Title: Transposon elements shape population divergence of rice blast fungus *Magnaporthe oryzae*

Lianyu Lin[a], Ting Sun[a], Jiayuan Guo[a,b], Lili Lin[a], Meilian Chen[b], Zhe Wang[a], Jiandong Bao[c], Justice Norvienyeku[d], Yijuan Han[b], Guodong Lu[a], Christopher Rensing[e], Huakun Zheng [a*], Zhenhui Zhong [a*], Zonghua Wang [a,b *]

[a] State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

[b] Fuzhou Institute of Oceanography, Minjiang University, Fuzhou 350108, China.

[c] State Key Laboratory for Managing Biotic and Chemical Treats to the Quality and Safety of Agro-Products, Institute of Plant Protection and Microbiology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.

[d] Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests, Ministry of Education, College of Plant Protection, Hainan University, Haikou 570228, China.

[e]Institute of Environmental Microbiology, College of Resource and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

# These authors contributed to this work equally.

* To whom correspondence should be addressed. Email: wangzh@fafu.edu.cn (Z.W.), zhenhuizhong@gmail.com (Z.Z.), huakunzheng@163.com (H.Z.)

ABSTRACT

Dynamic transposition of transposable elements (TEs) in fungal pathogens have been shown to have a significant impact on genome stability, gene expression, and virulence to the host. In *Magnaporthe oryzae*, genome plasticity resulting from TE insertion polymorphisms is a major driving force leading to the rapid evolution and diversification of this rice blast fungus. Despite their importance in *M. oryzae* population evolution and divergence, our understanding of TEs in this context remains limited. Here we conducted a genome-wide analysis of TE transposition dynamics in the eleven most abundant TE families and transcriptomic analysis in different *M. oryzae* populations. Our results show that these TEs have specifically expanded in recently isolated *M. oryzae* rice populations, with the presence/absence polymorphism of TE insertions highly concordant with population divergence on Geng/*Japonica* and Xian/*Indica* rice cultivars. Among the TE families, POT2 and Mg-SINE were shown to be overrepresented in clade-specific TE insertion events, indicating their essential roles in population divergence. Notably, the genes targeted by clade-specific TEs showed clade-specific expression patterns, suggesting a transcriptional regulation of TEs on targeted genes. Furthermore, population transcriptomic analysis revealed that genes disrupted by POT2 and Mg-SINE insertions exhibited significantly lower expression, implying a silencing function of these TEs on the targeted genes. Our study provides a comprehensive analysis of TEs in *M. oryzae* populations and demonstrates a crucial role of recent TE bursts in adaptive evolution and diversification of the *M. oryzae* rice-infecting lineage.

IMPORTANCE

M. oryzae is the causal agent of the destructive blast disease, which caused massive loss of yield annually worldwide. The fungus diverged into distinct clades during adaptation toward the two rice subspecies, Xian/indica and Geng/japonica. Although the role of TEs in the adaptive evolution was well established, mechanisms underlying how TEs promote the population divergence of *M. oryzae* remains largely unknown. In this study, we reported that TEs shape the population divergence of *M. oryzae* by differentially regulating gene expression between Xian/*Indica*-infecting and Geng/*Japonica*-infecting populations. Our results revealed a TE insertion mediated gene expression adaption that led to the divergence of *M. oryzae* population infecting different rice subspecies.

Key Words: transposable element; population divergence; rice subspecies adaptation; rice blast disease.

INTRODUCTION

Rice blast disease, caused by the ascomycete filamentous fungus *Magnaporthe oryzae* (syn: *Pyricularia oryzae*), poses a significant threat to rice production worldwide, resulting in annual yield losses of 10-30% (1, 2). The deployment of resistant rice varieties is the most cost-effective and environmentally friendly strategy for controlling rice blast disease. However, the effectiveness of such resistance can rapidly be diminished due to rapid mutations of the *M. oryzae* genome. Therefore, it is crucial to uncover the mechanisms by which *M. oryzae* rapidly evolves and evades the rice immune system.

Transposable elements (TEs) make up over 10% of the *M. oryzae* genome (3, 4). Many studies have shown that TEs are major contributors to genome plasticity and a significant source of genetic variation (5-9). Recent evidence has revealed that TEs can be inserted in or around important effectors and alter the virulence spectrum of *M. oryzae*. For example, the insertion of POT3 into the promoter region of AVR-Pita in *M. oryzae* led to the acquisition of virulence towards the resistant rice cultivar Yashiro-mochi (10). TEs are also able to affect gene expression networks, and TE-dependent transcriptional regulation of some essential effectors can facilitate the pathogen's transition in its life cycle. For instance, in *Phytophthora sojae*, the avirulence genes *PsAvr1a*, *1b*, and *3a/5* were found to be transcriptionally inactive due to TE insertions in their promoter or 3' UTR regions (11).

These studies have demonstrated the crucial role of TE-mediated genome mutations in pathogen adaptation. However, previous investigations have mainly focused on the impact of TEs on specific effectors or secreted proteins (4, 12-16). To gain a better understanding on the functions of TEs in the complex *M. oryzae*-rice pathosystem, we conducted a comprehensive analysis of TE insertion polymorphisms in 90 *M. oryzae* rice isolates previously described in our study (17) and systematically investigated the roles of TEs in the regulation of gene expression and population divergence in *M. oryzae*.

RESULTS

Recent large-scale TE bursts in the *M. oryzae* genome

To examine the impact of transposable elements (TEs) on the *M. oryzae* genome, we analyzed the insertion time of LTR-retrotransposons (LTR-RTs) by comparing the divergence between their 5' and 3' LTRs, which were identical at the time of TE insertion and gradually diverged in the process of neutral selection. Using a *de novo* based method, we identified 1,129 intact LTR-RTs in seven near chromosomal-level *M. oryzae* rice isolates, including 70-15, Guy11, FJ81278, FJ98099, FJ72ZC7-77, AV1-1-1, and Sar-2-20-1 (3, 4, 18). Surprisingly, 91.7% (1,036/1,129) of the LTR-RTs showed extremely low levels of divergence between their LTR pairs (over 99% identity), and 69.4% (784/1,129) of them possessed identical LTR pairs, indicating that these LTR-RTs were recently inserted and are still active in the *M. oryzae* genomes.

To assess the activity of the most abundant TE families, we analyzed 11 TE families, including six LTR-RTs (RETRO5, RETRO6, RETRO7, Maggy, MGLR-3, and Pyret), two non-LTR retrotransposons (MGL and Mg-SINE), and three DNA transposons (POT2, POT3, and Occan) (**Fig S1; Table S1**) (3, 19-22). We calculated the Kimura 2-Parameter genetic distance (k-value) to measure the divergence between TE sequences and their associated consensus sequences (23). Low k-values indicate that the TE fragments were generated through recent insertion events, while high k-values indicate that the TE fragments are divergent copies generated through ancient transposition events (24). Our analysis revealed that all 11 TE families, especially sequences of MGL, Mg-SINE, Maggy, and a subset of POT2, POT3, and Occan, exhibit very low k-values, and more than half of all TE contents consisted of newly emerged TEs (k-values less than 5) (**Fig. 1a and b**), indicating a recent, large-scale burst of TEs in the genome of *M. oryzae* rice isolates. Notably, we observed two or more k-value peaks in POT2, Mg-SINE, and Pyret, indicating that these TEs families

have undergone multiple rounds of amplification.

Furthermore, we investigated these TE families in genomes of four *M. oryzae* non-rice isolates, including three *M. oryzae* grass isolates (MZ5-1-6, NI907, NI919) and one *M. oryzae* wheat isolate (B71) (25, 26). We found that TEs only accounted for less than five percent of the genomes, which is significantly lower than that found in *M. oryzae* rice isolates. (**Table S1** and **S2**). Interestingly, the k-values of TEs were much larger and the proportion of newly emerged TEs was also significantly lower in *M. oryzae* non-rice isolates, indicating that the TEs were generated by more ancient insertion events and were inactive in *M. oryzae* non-rice isolates (**Fig. 1b and c**). Notably, the *M. oryzae* non-rice isolates contained only a few copies of fragmented Maggy and Mg-SINE, which were very abundant and possessed very low k-values in *M. oryzae* rice isolates, suggesting that the two TEs specifically amplified in *M. oryzae* rice isolates. In summary, our analysis demonstrated that the TEs were recently and specifically expanded in the genome of *M. oryzae* rice isolates and maintained high activity.

Whole genome landscape of TE dynamics in *M. oryzae* population

To examine the dynamics of transposable elements (TEs) in the *M. oryzae* population, we conducted a genome-wide analysis of TE insertion sites in 90 rice isolates that had previously been published (17), with two *M. oryzae* grass isolates from *Setaria viridis* as an outgroup (27). Using paired-end read mapping to the reference genome method, we identified a total of 11,163 TE insertion sites, with an average of 1,312 sites per isolate. To verify these insertion sites, we randomly selected 17 from Guy11 or FJ81278 isolates for PCR amplification (**Table S3**). The number (1,312 versus 739) and location of TE insertions differed significantly between the rice and *S. viridis* isolates (**Fig. S2**), reflecting the evolutionary divergence of these subspecies and their corresponding TEs, which is consistent with the finding that these two subspecies diverged ~10,000 years ago (27). More than half (6,040/11,163) of the TE insertion sites were specific to individual isolates (**Fig. 2**), indicating frequent TE transposition

events. Notably, the number of POT2 insertion sites was substantially higher than those of other TE families, suggesting a higher activity and variability of POT2 in the *M. oryzae* population.

Furthermore, we conducted a comprehensive analysis of the genomic distribution of TE insertion sites in the *M. oryzae* population. Of the 11,163 TE insertion sites identified, a significant proportion (77%; 8,582/11,163) was found to be located within 1 kb of the flanking regions of genes or intragenic regions, and over 40% of the genes (5,259/12,991) were disrupted by these TE insertions. Our enrichment analysis showed that the distribution of the 11 TE families is non-random in the *M. oryzae* genome, and each family displayed a distinct preference for specific genomic regions. For instance, Maggy, MGLR-3, RETRO5, RETRO7, Pyret, POT3, and Occan were predominantly distributed in intergenic regions, while POT2 displayed a preference for the gene flanking regions. Additionally, SINE, MGL, and RETRO6 were found to preferentially target intragenic regions (adjusted p-value < 0.01, **Table S4**). These findings provide valuable insights into the mechanisms underlying TE insertions and their potential impact on gene regulation in *M. oryzae*.

Higher frequency of POT2 and POT3 insertions in promoter of secreted proteins

Based on the two-speed genome model, filamentous fungi genomes had been shown to display a bipartite architecture comprising of a gene-rich compartment, which evolves slowly containing core genes encoding essential functions and metabolisms, and a repeat-rich compartment, which evolves rapidly containing important virulence effectors involved in pathogenicity (28). To avoid recognition by the plant immune system, some pathogen effectors such as avirulence genes were shown to be influenced or silenced by transposon elements (8, 29). In line with this finding, we observed that genes encoding secreted proteins were more closely associated with TE junctions (**Fig. 3a**), and the proportion of genes encoding secreted proteins with TE insertion within 1-kb flanking regions was significantly higher than in those of non-secreted proteins (**Table 1**). Moreover, enrichment analysis for TEs associated

171 with genes encoding secreted proteins showed that POT2 and POT3 were
172 overrepresented in promoters of genes encoding secreted protein (adjusted p-value <
173 0.01, **Table S5**), implying that genes encoding secreted proteins are more prone to
174 disruption by POT2 and POT3.

175 Previous studies have shown that genes with presence/absence variation (PAV) tended
176 to be located near transposable elements (TEs) in fungal pathogen genomes, while
177 core genes were located further away from the TE-rich compartments (30-32). In this
178 study, we compared the genomic distribution of core genes and PAV genes and found
179 that PAV genes tended to be located closer to TE insertion sites than core genes (**Fig.**
180 **3b**). Our results were consistent with previous findings and suggested that PAV genes
181 may be more susceptible to TE-mediated disruption, potentially contributing to their
182 faster evolution in the context of host-pathogen interactions. Together, these findings
183 suggested that TEs are play a role in the evolution of pathogen effectors and
184 contribute to the dynamic nature of host-pathogen interactions.

185 **Association of TEs with *M. oryzae* population divergence**

186 Considering that the *M. oryzae* population diverged within only one thousand years
187 (17) and that the large-scale TE burst happened not long ago, we thus raised the
188 question of whether the population divergence of *M. oryzae* is associated with recent
189 TE amplification events. To characterize correlations between the 90 isolates, we
190 estimated the distance for each of two isolates by calculating the identity of the TE
191 insertion sites. The pairwise TE insertion identities varied from 17.6% between
192 YN126441 and FJ12JN-084-3 to 87.2% between TW-PT3-1 and TW-PT6-1, with an
193 average of only 38.7%, thereby strongly implicating the high variability of TE
194 junctions between the different isolates. However, when we grouped these isolates
195 based on the pairwise TE insertion identities, we discovered two distinct clusters (**Fig.**
196 **4a**) matching the Clade2 and Clade3 isolates that we had previously defined based on
197 genome-wide SNPs (17). We noticed that the remaining isolates out of the two
198 clusters were also able to match the Clade1 isolates even though they showed a

relatively low pairwise TE insertion identity, which can be attributed to an earlier time of divergence of Clade1 isolates from the *M. oryzae* population when compared to the other two clade isolates. Furthermore, we found that the pairwise TE insertion identity between intra-clade isolates was significantly higher than that between inter-clade isolates (**Table S6**). Surprisingly, the hierarchical tree constructed using TE insertion sites showed a high degree of similarity to the phylogenetic tree constructed based on whole-genome SNPs (**Fig. 4b**) (17) , indicating that the recent TE amplification event has been a major force driving population divergence of *M. oryzae*. Consistently, principal component analysis of these TE insertion sites displayed a similar pattern (**Fig. 4c**). Considering that the TEs were largely amplified recently in the *M. oryzae* genome (**Fig. 1**), we presumed that the recent high activity of TEs has been one of the major forces driving population divergence of *M. oryzae*.

POT2 and Mg-SINE are critical for the divergence of Clade2 and Clade3 isolates

TEs are able to exert either beneficial or deleterious effects on host genomes, and the retention or elimination of TEs is largely determined by their impact on the host. Positive selection has been shown to drive the frequency of a TE locus to increase or decrease dramatically during a population bottleneck or in response to a new environment (33). We hypothesized that positive selection has fixed a portion of clade-specific TE insertion sites in the intra-clade isolates, contributing to the adaptive evolution of the clade isolates. To identify the clade-specific TE insertion sites, we screened out those that were present in at least 80% of the intra-clade isolates and absent in more than 80% of the other two clade isolates. A total of 11 Clade1-specific, 212 Clade2-specific, and 168 Clade3-specific TE insertion sites were identified, with the number of Clade1-specific TE insertion sites being too small for subsequent analysis. Enrichment analysis revealed that POT2 and Mg-SINE TE families were overrepresented in both Clade2 and Clade3 datasets, indicating that the insertion of these two TE families was positively selected for retention in subpopulations of *M. oryzae* (**Fig. 5a and b**).

227 We investigated the influence of clade-specific TE junctions on genes, which, for this
228 purpose, we referred to as clade-specific TE-associated genes. We identified a total of
229 238 Clade2-specific and 173 Clade3-specific TE-associated genes. Interestingly, less
230 than 10% of these genes overlapped, suggesting that the Clade2- and Clade3-specific
231 TEs have distinct targeted genes (**Fig. S2**). Gene ontology (GO) enrichment analysis
232 revealed that the Clade2- and Clade3-specific TE-associated genes were enriched
233 under similar GO terms (adjusted p-value < 0.01, **Fig. 5c, d**). Of note, the top
234 enriched term in both datasets was 'GO:0004497, monooxygenase activity,' which is
235 correlated to cytochrome P450s on the fungal genome. We further validated the GO
236 enrichment results by performing enrichment analysis for these genes using the Pfam
237 database (**Table S7**). Several previous studies have shown that fungal P450s possess
238 detoxifying functions towards compounds produced by host plants during pathogen
239 infection, thereby enhancing the fitness of the pathogenic fungus to specific host
240 genotypes (27, 34-38). Therefore, we suggest that TE-induced variations in different
241 members of P450s partially contribute to the differential pathogenicity of the two
242 clade isolates.

243 **Genes associated with TE are significantly lower expressed**

244 Previous studies have demonstrated that transposable elements (TEs) were able to
245 affect gene expression by inserting into gene promoters or intragenic regions (39-41).
246 Therefore, we investigated whether TE insertion polymorphisms could shape the gene
247 expression networks in *M. oryzae* populations. To this end, we selected 16 isolates
248 and extracted total RNA for sequencing (**Table S8**). We identified 2,282 genes
249 targeted by 4,236 TE insertion sites that were polymorphic between the 16 isolates.
250 We then grouped the genes based on whether they contained TE insertion sites
251 (TE-present or TE-absent) and compared the expression levels between these two
252 groups. We observed that the TE-present gene group displayed significantly lower
253 expression levels than the TE-absent gene group (**Fig. 6a**), suggesting that TEs have
254 negative effects on the expression of their target genes. We have identified 131 genes

that exhibited clade-specific expression patterns. Furthermore, based on the expression levels of these genes, we were able to divide 16 isolates into three distinct clusters using principal component analysis. These findings suggest that transposable element-mediated gene regulation had a profound impact on population divergence (**Fig. 6b and c**). Interestingly, we found that only genes with insertion polymorphisms of POT2 and Mg-SINE displayed transcriptional suppression, while other TE families displayed little impact on the expression of their target genes. Our results suggest that the insertion polymorphisms of POT2 and Mg-SINE were able to shape the gene expression network of *M. oryzae* by inducing transcriptional suppression.

DISCUSSION

The high genomic plasticity and rapid evolution of the plant pathogen *M. oryzae* presents a significant challenge for rice blast disease control. Previous studies have identified transposons as a major driving force for the adaptive evolution of fungal pathogen genomes (9). Insertion polymorphisms of transposable elements (TEs) have been shown to lead to genomic instability, increased chromosomal recombination, and accelerated gene evolution (13, 42). However, the precise role of TE dynamics during *M. oryzae* evolution has remained poorly understood. To address this, we conducted a population-scaled TE analysis of *M. oryzae* to investigate how TE insertion polymorphisms contributed in shaping population structure and divergence.

We employed two methods to assess the activity of transposable elements (TEs) in the genome of *M. oryzae*. First, we estimated the insertion time of LTR-RTs and secondly, we calculated the Kimura 2-parameter genetic distance for the 11 most abundant TE families. Both analyses indicate that TEs have undergone recent amplification in the *M. oryzae* genome and have remained highly active. Previous studies have suggested that the insertion of TEs having caused gene presence/absence variations to be the main evolutionary mechanism driving the divergence of host-specific *M. oryzae*

282 lineages (12). In this study, we investigated the abundance and activity of 11 TEs in
283 pathogens from wheat and grass infecting lineages, and found that these 11 TEs have
284 undergone specific expansion in *M. oryzae* rice isolates and have recently become
285 inactive in non-rice isolates. Based on these results, we propose that the recent burst
286 of TE activity is the primary factor responsible for the divergence of *M. oryzae* from
287 rice and other host infecting lineages.

288 In order to elucidate the role of TEs in genome evolution and divergence, we
289 conducted a genome-wide survey of TE insertion sites in 90 *M. oryzae* rice isolates
290 that were previously published (17). We utilized PoPoolationTE2, which is a validated
291 method for estimating TE insertion frequency in populations (43). The TE insertion
292 junctions varied widely among the 90 isolates, indicating the highly dynamic and
293 active nature of TEs in *M. oryzae* populations. Consistent with prior research, we
294 found that the distribution of TE junctions across the genome was not evenly
295 distributed, and different TE families or superfamilies displayed preferences for
296 insertion into distinct genomic regions (6, 13, 15).

297 TEs have been shown to be a major contributor in causing genomic variations.
298 Previous studies have reported that TE insertions in or around *Avr* genes in *M. oryzae*
299 were able to lead to transcriptional silencing and loss of avirulence function (10,
300 44-48). Our analysis revealed that TE junctions are frequently observed in the
301 flanking regions of genes encoding secreted proteins (SPs), and the proportion of
302 genes encoding SPs with TE insertions is higher than that of non-SPs. This is
303 consistent with previous findings that SPs are enriched in repeat-rich regions and are
304 prone to rapid evolution (4, 49). We propose that the variations in SPs, induced by TE
305 insertion polymorphisms, is able to facilitate adaptive evolution of *M. oryzae*.
306 Presence/absence variations in genes resulting from TE insertions have been
307 identified as the major force driving divergence of host-specific *M. oryzae* lineages
308 (12). Consistent with this, we found that genes exhibiting high gain/loss
309 polymorphisms in the 90 isolates were preferentially located near TE junctions,

suggesting that TE-mediated presence/absence variation constitutes a significant mechanism underlying differentiation of host-specific or intra-species *M. oryzae* isolates.

Through the comparison of TE junctions in 90 *M. oryzae* rice isolates, we observed a clustering pattern that was similar to the three previously defined *M. oryzae* clades. This led us to investigate the potential role of TEs in *M. oryzae* population divergence. We constructed a hierarchical tree based on the TE junctions and found that it closely resembled the phylogenetic tree constructed using genome-wide SNPs. Principal component analysis of the TE junctions also yielded a similar clustering pattern. These findings suggested that the transposition of TEs was strongly associated with *M. oryzae* population divergence. Given that the majority of TEs in the *M. oryzae* genome were generated recently, and that the population divergence occurred not long ago, we hypothesize the large-scale amplification of TEs to be a major driving force contributing to the *M. oryzae* population divergence.

TE loci that undergo positive selection during evolution will be retained and will exhibit high frequencies in a population. Based on the hierarchical tree and PCA results, we postulate that the intra-clade isolates have a fixed set of TE insertion sites that are specific to each clade. These clade-specific TE insertion sites are then able to serve as molecular markers to distinguish isolates belonging to the three clades. Interestingly, we found that POT2 and Mg-SINE were significantly enriched in clade-specific TE insertion sites, suggesting that these clade-specific insertions of POT2 and Mg-SINE were beneficial for the adaptive evolution of clade isolates. Furthermore, we noted that cytochrome P450 proteins were overrepresented in both Clade2- and Clade3-specific TE-targeted genes. Given the essential roles of P450s in detoxifying phytoalexins produced by host plants, we hypothesize that the differential pathogenicity of clade isolates may be partially due to the variations in P450s induced by clade-specific TE insertions.

TEs that insert in or near genes have been reported to induce silencing of gene

338 expression through the formation of heterochromatic regions (9, 41). To investigate
339 the correlation between TE insertion and gene expression regulation in *M. oryzae*, we
340 conducted RNA sequencing on 16 isolates from the three clades. Our analysis
341 revealed that genes containing a TE insertion displayed significantly lower expression
342 compared to genes without TE insertion. Notably, only POT2 and Mg-SINE insertions
343 led to substantial suppression of gene expression. Therefore, we hypothesized that
344 clade-specific insertions of POT2 and Mg-SINE contribute to the adaptive evolution
345 of clade isolates by regulating expression of specific genes.

346 Through a comprehensive analysis of TEs in populations of the rice-infecting fungus
347 *M. oryzae*, we have revealed the crucial roles played by recent TE insertional bursts in
348 the adaptive evolution and diversification of this fungal lineage. Our findings
349 highlight the significance of TE-mediated genetic changes in the regulation of gene
350 expression, which in turn contributes to clade divergence.
351

352 # MATERIALS AND METHODS

353 **RNA extraction, library generation, and sequencing**

354 The fungal strains were cultured in liquid CM medium by incubation at 28℃ under
355 shaking at 110 rpm for three days. The mycelium was filtered, washed with
356 double-deionized water, and dried before being ground in liquid nitrogen. Ground
357 samples were transferred into DNase/RNase-free Eppendorf tubes, suspended in 1 mL
358 Trizol, and vortexed vigorously. To eliminate proteins, 200 μL of chloroform was
359 added to the mixture, which was then shaken for 15 s. After centrifugation at 12,000
360 rpm for 15 min at 4℃, 400 μL of the supernatant was collected and mixed with 400
361 μL of cold isopropanol. The mixture was kept at -20℃ for at least two hours, then
362 centrifuged at 12,000 rpm for 15 min at 4℃. The supernatant was discarded, and the
363 precipitates were washed with 1 mL of 70% alcohol and centrifuged at 12,000 rpm for
364 5 min at 4℃. After air drying for 5 min at room temperature, the pellets were diluted

365 with 54 μL of DNase/RNase-free deionized water and treated with 2 μL of DNase
366 I at 37℃ for 30 min. The mixture was brought up to 800 μL with RNase-free water,
367 followed by the addition of an equal volume of chloroform. After gentle mixing, the
368 mixture was centrifuged at 12,000 rpm for 15 min at 4℃. About 500 μL of the
369 supernatant was collected and mixed with 500 μL of cold isopropanol. The mixture
370 was then kept at -20℃ for more than three hours, followed by centrifugation at
371 12,000 rpm for 15 min at 4℃. The precipitates were washed with 1 mL of 70%
372 alcohol, air-dried for 5 min at room temperature, and eluted with DNase/RNase-free
373 deionized water. The RNA samples were then stored at -80℃ for RNA sequencing
374 analysis.

375 **Estimation of TE activity**

376 LTR-Finder (50), LTR-harvest (51) and LTR-Retriever (52) were used for *de novo*
377 identification of full-length LTR-RTs and estimation of insertion time. Information on
378 TE classification is based on previous research (3) and the conserved domains of TE
379 consensus sequences were predicted by Conserved Domain Database (53). The
380 Kimura 2-Parameter genetic distances (k-values) of TE fragments were calculated by
381 RepeatMasker (54) with option '-a'.

382 **Prediction of secreted proteins and PAV genes**

383 To identify putative secreted proteins, several criteria were employed, including the
384 presence of a signal peptide cleavage site, the absence of a transmembrane domain,
385 and a protein length of less than 400 amino acids. SignalP 4.0 and TMHMM 2.0 were
386 utilized for signal peptide and transmembrane domain prediction, respectively (55,
387 56). The transcript sequences of the 70-15 reference genome were aligned to the
388 previously published assemblies of 90 isolates using NCBI-blastn with default
389 parameters. Genes that exhibited more than 90% similarity with no gaps longer than
390 50 bp when compared to the assemblies were marked as "present". Genes that did not
391 meet these criteria were marked as "absent". PAV (presence-absence variation) genes

were defined as those that were absent in more than five isolates, while core genes were defined as those present in all isolates.

Identification of TE insertion sites in *M. oryzae* populations

The pair-end reads were mapped to the *M. oryzae* 70-15 reference genome using BWA (57) with default parameters. The alignment files were exported to PoPoolation2 (58) with default parameter to identify TE insertion site for each isolate. The TE insertion sites with score less than 0.3 were filtered, and the insertion sites located within 50 bp were merged.

Construction of TE hierarchical tree, principal component analysis

The insertion sites that are present in at least 5 isolates were used for constructing hierarchical tree and principal component analysis (PCA). The TE insertion sites in presence/absence variation format were exported to the R package 'hclust' (59) for construction of TE hierarchical tree. Vcftools (60) and Plink (61) were used for PCA.

Functional enrichment analysis

GO annotation information of *M. oryzae* was obtained from the JGI database (62). Conserved domains of *M. oryzae* protein sequences were predicted using the Pfam database (63). Fisher right-tailed test was used for enrichment analysis and a cutoff p-value less than 0.05 was used to define significant enrichment.

RNA-seq analysis

The clean reads were mapped to the 70-15 reference genome using hisat2 v2.2.1 (64) with default parameters. Stingtie v2.1.4 was used for expression quantification (65). Gene expressions were normalized with transcripts per million (TPM). The expression of each gene in the isolates with or without TE insertion were counted, averaged and compared. Two-tailed Wilcoxon paired test was used to estimate the significance of expression difference.

DATA AVAILABILITY

All high-throughput sequencing data generated in this study are accessible at NCBI's Gene Expression Omnibus (GEO) via GEO Series accession number GSE205351 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE205351).

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (32272513 and 32172365). Open Access funding provided by Fujian Agriculture and Forestry University.

REFERENCES

1. **Wilson RA, Talbot NJ.** 2009. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. Nature reviews. Microbiology **7:**185-195.
2. **Asibi AE, Chai Q, Coulter JA.** 2019. Rice blast: A disease with implications for global food security. Agronomy **9:**451.
3. **Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW.** 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature **434:**980-986.
4. **Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, Jiang H, Zhang D, Miao C, Tang H, Zhang J, Lu G, Ming R, Norvienyeku J, Wang B, Wang Z.** 2017. PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in Magnaporthe oryzae. Molecular plant **10:**1465-1468.
5. **Chuong EB, Elde NC, Feschotte C.** 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nature reviews. Genetics **18:**71-86.
6. **Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS.** 2018. Ten things you should know about transposable elements. Genome biology **19:**1-12.
7. **Gout L, Kuhn ML, Vincenot L, Bernard-Samain S, Cattolico L, Barbetti M, Moreno-Rico O, Balesdent MH, Rouxel T.** 2007. Genome structure impacts molecular evolution at the AvrLm1 avirulence locus of the plant pathogen Leptosphaeria maculans. Environmental microbiology **9:**2978-2992.
8. **Raffaele S, Kamoun S.** 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nature reviews. Microbiology **10:**417-430.
9. **Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, Thomma BP.** 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome research **26:**1091-1100.

456 10. **Kang S, Lebrun MH, Farrall L, Valent B.** 2001. Gain of virulence caused by insertion of a Pot3
457 transposon in a Magnaporthe grisea avirulence gene. Molecular plant-microbe interactions :
458 MPMI **14:**671-674.
459 11. **Whisson S, Vetukuri R, Avrova A, Dixelius C.** 2012. Can silencing of transposons contribute to
460 variation in effector gene expression in Phytophthora infestans? Mobile genetic elements
461 **2:**110-114.
462 12. **Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa**
463 **Y, Cano LM, Kamoun S, Terauchi R.** 2016. Host specialization of the blast fungus
464 Magnaporthe oryzae is associated with dynamic gain and loss of genes linked to transposable
465 elements. BMC genomics **17:**370.
466 13. **Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, Dean RA.** 2006. The role of
467 transposable element clusters in genome evolution and loss of synteny in the rice blast
468 fungus Magnaporthe oryzae. Genome biology **7:**R16.
469 14. **Castanera R, Lopez-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J,**
470 **Perez G, Pisabarro AG, Grigoriev IV, Stajich JE, Ramirez L.** 2016. Transposable Elements
471 versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional
472 Profiles. PLoS genetics **12:**e1006108.
473 15. **Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K.** 2019. Transposable
474 elements contribute to fungal genes and impact fungal lifestyle. Scientific reports **9:**4307.
475 16. **Shirke MD, Mahesh HB, Gowda M.** 2016. Genome-Wide Comparison of Magnaporthe
476 Species Reveals a Host-Specific Pattern of Secretory Proteins and Transposable Elements. PloS
477 one **11:**e0162458.
478 17. **Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, Lin L, Lin Y, Somai R, Lu L, Zhang W, Chen J,**
479 **Hong Y, Chen X, Wang B, Shen W-C, Lu G, Norvienyeku J, Ebbole DJ, Wang Z.** 2018.
480 Population genomic analysis of the rice blast fungus reveals specific events associated with
481 expansion of three main clades. The ISME Journal.
482 18. **Zhong Z, Lin L, Zheng H, Bao J, Chen M, Zhang L, Tang W, Ebbole DJ, Wang Z.** 2020.
483 Emergence of a hybrid PKS-NRPS secondary metabolite cluster in a clonal population of the
484 rice blast fungus Magnaporthe oryzae. Environmental microbiology.
485 19. **Sanchez E, Asano K, Sone T.** 2011. Characterization of Inago1 and Inago2 retrotransposons in
486 Magnaporthe oryzae. Journal of general plant pathology **77:**239-242.
487 20. **Kachroo P, Leong SA, Chattoo BB.** 1994. Pot2, an inverted repeat transposon from the rice
488 blast fungus Magnaporthe grisea. Molecular & general genetics : MGG **245:**339-348.
489 21. **Farman ML, Tosa Y, Nitta N, Leong SA.** 1996. MAGGY, a retrotransposon in the genome of the
490 rice blast fungus Magnaporthe grisea. Molecular & general genetics : MGG **251:**665-674.
491 22. **Dobinson KF, Harris RE, Hamer JE.** 1993. Grasshopper, a long terminal repeat (LTR)
492 retroelement in the phytopathogenic fungus Magnaporthe grisea. Molecular plant-microbe
493 interactions : MPMI **6:**114-126.
494 23. **Kimura M.** 1980. A simple method for estimating evolutionary rates of base substitutions
495 through comparative studies of nucleotide sequences. Journal of molecular evolution
496 **16:**111-120.
497 24. **Chalopin D, Naville M, Plard F, Galiana D, Volff JN.** 2015. Comparative analysis of
498 transposable elements highlights mobilome diversity and evolution in vertebrates. Genome
499 biology and evolution **7:**567-580.

500 25. **Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, Tang H, Farman M, Cook D, White**
501 **FF, Valent B, Liu S.** 2019. Effector gene reshuffling involves dispensable mini-chromosomes in
502 the wheat blast fungus. PLoS genetics **15:**e1008272.
503 26. **Gómez Luciano LB, Tsai IJ, Chuma I, Tosa Y, Chen Y-H, Li J-Y, Li M-Y, Lu M-YJ, Nakayashiki H,**
504 **Li W-H.** 2019. Blast fungal genomes show frequent chromosomal changes, gene gains and
505 losses, and effector gene turnover. Molecular biology and evolution **36:**1148-1161.
506 27. **Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, Lin Y, Wu X, Cai Z, Zhang Q, Lin X,**
507 **Hong Y, Huang J, Xu L, Zhang H, Chen L, Tang W, Zheng H, Chen X, Wang Y, Lian B, Zhang L,**
508 **Tang H, Lu G, Ebbole DJ, Wang B, Wang Z.** 2016. Directional Selection from Host Plants Is a
509 Major Force Driving Host Specificity in Magnaporthe Species. Scientific reports **6:**25591.
510 28. **Dong S, Raffaele S, Kamoun S.** 2015. The two-speed genomes of filamentous pathogens:
511 waltz with plants. Current opinion in genetics & development **35:**57-65.
512 29. **Seidl MF, Thomma B.** 2017. Transposable Elements Direct The Coevolution between Plants
513 and Microbes. Trends in genetics : TIG **33:**842-851.
514 30. **Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E.** 2021. Dynamics of transposable
515 elements in recently diverged fungal pathogens: lineage-specific transposable element
516 content and efficiency of genome defenses. G3 **11:**jkab068.
517 31. **Plissonneau C, Hartmann FE, Croll D.** 2018. Pangenome analyses of the wheat pathogen
518 Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. BMC
519 Biol **16:**5.
520 32. **Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D.** 2020. A 19-isolate
521 reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. BMC
522 Biol **18:**12.
523 33. **Torres DE, Oggenfuss U, Croll D, Seidl MF.** 2020. Genome evolution in fungal plant pathogens:
524 looking beyond the two-speed genome model. Fungal Biology Reviews.
525 34. **Schuler MA.** 2015. P450s in plants, insects, and their fungal pathogens, p. 409-449,
526 Cytochrome P450. Springer.
527 35. **Ide M, Ichinose H, Wariishi H.** 2012. Molecular identification and functional characterization
528 of cytochrome P450 monooxygenases from the brown-rot basidiomycete Postia placenta.
529 Archives of microbiology **194:**243-253.
530 36. **Nazir KH, Ichinose H, Wariishi H.** 2011. Construction and application of a functional library of
531 cytochrome P450 monooxygenases from the filamentous fungus Aspergillus oryzae. Applied
532 and environmental microbiology **77:**3147-3150.
533 37. **Doddapaneni H, Chakraborty R, Yadav JS.** 2005. Genome-wide structural and evolutionary
534 analysis of the P450 monooxygenase genes (P450ome) in the white rot fungus
535 Phanerochaete chrysosporium: evidence for gene duplications and extensive gene clustering.
536 BMC genomics **6:**92.
537 38. **Zhang S, Widemann E, Bernard G, Lesot A, Pinot F, Pedrini N, Keyhani NO.** 2012. CYP52X1,
538 representing new cytochrome P450 subfamily, displays fatty acid hydroxylase activity and
539 contributes to virulence and growth on insect cuticular substrates in entomopathogenic
540 fungus Beauveria bassiana. The Journal of biological chemistry **287:**13477-13486.
541 39. **Santana M, Queiroz M.** 2015. Transposable elements in fungi: a genomic approach. Scientific
542 J Genetics Gen Ther **1:**012.
543 40. **Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA,**

544 **Sanchez-Vallet A.** 2018. Transposable element insertions shape gene regulation and melanin
545 production in a fungal pathogen of wheat. BMC Biol **16:**78.
546 41. **Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schubeler D, Selker EU.**
547 2009. Relics of repeat-induced point mutation direct heterochromatin formation in
548 Neurospora crassa. Genome research **19:**427-437.
549 42. **Chadha S, Sharma M.** 2014. Transposable elements as stress adaptive capacitors induce
550 genomic instability in fungal pathogen Magnaporthe oryzae. PloS one **9:**e94415.
551 43. **Kofler R, Gomez-Sanchez D, Schlotterer C.** 2016. PoPoolationTE2: Comparative Population
552 Genomics of Transposable Elements Using Pool-Seq. Molecular biology and evolution
553 **33:**2759-2764.
554 44. **Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, Zhang Z, Zhao Q, Feng Q, Zhang H, Wang Z, Wang G,**
555 **Han B, Wang Z, Zhou B.** 2009. The Magnaporthe oryzae avirulence gene AvrPiz-t encodes a
556 predicted secreted protein that triggers the immunity in rice mediated by the blast resistance
557 gene Piz-t. Molecular plant-microbe interactions : MPMI **22:**411-420.
558 45. **Zhang S, Wang L, Wu W, He L, Yang X, Pan Q.** 2015. Function and evolution of Magnaporthe
559 oryzae avirulence gene AvrPib responding to the rice blast resistance gene Pib. Scientific
560 reports **5:**11642.
561 46. **Chen C, Chen M, Hu J, Zhang W, Zhong Z, Jia Y, Allaux L, Fournier E, Tharreau D, Wang G-L.**
562 2014. Sequence variation and recognition specificity of the avirulence gene AvrPiz-t in
563 Magnaporthe oryzae field populations. Fungal Genomics & Biology **4**.
564 47. **Miki S, Matsui K, Kito H, Otsuka K, Ashizawa T, Yasuda N, Fukiya S, Sato J, Hirayae K, Fujita Y,**
565 **Nakajima T, Tomita F, Sone T.** 2009. Molecular cloning and characterization of the AVR-Pia
566 locus from a Japanese field isolate of Magnaporthe oryzae. Molecular plant pathology
567 **10:**361-374.
568 48. **Hu ZJ, Huang YY, Lin XY, Feng H, Zhou SX, Xie Y, Liu XX, Liu C, Zhao RM, Zhao WS, Feng CH,**
569 **Pu M, Ji YP, Hu XH, Li GB, Zhao JH, Zhao ZX, Wang H, Zhang JW, Fan J, Li Y, Peng YL, He M, Li**
570 **DQ, Huang F, Peng YL, Wang WM.** 2022. Loss and Natural Variations of Blast Fungal
571 Avirulence Genes Breakdown Rice Resistance Genes in the Sichuan Basin of China. Frontiers
572 in plant science **13:**788876.
573 49. **Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, Ponaya A, Xiao G, Li J, Li C, Song MY, Cumagun CJ,**
574 **Deng Q, Lu G, Jeon JS, Naqvi NI, Zhou B.** 2015. Comparative genomics identifies the
575 Magnaporthe oryzae avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in
576 rice. The New phytologist **206:**1463-1475.
577 50. **Xu Z, Wang H.** 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR
578 retrotransposons. Nucleic acids research **35:**W265-268.
579 51. **Ellinghaus D, Kurtz S, Willhoeft U.** 2008. LTRharvest, an efficient and flexible software for de
580 novo detection of LTR retrotransposons. BMC bioinformatics **9:**18.
581 52. **Ou S, Jiang N.** 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification
582 of Long Terminal Repeat Retrotransposons. Plant physiology **176:**1410-1422.
583 53. **Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI,**
584 **Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ,**
585 **Marchler-Bauer A.** 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic
586 acids research **48:**D265-D268.
587 54. **Tempel S.** 2012. Using and understanding RepeatMasker. Methods in molecular biology

588 **859:**29-51.
589 55. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal
590 peptides from transmembrane regions. Nature methods **8:**785-786.
591 56. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein
592 topology with a hidden Markov model: application to complete genomes. Journal of
593 molecular biology **305:**567-580.
594 57. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler
595 transform. Bioinformatics **25:**1754-1760.
596 58. **Kofler R, Pandey RV, Schlotterer C.** 2011. PoPoolation2: identifying differentiation between
597 populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics
598 **27:**3435-3436.
599 59. **Galili T.** 2015. dendextend: an R package for visualizing, adjusting and comparing trees of
600 hierarchical clustering. Bioinformatics **31:**3718-3720.
601 60. **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,**
602 **Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G.** 2011. The variant
603 call format and VCFtools. Bioinformatics **27:**2156-2158.
604 61. **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de**
605 **Bakker PI, Daly MJ, Sham PC.** 2007. PLINK: a tool set for whole-genome association and
606 population-based linkage analyses. American journal of human genetics **81:**559-575.
607 62. **Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV,**
608 **Dubchak I.** 2014. The genome portal of the Department of Energy Joint Genome Institute:
609 2014 updates. Nucleic acids research **42:**D26-31.
610 63. **Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE,**
611 **Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A.** 2021. Pfam: The protein families
612 database in 2021. Nucleic acids research **49:**D412-D419.
613 64. **Kim D, Paggi JM, Park C, Bennett C, Salzberg SL.** 2019. Graph-based genome alignment and
614 genotyping with HISAT2 and HISAT-genotype. Nature biotechnology **37:**907-915.
615 65. **Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL.** 2015. StringTie
616 enables improved reconstruction of a transcriptome from RNA-seq reads. Nature
617 biotechnology **33:**290-295.
618
619
620

Figure legends

FIG 1 a) The Kimura 2-Parameter genetic distance of 11 TE families in three different *Magnaporthe* species were shown in density plot. **b, c)** The k-value proportions of 11 TE families in *M. oryzae* rice isolates (**b**) and non-rice isolates (**c**). The numbers of TE fragment were marked at the top of bars.

FIG 2 The circos diagram (from outer to inner) displays chromosomes, gene distribution, distribution of DNA transposon, Non-LTR-RT and LTR-RT junctions. The histogram shows the insertion site numbers of different TE families. The non-transparent and transparent colors represent the numbers of isolate-specific and isolate-shared TE junctions, respectively

FIG 3 The distance of TEs to the 5' and 3' ends of secreted proteins and non-secreted proteins (**a**) or core genes and PAV genes (**b**).

FIG 4 a) The heatmap showing the identity of TE insertion site between each two isolates. **b)** The hierarchical tree built by using the TE insertion sites of *M. oryzae* population. Three distinct clades were marked in red, blue and green colors respectively. **c)** Principal component analysis using the TE insertion sites.

FIG 5 a, b) Enrichment analysis of 11 TE families in Clade2- (**a**) and Clade3-specific (**b**) insertion sites. Fisher's exact test was used for significance test. **c**) GO enrichment analysis for the clade2- (**c**) and clade3-specific (**d**) TE-associated genes.

FIG 6 a) Comparison of gene expressions between the isolates present or absent with TE insertion. **b)** PCA clustering of 16 isolates based on expression level of 131 clade-specific DEGs. **c)** Heatmap showing expression level of 131 clade-specific DEGs in 16 isolates.

TABLE 1 The numbers of secreted protein and non-secreted protein that have TE insertion in 1-kb flanking regions or within gene.

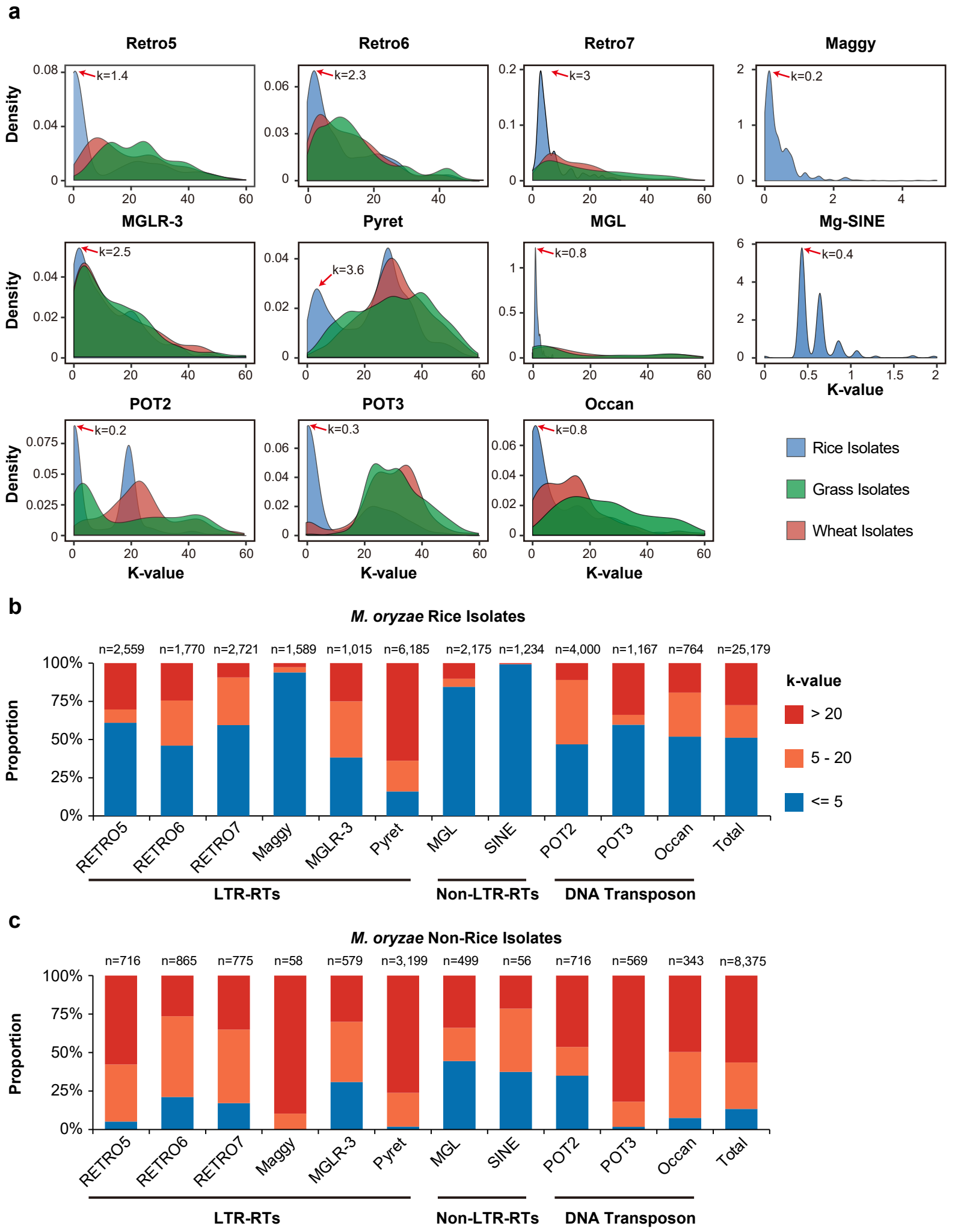
SUPPLEMENTAL MATERIAL

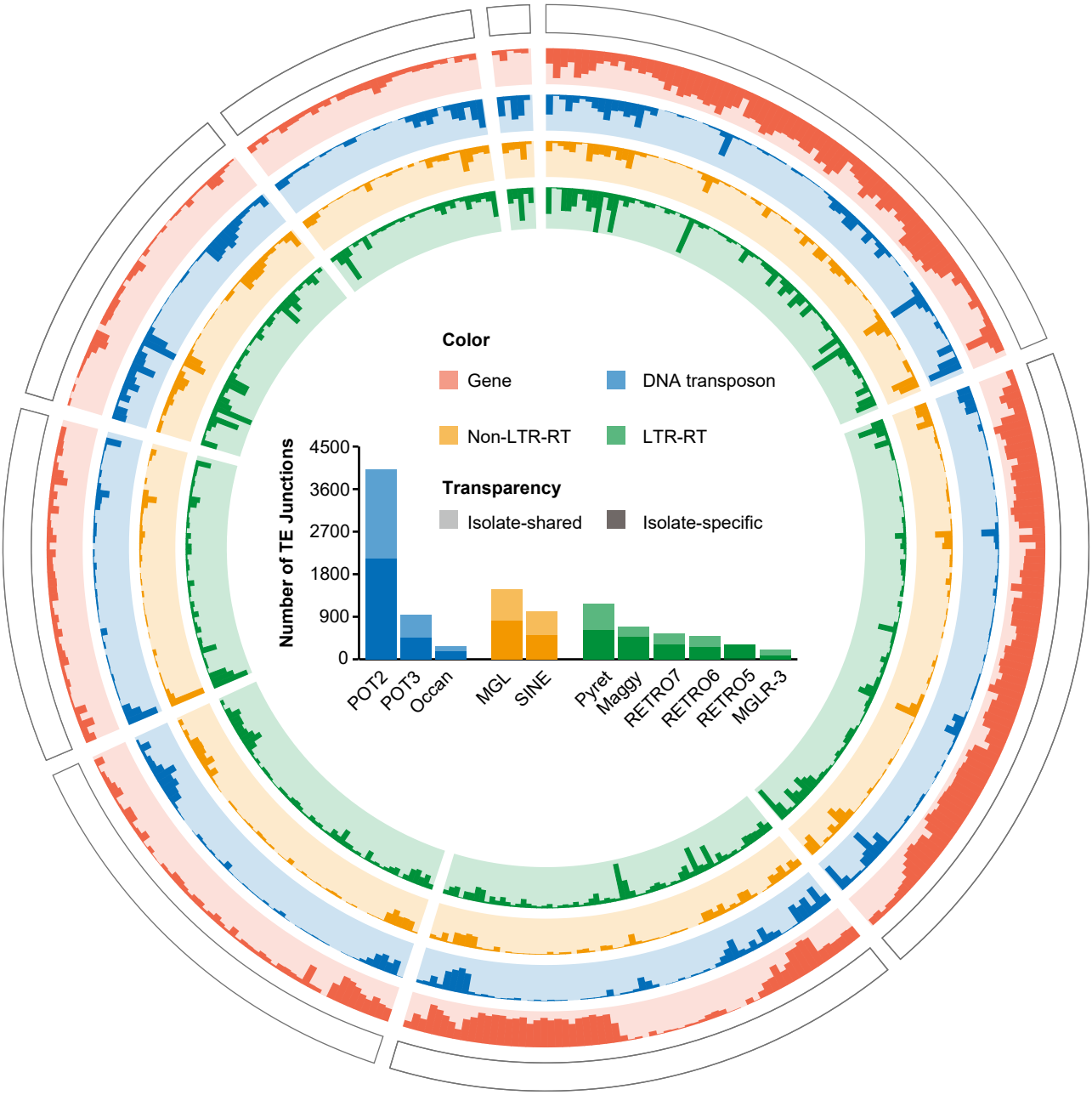
FIG S1 Schematic of the 11 most abundant TE families on the genome of the *M. oryzae* rice isolate. The conserved domains in the relative location of the TE consensus were highlighted using different colors.

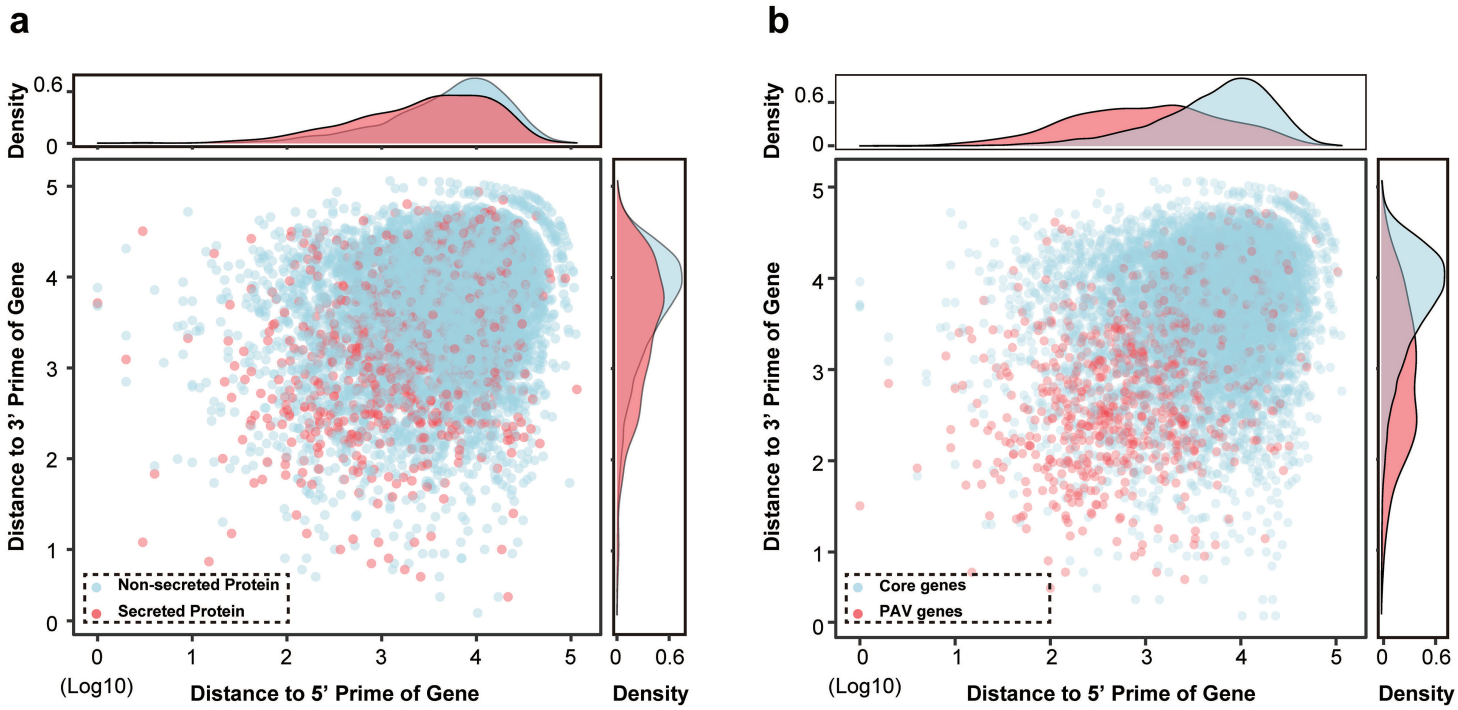
FIG S2 The venn chart compares the TE insertion loci of 90 *M. oryzae* rice isolates and two *S. viridis* isolates.

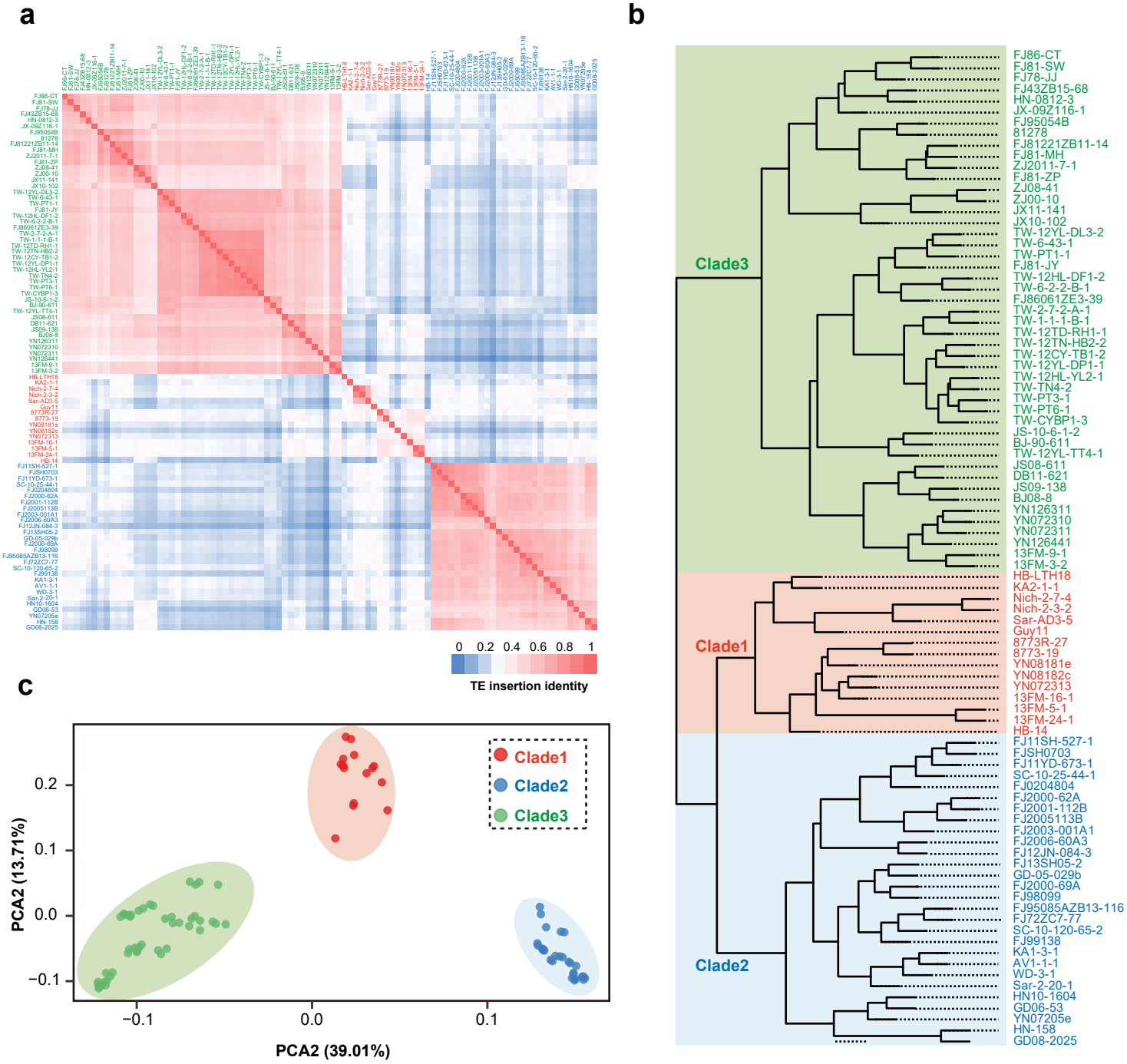
FIG S3 The venn chart displays the intersection of clade2-specific and clade3-specific TE-associated genes.

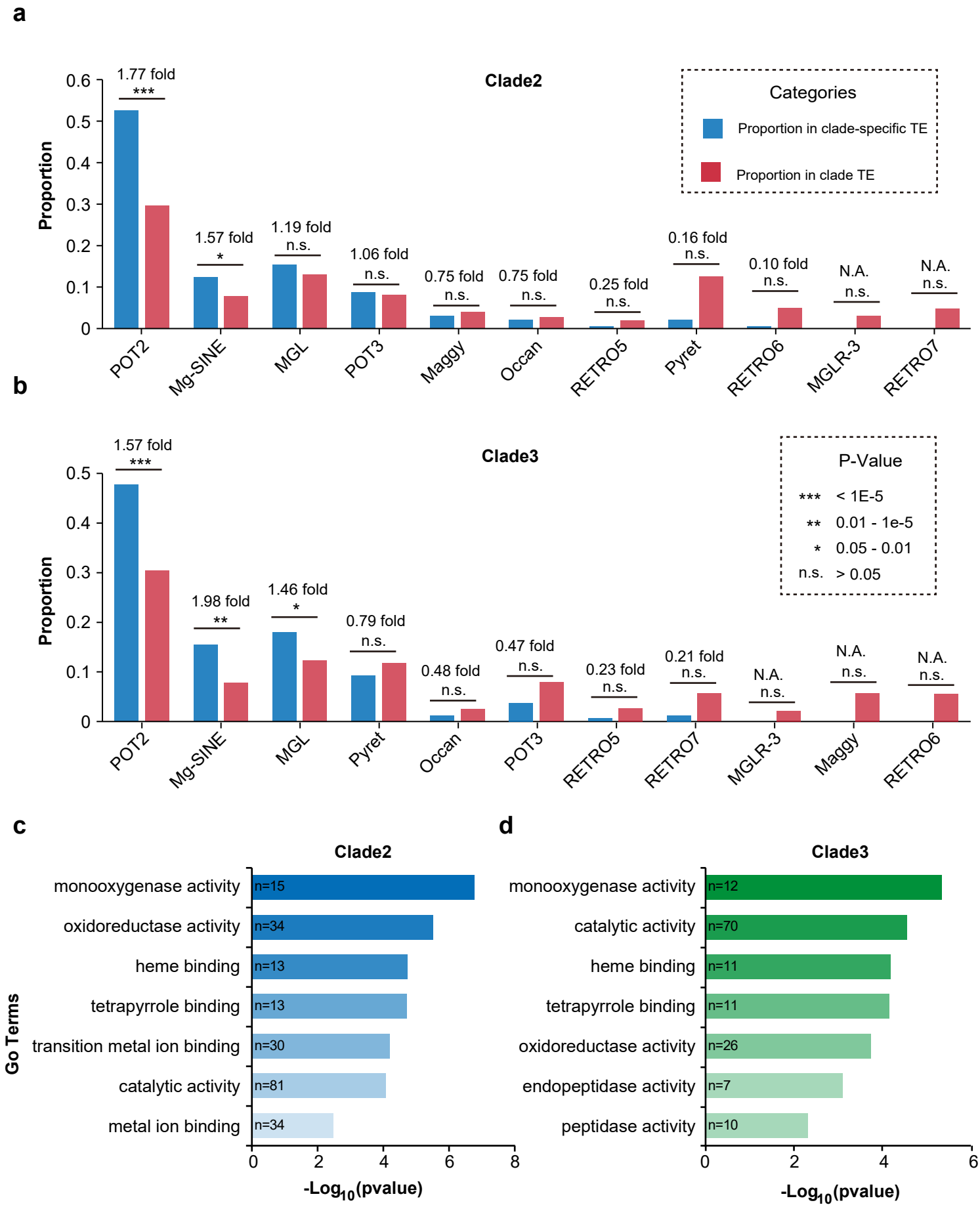
662 **TABLE S1** The proportions of 11 TE families on the genomes of seven *M. oryzae*
663 rice isolates.
664
665 **TABLE S2** The proportions of 11 TE families on the genomes of a *M. oryzae* wheat
666 isolate and three grass isolates.
667
668 **TABLE S3** Validation of 17 TE insertion sites by PCR amplification.
669
670 **TABLE S4** Insertion preference of various TEs in different genomic regions.
671
672 **TABLE S5** Enrichment analysis of TEs in the promoter of secreted proteins.
673
674 **TABLE S6** Identity of TE insertion sites between intra- or inter-clade isolates.
675
676 **TABLE S7** Enrichment analysis of the clade2- and clade3-specific TE-associated
677 genes using conserved domains.
678
679 **TABLE S8** GEO information of 16 RNA seq samples.
680
681


a
Retro5
Retro6
Retro7
Maggy
MGLR-3
Pyret
MGL
Mg-SINE
POT2
POT3
Occan
k=1.4
k=2.3
k=3
k=0.2
k=2.5
k=3.6
k=0.8
k=0.4
k=0.2
k=0.3
k=0.8
Density
K-value
Rice Isolates
Grass Isolates
Wheat Isolates
b
M. oryzae Rice Isolates
n=2,559
n=1,770
n=2,721
n=1,589
n=1,015
n=6,185
n=2,175
n=1,234
n=4,000
n=1,167
n=764
n=25,179
Proportion
k-value
> 20
5 - 20
<= 5
RETRO5
RETRO6
RETRO7
Maggy
MGLR-3
Pyret
MGL
SINE
POT2
POT3
Occan
Total
LTR-RTs
Non-LTR-RTs
DNA Transposon
c
M. oryzae Non-Rice Isolates
n=716
n=865
n=775
n=58
n=579
n=3,199
n=499
n=56
n=716
n=569
n=343
n=8,375
Proportion
RETRO5
RETRO6
RETRO7
Maggy
MGLR-3
Pyret
MGL
SINE
POT2
POT3
Occan
Total
LTR-RTs
Non-LTR-RTs
DNA Transposon


Color
Gene
DNA transposon
Non-LTR-RT
LTR-RT
Transparency
Isolate-shared
Isolate-specific
Number of TE Junctions
4500
3600
2700
1800
900
0
POT2
POT3
Occan
MGL
SINE
Pyret
Maggy
RETRO7
RETRO6
RETRO5
MGLR-3


a
Density
0.6
0
Distance to 3' Prime of Gene
5
4
3
2
1
0
Non-secreted Protein
Secreted Protein
0
1
2
3
4
5
(Log10)
Distance to 5' Prime of Gene
0
0.6
Density
b
Density
0.6
0
Distance to 3' Prime of Gene
5
4
3
2
1
0
Core genes
PAV genes
0
1
2
3
4
5
(Log10)
Distance to 5' Prime of Gene
0
0.6
Density


a
0
0.2
0.4
0.6
0.8
1
TE insertion identity
b
Clade3
Clade1
Clade2
FJ86-CT
FJ81-SW
FJ78-JJ
FJ43ZB15-68
HN-0812-3
JX-09Z116-1
FJ95054B
81278
FJ81221ZB11-14
FJ81-MH
ZJ2011-7-1
FJ81-ZP
ZJ08-41
ZJ00-10
JX11-141
JX10-102
TW-12YL-DL3-2
TW-6-43-1
TW-PT1-1
FJ81-JY
TW-12HL-DF1-2
TW-6-2-2-B-1
FJ86061ZE3-39
TW-2-7-2-A-1
TW-1-1-1-B-1
TW-12TD-RH1-1
TW-12TN-HB2-2
TW-12CY-TB1-2
TW-12YL-DP1-1
TW-12HL-YL2-1
TW-TN4-2
TW-PT3-1
TW-PT6-1
TW-CYBP1-3
JS-10-6-1-2
BJ-90-611
TW-12YL-TT4-1
JS08-611
DB11-621
JS09-138
BJ08-8
YN126311
YN072310
YN072311
YN126441
13FM-9-1
13FM-3-2
HB-LTH18
KA2-1-1
Nich-2-7-4
Nich-2-3-2
Sar-AD3-5
Guy11
8773R-27
8773-19
YN08181e
YN08182c
YN072313
13FM-16-1
13FM-5-1
13FM-24-1
HB-14
FJ11SH-527-1
FJSH0703
FJ11YD-673-1
SC-10-25-44-1
FJ0204804
FJ2000-62A
FJ2001-112B
FJ2005113B
FJ2003-001A1
FJ2006-60A3
FJ12JN-084-3
FJ13SH05-2
GD-05-029b
FJ2000-69A
FJ98099
FJ95085AZB13-116
FJ72ZC7-77
SC-10-120-65-2
FJ99138
KA1-3-1
AV1-1-1
WD-3-1
Sar-2-20-1
HN10-1604
GD06-53
YN07205e
HN-158
GD08-2025
c
Clade1
Clade2
Clade3
0.2
0.1
0.0
−0.1
PCA2 (13.71%)
−0.1
0.0
0.1
PCA2 (39.01%)


a
Clade2
Proportion
0.6
0.5
0.4
0.3
0.2
0.1
0
1.77 fold
***
1.57 fold
*
1.19 fold
n.s.
1.06 fold
n.s.
0.75 fold
n.s.
0.75 fold
n.s.
0.25 fold
n.s.
0.16 fold
n.s.
0.10 fold
n.s.
N.A.
n.s.
N.A.
n.s.
POT2
Mg-SINE
MGL
POT3
Maggy
Occan
RETRO5
Pyret
RETRO6
MGLR-3
RETRO7
Categories
Proportion in clade-specific TE
Proportion in clade TE
b
Clade3
Proportion
0.5
0.4
0.3
0.2
0.1
0
1.57 fold
***
1.98 fold
**
1.46 fold
*
0.79 fold
n.s.
0.48 fold
n.s.
0.47 fold
n.s.
0.23 fold
n.s.
0.21 fold
n.s.
N.A.
n.s.
N.A.
n.s.
N.A.
n.s.
POT2
Mg-SINE
MGL
Pyret
Occan
POT3
RETRO5
RETRO7
MGLR-3
Maggy
RETRO6
P-Value
*** < 1E-5
** 0.01 - 1e-5
* 0.05 - 0.01
n.s. > 0.05
c
Clade2
Go Terms
monooxygenase activity
oxidoreductase activity
heme binding
tetrapyrrole binding
transition metal ion binding
catalytic activity
metal ion binding
n=15
n=34
n=13
n=13
n=30
n=81
n=34
0
2
4
6
8
-Log10(pvalue)
d
Clade3
monooxygenase activity
catalytic activity
heme binding
tetrapyrrole binding
oxidoreductase activity
endopeptidase activity
peptidase activity
n=12
n=70
n=11
n=11
n=26
n=7
n=10
0
2
4
6
-Log10(pvalue)

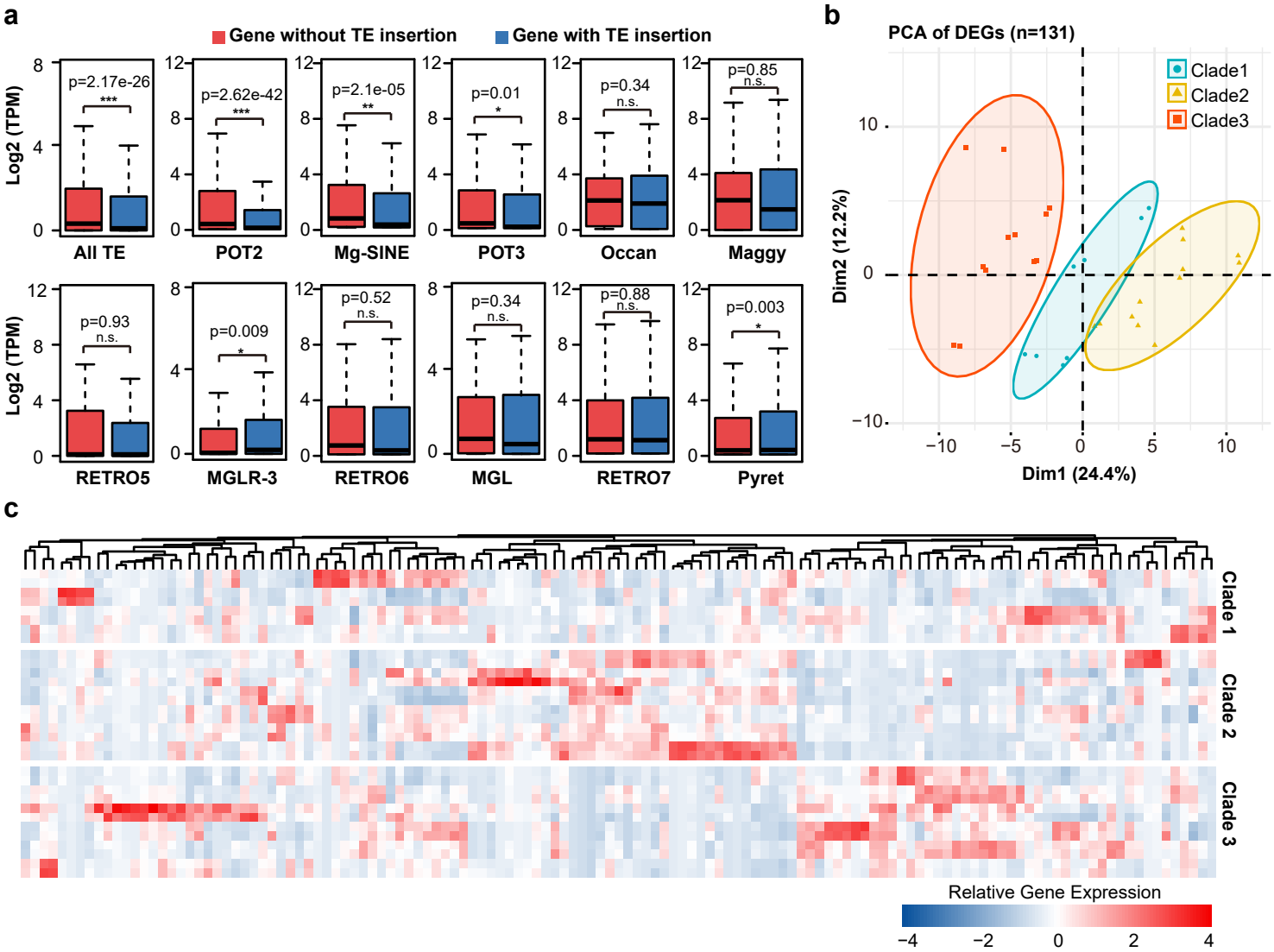

a
Gene without TE insertion
Gene with TE insertion
Log2 (TPM)
All TE
p=2.17e-26
***
POT2
p=2.62e-42
***
Mg-SINE
p=2.1e-05
**
POT3
p=0.01
*
Occan
p=0.34
n.s.
Maggy
p=0.85
n.s.
RETRO5
p=0.93
n.s.
MGLR-3
p=0.009
*
RETRO6
p=0.52
n.s.
MGL
p=0.34
n.s.
RETRO7
p=0.88
n.s.
Pyret
p=0.003
*
b
PCA of DEGs (n=131)
Clade1
Clade2
Clade3
Dim2 (12.2%)
Dim1 (24.4%)
c
Clade 1
Clade 2
Clade 3
Relative Gene Expression
−4
−2
0
2
4

TABLE 1 The numbers of secreted protein and non-secreted protein that have TE insertion in 1-kb flanking regions or within gene.

	Total number	Embed	Upstream	Downstream	Total
Non-SP	11,445	2,247	1,978	1,893	4,414
SP	1,546	358	482	475	845
All	12,991	2,605	2,460	2,368	5,259

#SP: secreted protein