Analyses of transposable elements in arbuscular mycorrhizal fungi support evolutionary parallels with filamentous plant pathogens

Jordana I. N. Oliveira[1], Catrina Lane[1], Ken Mugambi[1], Gokalp Yildirir[1], Ariane M. Nicol[1], Vasilis Kokkoris[1,2,¥], Claudia Banchini[2], Kasia Dadej[2], Jeremy Dettman[2], Franck Stefani[2], Nicolas Corradi*[,1]

[1] Department of Biology, University of Ottawa, Ottawa, ON, K1N 6N5, Canada

[2]Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON, K1A 0C6, Canada

[¥] Current address: Amsterdam Institute for Life and Environment (A-LIFE), Faculty of Science, 1081 HV, Amsterdam, the Netherlands

* Corresponding author ncorradi@uottawa.ca

Abstract

Transposable elements (TEs) are repetitive DNA sequences that excise or create copies that are inserted elsewhere in the genome. Their expansion shapes genome variability and evolution by impacting gene expression and rearrangement rates. Arbuscular mycorrhizal fungi (AMF) are beneficial plant symbionts with large, TE-rich genomes, and recent findings showed these elements vary significantly in abundance, evolution, and regulation among model AMF strains. Here, we aimed to obtain a more comprehensive understanding of TE function and evolution in AMF by investigating assembled genomes from representatives of all known families. We uncovered multiple, family-specific bursts of insertions in different species, indicating variable past and ongoing TE activity contributing to the diversification of AMF lineages. We also found that TEs are preferentially located within and around candidate effectors/secreted proteins, as well as in proximity to promoters. Altogether, these findings support the role of TEs in promoting the diversity in proteins involved in molecular dialogues with hosts and, more generally, in driving gene regulation. The mechanisms of TEs evolution we observed in these prominent plant symbionts bear striking similarities to those of many filamentous plant pathogens.

Keywords

Plant symbionts, repetitive DNA, effector proteins, mobile genome

Introduction

Arbuscular Mycorrhizal Fungi (AMF) are important plant symbionts in the sub-phylum Glomeromycotina[1]. These organisms support their hosts by transferring nutrients from the soil in exchange for carbohydrates and lipids, essential to their metabolism[2–4]. Fossil and phylogenomic evidence indicates that this successful partnership occurred over 400 million years ago when plants colonized land[2,5]. AMF hyphae and spores carry hundreds to thousands of nuclei[6], and these organisms have long been thought to be asexual due to the absence of formal sexual reproduction or structures[7]. However, this assumption has been challenged by the discovery of genes involved in sexual reproduction[8–10] and, more recently, by the discovery of homo-heterokaryotic life stages[11,12].

Genome size in AMF varies from about 50 to over 800 Megabases (Mb)[13], and correlates with genome composition, with larger genomes carrying significantly more repeated sequences, particularly transposable elements (TEs)[14]. TEs are repetitive DNA sequences that self-synthesize and insert copies of themselves within a genome, shuffling genetic information and creating functional novelties by inserting nearby promoters to impact gene expression[15] or in coding regions to alter protein sequences[16,17]. These elements are divided into two major classes based on their mobilization mechanisms: Class I (retrotransposons) use a “copy-and-paste” mechanism via an RNA intermediate to create new insertions, whereas Class II (DNA transposons) promote insertions by “cutting and pasting” DNA sequences[18]. TE expansions are controlled by DNA methylation and RNAi to maintain genome integrity. Still, external pressure, such as stress and variation of abiotic factors, can repress these silencing mechanisms and thus activate TE expansion[19]. Over time, TEs can also be domesticated to create new functions, a process known as molecular cooption, leading to increased protein variability that promotes adaptation to environmental changes[20].

In AMF, TE diversity has been detailed in a few species, mainly in the Gigasporaceae and Glomeraceae families[21,22]. More recently, their potential functional role has been uncovered in the model species *Rhizophagus irregularis*, where TE expression differs across life stages and is significantly up-regulated upon plant colonization[23,24]. Notably, TE families are differentially regulated depending on plant host identity, suggesting that these elements play an active, yet unknown, host-specific role during colonization[23].

Significant differences in TE abundance, genome location, and evolution exist between *R. irregularis* strains and species[13,23]. For instance, in the *R. irregularis* strain DAOM197198, the mobilization and retention of TEs can differ between related strains, and different species carry unique TE expansions, such as the long interspersed nuclear elements (LINEs) retroelements in *Gigaspora margarita*, which have been hypothesized to drive genome expansion in this genus[21,25]. TE insertion/retention also varies within AMF genomes. For example, in *R. irregularis*, the genome separates into two functionally distinct regions referred to as A- and B-compartments[12,26]. Most of the conserved genes are in the A-compartment, while the B-compartment is more repeat-dense and carries most genes potentially involved in the dialogue with the plant – i.e. secreted proteins (SP) and candidate effectors. Young TEs accumulate at higher rates in the latter but are eliminated early from the genome in the A-compartment[23].

It has been hypothesized that the abundance of TEs within the B-compartment in AMF mirrors what is observed in the genomes of some filamentous fungal and oomycete plant pathogens[24,26,27]. Specifically, these genomes can also separate into slowly evolving regions carrying conserved genes (analogous to the A-compartment), and others that evolve rapidly due to significantly higher TE and effector density (analogous to the B-compartment), leading to enhanced diversification of proteins involved in virulence[28,29]. As in some plant pathogens, B-compartment effectors are up-regulated in AMF upon colonization[26], and a significant positive correlation in the expression of TEs and B-compartment genes has been identified in the AMF *R. irregularis*[23].

Despite apparent similarities in the mode of TE evolution and regulation between AMF and distant plant pathogens, the mechanisms behind these have not been fully investigated. Furthermore, it is unknown whether the similarities observed between model AMF and plant pathogens are conserved among all Glomeromycotina lineages, as a detailed analysis of TEs in most known families has not yet been performed[23]. Here, we aim to improve our understanding of the diversity and evolution of Glomeromycotina TEs in comparison to pathogenic relatives by analyzing 24 AMF species with available genome assemblies, encompassing all known orders of this prominent symbiont group[30,31].

Results

Transposable elements diversity in Glomeromycotina

Transposable elements were annotated in 24 genome assemblies, including 3 Paraglomerales, 3 Archaeosporales, 5 Glomerales, 11 Diversisporales, and 2 Entrophosporales (**Table 1**). Only *Rhizophagus irregularis* have chromosome-level assemblies [26,32]. The quality of the other assemblies varies from a N50 of 1,501 to 1,306,809 in *Innospora majewskii* and *Rhizophagus clarus*, respectively (Table S1). To aid visualization of TE evolution, a phylogenetic tree was generated using OrthoFinder [33] that fully agrees with the most recent phylogenomic analyses of AMF (**Figure S1**) [34].

Table 1. Comparison between automated and customized TE annotation.

Group	Species	X̄% masked TE before	X̄% unknown TE before	X̄% known TE before	X̄% known TE after
Paraglomerales	3	22.3 ± 2.2	18.4 ± 2.1	**3.9** ± 2.15	**8.6** ± 0.32
Archaeosporales	3	62.5 ± 6.66	18.8 ± 6.75	**43.7** ± 5.47	**49.2** ± 9.99
Glomerales	5	45.8 ± 3.32	30.5 ± 1.55	15.3 ± 1.3	30.5 ± 2.85
Diversisporales	11	68.5 ± 2.62	45.3 ± 2.59	23.2 ± 0.98	43.9 ± 2.11
Entrophosporales	2	34.3 ± 0.3	25.8 ± 0.45	8.6 ± 0.38	14.1 ± 0.82

The mean percentage of the genome (±SE) covered by known or unknown TEs in each order was calculated for each of the two annotations. In the first annotation, the masked portion represents known and unknown TE sequences. There is an increase in known sequences for all orders after the curation. The highest and lowest mean percentages of known sequences are shown in green and red, respectively.

Using the automated method RepeatModeler to generate the repeat libraries, between 18% to 45% of genomes were covered by unknown TEs. This high number of unclassified repeats is mainly due to the lack of information on TEs from fungi in the main databases used as reference[35]. Building a customized curation over the automated repeat library, significantly improved the annotation of transposable elements in all genomes ($p < 0.01$), except for *Geosiphon pyriformis* and *Acaulospora colombiana*, where most TEs were already derived from known sequences. With this curated dataset, the proportion of TEs within AMF genomes now ranges from a minimum of 6.82% in *Innospora majewskii* to a maximum of 58.98% in *Ambispora leptoticha* (**Figure 1**). AMF genomes also contain several expanded gene families, such as Kinases, Kelch, Sel1, TPR, and BTB, which represent on average 3.7% of the AMF curated datasets, with *R. irregularis* containing the highest number of repeated genes at 8%[24,36] (**Figure 1a**).

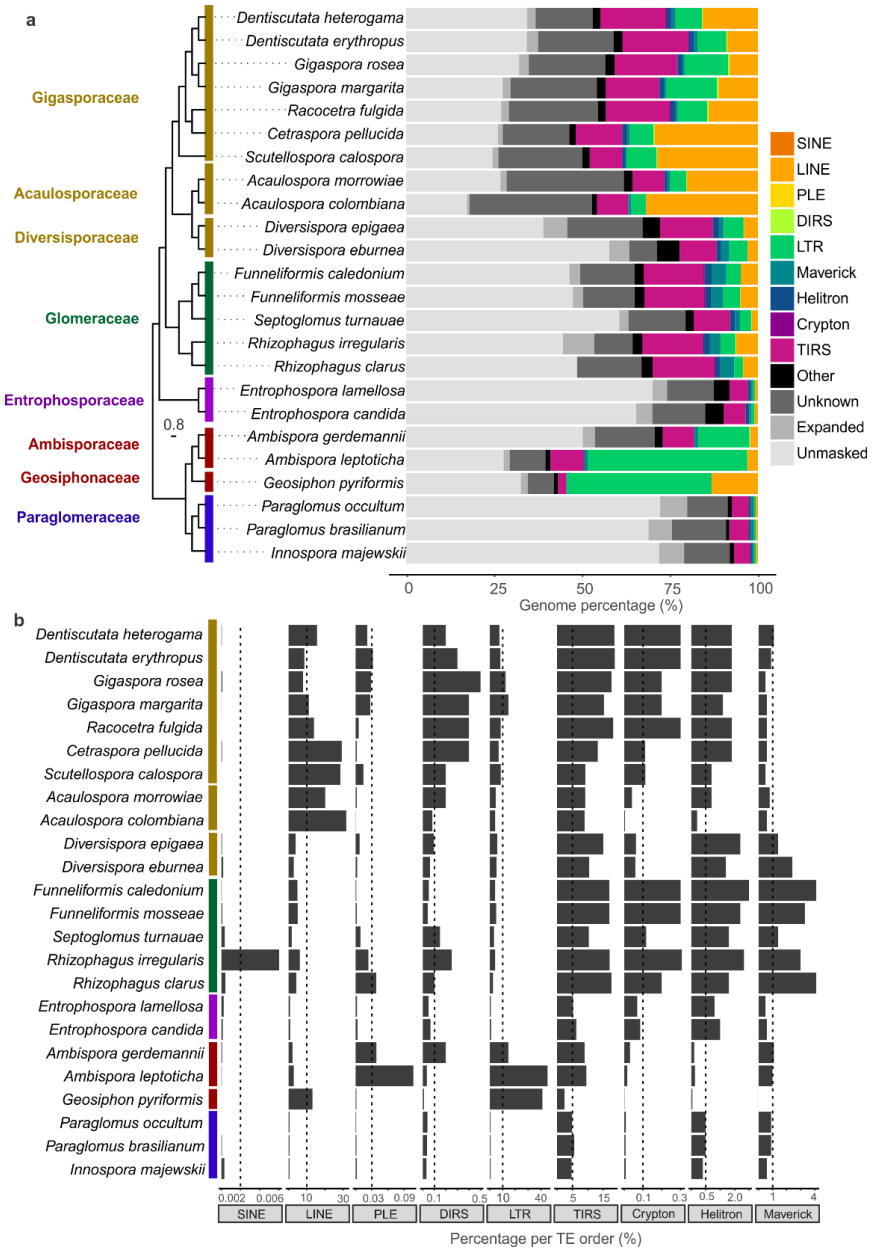

Figure 1 – Genome composition in Glomeromycotina species. **a)** Relative distribution of different types of sequences in AMF species. The phylogenetic tree is based on ortholog genes, and the bar indicates the level of nucleotide substitution, representing the distances between the taxa. Bootstrap support can be found in Figure S1. The tree is consistent with the most AMF phylogeny built by Rosling et al. 2024. The colors on the genome composition represent different TE orders. In black, are other repeats, such as satellite and low complexity; in dark grey are the expanded gene families. The "unknown" buffer represents the remained portion of masked repeats from the first annotation (without curation). The "unmasked" region represents the genome portion other than repeats. **b)** Relative content per order of transposable element in relation to the entire genome. The *x-axis* shows the genome percentage covered by each TE order.

In general, species within the same order share very similar TE diversity. Paraglomerales species always have the lowest content of TEs, ranging from 6.8% to 8.1%. The Archeosporales carry long-terminal repeats (LTRs) expansions especially *G. pyriformis* with 41.2% of the genome covered by LTR sequences. In the Glomerales, DNA transposons dominate TE abundance and carry unique expansions of Maverick

elements. The expansion of LINE elements found in *Gigaspora* species, previously hypothesized to be the cause of genome expansion in these genera[21,22], is common to all Diversisporales species (i.e. Acaulosporaceae, Diversisporaceae, and Gigasporaceae families).

A core of highly expanded TEs in Glomeromycotina

Of the 6,772 non-redundant consensus sequences curated in this study, 1,295 (19.12%) are shared among all species ("core" families) (**Figure 2a**), an indication that the 4,477 sequences either originated and diversified after species differentiation, or were lost secondarily in different lineages. The AMF "core" families represent on average 50% of the TEs copies in AMF genomes (**Figure 2b**). Among the core sequences, LTR/Bel has two families that are present in all genomes analyzed. In contrast, LTR/Copia, DNA/Academ, DNA/Dada, Crypton, and SINE are the superfamilies with a higher number of sequences that are "non-core" families.

Overall, species from Glomerales and Diversisporales have more expansions of clade-specific sequences and carry a greater number of non-core families compared to the other AMF groups. Pairwise comparisons showed that the non-core families are more likely to be shared between closely related species, and could represent clade-specific TE invasions and expansions (**Figure 2c**).

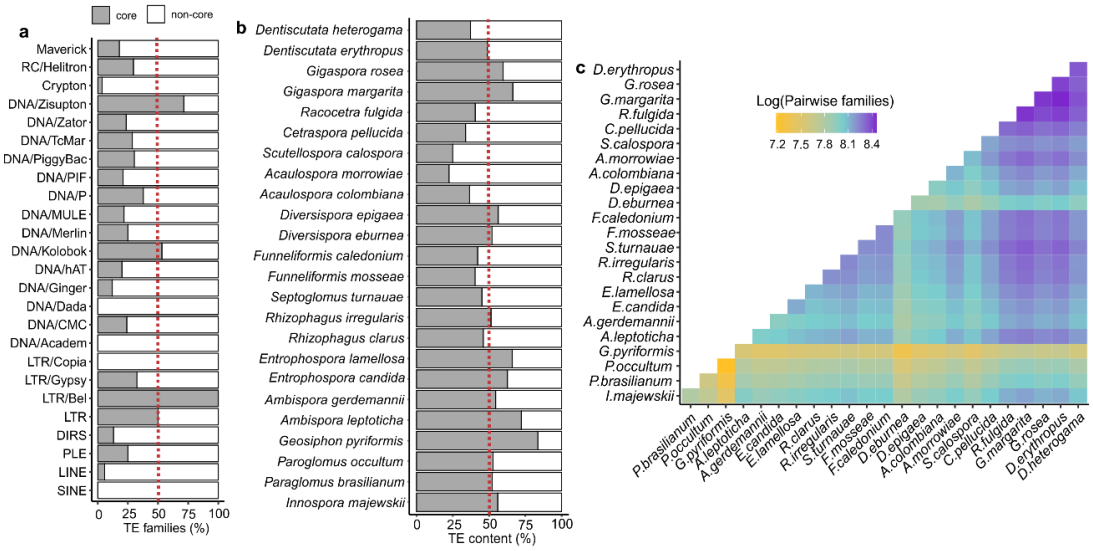

Figure 2 – TEs families shared across Glomeromycotina species. **a)** Percentages in relation to the total number of each superfamily that are present in all species (core) and in blank are the superfamily that are not present in all species (specific). **b)** Relation of core and specific families in the TE annotation. **c)** Pairwise comparison of shared families. The number of shared families between two species is expressed in log scale to minimize the bias of certain species having more families.

Transposable elements expansions and degeneration across AMF species

To determine how TEs are retained and mobilized in AMF, we constructed their repeat landscapes based on the Kimura substitution calculation. Lower Kimura substitution rates indicate recent insertion events, with sequences showing a low level of divergence from the consensus, whereas higher Kimura substitution rates uncover older TE insertions, meaning a high-level divergence between the copies and the consensus.

Clear patterns of conservation and distinctions in TE mobilization and retention exist among analyzed families (**Figure 3**). Specifically, the Paraglomeraceae and Entrophosporaceae have a flat distribution of TEs (**Figure 3a**), suggesting that minimal TE bursts have occurred in these two distinct lineages. In contrast, other families show specific waves of TE expansion and degeneration leading to genome diversification. For example, Geosiphonaceae and Ambisporaceae underwent similar and striking expansions of LTRs, while LINEs have similarly invaded Acaulosporaceae and Gigasporaceae. Of all the AMF families analyzed, Glomeraceae and Diversisporaceae were the only ones to show obvious recent expansions (peaks between 1 to 5 sequence divergence) (**Figure 3e, f**), suggesting ongoing TE activity in their genomes.

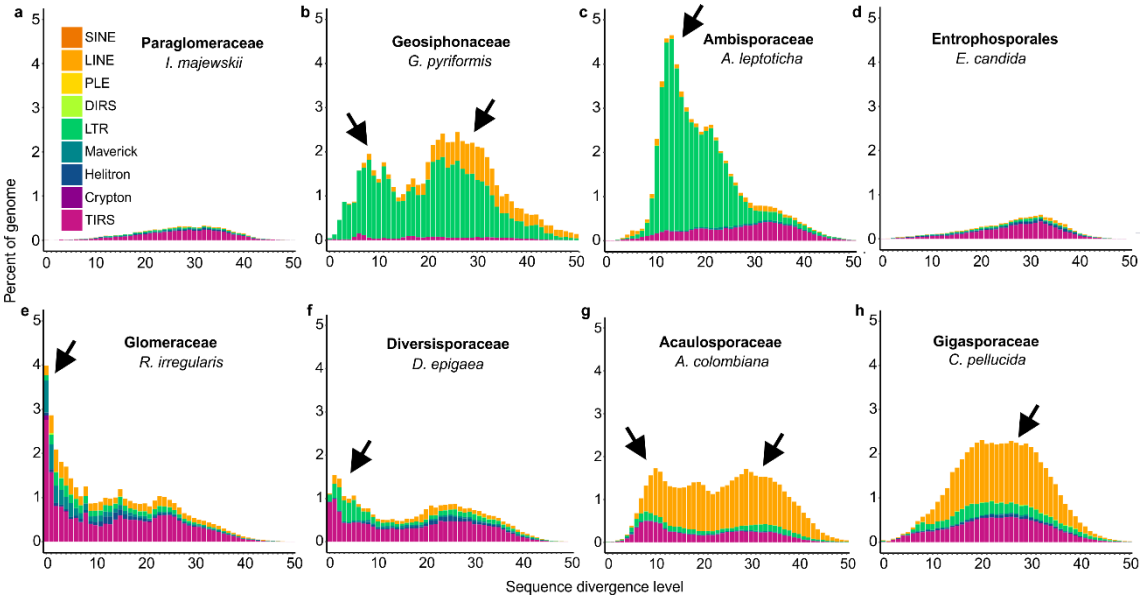

Figure 3 – Transposable elements landscape across AMF species. Each histogram represents the genome percentage (y-axis) of copies with different level of divergence (x-axis). The arrows highlight the specific past events of TE that might contributed to the diversification of the genome.

182 *TEs are preferentially located upstream of genes in Glomeromycotina*

183 In filamentous plant pathogens, TEs located within 1000 bp upstream to their transcript
184 start site (-1Kb TSS), - i.e. in obvious vicinity to promoters and/or regulatory sequences
185 that influence gene expression - are co-express with the genes they co-locate with [37,38].
186 In the model AMF *R. irregularis*, genes within the -1Kb TSS also co-express with the
187 closest TEs [23].

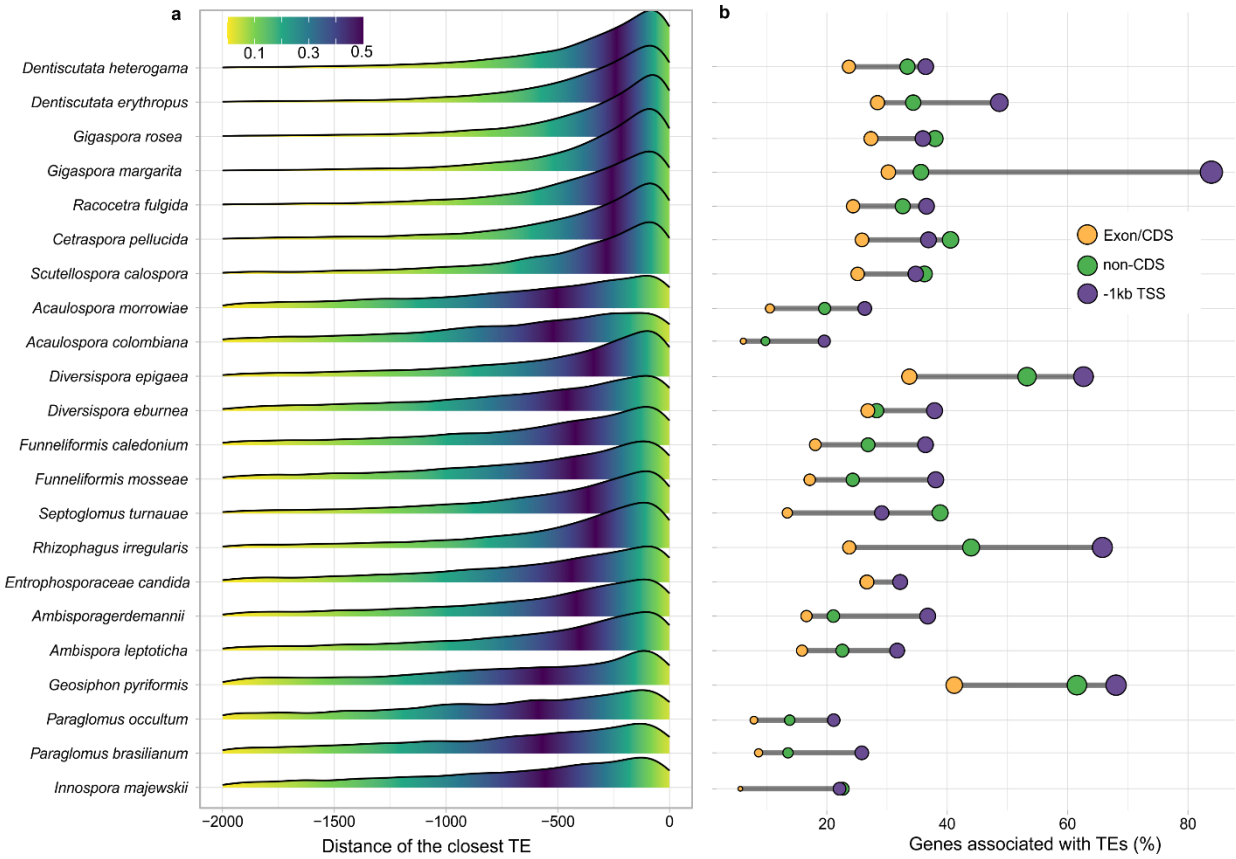

Figure 4 – Transposable elements association within genes. **a)** Closest TE association in the upstream region of coding genes. The x-axis represents the base-pair distance from the transcription start site (TSS). The y-axis shows the density of genes with the closest TE at a given position (x-axis). The gradient color indicates the tail probability, highlighting regions with significant TE proximity. **b)** Percentages of genes harboring TE insertions upstream to the transcription start site (TSS), in UTRs and introns (non-CDS) and exon/CDS regions.

188 By investigating available datasets, we confirm that, in all AMF orders, the closest
189 TE to a gene locates within the 1Kb-TSS (average -872 nucleotides upstream of the TSS,
190 (SD = 399) (**Figure 4a**). In all cases, TE insertions are significantly more common within
191 -1Kb TSS regions compared to both non-coding regions (UTRs and introns, average of
192 38% *vs* 20%; p-value <0.001) and exons (average 38% *vs* 30%; p-value <0.001) (**Figure**
193 **4b**). For the -1Kb TSS insertions, notable differences exist among families (SD 17%).

194 For example, 20% of genes in Paraglomerales 60% of TEs are within the TSS compared
195 to up to 60% in Glomerales and Diversisporales. Overall, *Paraglomus* spp. and *A.*
196 *colombiana* have fewer genes with TE sequences within CDS/exons at 13% and 9%
197 respectively, whereas *D. epigea* and *G. pyriformis* have a higher proportion with 53% and
198 61% of TEs. *G. pyriformis* is also characterized by a much higher number of genes with
199 TEs insertions within exons (**Figure 4b**). In Glomeromycotina, LINE, LTR/Gypsy,
200 DNA/CMC, DNA/hAT, and DNA/MULE elements are the TE families commonly found
201 either near or within genes (**Figure S2**).

202 *Effectors locate within TE-rich regions in Glomeromycotina*

203 The effector genes of several filamentous plant pathogens are often associated with TEs
204 within so-called "adaptive regions" that promote effector variability and
205 diversification[39]. These genes are important to establish communication with the host
206 and start the infection process.

207 With the exception of *P. occultum*, *G. pyriformis,* and *A. marrowiae*, the density
208 of TE insertions, defined by sequence overlaps with coding sequence and co-location
209 within 1Kb up or downstream genes, is significantly higher (p-value <0.05) within
210 candidate effectors compared to other protein-coding genes **(Figure 5, Table S4**).
211 Effectors are predominantly found in TE-rich regions, while non-secreted genes are more
212 common in TE-poor areas, with secreted proteins showing an intermediate pattern. In
213 most species, effectors exhibit a right-skewed TE frequency distribution, supporting
214 genome compartmentalization where dynamic, TE-dense regions harbor SPs and effector
215 genes. This pattern aligns with the "two-speed genome" model observed in plant
216 pathogens [35,37,38,40], and the A/B genome compartments reported in the model AMF
217 *R. irregularis* [12,23,26].

218 It is noteworthy that the degree of fragmentation of the genome assembly does not
219 affect the general observations; in almost all cases, TEs are more likely to be found near
220 or within promoter regions (-1Kb) and in association with secreted proteins and effectors.
221 Even species with the lowest BUSCO scores (<50%), such as *Scutellospora calospora,*
222 *Dentiscutata heterogama, Entrophospora lamelosa,* and *Innospora majewskii*, presented
223 the observed trends are like those identified in more complete assemblies (**Table S1**).

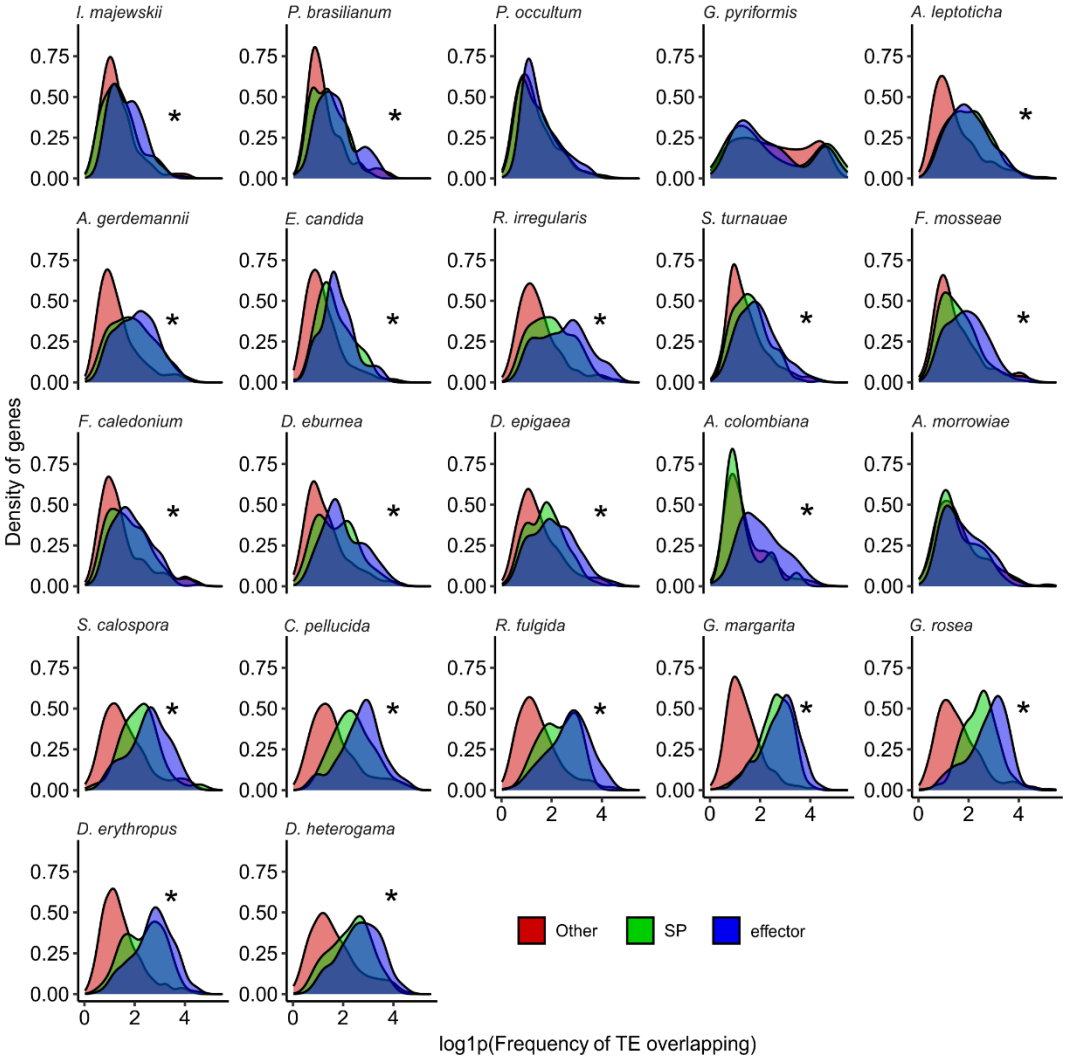

Figure 5 – TE overlap frequency within and around Genes. The x-axis represents the log-transformed frequency of TE insertions overlapping gene regions, while the y-axis shows the density of genes with a given TE overlap. Effectors (blue) are compared to secreted proteins (SP, green) and other protein-coding genes (red). Asterisks (*) indicate significant differences (t-test $p < 0.05$) between effectors and other genes. For additional statistical comparisons, see Table S4.

Discussion

A curated view of TE diversity and evolution in the Glomeromycotina

The current work provides a curated catalogue of TEs for all Glomeromycotina with available genome assemblies. Although the curation improved the TE annotations performed with tools based on sequence similarity[41,42], most of the families we analyzed, except *R. irregularis*, still have a proportion of putative unknown repeats. This is likely because *R. irregularis* is the only AMF species with chromosome-level assemblies, while others have fragmented assemblies or collapsed repeat regions[30], which are known to limit the detection of TEs annotation[35]. Hopefully, more chromosome-level assemblies will become available in the future, allowing the identification of species-specific TEs that may have been overlooked in our screening.

236 In general, AMF and other fungi carry expansions of retrotransposons such as
237 LINEs and LTRs, however, some species deviate from this trend[21,35]. For example,
238 insertions from DNA transposons such as TIRS, Crypton, Helitron, and Maverick
239 elements are abundant in Glomerales. The patterns of TE evolution we observed also
240 confirm that, together with the number of genes[13], TEs can vary dramatically
241 in presence/absence and relative abundance between AMF relatives and within TE
242 families. For example, the virus-like element Maverick is expanded exclusively in
243 Glomerales. This element has only been reported in the plant pathogen *Phakosora*
244 *pachyrhizi*, in the AMF *Rhizophagus (Glomus) intraradices*[43], and in the chytrid
245 *Batrachochytrium dendrobatidis* [44], and it is thus intriguing to speculate that it may
246 have originated via horizontal gene transfer in the Glomerales. Another, perhaps less
247 realistic scenario, is that this TE family was lost in all but one of known AMF orders.

248 In general, our work supports the view that obligate biotrophs carry more
249 transposable elements than non-obligate fungi, linking the pattern of TE insertions to their
250 lifestyle[35,45]. Within this context, a similar trend was observed in the cyanobacteria
251 endosymbiont *G. pyriformis*, which showed a greater expansion of mobile genetic
252 elements compared to any free-living relatives[46].

253 *TE diversity and distribution often reflect family evolutionary history*

254 In many cases, TE diversity and distributions were very similar between members of the
255 same AMF family. This supports events of TE insertions and retention that occurred in
256 the most recent common ancestor (MRCA) of these lineages. For example, all members
257 of the Archaeosporales that we analyzed showed similar bursts of LTRs and differed only
258 in how each of these TEs was retained over time. Similarly, members of the sister orders
259 Glomerales and Diversisporales underwent strikingly similar bursts of LINEs and
260 DNA/TIRs, again suggesting that these were inserted before these lineages diverged.

261 Remarkably, the evolution of these elements can sometimes go beyond
262 phylogenetic relationships. For example, members of phylogenetically distinct families,
263 Paraglomeraceae and Entrophosporaceae, can have virtually identical TE distributions.
264 This similarity may reflect independent and adaptive TE retentions due to similarities in
265 (epi) genomic architecture, such as methylation, or ecological traits[35].

266

267

268 *Evidence of A/B compartments across all Glomeromycotina*

269 The chromosomes of the model species *R. irregularis* are separated into two main regions,
270 called A/B compartments[12,26], and our analysis strongly indicates that this
271 organization is shared by most Glomeromycotina. In particular, TE density is always
272 significantly higher near secreted proteins and candidate effectors in AMF genomes; a
273 characteristic that is highly reminiscent of the B-compartment and that shares further
274 highlights striking similarities in the genome biology of AMF and many filamentous
275 fungal plant pathogens[24,29].

276 Specifically, these independent symbiotic and pathogenic lineages both carry
277 genes involved in dialogue with the host (effectors) within fast-evolving and repeat-rich
278 regions of the genome. As in fungal pathogens, the association between effectors and TE
279 in AMF is likely to diversify these molecules, allowing for improved adaptation and
280 symbiosis with a wider range of hosts. Notably, this organization can also lead to a
281 "devil's bargain" situation, as TE insertions near or within effectors could degenerate
282 these genes by increasing the chances of deleterious mutations for the symbiosis[39].

283 Future studies based on analyses chromosome-level assemblies and chromatin-
284 capture (Hi-C) information from representatives of the orders Diversisporales,
285 Archeasporales, and Paraglomerales will hopefully provide a more detailed view of the
286 shared and lineage-specific regulatory processes affected by each compartment in
287 Glomeromycotina.

288 *Are TE bursts linked with AMF species diversification and adaptation?*

289 Bursts of TEs promote genomic rebuilding, and this mechanism is often associated with
290 higher speciation rates[47]. In this context, it is intriguing to speculate that the reduced
291 TE load of Paraglomerales, together with its flat landscape distribution (indicating low
292 TE activity and highly stable genome) is linked with a limited species richness within this
293 family[48]. Indeed, to date, Paraglomerales has fewer species formally recognized
294 compared to other Glomeromycotina, including other early diverging lineages like
295 Archaeosporales – e.g. 11 Paraglomerales and 19 Archaeosporales[48]. In contrast, more
296 divergent AMF orders such as Glomerales and Diversisporales featured high levels of TE
297 bursts within their genomes and are significantly more species-rich, with 268 and 70
298 species, respectively[48].

299 The lower TE abundance of Paraglomerales also supports the hypothesis that
300 fewer gene/TE associations with reduced mycorrhizal capabilities[14] recently reported
301 for this order compared to AMF relatives[49,50]. Here, lower TE loads may reduce
302 the chances of generating genomic diversity necessary for adapting to environmental
303 change[27,51], especially if mating is rare or absent[10,24]. In support of this, most other
304 AMF groups with high genomic diversity and adaptability[13], such as Glomeraceae,
305 have experienced notable TE bursts in the past and carry evidence of recent and ongoing
306 TE activity.

307 *Indirect evidence of gene regulation through TE in Glomeromycotina*

308 Our analyses also revealed that TEs are preferentially locate within promoter regulatory
309 regions in these symbionts. TE insertions near genes can impact gene expression by
310 inducing methylation or altering transcription factor binding sites[52], and even
311 promoting antifungal resistance as observed in a plant pathogen[53]. In the yeast
312 *Saccharomyces cerevisiae*, LTR elements are preferentially inserted 700 bp upstream to
313 coding genes and show synergistic expression during heat-stress induction[54].

314 In *R. irregularis*, co-expression has been observed between SP and upstream TEs
315 in the B-compartment, with TEs showing higher expression during symbiosis[23] than in
316 spores[24]. In addition, TEs showed strong methylation signals in germinating
317 spores[26], which coincides with high expression of an RNAi-related gene[23]. This
318 suggests that TEs may undergo demethylation during symbiosis, leading to increased
319 expression of both TEs and symbiosis-related genes, such as secreted proteins and
320 effectors. Again, this phenomenon has been observed in plant pathogens, where genes
321 with TE insertions within -1Kb are highly expressed during early infection, revealing a
322 synchrony between TEs and infection-related genes[37,54].

323 Remarkably, our analyses revealed that regions containing effector genes have
324 significantly more TE insertions compared to other coding sequences. This phenomenon
325 has also been observed in plant pathogens, where TE accumulation drives effector
326 diversification, thereby increasing fungal virulence[28,39]. On the other hand, the plant
327 evolves its immune system to combat infections, and again, TE-driven modifications
328 serve as a source for the fungus to overcome host defenses, in an ongoing coevolutionary
329 process[55]. Similarly, in AMF, diversified effectors could lead to broader adaptation to
330 different hosts and environments, as these cosmopolitan symbionts can associate with
331 many different hosts[56]. Given the high expression of TEs during symbiosis and frequent

332 association with SP, it would be interesting to investigate changes in methylation to
333 understand how these sequences become more actively transcribed upon contact with the
334 host. Additionally, it would be valuable to examine which factors, whether biotic or
335 abiotic stressors, can demethylate TEs in spores and reactivate TE mobilization as a
336 source of variability.

337 As our study uncovered the presence of numerous species and order-specific TE
338 families, it will be important that future analyses aim to determine their relative function
339 and regulation among various conditions and plant hosts. This will be a crucial step to
340 obtain a complete understanding of the evolution and role of TEs in these widespread root
341 symbionts. Within this context, Maverick and LINE elements, which present specific
342 expansions in Glomerales and Diversisporales, are good candidates for targeted
343 investigation.

344 Material and Methods

345 *Genome assemblies*

346 We downloaded 21 genome assemblies, protein, and gene annotation files from all
347 Glomeromycotina orders from NCBI and provided 3 new genome assemblies **Table S1**.
348 Despite other AMF species having genomes deposited in the databases, we opted to keep
349 only the assemblies with protein and gene annotation files which were used in the further
350 analysis.

351 *Hidden-Markov Models (HMM) of transposition domains*

352 We constructed HMM profiles for several transposition domains described in **Table S2**.
353 Sequences for each domain were aligned separately using mafft[57], then converted to
354 Stockholm format using esl-format and submitted to hmmbuild[58] to generate the HMM
355 profiles. The profiles were concatenated and prepared for hmmscan using hmmpress. All
356 models are available on https://github.com/jordana-olive/Transposable-Elements-in-AM-
357 Fungi-Evolutionary-Parallels-with-Fungal-Plant-Pathogens.

358 *Repeat library generation*

359 Each genome was submitted to RepeatModeler2.0[41] with the -LTRStruct option to
360 generate the repeat sequence files. All the repeat libraries were concatenated into a single
361 file from which sequences shorter than 1000 nucleotides were removed using seqkit [59].
362 The most representative sequence (consensus) for each putative TE family was

363 prospected by clustering the repeat library using cd-hit-est -aS 0.8 -c 0.8 [60], following
364 the 80-80-80 rule. This rule establishes that all sequences with at least 80% similarity
365 along 80% of the total length should belong to the same TE family [61]. The final file
366 containing the nonredundant TE-families was named "non-curated library".

367 *Transposable elements curation and annotation*

368 We predicted open reading frames (ORFs) in the "non-curated library" using getorf[62],
369 selecting translated ORFs of at least 300 nucleotides. We submitted the ORFs from the
370 "non-curated library" to hmmsearch [58] to retrieve the sequences containing the
371 transposition domains using the hmm models. We selected the best match of each
372 sequence using HmmPY.py[63] (e-value 1e-17, -c 0.35, -h 50). Sequences carrying
373 transposition domains were renamed with the corresponding family domain. Curation
374 generated the "final library" to be used for transposable element annotation. To compare
375 the impact of the curation on the TE resolution, the annotation was performed twice in
376 each genome using RepeatMasker2.0.4 -s -a modes[42] using the non-curated library and
377 the final library. The alignment files generated by the annotation process were used to
378 calculate the degree of divergence based on Kimura values using the scripts
379 calcDivergenceFromAlign.pl and createRepeatLandscape.pl provided by
380 the RepeatMasker tool.

381 *Orthologues genes and AMF phylogenetic tree*

382 The protein sequences of the studied genomes were submitted to OrthoFinder[33] to
383 construct a phylogenetic tree based on orthologous genes, and iTOL was used to visualize
384 the branch length and bootstrap support (**Figure S1**)[64]. The tree was rooted with species
385 from Paraglomerales and Archeosporales and visualized using the ggtree R package in
386 combination with the TE annotation data.

387 *TE association with genes*

388 We compared the intersection between TEs and genes using bedtools intersect[65] in gff
389 from Genebank and our TE annotation. To maximize the results, we rely on the species
390 with available gene annotation and protein sequences, which allow us to confirm the gene
391 structure by comparing coding sequences. We also split the annotations into three files:
392 1000 base pairs (bp) upstream of the gene's transcription target site, exons and other non-
393 coding regions (5' & 3'UTR and introns). For instance, using bedtools subtract, introns
394 and UTRs were prospected by subtracting the coding region information following the

395 protein sequences coordinates[65]. We considered a valid intersection when the TE
396 overlap was at least 50 nucleotides long and not identical to the size of TE and the gene
397 simultaneously.

398 *Prediction of secreted proteins and effectors*

399 Secreted proteins were predicted in all species using SignalP-6.0[66]. From the predicted
400 secreted proteins, we used the EffectorP-3.0[67] to annotate the candidate effectors.

401 **Acknowledgments**

402 Our research is funded by the Discovery program of the Natural Sciences and Engineering
403 Research Council (RGPIN2020-05643) and a Discovery Accelerator Supplements
404 Program (RGPAS-2020-00033). NC is a Research Chair in Microbial Genomics at the
405 University of Ottawa. JINO is funded by the Mitacs Accelerate Program (IT16902) and a
406 Discovery Accelerator Supplements Program (RGPAS-2020-00033).

407 **Author's contribution**

408 Experiment design, contextualization and writing (JINO and NC). Bioinformatic analysis
409 (JINO, CL, KM, GY). Repeat annotation and repeat landscapes (CL). Biological material,
410 sequencing and analyses for *I. majewskii, S. turnauae* and *E. lamellosa* (CB, VK, FS, KD,
411 JD). Data curation and downloading (AMN, GY). Github organization (KM). Supervision
412 (NC). All authors read and approved the final version of the manuscript.

413 **Supplementary information**

414 **Table S1 –** Genome information for the assemblies used in this study.

415 **Table S2 –** Transposable element domains used in the hmm models.

416 **Table S3 –** Genes, secreted proteins, and effectors in Glomeromycotina.

417 **Table S4 –** Significance (t-test) of TE frequency-per-gene comparisons.

418 **Figure S1 –** Phylogenetic tree of Glomeromycotina from ortholog genes.

419 **Figure S2 -** Association of genes per TE order.

420 **References**

421 1. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, et al. A phylum-
422 level phylogenetic classification of zygomycete fungi based on genome-scale data.
423 Mycologia. 2016;108:1028–46.

2. Rich MK, Vigneron N, Libourel C, Keller J, Xue L, Hajheidari M, et al. Lipid exchanges drove the evolution of mutualism during plant terrestrialization. Science. 2021;372:864–8.

3. Jiang Y, Wang W, Xie Q, Liu N, Liu L, Wang D, et al. Plants transfer lipids to sustain colonization by mutualistic mycorrhizal and parasitic fungi. Science. 2017;356:1172–3.

4. Keymer A, Pimprikar P, Wewer V, Huber C, Brands M, Bucerius SL, et al. Lipid transfer from plants to arbuscular mycorrhiza fungi. Elife [Internet]. 2017;6. Available from: http://elifesciences.org/lookup/doi/10.7554/eLife.29107

5. Remy W, Taylor TN, Hass H, Kerp H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proc Natl Acad Sci U S A [Internet]. 1994; Available from: http://dx.doi.org/10.1073/pnas.91.25.11841

6. Kokkoris V, Stefani F, Dalpé Y, Dettman J, Corradi N. Nuclear Dynamics in the Arbuscular Mycorrhizal Fungi. Trends Plant Sci. 2020;25:765–78.

7. Normark BB, Judson OP, Moran NA. Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society. 2003. p. 69–84.

8. Halary S, Daubois L, Terrat Y, Ellenberger S, Wöstemeyer J, Hijri M. Mating Type Gene Homologues and Putative Sex Pheromone-Sensing Pathway in Arbuscular Mycorrhizal Fungi, a Presumably Asexual Plant Root Symbiont. PLoS One. 2013;8:e80729.

9. Halary S, Malik S-B, Lildhar L, Slamovits CH, Hijri M, Corradi N. Conserved Meiotic Machinery in Glomus spp., a Putatively Ancient Asexual Fungal Lineage. Genome Biol Evol. 2011;3:950–8.

10. Yildirir G, Malar C M, Kokkoris V, Corradi N. Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. Trends Microbiol. 2020;xx:1–3.

11. Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, et al. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nature Microbiology. 2016;1.

12. Sperschneider J, Yildirir G, Rizzi YS, Malar C M, Mayrand Nicol A, Sorwar E, et al. Arbuscular mycorrhizal fungi heterokaryons have two nuclear populations with distinct roles in host-plant interactions. Nat Microbiol [Internet]. 2023; Available from: https://www.nature.com/articles/s41564-023-01495-8

13. Oliveira J, Yildirir G, Corradi N. From Chaos Comes Order: Genetics and Genome Biology of Arbuscular Mycorrhizal Fungi. Annu Rev Microbiol [Internet]. 2024; Available from: http://dx.doi.org/10.1146/annurev-micro-041522-105143

14. Malar C M, Wang Y, Stajich JE, Kokkoris V, Villeneuve-Laroche M, Yildirir G, et al. Early branching arbuscular mycorrhizal fungus Paraglomus occultum carries a small and repeat-poor genome compared to relatives in the Glomeromycotina. Microb Genom [Internet]. 2022;8. Available from: http://dx.doi.org/10.1099/mgen.0.000810

462 15. Simonti CN, Pavlicev M, Capra JA. Transposable Element Exaptation into Regulatory
463 Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints.
464 Mol Biol Evol. 2017;34:2856–69.

465 16. Sela N, Mersch B, Hotz-Wagenblatt A, Ast G. Characteristics of transposable element
466 exonization within human and mouse. PLoS One. 2010;5:e10907.

467 17. Schmitz J, Brosius J. Exonization of transposed elements: A challenge and
468 opportunity for evolution. Biochimie. 2011;93:1928–34.

469 18. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified
470 classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–
471 82.

472 19. Oggenfuss U, Croll D. Recent transposable element bursts are associated with the
473 proximity to genes in a fungal plant pathogen. PLoS Pathog. 2023;19:e1011130.

474 20. Cosby RL, Chang N-C, Feschotte C. Host-transposon interactions: conflict,
475 cooperation, and cooption. Genes Dev. 2019;33:1098–116.

476 21. Venice F, Ghignone S, Salvioli di Fossalunga A, Amselem J, Novero M, Xianan X, et
477 al. At the nexus of three kingdoms: the genome of the mycorrhizal fungus Gigaspora
478 margarita provides insights into plant, endobacterial and fungal interactions. Environ
479 Microbiol. 2020;22:122–41.

480 22. Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, et al.
481 Comparative genomics of Rhizophagus irregularis, R. cerebriforme, R. diaphanus and
482 Gigaspora rosea highlights specific genetic features in Glomeromycotina. New Phytol.
483 2019;222:1584–98.

484 23. Oliveira JIN, Corradi N. Strain-specific evolution and host-specific regulation of
485 transposable elements in the model plant symbiont Rhizophagus irregularis. G3.
486 2024;14:jkae055.

487 24. Dallaire A, Manley BF, Wilkens M, Bista I, Quan C, Evangelisti E, et al.
488 Transcriptional activity and epigenetic regulation of transposable elements in the
489 symbiotic fungus Rhizophagus irregularis. Genome Res. 2021;31:2290–302.

490 25. Silvestri A, Turina M, Fiorilli V, Miozzi L, Venice F, Bonfante P, et al. Different
491 Genetic Sources Contribute to the Small RNA Population in the Arbuscular Mycorrhizal
492 Fungus Gigaspora margarita. Front Microbiol. 2020;11:1–15.

493 26. Yildirir G, Sperschneider J, Malar C M, Chen ECH, Iwasaki W, Cornell C, et al. Long
494 reads and Hi-C sequencing illuminate the two-compartment genome of the model
495 arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 2022;233:1097–
496 107.

497 27. Reinhardt D, Roux C, Corradi N, Di Pietro A. Lineage-Specific Genes and Cryptic
498 Sex: Parallels and Differences between Arbuscular Mycorrhizal Fungi and Fungal
499 Pathogens. Trends Plant Sci. 2021;26:111–23.

500 28. Seidl MF, Thomma BPHJ. Transposable Elements Direct The Coevolution between
501 Plants and Microbes. Trends Genet. 2017;33:842–51.

29. Faino L, Seidl MF, Shi-Kunne X, Pauper M, Van Den Berg GCM, Wittenberg AHJ, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2016;26:1091–100.

30. Montoliu-Nerin M, Sánchez-García M, Bergin C, Kutschera VE, Johannesson H, Bever JD, et al. In-depth phylogenomic analysis of arbuscular mycorrhizal fungi based on a comprehensive set of de novo genome assemblies. Front Fungal Biol [Internet]. 2021;2. Available from: http://dx.doi.org/10.3389/ffunb.2021.716385

31. Sahraei SE, Sánchez-García M, Montoliu-Nerin M, Manyara D, Bergin C, Rosendahl S, et al. Whole genome analyses based on single, field collected spores of the arbuscular mycorrhizal fungus Funneliformis geosporum. Mycorrhiza. 2022;32:361–71.

32. Manley BF, Lotharukpong JS, Barrera-Redondo J, Llewellyn T, Yildirir G, Sperschneider J, et al. A highly contiguous genome assembly reveals sources of genomic novelty in the symbiotic fungus Rhizophagus irregularis. G3 (Bethesda) [Internet]. 2023;13. Available from: http://dx.doi.org/10.1093/g3journal/jkad077

33. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.

34. Rosling A, Eshghi Sahraei S, Kalsoom Khan F, Desirò A, Bryson AE, Mondo SJ, et al. Evolutionary history of arbuscular mycorrhizal fungi and genomic signatures of obligate symbiosis. BMC Genomics. 2024;25:529.

35. Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Rep. 2019;9:4307.

36. Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, Ndikumana S, et al. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 2018;220:1161–71.

37. Fouché S, Badet T, Oggenfuss U, Plissonneau C, Francisco CS, Croll D. Stress-Driven Transposable Element De-repression Dynamics and Virulence Evolution in a Fungal Pathogen. Mol Biol Evol. 2019;37:221–39.

38. Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, et al. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biol. 2018;16:78.

39. Fouché S, Oggenfuss U, Chanclud E, Croll D. A devil's bargain with transposable elements in plant pathogens. Trends Genet. 2022;38:222–30.

40. Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E. Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. G3 (Bethesda) [Internet]. 2021;11. Available from: http://dx.doi.org/10.1093/g3journal/jkab068

41. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences. 2020;117:9451–7.

42. Smit AFA, Hubley R, Green P. RepeatMasker [Internet]. 1996 [cited November, 28 2023]. Available from: https://www.repeatmasker.org/

43. Pritham EJ, Putliwala T, Feschotte C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene. 2007;390:3–17.

44. Helmstetter N, Harrison K, Gregory J, Harrison J, Ballou E, Farrer RA. A near-complete telomere-to-telomere genome assembly for Batrachochytrium dendrobatidis GPL JEL423 reveals a larger CBM18 gene family and a smaller M36 metalloprotease gene family than previously recognised. G3 [Internet]. 2024; Available from: https://academic.oup.com/g3journal/advance-article-abstract/doi/10.1093/g3journal/jkae304/7930337

45. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, et al. Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet. 2016;12:e1006108.

46. Sorwar E, Oliveira JIN, Malar C M, Krüger M, Corradi N. Assembly and comparative analyses of the Geosiphon pyriformis metagenome. Environ Microbiol. 2024;26:e16681.

47. Belyayev A. Bursts of transposable elements as an evolutionary driving force. J Evol Biol. 2014;27:2573–84.

48. Tedersoo L, Magurno F, Alkahtani S, Mikryukov V. Phylogenetic classification of arbuscular mycorrhizal fungi: new species and higher-ranking taxa in Glomeromycota and Mucoromycota (class Endogonomycetes). MycoKeys. 2024;107:273–325.

49. Gosling P, Proctor M, Jones J, Bending GD. Distribution and diversity of Paraglomus spp. in tilled agricultural soils. Mycorrhiza. 2014;24:1–11.

50. Säle V, Palenzuela J, Azcón-Aguilar C, Sánchez-Castro I, da Silva GA, Seitz B, et al. Ancient lineages of arbuscular mycorrhizal fungi provide little plant benefit. Mycorrhiza. 2021;31:559–76.

51. Mathieu S, Cusant L, Roux C, Corradi N. Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. New Phytol. 2018;220:1129–34.

52. Carducci F, Biscotti MA, Barucca M, Canapa A. Transposable elements in vertebrates: species evolution and environmental adaptation. The European Zoological Journal. 2019;86:497–503.

53. Omrane S, Audéon C, Ignace A, Duplaix C, Aouini L, Kema G, et al. Plasticity of the MFS1 promoter leads to multidrug resistance in the wheat pathogen Zymoseptoria tritici. mSphere [Internet]. 2017;2. Available from: http://dx.doi.org/10.1128/mSphere.00393-17

54. Feng G, Leem Y-E, Levin HL. Transposon integration enhances expression of stress response genes. Nucleic Acids Res. 2013;41:775–89.

55. Macquet J, Mounichetty S, Raffaele S. Genetic co-option into plant-filamentous pathogen interactions. Trends Plant Sci. 2022;27:1144–58.

579 56. Stürmer SL, Bever JD, Morton JB. Biogeography of arbuscular mycorrhizal fungi
580 (Glomeromycota): a phylogenetic perspective on species distribution patterns.
581 Mycorrhiza. 2018;28:587–603.

582 57. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-
583 scale multiple sequence alignments. Bioinformatics. 2018;34:2490–2.

584 58. Eddy S. HMMER user's guide. Department of Genetics, Washington University
585 School of Medicine. 1992;2:13.

586 59. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q
587 file manipulation. PLoS One. 2016;11:e0163962.

588 60. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-
589 generation sequencing data. Bioinformatics. 2012;28:3150–2.

590 61. Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV. A beginner's guide
591 to manual curation of transposable elements. Mob DNA. 2022;13:7.

592 62. EMBOSS: getorf manual [Internet]. 2000. Available from:
593 https://embossgui.sourceforge.net/demo/manual/getorf.html

594 63. Guerrero-Araya E. HmmPy: A Pure-Python parser of HMMER3 output [Internet].
595 Github; 2016 [cited 2024 Oct 15]. Available from:
596 https://github.com/EnzoAndree/HmmPy

597 64. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
598 tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.

599 65. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
600 features. Bioinformatics. 2010;26:841–2.

601 66. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD,
602 et al. SignalP 6.0 predicts all five types of signal peptides using protein language models.
603 Nat Biotechnol. 2022;40:1023–5.

604 67. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of apoplastic and cytoplasmic
605 effectors in fungi and Oomycetes. Mol Plant Microbe Interact. 2022;35:146–56.