Adaptive evolution in virulence effectors of the rice blast fungus *Pyricularia oryzae*

Marie Le Naour--Vernet[1*], Florian Charriat[1*], Jérôme Gracy[2], Sandrine Cros-Arteil[1], Sébastien Ravel[1,3], Florian Veillet[1], Isabelle Meusnier[1], André Padilla[2], Thomas Kroj[1], Stella Cesari[1**], Pierre Gladieux[1**]

[1]PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

[2]Centre de Biologie Structurale (CBS), Univ Montpellier, INSERM, CNRS, 29 rue de Navacelles, 34090 Montpellier, France.

[3]CIRAD, UMR PHIM, 34090 Montpellier, France

* These authors contributed equally to this work

** Corresponding authors:

Pierre Gladieux

Tel: +33 4 99 62 48 17

Email: pierre.gladieux@inrae.fr

Stella Cesari

Tel: +33 4 99 62 48 62

Email: stella.cesari@inrae.fr

Short title: MAX effector evolution in the rice blast fungus

ABSTRACT

Plant pathogens secrete proteins called effectors that target host cellular processes to promote disease. Recently, structural genomics has identified several families of fungal effectors that share a similar three-dimensional structure despite remarkably variable amino-acid sequences and surface properties. To explore the selective forces that underlie the sequence variability of structurally-analogous effectors, we focused on MAX effectors, a structural family of effectors that are major determinants of virulence in the rice blast fungus *Pyricularia oryzae*. Using structure-informed gene annotation, we identified 58 to 78 MAX effector genes per genome in a set of 120 isolates representing seven host-associated lineages. The expression of MAX effector genes was primarily restricted to the early biotrophic phase of infection and strongly influenced by the host plant. Pangenome analyses of MAX effectors demonstrated extensive presence/absence polymorphism and identified gene loss events possibly involved in host range adaptation. However, gene knock-in experiments did not reveal a strong effect on virulence phenotypes suggesting that other evolutionary mechanisms are the main drivers of MAX effector losses. MAX effectors displayed high levels of standing variation and high rates of non-synonymous substitutions, pointing to widespread positive selection shaping the molecular diversity of MAX effectors. The combination of these analyses with structural data revealed that positive selection acts mostly on residues located in particular structural elements and at specific positions. Our work provides unique insights into the evolutionary history of an extended fungal effector family and opens up new research avenues to deepen our understanding of the molecular coevolutionary interactions of fungi with plant hosts.

AUTHOR SUMMARY

Fungal plant pathogens use small secreted proteins, called effectors, to manipulate to their own advantage their host's physiology and immunity. The evolution of these effectors, whether spontaneously or in response to human actions, can lead to epidemics or the emergence of new diseases. It is therefore crucial to understand the mechanisms underlying this evolution. In this article, we report on the evolution of effectors in one of the prime experimental model systems of plant pathology, the fungus causing blast diseases in rice, wheat, and other cereals or grasses. We identify a particular class of effectors, the MAX effectors, using structural models based on experimental protein structures of effectors previously shown to have a major role in fungal virulence. We show that this class of effector is produced by the pathogen during the early stages of infection, when plant cells are still alive. By comparing the gene content of isolates infecting different plant species, we show that the MAX effector arsenal is highly variable from one isolate to another. Finally, using the inferential framework of population genetics, we demonstrate that MAX effectors exhibit very high genetic variability and that this results from the action of natural selection.

68 **INTRODUCTION**

69 Plant pathogens secrete effector proteins to manipulate the physiology and metabolism of their
70 host and to suppress its immunity. Consequently, effectors are expected to engage in
71 coevolutionary interactions with plant defense molecules. The proximate mechanisms of
72 effector-driven adaptation are relatively well-characterized: plant pathogens adapt to new hosts
73 through changes in effector repertoires and effector sequences [1, 2]. However, the ultimate
74 (eco-evolutionary) mechanisms underlying effector diversification have remained elusive. The
75 concept of coevolution posits that adaptation in one partner drives counter-adaptations in the
76 coevolving partner [3-5]. Under the co-evolutionary arms race model, variation for disease
77 resistance and pathogen virulence is transient, resulting in a turnover of sequence variation
78 through repeated episodes of strong directional selection [6]. In agricultural systems, because
79 pathogens tend to be ahead of their hosts in the arms race owing to their larger populations and
80 shorter generation times, the co-evolutionary arms race tends to result in so-called boom and
81 bust cycles [7]. Under the alternative, 'trench warfare' hypothesis, advances and retreats of
82 resistance or virulence genes frequencies maintain variation as dynamic polymorphisms [8, 9].
83 The maintenance of genetic polymorphisms is called 'balancing selection', a process by which
84 different alleles or haplotypes are favored in different places (via population subdivision)
85 and/or different times (via frequency-dependent negative selection). While there is a growing
86 body of data demonstrating the nature and prevalence of the selective pressures that shape the
87 diversity of immune systems in plants [6, 10-12], we still lack a clear picture of the co-
88 evolutionary mechanisms underlying the molecular evolution of virulence factors in their
89 interacting antagonists [13].

90 Effectors from plant pathogenic fungi are typically cysteine-rich secreted proteins
91 smaller than 200 amino acids with an infection-specific expression pattern. Effectors are
92 numerous in fungal genomes (several hundred to more than a thousand per genome), and rarely
93 show homologies with known proteins or domains. They are also highly variable in sequence
94 and do not form large families of sequence homologs. Based on similarity analyses, fungal

effectors can form small groups of paralogs (typically with less than five members), but they are most often singletons. This apparent lack of larger effector families has hindered attempts to probe into the evolutionary factors underlying their diversification. In addition, the high diversity of fungal effectors has hampered functional analyses due to the lack of good criteria for prioritizing them and our inability to predict their physiological role. Consequently, the virulence function and evolutionary history of most fungal effectors remain unknown.

Recently, the resolution of the three-dimensional (3D) structure of fungal effectors combined with Hidden Markov Model (HMM) pattern searches and structure modeling revealed that fungal effector repertoires are, despite their hyper-variability, dominated by a limited number of families gathering highly sequence-diverse proteins with shared structures and, presumably, common ancestry [14-17]. One such structurally-conserved but sequence-diverse fungal effector family is the MAX (*Magnaporthe* Avrs and ToxB-like) effector family. MAX effectors are specific to ascomycete fungi and show massive expansion in *Pyricularia oryzae* (synonym: *Magnaporthe oryzae*) [17], the fungus causing rice blast disease, one of the most damaging diseases of rice [18, 19]. MAX effectors are characterized by a conserved structure composed of six β-strands organized into two antiparallel β-sheets that are stabilized in most cases by one or two disulfide bridges. The amino acid sequence of MAX effectors is very diverse and they generally have less than 15% identity, which makes them a family of analogous, not homologous effectors. MAX effectors are massively expressed during the biotrophic phase of infection, suggesting an important role in disease development and fungal virulence [17]. Remarkably, about 50% of the known avirulence (AVR) effectors of *P. oryzae* belong to the MAX family, indicating that these effectors are closely monitored by the host plant immune system [17].

Pyricularia oryzae is a multi-host, poly-specialist pathogen that infects more than 50 monocotyledonous plants, including major cereal crops such as rice, maize, wheat, or barley [20-23]. *Pyricularia oryzae* has repeatedly emerged on new hosts [21, 24], in new geographical areas [25, 26], and phylogenomic analyses have revealed that it can be subdivided into several genetic

122 lineages, each preferentially associated with a specific or restricted set of host plant genera [27].
123 In *P. oryzae*, effectors can play a major role in host-shifts or host-range expansions [28-30]. For
124 example, loss of function of the PWT3 effector in *Lolium*-infecting strains contributed to gain of
125 virulence on wheat [29]. Similarly, loss of the MAX effector AVR1-CO39 is thought to have
126 contributed to the emergence of rice blast from foxtail-millet infecting isolates [20, 31]. This
127 indicates that MAX effectors may be important determinants of host specificity in *P. oryzae*.

128 In this study, we characterized the genetic diversity of MAX effectors in *P. oryzae* and
129 within its different host-specific lineages. We explored the evolutionary drivers of the
130 diversification of MAX effectors and tested whether MAX effectors represent important
131 determinants of *P. oryzae* host specificity. To this aim, we assembled and annotated 120 high-
132 quality *P. oryzae* genomes from isolates representing seven main host-specific lineages. We
133 mined these genomes for putative effectors and used hidden Markov models based on fold-
134 informed protein alignments to annotate putative MAX effectors. We identified 58 to 78 putative
135 MAX effector genes per individual genome distributed in 80 different groups of MAX homologs.
136 We showed that the expression of *MAX* effector genes is largely restricted to the early biotrophic
137 phase of infection and strongly influenced by the host plant. Our evolutionary analyses showed
138 that MAX effectors harbor more standing genetic variation than other secreted proteins and
139 non-effector genes, and high rates of non-synonymous substitutions, pointing to positive
140 selection as a potent evolutionary force shaping their sequence diversity. Pangenome analyses of
141 MAX effectors demonstrated extensive presence/absence polymorphism and identified several
142 candidate gene loss events possibly involved in host range adaptation. Our work demonstrates
143 that MAX effectors represent a highly dynamic compartment of the genome of *P. oryzae*, likely
144 reflecting intense co-evolutionary interactions with host molecules.

145 **RESULTS**

146 *Genome assembly and prediction of MAX effector genes.*

147 We assembled the genomes of a worldwide collection of 120 haploid isolates of *Pyricularia*
148 *oryzae* fungi from 14 host genera: *Oryza* (n=52), *Triticum* (n=21), *Lolium* (n=12), *Setaria*
149 (n=8), *Eleusine* (n=8), *Echinochloa* (n=4), *Zea* (n=4), *Bromus* (n=2), *Brachiaria* (n=2),
150 *Festuca* (n=2), *Stenotaphrum* (n=2), *Eragrostis* (n=1), *Hordeum* (n=1), and *Avena* (n=1) (S1
151 Table). Assembly size ranged from 37Mb to 43.2Mb, with an average size of 40.2 Mb (standard
152 deviation [s.d.]: 1.9Mb). L50 ranged from five to 411 contigs (mean: 97.1; s.d.: 83.2) and N50
153 from 28Kb to 4.0Mb (mean: 238.6Kb; s.d.: 43.8Kb; S1 Table). Gene prediction based on protein
154 sequences from reference 70-15 and RNAseq data identified 11,520 to 12,055 genes per isolate
155 (mean: 11,763.2; s.d.: 103.7). The completeness of assemblies, as estimated using BUSCO [32],
156 ranged between 93.4 and 97.0% (mean: 96.4%; s.d.: 0.6%; S1 Table).

157 MAX effectors were identified among predicted secreted proteins using a combination of
158 similarity searches [33, 34] and structure-guided alignments [35] as summarized in Figure 1. To
159 assess variation in the MAX effector content of *P. oryzae*, we constructed groups of homologous
160 genes (i.e., "orthogroups" or OG) using the clustering algorithm implemented in ORTHOFINDER
161 [36]. A given orthogroup was classified as secreted proteins or MAX effectors if 10% of
162 sequences in the group were identified as such by functional annotation. Sequences were
163 grouped in 14,767 orthogroups, of which 80 were classified as encoding MAX effectors, and
164 3,283 as encoding other types of secreted proteins (Figure 1). The number of MAX orthogroups
165 per isolate ranged from 58 to 73 (average: 65.8; s.d.: 2.8), representing between 58 to 78 *MAX*
166 genes per isolate (average: 68.4; s.d.: 3.6). The 80 orthogroups of MAX effectors were further
167 split into 94 groups of orthologs, by identifying paralogs using gene genealogies inferred with
168 RAXML v8 [37] (S2 Table).

169

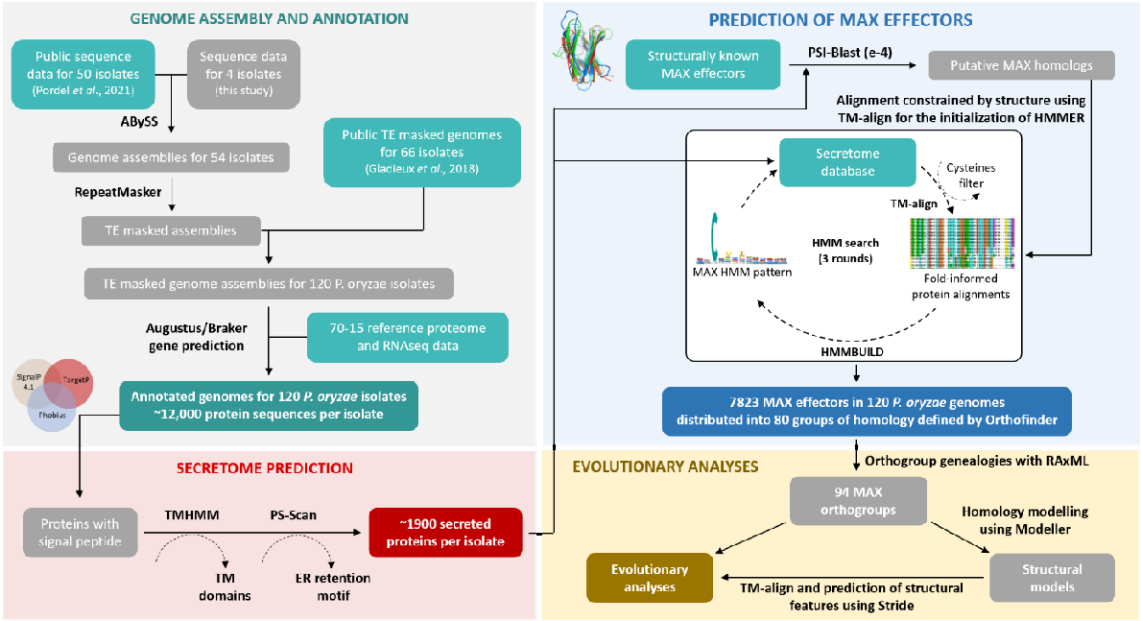

Figure 1. Schematic representation of the key steps of the bioinformatic pipeline used to predict genes in 120 genomes of *Pyricularia oryzae*, and to identify genes encoding MAX effectors. References: Gladieux et al. 2018 [27]; Pordel et al. 2021 [21].

Expression of MAX effector genes during rice infection.

To determine whether these putative MAX effectors are deployed by *P. oryzae* during plant infection, we analyzed the expression patterns of the 67 *MAX* genes predicted in the genome of the reference isolate Guy11 by qRT-PCR (Figure 2A). Using RNA samples from Guy11 mycelium grown on artificial media, we found that 94% of the *MAX* genes (63 genes) were not, or very weakly expressed during axenic culture, and only four (i.e., *MAX24*, *MAX29*, *MAX59*, and *MAX66*) showed weak, medium or strong constitutive expression (Figure 2A). *MAX* genes were, therefore, predominantly repressed in the mycelium of *P. oryzae*.

Following spray inoculation of Guy11 on the rice cultivar Maratelli, which is highly susceptible to *P. oryzae*, 67% of the *MAX* genes (45 genes) were expressed (Figure 2A). Among them, three were also expressed in the mycelium (i.e., *MAX31*, *MAX59*, and *MAX94*). *MAX31* was over-expressed under infection conditions, whereas the other two showed similar levels of expression *in vitro* and during infection. 64% of the *MAX* genes (42 genes) showed an infection-specific expression profile with relative expression levels ranging from very low (0.008-0.04) to

very high (>5). Like the *Bas3* gene, encoding a *P. oryzae* effector specifically induced during the biotrophic phase of infection, all *MAX* genes showed maximal expression between the second and fourth day post-inoculation (Figure 2A; S1 Figure).

To test whether the genotype of the host plant could influence the expression of *MAX* genes, we analyzed their expression patterns upon infection of the rice cultivar Kitaake, which has a higher basal resistance to *P. oryzae* than Maratelli (Figure 2A; Figure 2B). During Kitaake infection, 78% of the *MAX* genes (52 genes) were upregulated compared to the *in vitro* condition, while only 64% (43 genes) were induced upon infection of Maratelli (Figure 2A). Some *MAX* genes not expressed in Maratelli were induced in Kitaake (e.g., *MAX24, MAX30, MAX32, MAX43, MAX71,* and *MAX73*) (Figure 2B, S2 Figure). Others were significantly upregulated in Kitaake compared to Maratelli (i.e., *MAX22, MAX44, MAX55, MAX57, MAX69*, and *MAX91*). However, a few genes, such as *MAX15, MAX37* and *MAX62*, among the most strongly expressed effectors in Maratelli, showed weaker levels of expression in Kitaake. These results show that Guy11 deploys a wider diversity of MAX effectors during the infection of Kitaake compared to that of Maratelli, and that MAX effectors are subject to host-dependent expression polymorphism.

Taken together, our data revealed that during the biotrophic phase of rice infection, *P. oryzae* massively deploys MAX effectors in a plastic manner suggesting that they have an important function in fungal virulence.

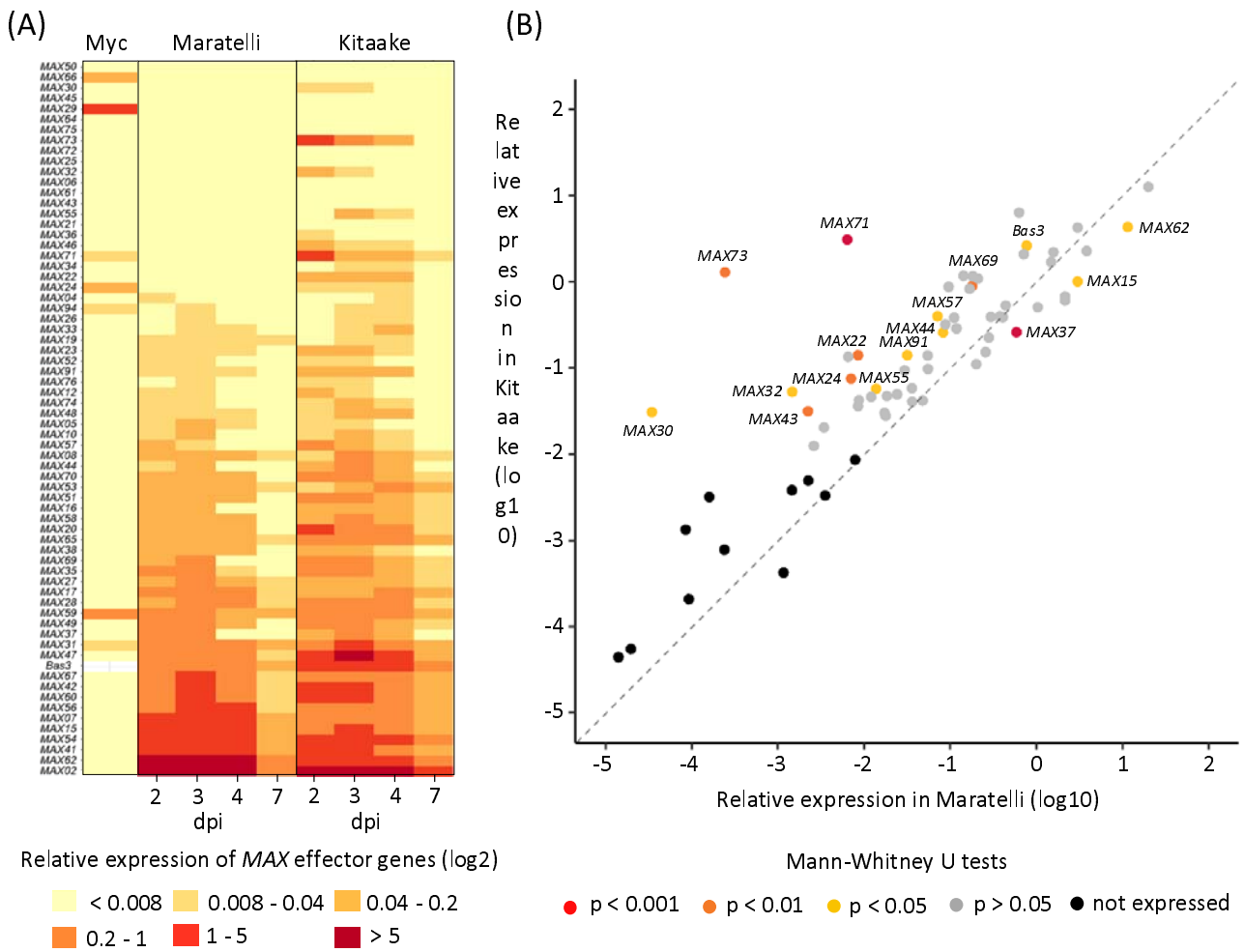

Figure 2. The expression of *MAX* genes is biotrophy-specific and is influenced by the host plant. (A) Transcript levels of *MAX* genes and the biotrophy marker gene *Bas3* were determined by qRT-PCR in the mycelium (Myc) of the *P. oryzae* isolate Guy11 grown for 5 days in liquid culture, and in infected leaves of the rice cultivars Maratelli and Kitaake at 2, 3, 4, and 7 days post inoculation (dpi) with Guy11. Relative expression levels were calculated using the constitutively expressed *MoEF1α* (*Elongation Factor 1α*) gene as a reference. The heatmap shows the median relative expression value for each gene (in log2 scale), calculated from 6 independent biological samples for the Myc condition, and 3 independent inoculation experiments (each with 5 independent leaf samples per time point) for each rice cultivar. Effectors were ranked from top to bottom by increasing relative expression values in Maratelli. Relative expression values were assigned to six categories: not expressed (<0.008), very weakly (0.008-0.04), weakly (0.04-0.2), moderately (0.2-1), strongly (1-5) and very strongly expressed (>5). (B) Scatter plot comparing the relative expression levels of MAX genes in Guy11-infected Maratelli and Kitaake cultivars. Each point shows the maximum median relative expression value (in log10 scale) calculated in the infection kinetics described in (A). Difference in effector relative expression levels between the two conditions was assessed by Mann-Whitney U tests and dots were

224 colored according to significance results: grey (p>0.05), yellow (p<0.05), orange (p<0.001), red (p<0.0001), black
225 (effectors not expressed in both conditions).
226
227
228 *Variability of the MAX effector repertoire.*
229 To investigate the genetic diversity of MAX effectors in *P. oryzae*, we analyzed their nucleotide
230 diversity per base pair (π), their ratio of non-synonymous to synonymous nucleotide diversity
231 (π_N/π_S), and their presence-absence polymorphism. Compared to other secreted proteins or
232 other genes, MAX effector orthogroups had higher π, and π_N/π_S values, and lower presence
233 frequency (S3 Figure). Orthogroups including known avirulence genes like *AVR1-CO39*, *AvrPiz-t*
234 and *AVR-Pik* featured among the most diverse orthogroups of MAX effectors (S1 Data).
235 We categorized genes in the pangenome according to their presence frequencies [38],
236 with core genes present in all isolates, softcore genes present in >99% isolates, cloud genes
237 present in 1-99% isolates and shell genes present in <1% isolates. The majority of MAX effector
238 genes were classified as cloud (64/80 [80%] orthogroups), while the majority of other secreted
239 proteins or other genes were classified as core or softcore (1650/3283 [50.2%] and
240 6714/11404 [58.9%] orthogroups, respectively) (Figure 3A). Only a minority of genes were
241 present in multiple copies (MAX: 15/80 [18.8%]; other effectors: 746/3283 [22.7%]; other
242 genes: 1439/11404 [12.6%]; Figure 3A). Assessment of the openness of the pan-genome by
243 iteratively subsampling isolates revealed a closed pangenome with a limited number of pan and
244 core genes for MAX effectors, other secreted proteins and the remainder of the gene space
245 (Figure 3B). Nucleotide diversity differed significantly between categories of the pangenome for
246 non-MAX effectors (Kruskal-Wallis test: H=181.17, d.f.=2, p<0.001) and other genes (Kruskal-
247 Wallis test: H=225.25, d.f.=2, p<0.001), but not for MAX effectors (Kruskal-Wallis test: H=2.50,
248 d.f.=2, p>0.05). For non-MAX effectors and other genes, nucleotide diversity π was significantly
249 higher in the cloud genes than in softcore genes and core genes (Post-hoc Mann-Whitney U-
250 tests, p<0.001; Figure 3C).

Together, these analyses show that the MAX effector repertoire is highly plastic compared to other gene categories, both in terms of the presence/absence of orthogroups and the sequence variability within orthogroups.

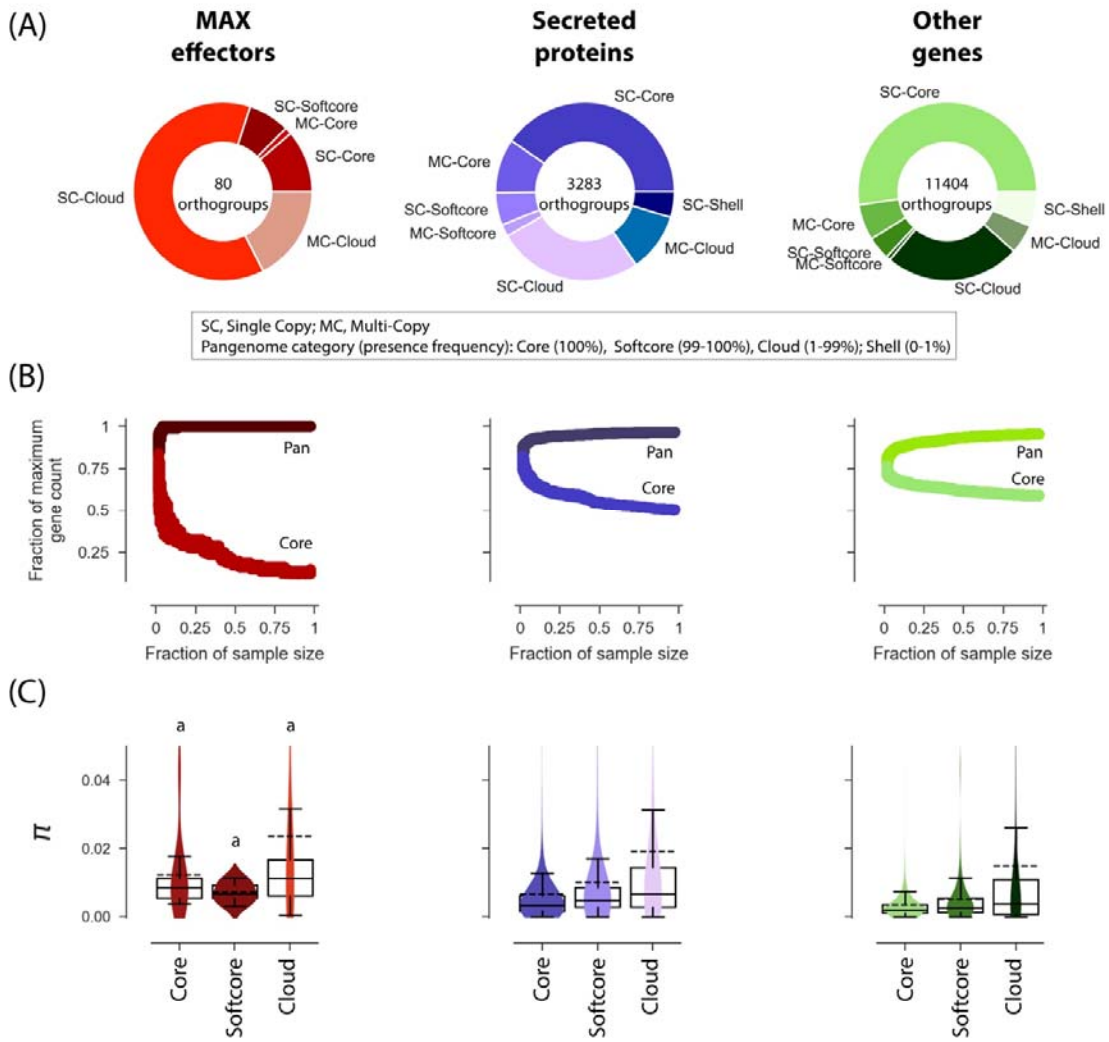

Figure 3. The pan-genome of *P. oryzae*. (A) Composition of the pangenome of MAX effectors, other secreted proteins, and other genes. (B) Rarefaction analysis of the size of pan- and core-genomes. For k ranging from two to the sample size minus one, pan- and core-genome sizes were computed for 1000 random combinations of k isolates. Subsample size is represented as a fraction of the sample size (n=121), and pan- and core-genome sizes as a fraction of maximum gene counts (reported at the center of donut plots in panel A). "core" genes are present in all isolates of a pseudo-sample of size k; "pan" qualifies genes that are not "core". (C) Nucleotide diversity per base pair (π) in core, softcore, and cloud genes. A number of data points were cropped from the nucleotide diversity plot for visually optimal presentation but included in statistical tests. In box plots, the dashed black line is the mean, the solid black line is the

median. Shell genes were not included in the nucleotide diversity plot because it was not computable due to the small sample size or lack of sequence after filtering for missing data. Shared superscripts indicate non-significant differences (Post-hoc Mann-Whitney U-tests).

268 *Population subdivision.*

269 To investigate signatures of positive selection in the genome of *P. oryzae*, and identify candidate
270 loci involved in host specificity, we first identified the divergent lineages represented in our
271 dataset. We inferred population structure from 5.04e5 SNPs in single-copy core orthologs, using
272 complementary approaches that make no assumption about random mating or linkage
273 equilibrium. Both clustering analyses with the sNMF software [38] (Figure 4A) and neighbor-net
274 phylogenetic networks [39] (Figure 4B) revealed consistent patterns that split genetic variation
275 primarily by host of origin, with seven major lineages mainly associated with rice (*Oryza*),
276 foxtail millet (*Setaria*), wheat (*Triticum*), ray-grass (*Lolium*), goosegrass (*Eleusine*), barnyard
277 grass and maize (*Echinochloa* and *Zea*), and St. Augustine grass (*Stenotaphrum*).

278 Population subdivision inferred from MAX effectors using either 130 SNPs without
279 missing data in single-copy core MAX effectors (Figure 4C) or presence/absence variation of all
280 80 MAX effector orthogroups (Figure 4D) revealed essentially the same groups as the analysis of
281 the single-copy core orthologs. This indicates that genome-wide nucleotide variation, variation
282 in MAX effector content, and nucleotide variation at MAX effectors reflected similar genealogical
283 processes. The *Oryza* and *Setaria* lineages displayed exceptionally high presence/absence
284 variation of MAX effectors (average Hamming distance between pairs of isolate: 0.123 and
285 0.095; Figure 4D), but only limited sequence variation at single copy core MAX effectors
286 (average Hamming distance between pairs of isolate: 0.017 and 0.012; Mann-Whitney U-tests,
287 $p<0.05$; Figure 4C).

288 Population subdivision was also apparent at the level of the global characteristics of
289 genomes. Assembly size differed significantly among lineages (Kruskal-Wallis test: H=72.9,
290 d.f.=4, p<0.0001), and genome assemblies for the *Oryza*-infecting lineage, in particular, were
291 significantly shorter than assemblies from other groups (Post-hoc Mann-Whitney U-tests,
292 p<0.001; S4 Figure). The number of predicted genes, which was positively and significantly
293 correlated with assembly size (Spearman's ρ=0.31, $p<0.001$), also differed significantly among
294 lineages and was the highest in the *Lolium*-infecting lineage and the lowest in the *Eleusine*-

infecting lineage (Kruskal-Wallis test: H=19.4, d.f.=4, p<0.0001; S4 Figure). The distribution of the number of orthogroups in lineages was largely similar to the distribution of the number of genes, with more orthogroups in the *Lolium*-infecting lineage and fewer in the *Eleusine*-infecting lineage (S4 Figure).

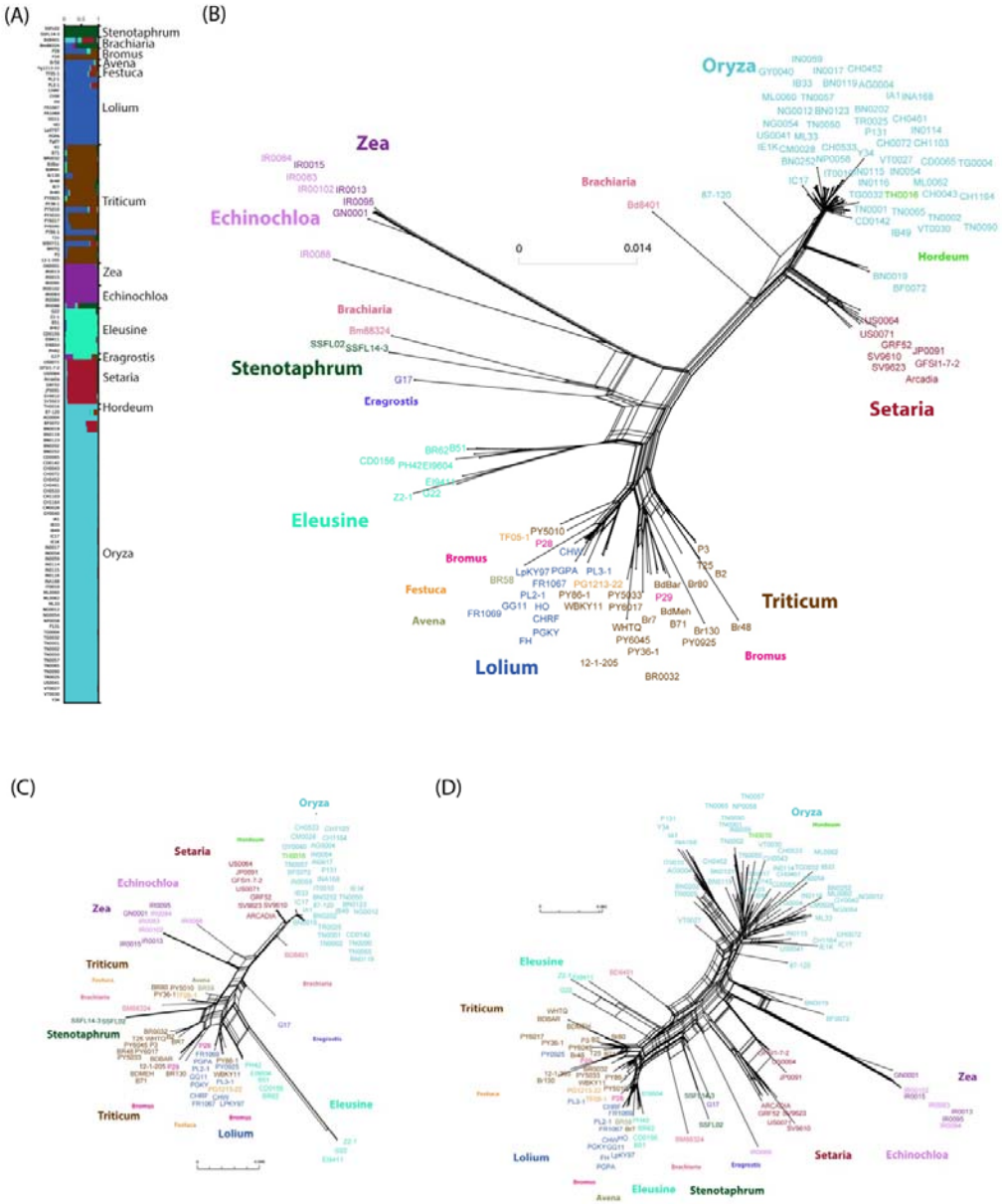

Figure 4. Population subdivision in 120 isolates of *Magnaporthe oryzae*. Population subdivision was inferred from (A-B) 5.04e5 single nucleotide polymorphisms (SNPs) without missing data identified in coding sequences of single-copy core orthologs, (C) 130 SNPs without missing data identified in coding sequences of single-copy core MAX effectors, (D) a table of presence/absence data (coded as 0 and 1) for all 80 MAX effector orthogroups. (A) Genetic ancestry

304 proportions for individual isolates in K=7 ancestral populations, as estimated using the sNMF clustering algorithm
305 [38]; each isolate is depicted as a horizontal bar divided into K segments representing the proportion of ancestry in
306 K=7 inferred ancestral populations; host genus of origin is indicated on the right side. (B-D) Neighbor-net
307 phylogenetic networks estimated with SplitsTree [39], with isolate names colored according to their host of origin. In
308 (C), haplotypes repeated multiple times were represented by the first items in the following lists: (BN0252, CD0065,
309 CH0043, CH0072, CH0452, CH0461, IN0114, IN0115, IN0116), (IB33, ML0060, ML0062, ML33, NG0054, NP0058,
310 TG0004, TG0032, US0041, VT0027, VT0030), (TR0025, Y34), (CHRF, HO, LPKY97, P28).

311
312 *Loss of MAX effectors in specific lineages does not necessarily associate with host specificity.*

313 The comparison of the MAX effector content in the genomes of 120 *P. oryzae* isolates revealed
314 extensive presence/absence polymorphism between host-specific groups (S3 Table). To address
315 the underlying evolutionary mechanisms, we tested experimentally the hypothesis that MAX
316 effector losses are massively related to escape from receptor-mediated non-host resistance.
317 Indeed, the loss of MAX effectors in specific lineages of *P. oryzae* could primarily serve to escape
318 from non-host resistance during infections of novel plant species carrying immune receptors
319 specifically recognizing these effectors. To test this hypothesis, we focused on the *Oryza*- and
320 *Setaria*-infecting lineages, as previous investigations suggested that the *Oryza*-infecting lineage
321 emerged by a host shift from *Setaria* and we found both groups to be closely related (Figure 4)
322 [21, 40]. Our strategy was to introduce into the *Oryza*-isolate Guy11 MAX effectors absent from
323 the *Oryza* lineage but present in the *Setaria* lineage, and to assess the ability of these transgenic
324 isolates to infect rice.

325 We identified three MAX orthogroups that were largely or completely absent from the
326 *Oryza* lineage, but present in the majority of isolates of the other lineages (S3 Table).
327 Orthogroup MAX79 (OG0011591-1) was absent in all 52 *Oryza*-infecting isolates, while MAX83
328 (OG0011907), and MAX89 (OG0012141) were absent in 50 and 46 of them, respectively (S3
329 Table). Constructs carrying the genomic sequence of *MAX79*, *MAX83* or *MAX89* derived from the
330 *Setaria* isolate US0071 and under the control of the strong infection specific promoter of the
331 effector *AVR-Pia* were generated and stably introduced into Guy11. For each construct, three
332 independent transgenic lines were selected. Transgene insertion was verified by PCR and the

333 expression of transgenes was measured by qRT-PCR (S5 Figure). To test whether the selected
334 MAX effectors trigger immunity in rice, the transgenic isolates were spray-inoculated onto a
335 panel of 22 cultivars representative of the worldwide diversity of rice (S4 Table).

336 As controls, we used the MAX effectors *AVR-Pia*, which is rare outside the *Oryza* and
337 *Setaria* lineages, and *AVR1-CO39*, which is absent or pseudogenized in the *Oryza* lineage, but
338 present in all other host-specific lineages including *Setaria*. Both effectors are detected in rice by
339 the paired NLR immune receptors RGA4 and RGA5 from the *Pi-a/Pi-CO39* locus and thereby
340 contribute, respectively, to host or non-host resistance in this plant species [41, 42].

341 As expected, isolates expressing *AVR1-CO39* or *AVR-Pia* triggered resistance in the rice
342 variety Aichi Asahi that carries *Pi-a*, but caused disease on Nipponbare (*pi-a*$^-$) and other varieties
343 lacking this *R* locus (Figure 5; S4 Table). Unlike the positive controls, the effectors MAX79,
344 MAX83 and MAX89 were not recognized and did not induce resistance in any of the tested rice
345 cultivars (Figure 5; S4 Table). The disease symptoms caused by the transgenic isolates carrying
346 these effectors were similar to those observed for wild-type Guy11 or Guy11 isolates carrying an
347 *RFP* (red fluorescent protein) construct. This suggests that these effectors do not significantly
348 increase the virulence of Guy11.

349 These experiments show that despite their loss in the *Oryza*-infecting lineage of *P.*
350 *oryzae,* and unlike AVR1-CO39, the effectors MAX79, MAX83, and MAX89 do not seem to induce
351 non-host resistance in rice. Consequently, other mechanisms than escape from host immunity
352 contributed to the loss of these MAX effectors during the putative host shift of *P. oryzae* from
353 *Setaria* to *Oryza*.
354

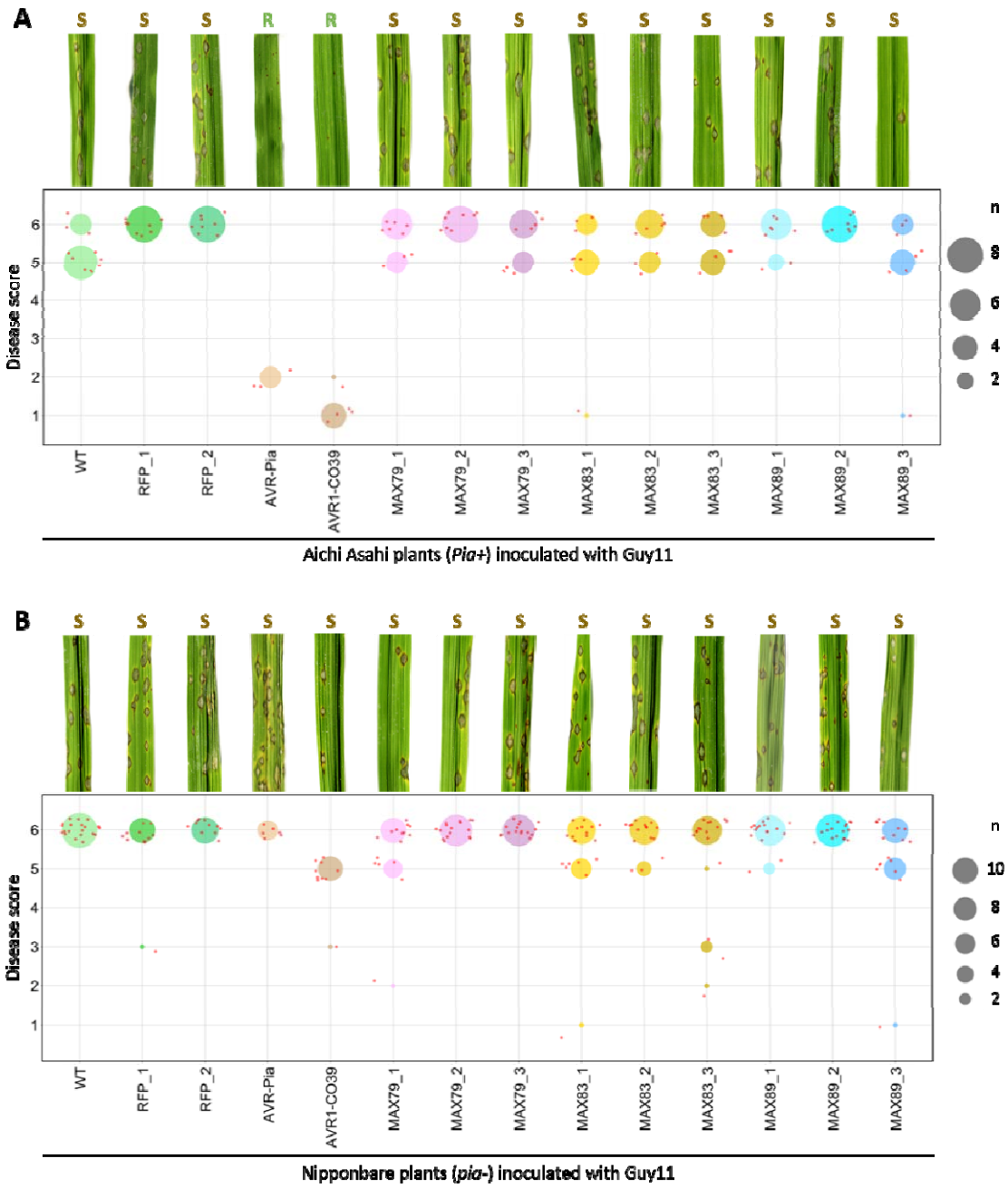

Figure 5. AVR1-CO39 contributes to non-host specificity in rice but not MAX79, MAX83 or MAX89. Wild type and transgenic isolates of *P. oryzae* Guy11 expressing the *RFP* (*red fluorescent protein*), *AVR-Pia, AVR1-CO39, MAX79*, *MAX83* or *MAX89* gene were spray-inoculated at 40 000 spores/ml on three-week-old rice plants of the cultivars Aichi Asahi (A) and Nipponbare (B). For each condition, representative disease phenotypes on rice leaves at seven days post-inoculation are shown (top panels, R: resistance, S: susceptibility). Disease phenotypes were also scored (from 1 [complete resistance] to 6 [high susceptibility]) on leaves from 3 to 5 individual rice plants and data are shown as dot plots (bottom panels). The size of each circle is proportional to the number of replicates (n) matching the corresponding score for each condition. Small red dots correspond to individual measurements. The experiment was

363 performed twice for Aichi Asahi and four times for Nipponbare for all isolates except for WT, AVR-Pia, and AVR1-
364 CO39 control isolates. For these isolates, experiments were performed once on Aichi Asahi and twice on Nipponbare
365 because disease phenotypes are well characterized in the literature.
366
367 *Signatures of balancing selection at MAX effectors.*
368 To investigate the impact of balancing selection on MAX effector evolution, we focused on single-
369 copy core, softcore, and cloud orthogroups to avoid the possible effect of gene paralogy. We then
370 computed π (nucleotide diversity per bp), F_{ST} (the amount of differentiation among lineages
371 [43]), π_N (non-synonymous nucleotide diversity), π_S (synonymous nucleotide diversity), and
372 π_N/π_S (the ratio of non-synonymous to synonymous nucleotide diversity). Large values of π and
373 π_N/π_S, in particular, are possible indicators of a gene being under balancing selection.
374 Nucleotide diversity (π) differed significantly between groups of genes (Kruskal-Wallis
375 test, H=509.9, d.f.=2, p <0.001; Figure 6A; S5 Table). π was significantly higher for the set of
376 MAX effectors (average π: 0.0104, standard deviation: 0.0137), than for other secreted proteins
377 (average π: 0.0079, standard deviation: 0.020), and other genes (average π: 0.0049, standard
378 deviation: 0.014; Mann-Whitney U-tests, p<0.05), showing that MAX effectors, and to a smaller
379 extent other secreted proteins, are more variable than a typical gene. At the lineage level,
380 however, nucleotide diversity at MAX effectors tended to not significantly differ from other
381 putative effectors, or other genes (S5 Table).
382 In addition to having greater nucleotide variation than other genes at the species level,
383 MAX effectors also displayed a higher ratio of non-synonymous to synonymous nucleotide
384 diversity (Figure 6B; S5 Table). The π_N/π_S ratio differed significantly between groups of genes
385 (Kruskal-Wallis tests H=101.4, $d.f.$=2, p<0.001), and the excess of non-synonymous diversity
386 was significantly, and markedly, higher for MAX effectors (average π_N/π_S: 1.826, standard
387 deviation: 3.847) than for other effectors (average π_N/π_S: 0.461, standard deviation: 1.600), and
388 other genes (average π_N/π_S: 0.448, standard deviation: 1.463; Mann-Whitney U-tests, p<0.05).
389 The higher π_N/π_S of MAX effectors was mostly driven by differences in π_N (Figure 6A; S5 Table).
390 Twenty MAX effectors displayed values in the top 5% percentile of non-effector genes, far

391 exceeding the four genes expected by chance ($p<0.05$). More specifically, 26 MAX effectors
392 displayed π_N/π_S values greater than 1, which is the value expected under neutrality. This
393 included three well-known avirulence genes: *AVR1-CO39* ($\pi_N/\pi_S=2.564$), *AVR-Pik*
394 ($\pi_N/\pi_S=15.574$), and *AvrPiz-t* ($\pi_N/\pi_S=1.431$). The average π_N/π_S ratio was also higher at MAX
395 effectors than other secreted proteins and other genes in all lineages, with significant differences
396 in four lineages (Mann-Whitney U-tests, $p<0.05$), and the average π_N/π_S was greater than one in
397 the *Oryza* lineage (Figure 6B; S5 Table). Seven to eleven MAX effectors had $\pi_N/\pi s>1$ at the
398 lineage level, representing 8% (*Setaria*-infecting lineage) to 41% (*Lolium*-infecting lineage) of
399 MAX effectors with a defined $\pi_N/\pi s$ ratio (*Zea*/*Echinochloa* lineage excluded, as only three MAX
400 effectors had a defined $\pi_N/\pi s$; S2 Data).

401 $\pi_N/\pi_S>1$ is a strong indication of multiallelic balancing selection (*i.e.*, multiple alleles at
402 multiple sites are balanced), as single sites under very strong balancing selection cannot
403 contribute enough non-synonymous variability to push the π_N/π_S ratio above one [44]. To assess
404 whether the adaptation of lineages to their respective hosts may contribute to the species-wide
405 excess of non-synonymous diversity detected at MAX effectors, we estimated population
406 differentiation. The differentiation statistic F_{ST} differed significantly between groups of genes
407 (Kruskal-Wallis tests $H=8.731$, $d.f.=2$, $p=0.013$), and differentiation was significantly higher for
408 MAX effectors than for other secreted proteins and other genes (Figure 6A; S5 Table). F_{ST} was
409 also significantly, albeit relatively weakly, correlated with π_N/π_S at MAX effectors (Spearman's ρ:
410 0.304, $p=0.007$; S6 Figure). These observations indicate that between-lineages differences in
411 allele frequencies are greater for MAX effectors than for other secreted proteins or other genes,
412 which may result from divergent selection exerted by hosts.
413
414

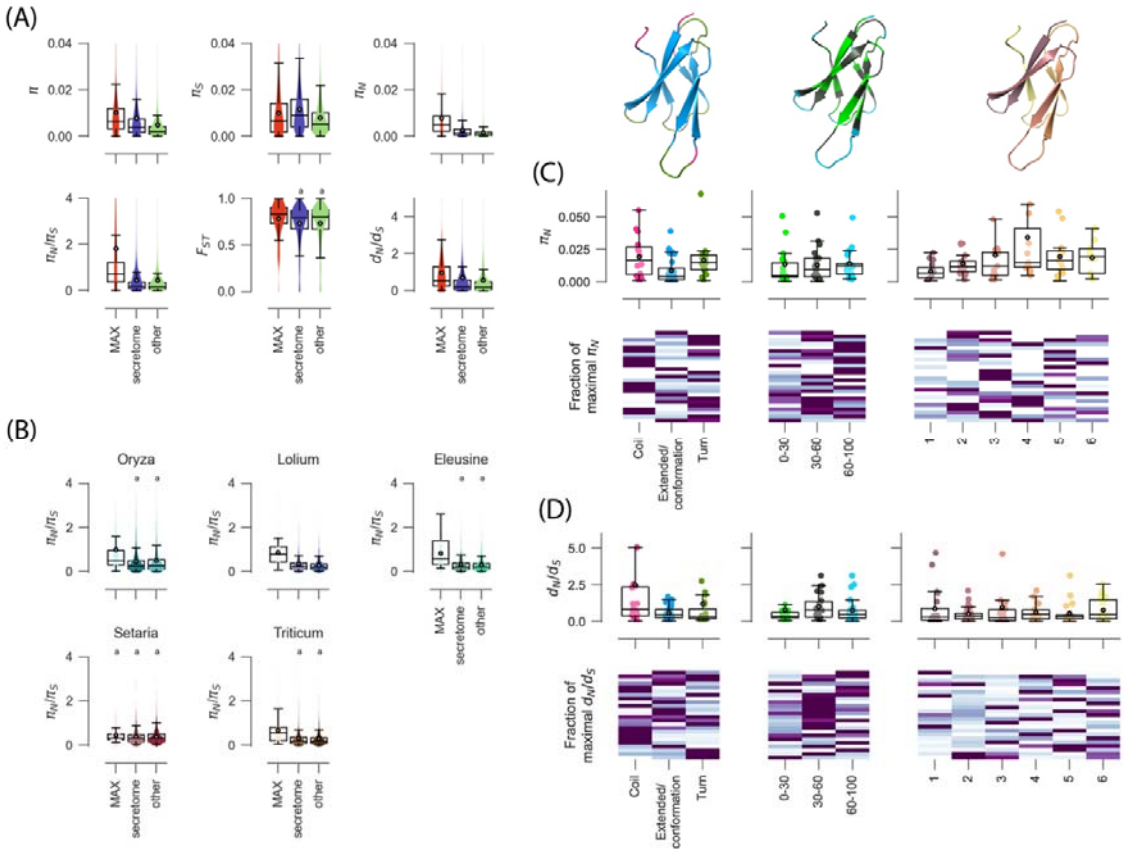

Figure 6. Summary statistics of polymorphism and divergence at MAX effectors, other secreted proteins (i.e., secretome), and other genes of *P. oryzae*. (A) Species-wide estimates of π (nucleotide diversity per bp), F_{ST} (the amount of differentiation among lineages), π_N (non-synonymous nucleotide diversity per bp), π_S (synonymous nucleotide diversity per bp), π_N/π_S (the ratio of non-synonymous to synonymous nucleotide diversity), d_N/d_S (the ratio of non-synonymous to synonymous rates of substitutions). (B) Lineage-specific estimates of π_N/π_S. (C) and (D) Species-wide estimates of π_N and d_N/d_S computed at MAX effectors with signatures of balancing selection (π_N/π_S>1; panel C) and signatures of directional selection (d_N/d_S>1; panel D) for different classes of secondary structure annotations highlighted on the three-dimensional structure of *AVR-Piz-t* above panel C: (i) structural features, with three subclasses: "extended conformation", "coils", and "turns"; (ii) solvent accessibility percentage of the Van der Waals surface of the amino acid sidechain, with three sub-classes: 0-30% (buried), 30-60% (intermediate), and 60-100% (exposed); (iii) structural domains, with six subclasses that grouped the coil, extended conformation and turn residues that define the six beta strands characteristic of MAX effectors. In the heatmaps, each line represents a MAX effector. For a given MAX effector and a given structural feature, the darkest color indicates the class of the secondary structure annotation for which the summary statistic is the highest. Only single-copy core, softcore, and cloud groups of orthologous genes were included in calculations. Shared superscripts indicate non-significant differences (post-hoc Mann-Whitney U-tests, p>0.05). A number of data points were cropped from plots in (A) and (B) for visually optimal presentation but included in statistical tests. In box plots, the black circle is the mean, the black line is the median.

436

437 *Signatures of recurrent directional selection at MAX effectors*

438 To detect adaptive molecular evolution, we collected orthologous sequences from outgroup
439 *Pyricularia sp.* LS [45, 46] and estimated the d_N/d_S ratio (the ratio of non-synonymous to
440 synonymous substitution rates) using a maximum likelihood method [47]. Outgroup sequences
441 could be retrieved for 10,174 out of 14,664 single-copy orthogroups, including 66 out of 94
442 single-copy orthologs of MAX effectors. The d_N/d_S ratio differed significantly between groups of
443 genes (Kruskal-Wallis tests H=45.812, $d.f.$=2, p<0.001; Figure 6A; S5 Table), and was higher for
444 MAX effectors (average d_N/d_S: 0.977, s.d.: 1.316) than for other secreted proteins (average
445 d_N/d_S: 0.711, s.d.: 1.722), and other genes (average d_N/d_S: 0.584, s.d.: 1.584; Mann-Whitney U-
446 tests, p<0.05). The same pattern of higher d_N/d_S for MAX effectors was observed at the lineage
447 level (S5 Table). Twenty-four of the 66 MAX effectors with outgroup sequence (i.e., 36.4%)
448 showed d_N/d_S>1 (S2 Data), which is a strong indication of directional selection. d_N/d_S>1 is only
449 expected for genes that have experienced repeated bouts of directional selection which led to
450 repeated fixations of amino-acid substitutions [44]. Eleven MAX effectors displayed signatures of
451 both multiallelic balancing selection (π_N/π_S>1) and multiallelic directional selection (d_N/d_S>1).

452 The divergence data, therefore, indicate that a scenario of molecular co-evolution
453 involving repeated selective sweeps may apply to a substantial fraction (at least one-third) of
454 MAX effectors.

455

456 *Structural determinants of polymorphism and divergence at MAX effectors.*

457 Different parts of proteins can be under different selective forces. To investigate the
458 relationships between the structural properties of MAX effectors and signatures of balancing or
459 directional selection, we focused on MAX effectors with d_N/d_S and π_N/π_S ratios >1 and we
460 computed π_N and d_N/d_S for different classes of secondary structure annotations, as computed by
461 STRIDE from MAX effector structures predicted by homology modeling for each orthologous
462 group [48] (S7 Figure). We used π_N and not π_N/π_S because the latter tended to be undefined due

463 to relatively short sequence lengths. We defined three classes of secondary structure
464 annotations: (1) structural features, with three subclasses: "extended conformation", "coils", and
465 "turns"; (2) solvent accessibility percentage of the Van der Waals surface of the amino acid
466 sidechain, with three sub-classes: 0-30% (buried), 30-60% (intermediate), and 60-100%
467 (exposed); (3) structural domains, with six subclasses that grouped the coil, extended
468 conformation and turn residues that define the six beta strands characteristic of MAX effectors.
469 We used the structure-guided alignment generated by TM-ALIGN to extract the secondary
470 structure annotations of each amino acid of the MAX effector sequences (S7 Figure).

471 We found that π_N significantly differed (Kruskal-Wallis tests H=8.504, $d.f.$=2, p=0.014)
472 between subclasses of structural features (Figure 6C; S6 Table). In the 25 MAX effectors under
473 multiallelic balancing selection that passed our filter on sequence length (at least ten sequences
474 longer than 10bp, see Methods), π_N was significantly higher at coils and turns, than at extended
475 conformations (coils: π_N=0.0191; turns: π_N=0.0167; extended conformations: π_N=0.0086;
476 posthoc Mann-Whitney U-tests, $p<0.05$), and 10 and 9 MAX effectors displayed their highest
477 values of π_N in coils and turns, respectively. π_N did not significantly differ between relative
478 solvent accessibility subclasses (Kruskal-Wallis tests H=2.308, $d.f.$=2, p=0.315), but differences
479 were marginally significant between structural domains (Kruskal-Wallis tests H=11.035, $d.f.$=5,
480 p=0.051). The third, fourth and fifth beta strands displayed the highest levels of non-
481 synonymous diversity (π_N=0.0341, π_N=0.0341, and π_N=0.0341, respectively), and 18 out of 25
482 MAX effectors displayed their highest values of π_N at one of these three beta strands (S6 Table).

483 In the 23 MAX effectors under multiallelic directional selection that passed our filter on
484 sequence length (at least ten sequences longer than 10bp, see Methods), differences in d_N/d_S
485 were most pronounced between subclasses of structural features (Kruskal-Wallis tests
486 H=5.499, $d.f.$=2, p=0.064), with higher average d_N/d_S values for coils and turns (d_N/d_S=2.490
487 and d_N/d_S=1.184, respectively) than extended conformations (d_N/d_S=0.573) (Figure 6D; S6
488 Table). The average d_N/d_S was also close to one for the 30-60% subclass of relative solvent

489 accessibility (d_N/d_S=0.994), and 12 MAX effectors with signatures of directional selection had
490 their highest d_N/d_S values for this subclass, although differences were not significant.
491 Overall, these analyses show that multiallelic balancing and directional selection acted
492 preferentially on coils and turns, but that the impact of two forms of selection on structural
493 domains and solvent accessibility subclasses differs.
494
495
496 **DISCUSSION**
497 *MAX effectors as model systems to investigate effector evolution.*
498 Effectors involved in coevolutionary interactions with host-derived molecules are expected to
499 undergo non-neutral evolution. Yet, the role of natural selection in shaping polymorphism and
500 divergence at effectors has remained largely elusive [2]. Despite the prediction of large and
501 molecularly diversified repertoires of effector genes in many fungal genomes, attempts to probe
502 into the evolutionary drivers of effector diversification in plant pathogenic fungi have been
503 hindered by the fact that, until recently, no large effector families had been identified. In this
504 study, we overcome the methodological and conceptual barrier imposed by effector hyper-
505 diversity by building on our previous discovery [17] of an important, structurally-similar, but
506 sequence-diverse family of fungal effectors called MAX. We used a combination of structural
507 modeling, evolutionary analyses, and molecular plant pathology experiments to provide a
508 comprehensive overview of polymorphism, divergence, gene expression, and presence/absence
509 at MAX effectors. When analyzed species-wide or at the level of sub-specific lineages, ratios of
510 non-synonymous to synonymous nucleotide diversity, as well as ratios of non-synonymous to
511 synonymous substitutions, were consistently higher at MAX effectors than at other loci. At the
512 species level, the two ratios were also significantly higher than expected under the standard
513 neutral model for a large fraction of MAX effectors. The signatures of adaptive evolution
514 detected at MAX effectors, combined with their extensive presence/absence variation, are

515 consistent with their central role in coevolutionary interactions with host-derived ligands that
516 impose strong selection on virulence effectors.
517
518 *Adaptive evolution at MAX effectors*
519 Rates of evolution determined from orthologous comparisons with outgroup sequences revealed
520 that, for a large fraction of MAX effectors, non-synonymous changes have accumulated faster
521 than synonymous changes. The fast rate of amino-acid change at MAX effectors is consistent
522 with a classic arms race scenario, which entails a series of selective sweeps as new virulent
523 haplotypes - capable of avoiding recognition by plant immune receptors that previously
524 prevented pathogen multiplication - spread to high frequency [49, 50]. Furthermore, it is
525 important to note that although large values of the d_N/d_S ratio provide strong evidence for
526 directional selection, small values do not necessarily indicate the lack thereof, as d_N/d_S ratios
527 represent the integration of genetic drift, constraint, and adaptive evolution [50][51]. Much of
528 the adaptive changes at MAX effectors probably took place before the radiation of *P. oryzae* on
529 its various hosts. However, the observation that d_N/d_S values determined from orthologous
530 comparisons with outgroup are higher at the species level than at the sub-specific lineage level
531 indicates that part of the signal of directional selection derives from inter-lineage amino acid
532 differences associated with host-specialization. Our structural modeling indicates that it is
533 preferentially "turns" and "coils", but also residues with intermediate solvent accessibility,
534 which often evolve at an unusually fast rate, and therefore that these are probably the residues
535 of MAX proteins preferentially involved in coevolutionary interactions with host-derived
536 molecules.
537 MAX effectors are characterized by a remarkable excess of non-synonymous
538 polymorphism, compared to synonymous polymorphism, at the species level, but also - albeit to
539 a lesser extent - at the sub-specific lineage level. This raises the question of how polymorphisms
540 are maintained in the face of adaptive evolution, given that selective sweeps under a classic arms
541 race scenario are expected to erase variation [6, 9]. Directional selection restricted to host-

542 specific lineages - i.e., local adaptation - may contribute to the signature of multiallelic balancing
543 selection observed at the species level. The observation of a positive correlation between π_N/π_S
544 and the differentiation statistic F_{ST}, together with the fact that most MAX effectors are
545 monomorphic at the lineage level, are consistent with a role of divergent selection exerted by
546 hosts in the maintenance of species-wide diversity at MAX effectors. However, the finding that
547 MAX effectors with a defined π_N/π_S at the sub-specific lineage level (i.e., MAX effectors with
548 $\pi_S \neq 0$) present a higher ratio than the other genes also indicates that the adaptive evolution
549 process is not simply one of successive selective sweeps. This is consistent with balancing
550 selection acting at the lineage level, through which polymorphisms in MAX virulence effectors
551 are maintained due to spatiotemporal variation in selection pressures posed by the hosts – a
552 process known as the trench-warfare model [49]. Our structural modeling suggests in particular
553 that the "coils" and "turns" are the preferred substrate of these coevolutionary interactions
554 leading to the maintenance of elevated polymorphism at MAX virulence effectors.
555
556
557 *Expression kinetics of MAX effectors*
558 Expression profiling showed that the MAX effector repertoire was induced specifically and
559 massively during infection. Depending on the host genotype, between 64 and 78% of the MAX
560 effectors were expressed and expression was particularly strong during the early stages of
561 infection. These findings are consistent with previous studies that analyzed genome-wide gene
562 expression during rice infection or specifically addressed MAX effector expression, and they
563 reinforce the hypothesis that MAX effectors are crucial for fungal virulence and specifically
564 involved in the biotrophic phase of infection [17, 52].
565 How this coordinated deployment of the MAX effectors is regulated remains largely
566 unknown. Genome organization does not seem to be a major factor, since MAX effectors do not
567 colocalize and more generally, there is no clustering of effectors in the *P oryzae* genome, only a
568 slight enrichment in subtelomeric regions [52, 53]. This differs from other pathogenic fungi,

569 such as *Leptosphaeria maculans*, for which early-expressed effectors are clustered in AT-rich
570 isochores, and co-regulated by epigenetic mechanisms [54]. Analysis of promoter regions of
571 MAX effectors did not identify common DNA motifs that may be targeted by transcription
572 factors, and no such transcriptional regulators that would directly regulate large fractions of the
573 effector complement of *P. oryzae* have been identified yet. The few known transcriptional
574 networks controlled by regulators of *P. oryzae* pathogenicity generally comprise different
575 classes of fungal virulence genes such as secondary metabolism genes or carbohydrate-active
576 enzymes; they are not restricted to effectors or enriched in MAX effectors. For instance, Rgs1, a
577 regulator of G-protein signaling necessary for appressorium development, represses the
578 expression of 60 temporally co-regulated effectors in axenic culture and during the pre-
579 penetration stage of plant infection [55]. However, among them are only two MAX effectors,
580 *MAX15* (*MGG05424*) and *MAX67* (*MGG16175*). This suggests that multiple complementary
581 mechanisms contribute to the precise coordination of MAX effector expression during rice
582 invasion.

583 Expression profiling also revealed that the plant host genotype strongly influenced the
584 expression of the MAX effector repertoire, suggesting that plasticity in effector expression may
585 contribute to the adaptation of *P. oryzae* to its hosts. MAX effectors were stronger expressed in
586 the more resistant Kitaake rice variety than in highly susceptible Maratelli rice. This is
587 reminiscent of other pathogenic fungi, such as *Fusarium graminearum* and *L. maculans,* for
588 which a relationship between host resistance levels and effector expression was established [56,
589 57]. Specific experiments will have to be performed to analyze in more detail this potential link
590 between plant resistance and MAX effector expression.

591

592 *Presence/absence polymorphism of MAX effectors*

593 Pangenome analyses demonstrated extensive variability in the MAX effector repertoire. In cases
594 where MAX effectors are specifically absent from some lineages, but present in most or all
595 others, it is tempting to hypothesize that they experienced immune-escape loss-of-function

596 mutations that directly contributed to host range expansion or host shifts. A possible example of
597 such a mechanism is the non-MAX effector *PWT3* of *P. oryzae* that is specifically absent from the
598 *Triticum*-infecting lineage [29]. PWT3 triggers resistance in wheat cultivars possessing the
599 *RWT3* resistance gene [58], and its loss was shown to coincide with the widespread deployment
600 of *RWT3* wheat. Similarly, the loss of the effector *AVR1-CO39* (*MAX86*), which is specifically
601 absent from the *Oryza*-infecting lineage and that is detected by the rice NLR immune receptor
602 complex RGA4/RGA5, has been suggested to have contributed to the initial colonization of rice
603 by the *Setaria*-infecting lineage [20, 31, 59]. Two other orthologous *P. oryzae* effectors, *PWL1*
604 and *PWL2*, exclude Eleusine and rice-associated isolates from infecting *Eragrostis curvula*, and
605 can, therefore, also be considered as host-specificity determinants [28, 60]. Interestingly,
606 ALPHAFOLD predicts PWL2 to adopt a MAX effector fold [61]. In our study, however, gene knock-
607 in experiments with *MAX79*, *MAX83,* and *MAX89* - specifically absent from the *Oryza*-infecting
608 lineage - did not reveal a strong effect on virulence towards a large panel of rice varieties. Hence,
609 unlike AVR1-CO39, these effectors are not key determinants of host-specificity. This suggests
610 that overcoming non-host resistance is not the only and maybe not the main evolutionary
611 scenario behind the specific loss of MAX effectors in the *Oryza*-infecting lineage. A possible
612 alternative mechanism that can explain massive MAX effector loss during host shifts is a lack of
613 functionality in the novel host. Some MAX effectors from a given lineage may have no function in
614 the novel host, simply because their molecular targets are absent or too divergent in the novel
615 host. Cellular targets of fungal effectors remain unknown for the most part, but knowledge of the
616 molecular interactors of MAX effectors may help shed light on the drivers of their
617 presence/absence polymorphism.
618
619 *Concluding remarks*
620 The discovery of large, structurally-similar, effector families in pathogenic fungi and the
621 increasing availability of high-quality whole genome assemblies and high-confidence annotation
622 tools, pave the way for in-depth investigations of the evolution of fungal effectors by

interdisciplinary approaches combining state-of-the-art population genomics, protein structure analysis, and functional approaches. Our study on MAX effectors in the model fungus and infamous cereal killer *P. oryzae* demonstrates the power of such an approach. Our investigations reveal the fundamental role of directional and balancing selection in shaping the diversity of MAX effector genes and pinpoint specific positions in the proteins that are targeted by these evolutionary forces. This type of knowledge is still very limited on plant pathogens, and there are very few studies compared to the plethoric literature on the evolution of virulence factors in human pathogens. Moreover, by revealing the concerted and plastic deployment of the MAX effector repertoire, our study highlights the current lack of knowledge on the regulation of these processes. A major challenge will now be to identify the regulators, target proteins and mode of action of MAX effectors, in order to achieve a detailed understanding of the relationships between the structure, function and evolution of these proteins.

METHODS

Genome assemblies, gene prediction, and pan-genome analyses

Among the 120 genome assemblies included in our study, 66 were already publicly available, and 54 were newly assembled (S1 Table). For the 54 newly assembled genomes, reads were publicly available for 50 isolates, and four additional isolates were sequenced (available under BioProject PRJEB47684). For the four sequenced isolates, DNA was extracted using the same protocol as in ref. [62]. For the 54 newly generated assemblies, CUTADAPT [63] was used for trimming and removing low-quality reads, reads were assembled with ABYSS 2.2.3 [64] using eight different K-mer sizes, and we chose the assembly produced with the K-mer size that yielded the largest N50. For all 120 genome assemblies, genes were predicted by BRAKER 1 [65] using RNAseq data from ref. [21] and protein sequences of isolate 70-15 (Ensembl Fungi release 43). To complement predictions from BRAKER, we also predicted genes using AUGUSTUS 3.4.0 [56][3] with RNAseq data from ref. [21] , protein sequences of isolate 70-15 (Ensembl Fungi release 43), and *Magnaporthe grisea* as the training set. Gene predictions from BRAKER and AUGUSTUS were merged by removing the genes predicted by AUGUSTUS that overlapped with genes predicted by BRAKER. Repeated elements were masked with REPEATMASKER 4.1.0 (http://www.repeatmasker.org/). The quality of genome assembly and gene prediction was checked using BUSCO [32]. The homology relationships among predicted genes were identified using ORTHOFINDER v2.4.0 [36]. The size of pan- and core-genomes was estimated using rarefaction, by resampling combinations of one to 119 genomes, taking a maximum of 100 resamples by pseudo-sample size. Sequences for each orthogroup were aligned at the codon level (i.e., keeping sequences in coding reading frame) with TRANSLATORX [66], using MAFFT v7 [67] as the aligner and default parameters for GBLOCKS [68].

Identification of effectors sensu lato, and MAX effectors

We predicted the secretome by running SIGNALP 4.1 [69], TARGET [70], and PHOBIUS [62] to identify signal peptides in the translated coding sequences of 12000 orthogroups. Only proteins

663 predicted to be secreted by at least two methods were retained. Transmembrane domains were
664 identified using TMHMM [71] and proteins with a transmembrane domain outside the 30 first
665 amino acids were excluded from the predicted secretome. Endoplasmic reticulum proteins were
666 identified with PS-SCAN (https://ftp.expasy.org/databases/prosite/ps_scan/), and excluded.

667 To identify MAX effectors, we used the same approach as in the original study that
668 described MAX effectors [17]. We first used PSI-BLAST [33] to search for homologs of known
669 MAX effectors (AVR1-CO39, AVR-Pia, AvrPiz-t, AVR-PikD, and ToxB) in the predicted secretome.
670 Significant PSI-BLAST hits (e-value < e-4) were aligned using a structural alignment procedure
671 implemented in TM-ALIGN [35]. Three rounds of HMMER [34] searches were then carried out,
672 each round consisting of alignment using TM-ALIGN, model building using HMMBUILD, and HMM
673 search using HMMSEARCH (e-value < e-3). Only proteins with two expected, conserved Cysteines
674 less than 33-48 amino acids apart were retained in the first two rounds of HMMER searches, as
675 described in ref. [17].

676 Subsequent evolutionary analyses were conducted on three sets of orthogroups: MAX
677 effectors, putative effectors, and other genes. The "MAX" group corresponded to 80 orthogroups
678 for which at least 10% of sequences were identified as MAX effectors. The "secreted proteins"
679 groups corresponded to 3283 orthogroups that were not included in the MAX group, and for
680 which at least 10% of sequences were predicted to be secreted proteins. The last group included
681 the remaining 11404 orthogroups.

682 For missing MAX effector sequences, we conducted an additional similarity search to
683 correct for gene prediction errors. For a given MAX orthogroup and a given isolate, if a MAX
684 effector was missing, we used BLAST-N to search for significant hits using the longest sequence of
685 the orthogroup as the query sequence, and the isolate's genome assembly as the subject
686 sequence (S3 Table). We also corrected annotation errors, such as the presence of very short
687 (typically <50bp) or very long (typically >500bp) introns, missing terminal exons associated
688 with premature stops, or frameshifts caused by indels. All these annotation errors were checked,
689 and corrected manually if needed, using the RNAseq data used in gene prediction in the

690 INTEGRATIVE GENOME VIEWER [72, 73]. We also found that some orthogroups included chimeric
691 gene resulting from erroneous merging of two genes that were adjacent in assemblies. This was
692 the case for orthogroups OG0000093 and OG0010985, and we used RNA-seq data in the
693 INTEGRATIVE GENOME VIEWER to split the merged genic sequences and keep only the sequence that
694 corresponded to a MAX effector.

695 For evolutionary analyses conducted on single-copy orthologs, the 11 orthogroups that
696 included paralogous copies of MAX effectors were split into sets of orthologous sequences using
697 genealogies inferred using RAXML v8 [37], yielding a total of 94 single-copy MAX orthologs, of
698 which 90 orthologs passed our filters on length and sample size to be included in evolutionary
699 analyses (see below). For each split orthogroup, sets of orthologous sequences were assigned a
700 number that was added to the orthogroup's identifier as a suffix (for instance paralogous
701 sequences of orthogroup OG0000244 were split into orthogroups OG0000244_1 and
702 OG0000244_2). Sequences were re-aligned using TRANSLATORX (see above) after splitting
703 orthogroups.

704 All genome assemblies, aligned coding sequences for all orthogroups and single-copy
705 orthologs are available in Zenodo, doi: 10.5281/zenodo.7689273.

706

707 *Homology modeling of MAX effectors*

708 To check that orthogroups predicted to be MAX effectors had the typical 3D structure of MAX
709 effectors with two beta sheets of three beta strands each, eight experimental structures with
710 MAX-like folds were selected as 3D templates for homology modeling (PDB identifiers of the
711 templates: 6R5J, 2MM0, 2MM2, 2MYW, 2LW6, 5A6W, 5Z1V, 5ZNG). For each of the 94 MAX
712 orthologous groups, one representative protein was selected and homology models of this 1D
713 query relative to each 3D template were built using MODELLER [74] with many alternative query-
714 template threading alignments. The structural models generated using the alternative
715 alignments were evaluated using a combination of different structural scores (DFIRE [75], GOAP
716 [76], QMEAN [77]). Our homology modeling procedure will be described with more details in a

717 manuscript currently in preparation. The best structural models for the 94 representative
718 sequences of each group of MAX orthologs are available at
719 https://pat.cbs.cnrs.fr/magmax/model/. The correspondence between MAX orthogroups
720 identifiers used in homology modeling and MAX orthogroups identifiers resulting from gene
721 prediction is given in S2 Table.
722
723
724
725 *Evolutionary analyses*
726 Lineage-level analyses were conducted on a dataset from which divergent or introgressed
727 isolates were removed (G17 from *Eragrostis*, Bm88324 & Bd8401 from *Setaria*, 87-120; BF0072
728 and BN0019 from *Oryza*; IR0088 from *Echinochloa*), to limit the impact of population
729 subdivision within lineages. The *Stenotaphrum*-infecting lineage was not included in lineage-
730 level analyses due to the small sample size.
731 Nucleotide diversity [78] and population differentiation [43] were estimated using
732 EGGLIB v3 [79]. Sites with more than 30% missing data were excluded. Orthogroups with less
733 than 10 sequences (*nseff*<10, *nseff* being the average number of used samples among sites that
734 passed the missing data filter) or shorter than 30bp (*lseff*<30, *lseff* being the number of sites used
735 for analysis after filtering out sites with too many missing data) were excluded from computations.
736 For analyses of polymorphism at secondary structure annotations, the cutoff on *lseff* was set at
737 10bp.
738 For the computation of d_N/d_S and quantification of adaptive evolution, we used isolate
739 NI919 of *Pyricularia* sp. LS [45, 46] as the outgroup (assembly GCA_004337975.1, European
740 Nucleotide Archive). Genes were predicted in the outgroup assembly using EXONERATE V2.2
741 CODING2GENOME [80]. For each gene, the query sequence was a *P. oryzae* sequence randomly
742 selected among sequences with the fewest missing data. In parsing EXONERATE output, we

743 selected the sequence with the highest score, with a length greater than half the length of the
744 query sequence.

745 The d_N/d_S ratio was estimated using a maximum likelihood approach (runmode=-2,
746 CodonFreq=2 in CODEML [81]), in pairwise comparisons of protein coding sequences (*i.e.*,
747 without using a phylogeny). For each d_N/d_S we randomly selected 12 ingroup sequences and
748 computed the average d_N/d_S across the 12 ingroup/outgroup pairs.

749 Kruskal-Wallis tests were performed using the SCIPY.STATS.KRUSKAL library in PYTHON 3.7.
750 Posthoc Mann-Whitney U-tests were performed using the SCIKIT_POSTHOCS library in PYTHON 3.7,
751 with p-values adjusted using the Bonferroni-Holm method.

752

753 *Constructs for the transformation of fungal isolates*

754 PCR products used for cloning were generated using the Phusion High-Fidelity DNA Polymerase
755 (Thermo Fisher) and the primers listed in S7 Table. Details of the constructs are given in S8
756 Table. Briefly, the pSC642 plasmid (derived from the pCB1004 vector), containing a cassette for
757 the expression of a gene of interest under the control of the *AVR-Pia* promoter (*pAVR-Pia*) and
758 the *Neurospora crassa β-tubulin* terminator (*t-tub*), was amplified by PCR with primers oML001
759 and oTK609 for the insertion of MAX genes listed in S9 Table. The MAX genes
760 *Mo_US0071_000070* (*MAX79*), *Mo_US0071_046730* (*MAX89*) and *Mo_US0071_115900*
761 (*MAX83*), amplified by PCR from genomic DNA of the *P. oryzae* isolate US0071, were inserted
762 into this vector using the Gibson Assembly Cloning Kit (New England BioLabs). The final
763 constructs were linearized using the KpnI restriction enzyme (Promega) before *P. oryzae*
764 transformation.

765

766 *Plant and fungal growth conditions*

767 Rice plants (*Oryza sativa*) were grown in a glasshouse in a substrate of 31% coconut peat, 30%
768 Baltic blond peat, 15% Baltic black peat, 10% perlite, 9% volcanic sand, and 5% clay,
769 supplemented with 3.5 $g.L^{-1}$ of fertilizer (Basacote® High K 6M, NPK 13-5-18). Plants were

770 grown under a 12h-light photoperiod with a day-time temperature of 27°C, night-time
771 temperature of 21°C, and 70% humidity. For spore production, the wild-type and transgenic
772 isolates of *P. oryzae* Guy11 were grown for 14 days at 25°C under a 12h-light photoperiod on
773 rice flour agar medium (20 $g.L^{-1}$ rice seed flour, 2.5 $g.L^{-1}$ yeast extract, 1.5% agar, 500.000U
774 penicillin g), supplemented with 240 $\mu g.ml^{-1}$ hygromycin for transgenic isolates. For mycelium
775 production, plugs of mycelium of *P. oryzae* Guy11 were grown in liquid medium (10 $g.L^{-1}$
776 glucose, 3 $g.L^{-1}$ KNO_3, 2 $g.L^{-1}$ KH_2PO_4, 2,5 $g.L^{-1}$ yeast extract, 500 000U penicillin g) for 5 days at
777 25°C in the dark under agitation.
778
779 *Fungal transformation*
780 Protoplasts from the isolate Guy11 of *P. oryzae* were transformed by heat shock with 10µg of
781 KpnI-linearized plasmids for the expression of MAX effectors or RFP as described previously
782 [74]. After two rounds of antibiotic selection and isolation of monospores, transformed isolates
783 were genotyped by Phire Plant Direct PCR (Thermo Scientific) using primers described in S7
784 Table. The Guy11 transgenic isolates expressing *AVR-Pia* and *AVR1-CO39* were previously
785 generated [82, 83].
786
787 *Fungal growth and infection assays*
788 For the analysis of interaction phenotypes, leaves of three-week-old rice plants were spray-
789 inoculated with conidial suspensions (40 000 $conidia.ml^{-1}$ in water with 0.5% gelatin). Plants
790 were incubated for 16 hours in the dark at 25°C and 95% relative humidity, and then grown for
791 six days in regular growth conditions. Seven days after inoculation, the youngest leaf that was
792 fully expanded at the time of inoculation was collected and scanned (Scanner Epson Perfection
793 V370) for further symptoms analyses. Phenotypes were qualitatively classified according to
794 lesion types: no lesion or small brown spots (resistance), small lesions with a pronounced
795 brown border and a small gray center (partial resistance), and larger lesions with a large gray
796 center or dried leaves (susceptibility). For the analysis of gene expression, plants were spray-

797 inoculated with conidial suspensions at 50 000 conidia.ml^{-1} (in water with 0.5% gelatin), and
798 leaves were collected three days after inoculation.
799
800 *RNA extraction and qRT-PCR analysis*
801 Total RNA extraction from rice leaves or Guy11 mycelium and reverse transcription were
802 performed as described by ref. [84]. Briefly, frozen leaves and mycelium were mechanically
803 ground. RNA was extracted using TRI-reagent® (Sigma-Aldrich) and chloroform separation.
804 Denaturated RNA (5µg) was retrotranscribed and used for quantitative PCR using GoTaq qPCR
805 Master Mix according to the manufacturer's instructions (Promega) at a dilution of 1/10 for
806 mycelium and 1/7 for rice leaves. The primers used are described in S7 Table. Amplification was
807 performed as described by ref. [84] using a LightCycler480 instrument (Roche), and data were
808 extracted using the instrument software. To calculate *MAX* gene expressions, the $2^{-\Delta\Delta CT}$ method
809 and primers measured efficiency were used. Gene expression levels are expressed relative to the
810 expression of constitutive reference gene *MoEF1α*.
811
812 *Statistical analyses of phenotypic data*
813 For expression comparison between Kitaake and Maratelli infection, all analyses were
814 performed using R (http://www.r-project.org). The entire kinetic experiment was repeated
815 three times with five biological replicates for each time point. For each variety, gene, and
816 experimental replicate, values corresponding to the day post-inoculation with the highest
817 median expression were extracted for statistical analyses. Expression data were not normally
818 distributed so for each gene, differences between varieties were evaluated using non-parametric
819 Mann-Whitney U-tests.
820
821 **SUPPLEMENTARY MATERIALS**
822 S1 Table. Genomic assemblies with metadata.
823 S2 Table. Nomenclature of MAX effectors.

S3 Table. Presence/absence of MAX effector orthologs.

S4 Table. The expression of MAX79, MAX83 and MAX89 in Guy11 does not trigger recognition in a panel of rice varieties.

S5 Table. Gene average of summary statistics of polymorphism, differentiation and divergence.

S6 Table. π_N and d_N/d_S in different classes of secondary structure annotations for MAX effectors with $\pi_N/\pi_S>1$ and $d_N/d_S>1$, respectively.

S7 Table. Primers for cloning and expression analyses.

S8 Table. Vector constructs.

S9 Table. Sequences of the MAX effectors in the isolate US0071 that were used for the complementation of Guy11.

S1 Figure. Expression patterns of MAX effectors during rice infection.

S2 Figure. Differential expression levels of MAX effectors upon infection of two different rice cultivars.

S3 Figure. Nucleotide diversity (π), ratio of non-synonymous to synonymous nucleotide diversity (π_N/π_S), orthogroup frequency for MAX effectors, other secreted proteins, and other genes.

S4 Figure. Assembly size, number of predicted genes, and number of orthogroups in lineages of *M. oryzae*.

S5 Figure. MAX79, MAX83 and MAX89 are expressed in the transgenic Guy11 isolates upon rice inoculation.

S6 Figure. F_{ST} versus π_N/π_S at MAX effectors

S7 Figure. Secondary structure annotations of MAX effectors aligned with TM-ALIGN.

S1 Data. Summary statistics per orthogroup.

S2 Data. Summary statistics per MAX effector ortholog, species wide, and per lineage.

REFERENCES

1. Schulze-Lefert P, Panstruga R. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science. 2011;16(3):117-25. doi: 10.1016/j.tplants.2011.01.001.
2. Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, et al. The genome biology of effector gene evolution in filamentous plant pathogens. Annual review of phytopathology. 2018;56:21-40.
3. Haldane JBS. Disease and evolution. Ricerca Scient 1949;19:68-76.
4. Flor HH. Inheritance of pathogenicity in Melampsora lini. Phytopathology. 1942;32:653-69.
5. Barrett JA. Frequency-dependent selection in plant-fungal interactions. Philosophical Transactions of the Royal Society of London B, Biological Sciences. 1988;319(1196):473-83.
6. Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant R-genes. Science. 2001;292(5525):2281-5.
7. Brown JKM. Chance and selection in the evolution of barley mildew. Trends in Microbiology. 1994;2(12):470-5. PubMed PMID: 29.
8. Brown JKM. Durable resistance of crops to disease: a Darwinian perspective. Annual review of phytopathology. 2015;53:513-39.
9. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature. 1999;400(6745):667-71.
10. Gladieux P, van Oosterhout C, Fairhead S, Jouet A, Ortiz D, Ravel S, et al. Extensive immune receptor repertoire diversity in disease-resistant rice landraces. bioRxiv. 2022:2022-12.
11. Bakker EG, Toomajian C, Kreitman M, Bergelson J. A genome-wide survey of R gene polymorphisms in Arabidopsis. The Plant cell. 2006;18(8):1803-18.
12. Bakker EG, Traw MB, Toomajian C, Kreitman M, Bergelson J. Low levels of polymorphism in genes that control the activation of defense response in Arabidopsis thaliana. Genetics. 2008;178(4):2031-43.
13. Ebert D, Fields PD. Host–parasite co-evolution and its genomic signature. Nature Reviews Genetics. 2020;21(12):754-68.
14. Ebbole DJ, Chen M, Zhong Z, Farmer N, Zheng W, Han Y, et al. Evolution and Regulation of a Large Effector Family of Pyricularia oryzae. Molecular Plant-Microbe Interactions. 2021;34(3):255-69.
15. Seong K, Krasileva KV. Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen Magnaporthe oryzae. Molecular Plant-Microbe Interactions. 2021;34(11):1267-80.
16. Seong K, Krasileva KV. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. Nature Microbiology. 2023;8(1):174-87.
17. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. PLoS Pathog. 2015;11(10):e1005228.
18. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. Nature ecology & evolution. 2019;3(3):430.
19. Fernandez J, Orth K. Rise of a cereal killer: the biology of Magnaporthe oryzae biotrophic growth. Trends in microbiology. 2018;26(7):582-97.
20. Couch BC, Fudal I, Lebrun M-H, Tharreau D, Valent B, van Kim P, et al. Origins of Host-Specific Populations of the Blast Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of Pandemic Clones on Rice and Weeds of Rice. Genetics. 2005;170(2):613-30. PubMed PMID: 61.
21. Pordel A, Ravel S, Charriat F, Gladieux P, Cros-Arteil S, Milazzo J, et al. Tracing the origin and evolutionary history of Pyricularia oryzae infecting maize and barnyard grass. Phytopathology. 2021;111(1):128-36.
22. Kato H, Yamamoto M, Yamaguchi-Ozaki T, Kadouchi H, Iwamoto Y, Nakayashiki H, et al. Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of Pyricularia populations isolated from Gramineae, Bambusideae and Zingiberaceae plants. Journal of General Plant Pathology. 2000;66:30-47.
23. Urashima AS, Igarashi S, Kato H. Host range, mating type, and fertility of Pyricularia grisea from wheat in Brazil. Plant Disease. 1993;77(12):1211-6.
24. Igarashi S. Pyricularia em trigo. 1. Ocorrencia de Pyricularia sp noestado do Parana. Fitopatol Bras. 1986;11:351-2.
25. Milazzo J, Pordel A, Ravel S, Tharreau D. First scientific report of Pyricularia oryzae causing gray leaf spot disease on perennial ryegrass (Lolium perenne) in France. Plant Disease. 2019;103(5):1024-.

26. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. BMC biology. 2016;14(1):84.
27. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae. mBio. 2018;9(1).
28. Sweigard JA, Carroll AM, Kang S, Farrall L, Chumley FG, Valent B. Identification, Cloning, and Characterization of *PWL2*, a Gene for Host Species Specificity in the Rice Blast Fungus. The Plant cell. 1995;7(8):1221-33. PubMed PMID: 260.
29. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 2017;357(6346):80-3.
30. Asuke S, Tanaka M, Hyon G-S, Inoue Y, Vy TTP, Niwamoto D, et al. Evolution of an Eleusine-specific subgroup of Pyricularia oryzae through a gain of an avirulence gene. Molecular Plant-Microbe Interactions. 2020;33(2):153-65.
31. Zheng Y, Zheng W, Lin F, Zhang Y, Yi Y, Wang B, et al. AVR1-CO39 is a predominant locus governing the broad avirulence of Magnaporthe oryzae 2539 on cultivated rice (Oryza sativa L.). Molecular plant-microbe interactions. 2011;24(1):13-7.
32. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-2.
33. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-402.
34. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic acids research. 2011;39(suppl_2):29-37.
35. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic acids research. 2005;33(7):2302-9.
36. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology. 2015;16(1):157.
37. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.
38. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014;196(4):973-83.
39. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution. 2006;23(2):254-67.
40. Tosa Y, Osue J, Eto Y, Oh H-S, Nakayashiki H, Mayama S, et al. Evolution of an avirulence gene, AVR1-CO39, concomitant with the evolution and differentiation of Magnaporthe oryzae. Molecular Plant-Microbe Interactions. 2005;18(11):1148-60.
41. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. The Plant cell. 2013;25(4):1463-81. Epub 2013/04/04. doi: 10.1105/tpc.112.107201. PubMed PMID: 23548743; PubMed Central PMCID: PMCPMC3663280.
42. Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, et al. A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. The Plant Journal. 2011;66(3):467-79.
43. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358-70. PubMed PMID: 283.
44. Hahn MW. Molecular population genetics: Oxford University Press; 2018.
45. Hirata K, Kusaba M, Chuma I, Osue J, Nakayashiki H, Mayama S, et al. Speciation in Pyricularia inferred from multilocus phylogenetic analysis. Mycological Research. 2007;111(7):799-808.
46. Gómez Luciano LB, Tsai IJ, Chuma I, Tosa Y, Chen Y-H, Li J-Y, et al. Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. Molecular Biology and Evolution. 2019;36(6):1148-61.
47. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular biology and evolution. 1998;15(5):568-73.
48. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins: Structure, Function, and Bioinformatics. 1995;23(4):566-79.
49. Clay K, Kover PX. The Red Queen hypothesis and plant/pathogen interactions. Annual review of Phytopathology. 1996;34(1):29-50.
50. Van Valen L. A new evolutionary law. 1973.

51. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends in ecology & evolution. 2000;15(12):496-503.
52. Yan X, Tang B, Ryder LS, MacLean D, Were VM, Eseola AB, et al. The transcriptional landscape of plant infection by the rice blast fungus Magnaporthe oryzae reveals distinct families of temporally co-regulated and structurally conserved effectors. bioRxiv. 2022:2022-07.
53. Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, et al. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus Magnaporthe oryzae Attacking Different Host Plants. Genome Biol Evol. 2015;7(10):2896-912. Epub 2015/10/11. doi: 10.1093/gbe/evv187. PubMed PMID: 26454013; PubMed Central PMCID: PMCPMC4684704.
54. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, et al. Epigenetic control of effector gene expression in the plant pathogenic fungus Leptosphaeria maculans. PLoS genetics. 2014;10(3):e1004227.
55. Tang B, Yan X, Ryder LS, Cruz-Mireles N, Soanes DM, Molinari C, et al. Rgs1 is a regulator of effector gene expression during plant infection by the rice blast fungus Magnaporthe oryzae. bioRxiv. 2022:2022-09.
56. Fall LA, Salazar MM, Drnevich J, Holmes JR, Tseng M-C, Kolb FL, et al. Field pathogenomics of Fusarium head blight reveals pathogen transcriptome differences due to host resistance. Mycologia. 2019;111(4):563-73.
57. Sonah H, Zhang X, Deshmukh RK, Borhan MH, Fernando WGD, Belanger RR. Comparative transcriptomic analysis of virulence factors in Leptosphaeria maculans during compatible and incompatible interactions with canola. Frontiers in plant science. 2016;7:1784.
58. Arora S, Steed A, Goddard R, Gaurav K, O'Hara T, Schoen A, et al. A wheat kinase and immune receptor form host-specificity barriers against the blast fungus. Nature Plants. 2023:1-8.
59. Farman ML, Eto Y, Nakao T, Tosa Y, Nakayashiki H, Mayama S, et al. Analysis of the structure of the *AVR1-Co39* avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. Molecular Plant-Microbe Interactions. 2002;15:6-16. PubMed PMID: 90.
60. Kang S, Sweigard Ja Fau - Valent B, Valent B. The PWL host specificity gene family in the blast fungus Magnaporthe grisea. Molecular Plant Microbe Interactions. 1995;(0894-0282 (Print)).
61. Brabham HJ, Gómez De La Cruz D, Were V, Shimizu M, Saitoh H, Hernández-Pinzón I, et al. Barley MLA3 recognizes the host-specificity determinant PWL2 from rice blast (M. oryzae). bioRxiv. 2022:2022-10.
62. Thierry M, Charriat F, Milazzo J, Adreit H, Ravel S, Cros-Arteil S, et al. Maintenance of divergent lineages of the Rice Blast Fungus Pyricularia oryzae through niche separation, loss of sex and post-mating genetic incompatibilities. PLoS pathogens. 2022;18(7):e1010687.
63. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-2.
64. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. Genome research. 2009;19(6):1117-23.
65. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2015;32(5):767-9.
66. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research. 2010;38(suppl_2):W7-W13.
67. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics. 2008;9(4):286-98. doi: 10.1093/bib/bbn013. PubMed PMID: ISI:000256756400004.
68. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution. 2000;17(4):540-52.
69. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods. 2011;8(10):785-6.
70. Emanuelsson O, Nielsen H, Brunak S, Von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology. 2000;300(4):1005-16.
71. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001;305(3):567-80.
72. Robinson JT, Thorvaldsdóttir H, Turner D, Mesirov JP. igv. js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics. 2023;39(1):btac830.
73. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24-6.

74. Webb B, Sali A. Protein Structure Modeling with MODELLER. Methods Mol Biol. 2021;2199(1940-6029 (Electronic)):239–55.
75. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein science. 2002;11(11):2714-26.
76. Zhou H, Skolnick J. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. Biophysical journal. 2011;101(8):2043-52.
77. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics. 2008;71(1):261-77.
78. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;105(2):437-60.
79. Siol M, Coudoux T, Ravel S, De Mita S. EggLib 3: A python package for population genetics and genomics. Molecular Ecology Resources. 2022;22(8):3176-87.
80. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC bioinformatics. 2005;6(1):1-11.
81. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences. 1997;13(5):555-6.
82. Ribot C, Césari S, Abidi I, Chalvon V, Bournaud C, Vallet J, et al. The M agnaporthe oryzae effector AVR 1–CO 39 is translocated into rice cells independently of a fungal-derived machinery. The Plant Journal. 2013;74(1):1-12.
83. Ortiz D, De Guillen K, Cesari S, Chalvon V, Gracy J, Padilla A, et al. Recognition of the Magnaporthe oryzae effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. The Plant cell. 2017;29(1):156-68.
84. Pélissier R, Buendia L, Brousse A, Temple C, Ballini E, Fort F, et al. Plant neighbour-modulated susceptibility to pathogens in intraspecific mixtures. Journal of experimental botany. 2021;72(18):6570-80.