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 GHJ KEMA[3], Daniel CROLL[4], Joëlle AMSELEM[2], Andrea SANCHEZ-VALLET[5], Thierry C MARCEL*[1]

[1] Université Paris-Saclay, INRAE, AgroParisTech, UMR BIOGER, Thiverval-Grignon, France

[2] UR URGI, INRA, Université Paris-Saclay, Versailles, France

[3] Plant Research International B.V., Wageningen, The Netherlands

[4] Université de Neuchâtel, Department of Ecology and Evolution, Neuchâtel, Switzerland

[5] CBGP, INIA, Campus de Montegancedo UPM, Pozuelo de Alarcón, Madrid, Spain

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

Word count (main text): 4,932

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 interactions. 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 performed whole-genome sequencing of 103 isolates and quantified pathogenicity traits on 12 cultivars harbouring different resistance genes. A multi-host genome-wide association studies identified 58 candidate genes associated with pathogenicity, of which nineteen were highly expressed and/or differentially expressed *in planta*. Two of these had large effects and three were shared among more than one cultivar, suggesting that *Z. tritici* pathogenicity is often 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 element 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[1]. Understanding the molecular dialogue between pathogens and their hosts is essential to design durable and effective control strategies. Major molecular factors in the ability of fungal pathogens to cause disease are effectors. Those are proteins delivered into the host apoplast or translocated inside cells to manipulate host physiology or suppress its immunity to favour infection[2]. In turn, plants respond to pathogen invasion by activating defence mechanisms based on recognition of pathogen-associated molecular patterns and effectors by cell surface and intracellular receptors. The discovery of resistance genes (*R*) that encode receptors has provided potential means to control fungal diseases through resistance breeding[3]. However, pathogens have repeatedly overcome host resistance by evading recognition or by suppression of host immunity. These rapid adaptations are mainly driven by genetic variation at pathogenicity genes and genome evolution[4]. Gene diversification, deletions or horizontal acquisitions have generated adaptive variants at loci encoding effectors. Transposable elements (TEs), high mutation and recombination rates are thought to contribute considerably to the extensive genome variation in many fungal species[5-8].

In the case of antagonistic host-pathogen interactions, the fate of genetic variation is determined partly 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[9-12]. Moreover, the diversity of host genotypes in *R*-gene content exerts heterogeneous selection pressure on pathogen populations, giving rise to the co-maintenance of multiple pathogen variants better adapted to specific subsets of hosts[3,13]. Hence, in addition to the identification of effector genes, elucidating the mechanisms contributing to gene diversification provides critical information about their evolvability. This knowledge has tremendous applications for the durability of crop protection strategies[14].

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[15]. 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[16]. For instance, small noncoding RNAs and fungal secondary metabolites play similar effector roles in various plant pathogenic fungi[17]. 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[18], protease inhibitors[19] or disruptors of the hormone-signalling pathway[20].

Effector discovery also led to the systematic search for host genes conferring resistance[15]. 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[14]. 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 causes the Septoria tritici blotch (STB) disease on wheat. *STB* is a major threat to wheat crops in Europe and in temperate regions around the world[21]. *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[22]. This stage is accompanied by the development of pycnidia containing asexual spores that further spread the disease[23,24]. To date, three *Z. tritici* pathogenicity genes that encode avirulence factors have been cloned. These include *AvrStb6* and *AvrStb9* that are recognized by the *R* genes *Stb6* and *Stb9*, respectively, and lead to a strong qualitative resistance response[12,25,26], and *Avr3D1* which triggers quantitative resistance on cultivars harbouring *Stb7* or *Stb12*[10]. In addition to the above-mentioned *R* genes, 18 other race-specific *Stb* resistance genes and numerous quantitative resistance loci have been genetically mapped in wheat[27], but only *Stb6* and *Stb16q* have been cloned to date[28,29]. An ideal approach to identify genetic determinants of complex traits is through Genome-Wide Association Studies (GWAS). GWAS aims to associate sequence polymorphisms to phenotypic variation, and relies on linkage disequilibrium (LD) –the non-random association of alleles at different loci– to identify causal variants. The approach has been used extensively in plants but remains at its infancy on fungal populations[30,31].

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

complex genetic architecture of quantitative pathogenicity and pathogen adaptation to host resistances. Beyond this model, our findings may have broad implications for sustainable disease management.

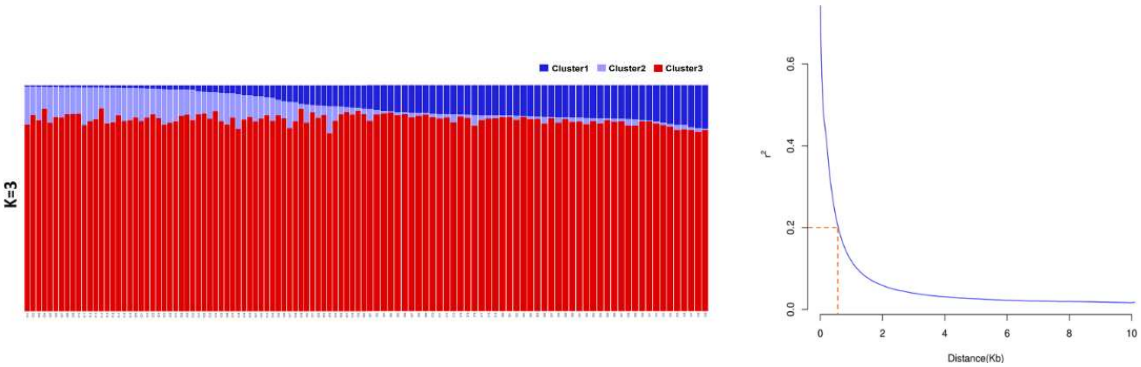
Results

Characterization of the fungal collection. To efficiently perform GWAS and reduce confounding effects, it is important to select a population that is genetically and phenotypically diverse with little stratification. Following these guidelines, we focused on a subset of 103 French *Z. tritici* isolates selected from a larger population for their genetic diversity (data not shown). We sequenced their genomes and generated a total of 247 Gb of sequence data with an average depth of 61x. After quality assessment, we obtained a total of 1,463,638 Single Nucleotide Polymorphisms (SNPs) that we further filtered for Minor Allele Frequency (MAF) and missing rates, resulting in a final set of 718,810 SNPs. Analysis of functional variant effect revealed that 59,318 SNPs (8.25%) induced missense mutations and 2,493 (0.3%) resulted in a loss of function such as premature stop codons, loss of stop or start codons (Table 1). As expected, principal component analysis (PCA) using full polymorphism data did not show any genetic structure; the first three PCs explained less than 4% of the total variance (Fig. S1). We further confirmed this with STUCTURE using a subset of evenly distributed SNPs (1 SNP per 5 kb) and like the PCA analysis, Bayesian clustering failed to infer sub-populations (Fig. 1A). Similarly, individual isolates showed very low genetic relatedness (Fig. S2). Genome-wide LD decayed (r^2<0.2) over physical distance at ~0.5 kb (Fig. 1B). Given our high SNP density (1 SNP every 55 bp) we expected marker-trait associations to be resolved to the individual gene level. When pathotyped on wheat differential cultivars, isolates showed a wide virulence spectrum depending on the host cultivar, and strong isolate × cultivar interactions (Tables S1, S2). Overall, strains isolated from the cultivar 'Premio' were more aggressive compared to those isolated from the susceptible cultivar 'Apache', even when tested on the cultivar 'Apache' (Fig. S3).

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

138 **A** **B**



139

140 **Fig. 1 |** Population structure and Linkage Disequilibrium (LD) in the *Z. tritici* population. **A.** Admixture plot
141 of the fungal population. Each vertical bar represents a single isolate and is colored according to the
142 membership coefficient (Q_i) to the three sub-population clusters from STRUCTURE. **B.** LD decay over
143 physical distance. The r^2 values were calculated between pairs of SNPs up to a physical distance of 20 kb
144 and fitted against the physical distance using a non-linear model. The dotted orange line shows the distance
145 (0.5 kb) at which LD decays at the $r^2 < 0.2$ level.

146

147 **A complex genetic architecture and host-specificity underlies pathogenicity.** To elucidate the
148 genetic basis of the observed pathogenicity differences, we investigated two key quantitative phenotypes
149 (PLACN and PLACP) using mixed linear models for marker-trait association tests, performed
150 independently on each of the studied cultivars. Collectively, we identified 94 genome-wide significant
151 SNPs and an additional 53 at the 10% False Discovery Rate (FDR) threshold, defining 58 distinct genes
152 (Fig. 2, Table S4). The strongest associations were at the *AvrStb6* and *AvrStb9* loci and were detected
153 in all *Stb6*- and *Stb9*- containing cultivars, respectively. Initially not thought to harbour *Stb9*, PLACN
154 on cultivar 'Cadenza' was highly associated (P=3.57e-13) with polymorphisms in *AvrStb9*, suggesting
155 that the cultivar possesses the corresponding *R* gene. The presence of *Stb9* in 'Cadenza' was recently
156 confirmed by its tightly linked genetic markers (data not shown). Interestingly, the single-locus model
157 failed to detect *AvrStb9* and demonstrates the usefulness of the multi-locus mixed model in detecting
158 loci with larger effects. However, if multiple SNPs were in strong LD, biases may be created while
159 estimating the genetic variance. Nevertheless, the multi-locus approach in our study generally
160 outperformed the traditional single-locus tests (Fig. S4). Candidate pathogenicity loci were identified in
161 the thirteen *Z. tritici* core chromosomes. Despite recent suggestions of accessory chromosomes being
162 enriched in putative effectors[32], we could not map any gene residing on accessory chromosomes in this
163 multi-host GWAS (Fig. 2, Table S4). Only eleven genes were shared among PLACP and PLACN (Fig.
164 2, Table S4), which was expected, since necrosis coverage and pycnidia production levels were highly
165 variable (Fig. S3). Hence, considering both traits in association analyses could reveal genetic
166 determinants influencing both, the conidial production and host susceptibility. The phenotypic variance
167 explained by the identified loci associated with pathogenicity traits on each cultivar ranged from 9.1%

to 65.7% (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, analysis of virulent/avirulent alleles showed quantitative differences (Table S4, Fig. 3B). Further, the number of associated genes varied greatly between cultivars (three candidates for 'Bulgaria-88' and 'Toisondor' *vs* ten for 'Israel-493') and ~76% were specific to a single host cultivar (Table S4, Fig. 3A). This weak overlap observed between genes is a strong indication of host-specificity, in consistence with the different *Stb* genes or *Stb* combinations harboured by the cultivar set used here.

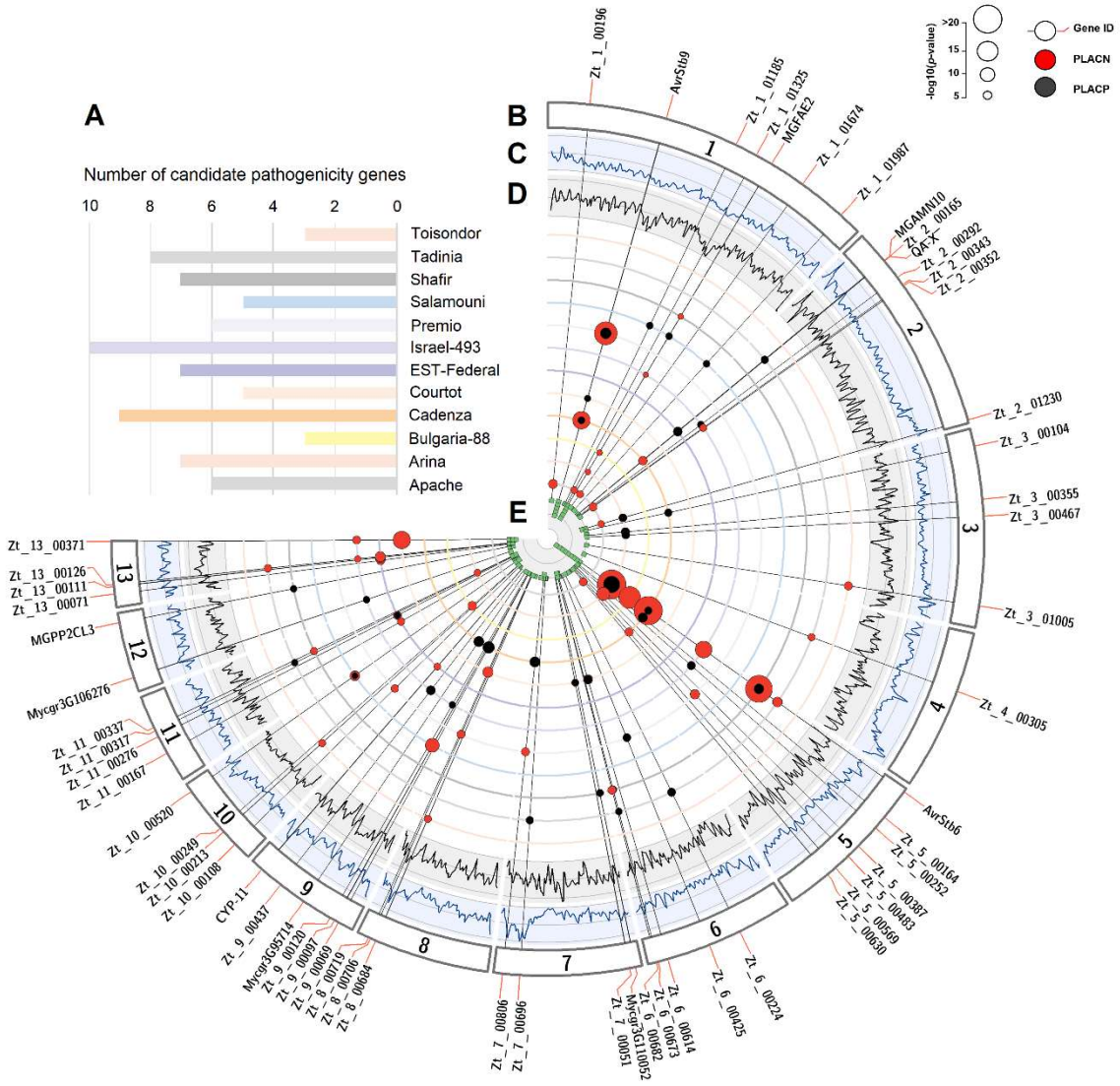
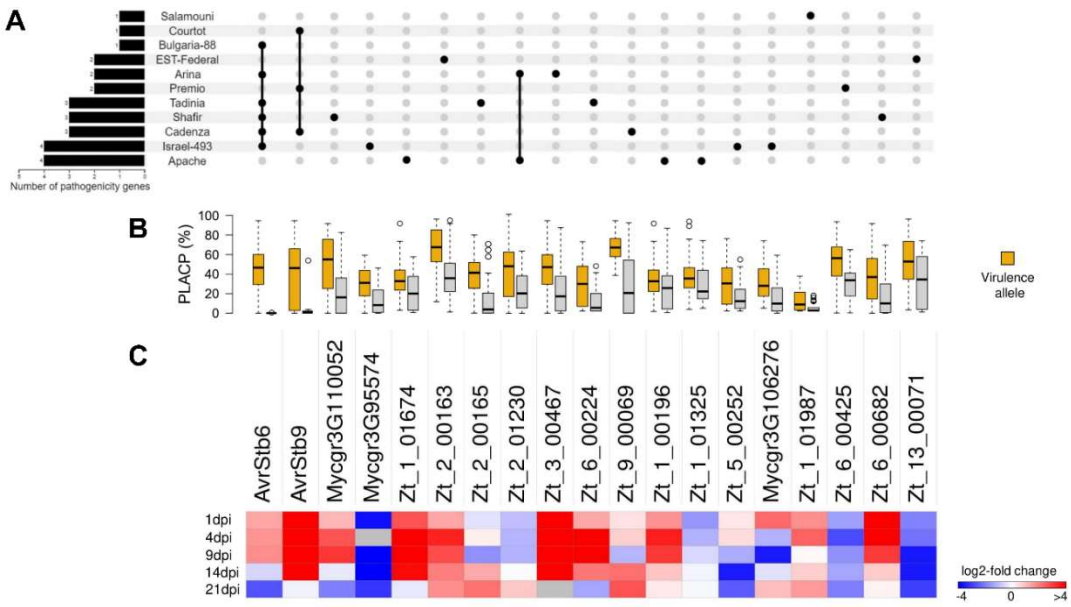


Fig. 2 | Overview of the identified candidate pathogenicity genes from the multi-host GWAS. **A.** Number of identified genes per cultivar. **B.** Gene IDs and their significant associations. Black dots correspond to PLACP (percentage of leaf area covered by pycnidia) and red dots to PLACN (percentage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

183 sliding windows with a step size of 25 kb. Genomic positions are displayed on the x-axis. Tajima's D values
184 are displayed on the y-axis. The y-axis ranges from -2.5 to 2. **E.** The number of associations for each gene
185 (PLACN/PLACP per cultivar).

186 ***In planta* gene expression and functional annotation.** To reveal the expression profile of the identified
187 candidate pathogenicity genes using GWAS, we calculated the relative abundance of *Z. tritici*
188 transcripts, using RNAseq data generated during different *Z. tritici* infection time courses[22]. Of the 58
189 GWAS candidates, ten genes exhibited no expression (FPKM = 0) or very low expression (FPKM < 5)
190 throughout the infection cycle, while another ten genes were found to be highly expressed (FPKM above
191 the 95th percentile of the mean expression value) either at specific time points or throughout the infection
192 cycle (Fig. 3C, Table S5). Interestingly, *AvrStb6*, *Zt_6_00224* and *Zt_3_00467* were among the most
193 highly expressed *Z. tritici* genes at the disease transitional phase (9 dpi; Table S5). Differentially
194 expressed genes *in planta* were identified by examining the ratio of expression between the Czapek-
195 Dox broth (CDB) culture medium and the five time points of wheat infection. In our set of candidate
196 genes, sixteen were differentially expressed with a FDR below 0.10 and a log2-fold change greater than
197 2. Among them, eleven were up-regulated at specific time points of the infection (Fig. 3C; Table S5).
198 For subsequent investigations, we considered a set of nineteen genes as high-confidence candidates,
199 whose expression profiles supported their potential role in pathogenicity; this included the sixteen
200 differentially expressed genes and three additional ones that were highly expressed during the infection
201 (Fig. 3; Table S5).



202 **Fig. 3 |** The quantitative nature and host-specificity of *Z. tritici* pathogenicity. **A.** An upset plot of candidate
203 pathogenicity genes across host cultivars. **B.** Boxplots showing variation in quantitative pathogenicity
204 (PLACP) of the isolates carrying different lead SNP alleles of the candidate pathogenicity genes. **C.**
205 Differential gene expression shown as log2-fold change compared to the culture medium. Red shades indicate
206 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 analysed 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 other
223 protein-coding genes (median $\pi_{Pathogenicity}$=0.0212 vs. $\pi_{Genome\text{-}wide}$=0.0079; $P < 0.01$; Fig. 4A). We found
224 no significant impact of loss of function mutations or presence/absence polymorphisms, except for
225 *Mycgr3G110052*. In addition, we wanted to examine which mechanisms were likely to contribute to the
226 high sequence diversity in our candidates. As TE dynamics have been observed to contribute to fungal
227 genome plasticity and diversity, we calculated the proximity of candidate pathogenicity genes to the
228 nearest TEs. 52.6% of these genes were flanked by or adjacent to TEs (*i.e.* at less than 5 kb distance;
229 Table S7). Comparison with genome-wide genes with regards to TE proximity revealed that candidate
230 pathogenicity genes were significantly closer to TEs ($P < 0.05$; Fig. 4B). Gene diversification is also
231 accelerated by meiotic recombination, which influences both natural population dynamics and selection.
232 We used a combination of algorithms in two independent software (RDP4 and GARD) to detect
233 signatures of recombination. We found signatures of recombination supported by both detection
234 methods in a significant proportion (63.2%) of the putative pathogenicity genes (Table S8).
235 Interestingly, genes with low or no signatures of recombination were in close proximity to TEs, whereas
236 those with higher recombination events were found at greater distances from TEs (Fig. 4C). We
237 hypothesize that recombination in particular and proximity to TEs promotes allelic diversity in genes
238 that are putatively involved in quantitative pathogenicity.

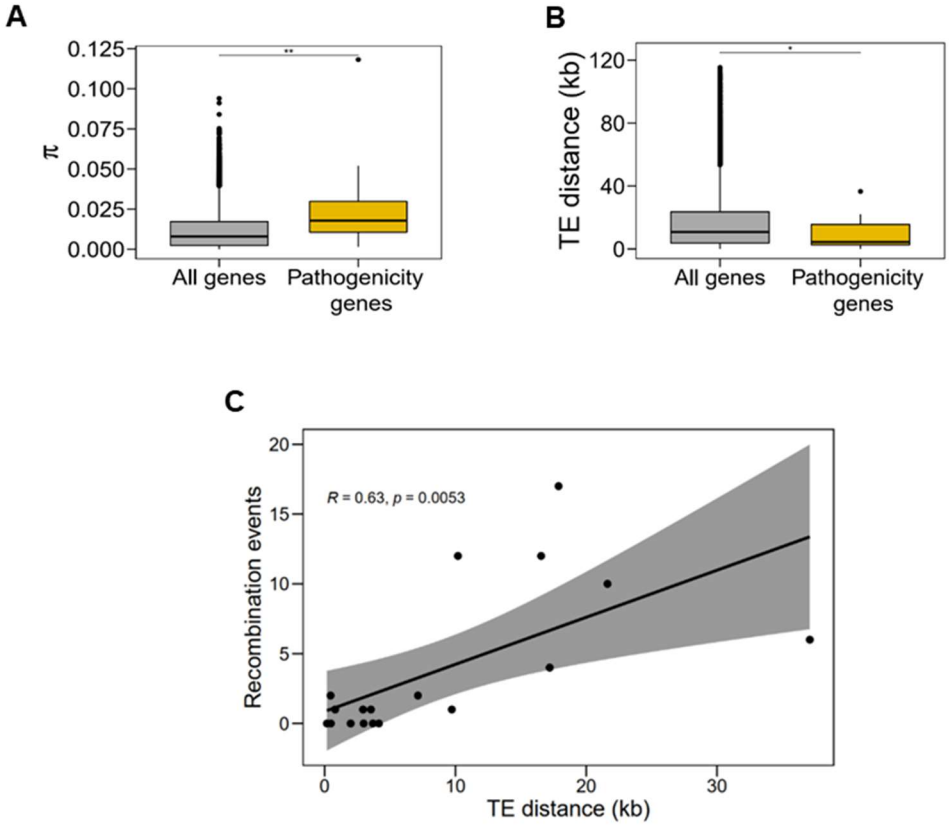


Fig. 4 | Genetic diversity of quantitative pathogenicity genes in *Z. tritici* and factors contributing to gene diversification. **A.** Distribution of nucleotide diversity (π) among candidate pathogenicity genes and the rest of the protein-coding genes. Statistical significance was determined using a permutation test with 1000 iterations; $P<0.01$**. **B.** TE distance compared between pathogenicity genes and the rest of the protein-coding genes in *Z. tritici*. Statistical significance was determined using a permutation test with 1000 iterations; $P<0.05$*. **C.** the relationship between TE distance and recombination events 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 the case of *Zt_1_01987* to 10.97% for *AvrStb6* (Table S7). We also found four genes with a $d_N/d_S > 1$ (*Zt_3_00467*, ω=7.03; *AvrStb6*, ω=2.84; *Mycgr3G106276*, ω=1.18 and *AvrStb9*, ω=1.04), that could be considered as genes under strong selection pressure. Since ω averaged across all gene-tree branches rarely surpasses 1, and as the selective pressure could affect codon sites differently, we tested different codeML models to detect positively-selected codons. For all model comparisons, we found signatures of positive diversifying selection in 68.4% of candidate genes using both, the more conservative M1a vs. M2a and the M7 vs. M8 model

260 pairs (Table S7). Together, this shows that pathogenicity genes are hotspots of positive diversifying
261 selection, likely as a result of a selection pressure exerted by host *R* genes.

262 **Two functionally distinct genes mediate quantitative pathogenicity.** To determine the role of the
263 candidate genes in quantitative pathogenicity, we selected two distinct genes for functional validation:
264 *Zt_3_00467* (Fig. 5A), which encodes a SSP that is highly expressed *in planta*, with a peak at the disease
265 transition phase (Fig. 5B); and *Zt_6_00682* (Fig. 6A), which encodes a S-adenosylmethionine-
266 dependent methyltransferase (SAM-MTase) that is differentially expressed during the disease
267 asymptomatic phase (Fig. 6B). For *Zt_3_00467*, we ectopically expressed the avirulent (IPO10273) and
268 virulent (3D7) alleles in the 3D7 strain background, and assessed quantitative pathogenicity traits on the
269 host cultivars 'Arina' and 'Riband'. Overall, Mutants expressing the avirulent allele (*3D7* +
270 $Zt_3_00467_{IPO10273}$) showed a slight reduction of PLACP levels compared to those expressing the
271 virulent allele (*3D7* + $Zt_3_00467_{3D7}$; Fig. 5C). When compared individually, one mutant expressing
272 the avirulent allele showed a phenotype similar to the wild-type, while another expressing the virulent
273 allele had higher PLACP compared to the wild-type (Fig. S7). The difference in phenotypic responses
274 between ectopic mutants is a well-known phenomenon that may be caused by distinct allele integration
275 sites in the different mutant lines. Testing transformants expressing the avirulent allele on the cultivar
276 'Riband' led to lower development of pycnidia compared to the virulent and the wild-type strains (Fig.
277 S8). 'Arina' and 'Riband' share a common resistance gene known as *Stb15*. These phenotypic
278 observations suggest that *Zt_3_00467* recognition leads to quantitative resistance which could be
279 attributed to *Stb15*. Comparison of the avirulent and virulent protein sequences showed seven amino
280 acid substitutions, all positively selected (Fig. 5D). Analysis of *Zt_3_00467* in the global *Z. tritici*
281 collection showed that the gene was conserved and that most protein variants were population specific
282 and segregating by geographical origin (Fig. S9).

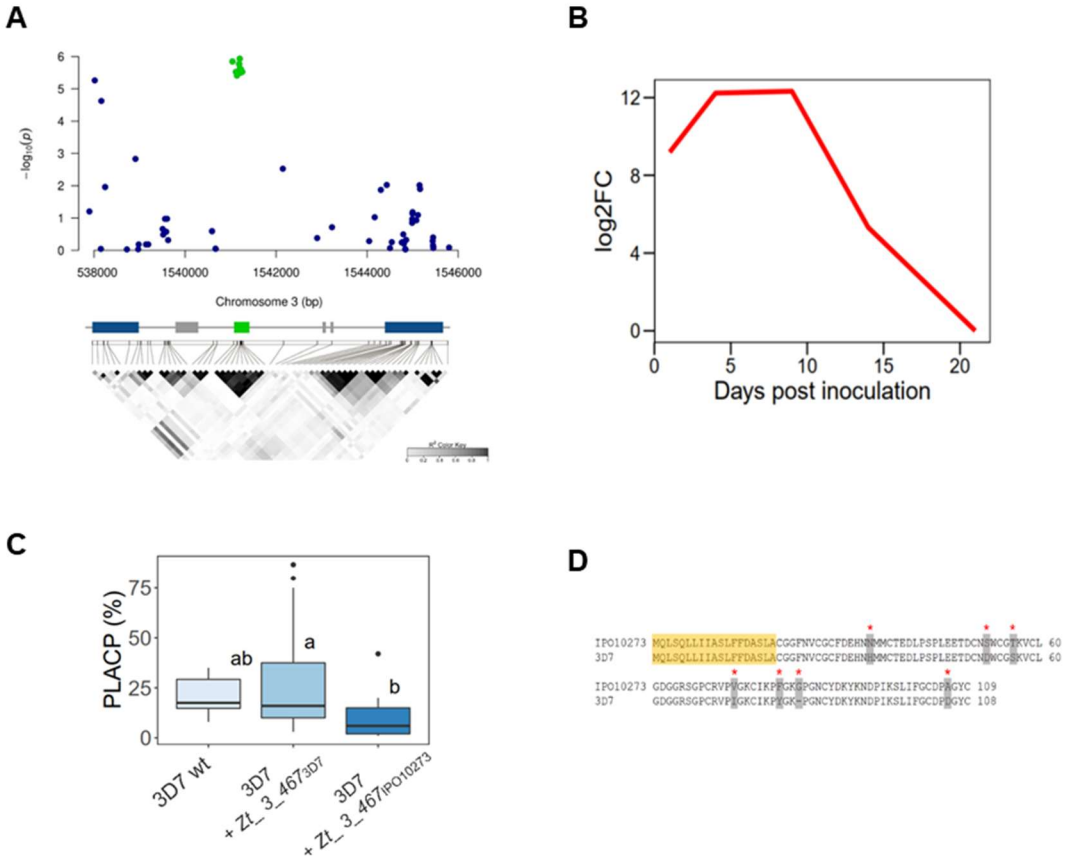


Fig. 5 | An effector gene encoding a positively-selected small secreted protein is involved in quantitative pathogenicity of *Z. tritici*. **A.** Regional association plot of a candidate effector protein detected on cultivar ''Arina''. X and Y axes indicate positions on chromosome 6 and −log10 (*P*-value) for associations with PLACP, respectively. Annotations on the reference genome are shown at the bottom with the candidate gene represented by a green box, other genes by blue boxes and transposable elements by grey boxes. A linkage disequilibrium plot is shown below the annotation bar. **B.** log2-fold expression changes of *Zt_3_00467* based on RNAseq data collected throughout the time course of *Z. tritici* infection on wheat. **C.** percentage of leaf area covered by pycnidia (PLACP) produced by the wild-type strain 3D7 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.

For *Zt_6_00682*, we generated knockout (KO) mutants of the 3D7 strain by targeted gene disruption and ectopic transformants with the virulent (3D7) and avirulent (3D1) alleles in the KO background. This allowed us to assess allele-dependent effects of *Zt_6_00682.* Three independent KO and ectopic transformants for each allele were tested on the cultivar 'Shafir', where the association was detected. Compared to the wild-type, KO transformants resulted in unexpectedly high PLACP values ranging from 67% to 100% (Fig. 6C). Ectopic transformation of the 3D1 allele in the KO background (3D7Δ*Zt_ 6_00682+ Zt_ 6_00682*$_{3D1}$) resulted in reduced levels of PLACP, while ectopic lines

transformed with the 3D7 allele (3D7Δ*Zt_ 6_00682*+ *Zt_ 6_00682*$_{3D7}$) showed a slightly higher PLACP than the wild-type (Fig. 6C). Besides 'Shafir', the deletion mutants were aggressive on another four cultivars, but phenotypic differences between ectopic transformants and wild-type were not significant (Fig. S6). Five amino acid differences were found between the avirulent and virulent protein sequences, of which three were under positive selection (Fig. 6D). Similar to *Zt_3_00467, Zt_ 6_00682* was conserved in the global population but was highly polymorphic (Fig. S9).

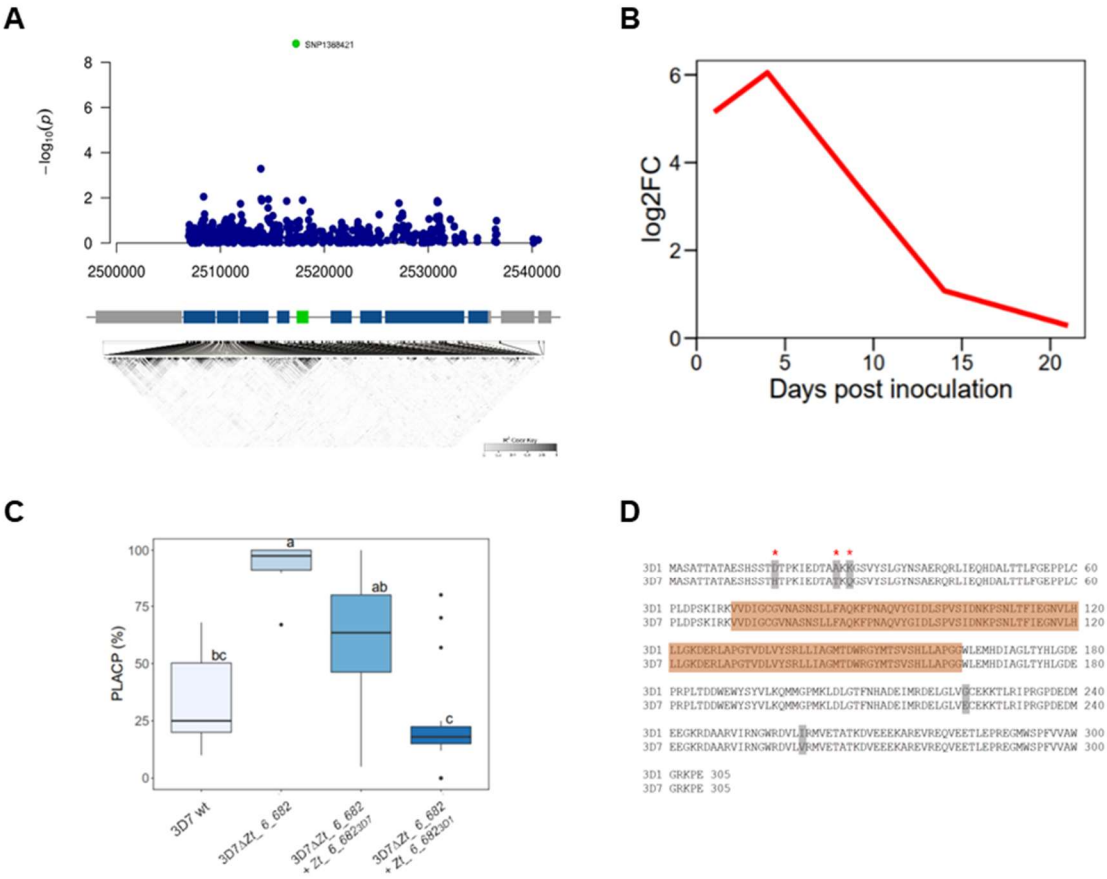


Fig. 6 | A gene encoding a methyltransferase mediates quantitative pathogenicity in *Z. tritici*. **A.** Regional association plot of the pathogenicity gene detected on cultivar ''Shafir''. **B.** log2-fold expression changes of *Zt_6_00682* based on RNAseq transcription data collected throughout the time course of *Z. tritici* infection on wheat. **C.** Percentage of leaf area covered by pycnidia (PLACP) produced by the wild-type strain 3D7, 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.

321 **Discussion**

322 Our results point to a complex genetic architecture of pathogenicity in *Z. tritici*, involving many genes
323 with variable additive effects, resulting in quantitative phenotypes on the host cultivar. These findings
324 are consistent with previous studies[33,34] and provide further evidence that pathogenicity in this
325 pathosystem is complex and results from the contribution of both large- and small- effect genes. We
326 highlighted nineteen differentially and highly expressed genes *in planta* and focused on their features
327 for subsequent investigations. The vast majority of these genes had a peak expression in the phases
328 corresponding to the colonization and the pre-pycnidia formation, before the appearance of symptoms.
329 During colonization, hyphae penetrate into host tissue *via* stomata and colonize the substomatal cavity,
330 but remain exclusively in the intercellular space. This is followed by hyphal invasion of the apoplast
331 leading to pre-pycnidia formation in substomatal cavities[23]. It was shown that these are the key phases
332 of effector gene deployment by plant pathogens[35]. These generally encode small, secreted, cysteine-rich
333 proteins with no sequence homology in other species and play a vital role in successful pathogen
334 invasion, suppression of host immune perception and manipulation of its physiology[36]. Here, we mapped
335 a set of diverse secreted proteins. While effector-like proteins may initiate an effector-triggered immune
336 response, as was shown for the gene *Zt_3_00467*, which triggered quantitative resistance in cultivars
337 'Riband' and 'Arina', secreted molecules with a known functional domain could also promote disease.

338 Of the functionally annotated secreted proteins, we identified domains associated with macromolecule
339 degradation, fungal metabolism or stress response, among others. Some genes encoded proteins that
340 show strong homology with functionally characterized virulence factors from other pathogenic fungi.
341 For instance, *Zt_1_01674* encodes a Thioredoxin reductase (*Trr*), a flavoenzyme that plays a central role
342 in the thioredoxin oxidative stress resistance pathway and is involved in regulating gene transcription,
343 cell growth and apoptosis[37]. Deletion of the *Trr* ortholog resulted in reduced hyphal growth, sexual
344 reproduction and virulence in *Fusarium graminearum*[38] and profoundly affected pathogenicity of the
345 entomopathogenic fungus *Beauveria bassiana*[39]. We hypothesize that *Zt_1_01674* may play similar
346 roles and could be an important factor in *Z. tritici* development and pathogenicity. We also identified a
347 cell-wall degrading enzyme (*Zt_2_00163*) belonging to the Glycoside Hydrolase protein family.
348 Glycoside hydrolases are known to play important roles in fungal invasion and nutrition acquisition via
349 the degradation or modification of plant cell wall components, such as cellulose or hemicellulose[40].

350 Genes encoding proteins predicted to be non-secreted were common in our candidate list, highlighting
351 not only the polygenic nature of pathogenicity, but also the potential role of the non-secretory pathway
352 in fungal pathogenicity. This was demonstrated by the functional studies of *Zt_6_00682*, which encodes
353 a SAM-dependent Methyltransferase (SAM-MTase). SAM-MTases catalyse the transfer of methyl
354 groups from SAM to a large variety of substrates, ranging from metabolites to macromolecules,
355 including proteins[41]. They are involved in epigenetic regulation with functions ranging from the control
356 of gene expression, silencing or activation of transposon control[42]. *Zt_6_00682* shares some sequence

357 similarities with the putative SAM-MTase gene *LaeA*, a component of the velvet transcription factor
358 complex that functions as a global regulator of fungal secondary metabolism and development for a
359 handful of fungi[43]. Recent work showed that mutants lacking the velvet gene *velB* in *Z. tritici* have
360 defects in vegetative growth and asexual reproduction[44], while the loss of *LaeA* attenuates pathogenicity
361 in a variety of filamentous fungi[45]. However, the inactivation of *Zt_6_00682* promoted pycnidia
362 formation, suggesting that it functions differently from *LaeA* and may negatively regulate the
363 transcriptional activity of genes or metabolites associated with host colonization. However, how this
364 SAM-MTase affects the regulation of pathogenicity factors needs to be investigated as a key step
365 towards understanding the evolutionary mechanisms associated with the emergence of highly aggressive
366 strains.

367 In addition, we identified three genes with predicted transmembrane domains. For instance, the *DPM2*
368 ortholog (*Zt_1_01325*) was highly expressed throughout the infection cycle, highlighting its potential
369 role in *Z. tritici* pathogenicity. *DPM2* is a regulatory subunit of the dolichol-phosphate mannose synthase
370 complex, key enzymes for GPI anchor biosynthesis[46]. The alteration of the *DPM* complex affected cell
371 wall integrity, fungal growth and increased chitin levels in *Candida albicans*[46]. Traditionally,
372 transmembrane proteins are systematically discarded in effector gene prediction studies. There is
373 however a growing evidence that they could play key roles in pathogenicity of filamentous fungi and
374 our results support this hypothesis[47,48].

375 We found substantial evidence for host specificity demonstrated by a highly significant isolate × cultivar
376 interaction and few shared associations between host cultivars. It is also probable that the weak overlap
377 in candidate pathogenicity genes among host cultivars is partly due to a lack of statistical power to detect
378 variants of small effect[49]. However, besides *Stb6*- and *Stb9*-carrying cultivars, there was no meaningful
379 correlation among phenotypes on different hosts, which further supports host specificity. In several well
380 studied host-pathogen systems, specificity is determined mainly by the matching of resistance or
381 susceptibility genes and their corresponding effectors[50-52], following a GFG relationship (or inverse GFG
382 in the case of necrotrophic fungal pathogens).

383 In *Z. tritici*, pathogenicity can be both qualitative[26,53,54] and quantitative[10,55]. Here, we showed that
384 pathogenicity is predominantly quantitative based on multiple genes of small-to-moderate effects. This
385 was demonstrated with the allele analysis of candidate genes and functional studies of *Zt_3_00467* and
386 *Zt_6_00682*. In contrast to what has been shown for most GFG interactions, the recognition of the
387 *Zt_3_00467* and *Zt_6_00682* avirulent alleles does not trigger strong resistance, but rather a reduction
388 at the onset of disease symptoms. Furthermore, the fact that avirulent alleles are only recognized by
389 certain wheat cultivars suggests that the interaction also follows a GFG model and demonstrates that
390 *Zt_3_00467* and *Zt_6_00682* are likely quantitative avirulence factors interacting with specific host
391 resistance genes. This is further supported by recent work in *Z. tritici* as shown for *Avr3D1* and a recently
392 reported effector gene (*G_07189*) that triggers quantitative *Stb20q* resistance[11,56]. This is not uncommon

393 in other filamentous fungi. For instance, the late effector LmSTEE98 from *Leptosphaeria maculans*
394 triggers quantitative resistance following a GFG model[57], similarly, silencing three *Blumeria graminis*
395 f sp. *tritici* effectors lead to quantitative gain of resistance[58]. Since *Zt_3_00467* encodes a typical
396 effector protein and is recognized by hosts harbouring the *Stb15* resistance gene, we suggest that
397 *Zt_3_00467* could be *AvrStb15* and trigger an effector-induced defence. We hypothesise that *Stb15*
398 resistance is not strong enough to block pathogen infection, although we cannot completely rule out the
399 existence of a *Zt_3_00467* haplotype that triggers qualitative resistance. However, additional genetic
400 factors might also contribute to differences in pathogenicity between avirulent and virulent strains. On
401 the other hand, *Zt_6_00682* encodes an atypical avirulence factor and may interact indirectly with the
402 host immune system. The fact that *Zt_6_00682* resembles genes involved in transcriptional regulation
403 and that its deletion mutants are more aggressive on some cultivars, suggests that *Zt_6_00682* negatively
404 regulates pathogenicity factors that are specifically recognized by one or more (unknown) host
405 quantitative resistance gene(s). Complementation mutants further support host specificity, where
406 recognition of the avirulent allele leads to a dramatic reduction in pycnidia formation. Unlike deletion
407 mutants, this was only observed on the cultivar where the gene was detected initially ('Shafir').
408 However, the resistance gene involved and the underlying mechanisms through which the avirulent
409 allele is recognized remain unknown.

410 In addition to *AvrStb6* and *AvrStb9*, we found the SSP *Zt_2_01230* to be significantly associated with
411 variation in quantitative pathogenicity on two cultivars, suggesting a common yet unknown *Stb*
412 resistance gene. We also identified a number of secreted and species specific proteins. Since canonical
413 fungal effectors are generally thought to be secreted, species-specific and without conserved protein
414 domain[59], we propose that the ones identified here constitute a set of candidate effectors interacting with
415 their cognate host *Stb* resistance products, and should be a priority target for functional studies.
416 Unlocking the molecular basis underpinning these interactions and how it could impact host defence
417 would lead to a better exploitation of quantitative resistance, which is thought to be more durable[60].

418 Evolution and diversification of pathogenicity genes enables adaptation to new host cultivars and, in
419 response to host population changes, they undergo changes in rapid allele frequency[61]. We observed
420 high levels of genetic diversity in our candidate pathogenicity genes when compared to the genome-
421 wide gene content. We found a single case of gene deletion and no significant impact of loss of function.
422 Because of the intact and functional coding sequences for nearly all of the identified genes, we suggest
423 that sequence polymorphism is the primary mechanism behind the rapid adaptation of *Z. tritici* to host
424 populations, permitting the pathogen to circumvent host resistance while conserving the presumable
425 function of the pathogenicity protein. We next investigated what could contribute to this extensive
426 genetic diversity. Recombination is a central source of diversity in sexual fungal pathogens[62]. Here, we
427 detected recombination events for a significant number of candidates, which suggests a role in the
428 observed genetic diversity of pathogenicity genes. In *Z. tritici*, it was reported that recombination is a

429 major driver of its evolution and is highly heterogeneous, concentrated in narrow genomic regions that
430 are enriched in putative effector genes[8,63]. TEs are also thought to contribute to gene diversification by
431 creating fast-evolving genomic regions through non-homologous recombination or repeat-induced point
432 mutations[64]. Here, we found that candidate pathogenicity genes are indeed at closer distances to TEs,
433 suggesting their potential role in sequence diversity. We also observed that TE distance was correlated
434 with recombination events in putative pathogenicity genes, and that the levels of genetic diversity were
435 high for genes with the highest recombination events, regardless of their distances to TEs. We
436 hypothesize that recombination is likely the major driver underlying the diversification of pathogenicity
437 genes. A recent study supports this hypothesis where, unlike distance to TEs, the recombination rate
438 was strongly associated with adaptive substitutions in *Z. tritici* genomes[4].

439 A key determinant of genetic diversity at pathogenicity genes is selection pressure. Understanding the
440 evolution of pathogenicity in fungal populations with respect to the selection pressure imposed by host
441 populations is necessary for the sustainable management of *R* genes in the field. This can be achieved
442 by searching for signatures of positive diversifying selection at the risk of recovering some false positive
443 signals. In this study, we used conservative maximum likelihood models and models that have been
444 shown to be robust to recombination and significantly reduce false positives[65]. More than two thirds of
445 candidate genes harbour signatures of positive selection acting on specific codons, suggesting that the
446 diversifying selection is likely the primary evolutionary force maintaining non-synonymous
447 substitutions relevant for *Z. tritici*-wheat quantitative interactions. For instance, the vast majority of
448 different mutations between the avirulent and virulent versions of *Zt_3_00467* and *Zt_6_00682* have
449 been identified as positively selected. Such selection pressure could be the result of host *R* gene
450 recognition, and may have greatly contributed to the accumulation of beneficial mutations. This agrees
451 with a previous study, which demonstrated that rapidly evolving genes play a prominent role in host-
452 specific disease development[66]. Together, these results suggest that host–pathogen coevolution drives
453 the emergence and maintenance of allelic diversity at loci likely involved in direct (or indirect)
454 interactions with host *R* genes. Future surveys of the diversity and structure of *R* genes in wheat
455 populations will provide a deeper insight into reciprocal patterns of selection, coevolution, and the
456 maintenance of diversity in resistance and pathogenicity in natural populations.

457 **Methods**

458 **Fungal material.** We analysed 103 *Z. tritici* isolates in this study deriving predominantly from a large
459 collection of French isolates collected in 2009 and 2010. This genetically diverse sample was mainly
460 isolated from the French cultivars 'Apache' (n=44) and 'Premio' (n=42). The remaining were isolated
461 from cultivars 'Soissons' (n=9), 'Caphorn' (n=3), 'Alixan' (n=1), 'Bermude' (n=1), 'Dinosor' (n=1),
462 and 'Garcia' (n=1). One isolate was collected in the United Kingdom in 2009 on cultivar 'Humbert'.
463 All strains are monospore non-clonal isolates were originally isolated either at the Plant Research
464 International in the Netherlands or at INRAE UR BIOGER in France. For the functional validation of

465 candidate pathogenicity genes, we used the two strains 3D1 (*i.e.*, ST99CH3D1) and 3D7 (*i.e.*,
466 ST99CH3D7) collected in 1999 from a Swiss wheat field and previously used for the functional
467 validation of *Avr3D1*[10].
468
469 **Plant material.** The pathogenicity of the 103 isolates was initially evaluated on 29 wheat cultivars,
470 including 15 carrying known *Stb* resistance genes. After the first replicate, only cultivars on which we
471 observed phenotypic variability were kept for second and third replicates. These were the cultivars
472 'Bulgaria-88' (*Stb1* and *Stb6*), 'Israel-493' (*Stb3* and *Stb6*), 'Tadinia' (*Stb4* and *Stb6*), 'Shafir' and
473 'Cadenza' (*Stb6*), 'Estanzuela Federal' (*Stb7*), 'Courtot' (*Stb9*), 'Apache' (*Stb4/5* and *Stb11*),
474 'Salamouni' (*Stb13*, *Stb14* and *StbSm3*) and 'Arina' (*Stb6* and *Stb15*); the resistance background of
475 cultivars 'Premio' and 'Toisondor' was not known. Details on the differential lines are available in Table
476 S3. This genetic material is available from the INRAE Small grain cereals Biological Resources Centre
477 in Clermont-Ferrand (<https://www6.clermont.inrae.fr/umr1095_eng/Organisation>).
478
479 **Quantitative pathogenicity assays.** For inoculum preparation, blastospores of *Z. tritici* stored at -80°C
480 were used to inoculate a liquid growth medium (Difco™ YPD Broth Cat. No. 242820) agitated at 140
481 rpm in a 17°C climate cell for 144 h and transferred to agar growth medium (Difco™ YPD Agar Cat.
482 No. 242720) for 96 h. Fresh blastospores were recovered with loops, resuspended in sterile water, and
483 the inoculum was adjusted to a concentration of 10^6 spores.L^{-1}, to which one drop of Tween®20 (Sigma
484 Aldrich® Cat. No. 9005-64-5) was added per 15 mL before inoculation. Five seeds from each cultivar
485 were sown in 90*90*95 mm pots containing Floradur B soil (80% blond peat / 20% black peat, with 1.2
486 kg.m^{-3} fertilizer 18/10/20). One day before inoculation, we kept three homogeneous seedlings on which
487 a 75-mm-length portion was delimited on the first true leaf using a black marker. Sixteen days old
488 seedlings were inoculated by applying the inoculum six times on the marked leaves with a flat tipped
489 paintbrush (synthetic bristles, 12 mm width), depositing approximately 30,000 spores.cm^{-2} on the leaf
490 surface. After inoculation, plants were watered and enclosed in transparent polyethylene high-density
491 bags for 72 h to initiate the infection. Ten days post-inoculation, leaves were removed keeping only the
492 marked leaf to ensure light homogeneity and longevity of the inoculated leaf. From sowing to disease
493 scoring, plants were kept in growth chambers regulated at 80% relative humidity, 16 h of light per 24 h
494 at a photon flux of 300 µmol.m^{-2}.s^{-1}, 22°C temperature under light and 18°C in the darkness. Quantitative
495 pathogenicity was assessed as percentages of leaf area covered by pycnidia (PLACP) and by necrosis
496 (PLACN) at 21 days post inoculation, averaged over three leaves and three replicates (nine leaves per
497 interaction). Phenotypic characterization of the generated deletion and ectopic transformants was done
498 following the same procedure described above in two replicates.
499
500 **Whole-genome resequencing.** Blastospores were harvested after seven days growth in liquid cultures
501 on yeast extract-peptone-dextrose Broth (Difco™ YPD Broth), and then lyophilised for 24 h. High

502 molecular weight DNA was extracted using a phenol-chloroform-based procedure[67]. Illumina paired-
503 end libraries were prepared for each sample and DNA fragments were re-sequenced with 2x100 bp reads
504 on a HiSeq-2000 sequencing system at Genewiz, Inc. (formerly Beckman Coulter Genomics). Depth of
505 sequence coverage was variable depending on the isolate, ranging from 16 to 303 genome equivalents
506 with an average of 61 genome equivalents.
507
508 **SNP calling and quality filtering.** For each isolate, reads were trimmed using trimmomatic v0.32[68] and
509 then mapped to the reference genome IPO323[69] using the mem algorithm from BWA v0.7.7 with default
510 settings[70]. Samtools v0.1.19 and Picard tools v1.106 were used to filter reads[71]
511 (<http://broadinstitute.github.io/picard/>); to remove optical duplicates, secondary alignments, reads
512 with a mapping quality below 30 and to keep only pairs in which both reads met quality checks. SNP
513 calling was done with Freebayes v0.9 using the option --report-monomorphic[72] and all positions
514 detected as low complexity regions or TEs were excluded. Low-complexity regions and TEs were
515 detected using RepeatMasker with default settings (RepeatMasker Open-4.0
516 <http://www.repeatmasker.org>) and REPET[73], respectively. Quality filters were applied to VCF files
517 using in-house scripts. Depth of coverage (DP) calculated at each position followed a normal distribution
518 around the depth main peak. Hence, a DP filter was set at twice the standard deviation of the mean. All
519 VCFs were stored as a matrix reduced to keep only positions where at least one isolate carries an
520 alternative allele at a non-filtered position. The SNP matrix was parsed as a .tped format, converted to
521 .ped/.map. The final matrix was filtered for less than 50% missing data and more than 10% Minor Allele
522 Frequency (MAF) using PLINK v.1.9[74] <https://cog-genomics.org/plink/>.

523 **Linkage disequilibrium (LD).** To assess the mapping power and resolution of GWAS, we estimated
524 genome-wide LD decay using PopLDdecay[75]. The average squared Pearson's correlation coefficient (r^2)
525 was computed between pairs of SNPs up to 20 kb, and plotted over physical distance by fitting a
526 nonlinear regression model.

527 **Population structure analyses.** Two approaches were used to investigate the fungal population
528 structure: (1) we performed a principal component analysis (PCA) using independent SNPs ($r^2 < 0.2$)
529 with the --pca command PLINK v.1.9 and (2) the admixture model in STRUCTURE v.2.2 software[76]
530 with a 10^5 burn-in period, 5×10^5 MCMC steps and 5 iterations for each k value from 1 to 10. The optimal
531 number of sub-populations was determined with the Evanno method[77].

532 **GWAS.** For each quantitative trait, we conducted GWAS using single- and multi-locus linear mixed
533 models based on the EMMA algorithm implemented in GAPIT v.3.0[78-80]. We used a VanRaden derived
534 kinship matrix as a random effect, and set the PCA option to 'true' to determine the optimum number
535 of PCs to be included in the model, following the Bayesian Information Criterion (BIC). The most robust
536 model was chosen by visualizing Quantile-Quantile plots that illustrate the observed vs expected *P*
537 values. A uniform threshold of $P=1/n$ (n being the effective number of independent SNPs) was used as

538 the genome-wide significance threshold. The effective number of independent SNPs was calculated
539 using the Genetic type 1 Error Calculator; <http://grass.cgs.hku.hk/gec/download.php>. The suggestive
540 genome-wide threshold for the 718,810 SNPs was 3.96×10^{-6} and the significant *P* value was 1.98×10^{-7}.
541 We also adjusted SNP *P*-values to account for multiple testing correction using the false discovery rate
542 (FDR) procedure. SNPs with an FDR-adjusted *P* value < 0.10 were deemed significant. For LD analysis,
543 we first calculated the pairwise LD with the lead SNP in a window size of 1 Mb using the --r^2 command
544 in PLINK v.1.9. We set a cut-off of $r^2 > 0.5$ to determine candidate genomic regions and assigned
545 predicted genes using a custom Perl script. In case more than one gene is assigned to the candidate
546 region, we picked the closest one from the lead SNP having pathogenicity-like properties (*e.g.*, secretion
547 signal, differential expression *in planta*). If no gene fell in the LD region, we selected the closest gene
548 to the lead SNP. LD heatmaps were constructed using Haploview v4.2[81].

549 **Functional annotation.** We inferred gene models from the *ab initio Z. tritici* annotation[82] in the
550 candidate genomic regions identified by GWAS. We used RNAseq datasets of the reference isolate
551 IPO323[22] to check for missing genes or manually refine gene models (*e.g.*, splice junctions, missing
552 exons). N-terminal secretion signals were predicted using the Hidden Markov Model (HMM) scoring
553 method implemented in SignalP 3.0[83]. Transmembrane domains were identified with TMHMM 2.0[84].
554 Effector probabilities were predicted through EffectorP 3.0[85]. Conserved protein domains were assigned
555 using HMM scan against the PFAM 35.0 database and blastp was used to search for homologs from
556 other fungi. Finally, we used PHIB-BLAST, a blast server for PHI-base (Pathogen-Host Interactions),
557 to search for homologs against functionally characterized pathogenicity factors.

558 **Gene expression analysis.** RNAseq data from *in vitro* and *in planta* (1-, 4-, 9-, 14- and 21-days post
559 infection; dpi) were obtained from a previous study[22]Cliquez ou appuyez ici pour entrer du texte.. We
560 computed gene expression quantified as fragments per kilobase of exon per million fragments mapped
561 (FPKM) using the improved gene models with Cuffdiff v.2.2.1[86]. We inferred pairwise log2-fold
562 expression changes and *P* value adjustment for *in planta* RNAseq samples compared to the Czapek-Dox
563 broth (CDB) culture medium. Genes with an FDR-adjusted *P*-value < 0.10 and a log2|fold change|> 2
564 were considered as differentially expressed during the infection.

565

566 ***De novo* assembly and sequence diversity.** We used SPAdes v.3.14.1 to produce draft assemblies for
567 each isolate[87]. We ran the tool with the following settings: -k 21,33,55,75,95 --careful to reduce
568 mismatches and indel errors. We performed blastn[88] searches using ~1000 bp of the target locus as a
569 query to retrieve matching contig segments in each assembly. High-confidence matches were first
570 extracted from the contig set using samtools v1.10 and subsequently aligned with MAFFT v7.464 using
571 the --auto option and 1,000 iterative refinement cycles[89]. Alignments were inspected using JalView[90]
572 and ambiguous regions were removed with trimAL[91] with less stringent parameters. We used the batch

573 mode of DNAsp v.6[92] to calculate summary statistics of population genetic parameters associated with
574 candidate genes such as nucleotide diversities (π)[93] and Tajima's D[94]. TE locations for *Z. tritici* were
575 retrieved from a previous study[82] and we parsed gene distances to TEs using the command 'closest'
576 from bedtools[95].

577

578 **Evolutionary analysis.** Before testing for positive diversifying selection, we checked for evidence of
579 recombination, as this can cause false positives. We used the software packages RDP4[96] and GARD[97]
580 for recombination detection. We considered genes with recombination events if both detection methods
581 found evidence of recombination. We then constructed non-recombinant phylogenetic trees of gene
582 coding sequences using the RAxML algorithm[98] implemented in RDP4. Statistical support for the nodes
583 was obtained after 100 bootstrap replicates. We used codeml of the PAML package[99] to estimate the
584 evolutionary rates of candidate genes, based on Markov models of codon evolution. Branch lengths of
585 the inferred ML trees were estimated by using the model M0 and then fixed in further analyses. The
586 model M0 assumes constant selective pressure across codon sites and over time. The selective pressure
587 at the coding-sequence level was measured by ω, the ratio of non-synonymous to synonymous rates
588 (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
589 diversifying selection, respectively. Models M1a, M2a, M7 and M8 of variable selective pressure across
590 codon sites were used to estimate selective pressure and test for positive selection[100,101]. For each
591 candidate gene, we performed pairwise likelihood ratio tests (LRTs) for positive selection, comparing
592 models that allow sites with $\omega > 1$ (M2a and M8) with models that do not (M1a and M7). We considered
593 genes under positive diversifying selection if both model comparisons in LRT were significant. Finally,
594 positively selected sites were inferred with a Bayes Empirical Bayes (BEB) probability[101] and codons
595 with $\omega > 1$ at a probability > 0.95 were considered positively-selected.

596

597 **Functional validation of candidate pathogenicity genes.** For *Zt_6_00682*, we generated knockout
598 mutants in the 3D7 isolate and ectopic transformants expressing the virulent and avirulent alleles in the
599 knock-out background. To create plasmid constructs for gene disruption, flanking regions (~1 kb before
600 start and stop codons) of *Zt_6_00682* were amplified from gDNA using primers listed in Table S9. The
601 hygromycin resistance gene cassette was amplified from the plasmid pNOV2114_HygR_attP1_attP2
602 and cloned into a pNOV2214 vector. The vector was then digested with *Hind*III restriction enzyme (New
603 England Biolabs), amplified and used for the cloning of PCR fragments with a Gibson Assembly cloning
604 kit (New England Biolabs) following the manufacturer's instructions. Constructs were transformed into
605 *Escherichia coli* NEB 5-α using a heat shock transformation, screened by PCR and verified by Sanger
606 sequencing using primers listed in Table S9. Confirmed plasmids were transformed into *Agrobacterium*
607 *tumefaciens* cells by heat shock and validated by PCR. The vector was used to transform *Z. tritici* via *A.*
608 *tumefaciens*-mediated transformation (ATMT) following a procedure adapted from Zwiers & de
609 Waard[102]. Transformed *Z. tritici* colonies growing in the presence of hygromycin were tested by PCR to

610 confirm the successful deletion of the gene. To validate the allele-dependent effect of *Zt_6_00682*, we
611 generated and validated ectopic transformants by placing both the virulent (3D7) and the avirulent (3D1)
612 alleles into the 3D7 KO background, following the same procedure described above with the following
613 modifications: (*i*) we used the vector pNOV_3Gate_SUL with sulfonylurea as a resistance marker and
614 the digestion was achieved with *Hind*III and *Kpn*I restriction enzymes. The resulting vectors,
615 pNOV_sulf_3D7g.7232 and pNOV_sulf_3D1g.7155, were used to transform *Z. tritici* via ATMT
616 following the procedure described above. Primers used to verify gene disruptions and ectopic
617 integrations are listed in Table S9. Finally, three independent transformants carrying the different
618 virulence alleles and three knock-out mutants were used for phenotypic characterization *in planta*. For
619 *Zt_3_00467*, we ectopically expressed the virulent (3D7) and avirulent (IPO10273) alleles in the 3D7
620 background. A fragment containing the *Zt_3_00467* gene, including 1800 bp upstream of the start codon
621 and 463 bp downstream of the stop codon was amplified using Phusion DNA polymerase (NEB) and
622 the primers included in table S9. This fragment was cloned into a pCGEN vector[103] previously digested
623 with *Kpn*I (New England Biolabs), using the In-Fusion HD Cloning Kit (Takara Bio) and following the
624 manufacturer's instructions. Vectors were transformed using heat-shocked *E. coli* Stellar cells and
625 verified using Sanger sequencing. Transformation via ATMT was performed as described above and
626 selection for the *Z. tritici* mutants was performed on geneticin (150 $\mu g.mL^{-1}$)-containing plates.
627

628 **Data availability**

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

631 **Acknowledgements**

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

640 **Financial Disclosure**

641 This work was supported by the French National Research Agency (program BIOADAPT, ANR-12-
642 ADAP-0009-04), the FSOV 2008 B project and BASF France SAS for R.A. postdoctoral fellowship.
643 T.C.M. was supported by a Short-Term Scientific Mission granted by the COST SUSTAIN action
644 FA1208. A.S.V. was supported by the Ministerio de Ciencia e Innovación (Ramón y Cajal program,

grant no. RYC2018-025530-I). INRAE BIOGER benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Author contributions

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

Competing Interests

The authors declare they have no competing interests.

References

1. Savary, S., et al. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **3**, 430–439 (2019). https://doi.org/10.1038/s41559-018-0793-y
2. Franceschetti, M., et al. Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiol. Mol. Biol. Rev.* **81**, e00066-16 (2022). https://doi.org/10.1128/MMBR.00066-16
3. lo Presti, L., et al. Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* **66**, 513–545 (2015). https://doi.org/10.1146/annurev-arplant-043014-114623
4. Grandaubert, J., Dutheil, J. Y., & Stukenbrock, E. H. The genomic determinants of adaptive evolution in a fungal pathogen. *Evol. Lett.* **3**, 299–312 (2019). https://doi.org/10.1002/evl3.117
5. Möller, M., & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* **15**, 756–771 (2017). https://doi.org/10.1038/nrmicro.2017.76
6. Raffaele, S., & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* **10**, 417–430 (2012). https://doi.org/10.1038/nrmicro2790
7. Rouxel, T., et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. *Nat. Commun.* **2**, 202 (2011). https://doi.org/10.1038/ncomms1189
8. Stukenbrock, E. H., & Dutheil, J. Y. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics* **208**, 1209–1229 (2018). https://doi.org/10.1534/genetics.117.300502
9. Singh, P.K., et al. Co-evolutionary interactions between host resistance and pathogen avirulence genes in rice-*Magnaporthe oryzae* pathosystem, *Fungal Genet. Biol.* **115**, 9-19 (2018). https://doi.org/10.1016/j.fgb.2018.04.005
10. Meile, L., et al. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. *New Phytol.* **219**, 1048–1061 (2018). https://doi.org/10.1111/nph.15180

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

736 29. Saintenac, C., et al. A wheat cysteine-rich receptor-like kinase confers broad-spectrum resistance
737 against Septoria tritici blotch. *Nat. Commun.* **12**, 433 (2021). https://doi.org/10.1038/s41467-020-
738 20685-0

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

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

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

748 33. Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. A fungal wheat pathogen
749 evolved host specialization by extensive chromosomal rearrangements. *ISME J.* **11**, 1189–1204
750 (2017). https://doi.org/10.1038/ismej.2016.196

751 34. Stewart, E. L., et al. Quantitative trait locus mapping reveals complex genetic architecture of
752 quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.* **19**, 201–216
753 (2018). https://doi.org/10.1111/mpp.12515

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

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

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

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

764 39. Zhang, L.-B., Tang, L., Ying, S.-H., & Feng, M.-G. Distinct roles of two cytoplasmic thioredoxin
765 reductases (Trr1/2) in the redox system involving cysteine synthesis and host infection of *Beauveria*
766 *bassiana*. *Appl. Microbiol. Biotechnol.* **100**, 10363–10374 (2016). https://doi.org/10.1007/s00253-
767 016-7688-0

768 40. Esquerré-Tugayé, M.-T., Boudart, G., & Dumas, B. Cell wall degrading enzymes, inhibitory
769 proteins, and oligosaccharides participate in the molecular dialogue between plants and pathogens.
770 *Plant Physiology and Biochemistry*, *38*(1), 157–163 (2000). https://doi.org/10.1016/S0981-
771 9428(00)00161-3

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

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

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

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

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

46. Juchimiuk, M., Kruszewska, J., & Palamarczyk, G. Dolichol phosphate mannose synthase from the pathogenic yeast *Candida albicans* is a multimeric enzyme. *Biochim. Biophys. Acta Gen. Subj.* **1850**, 2265–2275 (2015). https://doi.org/10.1016/j.bbagen.2015.08.012

47. Weiland, P., & Altegoer, F. Identification and characterization of two transmembrane proteins required for virulence of *Ustilago maydis*. *Front. Plant Sci.* **12**, 669835 (2021). https://doi.org/10.3389/fpls.2021.669835

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

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

50. Bourras, S., et al. Multiple avirulence loci and allele-specific effector recognition control the *Pm3* race-specific resistance of wheat to powdery mildew. *The Plant Cell* **27**, 2991–3012 (2015). https://doi.org/10.1105/tpc.15.00171

51. Faris, J. D., et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13544–13549 (2010). https://doi.org/10.1073/pnas.1004090107

52. Liu, W., et al. Recent progress in understanding PAMP- and effector-triggered immunity against the rice blast fungus *Magnaporthe oryzae*. *Mol. Plant* **6**, 605–620 (2013). https://doi.org/10.1093/mp/sst015

53. Brading, P. A., Verstappen, E. C. P., Kema, G. H. J., & Brown, J. K. M. A gene-for-gene relationship between wheat and *Mycosphaerella graminicola*, the Septoria tritici blotch pathogen. *Phytopathology* **92**, 439–445 (2012). https://doi.org/10.1094/PHYTO.2002.92.4.439

54. Kema, G. H. J., Verstappen, E. C. P., & Waalwijk, C. Avirulence in the wheat Septoria tritici leaf blotch fungus *Mycosphaerella graminicola* is controlled by a single locus. *Mol. Plant-Microbe Interact.* **13**, 1375–1379 (2000). https://doi.org/10.1094/MPMI.2000.13.12.1375

55. Zhan, F., et al. Linear correlation analysis of *Zymoseptoria tritici* aggressiveness with *in vitro* growth rate. *Phytopathology* **106**, 1255–1261 (2016). https://doi.org/10.1094/PHYTO-12-15-0338-R

56. Meile L, et al. Natural variation in Avr3D1 from *Zymoseptoria* sp. contributes to quantitative gene-for-gene resistance and to host specificity. *New Phytol.* **238**, 1562-1577 (2023). https://doi.org/10.1111/nph.18690

817 57. Jiquel, A., et al. A gene-for-gene interaction involving a 'late' effector contributes to quantitative
818 resistance to the stem canker disease in *Brassica napus*. *New Phytol.* **231**, 1510–1524 (2021).
819 https://doi.org/10.1111/nph.17292

820 58. Schaefer, L.K., et al. Cross-kingdom RNAi of pathogen effectors leads to quantitative adult plant
821 resistance in wheat. *Front. Plant Sci.* **11**, 253 (2020). https://doi.org/10.3389/fpls.2020.00253

822 59. Kim, K.-T., et al. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their
823 potential role in host association. *Front. Plant Sci.* **7**, 186 (2016).
824 https://www.frontiersin.org/article/10.3389/fpls.2016.00186

825 60. Kettles, G. J., & Kanyuka, K. Dissecting the molecular interactions between wheat and the fungal
826 pathogen *Zymoseptoria tritici*. *Front. Plant Sci.* **7**, 508 (2016).
827 https://www.frontiersin.org/article/10.3389/fpls.2016.00508

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

831 62. Burdon, J. J., & Silk, J. Sources and patterns of diversity in plant-pathogenic fungi. *Phytopathology*
832 **87**, 664–669 (1997). https://doi.org/10.1094/PHYTO.1997.87.7.664

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

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

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

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

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

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

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

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

855 71. Li, H., et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
856 (2009). https://doi.org/10.1093/bioinformatics/btp352

857 72. Garrison, E., & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv*,
858 1207.3907 (2012). http://arxiv.org/abs/1207.3907

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

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

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

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

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

78. Kang, H. M., et al. Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008). https://doi.org/10.1534/genetics.107.080101

79. Lipka, A. E., et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012). https://doi.org/10.1093/bioinformatics/bts444

80. Segura, V., et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **44**, 825–830 (2012). https://doi.org/10.1038/ng.2314

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

82. Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. 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 Genom. Genet.* **5**, 1323-1333 (2015). https://doi.org/10.1534/g3.115.017731

83. Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004). https://doi.org/10.1016/j.jmb.2004.05.028

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

85. Sperschneider, J., & Dodds, P. N. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microbe Interact.* **35**, 146–156 (2021). https://doi.org/10.1094/MPMI-08-21-0201-R

86. Trapnell, C., et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012). https://doi.org/10.1038/nprot.2012.016

87. Bankevich, A., et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012). https://doi.org/10.1089/cmb.2012.0021

88. Camacho, C., et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009). https://doi.org/10.1186/1471-2105-10-421

89. Katoh, K., & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). https://doi.org/10.1093/molbev/mst010

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

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

907 92. Rozas, J., et al. DnaSP 6: DNA Sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.*
908 **34**, 3299–3302 (2017). https://doi.org/10.1093/molbev/msx248
909 93. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
910 https://doi.org/10.7312/nei-92038
911 94. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
912 *Genetics* **123**, 585–595 (1989). https://doi.org/10.1093/genetics/123.3.585
913 95. Quinlan, A. R., & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
914 *Bioinformatics* **26**, 841–842 (2010). https://doi.org/10.1093/bioinformatics/btq033
915 96. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. RDP4: Detection and analysis of
916 recombination patterns in virus genomes. *Virus Evol.* **1** (2015). https://doi.org/10.1093/ve/vev003
917 97. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. GARD: a
918 genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098 (2006).
919 https://doi.org/10.1093/bioinformatics/btl474
920 98. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
921 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). https://doi.org/10.1093/bioinformatics/btu033
922 99. Yang, Z., & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol.*
923 *Evol.* **15**, 496–503 (2000). https://doi.org/10.1016/S0169-5347(00)01994-7
924 100. Nielsen, R., & Yang, Z. Likelihood models for detecting positively selected amino acid sites
925 and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
926 http://www.genetics.org/content/148/3/929.abstract
927 101. Yang, Z., Wong, W. S. W., & Nielsen, R. Bayes empirical bayes inference of amino acid sites
928 under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005).
929 https://doi.org/10.1093/molbev/msi097
930 102. Zwiers, L.-H., & de Waard, M. A. Efficient Agrobacterium tumefaciens-mediated gene
931 disruption in the phytopathogen *Mycosphaerella graminicola*. *Curr. Genet.* **39**, 388–393 (2001).
932 https://doi.org/10.1007/s002940100216
933 103. Motteram, J., et al. Aberrant protein N-glycosylation impacts upon infection-related growth
934 transitions of the haploid plant-pathogenic fungus *Mycosphaerella graminicola*. *Mol. Microbiol.* **81**,
935 415–433 (2011). https://doi.org/10.1111/j.1365-2958.2011.07701.x
936