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 distinct 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, their 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 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 identified in 24 genome assemblies, including 3 Paraglomerales, 3 Archaeosporales, 5 Glomerales, 11 Diversisporales, and 2 Entrophosporales (**Table 1**).

Table 1. Comparison between general and personalized transposable element 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.

Before curation, between 18% to 45% of genomes were covered by unknown TEs. Curation 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 repeats (annotated TEs and satellites) 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% [24,32] (**Figure 1a**).

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 also 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], are common to all Diversisporales species (i.e. Acaulosporaceae, Diversisporaceae, and Gigasporaceae families).

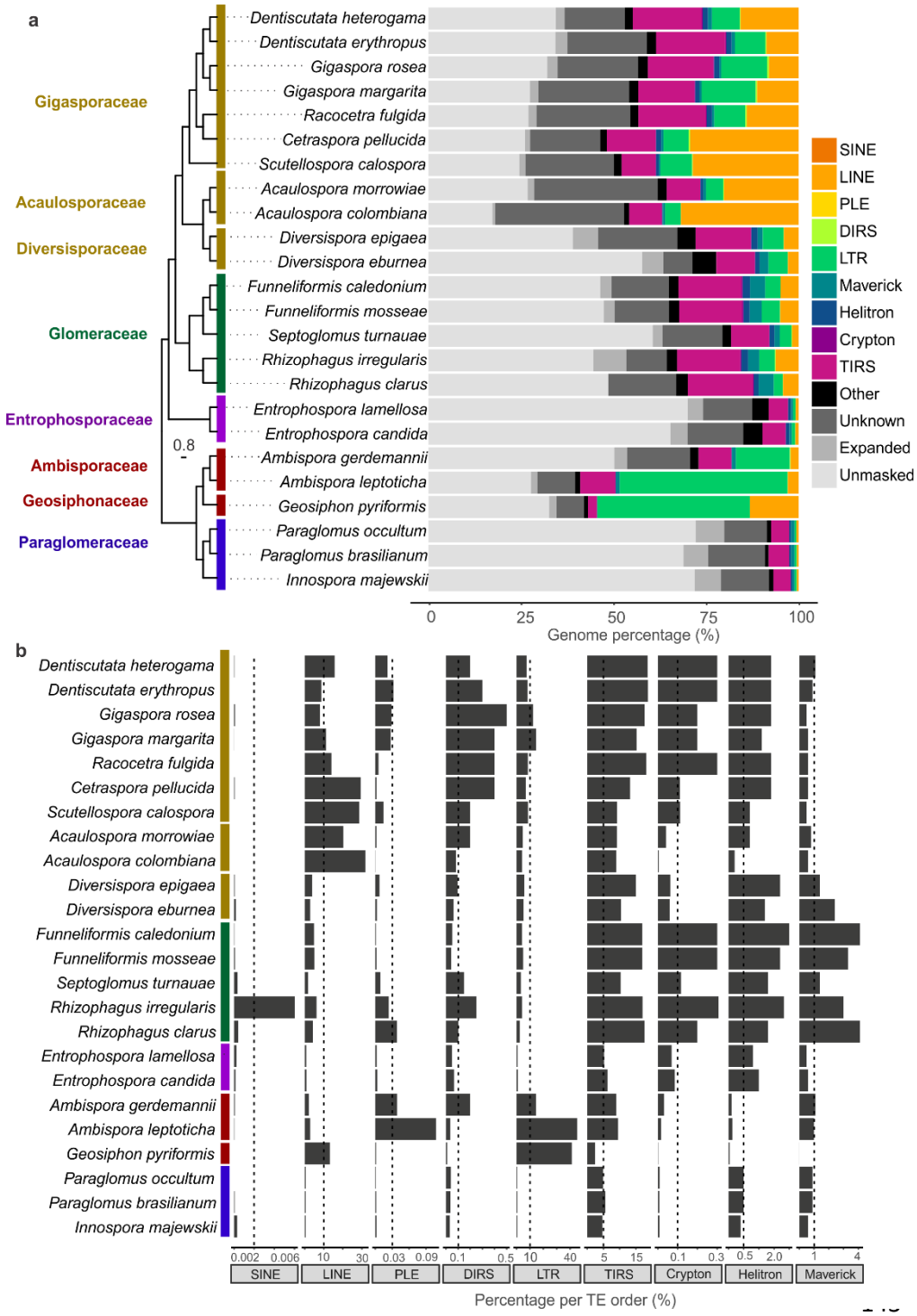


Figure 1 – Genome composition in Glomeromycotina species. **a)** Relative distribution of different types of sequences in AMF species. The phylogenetic tree is based on ortholog genes, and the bar indicates the level of nucleotide substitution, representing the distances between the taxa. The colors 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.

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 (**Figure 2a**). These AMF "core" elements represent on average 50% of the TEs copies in AMF genomes (**Figure 2b**). Among the core elements, 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" elements.

Overall, species from Glomerales and Diversisporales have more expansions of clade-specific elements 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 it could represent clade-specific TE invasions and expansions (**Figure 2c**).

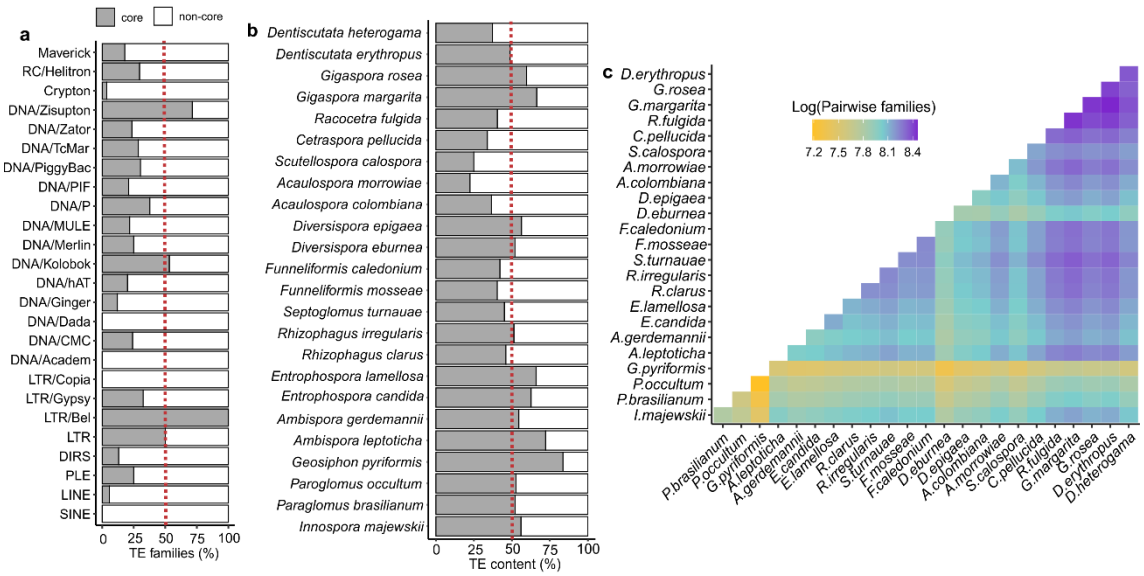


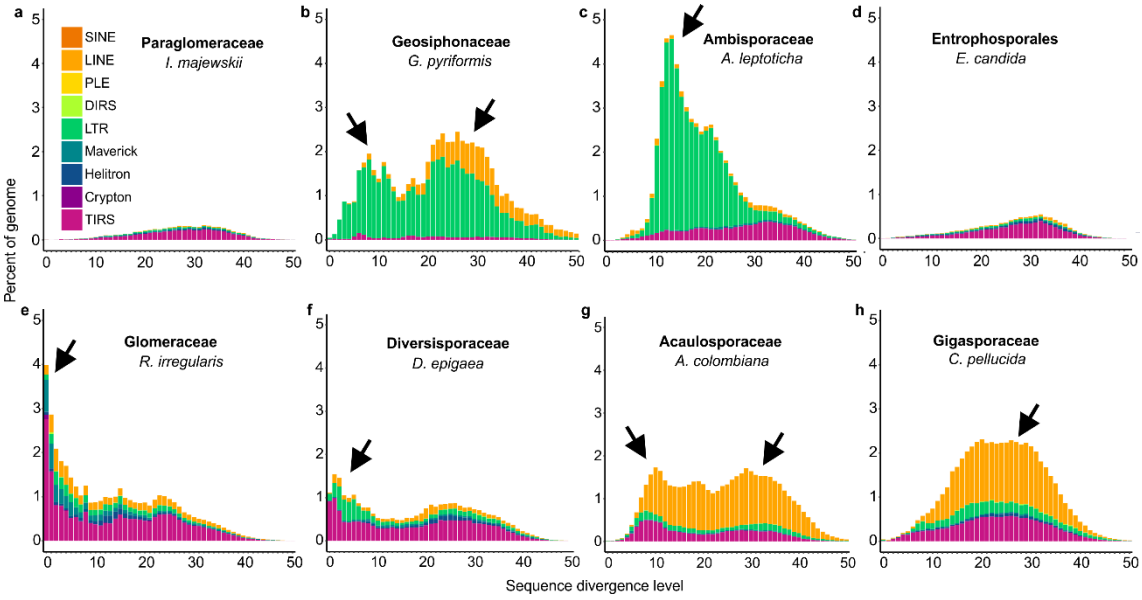
Figure 2 – TEs families shared across Glomeromycotina species. **a)** Percentages in relation to the total number of each superfamily that are present in all species (core) and in blank are the superfamily that are not present in all species (specific). **b)** Relation of core and specific families in 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 the 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.

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

173



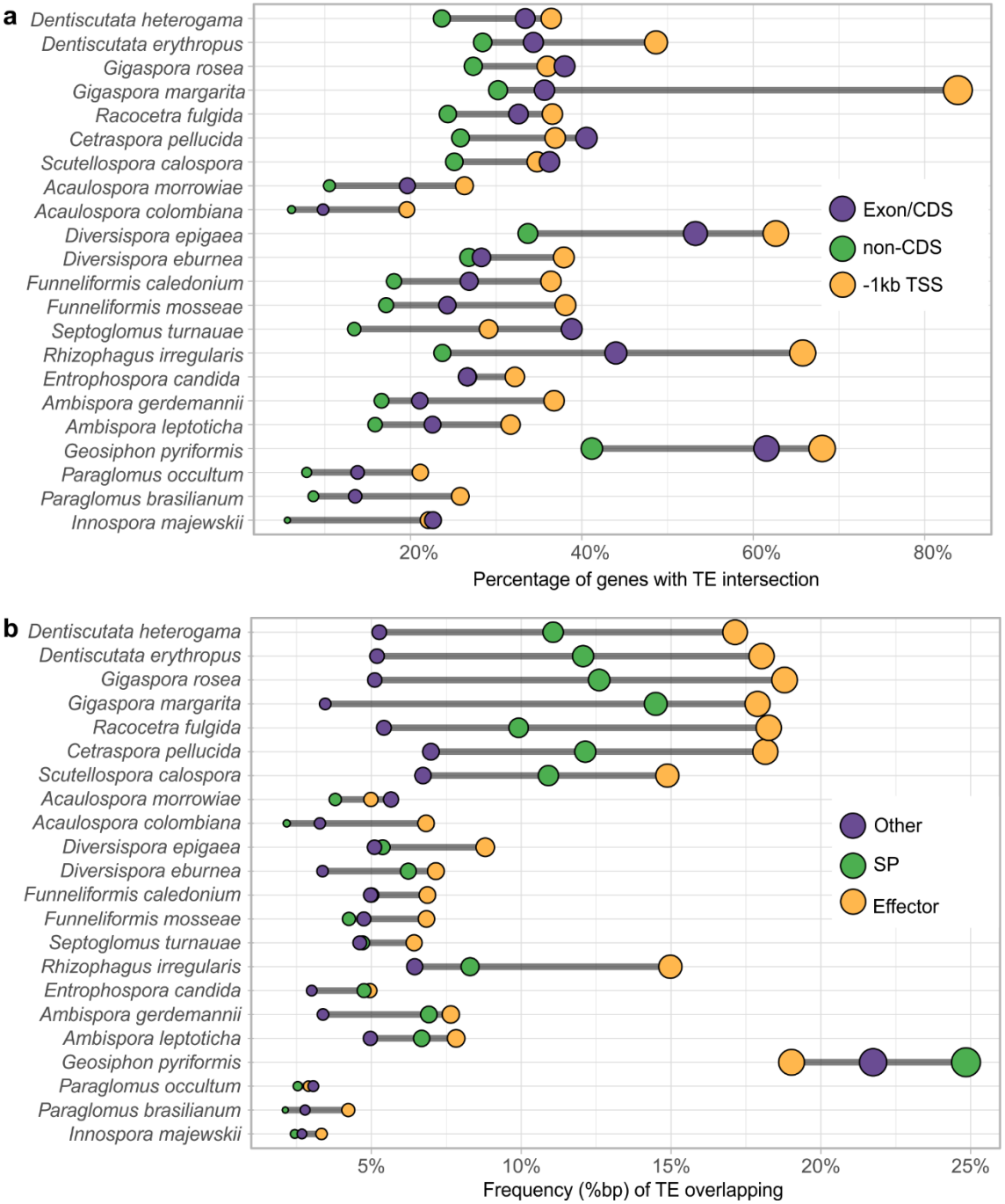
174

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.

175 *TEs are preferentially associated with promoters and effector genes in Glomeromycotina*

176 To determine TEs-gene associations, we investigated the frequency of genes intersecting
177 within 1000 bp upstream to their transcript start site (-1Kb TSS) – i.e. close to
178 the promoter and/or regulatory sequences that influence gene expression [33,34], as well
179 as within coding sequences (CDS/exons) and in non-coding sequences (untranslated
180 regions, UTRs; introns) (**Figure 4a**). In Glomeromycotina, LINE, LTR/Gypsy,

181 DNA/CMC, DNA/hAT, and DNA/MULE elements are the TE families commonly found
182 either near or within genes (**Figure S1**).



183

Figure 4 – Transposable elements association within genes. **a)** Percentages of genes harboring TE insertions in upstream to the transcription start site (TSS), in UTRs and introns (non-CDS) and exon/CDS regions. **b)** Frequency of TE overlapping within and surrounding genes: SP, secreted proteins, effectors and genes other than SPs; %bp.

184 Overall, TE insertions are significantly more common within -1Kb TSS regions
185 compared to both non-coding regions (UTRs and introns, average of 38% *vs* 20%; p-
186 value <0.001) and exons (average 38% *vs* 30%; p-value <0.001). For the -1Kb TSS
187 insertions, notable differences exist among families (SD 17%). For example, 20% of

genes in Paraglomerales have TEs in promoter regions compared to up to 60% in Glomerales and Diversisporales. For CDS/exons, *Paraglomus* spp. and *A. colombiana* have fewer genes with TE-associated exons at 13% and 9% respectively, whereas *D. epigea* and *G. pyriformis* have a higher proportion with 53% and 61% of exons associated with TEs.

We also found that the frequency of TE insertions is significantly higher (p-value <0.001) near (i.e. + or –1Kb) or within secreted proteins and candidate effectors compared to other protein-coding genes. Candidate effectors are also significantly more likely to be associated with TEs than with secreted proteins (p-value <0.001) (**Figure 4b**). *G. pyriformis* is also characterized by a much higher number of genes with TEs insertions within exons (**Figure 4a**) and non-secreted proteins (**Figure 4b**). It is noteworthy that the degree of fragmentation of the genome assembly does not affect the general observations; in almost all cases, TEs are more likely to be found near or within promoter regions (-1Kb) and in association with secreted proteins and effectors.

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 [35,36], 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 [37]. 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.

In general, AMF and other fungi carry expansions of retrotransposons such as LINEs and LTRs, however, some species deviate from this trend [21,37]. For example, insertions from DNA transposons such as TIRS, Crypton, Helitron, and Maverick elements are abundant in Glomerales. The patterns of TE evolution we observed also confirm that, together with the number of genes [13], TEs can vary dramatically in presence/absence and relative abundance between AMF relatives and within TE families. For example, the virus-like element Maverick is expanded exclusively in Glomerales. This element has only been reported in the plant pathogen *Phakosora*

221 *pachyrhizi* and in the AMF *Rhizophagus (Glomus) intraradices* [38] and it is thus
222 intriguing to speculate that it may have originated via horizontal gene transfer or lost in
223 the other fungal lineages.

224 In general, our work supports the view that obligate biotrophs carry more
225 transposable elements than non-obligate fungi, linking the pattern of TE insertions to their
226 lifestyle [37,39]. Within this context, a similar trend was observed in the cyanobacteria
227 endosymbiont *G. pyriformis*, which showed a greater expansion of mobile genetic
228 elements compared to any free-living relatives [40].

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

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

237 Remarkably, the evolution of these elements can sometimes go beyond
238 phylogenetic relationships. For example, members of phylogenetically distinct families,
239 Paraglomeraceae and Entrophosporaceae, can have virtually identical TE distributions.
240 This similarity may reflect independent and adaptive TEs retentions due to similarities in
241 (epi) genomic architecture, such as methylation, or ecological traits [37].

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

243 The chromosomes of the model species *R. irregularis* are separated into two main regions,
244 called A/B compartments [12,26], and our analysis indicates that this organization is
245 shared by most Glomeromycotina. In particular, TE density is always significantly higher
246 near secreted proteins and candidate effectors in AMF genomes; a characteristic that is
247 highly reminiscent of the B-compartment and that shares further highlights striking
248 similarities in the genome biology of AMF and many filamentous fungal plant pathogens
249 [24,29].

250 Specifically, these independent symbiotic and pathogenic lineages both carry
251 genes involved in dialogue with the host (effectors) within fast-evolving and repeat-rich
252 regions of the genome. As in fungal pathogens, the association between effectors and TE

253 in AMF is likely to diversify these molecules, allowing for improved adaptation and
254 symbiosis with a wider range of hosts. Notably, this organization can also lead to a
255 "devil's bargain" situation, as TE insertions near or within effectors could degenerate
256 these genes by increasing the chances of deleterious mutations for the symbiosis [41]

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

258 Bursts of TEs promote genomic rebuilding, and this mechanism is often associated with
259 higher speciation rates [42]. In this context, it is intriguing to speculate that the reduced
260 TE load of Paraglomerales, together with its flat landscape distribution (indicating low
261 TE activity and highly stable genome) is linked with a limited species richness within this
262 family [43]. Indeed, to date, Paraglomerales has fewer species formally recognized
263 compared to other Glomeromycotina, including other early diverging lineages like
264 Archaeosporales – e.g. 11 Paraglomerales and 19 Archaeosporales [43]. In contrast, more
265 divergent AMF orders such as Glomerales and Diversisporales featured high levels of TE
266 bursts within their genomes and are significantly more species-rich, with 268 and 70
267 species, respectively [43].

268 The lower TE abundance of Paraglomerales also supports the hypothesis that
269 fewer gene/TE associations with reduced mycorrhizal capabilities [14] recently reported
270 for this order compared to AMF relatives [44,45]. Here, lower TE loads may reduce
271 the chances of generating genomic diversity necessary for adapting to environmental
272 change [27,46], especially if mating is rare or absent [10,24]. In support of this, most
273 other AMF groups with high genomic diversity and adaptability [13], such as
274 Glomeraceae, have experienced notable TE bursts in the past and carry evidence of recent
275 and ongoing TE activity.

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

277 Our analyses also revealed that TEs are preferentially located near promoter regions when
278 associated with genes. TE insertions near genes may impact gene expression by inducing
279 methylation or altering transcription factor binding sites [47], and even promoting
280 antifungal resistance as observed in a plant pathogen [48]. In the yeast *Saccharomyces*
281 *cerevisiae*, LTR elements are preferentially inserted 700 bp upstream to coding genes and
282 show synergistic expression during heat-stress induction [49].

283 In *R. irregularis*, co-expression has been observed between SP and upstream TEs
284 in the B-compartment, with TEs showing higher expression during symbiosis [23] than

285 in spores [24]. In addition, TEs showed strong methylation signals in germinating spores
286 [26], which coincides with high expression of an RNAi-related gene [23]. This suggests
287 that TEs may undergo demethylation during symbiosis, leading to increased expression
288 of both TEs and symbiosis-related genes, such as secreted proteins and effectors. Again,
289 this phenomenon has been observed in plant pathogens, where genes with TE insertions
290 within -1Kb are highly expressed during early infection, revealing a synchrony between
291 TEs and infection-related genes [33,49].

292 Remarkably, our analyses revealed that regions containing effector genes have
293 significantly more TE insertions compared to other coding sequences. This phenomenon
294 has also been observed in plant pathogens, where TE accumulation drives effector
295 diversification, thereby increasing fungal virulence [28,41]. On the other hand, the plant
296 evolves its immune system to combat infections, and again, TE-driven modifications
297 serve as a source for the fungus to overcome host defenses, in an ongoing coevolutionary
298 process [50]. Similarly, in AMF, diversified effectors could lead to broader adaptation to
299 different hosts and environments, as these cosmopolitan symbionts can associate with
300 many different hosts [51]. Given the high expression of TEs during symbiosis and
301 frequent association with SP, it would be interesting to investigate changes in methylation
302 to understand how these sequences become more actively transcribed upon contact with
303 the host. Additionally, it would be valuable to examine which factors, whether biotic or
304 abiotic stressors, can demethylate TEs in spores and reactivate TE mobilization as a
305 source of variability.

306 **Material and Methods**

307 *Genome assemblies*

308 We downloaded 21 genome assemblies, protein, and gene annotation files from all
309 Glomeromycotina orders from NCBI and provided 3 new genome assemblies **Table S1**.

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

311 We constructed HMM profiles for several transposition domains described in **Table S2**.
312 Sequences for each domain were aligned separately using mafft [52], then converted to
313 Stockholm format using esl-format and submitted to hmmbuild [53] to generate the HMM
314 profiles. The profiles were concatenated and prepared for hmmscan using hmmpress. All
315 models are available on https://github.com/jordana-olive/Transposable-Elements-in-AM-
316 Fungi-Evolutionary-Parallels-with-Fungal-Plant-Pathogens.

317 *Repeat library generation*

318 Each genome was submitted to RepeatModeler2.0 [35] with the -LTRStruct option to
319 generate the repeat sequence files. All the repeat libraries were concatenated into a single
320 file from which redundant sequences and sequences shorter than 1000 nucleotides were
321 removed using seqkit rmdup. These files were named "non-curated library".

322 *Transposable elements curation and annotation*

323 We predicted open reading frames (ORFs) in the "non-curated library" using getorf,
324 selecting translated ORFs of at least 300 nucleotides. We submitted the ORFs from the
325 "non-curated library" to hmmsearch to retrieve the sequences containing the transposition
326 domains using the hmm models. We selected the best match of each sequence using
327 HmmPY.py [54] (e-value 1e-17, -c 0.35, -h 50). Sequences carrying transposition
328 domains were renamed with the corresponding family domain. Curation generated the
329 "final library" to be used for transposable element annotation. TE annotation was
330 performed in each genome using RepeatMasker2.0.4 -s -a modes [36]. The alignment
331 files generated by the annotation process were used to calculate the degree of divergence
332 based on Kimura values using the scripts calcDivergenceFromAlign.pl and
333 createRepeatLandscape.pl provided by the RepeatMasker tool.

334 *Orthologues genes and AMF phylogenetic tree*

335 The protein sequences of the studied genomes were submitted to OrthoFinder [55] to
336 construct a phylogenetic tree based on orthologous genes. The tree was rooted with
337 species from Paraglomerales and Archeosporales and visualized using the ggtree R
338 package.

339 *TE association with genes*

340 We compared the intersection between TEs and genes using bedtools intersect [56] in gff
341 from Genebank and our TE annotation. We also split the annotations into three files: 1000
342 base pairs (bp) upstream of the gene's transcription target site, exons and other non-
343 coding regions (5' & 3'UTR and introns). We considered a valid intersection when the
344 TE overlap was at least 50 nucleotides long.

345 *Prediction of secreted proteins and effectors*

346 Secreted proteins were predicted in all species using SignalP-6.0 [57]. From the predicted
347 secreted proteins, we used the EffectorP-3.0 [58] to annotate the candidate effectors.

Acknowledgments

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

Author's contribution

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

Supplementary information

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

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

Figure S1 – Association of genes per TE order.

References

1. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, et al. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia. 2016;108: 1028–1046.

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: 864–868. doi:10.1126/science.abg0929

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: 1172–1173. doi:10.1126/science.aam9970

4. Keymer A, Pimprikar P, Wewer V, Huber C, Brands M, Bucerius SL, et al. Lipid transfer from plants to arbuscular mycorrhiza fungi. Elife. 2017;6. 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. 1994. doi: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: 765–778. doi:10.1016/j.tplants.2020.05.002

7. Normark BB, Judson OP, Moran NA. Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society. 2003. pp. 69–84. doi:10.1046/j.1095-8312.2003.00182.x

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. doi:10.1371/journal.pone.0080729

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: 950–958. doi:10.1093/gbe/evr089

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. doi:10.1016/j.tim.2020.03.013

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. 2023. doi:10.1038/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. 2024. doi: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. 2022;8. doi:10.1099/mgen.0.000810

15. Simonti CN, Pavlicev M, Capra JA. Transposable Element Exaptation into Regulatory Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints. Mol Biol Evol. 2017;34: 2856–2869. doi:10.1093/molbev/msx219

16. Sela N, Mersch B, Hotz-Wagenblatt A, Ast G. Characteristics of transposable element exonization within human and mouse. PLoS One. 2010;5: e10907. doi:10.1371/journal.pone.0010907

17. Schmitz J, Brosius J. Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie. 2011;93: 1928–1934. doi:10.1016/j.biochi.2011.07.014

18. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8: 973–982. doi:10.1038/nrg2165

19. Oggenfuss U, Croll D. Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. PLoS Pathog. 2023;19: e1011130. doi:10.1371/journal.ppat.1011130

20. Cosby RL, Chang N-C, Feschotte C. Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 2019;33: 1098–1116. doi:10.1101/gad.327312.119

21. Venice F, Ghignone S, Salvioli di Fossalunga A, Amselem J, Novero M, Xianan X, et al. At the nexus of three kingdoms: the genome of the mycorrhizal fungus Gigaspora margarita provides insights into plant, endobacterial and fungal interactions. Environ Microbiol. 2020;22: 122–141. doi:10.1111/1462-2920.14827

22. Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, et al. Comparative genomics of Rhizophagus irregularis, R. cerebriforme, R. diaphanus and Gigaspora rosea highlights specific genetic features in Glomeromycotina. New Phytol. 2019;222: 1584–1598. doi:10.1111/nph.15687

23. Oliveira JIN, Corradi N. Strain-specific evolution and host-specific regulation of transposable elements in the model plant symbiont Rhizophagus irregularis. G3 . 2024. doi:10.1093/g3journal/jkae055

24. Dallaire A, Manley BF, Wilkens M, Bista I, Quan C, Evangelisti E, et al. Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus Rhizophagus irregularis. Genome Res. 2021;31: 2290–2302. doi:10.1101/gr.275752.121

25. Silvestri A, Turina M, Fiorilli V, Miozzi L, Venice F, Bonfante P, et al. Different Genetic Sources Contribute to the Small RNA Population in the Arbuscular Mycorrhizal Fungus Gigaspora margarita. Front Microbiol. 2020;11: 1–15. doi:10.3389/fmicb.2020.00395

26. Yildirir G, Sperschneider J, Malar C M, Chen ECH, Iwasaki W, Cornell C, et al. Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 2022;233: 1097–1107. doi:10.1111/nph.17842

27. Reinhardt D, Roux C, Corradi N, Di Pietro A. Lineage-Specific Genes and Cryptic Sex: Parallels and Differences between Arbuscular Mycorrhizal Fungi and Fungal Pathogens. Trends Plant Sci. 2021;26: 111–123. doi:10.1016/j.tplants.2020.09.006

28. Seidl MF, Thomma BPHJ. Transposable Elements Direct The Coevolution between Plants and Microbes. Trends Genet. 2017;33: 842–851. doi:10.1016/j.tig.2017.07.003

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: 1091–1100. doi:10.1101/gr.204974.116

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. 2021;2. doi: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: 361–371. doi:10.1007/s00572-022-01091-4

32. 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: 1161–1171. doi:10.1111/nph.14989

33. 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: 221–239. doi:10.1093/molbev/msz216

34. 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. doi:10.1186/s12915-018-0543-2

35. 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: 9451–9457. doi:10.1073/pnas.1921046117

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

37. Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Rep. 2019;9: 4307. doi:10.1038/s41598-019-40965-0

38. 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. doi:10.1016/j.gene.2006.08.008

39. 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. doi:10.1371/journal.pgen.1006108

40. 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. doi:10.1111/1462-2920.16681

41. Fouché S, Oggenfuss U, Chanclud E, Croll D. A devil's bargain with transposable elements in plant pathogens. Trends Genet. 2022;38: 222–230. doi:10.1016/j.tig.2021.08.005

42. Belyayev A. Bursts of transposable elements as an evolutionary driving force. J Evol Biol. 2014;27: 2573–2584. doi:10.1111/jeb.12513

43. 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. doi:10.3897/mycokeys.107.125549

44. Gosling P, Proctor M, Jones J, Bending GD. Distribution and diversity of Paraglomus spp. in tilled agricultural soils. Mycorrhiza. 2014;24: 1–11. doi:10.1007/s00572-013-0505-z

45. 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: 559–576. doi:10.1007/s00572-021-01042-5

46. Mathieu S, Cusant L, Roux C, Corradi N. Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. New Phytol. 2018;220: 1129–1134. doi:10.1111/nph.15275

47. 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. doi:10.1080/24750263.2019.1695967

48. 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. 2017;2. doi:10.1128/mSphere.00393-17

49. Feng G, Leem Y-E, Levin HL. Transposon integration enhances expression of stress response genes. Nucleic Acids Res. 2013;41: 775–789. doi:10.1093/nar/gks1185

50. Macquet J, Mounichetty S, Raffaele S. Genetic co-option into plant-filamentous pathogen interactions. Trends Plant Sci. 2022;27: 1144–1158. doi:10.1016/j.tplants.2022.06.011

51. Stürmer SL, Bever JD, Morton JB. Biogeography of arbuscular mycorrhizal fungi (Glomeromycota): a phylogenetic perspective on species distribution patterns. Mycorrhiza. 2018;28: 587–603. doi:10.1007/s00572-018-0864-6

52. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics. 2018;34: 2490–2492. doi:10.1093/bioinformatics/bty121

53. Eddy S. HMMER user's guide. Department of Genetics, Washington University School of Medicine. 1992;2: 13. Available: http://www.eddylab.org/software/hmmer/Userguide.pdf

54. Guerrero-Araya E. HmmPy: A Pure-Python parser of HMMER3 output. Github; 2016. Available: https://github.com/EnzoAndree/HmmPy

55. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

540 56. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
541 features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

542 57. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD,
543 et al. SignalP 6.0 predicts all five types of signal peptides using protein language
544 models. Nat Biotechnol. 2022;40: 1023–1025. doi:10.1038/s41587-021-01156-3

545 58. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of apoplastic and cytoplasmic
546 effectors in fungi and Oomycetes. Mol Plant Microbe Interact. 2022;35: 146–156.
547 doi:10.1094/MPMI-08-21-0201-R