Whole-genome sequencing reveals diverse mechanisms underlying quantitative pathogenicity and host adaptation in a fungal plant pathogen

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, 91120 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; reda.amezrou@inrae.fr

Word count (main text): 5,364

Abstract

Knowledge of genetic determinism and evolutionary dynamics mediating host-pathogen interactions is essential to manage fungal plant diseases. However, the genetic architecture of fungal pathogenicity remains poorly understood, and studies often focus on large-effect effector genes triggering strong, qualitative resistance. It is not clear how this translates to predominately quantitative pathogens. Here, we used the *Zymoseptoria tritici*-wheat model to elucidate the genetic architecture of quantitative pathogenicity and mechanisms mediating host adaptation. *Z. tritici* is a globally occurring pathogen that causes severe yield losses on wheat. We perform whole-genome sequencing of 103 isolates and quantified pathogenicity traits on 12 cultivars harbouring different resistance genes. We perform a multi-host GWAS and identified 58 candidate genes associated with pathogenicity, of which nineteen are highly expressed and/or differentially expressed *in planta*. Two of these had large effects and three were shared in more than one cultivar, suggesting that *Z. tritici* pathogenicity is predominantly quantitative and host-specific. Analysis of genetic diversity revealed that sequence polymorphism is the main evolutionary process mediating differences in quantitative pathogenicity, a process that is likely facilitated by genetic recombination and transposable elements dynamics. We found signatures of positive diversifying selection in ~68% of the candidate genes acting on specific amino acid substitutions, likely responsible for evasion of host recognition. Finally, we used functional approaches to confirm the role of an effector-like gene and a methyltransferase in quantitative pathogenicity. 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 (Savary et al., 2019). 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 (Marina et al., 2022). In turn, plants respond to pathogen invasion by activating defence mechanisms based on cell surface and intracellular receptors recognition of pathogen-associated molecular patterns and effectors, respectively. The discovery of resistance genes (*R*) encoding receptors has provided valuable means to control fungal diseases through resistance breeding (lo Presti et al., 2015). However, pathogens have repeatedly overcome host resistance by evading recognition or suppression of host immunity. These rapid adaptations are mainly driven by genetic variation at effector loci and genome evolution. Gene diversification, deletions or horizontal acquisitions generated adaptive variants at loci encoding effectors. Transposable elements (TEs), high mutation and recombination rates are thought to considerably contribute to the extensive genome variation in many fungal species (Möller & Stukenbrock, 2017; Raffaele & Kamoun, 2012; Rouxel et al., 2011; Stukenbrock & Dutheil, 2018). In the case of antagonistic host-pathogen interactions, the fate of genetic variation is partly determined by the selection pressure exerted by host *R* proteins. This is especially true when *R* genes and effectors are involved in a gene-for-gene (GFG) relationship. Moreover, the diversity of host genotypes in *R* gene content exert heterogeneous selection pressure on pathogen populations, giving rise to the co-maintenance of multiple pathogen variants better adapted on specific subsets of hosts (lo Presti et al., 2015; Stukenbrock & McDonald, 2009). Hence, in addition to effector genes identification, elucidating the mechanisms contributing to gene diversification provide critical information about their evolvability. This knowledge has tremendous applications for the durability of crop protection strategies (van de Wouw & Idnurm, 2019).

The first fungal effectors to be identified were small secreted proteins (SSP). SSPs trigger strong resistance (*i.e.* encoded by avirulence genes) as a result of their recognition by plant immune receptors following a typical GFG relationship (Domazakis et al., 2017). The increasing availability of fungal genomes over the past decade allowed significant advances in investigating molecular plant–fungal interactions and facilitated effector discovery. However, effector biology has focused almost exclusively on SSPs despite the growing evidence that other genetic factors are also involved in host colonization (Varden et al., 2017). For instance, small noncoding RNAs and fungal secondary metabolites play similar effector roles in various plant pathogenic fungi (Collemare et al., 2019). Other fungal proteins were shown to interact with the host immune system or act to promote pathogenicity; these include for example cell wall–degrading enzymes (Kubicek et al., 2014), protease inhibitors (van der Hoorn, 2008) or disruptors of the hormone-signalling pathway (Ma & Ma, 2016).

Effector discovery also led to the systematic search of host genes conferring resistance (Domazakis et al., 2017). Resistance to many fungal diseases can be divided into two categories: (*i*) qualitative (or race-specific) resistance, which is usually controlled by large-effect genes and is effective against avirulent strains; and (*ii*) quantitative (or non–race-specific) resistance, usually controlled by genes with small-to-moderate effects and is thought to be less specific. While cloning plant immune receptors remains challenging, the discovery of the interacting fungal effectors is considered more straightforward (van de Wouw & Idnurm, 2019). However, fungal pathogenicity studies often focus on large effect genes, whose interaction with genetic factors from the host trigger a strong immune response, leading to qualitative resistance. It remains unclear how these findings translate to quantitative models, where the interaction of quantitative components of pathogenicity (also called aggressiveness in the plant pathology literature) and plant resistance trigger a partial/incomplete immune response. By contrast to quantitative resistance in plants, the genetic basis of quantitative pathogenicity in pathogens and the process of host adaptation have received relatively little attention. 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 cause 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 (O'Driscoll et al., 2014). *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. This stage is accompanied by the development of pycnidia containing asexual spores that further spread the disease (Duncan & Howard, 2000; Kema et al., 1996). To date, three *Z. tritici* pathogenicity genes encoding 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 (Amezrou *et al.*, 2022; Kema et al., 2018; Zhong et al., 2017), and *Avr3D1* which triggers quantitative resistance on cultivars harbouring *Stb7* or *Stb12* (Meile et al., 2018). 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 (Brown et al., 2015), but only *Stb6* and *Stb16q* have been cloned to date (Saintenac et al., 2018, 2021). 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 extensively used in plants but remains at its infancy on fungal populations (Bartoli & Roux, 2017; Sánchez-Vallet et al., 2018).

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. To this end, we performed whole-genome sequencing of 103 *Z. tritici*

112 isolates that we quantitatively phenotyped on 12 wheat genotypes harbouring multiple *Stb* resistance
113 genes. We used association genetics to identify candidate genes responsible for differences in
114 pathogenicity and we provide multiple lines of evidence including transcriptomics, allelic analyses and
115 functional annotation to draw a short-list of the most robust candidates. In addition, we used population
116 genomics approaches to investigate the diversification and evolution of quantitative pathogenicity
117 genes. Finally, we confirmed the role of two genes putatively mediating quantitative pathogenicity using
118 functional approaches. Together, our results generate a comprehensive view on the complex genetic
119 architecture of quantitative pathogenicity and pathogen adaptation to host resistances. Beyond this
120 model, our findings may have broad implications for sustainable disease management.
121

122 **Results**

123 **Characterization of the fungal collection.** To efficiently perform GWAS and reduce confounding
124 effects, it is important to select a population that is genetically and phenotypically diverse with little
125 stratification. Following these guidelines, we focused on a subset of 103 French *Z. tritici* isolates selected
126 from a larger population for their genetic diversity (data not shown). We sequenced their genomes and
127 generated a total of 247 Gb of sequence data with an average depth of 61x. After quality assessment, we
128 obtained a total of 1,463,638 SNPs that we further filtered for MAF and missingness, resulting in a final
129 set of 718,810 SNPs. Analysis of functional variant effect revealed that 59,318 SNPs (8.25%) induced
130 missense mutations and 2,493 (0.34%) resulted in a loss of function such as premature stop codons, loss
131 of stop or start codons (Table 1). As expected, principal component analysis (PCA) using full
132 polymorphism data did not show any genetic structure; the first three PCs explained less than 4% of
133 total variance (Figure S1). We further confirmed this with STUCTURE using a subset of evenly
134 distributed SNPs (1 SNP per 5kb) and like the PCA analysis, Bayesian clustering failed to infer sub-
135 populations (Figure 1a). Similarly, individual isolates showed very low genetic relatedness (Figure S2).
136 Genome-wide LD decayed (r^2<0.2) over physical distance at ~0.5kb (Figure 1b). Given our high SNP
137 density (1 SNP every 55bp) we expected marker-trait associations to be resolved to individual gene
138 level. When pathotyped on wheat differential cultivars, isolates showed a wide virulence spectrum
139 depending on the host cultivar, and strong isolate × cultivar interactions (Tables S1, S2). Overall, strains
140 isolated from the cultivar 'Premio' were more aggressive compared to the ones isolated from the
141 susceptible cultivar 'Apache', even when tested on the cultivar 'Apache' (Figure S3).

142 **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

A **B**

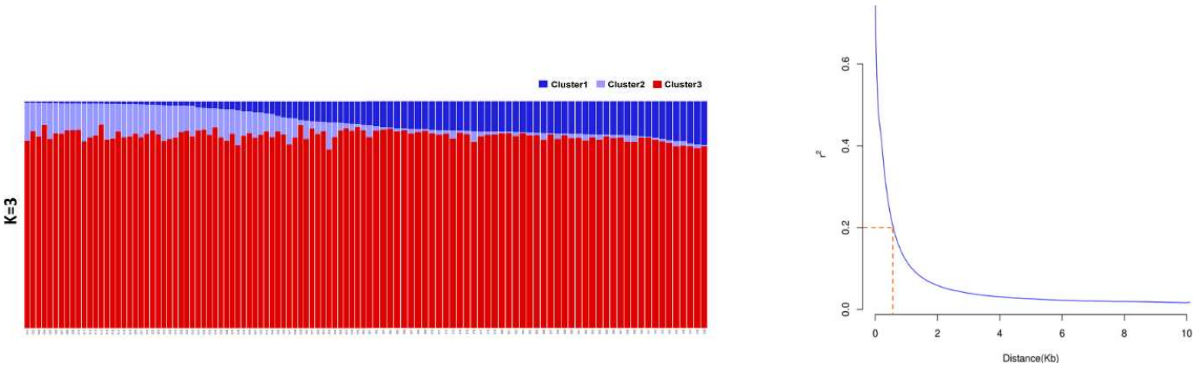

Figure 1. Population structure and Linkage Disequilibrium (LD) in the *Z. tritici* population. **A.** Admixture plot of the fungal population. Each vertical bar represents a single isolate and is colored according to the membership coefficient (Q_i) to the sub-population clusters from STRUCTURE. **B.** 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 (0.5kb) at which LD decays at $r^2 < 0.2$ 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 association tests. We identified 94 genome-wide significant SNPs and an additional 53 at the 10% FDR threshold, defining 58 distinct genes (Figure 2, Table S4). The strongest associations were at the *AvrStb6* and *AvrStb9* loci and were detected in all *Stb6*- and *Stb9*- 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 (data not shown). Interestingly, the single-locus model failed to detect *AvrStb9* and demonstrates the usefulness of multi-locus mixed model in detecting loci with larger effect. 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 (Figure S4). Candidate pathogenicity loci were identified in the thirteen *Z. tritici* core chromosomes. Despite recent suggestions of accessory chromosomes being enriched in putative effectors (Habig et al., 2017), we could not map any gene residing on accessory chromosomes in this

multi-host GWAS (Figure 2, Table S4). Only eleven genes were shared among PLACP and PLACN (Figure 2, Table S4), which was expected, since necrosis coverage and pycnidia production levels were highly variable (Figure S3). Hence, considering both traits in association analyses could reveal genetic determinants influencing both, the asexual reproduction and host susceptibility. The phenotypic variance explained by the identified loci ranged from 9.08% to 65.70% (Table S4). With the exception of *AvrStb6* and *AvrStb9*, all of the detected genes accounted for small-to-moderate effects (R^2<30) and similarly, virulent/avirulent alleles analysis showed rather quantitative differences (Table S4, Figure 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 (Table S4, Figure 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.

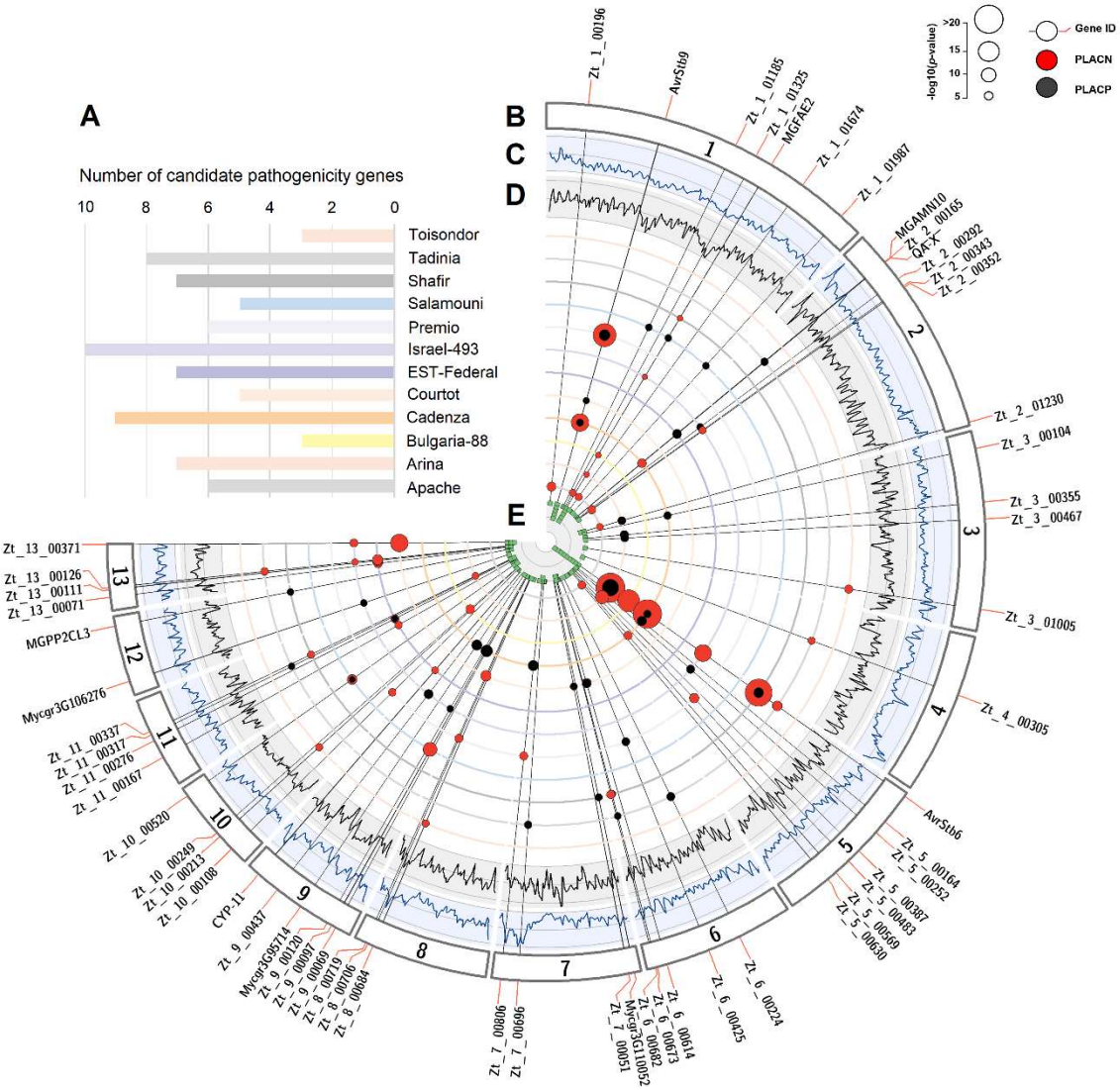

Figure 2. Overview of the identified candidate pathogenicity genes from the multi-host GWAS. **A.** Number of identified genes per cultivar. **B.** Gene ID and their significant associations. Dots correspond to PLACP (black color) and PLACN (red color) marker-trait associations. The larger the dot the 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).

***In planta* gene expression and functional annotation.** To reveal the expression profile of the 58 candidate pathogenicity genes, we calculated the relative abundance of *Z. tritici* transcripts, using RNAseq data generated during different *Z. tritici* infection time courses (Rudd et al., 2015). Ten genes exhibited no expression (FPKM = 0) or very low expression (FPKM < 5) throughout the infection cycle. Another ten genes were found to be highly expressed (FPKM above the 95th percentile of mean expression value) either at specific time points or throughout the infection cycle (Figure 3c; Table S5). Interestingly, *AvrStb6*, *Zt_6_00224* and *Zt_3_00467* were among the top 100 highly expressed *Z. tritici* genes at the disease transitional phase (9dpi; Table S5). To identify genes differentially expressed *in planta*, we examined the ratio of expression between the CDB culture medium and the five time points of wheat infection. We found sixteen differentially expressed genes (DEGs) with a FDR below 0.10 and a log2|fold change| greater than 2. Among them, eleven were up-regulated at specific time points (Figure 3c; Table S5). For subsequent investigations, we considered a set of nineteen genes as high-confidence candidates, whose expression profile supported their potential role in pathogenicity: this included the sixteen DEGs and three additional ones highly expressed during the infection (Figure 3).

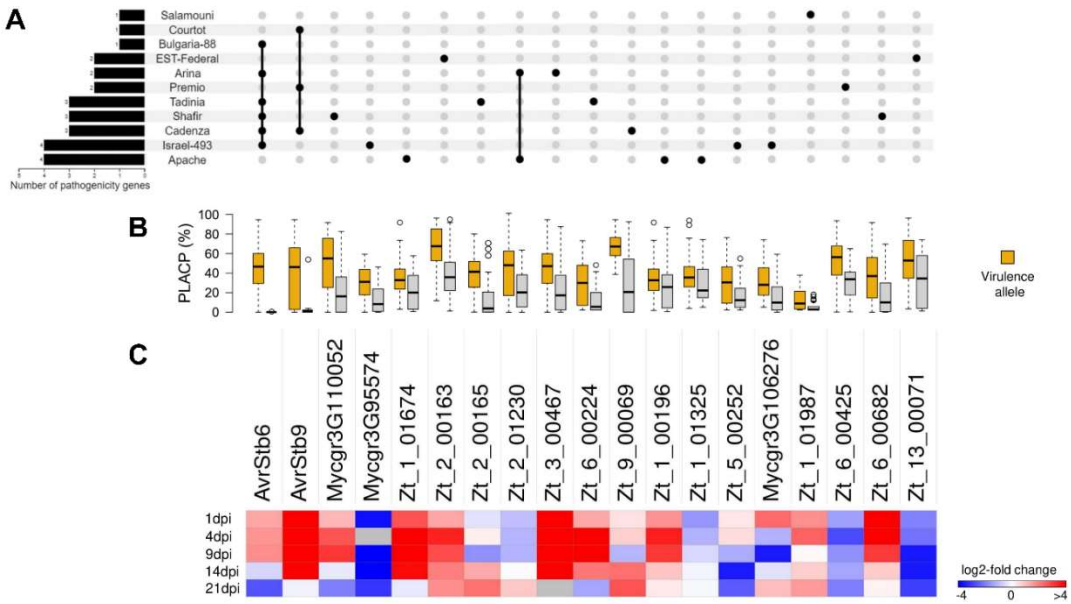

Figure 3. The quantitative nature and host-specificity of *Z. tritici* pathogenicity. **A.** An upset plot of candidate pathogenicity genes across host cultivars. **B.** Boxplots showing variation in quantitative pathogenicity (PLACP) of the isolates carrying different lead SNP alleles of the candidate pathogenicity genes **C.** Differential gene expression shown as log2-fold change compared to the culture medium, CDB. Red shades indicate up-regulated genes and blue shades indicate down-regulated genes during the infection cycle.

207 Eleven protein products out of the nineteen retained as best candidates were predicted to be secreted
208 (Table S6). These secreted proteins could be effectors promoting disease or interacting with the cognate
209 *Stb* resistance genes used in this study. We also found three proteins with predicted transmembrane
210 helices (THMs), a class that is systematically discarded in effector discovery studies. EffectorP predicted
211 seven as putative effectors, with five (excluding *AvrStb6*) lacking homology with other fungal species
212 and might be considered as *Z. tritici*-specific candidate effectors (Table S6). The identified conserved
213 protein domains included functions such as hydrolase (*e.g.* Glycosyl hydrolase), oxidoreductase (*e.g.*
214 Thioredoxin reductase) or transferase activities (*e.g.* Methyltransferase; Table S6). Using PHIB-Blast,
215 we found that seven candidates shared homology with known pathogenicity factors from other
216 pathogenic fungi and therefore constitute promising candidates mediating quantitative pathogenicity in
217 *Z. tritici* (Table S6).

218 **Factors of genetic diversity in pathogenicity genes.** Gain in virulence and/or escape from host
219 recognition is often mediated by polymorphisms in pathogenicity genes. We investigated sequence
220 variation and diversity in the identified genes and detailed results are summarized in Table S7. In
221 particular, we compared patterns of nucleotide diversity in candidate pathogenicity genes with the
222 genome-wide levels. We observed higher genetic diversity in pathogenicity genes compared to the rest
223 of protein-coding genes (median $\pi_{Pathogenicity}$=0.0212 vs. $\pi_{Genome\text{-}wide}$=0.0079; $P < 0.01$; Figure 4A). We
224 found no significant impact of loss of function mutations or presence/absence polymorphism, except for
225 *Mycgr3G110052*. This suggests that differences in quantitative pathogenicity is mainly due to sequence
226 polymorphism. In addition, we wanted to examine which mechanisms were likely to contribute to the
227 high sequence diversity in our candidates. As TE dynamics have been observed to contribute to fungal
228 genome plasticity and diversity, we calculated the proximity of candidate pathogenicity genes to the
229 nearest TEs. 52.63% of these genes flanked by or adjacent to TEs (*i.e.* at less than 5kb distance; Table
230 S7). Comparison with genome-wide genes with regards to TEs proximity revealed that candidate
231 pathogenicity genes were significantly closer to TEs ($P < 0.05$; Figure 4B). Gene diversification is also
232 accelerated by meiotic recombination, which influences both natural population dynamics and selection.
233 We used a combination of algorithms in two independent software (RDP4 and GARD) to detect
234 signatures of recombination. We found evidence of recombination with both detection methods in
235 63.15% of the putative pathogenicity genes (Table S8). Interestingly, genes with little or no signatures
236 of recombination were in close proximity to TEs, whereas genes with higher recombination events were
237 found at greater distance from TEs (Figure 4C). We hypothesize that recombination in particular and
238 proximity to TEs promote allelic diversity in genes putatively involved in quantitative pathogenicity.

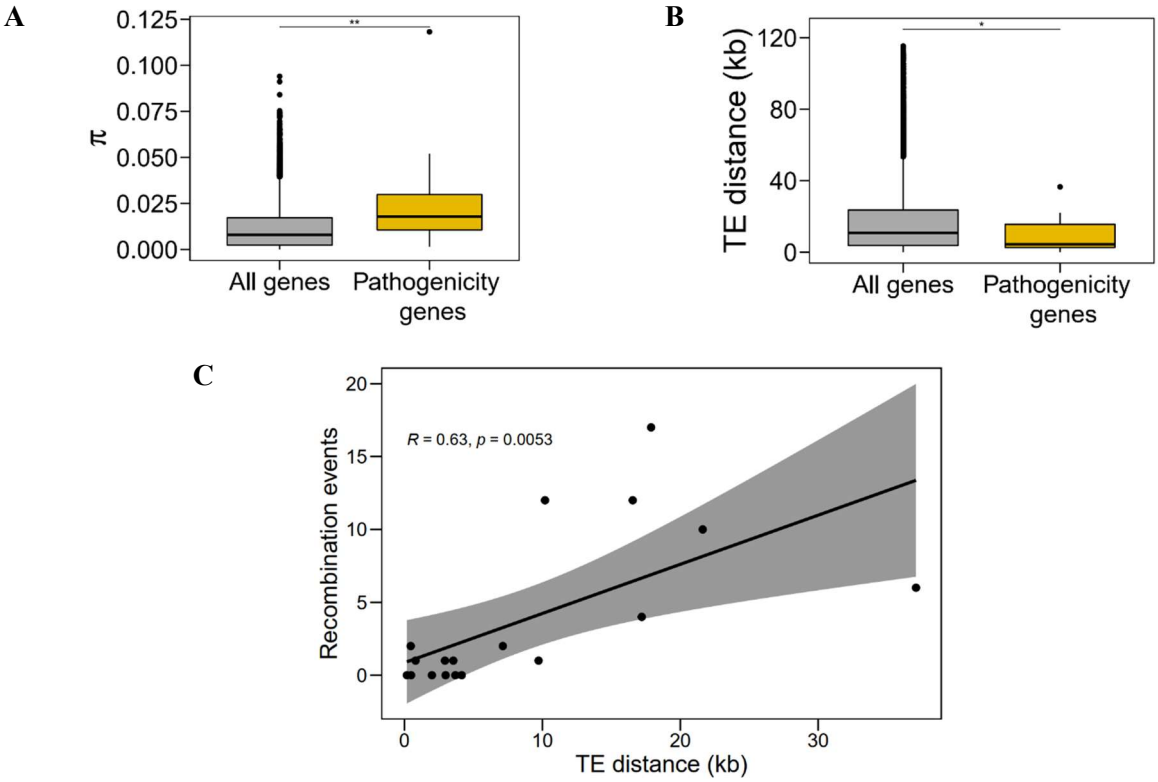

Figure 4. Genetic diversity of quantitative pathogenicity genes and factors contributing to gene diversification. **A.** Distribution of nucleotide diversity (π) among candidate pathogenicity genes and the rest of the protein-coding genes. Statistical significance was determined with the Mann-Whitney test; $P<0.01$**. **B.** TE distance compared between pathogenicity genes and the rest of the protein-coding genes in *Z. tritici*. Statistical significance was determined with the Mann-Whitney test; $P<0.05$*. **C.** the relationship between TE distance and recombination events detected through RDP in candidate pathogenicity genes.

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 ($\omega > 1$) compared to the expectation under neutral evolution, $\omega = 1$. The proportion of positively-selected sites ($\omega > 1$) varied greatly among genes and ranged from 0.42% in case of *Zt_1_01987* to 10.97% for *AvrStb6* (Table S7). We also found four genes with a predicted signal peptide having 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. Nevertheless, since ω averaged across all gene-tree branches rarely surpasses 1, and as the selective pressure could affect codon sites differently, we used different codeML models to detect positively-selected codons. For all model comparisons, we found signatures of positive diversifying selection in ~76.5% of candidate genes using both, the more conservative M1a vs. M2a and the M7 vs. M8 model pairs (Table S7). Together, this shows that pathogenicity genes are hotspots of positive diversifying selection, likely as a result of a selection pressure exerted by host *R* genes.

271 **Two functionally distinct genes mediate quantitative pathogenicity.** To determine the role of the
272 candidate genes in quantitative pathogenicity, we selected two distinct genes for functional validation:
273 *Zt_3_00467* (Figure 5A), which encodes a SSP highly expressed *in planta*, with a peak at the disease
274 transition phase (Figure 5B); and *Zt_6_00682* (Figure 6A), which encodes a S-adenosylmethionine-
275 dependent methyltransferase (SAM-MTase) that is differentially expressed during the disease
276 asymptomatic phase (Figure 6B). For *Zt_3_00467*, we ectopically expressed the avirulent (IPO10273)
277 and virulent (3D7) alleles in the 3D7 strain background, and assessed quantitative pathogenicity traits
278 on the host cultivars 'Arina' and 'Riband'. Overall, Mutants expressing the avirulent allele (*3D7* +
279 $Zt_3_00467_{IPO10273}$) showed a slight reduction of PLACP levels compared to the ones expressing the
280 virulent allele (*3D7* + $Zt_3_00467_{3D7}$; Figure 5C). When compared individually, one mutant expressing
281 the avirulent allele showed a phenotype similar to the wild-type, while another expressing the virulent
282 allele had higher PLACP compared to the wild-type (Figure S7). The difference in phenotypic responses
283 between ectopic mutants is a well-known phenomenon that may be caused by distinct allele integration
284 sites in the different mutant lines. Testing transformants expressing the avirulent allele on the cultivar
285 'Riband' led to lower development of pycnidia compared to the virulent and the wild-type strains (Figure
286 S8). 'Arina' and 'Riband' share a common resistance gene known as *Stb15*. These phenotypic
287 observations suggest that *Zt_3_00467* recognition leads to quantitative resistance which could be
288 attributed to *Stb15*. Comparison of the avirulent and virulent protein sequences showed seven amino
289 acid substitutions, all positively-selected (Figure 5D). Analysis of *Zt_3_00467* in the global *Z. tritici*
290 collection showed that the gene was conserved and that most protein variants were population-specific
291 and segregating by geographical origin (Figure 5E).

292 For *Zt_6_00682*, we generated knockout (KO) mutants of the 3D7 strain by targeted gene disruption
293 and ectopic transformants with the virulent (3D7) and avirulent (3D1) alleles in the KO background.
294 This allow to assess allele-dependent effects of *Zt_6_00682.* Three independent KO and ectopic
295 transformants for each allele were tested on the cultivar 'Shafir', where the association was detected.
296 Compared to the wild-type, KO transformants resulted in unexpectedly high PLACP values ranging
297 from 67% to 100% (Figure 6C). Ectopic transformation of 3D7 allele into 3D1 in the KO background
298 (3D7Δ*Zt_ 6_00682*+ $Zt_6_00682_{3D1}$) resulted in reduced levels of PLACP, while ectopic lines
299 transformed with the 3D7 allele (3D7Δ*Zt_ 6_00682*+ $Zt_6_00682_{3D7}$) showed a slightly higher PLACP
300 than the wild-type (Figure 6C). Besides 'Shafir', deletion mutants were aggressive on another four
301 cultivars, but phenotypic differences between ectopic transformants and wild-type were not significant
302 (Figure S6). Five amino acid differences were found between the avirulent and virulent protein
303 sequences, of which three were under positive selection (Figure 6D). Similar to *Zt_3_00467,*
304 *Zt_ 6_00682* was conserved in the global population but was highly polymorphic (Figure 6E).

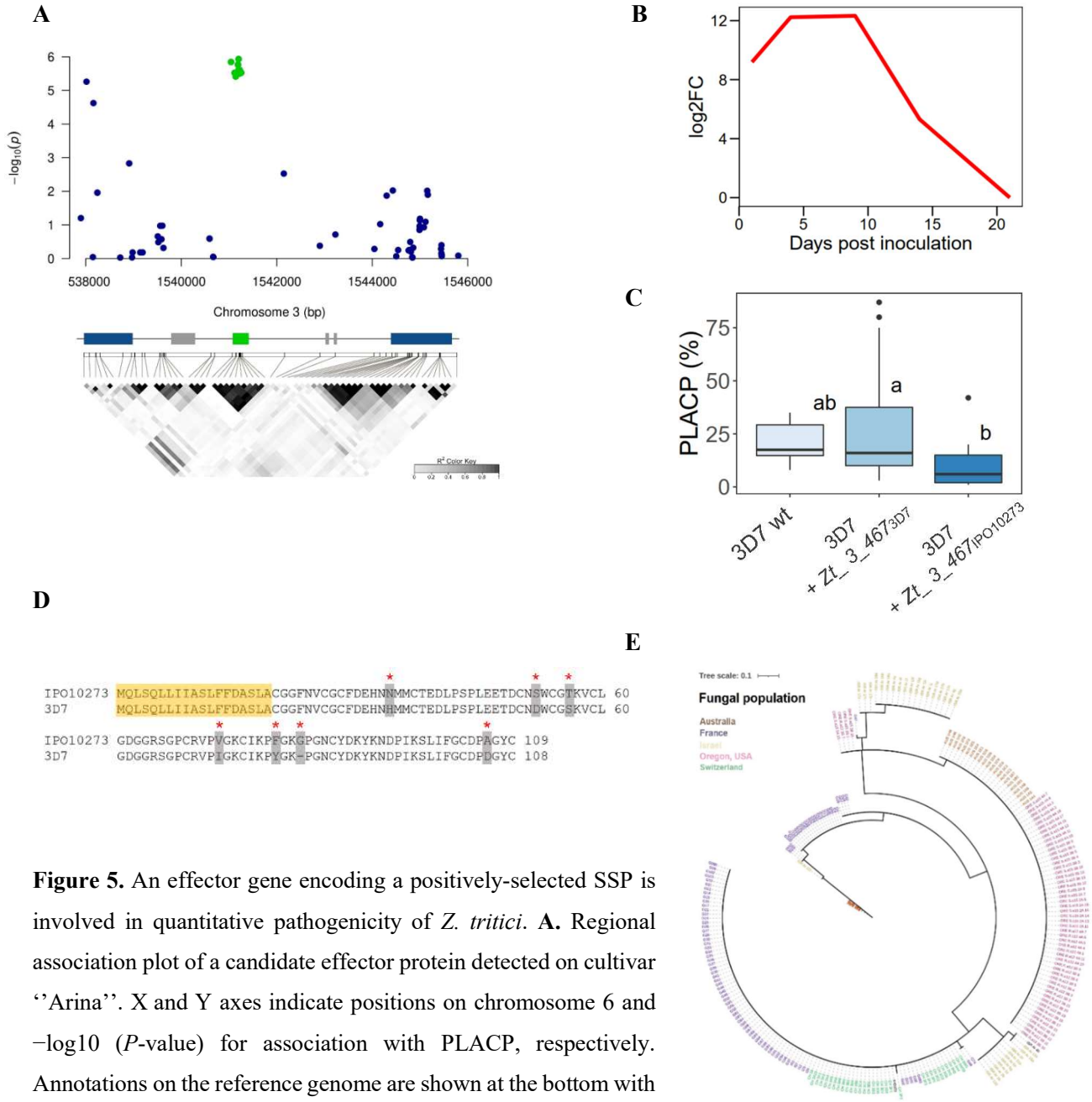

Figure 5. An effector gene encoding a positively-selected SSP is involved in quantitative pathogenicity of *Z. tritici*. **A.** Regional association plot of a candidate effector protein detected on cultivar ''Arina''. X and Y axes indicate positions on chromosome 6 and −log10 (*P*-value) for association 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. 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.** PLACP produced by the wild-type strain 3D7 and the ectopic transformants expressing the virulent (3D7) and avirulent (IPO10273) alleles. Statistical significance was determined using 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. **E.** Phylogenetic tree of *Zt_3_00467* generated from the protein sequence of the global *Z. tritici* isolates.

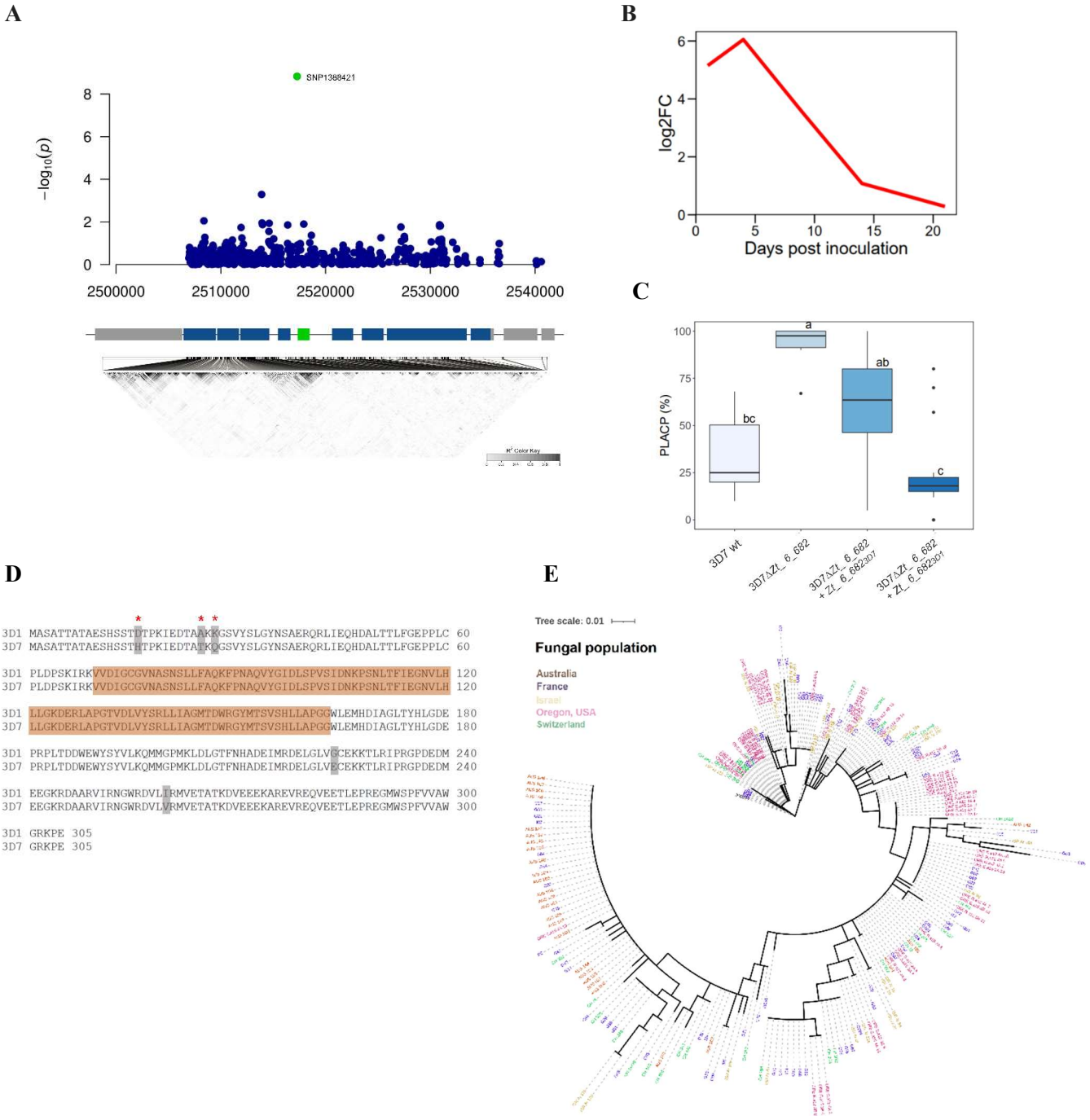

Figure 6. A gene encoding a methyltransferase mediates quantitative pathogenicity in *Z. tritici*. **A.** Regional association plot of the pathogenicity gene detected on cultivar ''Shafir'', as described in Figure 5. **B.** log2-fold expression changes of *Zt_6_00682* based on RNAseq transcription data collected throughout the time course of *Z. tritici* infection on wheat. **C.** PLACP produced by the wild-type strain 3D7, knockout mutants and the complementation mutants carrying the virulent (3D7) and avirulent (3D1) alleles. Statistical significance was determined using a Tukey's HSD test. **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. **E.** Phylogenetic tree of *Zt_6_00682* generated from the protein sequence of the global *Z. tritici* isolates.

352 **Discussion**

353 This work describes the genome-wide identification of genes involved in pathogenicity of an
354 economically important plant pathogen. Our method involves whole-genome sequencing, quantitative
355 pathogenicity measurements on different wheat cultivars, analysis of gene expression and extensive
356 functional annotation to provide clues about the genetic determinants of quantitative pathogenicity. The
357 candidate genes identified using these complementary approaches exhibited high levels of genetic
358 diversity and signatures of positive diversifying selection. We show that proximity to TEs and/or
359 recombination are likely important contributing factors to candidate gene diversification, mainly
360 through the accumulation of non-synonymous SNP mutations. Using functional approaches, we were
361 able to demonstrate the involvement of two positively-selected genes in quantitative pathogenicity,
362 which further support our findings.

363 **Pathogenicity has a complex polygenic architecture.** Our results point to a complex genetic
364 architecture of pathogenicity in *Z. tritici*, involving many genes with variable additive effects, resulting
365 in quantitative phenotypes on the host cultivar. These findings are consistent with previous studies
366 (Hartmann et al., 2017; Stewart et al., 2018) and provides further evidence that pathogenicity in *Z. tritici*
367 is complex and results from the contribution of both large and small effect genes. Using RNAseq data,
368 we highlighted nineteen differentially and highly expressed genes *in planta* and focused on their features
369 for subsequent investigations. The vast majority of these genes had a peak expression at 4dpi and/or
370 9dpi, which corresponds to the colonization and the pre-pycnidia formation phases, respectively, before
371 the apparition of symptoms. During colonization, hyphae penetrates into host tissue *via* stomata and
372 colonize the substomatal cavity, but remain exclusively in the intercellular space. This is followed by
373 hyphal invasion of the apoplast leading to pre-pycnidia formation in substomatal cavities (Duncan &
374 Howard, 2000). It was shown that these are the key phases of effector genes deployment by plant
375 pathogens (Mirzadi Gohari et al., 2015). These generally encode for small secreted cysteine-rich proteins
376 with no homology in other species and play a vital role in successful pathogen invasion, suppression of
377 host immune perception and manipulation of its physiology (Toruño et al., 2016). Here, we mapped a
378 set of diverse secreted proteins. While effector-like proteins may initiate an effector-triggered immune
379 response, as was shown for the gene *Zt_3_00467*, which triggered quantitative resistance in cultivars
380 'Riband' and 'Arina', secreted molecules with a known functional domain could also promote disease.

381 Of the functionally annotated secreted proteins, we identified domains associated with macromolecule
382 degradation, fungal metabolism or stress response, among others. Some genes encoded proteins that
383 show strong homology with functionally characterized virulence factors from other pathogenic fungi.
384 For instance, *Zt_1_01674* encodes a Thioredoxin reductase (*Trr*), a flavoenzyme that plays a central role
385 in the thioredoxin oxidative stress resistance pathway and is involved in regulating gene transcription,
386 cell growth and apoptosis (Arnér & Holmgren, 2000). Deletion of the *Trr* ortholog resulted in reduced
387 hyphal growth, sexual reproduction and virulence in *Fusarium graminearum* (Fan et al., 2019) and

388 profoundly affected pathogenicity of the entomopathogenic fungus *Beauveria bassiana* (L.-B. Zhang et
389 al., 2016). We hypothesize that *Zt_1_01674* may play similar roles and could be an important factor in
390 *Z. tritici* development and pathogenicity. We also identified a cell wall degrading enzyme (*Zt_2_00163*)
391 belonging to the Glycoside Hydrolase protein family. Glycoside hydrolases are known to play important
392 roles in fungal invasion and nutrition acquisition via the degradation or modification of plant cell wall,
393 such as cellulose or hemicellulose (Esquerré-Tugayé et al., 2000).

394 Nonetheless, genes encoding proteins predicted to be non-secreted were largely present in our candidate
395 list, highlighting not only the polygenic nature of pathogenicity, but also the potential role of the non-
396 secretory pathway in fungal pathogenicity. This was demonstrated by the functional studies of
397 *Zt_6_00682*, which encodes for a SAM-dependent Methyltransferase (SAM-MTase). SAM-MTases
398 catalyse the transfer of methyl groups from SAM to a large variety of substrates, ranging from
399 metabolites to macromolecules, including proteins (J. Zhang & Zheng, 2016). They are involved in
400 epigenetic regulation with functions ranging from the control of gene expression, silencing or activation
401 of transposon control (Martienssen & Vincent, 2001). *Zt_6_00682* shares some sequence similarities
402 with the putative SAM-MTase *LaeA* gene, a component of the velvet transcription factor complex that
403 functions as a global regulator of fungal secondary metabolism and development for a handful of fungi
404 (Bayram & Braus, 2012). While the loss of *LaeA* attenuates pathogenicity in a variety of filamentous
405 fungi (reviewed in(Jain & Keller, 2013), the inactivation of *Zt_6_00682* promoted pycnidia formation,
406 suggesting that it functions differently from *LaeA* and may negatively regulate the transcriptional
407 activity of genes or metabolites associated with host colonization. However, how this SAM-MTase
408 affects the regulation of pathogenicity factors needs to be investigated as a key step towards
409 understanding the evolutionary mechanisms associated with emergence of highly aggressive strains. In
410 addition, we identified three genes with predicted transmembrane domains. For instance, the *DPM2*
411 ortholog (*Zt_1_01325*) was highly expressed throughout the infection cycle, highlighting its potential
412 role in *Z. tritici* pathogenicity. *DPM2* is a regulatory subunit of the dolichol-phosphate mannose synthase
413 complex, key enzymes for GPI anchor biosynthesis (Juchimiuk et al., 2015). The alteration of the *DPM*
414 complex affected cell wall integrity, fungal growth and increased chitin levels in *Candida albicans*
415 (Juchimiuk et al., 2015). Traditionally, transmembrane proteins are systematically discarded in effector
416 gene prediction studies. There is however a growing evidence that they could play key roles in
417 pathogenicity of filamentous fungi and our results support this hypothesis (Weiland & Altegoer, 2021;
418 Gu et al., 2015).

419 **Pathogenicity is predominantly host specific and quantitative.** Host specificity in plant pathogenic
420 fungi is mediated by a complex process involving multiple genes and mechanisms. We found substantial
421 evidence for host specificity demonstrated by a highly significant isolate × cultivar interaction and few
422 shared associations between host cultivars. It is also probable that the weak overlap in candidate
423 pathogenicity genes among host cultivars is partly due to a lack in statistical power to detect variants of

424 small effect (Korte & Farlow, 2013). However, besides *Stb6*- and *Stb9*-carrying cultivars, there was no
425 meaningful correlation among phenotypes on different hosts, which further supports host specificity. In
426 several well studied host-pathogen systems, specificity is determined mainly by the matching of
427 resistance or susceptibility genes and their corresponding effectors (Bourras et al., 2015; Faris et al.,
428 2010; Liu et al., 2013), following a GFG relationship (or inverse GFG in case of necrotrophic fungal
429 pathogens).

430 In *Z. tritici*, pathogenicity can be both qualitative (Brading et al., 2002; Kema et al., 2000; Zhong et al.,
431 2017) and quantitative (Meile et al., 2018; Zhan et al., 2016). Here, we showed that pathogenicity is
432 predominantly quantitative based on multiple genes of small-to-moderate effects. This was
433 demonstrated with the allele analysis of candidate genes and functional studies of *Zt_3_00467* and
434 *Zt_6_00682*. In contrast to what has been shown for most GFG interactions, the recognition of
435 *Zt_3_00467* and *Zt_6_00682* avirulent alleles does not trigger strong resistance, but rather a reduction
436 at the onset of disease symptoms. Furthermore, the fact that avirulent alleles are only recognized by
437 certain wheat cultivars suggests that the interaction also follows a GFG model and demonstrates that
438 *Zt_3_00467* and *Zt_6_00682* are likely quantitative avirulence factors interacting with specific host
439 resistance genes. Since *Zt_3_00467* encodes a typical effector protein and is recognized by hosts
440 harbouring the *Stb15* resistance gene, we suggest that *Zt_3_00467* could be *AvrStb15* and trigger an
441 effector-induced defence. We hypothesise that *Stb15* resistance is not strong enough to block pathogen
442 infection, although we cannot completely rule out the existence of a *Zt_3_00467* haplotype that triggers
443 qualitative resistance. However, additional genetic factors might also contribute to differences in
444 pathogenicity between avirulent and virulent strains. On the other hand, *Zt_6_00682* encodes an atypical
445 avirulence factor and may interact indirectly with the host immune system. The fact that *Zt_6_00682*
446 resembles genes involved in transcriptional regulation and that its deletion mutants are more aggressive
447 on some cultivars, suggests that *Zt_6_00682* negatively regulates pathogenicity factors that are
448 specifically recognized by one or more (unknown) host quantitative resistance gene(s).
449 Complementation mutants further support host specificity, where the avirulent allele recognition leads
450 to a dramatic reduction of pycnidia formation. Unlike deletion mutants, this was only observed on the
451 cultivar where the gene was initially detected ('Shafir'). However, the resistance gene involved and the
452 underlying mechanisms through which the avirulent allele is recognized remain unknown.

453 In addition to *AvrStb6* and *AvrStb9*, we found the SSP *Zt_2_01230* to be significantly associated with
454 variation in quantitative pathogenicity on two cultivars, suggesting a common yet unknown *Stb*
455 resistance gene. In addition to the functionally validated effectors, we also identified a number of
456 secreted and species-specific proteins. Since canonical fungal effectors are generally thought to be
457 secreted, species-specific and without conserved protein domain (Kim et al., 2016), we propose that the
458 ones identified here constitute a set of candidate effectors interacting with their cognate host *Stb*
459 resistance products, and should be a priority target for functional studies. Unlocking the molecular basis

460 underpinning these interactions and how it could impact host defence would lead to a better exploitation
461 of quantitative resistance, which is thought to be more durable (Kettles & Kanyuka, 2016).

462 **Diversification of pathogenicity genes as an adaptive response to host recognition.** Evolution and
463 diversification of pathogenicity genes enables adaptation to new host cultivars and in response to host
464 population changes, they undergo rapid allele frequency changes (L. G. Barrett et al., 2009). We
465 observed high levels of genetic diversity in our candidate pathogenicity genes when compared to the
466 genome-wide gene content. We found a single case of gene deletion and no significant impact of loss of
467 function. Because of the intact and functional coding sequences for nearly all of the identified genes, we
468 suggest that sequence polymorphism is the primary mechanism behind *Z. tritici* rapid adaptation to host
469 populations, permitting the pathogen to circumvent host resistance while conserving the presumable
470 function of the pathogenicity protein. We next investigated what could contribute to this extensive
471 genetic diversity. Recombination is a central source of diversity in sexual fungal pathogens (Burdon &
472 Silk, 1997). Here, we detected recombination events for a significant number of candidates, which
473 suggests its role in the observed genetic diversity of pathogenicity genes. In *Z. tritici*, it was reported
474 that recombination is a major driver of its evolution and is highly heterogeneous, concentrated in narrow
475 hotspots genomic regions enriched in putative effector genes (Croll et al., 2015; Stukenbrock & Dutheil,
476 2018). TEs are also thought to contribute to gene diversification by creating fast-evolving genomic
477 regions through non-homologous recombination or RIP mutations (Seidl & Thomma, 2017). Here, we
478 found that candidate pathogenicity genes are indeed at closer distance to TEs, suggesting their potential
479 role in sequence diversity. We also observed that TE distance was correlated with recombination events
480 in putative pathogenicity genes, and that the levels of genetic diversity were high for genes with the
481 highest recombination events, regardless of their distance to TE. We hypothesize that recombination is
482 likely the major driver underlying the diversification of pathogenicity genes. A recent study supports
483 this hypothesis where unlike distance to TEs, recombination rate was found to be strongly associated
484 with adaptive substitutions in *Z. tritici* genomes (Grandaubert et al., 2019).

485 A key determinant of genetic diversity at pathogenicity genes is selection pressure. Understanding the
486 evolution of pathogenicity in fungal populations with respect to the selection pressure imposed by host
487 populations is necessary for the sustainable management of *R* genes in the field. This can be achieved
488 by searching for signatures of positive diversifying selection with the risk of recovering some false
489 positive signals. In this study, we used conservative maximum likelihood models and models that have
490 been shown to be robust to recombination and significantly reduce false positives (Anisimova et al.,
491 2003). We found that more than two thirds of candidate genes harbour signatures of positive selection
492 acting on specific codons, suggesting that the diversifying selection is likely the evolutionary force
493 maintaining non-synonymous substitutions relevant for *Z. tritici*-wheat quantitative interactions. For
494 instance, the vast majority of different mutations between the avirulent and virulent versions of
495 *Zt_3_00467* and *Zt_6_00682* have been identified as positively-selected. Such selection pressure could

496 be the result of host *R* gene recognition, and may have greatly contributed to the accumulation of
497 beneficial mutations. This agrees with a previous study, which demonstrated that rapidly evolving genes
498 play a prominent role in host-specific disease development (Poppe et al., 2015). Together, these results
499 suggest that host–pathogen coevolution drives the emergence and maintenance of allelic diversity at loci
500 likely involved in direct (or indirect) interactions with host *R* genes. Future surveys of *R* genes diversity
501 and structure in wheat populations will provide a deeper insight into reciprocal patterns of selection,
502 coevolution, and the maintenance of diversity in resistance and pathogenicity in natural populations.

503 **Material and Methods**

504 **Fungal material.** We used a subset of 103 *Z. tritici* isolates in this study deriving predominantly from
505 a large collection of French isolates collected in 2009 and 2010. This genetically diverse subset was
506 mainly isolated from the French cultivars 'Apache' (n=44) and 'Premio' (n=42). The remaining were
507 isolated from cultivars 'Soissons' (n=9), 'Caphorn' (n=3), 'Alixan' (n=1), 'Bermude' (n=1), 'Dinosor'
508 (n=1), and 'Garcia' (n=1). One isolate was collected in the United Kingdom in 2009 on cultivar
509 'Humbert'. All strains are monospore non-clonal isolates that have been originally isolated either at the
510 Plant Research International in the Netherlands or at INRAE UR BIOGER in France. For the functional
511 validation of candidate pathogenicity genes, we used the two strains 3D1 (*i.e.* ST99CH3D1) and 3D7
512 (*i.e.* ST99CH3D7) collected in 1999 from a Swiss wheat field and previously used for the functional
513 validation of *Avr3D1* (Meile *et al.*, 2018).
514

515 **Plant material.** The pathogenicity of the 103 isolates was initially evaluated on 29 wheat cultivars,
516 including 15 carrying known *Stb* resistance genes. After the first replicate, only cultivars on which we
517 observed phenotypic variability were kept for a second and a third replicates. These were the cultivars
518 'Bulgaria-88' (*Stb1* and *Stb6*), 'Israel-493' (*Stb3* and *Stb6*), 'Tadinia' (*Stb4* and *Stb6*), 'Shafir' and
519 'Cadenza' (*Stb6*), 'Estanzuela Federal' (*Stb7*), 'Courtot' (*Stb9*), 'Apache' (*Stb4/5* and *Stb11*),
520 'Salamouni' (*Stb13*, *Stb14* and *StbSm3*) and 'Arina' (*Stb6* and *Stb15*); the resistance background of
521 cultivars 'Premio' and 'Toisondor' was not known. Details on the differential lines are available in Table
522 S3. This genetic material is available from the INRAE Small grain cereals Biological Resources Centre
523 in Clermont-Ferrand (<https://www6.clermont.inrae.fr/umr1095_eng/Organisation>).
524

525 **Quantitative pathogenicity assays.** For inoculum preparation, blastospores of *Z. tritici* stored at -80°C
526 were used to inoculate a liquid growth medium (Difco™ YPD Broth Cat. No. 242820) agitated at 140
527 rpm in a 17°C climate cell for 144h and transferred to agar growth medium (Difco™ YPD Agar Cat.
528 No. 242720) for 96h. Fresh blastospores were recovered with loops, resuspended in sterile water, and
529 the inoculum was adjusted to a concentration of 10^6 spores.L^{-1}, to which one drop of Tween®20 (Sigma
530 Aldrich® Cat. No. 9005-64-5) was added per 15 mL before inoculation. Five seeds from each cultivar
531 were sown in 90*90*95 mm pots containing Floradur B soil (80% blond peat / 20% black peat, with 1.2
532 kg.m^{-3} fertilizer 18/10/20). One day before inoculation, we kept three homogeneous seedlings on which

533 a 75 mm length portion was delimited on the first true leaf using a black marker. Sixteen days old
534 seedlings were inoculated by applying six times the inoculum on the marked leaves with a flat tipped
535 paintbrush (synthetic bristles, 12mm width), depositing approximately 30,000 spores.cm^{-2} on the leaf
536 surface. After inoculation, plants were watered and enclosed in transparent polyethylene high-density
537 bags for 72h to initiate the infection. Ten days post-inoculation, leaves were cut keeping only the marked
538 leaf to ensure light homogeneity and longevity of the inoculated leaf. From sowing to disease scoring,
539 plants were kept in growth chambers regulated at 80% relative humidity, 16h of light per 24h at a photon
540 flux of 300 $\mu mol.m^{-2}.s^{-1}$, a 22°C temperature under light and 18°C in the obscurity. Quantitative
541 pathogenicity was assessed as percentages of leaf area covered by pycnidia (PLACP) and by necrosis
542 (PLACN) at 21 days post-inoculation, averaged over three leaves and three replicates (nine leaves per
543 interaction). Phenotypic characterization of the generated deletion and ectopic transformants was done
544 following the exact same procedure described above in two replicates.
545
546 **Whole-genome resequencing.** Blastospores were harvested after seven days growth in liquid cultures
547 on yeast extract-peptone-dextrose Broth (Difco™ YPD Broth), and then freezed-dried for 24h. High
548 molecular weight DNA was extracted using a phenol-chloroform based procedure (Fagudens et al.,
549 2020). Illumina paired-end libraries were prepared for each sample and DNA fragments were re-
550 sequenced with 2x100 bp reads on a HiSeq-2000 sequencing system at Genewiz, Inc. (formerly
551 Beckman Coulter Genomics). Depth of sequence coverage was very variable depending on the isolate,
552 ranging from 16 to 303 genome equivalents with an average of 61 genome equivalents.
553
554 **SNP calling and quality filtering.** For each isolate, reads were trimmed using trimmomatic v0.32
555 (Bolger et al., 2014) and then mapped to the reference genome IPO323 (Goodwin et al., 2011) using the
556 mem algorithm from BWA v0.7.7 with default settings (Li & Durbin, 2009). Samtools v0.1.19 and
557 Picard tools v1.106 were used to filter reads (Li et al., 2009); <http://broadinstitute.github.io/picard/>;
558 to remove optical duplicates, secondary alignments, reads with a mapping quality below 30 and to keep
559 only pairs in which both reads met quality checks. SNP calling was done with Freebayes v0.9 using
560 the option --report-monomorphic (Garrison & Marth, 2012) and all positions detected as low complexity
561 regions or TEs were excluded. Low complexity regions and TEs were detected using RepeatMasker
562 with default settings (RepeatMasker Open-4.0 <http://www.repeatmasker.org>) and REPET (Flutre et
563 al., 2011), respectively. Quality filters were applied to VCF files using in-house scripts. Depth of
564 coverage (DP) calculated at each position followed a normal distribution around the depth main peak.
565 Hence, a DP filter was set at twice the standard deviation of the mean. All VCFs were stored as a matrix
566 reduced to keep only positions where at least one isolate carries an alternative allele at a non-filtered
567 position. The SNP matrix was parsed as a .tped format, converted to .ped/.map. The final matrix was
568 filtered for less than 50% missing data and more than 10% Minor Allele Frequency (MAF) using PLINK
569 v.1.9 (Chang et al., 2015) <https://cog-genomics.org/plink/>.

570 **Linkage disequilibrium (LD).** To assess the mapping power and resolution of GWAS, we estimated
571 genome-wide LD decay using PopLDdecay (C. Zhang et al., 2019). The average squared Pearson's
572 correlation coefficient (r^2) was computed between pairs of SNPs up to 20 kb, and plotted over physical
573 distance by fitting a nonlinear regression model.

574 **Population structure analyses.** Two approaches were used to investigate the fungal population
575 structure: (1) we performed a principal component analysis (PCA) using independent SNPs ($r^2 < 0.2$)
576 with the --pca command PLINK v.1.9 and (2) the admixture model in STRUCTURE v.2.2 software
577 (Pritchard et al., 2000) with a 10^5 burn-in period, 5×10^5 MCMC steps and 5 iterations for each k value
578 from 1 to 10. The optimal number of sub-populations was determined with the Evanno method (Evanno
579 et al., 2005).

580 **GWAS.** For each quantitative trait, we conducted GWAS using single- and multi-locus linear mixed
581 models based on the EMMA algorithm implemented in GAPIT v.3.0 (Kang et al., 2008; Lipka et al.,
582 2012; Segura et al., 2012). We used a VanRaden derived kinship matrix as a random effect, and set the
583 PCA option to 'true' to determine the optimum number of PCs to be included in the model, following
584 the Bayesian Information Criterion (BIC). The most robust model was chosen by visualizing Quantile-
585 Quantile plots that illustrate the observed vs expected *P*-values. A uniform threshold of $P=1/n$ (n being
586 the effective number of independent SNPs) was used as the genome-wide significance threshold. The
587 effective number of independent SNPs was calculated using the Genetic type 1 Error Calculator;
588 <http://grass.cgs.hku.hk/gec/download.php>. The suggestive genome-wide threshold for the 718,810
589 SNPs was 3.96×10^{-6} and the significant *P*-value was 1.98×10^{-7}. We also adjusted SNP *P*-values to
590 account for multiple testing correction using the false discovery rate (FDR) procedure. SNPs with an
591 FDR-adjusted *P*-value < 0.10 were deemed significant. For LD analysis, we first calculated the pairwise
592 LD with the lead SNP in a window size of 1 Mb using the --r^2 command in PLINK v.1.9. We set a cut-
593 off of $r^2 > 0.5$ to determine candidate genomic regions and assigned predicted genes using a custom Perl
594 script. In case more than one gene is assigned to the candidate region, we picked the closest one from
595 the lead SNP having pathogenicity-like properties (*e.g.* secretion signal, differential expression *in*
596 *planta*). If no gene fell in the LD region, we selected the closest gene to the lead SNP. LD heatmaps
597 were constructed using Haploview v4.2 (J. C. Barrett et al., 2005).

598 **Functional annotation.** We inferred gene models from the *ab initio Z. tritici* annotation (Grandaubert
599 et al., 2015) in the candidate genomic regions identified by GWAS. We used RNAseq datasets of the
600 reference isolate IPO323 (Rudd et al., 2015) to check for missing genes or manually refined gene models
601 (*e.g.* splice junctions, missing exons). N-terminal secretion signals were predicted by using the Hidden
602 Markov Model (HMM) scoring method implemented in SignalP 3.0 (Dyrløv Bendtsen et al., 2004).
603 Transmembrane domains were identified with TMHMM 2.0 (Krogh et al., 2001). Effector probabilities
604 were predicted through EffectorP 3.0 (Sperschneider & Dodds, 2021). Conserved protein domains were
605 assigned using HMM scan against the PFAM 35.0 database and blastp was used to search for homologs

606 from other fungi. Finally, we used PHIB-BLAST, a blast server for PHI-base (Pathogen-Host
607 Interactions), to search for homologs against functionally characterized pathogenicity factors.
608
609 **Gene expression analysis.** RNAseq data from *in vitro* and *in planta* (1-, 4-, 9-, 14- and 21-days post
610 infection; dpi) were obtained from a previous study (Rudd et al., 2015). We computed gene expression
611 quantified as fragments per kilobase of exon per million fragments mapped (FPKM) using the improved
612 gene models with Cuffdiff v.2.2.1 (Trapnell et al., 2012). We inferred pairwise log2-fold expression
613 changes and *P*-value adjustment for *in planta* RNAseq samples compared to the Czapek-Dox broth
614 (CDB) culture medium. Genes with an FDR-adjusted *P*-value < 0.10 and a log2|fold change|> 2 were
615 considered as differentially expressed during the infection.
616
617 ***De novo* assembly and sequence diversity.** We used SPAdes v.3.14.1 to produce draft assemblies for
618 each isolate (Bankevich et al., 2012). We ran the tool with the following settings: -k 21,33,55,75,95 --
619 careful to reduce mismatches and indel errors. We performed blastn (Camacho et al., 2009) searches
620 using ~1000 bp of the target locus as a query to retrieve matching contig segments in each assembly.
621 High-confidence matches were first extracted from the contig set using samtools v1.10 and subsequently
622 aligned with MAFFT v7.464 using the --auto option and 1,000 iterative refinement cycles (Katoh &
623 Standley, 2013). Alignments were inspected using JalView (Waterhouse et al., 2009) and ambiguous
624 regions were removed with trimAL (Capella-Gutiérrez et al., 2009) with less stringent parameters. We
625 used the batch mode of DNAsp v.6 (Rozas et al., 2017) to calculate summary statistics of population
626 genetic parameters associated with candidate genes such as nucleotide diversities (π, (Nei, n.d.)) and
627 Tajima's D (Tajima, 1989). TE locations for *Z. tritici* were retrieved from a previous study (Grandaubert
628 et al., 2015) and we parsed gene distances to TE using the command 'closest' from bedtools (Quinlan
629 & Hall, 2010).
630
631 **Evolutionary analysis.** Before testing for positive diversifying selection, we checked for evidence of
632 recombination, as this can cause false positives. We used the software packages RDP4 (Martin et al.,
633 2015) and GARD (Kosakovsky Pond et al., 2006) for recombination detection. We considered genes
634 with recombination events if both detection methods found evidence of recombination. We then
635 constructed non-recombinant phylogenetic trees of gene coding sequences using the RAxML algorithm
636 (Stamatakis, 2014) implemented in RDP4. Statistical support for the nodes was obtained after 100
637 bootstrap replicates. We used codeml of the PAML package (Yang & Bielawski, 2000) to estimate the
638 evolutionary rates of candidate genes, based on Markov models of codon evolution. Branch lengths of
639 the inferred ML trees were estimated by using the model M0 and then fixed in further analyses. The
640 model M0 assumes constant selective pressure across codon sites and over time. The selective pressure
641 at the coding-sequence level was measured by ω, the ratio of non-synonymous to synonymous rates
642 (d_N/d_S). A d_N/d_S ratio of 1 ($\omega = 1$) indicates neutrality, whereas $\omega < 1$ and $\omega > 1$ suggests purifying and

643 diversifying selection, respectively. Models M1a, M2a, M7 and M8 of variable selective pressure across
644 codon sites were used to estimate selective pressure and test for positive selection (Nielsen & Yang,
645 1998; Yang et al., 2005). For each candidate gene, we performed pairwise likelihood ratio tests (LRTs)
646 for positive selection, comparing models that allow sites with $\omega > 1$ (M2a and M8) with models that do
647 not (M1a and M7). We considered genes under positive diversifying selection if both model
648 comparisons in LRT were significant. Finally, positively selected sites were inferred with a Bayes
649 Empirical Bayes (BEB) probability (Yang et al., 2005) and codons with $\omega > 1$ at a probability > 0.95
650 were considered positively-selected.
651
652 **Functional validation of candidate pathogenicity genes.** For *Zt_6_00682*, we generated knockout
653 mutants in the 3D7 isolate and ectopic transformants expressing the virulent and avirulent alleles in the
654 know-out background. To create plasmid constructs for gene disruption, flanking regions (~1kb before
655 start and stop codons) of *Zt_6_00682* were amplified from gDNA using primers listed in Table S9. The
656 hygromycin resistance gene cassette was amplified from the plasmid pNOV2114_HygR_attP1_attP2
657 and cloned into a pNOV2214 vector. The vector was then digested with *Hind*III restriction enzyme (New
658 England Biolabs), amplified and used for the cloning of PCR fragments with a Gibson Assembly cloning
659 kit (New England Biolabs) following the manufacturer's instructions. Constructs were transformed into
660 *Escherichia coli* NEB 5-α using a heat shock transformation, screened by PCR and verified by Sanger
661 sequencing using primers listed in Table S9. Confirmed plasmids were transformed into *Agrobacterium*
662 *tumefaciens* cells by heat shock and validated by PCR. The vector was used to transform *Z. tritici* via *A.*
663 *tumefaciens*-mediated transformation (ATMT) following a procedure adapted from Zwiers & de Waard
664 (2001). Transformed *Z. tritici* colonies growing in the presence of hygromycin were tested by PCR to
665 confirm the successful deletion of the gene. To validate the allele dependent effect of *Zt_6_00682*, we
666 generated and validated ectopic transformants by placing both the virulent (3D7) and the avirulent (3D1)
667 alleles into the 3D7 KO background, following the exact same procedure described above with the
668 following modifications: (*i*) we used the vector pNOV_3Gate_SUL with a sulfonylurea as a resistance
669 marker, (*ii*) the digestion was achieved with *Hind*III and *Kpn*I restriction enzymes. The resulting vectors,
670 pNOV_sulf_3D7g.7232 and pNOV_sulf_3D1g.7155, we used to transform *Z. tritici* via ATMT
671 following the procedure described above. Primers used to verify gene disruptions and ectopic
672 integrations are listed in Table S9. Finally, three independent transformants carrying the different
673 virulence alleles and three knock-out mutants were used for phenotypic characterization *in planta*. For
674 *Zt_3_00467*, we ectopically expressed the virulent (3D7) and avirulent (IPO10273) alleles in the 3D7
675 background. A fragment containing *Zt_3_00467* gene, including 1800 pb upstream of the start codon
676 and 463 pb downstream of the stop codon was amplified using the Phusion DNA polymerase (NEB)
677 and the primers included in table S9. This fragment was cloned into a pCGEN vector (Motteram et al.,
678 2011) previously digested with *Kpn*I (New England Biolabs), using In-Fusion HD Cloning Kit (Takara
679 Bio) and following the manufacturer's instructions. Vectors were transformed using heat-shocked *E. coli*

680 Stellar cells and verified using Sanger sequencing. Transformation via ATMT was performed as
681 described above and selection for the *Z. tritici* mutants was performed on geneticin (150 µg.mL^{-1})-
682 containing plates.
683

684 **Data availability**

685 Sequence data of the 103 French isolates are deposited at the NCBI Sequence Read Archive under the
686 accession number PRJNA777581.

687 **Acknowledgements**

688 We thank David Gouache (ARVALIS Institut du Végétal, Boigneville, France), Marc-Henri Lebrun,
689 Henriette Goyeau, Frédéric Suffert and Anne-Sophie Walker (INRAE UR BIOGER, Palaiseau, France)
690 for their help in establishing the fungal collection; Robert King and Jason Rudd (Rothamsted Research,
691 Harpenden, United Kingdom) for providing RNAseq data; Gabriel Scalliet (Syngenta Crop Protection
692 AG, Stein, Switzerland) for providing the vectors pNOV2114, pNOV_3Gate_SUL and
693 pNOV_3Gate_HYG used to clone the candidate pathogenicity genes; Martin Willigsecker and Béatrice
694 Beauzoone (INRAE UR BIOGER) for their help in the pathology assays; and Thierry Rouxel (INRAE
695 UR BIOGER) for critical reading of the manuscript and valuable feedback.

696 **Financial Disclosure**

697 This work was supported by the French National Research Agency (program BIOADAPT, ANR-12-
698 ADAP-0009-04), the FSOV 2008 B project and BASF France SAS for R.A. postdoctoral fellowship.
699 T.C.M. was supported by a Short-Term Scientific Mission granted by the COST SUSTAIN action
700 FA1208. A.S.V. was supported by the Ministerio de Ciencia e Innovación (Ramón y Cajal program,
701 grant no. RYC2018-025530-I). INRAE BIOGER benefits from the support of Saclay Plant Sciences-
702 SPS (ANR-17-EUR-0007).

703 **Author contributions**

704 T.C.M and R.A. conceived and designed the study. G.H.J.K. and H.G. isolated and provided most of the
705 natural *Z. tritici* isolates constituting the GWAS panel. A.D., T.C.M. and J.C. carried out virulence
706 assays on the natural isolates, and R.A. on the transformant isolates. A.D. and T.C.M. prepared the DNA
707 samples for whole genome resequencing. J.C., N.L. and J.A. performed the bioinformatics analyses,
708 including preparation of the SNP matrix and *de novo* genome assemblies. D.C. provided *Z. tritici*
709 genomic data. R.A., L.D., A.P. and A.S.V. performed the cloning and genetic transformations. R.A.
710 performed the genome wide association, gene expression and evolutionary sequence analyses. R.A. and
711 T.C.M. wrote the manuscript. All authors have reviewed the manuscript.

712 **Competing Interests**

713 The authors declare to have no competing interests.

714

715 **References**

716 Amezrou, R., Audeon, C., Compain, J., Gélisse, S., Ducasse, A., Saintenac, C., Lapalu, N., Orford, S.,
717 Croll, D., Amselem, J., Fillinger, S., Marcel, T.C. (2022). A secreted protease-like protein in
718 *Zymoseptoria tritici* is responsible for avirulence on *Stb9* resistance gene in wheat. *bioRxiv*,
719 2022.10.31.514577. https://doi.org/10.1101/2022.10.31.514577

720 Anisimova, M., Nielsen, R., & Yang, Z. (2003). Effect of Recombination on the Accuracy of the
721 Likelihood Method for Detecting Positive Selection at Amino Acid Sites. *Genetics*, *164*(3),
722 1229–1236. https://doi.org/10.1093/genetics/164.3.1229

723 Arnér, E. S. J., & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin
724 reductase. *European Journal of Biochemistry*, *267*(20), 6102–6109.
725 https://doi.org/10.1046/j.1432-1327.2000.01701.x

726 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M.,
727 Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. v, Sirotkin, A. v, Vyahhi, N., Tesler, G.,
728 Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and
729 Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
730 https://doi.org/10.1089/cmb.2012.0021

731 Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD
732 and haplotype maps. *Bioinformatics*, *21*(2), 263–265.
733 https://doi.org/10.1093/bioinformatics/bth457

734 Barrett, L. G., Thrall, P. H., Dodds, P. N., van der Merwe, M., Linde, C. C., Lawrence, G. J., &
735 Burdon, J. J. (2009). Diversity and Evolution of Effector Loci in Natural Populations of the Plant
736 Pathogen Melampsora lini. *Molecular Biology and Evolution*, *26*(11), 2499–2513.
737 https://doi.org/10.1093/molbev/msp166

738 Bartoli, C., & Roux, F. (2017). Genome-Wide Association Studies In Plant Pathosystems: Toward an
739 Ecological Genomics Approach. *Frontiers in Plant Science*, *8*, 763.
740 https://www.frontiersin.org/article/10.3389/fpls.2017.00763

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

744 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
745 sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

746 Bourras, S., McNally, K. E., Ben-David, R., Parlange, F., Roffler, S., Praz, C. R., Oberhaensli, S.,
747 Menardo, F., Stirnweis, D., Frenkel, Z., Schaefer, L. K., Flückiger, S., Treier, G., Herren, G.,
748 Korol, A. B., Wicker, T., & Keller, B. (2015). Multiple Avirulence Loci and Allele-Specific
749 Effector Recognition Control the *Pm3* Race-Specific Resistance of Wheat to Powdery Mildew.
750 *The Plant Cell*, *27*(10), 2991 LP – 3012. https://doi.org/10.1105/tpc.15.00171

751 Brading, P. A., Verstappen, E. C. P., Kema, G. H. J., & Brown, J. K. M. (2002). A Gene-for-Gene
752 Relationship Between Wheat and Mycosphaerella graminicola, the Septoria Tritici Blotch
753 Pathogen. *Phytopathology®*, *92*(4), 439–445. https://doi.org/10.1094/PHYTO.2002.92.4.439

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

757 Burdon, J. J., & Silk, J. (1997). Sources and Patterns of Diversity in Plant-Pathogenic Fungi.
758 *Phytopathology®*, *87*(7), 664–669. https://doi.org/10.1094/PHYTO.1997.87.7.664

759 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L.
760 (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.
761 https://doi.org/10.1186/1471-2105-10-421

762 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated
763 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.
764 https://doi.org/10.1093/bioinformatics/btp348

765 Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015).
766 Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*,
767 *4*(1). https://doi.org/10.1186/s13742-015-0047-8

768 Collemare, J., O'Connell, R., & Lebrun, M.-H. (2019). Nonproteinaceous effectors: the terra incognita
769 of plant–fungal interactions. *New Phytologist*, *223*(2), 590–596.
770 https://doi.org/10.1111/nph.15785

771 Croll, D., Lendenmann, M. H., Stewart, E., & McDonald, B. A. (2015). The Impact of Recombination
772 Hotspots on Genome Evolution of a Fungal Plant Pathogen. *Genetics*, *201*(3), 1213–1228.
773 https://doi.org/10.1534/genetics.115.180968

774 Domazakis, E., Lin, X., Aguilera-Galvez, C., Wouters, D., Bijsterbosch, G., Wolters, P. J., &
775 Vleeshouwers, V. G. A. A. (2017). Effectoromics-Based Identification of Cell Surface Receptors
776 in Potato. *In Methods in molecular biology*, 1578, 337–353. https://doi.org/10.1007/978-1-4939-
777 6859-6_29

778 Duncan, K. E., & Howard, R. J. (2000). Cytological analysis of wheat infection by the leaf blotch
779 pathogen *Mycosphaerella graminicola*. *Mycological Research*, *104*(9), 1074–1082.
780 https://doi.org/10.1017/S0953756299002294

781 Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved Prediction of Signal
782 Peptides: SignalP 3.0. *Journal of Molecular Biology*, *340*(4), 783–795.
783 https://doi.org/10.1016/j.jmb.2004.05.028

784 Esquerré-Tugayé, M.-T., Boudart, G., & Dumas, B. (2000). Cell wall degrading enzymes, inhibitory
785 proteins, and oligosaccharides participate in the molecular dialogue between plants and
786 pathogens. *Plant Physiology and Biochemistry*, *38*(1), 157–163. https://doi.org/10.1016/S0981-
787 9428(00)00161-3

788 Evanno, G., Regnaut, S., & Gouget, J. (2005). Detecting the number of clusters of individuals using
789 the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.
790 https://doi.org/10.1111/j.1365-294X.2005.02553.x

791 Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal
792 wheat pathogen Zymoseptoria tritici: A laboratory workflow. *Current Protocols in*
793 *Microbiology*, 59, e128. https://doi:10.1002/cpmc.128

Fan, X., He, F., Ding, M., Geng, C., Chen, L., Zou, S., Liang, Y., Yu, J., & Dong, H. (2019). Thioredoxin Reductase Is Involved in Development and Pathogenicity in *Fusarium graminearum*. *Frontiers in Microbiology*, 10, 393. https://www.frontiersin.org/article/10.3389/fmicb.2019.00393

Faris, J. D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., Fellers, J. P., Meinhardt, S. W., Rasmussen, J. B., Xu, S. S., Oliver, R. P., Simons, K. J., & Friesen, T. L. (2010). A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences*, *107*(30), 13544 LP – 13549. https://doi.org/10.1073/pnas.1004090107

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering Transposable Element Diversification in De Novo Annotation Approaches. *PLOS ONE*, *6*(1), e16526. https://doi.org/10.1371/journal.pone.0016526

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. 1–9. http://arxiv.org/abs/1207.3907

Goodwin, S. B., ben M'Barek, S., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., Foster, A. J., van der Lee, T. A. J., Grimwood, J., Aerts, A., Antoniw, J., Bailey, A., Bluhm, B., Bowler, J., Bristow, J., van der Burgt, A., Canto-Canché, B., Churchill, A. C. L., Conde-Ferràez, L., … Kema, G. H. J. (2011). Finished Genome of the Fungal Wheat Pathogen *Mycosphaerella graminicola* Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. *PLOS Genetics*, *7*(6), e1002070. https://doi.org/10.1371/journal.pgen.1002070

Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. (2015). RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus *Zymoseptoria* Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. *G3 Genes|Genomes|Genetics*, *5*(7), 1323–1333. https://doi.org/10.1534/g3.115.017731

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

Gu, Q., Chen, Y., Liu, Y., Zhang, C. and Ma, Z. (2015). The transmembrane protein FgSho1 regulates fungal development and pathogenicity via the MAPK module Ste50-Ste11-Ste7 in *Fusarium graminearum*. *New Phytologist*, *206*(1), 315–328. https://doi.org/10.1111/nph.13158

Habig, M., Quade, J., & Stukenbrock, E. H. (2017). Forward Genetics Approach Reveals Host Genotype-Dependent Importance of Accessory Chromosomes in the Fungal Wheat Pathogen *Zymoseptoria tritici*. *MBio*, *8*(6), e01919–17. https://doi.org/10.1128/mBio.01919-17

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME Journal*, *11*(5), 1189–1204. https://doi.org/10.1038/ismej.2016.196

Jain, S., & Keller, N. (2013). Insights to fungal biology through LaeA sleuthing. *Fungal Biology Reviews*, *27*(2), 51–59. https://doi.org/10.1016/j.fbr.2013.05.004

Juchimiuk, M., Kruszewska, J., & Palamarczyk, G. (2015). Dolichol phosphate mannose synthase from the pathogenic yeast *Candida albicans* is a multimeric enzyme. *Biochimica et Biophysica*

Acta (BBA) - General Subjects, *1850*(11), 2265–2275. https://doi.org/10.1016/j.bbagen.2015.08.012

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics*, *178*(3), 1709 LP – 1723. https://doi.org/10.1534/genetics.107.080101

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kema, G. H. J., Mirzadi Gohari, A., Aouini, L., Gibriel, H. A. Y., Ware, S. B., van den Bosch, F., Manning-Smith, R., Alonso-Chavez, V., Helps, J., ben M'Barek, S., Mehrabi, R., Diaz-Trujillo, C., Zamani, E., Schouten, H. J., van der Lee, T. A. J., Waalwijk, C., de Waard, M. A., de Wit, P. J. G. M., Verstappen, E. C. P., … Seidl, M. F. (2018). Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin resistance. *Nature Genetics*, *50*(3), 375–380. https://doi.org/10.1038/s41588-018-0052-9

Kema, G. H. J., Sayoud, R., Annone, J. G., & van Silfhout, C. H. (1996). Genetic variation for virulence and resistance in the wheat-*Mycosphaerella graminicola* pathosystem II. Analysis of interactions between pathogen isolates and host cultivars. *Phytopathology®*, 86(2), 213–220. https://doi.org/10.1094/Phyto-86-213

Kema, G. H. J., Verstappen, E. C. P., & Waalwijk, C. (2000). Avirulence in the Wheat Septoria tritici Leaf Blotch Fungus *Mycosphaerella graminicola* Is Controlled by a Single Locus. *Molecular Plant-Microbe Interactions®*, *13*(12), 1375–1379. https://doi.org/10.1094/MPMI.2000.13.12.1375

Kettles, G. J., & Kanyuka, K. (2016). Dissecting the Molecular Interactions between Wheat and the Fungal Pathogen *Zymoseptoria tritici*. *Frontiers in Plant Science*, *7*, *508*. https://www.frontiersin.org/article/10.3389/fpls.2016.00508

Kim, K.-T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G., Kang, S., & Lee, Y.-H. (2016). Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. *Frontiers in Plant Science*, *7*, 186. https://www.frontiersin.org/article/10.3389/fpls.2016.00186

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics*, *22*(24), 3096–3098. https://doi.org/10.1093/bioinformatics/btl474

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kubicek, C. P., Starr, T. L., & Glass, N. L. (2014). Plant Cell Wall–Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. *Annual Review of Phytopathology*, *52*(1), 427–451. https://doi.org/10.1146/annurev-phyto-102313-045831

874 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform.
875 *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

876 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin,
877 R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map
878 format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.
879 https://doi.org/10.1093/bioinformatics/btp352

880 Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., &
881 Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics*,
882 *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

883 Liu, W., Liu, J., Ning, Y., Ding, B., Wang, X., Wang, Z., & Wang, G.-L. (2013). Recent Progress in
884 Understanding PAMP- and Effector-Triggered Immunity against the Rice Blast Fungus
885 *Magnaporthe oryzae*. *Molecular Plant*, *6*(3), 605–620. https://doi.org/10.1093/mp/sst015

886 lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S.,
887 & Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant*
888 *Biology*, *66*(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

889 Ma, K.-W., & Ma, W. (2016). Phytohormone pathways as targets of pathogens to facilitate infection.
890 *Plant Molecular Biology*, *91*(6), 713–725. https://doi.org/10.1007/s11103-016-0452-0

891 Marina, F., Abbas, M., J, J.-D. M., G, P. H., Sophien, K., & J, B. M. (2022). Effectors of Filamentous
892 Plant Pathogens: Commonalities amid Diversity. *Microbiology and Molecular Biology Reviews*,
893 *81*(2), e00066-16. https://doi.org/10.1128/MMBR.00066-16

894 Martienssen, R. A., & Vincent, C. (2001). DNA Methylation and Epigenetic Inheritance in Plants and
895 Filamentous Fungi. *Science*, *293*(5532), 1070–1074.
896 https://doi.org/10.1126/science.293.5532.1070

897 Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and
898 analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1).
899 https://doi.org/10.1093/ve/vev003

900 Meile, L., Croll, D., Brunner, P. C., Plissonneau, C., Hartmann, F. E., McDonald, B. A., & Sánchez-
901 Vallet, A. (2018). A fungal avirulence factor encoded in a highly plastic genomic region triggers
902 partial resistance to septoria tritici blotch. *New Phytologist*, *219*(3), 1048–1061.
903 https://doi.org/10.1111/nph.15180

904 Mirzadi Gohari, A., Ware, S. B., Wittenberg, A. H. J., Mehrabi, R., ben M'Barek, S., Verstappen, E.
905 C. P., van der Lee, T. A. J., Robert, O., Schouten, H. J., de Wit, P. P. J. G. M., & Kema, G. H. J.
906 (2015). Effector discovery in the fungal wheat pathogen *Zymoseptoria tritici*. *Molecular Plant*
907 *Pathology*, *16*(9), 931–945. https://doi.org/10.1111/mpp.12251

908 Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant
909 pathogens. *Nature Reviews Microbiology*, *15*(12), 756–771.
910 https://doi.org/10.1038/nrmicro.2017.76

911 Motteram, J., Lovegrove, A., Pirie, E., Marsh, J., Devonshire, J., van de Meene, A., Hammond-
912 Kosack, K. and Rudd, J.J. (2011), Aberrant protein N-glycosylation impacts upon infection-

913 related growth transitions of the haploid plant-pathogenic fungus *Mycosphaerella graminicola*.
914 *Molecular Microbiology*, *81*(2), 415–433. https://doi.org/10.1111/j.1365-2958.2011.07701.x

915 Nei, M. (n.d.). *Molecular Evolutionary Genetics*. Columbia University Press.
916 https://doi.org/10.7312/nei-92038

917 Nielsen, R., & Yang, Z. (1998). Likelihood Models for Detecting Positively Selected Amino Acid
918 Sites and Applications to the HIV-1 Envelope Gene. *Genetics*, *148*(3), 929 LP – 936.
919 http://www.genetics.org/content/148/3/929.abstract

920 O'Driscoll, A., Kildea, S., Doohan, F., Spink, J., & Mullins, E. (2014). The wheat Septoria conflict: a
921 new front opening up? *Trends in Plant Science*, *19*(9), 602–610.
922 https://doi.org/10.1016/j.tplants.2014.04.011

923 Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly Evolving Genes Are Key
924 Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici*
925 (*Mycosphaerella graminicola*). *PLOS Pathogens*, *11*(7), e1005055.
926 https://doi.org/10.1371/journal.ppat.1005055

927 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
928 Multilocus Genotype Data. *Genetics*, *155*(2), 945 LP – 959.

929 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
930 features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

931 Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can
932 be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

933 Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez,
934 V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A.,
935 Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., … Howlett, B. J.
936 (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome
937 affected by Repeat-Induced Point mutations. *Nature Communications*, *2*(1), 202.
938 https://doi.org/10.1038/ncomms1189

939 Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.
940 E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large
941 Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.
942 https://doi.org/10.1093/molbev/msx248

943 Rudd, J. J., Kanyuka, K., Hassani-Pak, K., Derbyshire, M., Andongabo, A., Devonshire, J., Lysenko,
944 A., Saqi, M., Desai, N. M., Powers, S. J., Hooper, J., Ambroso, L., Bharti, A., Farmer, A.,
945 Hammond-Kosack, K. E., Dietrich, R. A., & Courbot, M. (2015). Transcriptome and Metabolite
946 Profiling of the Infection Cycle of *Zymoseptoria tritici* on Wheat Reveals a Biphasic Interaction
947 with Plant Immunity Involving Differential Pathogen Chromosomal Contributions and a
948 Variation on the Hemibiobiotrophic lifestyle definition. *Plant Physiology*, *167*(3), 1158 LP –
949 1185. https://doi.org/10.1104/pp.114.255927

950 Saintenac, C., Cambon, F., Aouini, L., Verstappen, E., Ghaffary, S. M. T., Poucet, T., Marande, W.,
951 Berges, H., Xu, S., Jaouannet, M., Favery, B., Alassimone, J., Sánchez-Vallet, A., Faris, J.,
952 Kema, G., Robert, O., & Langin, T. (2021). A wheat cysteine-rich receptor-like kinase confers

953 broad-spectrum resistance against Septoria tritici blotch. *Nature Communications*, *12*(1), 433.
954 https://doi.org/10.1038/s41467-020-20685-0

955 Saintenac, C., Lee, W.-S., Cambon, F., Rudd, J. J., King, R. C., Marande, W., Powers, S. J., Bergès,
956 H., Phillips, A. L., Uauy, C., Hammond-Kosack, K. E., Langin, T., & Kanyuka, K. (2018).
957 Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen
958 *Zymoseptoria tritici*. *Nature Genetics*, *50*(3), 368–374. https://doi.org/10.1038/s41588-018-0051-
959 x

960 Sánchez-Vallet, A., Hartmann, F. E., Marcel, T. C., & Croll, D. (2018). Nature's genetic screens:
961 using genome-wide association studies for effector discovery. *Molecular Plant Pathology*, *19*(1),
962 3–6. https://doi.org/10.1111/mpp.12592

963 Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The
964 global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, *3*(3),
965 430–439. https://doi.org/10.1038/s41559-018-0793-y

966 Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An
967 efficient multi-locus mixed-model approach for genome-wide association studies in structured
968 populations. *Nature Genetics*, *44*(7), 825–830. https://doi.org/10.1038/ng.2314

969 Seidl, M. F., & Thomma, B. P. H. J. (2017). Transposable Elements Direct The Coevolution between
970 Plants and Microbes. *Trends in Genetics*, *33*(11), 842–851.
971 https://doi.org/10.1016/j.tig.2017.07.003

972 Sperschneider, J., & Dodds, P. N. (2021). EffectorP 3.0: Prediction of Apoplastic and
973 Cytoplasmic Effectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interactions®*, *35*(2),
974 146–156. https://doi.org/10.1094/MPMI-08-21-0201-R

975 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
976 phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

977 Stewart, E. l., Croll, D., Lendenmann, M. H., Sanchez-Vallet, A., Hartmann, F. E., Palma-Guerrero, J.,
978 Ma, X., & McDonald, B. A. (2018). Quantitative trait locus mapping reveals complex genetic
979 architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. *Molecular*
980 *Plant Pathology*, *19*(1), 201–216. https://doi.org/10.1111/mpp.12515

981 Stukenbrock, E. H., & Dutheil, J. Y. (2018). Fine-Scale Recombination Maps of Fungal Plant
982 Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots. *Genetics*,
983 *208*(3), 1209 LP – 1229. https://doi.org/10.1534/genetics.117.300502

984 Stukenbrock, E. H., & McDonald, B. A. (2009). Population Genetics of Fungal and Oomycete
985 Effectors Involved in Gene-for-Gene Interactions. *Molecular Plant-Microbe Interactions®*,
986 *22*(4), 371–380. https://doi.org/10.1094/MPMI-22-4-0371

987 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA
988 polymorphism. *Genetics*, *123*(3), 585 LP – 595.

989 Toruño, T. Y., Stergiopoulos, I., & Coaker, G. (2016). Plant-Pathogen Effectors: Cellular Probes
990 Interfering with Plant Defenses in Spatial and Temporal Manners. *Annual Review of*
991 *Phytopathology*, *54*(1), 419–441. https://doi.org/10.1146/annurev-phyto-080615-100204

992 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L.,
993 Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-
994 seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578.
995 https://doi.org/10.1038/nprot.2012.016

996 van der Hoorn, R. A. L. (2008). Plant Proteases: From Phenotypes to Molecular Mechanisms. *Annual*
997 *Review of Plant Biology*, *59*(1), 191–223.
998 https://doi.org/10.1146/annurev.arplant.59.032607.092835

999 van de Wouw, A. P., & Idnurm, A. (2019). Biotechnological potential of engineering pathogen
1000 effector proteins for use in plant disease management. *Biotechnology Advances*, *37*(6), 107387.
1001 https://doi.org/10.1016/j.biotechadv.2019.04.009

1002 Varden, F. A., de la Concepcion, J. C., Maidment, J. H. R., & Banfield, M. J. (2017). Taking the stage:
1003 effectors in the spotlight. *Current Opinion in Plant Biology*, *38*, 25–33.
1004 https://doi.org/10.1016/j.pbi.2017.04.013

1005 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview
1006 Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9),
1007 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

1008 Weiland, P., & Altegoer, F. (2021). Identification and Characterization of Two Transmembrane
1009 Proteins Required for Virulence of *Ustilago maydis*. *Frontiers in plant science*, *12*, 669835
1010 https://doi.org/10.3389/fpls.2021.669835

1011 Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in*
1012 *Ecology & Evolution*, *15*(12), 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7

1013 Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes Empirical Bayes Inference of Amino Acid
1014 Sites Under Positive Selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118.
1015 https://doi.org/10.1093/molbev/msi097

1016 Zhan, F., Xie, Y., Zhu, W., Sun, D., McDonald, B. A., & Zhan, J. (2016). Linear Correlation Analysis
1017 of *Zymoseptoria tritici* Aggressiveness with In Vitro Growth Rate. *Phytopathology®*, *106*(11),
1018 1255–1261. https://doi.org/10.1094/PHYTO-12-15-0338-R

1019 Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (2019). PopLDdecay: a fast and effective
1020 tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*,
1021 *35*(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875

1022 Zhang, J., & Zheng, Y. G. (2016). SAM/SAH Analogs as Versatile Tools for SAM-Dependent
1023 Methyltransferases. *ACS Chemical Biology*, *11*(3), 583–597.
1024 https://doi.org/10.1021/acschembio.5b00812

1025 Zhang, L.-B., Tang, L., Ying, S.-H., & Feng, M.-G. (2016). Distinct roles of two cytoplasmic
1026 thioredoxin reductases (Trr1/2) in the redox system involving cysteine synthesis and host
1027 infection of Beauveria bassiana. *Applied Microbiology and Biotechnology*, *100*(24), 10363–
1028 10374. https://doi.org/10.1007/s00253-016-7688-0

1029 Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J.,
1030 Compain, J., Lapalu, N., Amselem, J., McDonald, B. A., Croll, D., & Palma-Guerrero, J. (2017).
1031 A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars

carrying the Stb6 resistance gene. *New Phytologist*, *214*(2), 619–631. https://doi.org/10.1111/nph.14434

Zwiers, L.-H., & de Waard, M. A. (2001). Efficient Agrobacterium tumefaciens-mediated gene disruption in the phytopathogen *Mycosphaerella graminicola*. *Current Genetics*, *39*(5), 388–393. https://doi.org/10.1007/s002940100216