Ancestral and recent bursts of transposition shaped the massive genomes of plant pathogenic rust fungi

Authors:

Emma Corre[1], Emmanuelle Morin[1], Sebastien Duplessis[1*] and Cecile Lorrain[2]*

*Corresponding authors: sebastien.duplessis@inrae.fr; cecile.lorrain@usys.ehz.ch

[1] Université de Lorraine, INRAE, UMR 1136 IAM, F-54000, Nancy, France

[2] Plant Pathology Group, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

Abstract

Background

Transposable elements (TEs) play a crucial role in genome evolution, influencing gene regulation, diversity, and genome architecture. Rust fungi of the order Pucciniales (Subphylum Pucciniomycotina) are the largest group of obligate biotrophic plant pathogens and harbor some of the largest and most TE-rich genomes - up to 1.2Gb - compared to other fungi. This global genomic expansion contrasts with the smaller genomes and minimal mobilome found in other Pucciniomycotina species. Despite the availability of high-quality genome assemblies, our understanding of TE dynamics in Pucciniales remains limited due to inconsistent and incomplete TE annotations.

Results

We analyzed the mobilomes of 13 Pucciniomycotina species, producing a manually curated TE library for each genome. In Pucciniales, TEs occupy 47–92% of the genome, whereas 2-36% of TEs are detected in other Pucciniomycotina genomes. The comparison of gene and TE repertoires indicates that TEs, particularly LTR-retrotransposons and TIR-DNA transposons, are the primary contributors to the genome expansion of the Pucciniales. We reconstructed the proliferation histories of TEs in the Pucciniales, combining sequence similarity, clustering, and molecular clock approaches. We highlight recent and ancient TE invasions with some LTR-Gypsy elements predating the divergence of Pucciniomycotina (~176 Mya), while most TE accumulation in Pucciniales occurred within the last 50 Mya. However, the TE invasions in the Pucciniales genomes do not seem to result from specific deficiencies in known TE-control mechanisms.

Conclusion

Our findings uncover extensive TE proliferation in Pucciniales, predominantly driven by recent LTR-Gypsy expansions. The retention of ancestral TEs and the consistently TE-rich

genomes observed in Pucciniales highlight TE proliferation as an ancestral genomic feature in rust fungi.

Keywords: transposable elements, fungal genomics, rust fungi, plant-pathogens

BACKGROUND

Rust fungi (order Pucciniales) are ubiquitous plant pathogens that infect a wide range of vascular plants, including some of the world's most important crops and trees, such as wheat, soybean, poplar, and coffee trees [1, 2]. Epidemics caused by rust diseases can lead to significant yield losses, posing a substantial threat to global food security [2–4]. The Pucciniales are classified within the subphylum Pucciniomycotina, phylum Basidiomycota, in the Fungal Kingdom and represent the largest order of obligate biotrophic fungi, comprising over 8,000 species [5]. The Pucciniales are estimated to have emerged between 175-230 million years ago [5, 6]. Phylogenetic studies suggest that the diversification of rust fungi is closely tied to their obligate host association and coevolution, with occasional host jumps to unrelated species [7]. The obligate host association has shaped the genomic architecture of rust fungi. For instance, Pucciniales genomes show a reduced repertoire of plant cell wall-degrading enzymes, secondary metabolites, and transporter genes, typical traits of obligate biotrophic fungi [8–10]. However, Pucciniales genomes are exceptionally large, ranging from approximately 80 Mb to over 1.2 Gb [11, 12], and are expected to present even greater genome sizes [13], contrasting with the average ~46.5 Mb genome size commonly found in Basidiomycota [14]. These expanded genomes are characterized by lineage-specific multigene families, an abundance of secreted proteins encoding genes, and a high proportion of repetitive elements [8, 9, 11]. The genome size expansion observed in Pucciniales parallels other filamentous plant pathogens. For instance, several Ascomycetes and Oomycetes, such as *Blumeria graminis* and *Phytophthora infestans*, also possess large genomes exceeding 100 Mb, similarly characterized by a significant proportion of repetitive elements [15]. The high repeat content of Pucciniales genomes has posed challenges in generating high-quality genome assemblies, which are crucial for studying their organization and evolutionary dynamics.

The increasing affordability of new-generation sequencing technologies has revolutionized the sequencing of high-quality reference genomes for numerous non-model organisms, including species within the Pucciniales order [11, 16, 17]. The most recent genome assemblies consist of gapless and phased chromosomes generated from dikaryotic spores [11, 18]. Analyses of the most recent Pucciniales genomes have confirmed that transposable elements (TEs) are major drivers of genome architecture and evolution in rust fungi [11, 17]. TEs are ubiquitous mobile DNA sequences capable of proliferating within their host genomes. TEs are classified into two main classes based on transposition mechanisms [19]. Class I elements, or retrotransposons, replicate through an RNA intermediate, enabling their insertion into new genomic positions. Class II elements, or DNA transposons, move either through direct excision from double-stranded DNA or via single-strand excision followed by a rolling-circle replication mechanism [19]. Classes are subdivided into Orders and Super-families (SF) of TEs based on sequence structures [19]. Finally, TE copies with similar sequences are grouped in TE families. Both classes contain autonomous elements, which possess the machinery for independent transposition, and non-autonomous elements, which rely on the transposition machinery of other elements to propagate [19].

TE-mediated genomic plasticity may confer significant adaptive advantages in plant-pathogen interactions. TEs contribute to genomic plasticity by facilitating chromosomal rearrangements, promoting compartmentalization, duplicating or deleting genes, and influencing gene expression [20]. In the soybean rust *Phakopsora pachyrhizi*, a subset of TEs is expressed alongside effectors during the early stages of infection, suggesting that transposition activity may be influenced by the plant-rust interaction, as also reported in other pathosystems [11, 21]. There is also evidence that fungal genomes have co-opted some TEs [22]. For instance, the effector AvrK1 in the TE-rich *B. graminis* f. sp. *hordei* genome is derived directly from a LINE retrotransposon [23]. The interplay between TE expansion and transposition regulation is largely unknown in TE-rich fungal species, including Pucciniales. However, these species appear to have evolved tolerant genomes

that can accommodate TE activity, unlike other fungal species with compact genomes [24]. This tolerance may play a crucial role in the adaptability of rust fungi, contributing to their success as plant pathogens.

TE propagation can harm host genomes, and species have evolved defense mechanisms to regulate TE proliferation. In fungi, TE activity is often controlled through epigenetic silencing mechanisms [25, 26]. TEs are frequently associated with heterochromatin histone marks and cytosine DNA methylation (5mC) [27]. Repeat-induced point mutation (RIP) is also a defense mechanism that has been so far specifically found within Ascomycetes [8]. RIP inactivates TEs through cytosine-to-thymine mutations and involves the DNA methyltransferases DIM2 and RID [25, 27, 28]. RNA interference (RNAi) adds another layer of TE control. The RNAi pathway relies on three core components: Dicer, which processes double-stranded RNA into small interfering RNAs (siRNAs); Argonaute, which incorporates these siRNAs to guide the RNA-induced silencing complex (RISC); and RNA-dependent RNA polymerase (Rdmd), which amplifies siRNAs to sustain the silencing response [29]. This pathway targets TE transcripts for degradation or chromatin modifications, thereby repressing TE expression and preventing transposition [30]. Despite its demonstrated importance in plants and animals, RNAi-mediated TE regulation in fungi remains comparatively less studied. Comparative genomic studies suggest that fungi exhibit significant diversity in their TE control mechanisms, with some species retaining these pathways while others have lost them through evolution [25, 27, 29].

The present study investigates the role of TEs in shaping Pucciniales genome architecture over evolutionary timescales. We tested four key hypotheses: (i) TE invasions in Pucciniales are the result of an ancestral transposition burst that has been preserved, (ii) TE invasions in Pucciniales arise from species-specific independent bursts, (iii) TE proliferation reflects a combination of ancestral bursts retained within the genome and recent species-specific transposition events, and (iv) TE proliferation is an ancestral characteristic of Pucciniomycotina evolution, which has been largely lost in other species but retained in

Pucciniales. To achieve this, we generated a comprehensive annotation and curation of mobilomes across 13 selected Pucciniomycotina species with comparative genomic analyses, focusing on TE diversity, abundance, and evolutionary dynamics. Additionally, we compared gene content and patterns of multigene family expansion and contraction across Pucciniomycotina genomes to detect potential evidence of TE control presence/absence.

RESULTS

The massive genome size of Pucciniales is determined by their extensive mobilomes and large gene catalogs

To explore the contributors to the genome size expansion in Pucciniales, we selected a panel of 13 Pucciniomycotina species with reference genomes comprising nine Pucciniales species and four other Pucciniomycotina species (Fig. 1A; Table S1). We also included the complete genomes of two *Puccinia graminis* f. sp. *tritici* strains to compare intra-specific variation. The size of the 10 selected Pucciniales genomes ranges from 109 Mb (*Melampsora larici-populina*) to 1.057 Gb (*P. pachyrhizi MT2006*), contrasting with the much smaller genomes of the four other Pucciniomycotina genomes, with sizes ranging from 20 Mb (*Rhodosporidium toruloides*) to 33 Mb (*Microbotryum lychnidis-dioicae*) (Table S1; Fig. 1B). We compared gene contents and their contribution to the genome inflation of Pucciniales. The Pucciniales display a minimum of 15,984 genes in *Melampsora americana* and a maximum of 19,618 genes in *P. pachyrhizi*. In contrast, the gene catalogs in the other Pucciniomycotina species range from 7,819 genes in *M. lychnidis-dioicae* to 9,854 in *Leucosporidia creatinivora* (Table S1). The difference in gene numbers between Pucciniales and other Pucciniomycotina species is significantly different (Wilcoxon test p-value < 0.05), indicating that the genes slightly contribute to the Pucciniales genome expansion (Table S1). The TE coverage in Pucciniales genomes ranges from 47% to 92.3% in *M. larici-populina* and *P. pachyrhizi*, respectively (Fig. 1B; Table S1). We observed variation of the TE

coverage within the Pucciniales order and between closely related species. For instance, in the Melampsoraceae, TE content represents 47.0% of the 109 Mb genome of *M. larici-populina* to 83.7% of the 482 Mb genome of *Melamspora lini*. In the Pucciniaceae, TEs covered from 49.4% of the 150 Mb genome of *Puccinia striiformis* f. sp. *tritici* to 77.3% of the 206 Mb genome of *Puccinia hordei* (Fig. 1B). In comparison, the TE coverage in the other Pucciniomycotina ranges between 2.3% in *R. toruloides* and 36% in *L. creatinivora*. Together, these results highlight the combined contributions of gene content and TEs to the extensive genome size variation and inflation observed in Pucciniales, distinguishing them from the other Pucciniomycotina species.

LTR retrotransposons are the major contributors to the extensive mobilome and genome size expansion in the Pucciniales

To detect if TE proliferation in the Pucciniales is linked to specific elements, we further examined the TE landscapes in the genomes of the selected Pucciniomycotina species using REPET v3.0 followed by three rounds of manual curation to build a high-quality Pucciniomycotina TE library. In total, we annotated 41,940 TE consensus sequences, constituting the Pucciniomycotina TE library. TE annotation before curation detected 24,165 (57.6%) classified TE consensus sequences, 17,775 (42.4%) unclassified TE consensus sequences, and 2,176 (5.2%) TE consensus sequences with conflictual annotations (Fig. S1). After the first round of manual curation, we reduced the number of conflictual annotations to 685 (1.6%) TE consensus sequences but could not resolve unclassified annotations. Therefore, we used MCHelper [31] and improved the annotation of ~62% (10,985) of previously unclassified TE consensus sequences (Fig. S1). We used a final step during which we clustered all TE consensus sequences from the Pucciniomycotina TE library to reduce redundancy and correct the remaining conflictual annotations. Our final curated Pucciniomycotina TE library comprised 29,050 (71.96%) classified TE consensus sequences and 11,319 (28.03%) unclassified TE consensus sequences (Tables S2-S4).

On average, the Pucciniales genomes are covered by 33.9% of Class I TEs, ranging from 15.4% (16.9 Mb) in *M. larici-populina* to 48.8% (515.8 Mb) in *P. pachyrhizi*, and 23.5% of Class II TEs ranging from 13.2% in *M. lini* haplotype C (31.6Mb) to 31.8% in *Puccinia triticina* haplotype A (39.3 Mb) (Fig. 2A; Table S2). By contrast, the Class I and Class II genome coverages were drastically lower within the other Pucciniomycotina, with an average of 13.7% and 3.5%, respectively (Fig. 2A; Table S2). Class I elements represent the most abundant elements in the Pucciniales except for the three genomes of *M. larici-populina* with 15.4% Class I (16.9 Mb), *Melampsora americana* with 18.8% Class I (21.1 Mb) and *P. striiformis* f. sp. *tritici* 18.1% Class I (13.7 Mb) (Fig. 2A; Table S2). These results indicate an important contribution of Class I elements in the genome expansions compared to the Class II elements in six of the nine Pucciniales species.

The TE Orders content of the Pucciniales genomes predominantly comprises LTR retrotransposons and TIR DNA transposons. LTRs represent 14.4% of the genome to 63.0% in *M. americana* (16.2 Mb) and *M. lini* haplotype C (15.1 Mb), respectively (Fig. 2B-C; Table S3). At the LTR super-family level, we mostly found Copia and Gypsy elements, representing 2.9% of genome coverage (3.2 Mb in *M. larici-populina*) to 12.5% (3 Mb in *M. lini* haplotype C) and 5.5% (6.2 Mb in *M. americana*) to 48.2% (11.6 Mb in *M. lini* haplotype C), respectively (Fig. 2 D-E; Table S4). The non-LTR retrotransposons, such as LINE elements, represent from 0.4% to 2.6% of the Pucciniales genomes. The non-autonomous elements such as SINE represent less than 0.2% of the Pucciniales genomes (Tables S3 and S4). In the other Pucciniomycotina genomes, the amount of LINE covers up to 5.2% of the genome of *M. lychnidis-dioicae* (1.4 Mb). TIRs represent most DNA transposons within the Pucciniales genomes, with 10.2% to 28.1%, followed by non-autonomous MITE representing 0.6% to 5.0% (Table S3). Interestingly, some TIRs super-family like CACTA, hAT, MuDR, PIF-harbinger, and Tc1-Mariner are found only in the Pucciniales and not in the other Pucciniomycotina (Fig. 2 D-E; Table S4). The Helitron elements cover a small proportion of the genome, from 0.2% in *M. lini* haplotype C (6.2 Mb) to 2.6% in *P. triticina*

haplotype A (3.2 Mb). These results indicate that at the TE-Order level, the Pucciniales display an overall similar TE landscape with a majority of LTRs and TIR elements, suggesting ancient proliferation of TEs in the order Pucciniales. However, the variable amount of LTR and TIR elements found between Pucciniales genomes, even at the haplotype level, support additional bursts of transposition that occurred throughout the evolution of the order Pucciniales.

Temporal dynamics of TE activity shaped the genome architecture of Pucciniales

We hypothesized that the proliferation of TEs observed in the Pucciniales could have emerged through different scenarios throughout the evolution of the Pucciniomycotina. First, we used a within genome sequence similarity-based approach to compare TEs from the oldest insertions to the most recent insertions as follows: i) divergent TE consensus sequences containing copies with less than 85% sequence identity, ii) intermediate TE consensus sequences with copies of 85% to 95% sequence similarity and iii) conserved TE consensus sequences comprising copies with more than 95% sequence identity [11, 12]. The average sequence identity of all TE consensus sequences per genome ranges from 79.0% (*R. toruloides*) to 91.9% (*M. lychnidis dioicae*). We observed 9.3-66.7% divergent (*M. lychnidis-dioicae* and *R. toruloides*, respectively), 33.3-76.4% intermediate (*R. toruloides* and *Microbotrium intermedium*, respectively), and 0.0-26.3% conserved (*R. toruloides* and *M. lychnidis-dioicae*, respectively) TE consensus sequences in the other Pucciniomycotina (Fig. 3A-I; Table S5). Within the Pucciniales, TE consensus sequences similarity profiles revealed 29.9-66.1% divergent (*P. pachyrhizi* and *M. larici-populina*, respectively), 27.4-65.6% intermediate (*M. americana* and *P. pachyrhizi*, respectively), and 2.6-7% conserved TE consensus sequences (*Melampsora allii-populina* and *P. triticina*, respectively) (Fig. 3A-I; Table S6). Among the Pucciniales genomes, we observed that LTRs and TIRs compose most of each relative TE age category (Fig. 3A-I; Table S6). LTR elements represent 17.4-34.2% of the divergent TEs (in *M. americana* and *P. hordei*, respectively), while TIR elements represent 24.2-38.0% of the divergent TEs (in *P. pachyrhizi* and *M. larici-populina*,

respectively). In the conserved TEs, LTR elements represent 0.1-34.4% (in *M. larici-populina* and *M. lini*, respectively), and TIR elements represent 7.6-33.3% (in *P. graminis f, sp, tritici Ug99 haplotype C* and *M. larici-populina*, respectively) (Fig.3A-I; Table S6). The heterogeneity in the composition of TEs with different relative conservation profiles (i.e., relative insertion ages) in the Pucciniales suggests that multiple waves of TE proliferation occurred since the early divergence between the Pucciniales and the other Pucciniomycotina species. The subset of highly conserved elements in all genomes also indicates recent TE transposition events.

We further employed a clustering approach to identify TE families shared among or specific to species within the Pucciniales order. We found no TE family shared across all the nine Pucciniales species. However, we identified three clusters that are specific to the Melampsoraceae and three that are specific to the Pucciniaceae (Fig. S2). These findings suggest that the accumulation of ancient TE copies predated the divergence of the family Melampsoraceae and the family Pucciniastraceae (58±17 million years ago, according to Aime and McTaggart 2021 [2]) on one hand and the divergence of the family Pucciniaceae and the family Sphaerophragmiaceae (approximately 68±17 million years ago, according to Aime and McTaggart 2021 [2]) on the other hand. Interestingly, the three Melampsoraceae-specific TE clusters were all TIR elements, whereas the Pucciniaceae-specific clusters comprised two LTR retrotransposon clusters and one TRIM cluster (Fig. S2; Tables S7 and S8). Additionally, we observed a predominance of species-specific TE clusters, ranging from 650 clusters (3.8% of the mobilome) in *P. graminis* f. sp. *tritici* 20-1 to 3,174 clusters (15.6% of the mobilome) in *P. triticina* (Fig. S2). Within these species-specific clusters, TIR elements accounted for 3.5% of the mobilome in *M. americana* (608 clusters) but only 0.05% (9 clusters) in *P. graminis* f. sp. *tritici* 21-0. Interestingly, we identified haplotype-specific TE clusters in *M. lini* exhibiting 744 clusters (4.3% of its mobilome), whereas *P. graminis* f. sp. *tritici* showed between 42 clusters in Ug99 and 45

clusters in 20-1 haplotypes (Fig. S2). These findings highlight the variation in TE composition and specificity across species and even between haplotypes within individuals.

Finally, we retraced the evolutionary history of LTR invasion using their terminal repeats with Kimura molecular clock estimates. Across all Pucciniales genomes, LTRs were found to have accumulated over a broad temporal range, from 618 Mya to the present (Fig. 3J-K). The oldest Gypsy elements were identified in *P. pachyrhizi*. 0.1% of LTRs accumulated before 176 Mya, thus predating the divergence between the Pucciniales and the other Pucciniomycotina. However, our analysis suggests that LTR accumulation within the Pucciniales genomes, more particularly, spiked in the last 50 Mya, with a notable acceleration reflecting LTR expansions occurring within the last 5 Mya (Fig. 3J-K). We also identified two distinct proliferation profiles in the Pucciniales genomes. The first profile, observed in *M. lini* and other Melampsoraceae species, exhibits sharp bursts of Gypsy LTR activity, with up to 2,000 elements accumulating during specific intervals (e.g., 7-9 Mya, 6-4 Mya, and 3-2 Mya) (Fig. 3J). The second profile, characteristic of Phakopsoraceae and Pucciniaceae species, demonstrates a more gradual accumulation of LTRs over time (Fig. 3K). These results collectively indicate that while genome inflation driven by TE invasion is a shared feature of Pucciniales, multiple waves of proliferation have shaped their evolutionary trajectories, reflecting diverse temporal dynamics of TE activity across lineages.

The TE invasion of the Pucciniales appears unrelated to deficiency in TE control pathways

We investigated whether the TE invasion in Pucciniales compared to the other Pucciniomycotina, might be due to a loss of control mechanisms. The most studied control mechanism against TE proliferation in fungi is the induction of cytosine-to-thymine mutations in repetitive regions through RIP mutations [25]. Using the composite RIP index, we quantified RIP-like signatures in TE family consensus sequences in the Pucciniales and the other Pucciniomycotina genomes. We estimated the average RIP composite index per TE

family ranging from -0.66 in *M. americana* to -0.15 in *L. creatinivora*, indicating the lack of RIP-like signatures in TE families among the Pucciniales and the other Pucciniomycotina TE family sequences (Table S9). Additionally, the two DNA methyltransferase genes associated with RIP (DIM2 and RID) are absent from both the Pucciniales and the other Pucciniomycotina genomes. We identified two other DNA methyltransferase genes DNMT5 and DNMT1 present across all Pucciniomycotina genomes, except for *P. hordei* in which we did not find DNMT1 (Fig. S3). Finally, we investigated the presence or absence of genes involved in the RNAi pathway, namely Argonaute, Dicer, and RNA-dependent RNA polymerase Rdmd. We identified 25-72 (*P. hordei* and *P. triticina*) genes related to Dicer-like genes in the Pucciniales and 25-29 Dicer-like in the other Pucciniomycotina genomes (Fig. S4). Similarly, we identified 17-60 (*P. hordeii* and *M. lini*) Rdmd-like genes in the Pucciniales and 12-15 Rdmd-like genes in the other Pucciniomycotina genomes (Fig. S4). We observed a remarkable expansion of AGO1 solely in the *P. pachyrhizi* genome that displays 572 copies compared to a mean of 91 copies in the other species (Fig. S4). However, it remains to determine whether these copies are truly functional. All Pucciniomycotina genomes, regardless of their TE content, are equipped with genes potentially involved in TE control through DNA methylation silencing and RNAi pathways. We conclude that the TE proliferation observed in the Pucciniales genomes compared to the other Pucciniomycotina is not attributable to the loss of genes involved in TE control mechanisms and related to RIP, DNA methylation, or the RNAi pathway.

Finally, we analyzed gene family dynamics within the Pucciniomycotina subphylum to investigate potential evidence of loss of genes that may relate to TE control mechanisms in Pucciniales outside of the canonical mechanisms investigated above. For this, we determined multigene families within the Pucciniomycotina using *Ustilago maydis* as an outgroup and identified 46,545 orthogroups (Table S10). Of these, 2,830 core orthogroups (49,535 genes) were shared across all genomes, including *U. maydis*, while 208 orthogroups (3,666 genes) were specific to Pucciniomycotina (Table S10). We observed substantial

variation in species-specific orthogroups ranging from 73 in *M. lychnidis-dioicae* to 1,276 in *P. hordei.* We identified 808 orthogroups in contraction and 15 orthogroups in expansion at the separation of the Pucciniales and the other Pucciniomycotina (Fig S5). The 808 orthogroups in contraction in the Pucciniales (i.e., in expansion in the other Pucciniomycotina) comprise a total of 4,523 genes, of which 39.7% (1,796 genes) are of unknown functions (Table S11; Fig. S6). The five most represented functions within the remaining orthogroups in contraction in the Pucciniales, include 333 genes related to "Carbohydrate transport and metabolism", 258 genes related to "Transcription", 233 genes related to "Chromatin structure and dynamics", 207 genes with function related to "Secondary metabolites biosynthesis, transport and catabolism" and 197 genes related to "Amino acid transport and metabolism" (Fig S6; Table S11). However, we identified only 15 orthogroups with a significant gene loss/gain between the Pucciniales and the other Pucciniomycotina genomes, respectively (p-value < 0.05; Fig S7A). Two of these orthogroups representing 58 genes, are significantly in contraction in the Pucciniales genomes, and none display functions associated with the "Chromatin structure and dynamics" category (Fig S7B). Based on the functional categories identified in this comparative analysis, it remains possible that certain gene losses may be associated with the loss of functions related to TE control among yet unidentified, functions that may belong to the "Chromatin structure and dynamics" category.

DISCUSSION

In this study, we provide a better understanding of how TEs have profoundly shaped the genomes of Pucciniales, contributing to their distinctive genome expansions within the Pucciniomycotina subphylum. Our comprehensive TE annotation across 13 Pucciniomycotina genomes demonstrates that the Pucciniales possess a mobilome landscape enriched in LTR Gypsy and TIR elements, a feature contrasting sharply with the smaller genomes and the minimal TE contribution in other Pucciniomycotina species. Our analyses examined four hypotheses on TE proliferation in Pucciniales and evidenced both

ancient and recent TE bursts, especially from Gypsy elements, aligning with the third hypothesis: TE proliferation is driven by a combination of ancestral transposition bursts retained in the genome and recent species-specific independent transposition events throughout Pucciniales evolution. Hence, we uncover ancestral TE proliferation likely predating the early divergent branch of the order Pucciniales, followed by further species-specific expansions. In addition, despite the extensive TE presence, our findings indicate no notable deficiency in known TE regulatory mechanisms in the Pucciniales genomes compared to the other Pucciniomycotina genomes, suggesting that Pucciniales may have rather evolved tolerance to TE proliferation or deployed other mechanisms to accommodate their abundant presence in the genome.

The accumulation of TEs in Pucciniales is an ancestral genomic trait. Pucciniales is a monophyletic group of species [5], all possessing inflated genomes enriched in TEs. We identified retrotransposon copies retained in Pucciniales genomes that predate this divergence, suggesting a long-term persistence of ancestral TEs within rust fungal genomes [5]. This TE retention pattern is consistent with observations in previous works of rust fungal genomes, which often harbor highly divergent, long-standing TE families [17]. Interestingly, we did not find any TE families shared among all nine *Pucciniales* species in our panel, suggesting that ancestral copies have diverged substantially over time. However, we detected TE families consistently retained at the genus level, including three (LTR-unknown, LTR-Copia, and TRIM) within the family *Pucciniaceae* and three others (TIR-Unknown, TIR-hat, and TIR-Tc1Mariner) within the family *Melampsoraceae*. The genus-specific conservation of certain TEs suggests that ancestral elements have been retained in distinct subgroups, which mirrors patterns observed in other fungal species. For example, a recent analysis of transposable element repertoires across 24 arbuscular mycorrhizal fungal genomes revealed the persistence of ancient TEs over long evolutionary periods [32]. Likewise, in large, complex plant genomes such as wheat and maize, retrotransposons are retained across extended timescales [33, 34]. The long-term retention of TEs across such

diverse fungal and plant lineages suggests a common pattern of TE evolutionary dynamics, underlying their potential role in shaping their host-genome architecture.

Gypsy retrotransposons represent the largest proportion of TE content across Pucciniales genomes, followed by *Copia* retrotransposons and TIR elements. This abundance of retrotransposons is a common feature in other TE-rich fungal genomes, such as those of *B. graminis* and *Entomophthoraceae* spp. [35, 36]. We hypothesize that ancient *Gypsy* elements found in the order Pucciniales may have served as progenitors for more recent lineage-specific TE expansions observed across the Pucciniales genomes. Indeed, the dynamics of *Gypsy* element accumulation display extreme variation among Pucciniales species. For example, we observed a burst accumulation pattern in *M. lini*, marked by a rapid increase in *Gypsy* copies over a few million years. In contrast, species such as *P. pachyrhizi* exhibit a more continuous pattern of *Gypsy* accumulations. We also detected haplotype-specific TE families in the phased genomes of the two species *M. lini* and *P. graminis* f. sp. *tritici*, indicating that TE dynamics vary not only between species but also within the two haplotypes in dikaryotic Pucciniales. These differences suggest that *Gypsy* elements may undergo distinct cycles of activity in each species, with some elements remaining active and driving ongoing TE expansion while others persist as inactive remnants within the genome. Notably, the presence of still active elements aligns with the observed expression of certain TEs during host infection in *P. pachyrhizi* [11], further supporting the hypothesis that active TEs may still be involved in ongoing genomic changes during pathogen-host interactions.

TE richness is a shared genomic feature among different groups of fungi, which appears as a trait associated with obligate parasitic and symbiotic lifestyles. Similarly to the Pucciniales, other obligate biotrophic plant pathogens, such as *B. graminis*, the causal agent of powdery mildew in cereals, are characterized by a high proportion of TEs [37]. Enrichment in TE could enable rapid evolution through genomic plasticity in response to fast-evolving host immune systems and changing environmental conditions [24]. Outside of plant pathogens,

obligate insect parasites within the fungal order Entomophthorales exhibit TE-rich genomes [36]. TEs are reported as the main drivers for genome expansion in this clade. TE-rich genomes extend beyond obligate parasites and represent a key feature of beneficial plant symbionts. Arbuscular mycorrhizal fungi and ectomycorrhizal fungi exhibit substantial TE content and expansions, although the role of these TEs in their symbiotic functions remains scarcely understood [32, 38]. While the evolutionary pressures differ between parasitism and symbiosis, large TE-rich genomes in both contexts may suggest that TEs could reflect an evolutionary strategy to facilitate genome plasticity and adaptation to the host environment.

We observed no evidence of specific gene loss associated with known TE regulation mechanisms in the Pucciniales genomes compared to other Pucciniomycotina. First, RIP-like signatures are absent in the Pucciniomycotina subphylum which aligns with other studies centered on basidiomycetes and with the absence of DIM2 and RID genes in basidiomycetes [27, 39]. In addition, we found the DNA methyltransferases DNMT1 and DNMT5 in all 13 genomes, suggesting that the whole Pucciniomycotina phylum may possess the potential for TE silencing mediated by DNA methylation. DNMT1 and DNMT5 could play a role in the establishment and/or maintenance of cytosine DNA methylation (5mC), as found in plants and animals [27]. In addition, Bewick et al. (2019) [27] demonstrated that species in the Phylum Basidiomycota displayed the highest 5mC levels across fungal genomes, with a weak correlation between 5mC levels and repeat content in the selected species. Similarly, we identified the essential genes of the RNAi pathway in all 13 Pucciniomycotina genomes. Additionally, we confirmed the expansion of Argonaute 1 (AGO1) previously reported in *P. pachyrhizi* [11]. Most of these expanded AGO1-like genes were described as pseudo-genes surrounded by TEs, suggesting a TE-driven inactivation of expanded AGO1 gene members in the family. This pattern could be similar to the recent loss of DIM2 activity found in the wheat fungal pathogen *Zymoseptoria tritici* where an expansion of the DIM2 genes, followed by inactivation, is thought to have led to TE-derepression [40]. As our comparative genomic approach could not identify a Pucciniales-specific loss of TE

control, other approaches, such as epigenomics or small RNA sequencing, will be required to better understand TE regulation mechanisms in rust fungi.

Despite advancements in TE annotation and the availability of high-quality rust genomes, the absence of a standardized and comprehensive TE library for rust fungi remains a significant obstacle in performing comparative genomics to address further evolutionary studies [41]. Manually curating TE libraries for highly repeated genomes is labor-intensive and time-consuming. Here, we generated a three-step approach to curate TE annotations, which can be applied to many other genomes. This curation has resolved all the conflictual annotations and reduced the unclassified TEs from 42.4% to 16% after three rounds of manual curation. We created a unified rust fungal TE library, enabling consistent and precise TE identification across different rust fungal species. Still, several TE annotations remain unresolved after our curation, mainly consisting of elements probably too degraded to be correctly classified and nested elements. Additional manual curation efforts, supported by additional genome assemblies from closely related species, could support the refinement of the classification of these elements by offering more diagnostic features. As large-scale genomic projects generate vast sequencing data, creating a unified TE library becomes increasingly crucial for advancing our understanding of the roles of TEs in rust fungal biology and their evolution. Expanding and refining fungal TE libraries is essential to fully capture the diversity and dynamics of transposable elements in these genomes, enabling more precise evolutionary and functional insights in future studies.

CONCLUSION

This study describes manually curated TE catalogs in thirteen Pucciniomycotina species and provides a new fungal TE database. We report the genome expansions in Pucciniales are mainly guided by TE proliferation. Our work highlights recent bursts of Gypsy LTR in Pucciniales while the content in such elements remains relatively poor in other Pucciniomycotina. We also describe a few conserved TEs within the Pucciniales order linked

to older proliferation events. TE accumulations do not seem to result from Pucciniales-specific loss of TE control mechanisms. We provided new insights into Pucciniales genome architecture and favored the hypothesis that TE proliferation is an ancestral characteristic of the Pucciniales evolution.

METHODS

Genomic data

We selected the genomes and gene annotations of 13 Pucciniomycotina species (14 genomes) publicly available in the JGI MycoCosm database and NCBI repositories (Table S1). We included ten genomes from the order Pucciniales, comprising four genomes from the family Melampsoraceae (*M. allii-populina*, *M. lini*, *M. larici-populina*, and *M. americana*), five genomes from the family Pucciniaceae (*P. graminis* f. sp. *tritici Ug99 and 21-0*, *P. hordei*, *P. triticina*, and *P. striiformis* f. sp. *tritici*), and one genome from the family Phakopsoraceae (*P. pachyrhizi*). Additionally, four genomes representing species from outside the order Pucciniales but within the subphylum Pucciniomycotina were included (*L. creatinivora*, *M. lychnidis-dioicae*, *M. intermedium*, and *R. toruloides*; Table S1). We assessed genome assembly completeness using BUSCO (Benchmarking Universal Single-Copy Orthologs) to ensure data quality The genome completeness was analyzed with BUSCO v5.4.7 using the basidiomycota_odb10 marker set. These metrics provided a robust measure of the quality and completeness of the selected genome assemblies (Table S1).

TE detection and annotation

We detected and annotated repeats in Pucciniomycotina genomes using REPET v3.0 (https://urgi.versailles.inra.fr/Tools/REPET, [33]). Repeated elements were identified and annotated for each genome following a two-step process. First, TE families were identified using the TEdenovo pipeline, which builds TE family sequences based on repetitive

elements detected in a 100 Mb subset of the genome. This subset-based approach was applied to enhance workflow efficiency in highly repeated genomes, as described in [34]. Subsequently, the whole genomes were annotated with TEannot using the TE families generated by TEdenovo. A second round of TEannot was performed as a curation step, focusing exclusively on TE families with at least one full-length copy. For haplotype-phased genome assemblies, TEdenovo was run separately for each haplotype (also using 100 Mb subsets), and the resulting TE family sequences were merged before proceeding with the TEannot steps. This approach allowed for the identification of potential haplotype-specific TEs.

Curation of TE annotations

After the automated detection and annotation of TEs, we performed additional steps of curations of libraries to improve the classifications of elements detected as “conflicts” and “unclassified” TEs by REPETv3.0. First, the conflictual TE annotations were manually curated and reassigned to the appropriate TE Class or Order based on sequence similarity, applying a 50% similarity threshold. In parallel, we used the Manual Curator Helper tool MCHelper [31] to address redundancy, fragmentation, false positives, and unclassified sequences in TE libraries for each genome. Briefly, MCHelper extends consensus sequences through iterative rounds of Blast, extract, and extend (BEE) and refines them into subfamilies based on distance clustering. False positives are filtered out by removing sequences with low-quality fragments or homology to multi-copy genes, and the remaining sequences are classified through homology-based methods. We then pooled the 14 TE libraries into a Pucciniomycotina TE library to perform an additional manual curation step by clustering the Pucciniomycotina TE library with CD-HIT v4.8.1 [42]. We grouped TE sequences with 80% similarity across 80% of the sequence length and re-assigned previously unclassified TEs based on annotated TEs of the same cluster. The workflow to

generate curated TE annotations from multiple genomes combining automated detection and curation results is available at https://github.com/Raistrawby/TE_Rust/tree/main.

TE relative insertion age

The relative age of TE insertions in the Pucciniomycotina genomes was estimated with the similarity-based approach of REPETv3.0, as described in Lorrain et al. (2021) [43]. Briefly, we analyzed sequence divergence between TE copies to their cognate family using the extent of sequence similarity as a proxy to the divergence time of the TE copies within the host genome. The TE copies were classified into three categories: i) conserved, with copies similar up to 95% against their family, ii) intermediate, with a similarity between 85% and 95% and, iii) divergent, with similarity of copies below 85%.

Insertion age of LTR-retrotransposons

The insertion ages of LTR-retrotransposons were assessed using a molecular clock approach, as described in Gupta et al. (2023) [11]. Full-length LTR-retrotransposons were detected using LTRHarvest v 1.6.1 [44] with default parameters (except for -similar 90 (default 85) and -mintsd 5 (default 4)). Gypsy and Copia LTR-retrotransposons were selected based on the annotations from the repbase_23.12 database. Each element's 3' and 5' long-terminal repeats were extracted and aligned using MAFFT v7.471 [45] to calculate Kimura's P distances [46]. Gypsy and Copia insertion dates (T) were estimated using the formula T = K / 2r, where K is the distance between the 2 LTRs and is calculated with the Kimura 2-Parameter distance formula -0.5*log(1 - 2p - q) * sqrt (1 - 2 q)) with p = transition frequency and q transversion frequency and r is the fungal substitution rate of 1.05 × 10-9 nucleotides per site per year [46].

Gene family annotations and evolution

We predicted orthologs between the 13 Pucciniomycotina genomes and the reference *U. maydis* genome [23], a representative of the Ustilaginomycotina (the closest subphylum

to Pucciniomycotina within Basidiomycete), which we used as an outgroup to identify genes restricted to the Pucciniomycotina subphylum. Orthogroup analysis was performed using the software OrthoFinder v2.4.0 (i=2 -M MSA -S diamond -a mafft)[47] on the 14 proteomes. To assess the gene family gains and losses throughout the Pucciniomycotina genomes, we first needed to build a phylogenetic tree; to this end, we used the single copy orthogroups (i.e. orthogroups containing only one protein of each species) and infer the tree with *U. maydis* as an outgroup with Raxml and 500 bootstraps [48]. We used this rooted phylogenetic tree and the number of genes found per family (i.e. orthogroup) to estimate rates of gene families gains (i.e., families in expansion) and losses (i.e. families in contraction) at the species-, genus-, family- and order-levels with CAFE v5.0 [49].

Functional gene annotations and identification of DNA methyltransferases and RNA-induced silencing complex

We used EggNOG-mapper v2.1.12 [50] to predict functional annotations for all Pucciniomycotina genomes. We retrieved the PFAM, KOG, KEGG, and GO annotations from EggNOG-mapper annotations to explore the potential functions of gene families gained or lost throughout the evolution of the Pucciniomycotina. To identify the DNMTs and the genes involved in the RISC pathway in the 14 Pucciniomycotina genomes, we extracted the genes with corresponding PFAM domains. For DNA methyltransferases, we selected genes with PF00145 (C-5 cytosine-specific DNA methylase). For the RNA-induced silencing complex, we selected genes with PF05183 (RNA dependent RNA polymerase), PF16486 (N-terminal domain of argonaute), PF22389 (Protein Dicer, dsRNA-binding domain), PF09692 (Argonaute siRNA chaperone (ARC) complex subunit Arb1), PF22749 (Arb2 domain), PF00636 (Ribonuclease III domain), PF04992 (RNA polymerase Rpb1, domain6), PF03871 (RNA polymerase Rpb5, N-terminal domain), PF01191 (RNA polymerase Rpb5, C-terminal domain). We searched for characterized DNMT and RISC proteins to refine the annotations and performed BLASTp searches on the putative Pucciniomycotina DNMT and RISC proteins.

RIP signatures analysis

RIP-like signatures were searched by estimating the RIP composite index [25] and GC content in TE family sequences of all Pucciniomycotina genomes, as described in Feurtey et al. (2023)[51]. RIP Composite index values were calculated using the formula: (TpA/ApT) – (CpA + TpG/ApC + GpT) applied per TE family sequence. A region is considered affected by RIP if the Composite index exceeds 0 [39].

Data availability

All the data supporting the findings of this study, including TE annotations and gene functional predictions, are openly available at 10.5281/zenodo.14616359. The curated library of TE consensus sequences has been deposited in the Dfam database (https://dfam.org/home).

Code availability

The TE annotation and manual curation pipeline, as well as additional scripts used to create the results presented in this manuscript, are available in the following GitHub repository: https://github.com/Raistrawby/TE_Rust/tree/main.

Authors contributions

EC and CL wrote the manuscript. SD and EM revised the manuscript. EC, CL, and SD designed, analyzed, and interpreted the data. EM provided comparative genomics expertise and bioinformatics technical support. CL and SD contributed to the conception of the work.

Acknowledgments

Jana Sperschneider and Peter Dodds are acknowledged for providing early access to the phased assembly and genome annotation of the *M. lini* reference genome. EC was supported by a PhD grant from the French National Institute for Agriculture, Food and Environment (INRAE). EC, EM and SD were supported by the French Plan Investissement

d'Avenir (PIA) Lab of Excellence ARBRE [ANR-11-LABX-0002- 01]. CL was funded by an SNSF Amizione grant (PZ00P3_209022).

AI language model assistance

ChatGPT (developed by OpenAI) was used to refine and improve the clarity and grammar of the text. The authors reviewed and revised all outputs to ensure accuracy and alignment with the intended message.

ETHICS DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No competing interests.

REFERENCES

1. Duplessis S, Lorrain C, Petre B, Figueroa M, Dodds PN, Aime MC. Host Adaptation and Virulence in Heteroecious Rust Fungi. Annu Rev Phytopathol. 2021;59:403–22.

2. Aime MC, McTaggart AR. A higher-rank classification for rust fungi, with notes on genera. Fungal Syst Evol. 2021;7:21–47.

3. Carnegie AJ, Pegg GS. Lessons from the Incursion of Myrtle Rust in Australia. Annu Rev Phytopathol. 2018;56 Volume 56, 2018:457–78.

4. Carefoot, Sprott GL ER. Famine on the wind; Man's Battle Against Plant Disease. 1967.

5. Aime MC, Bell CD, Wilson AW. Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. Stud Mycol. 2018;89:143–52.

6. Wood AR, Aime MC. Phylogenetics of the rust fungi (Pucciniales) of South Africa, with notes on their life histories and possible origins. Mycologia. 2024;0:1–27.

7. Aime MC, McTaggart AR, Mondo SJ, Duplessis S. Chapter Seven - Phylogenetics and Phylogenomics of Rust Fungi. In: Townsend JP, Wang Z, editors. Advances in Genetics. Academic Press; 2017. p. 267–307.

8. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A. 2011;108:9166–71.

9. Presti LL, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol. 2015;66 Volume 66, 2015:513–45.

10. Guerillot P, Salamov A, Louet C, Morin E, Frey P, Grigoriev IV, et al. A Remarkable Expansion of Oligopeptide Transporter Genes in Rust Fungi (Pucciniales) Suggests a Specialization in Nutrient Acquisition for Obligate Biotrophy. Phytopathology®. 2023;113:252–64.

11. Gupta YK, Marcelino-Guimarães FC, Lorrain C, Farmer A, Haridas S, Ferreira EGC, et al. The soybean rust pathogen Phakopsora pachyrhizi displays transposable element proliferation that correlates with broad host-range adaptation on legumes. 2022;:2022.06.13.495685.

12. Lorrain C, Gonçalves Dos Santos KC, Germain H, Hecker A, Duplessis S. Advances in understanding obligate biotrophy in rust fungi. New Phytol. 2019;222:1190–206.

13. Ramos AP, Tavares S, Tavares D, Silva MDC, Loureiro J, Talhinhas P. Flow cytometry reveals that the rust fungus, romyces bidentis (Pucciniales), possesses the largest fungal genome reported—2489 Mbp. Mol Plant Pathol. 2015;16:1006–10.

14. Mohanta TK, Bae H. The diversity of fungal genome. Biol Proced Online. 2015;17:8.

15. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012;10:417–30.

16. Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, et al. A chromosome-level, fully phased genome assembly of the oat crown rust fungus Puccinia

coronata f. sp. avenae: a resource to enable comparative genomics in the cereal rusts. G3 GenesGenomesGenetics. 2022;12:jkac149.

17. Tobias PA, Schwessinger B, Deng CH, Wu C, Dong C, Sperschneider J, et al. Austropuccinia psidii, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 GenesGenomesGenetics. 2021;11:jkaa015.

18. Liang J, Li Y, Dodds PN, Figueroa M, Sperschneider J, Han S, et al. Haplotype-phased and chromosome-level genome assembly of Puccinia polysora, a giga-scale fungal pathogen causing southern corn rust. Mol Ecol Resour. n/a n/a.

19. 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;8:973–82.

20. Dubin MJ, Mittelsten Scheid O, Becker C. Transposons: a blessing curse. Curr Opin Plant Biol. 2018;42:23–9.

21. 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. 2020;37:221–39.

22. 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.

23. Amselem J, Vigouroux M, Oberhaensli S, Brown JKM, Bindschedler LV, Skamnioti P, et al. Evolution of the EKA family of powdery mildew avirulence-effector genes from the ORF 1 of a LINE retrotransposon. BMC Genomics. 2015;16:917.

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

25. Gladyshev E. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. Microbiol Spectr. 2017;5.

26. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19:199.

27. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev IV, James TY, et al. Diversity of cytosine methylation across the fungal tree of life. Nat Ecol Evol. 2019;3:479–90.

28. van Wyk S, Wingfield BD, De Vos L, van der Merwe NA, Steenkamp ET. Genome-Wide Analyses of Repeat-Induced Point Mutations in the Ascomycota. Front Microbiol. 2021;11:622368.

29. Nicolás FE, Garre V. RNA Interference in Fungi: Retention and Loss. Microbiol Spectr. 2016;4:10.1128/microbiolspec.funk-0008–2016.

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

31. Orozco-Arias S, Sierra P, Durbin R, González J. MCHelper automatically curates transposable element libraries across eukaryotic species. 2024;:2023.10.17.562682.

32. Oliveira JIN, Lane C, Mugambi K, Yildirir G, Nicol AM, Kokkoris V, et al. Analyses of transposable elements in arbuscular mycorrhizal fungi support evolutionary parallels with filamentous plant pathogens. 2024;:2024.11.04.621924.

33. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, et al. Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biol. 2018;19:103.

34. Mirouze M, Vitte C. Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. J Exp Bot. 2014;65:2801–12.

35. Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, et al. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics. 2018;19:381.

36. Stajich JE, Lovett B, Lee E, Macias AM, Hajek AE, de Bivort BL, et al. Signatures of transposon-mediated genome inflation, host specialization, and photoentrainment in Entomophthora muscae and allied entomophthoralean fungi. eLife. 12:RP92863.

37. Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, et al. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. Nat Genet. 2013;45:1092–6.

38. Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, et al. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat Commun. 2020;11:5125.

39. van Wyk S, Harrison CH, Wingfield BD, De Vos L, van der Merwe NA, Steenkamp ET. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ. 2019;7:e7447.

40. Möller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, et al. Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. PLOS Genet. 2021;17:e1009448.

41. Baril T, Croll D. A pangenome-guided manually curated library of transposable elements for Zymoseptoria tritici. BMC Res Notes. 2023;16:335.

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

43. Lorrain C, Oggenfuss U, Croll D, Duplessis S, Stukenbrock E. Transposable Elements in Fungi: Coevolution With the Host Genome Shapes, Genome Architecture, Plasticity and Adaptation. In: Zaragoza Ó, Casadevall A, editors. Encyclopedia of Mycology. Oxford: Elsevier; 2021. p. 142–55.

44. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 2008;9:18.

45. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47:W5–10.

46. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16:111–20.

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

48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

49. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006;22:1269–71.

50. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38:5825–9.

51. Feurtey A, Lorrain C, McDonald MC, Milgate A, Solomon PS, Warren R, et al. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. Nat Commun. 2023;14:1059.

FIGURES LEGENDS

Fig. 1: **The phylogenetic tree and transposable element content in the Pucciniomycotina.** A) phylogenetic tree of the 13 selected Pucciniomycotina species based on a single-copy orthogroup from Orthofinder v2.4.0 and visualized using iTOL (https://itol.embl.de/). The Pucciniales are represented in yellow. The category *Other Pucciniomycotina* (blue) refers to the four species of Pucciniomycotina outside the order Pucciniales. B) The bar plot displays the TE content across the analyzed genomes. Dark blue bars represent the proportion of the genome covered by TEs, while light blue bars show the total genome size (in Mb). Percentages on the right indicate the TE coverage relative to the total genome size. Species are described as follow: Melap (*Melampsora allii-populina*), Melli (*Melampsora lini*), Mellp (*Melampsora larici-populina*), Melam (*Melampsora americana*), Phapa (*Phakopsora pachyrhizi*), Pucst (*Puccinia striiformis* f. sp. *tritici*), Puctr (*Puccinia triticina*), Pucho (*Puccinia hordeii*), PucgtrUg99 (*Puccinia graminis* f. sp. *tritici*), Micld (*Microbotryum lychnidis-dioicae*), Micin (*Microbotryum intermedium*), Leucr (*Leucosporidia creatinivora*), Rhoto (*Rhodosporidium toruloides*).

Fig. 2: **Composition of transposable elements in Pucciniomycotina genomes.** The stacked bar plots represent the percentage of transposable elements (TEs) annotated at the A) Class level, B) Class I order level, and C) Class II order level for each Pucciniomycotina genome. The n values indicate the number of distinct TE families identified in each genome. The percentage of TE genome coverage at the super-family (SF) level for E) Class I elements (purple) and F) Class II elements (blue).

Fig. 3: **Transposable element insertion ages.** TE families sequence identity distribution between TE copies and the respective consensus sequences per genome: A) Mellp (*M. larici-populina*), B) Melap (*M. allii-populina*), C) Melam (*M. americana*), D) MelliH (*M. lini* haplotype H), E) Phapa (*P. pachyrhizi*), F) PucstA (*P. striiformis* f. sp. *tritici* haplotype A), G)

PuctrA (*P. triticina* haplotype A), H) PucgtrUg99A (*P. graminis* f. sp. *tritici* Ug99 haplotype A), I) Pucho (*P. hordeii*). The vertical dashed lines represent the 85% and 95% mean sequence identity percentage thresholds. J-K) Number of LTR-Gypsy elements accumulated through time (million years ago - Mya), based on insertion age for the J) Melampsoraceae species and I) the Pucciniaceae species.

Fig. S1: **Improvement of TE annotations through manual curation** . The proportion of annotated TE families after REPET detection and annotation (RawR), after the first round of curation (Postpct), after the second round of curation with MCHelper (postMC), after a third round of curation with Pucciniomycotina-guided clustering (PostCDhit).

Fig. S2: **Shared and specific TE families between the Pucciniomycotina species.** Numbers of TE families shared among the 14 genomes based on CDHIT clustering. The bars on the left represent the total number of genes per species. The clusters shared among Pucciniaceae are colored in blue and the clusters shared among Melampsoraceae are colored in red.

Fig. S3: **TE controls involved in methylations are detected in all species**. The number of genes involved in TE control corresponds to DNMT1 and DNMT5.

Fig. S4: **TE controls involved in RNA-induced silencing complex are detected for all genomes.** The number of genes involved in TE control corresponds to the RNAi pathway: Dicer-like (DCL; DCR), RNA-dependent RNA polymerase (RDP; RNT; RPAB; RPB), Argonaute (AGO; ARB) for each genome.

Fig. S5: **Contraction and Expansion of multigene families in Pucciniomycotina.**
The phylogenetic tree of the selected Pucciniomycotina species based on single-copy orthogroups from Orthofinder v2.4.0 and infer with Raxml (500 bootstraps) and visualized

using iTOL, rooted with *Ustilago maydis.* Numbers in purple correspond to the number of orthogroups in contractions at each branch or node and yellow at expansions.

Fig. S6: **Contraction of Multigene Families in Pucciniales.** Functional annotation of multigene families in contraction in Pucciniales. Genes are re-annotated by EggNOG-mapper v2.1.12. A threshold is set to kept only species with at least 20 proteins per COG category. COG categories correspond to: "-": no hit; B: chromatin structure and dynamics; C: energy production and conversion; E: amino acid transport and metabolism; G: carbohydrate transport and metabolism; I: lipid transport and metabolism; K: transcription; O: posttranslational modification, protein turnover, chaperones, P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; S: function unknown; T: signal transduction mechanisms.

Fig. S7: **Significant contraction and expansion of multigene families in Pucciniomycotina.** A) The phylogenetic tree of the selected Pucciniomycotina species based on a single-copy orthogroup from Orthofinder v2.4.0 and visualized using iTOL, rooted with *Ustilago maydis.* Numbers in black correspond to the number of significant orthogroups in expansion/contraction at each branch or node. B) COG category annotation of genes involved in expansion and contraction between Pucciniales and other Pucciniomycotina. Contraction and expansion are detected with CAFE v5.0 and orthogroups with a *p*-value (0.05) are kept. Genes are re-annotated by EggNOG-mapper v2.1.12. The number of genes per species and categories. COG categories functions correspond to : A: RNA processing and modification; C: energy production and conversion; D: cell cycle control, cell division chromosome partitioning; G: carbohydrate transport and metabolism; I: lipid transport and metabolism; K: transcription; L: replication recombination and repair; M: cell wall, membrane, envelope biogenesis; N: cell motility; O: posttranslational modification, protein turnover, chaperones, Q: secondary metabolites biosynthesis, transport and catabolism; S: function unknown; T: signal transduction mechanisms.

Table S1: **General information of selected genomes**.

Table S2: **TE annotations at the Class level for the 14 Pucciniomycotina genomes**. ClassAnn represents the different TE Classes (I, II, Unclassified); nbTE represents the number of TE families in each Class; ncovP represents the percentage of genome coverage of each TE class in the species genome; nbcopie represents the number of copies associated with the TE family for each Class; Sumcov represents the percentage of coverage by all the TEs per genome.

Table S3: **TE annotations at the Order level for the 14 Pucciniomycotina genomes**. OrderAnn represents the different TE Orders; nbTE represents the number of TE families in each Order; ncovP represents the percentage of genome coverage of each TE Order in the species genome; nbcopie represents the number of copies associated with the TE family for each Order; Sumcov represents the percentage of coverage by all the TEs per genome.

Table S4: **TE annotations at the Super-family level for the 14 Pucciniomycotina genomes**. OrderAnn represents the different TE SF; nbTE represents the number of TE families in each SF; ncovP represents the percentage of genome coverage of each TE SF in the species genome; nbcopie represents the number of copies associated with the TE family for each SF; Sommecov represents the percentage of coverage by all the TEs per species.

Table S5: **Average of the TE sequence identity at the Class level**. The meanID represents the mean sequence identity (%) of individual copies to their respective TE family consensus sequence averaged per TE class (ClassAnn).

Table S6: **Average of the TE sequence identity at the Order level**. The meanID represents the mean sequence identity (%) of individual copies to their respective TE family consensus sequence averaged per TE Order (OrderAnn).

Table S7: **Shared TEs within the three Melampsoraceae genomes**.

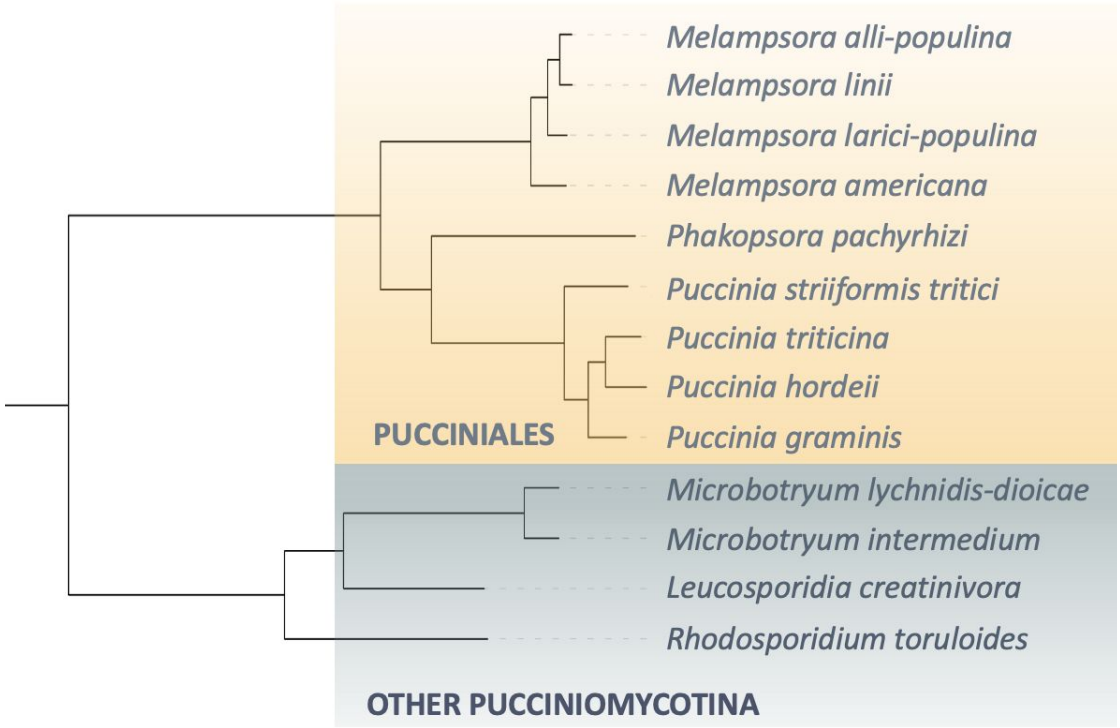
Table S8: **Shared TEs within the five Pucciniaceae genomes**.

Table S9: **RIP signature analysis**. RIP-like signatures were searched by estimating the RIP composite index in TE families consensus sequences for each Pucciniomycotina genome described in Feurtey et al. (2023)[51].

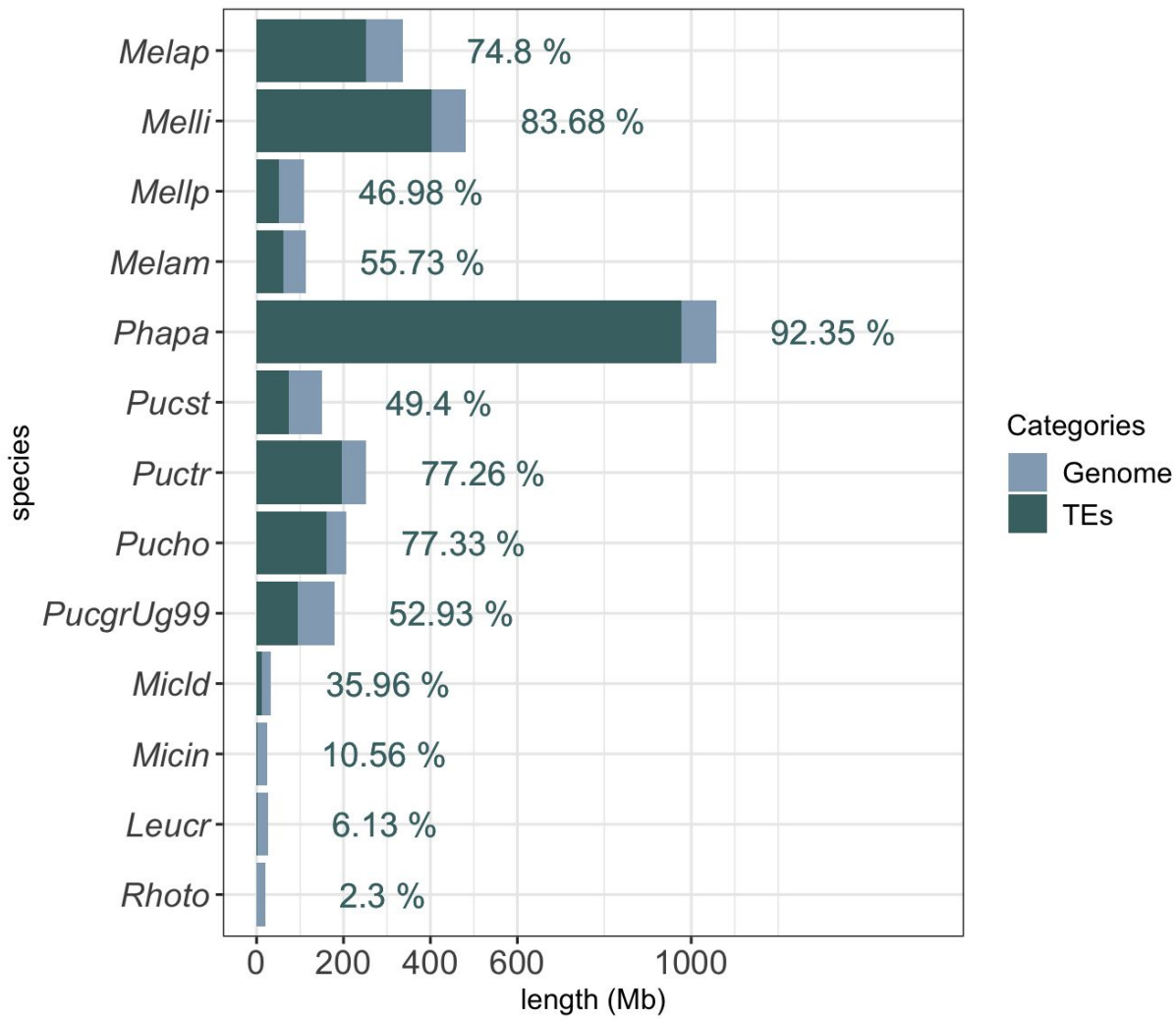
Table S10: **Orthogroups summary metrics of the 14 selected Pucciniomycotina genomes.**

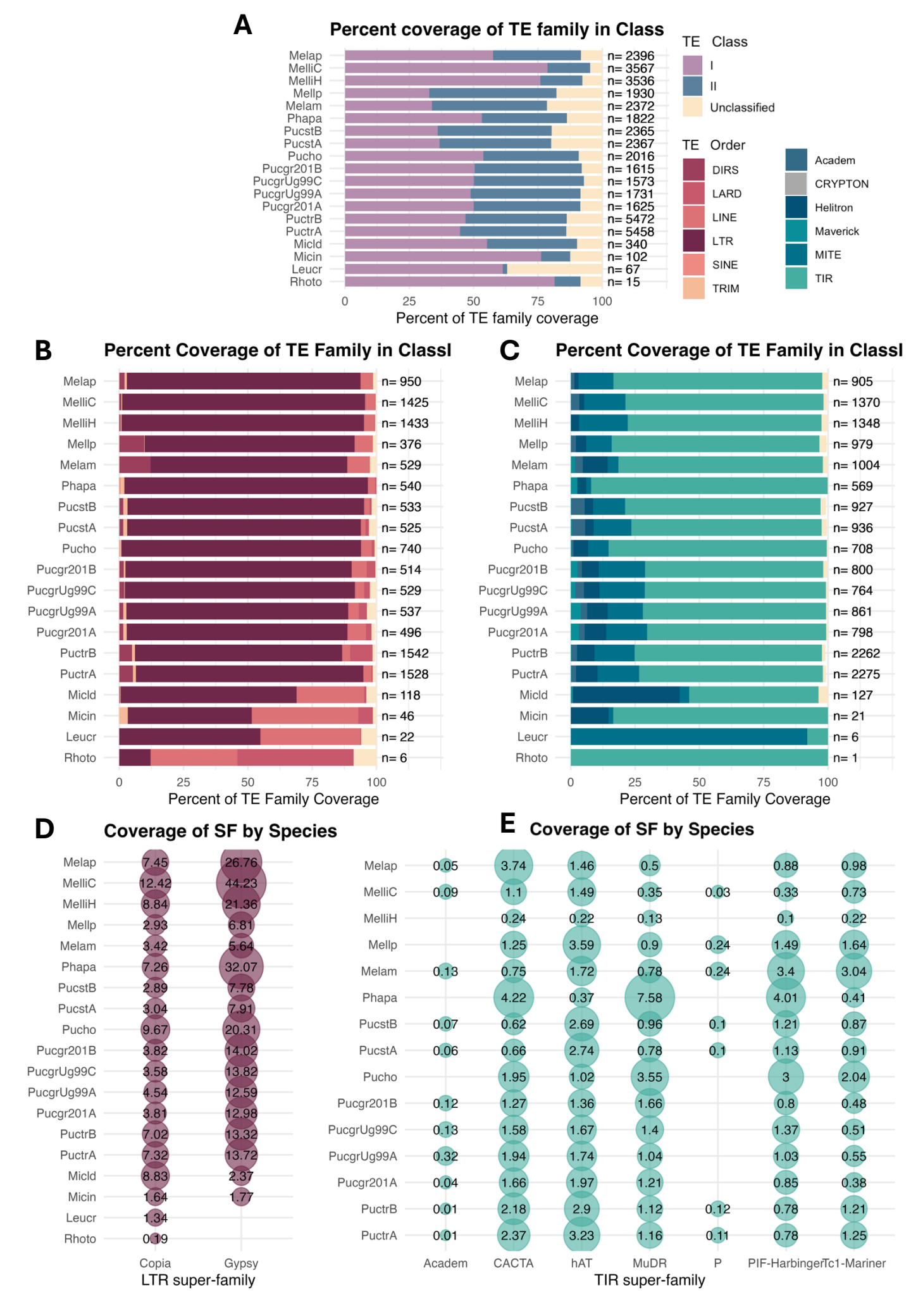
Table S11: **Orthogroups in contraction of the 14 selected Pucciniomycotina genomes.**

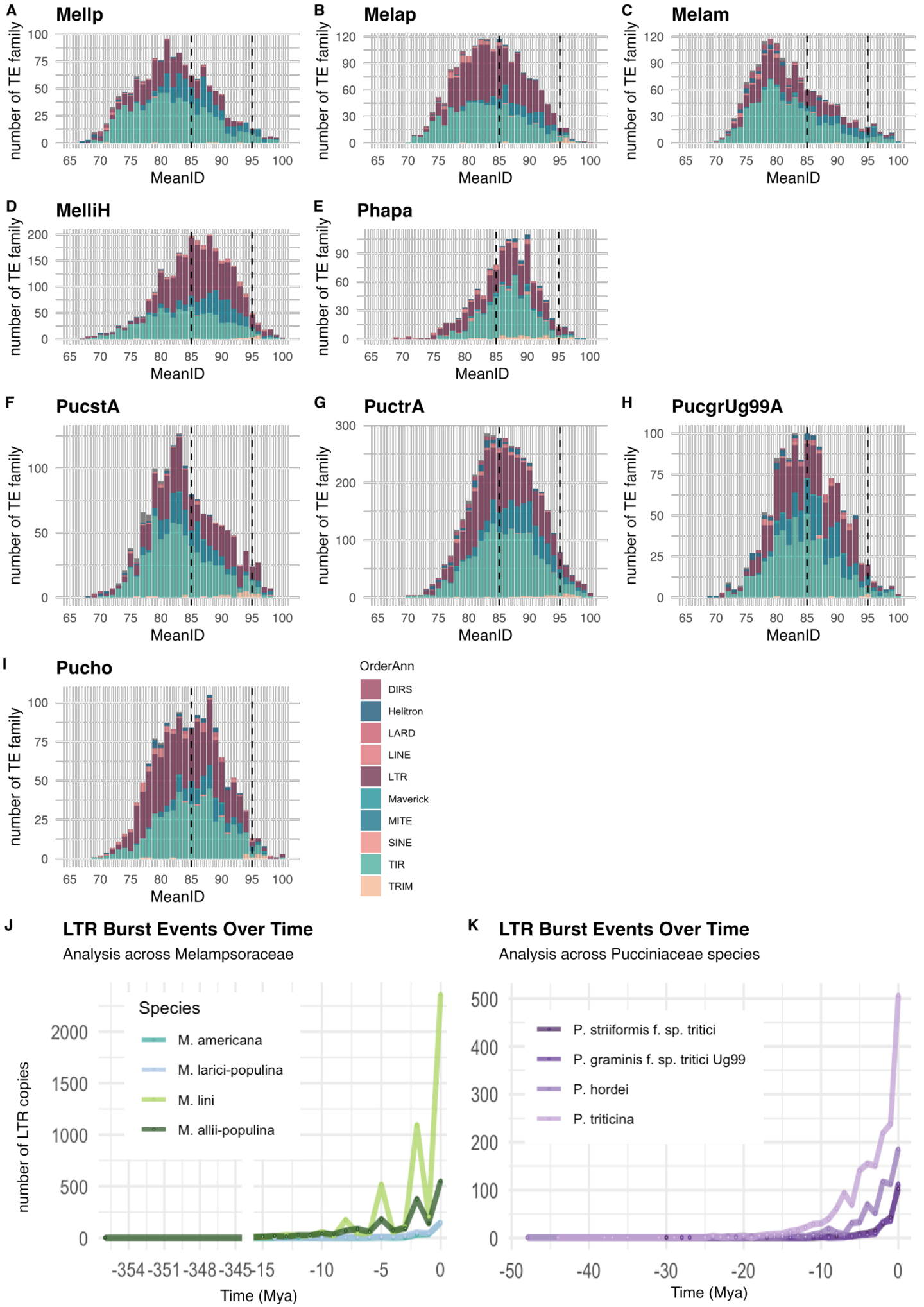
A


Melampsora alli-populina
Melampsora linii
Melampsora larici-populina
Melampsora americana
Phakopsora pachyrhizi
Puccinia striiformis tritici
Puccinia triticina
Puccinia hordeii
Puccinia graminis
PUCCINIALES
Microbotryum lychnidis-dioicae
Microbotryum intermedium
Leucosporidia creatinivora
Rhodosporidium toruloides
OTHER PUCCINIOMYCOTINA

B


Melap
Melli
Mellp
Melam
Phapa
Pucst
Puctr
Pucho
PucgrUg99
Micld
Micin
Leucr
Rhoto
species
74.8 %
83.68 %
46.98 %
55.73 %
92.35 %
49.4 %
77.26 %
77.33 %
52.93 %
35.96 %
10.56 %
6.13 %
2.3 %
0
200
400
600
1000
length (Mb)
Categories
Genome
TEs


A
Percent coverage of TE family in Class
Melap n= 2396
MelliC n= 3567
MelliH n= 3536
Mellp n= 1930
Melam n= 2372
Phapa n= 1822
PucstB n= 2365
PucstA n= 2367
Pucho n= 2016
Pucgr201B n= 1615
PucgrUg99C n= 1573
PucgrUg99A n= 1731
Pucgr201A n= 1625
PuctrB n= 5472
PuctrA n= 5458
Micld n= 340
Micin n= 102
Leucr n= 67
Rhoto n= 15
0 25 50 75 100
Percent of TE family coverage
TE Class
I
II
Unclassified
TE Order
DIRS
LARD
LINE
LTR
SINE
TRIM
Academ
CRYPTON
Helitron
Maverick
MITE
TIR
B
Percent Coverage of TE Family in ClassI
Melap n= 950
MelliC n= 1425
MelliH n= 1433
Mellp n= 376
Melam n= 529
Phapa n= 540
PucstB n= 533
PucstA n= 525
Pucho n= 740
Pucgr201B n= 514
PucgrUg99C n= 529
PucgrUg99A n= 537
Pucgr201A n= 496
PuctrB n= 1542
PuctrA n= 1528
Micld n= 118
Micin n= 46
Leucr n= 22
Rhoto n= 6
0 25 50 75 100
Percent of TE Family Coverage
C
Percent Coverage of TE Family in ClassI
Melap n= 905
MelliC n= 1370
MelliH n= 1348
Mellp n= 979
Melam n= 1004
Phapa n= 569
PucstB n= 927
PucstA n= 936
Pucho n= 708
Pucgr201B n= 800
PucgrUg99C n= 764
PucgrUg99A n= 861
Pucgr201A n= 798
PuctrB n= 2262
PuctrA n= 2275
Micld n= 127
Micin n= 21
Leucr n= 6
Rhoto n= 1
0 25 50 75 100
Percent of TE Family Coverage
D
Coverage of SF by Species
Melap 7.45 26.76
MelliC 12.42 44.23
MelliH 8.84 21.36
Mellp 2.93 6.81
Melam 3.42 5.64
Phapa 7.26 32.07
PucstB 2.89 7.78
PucstA 3.04 7.91
Pucho 9.67 20.31
Pucgr201B 3.82 14.02
PucgrUg99C 3.58 13.82
PucgrUg99A 4.54 12.59
Pucgr201A 3.81 12.98
PuctrB 7.02 13.32
PuctrA 7.32 13.72
Micld 8.83 2.37
Micin 1.64 1.77
Leucr 1.34
Rhoto 0.19
Copia Gypsy
LTR super-family
E
Coverage of SF by Species
Melap 0.05 3.74 1.46 0.5 0.88 0.98
MelliC 0.09 1.1 1.49 0.35 0.03 0.33 0.73
MelliH 0.24 0.22 0.13 0.1 0.22
Mellp 1.25 3.59 0.9 0.24 1.49 1.64
Melam 0.13 0.75 1.72 0.78 0.24 3.4 3.04
Phapa 4.22 0.37 7.58 4.01 0.41
PucstB 0.07 0.62 2.69 0.96 0.1 1.21 0.87
PucstA 0.06 0.66 2.74 0.78 0.1 1.13 0.91
Pucho 1.95 1.02 3.55 3 2.04
Pucgr201B 0.12 1.27 1.36 1.66 0.8 0.48
PucgrUg99C 0.13 1.58 1.67 1.4 1.37 0.51
PucgrUg99A 0.32 1.94 1.74 1.04 1.03 0.55
Pucgr201A 0.04 1.66 1.97 1.21 0.85 0.38
PuctrB 0.01 2.18 2.9 1.12 0.12 0.78 1.21
PuctrA 0.01 2.37 3.23 1.16 0.11 0.78 1.25
Academ CACTA hAT MuDR P PIF-Harbinger Tc1-Mariner
TIR super-family


A
MelIp
B
Melap
C
Melam
D
MelliH
E
Phapa
F
PucstA
G
PuctrA
H
PucgrUg99A
I
Pucho
number of TE family
MeanID
OrderAnn
DIRS
Helitron
LARD
LINE
LTR
Maverick
MITE
SINE
TIR
TRIM
J
LTR Burst Events Over Time
Analysis across Melampsoraceae
Species
M. americana
M. larici-populina
M. lini
M. allii-populina
number of LTR copies
Time (Mya)
K
LTR Burst Events Over Time
Analysis across Pucciniaceae species
P. striiformis f. sp. tritici
P. graminis f. sp. tritici Ug99
P. hordei
P. triticina
Time (Mya)