbev/mst010
896 93. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. Jalview Version
897 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191
898 (2009). https://doi.org/10.1093/bioinformatics/btp033
899 94. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. trimAl: a tool for automated alignment
900 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
901 https://doi.org/10.1093/bioinformatics/btp348
902 95. Rozas, J., et al. DnaSP 6: DNA Sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.*
903 **34**, 3299–3302 (2017). https://doi.org/10.1093/molbev/msx248
904 96. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
905 https://doi.org/10.7312/nei-92038
906 97. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
907 *Genetics* **123**, 585–595 (1989). https://doi.org/10.1093/genetics/123.3.585
908 98. Quinlan, A. R., & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
909 *Bioinformatics* **26**, 841–842 (2010). https://doi.org/10.1093/bioinformatics/btq033
910 99. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. RDP4: Detection and analysis of
911 recombination patterns in virus genomes. *Virus Evol.* **1** (2015). https://doi.org/10.1093/ve/vev003

100. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098 (2006). https://doi.org/10.1093/bioinformatics/btl474

101. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). https://doi.org/10.1093/bioinformatics/btu033

102. Yang, Z., & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000). https://doi.org/10.1016/S0169-5347(00)01994-7

103. Nielsen, R., & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998). http://www.genetics.org/content/148/3/929.abstract

104. Yang, Z., Wong, W. S. W., & Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005). https://doi.org/10.1093/molbev/msi097

105. Zwiers, L.-H., & de Waard, M. A. Efficient Agrobacterium tumefaciens-mediated gene disruption in the phytopathogen *Mycosphaerella graminicola*. *Curr. Genet.* **39**, 388–393 (2001). https://doi.org/10.1007/s002940100216

106. Motteram, J., et al. Aberrant protein N-glycosylation impacts upon infection-related growth transitions of the haploid plant-pathogenic fungus *Mycosphaerella graminicola*. *Mol. Microbiol.* **81**, 415–433 (2011). https://doi.org/10.1111/j.1365-2958.2011.07701.x

Acknowledgements

This work was supported by the French National Research Agency (program BIOADAPT, ANR-12-ADAP-0009-04) for T.C.M. and J.A., the FSOV 2008 B project for T.C.M., and BASF France SAS for R.A. postdoctoral fellowship. T.C.M. was supported by a Short-Term Scientific Mission granted by the COST SUSTAIN action FA1208. A.S.V. was supported by the Ministerio de Ciencia e Innovación (Ramón y Cajal program, grant no. RYC2018-025530-I). INRAE BIOGER benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). We thank David Gouache (ARVALIS Institut du Végétal, Boigneville, France), Marc-Henri Lebrun, Henriette Goyeau, Frédéric Suffert and Anne-Sophie Walker (INRAE UR BIOGER, Palaiseau, France) for their help in establishing the fungal collection; Robert King and Jason Rudd (Rothamsted Research, Harpenden, United Kingdom) for providing RNAseq data; Gabriel Scalliet (Syngenta Crop Protection AG, Stein, Switzerland) for providing the vectors pNOV2114, pNOV_3Gate_SUL and pNOV_3Gate_HYG used to clone the candidate pathogenicity genes; Martin Willigsecker and Béatrice Beauzoone (INRAE UR BIOGER) for their help in the pathology assays; and Thierry Rouxel (INRAE UR BIOGER) for critical reading of the manuscript and valuable feedback.

Author Contributions Statement

T.C.M and R.A. conceived and designed the study. G.H.J.K. and H.G. isolated and provided most of the natural *Z. tritici* isolates constituting the GWAS panel. A.D., T.C.M. and Jo.C. carried out pathogenicity assays on the natural isolates, and R.A. on the transformant isolates. A.D. and T.C.M. prepared the DNA

samples for whole-genome resequencing. Je.C., N.L. and J.A. performed the bioinformatics analyses, including SNP calling and *de novo* genome assemblies. D.C. provided sequence data of the global *Z. tritici* collection. R.A., L.D., A.P. and A.S.V. performed the cloning and genetic transformations. R.A. performed the genome-wide association, gene expression and evolutionary sequence analyses. R.A. and T.C.M. wrote the manuscript. All authors have reviewed the manuscript.

Competing Interests Statement

The authors declare they have no competing interests.