UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

Université de Lorraine, Collégium Sciences et Techniques
Ecole Doctorale Sciences et Ingénierie des Ressources Naturelles
UMR 1136 INRAE - Université de Lorraine, Interactions Arbres – Microorganismes

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention : Biologie et écologie des forêts et des agrosystèmes

par **EMMA CORRE**

Sous la direction de S. DUPLESSIS et C. LORRAIN

Invasion of Pucciniales genomes by transposable elements: How transposition events shape the evolution and the genomic architecture in rust fungi

Soutenue le 31 Octobre 2025

Membres du jury :

Présidente :	**Mme Mélanie MOREL-ROUHIER**	**Professeure, Université de Lorraine, Vandœuvre-lès-Nancy (France)**
Rapporteurs :	**Mr Daniel CROLL**	**Professeur, Université de Neuchâtel, Neuchâtel (Suisse)**
	Mme Gaëlle LELANDAIS	**Professeure, Université Paris-Saclay, Paris (France)**
Examinateurs :	**Mme Emmanuelle LERAT**	**Chargée de recherche, CNRS, Villeurbanne (France)**
	Mr Etienne DANCHIN	**Directeur de recherche, INRAE, Sophia Antipolis (France)**
Directeur(s) de thèse :	**Mr Sébastien DUPLESSIS**	**Directeur de recherche, Université de Lorraine / INRAE, Champenoux (France)**
	Mme Cécile LORRAIN	**Senior scientist, ETH, Zürich (Suisse)**

Declaration

In preparing this thesis, I used AI-based tools for supportive purposes. Grammarly, DeepL, and ChatGPT were employed for grammar checking, proofreading, and language improvement. ChatGPT was also used for debugging in R and Python. All writing and analyses were carefully reviewed by myself and subsequently verified by my supervisors.

Remerciements

Pour ne pas ajouter 200 pages à cette thèse, la liste des remerciements est non exhaustive par rapport à l'ampleur de ce que ces personnes ont pu m'apporter au quotidien.

Je tiens premièrement à remercier les membres du jury, **Daniel Croll, Emmanuelle Lerat, Étienne Danchin, Gaëlle Lelandais et Mélanie Morel-Rouhier,** d'avoir accepté de relire et d'évaluer cette thèse.
Je tiens aussi à remercier les membres de mon comité de thèse, **Alice Feurtey, Emmanuelle Morin, Florian Maumus et Pascal Frey,** qui ont pendant 3 ans suivi l'organisation de type "rock'n roll" de ce projet en m'apportant leurs précieux conseils.
Thank you to all the collaborators, notably **Jana Sperschneider** for her advices during the last genome assembly.

Merci à mes directeurs de thèse, **Sébastien et Cécile,** pour leur patience à toute épreuve, leur bienveillance, leurs conseils et leur accompagnement sur ces 3 ans.

Merci à toute l'**équipe écogénomique des interactions** pour leur bienveillance et leurs conseils. Merci **Claire** pour ta bonne humeur, **Aurélie** pour tes conseils de rush de fin de thèse, **Marc** pour ta bienveillance et tes conseils météo qui annoncent toujours un grand soleil sur Champenoux, **Stéphane** pour ta franchise et toutes les brioches pralinées, **Annegret** pour ton aide aux envois de colis, **Nicole**, **Marie-Claude et Solène** pour leur accompagnement (pour les ordres de mission, les frais de mission et la mirabelle), le duo **Claude et Cyrille** pour vos blagues de pauses café, **Manue** pour cette soirée JOBIM mémorable, tes conseils (je lis presque la doc en entier), et mes précieux gènes, **Sandrine** notre maman à tous (oui j'ai rangé mon bureau et repassé ma chemise), **Frédéric** pour ton hypocras, **Laure** pour ta gentillesse, **Mathilde** pour ces midis jeux de société, **Benoît** pour le rosé que nous n'aurons jamais bu, merci **Marceau**, premier fournisseur de mini-protections d'auriculaires et grand analyseur gazeux au MSMS, **Gaurav** pour ton maintien des plantes de bureau qui auraient pu finir en stress hydrique. Merci aussi aux précédents doctorants et post-docs, **Cintia**, tu es un soleil toujours optimiste, **Félix** d'avoir été mon co-bureau et qui, à partir de 15h, est autorisé à faire des blagues presque pas trop beauf. Thank you **Xiaoping** for supporting our jokes with Félix and being always happy in this office. Merci aux récents arrivants, **Héloïse et Aïsha**.

Merci à **Lulu**, pour tes conseils de vie, tes remarques franches et qui ne tombent jamais aux bons moments, tes mojitos revisités et ta bienveillance, grâce à toi la sécurité de l'INRAE est entre de bonnes mains. Merci **Emeline** pour ton aide dans l'organisation des journées et tes conseils bienveillants. Merci **Annie**, pour cette tentative de gâteau rouille-peuplier, pour ces soirées jeu Wasabi, pour ce premier test de four à pizza, qui est ultra fonctionnel !

Merci la plus connue des post-docs, **Elsa**, élue meilleure co-pilote, meilleure Papillon, pour ces soirées à base de Monbazillac, et d'occupation de **Chester** ce canapé inoubliable, merci un peu à **Miro** de ne pas l'avoir mangée en entier (le monde sait que tu es un chat diabolique, attention à toi). Merci Dr **Tatie Fadwa**, pour tes bons tuyaux, tes conseils tranchants mais efficaces, ton écoute et toutes les péripéties que tu nous as fait vivre, on ne s'est jamais ennuyé. Merci **Maell**, pour ton super appareil à croque-monsieur qui bat largement celui d'Elsa ! Merci **Lora**, pour tes appareils à crêpe et raclette high-tech design et ces fous rires jusqu'à 3h du matin juste en tapant sur des cartes.
Thank you **Ivona** for being my partner in crime, my support and being who you are. From apéro to the international congress surrounded by all the big brain, you will shine anywhere.
Merci **Lisa** pour ces soirées raclettes, ta bonne humeur et ton adresse légendaire.
Merci **Ségolène** d'avoir supporté nos "petits singes" qu'on met bien trop souvent sur nos épaules. Merci **Carla**, pour cette après-midi ski, ces soirées aux bars et d'être partie en congrès sous la neige !

Merci à mes camarades de master et notamment **Marie, Valou, Manea et Mélanie** pour ces vacances forcées sans réseau et ces colocations à base de cookies quand le moral était bas.
Merci **Damien**, pour les Dinosaurus et cette peluche qui me remonte toujours le moral.
Merci **Alice**, la meilleure des colocs, pour m'avoir supportée, entendue râler, soutenue dans les meilleurs comme dans les pires moments de cette thèse.
Merci mes **parrains** de m'avoir toujours supportée et conseillée dans mes choix d'études aux quatre coins de la France, sans ces révisions je ne serais pas ici.
Merci **Coco et Fanny**, pour cette soirée Tortuga (je l'aurai ce bateau), et d'avoir été là pour me faire râler sur autre chose que cette thèse !

Merci ma **Subaru WRX STI** d'être la meilleure Fiat de France et d'avoir roulé par tout temps sans jamais totalement mourir, peu importe ma délicatesse de conduite.
Merci au propriétaire de la superbe **206 édition Xbox**, **jantes alu et 205 pingouins**, pour m'avoir fourni un bureau de camping ergonomique lors de cette rédaction, ainsi que des plats "réconfortables", de me suivre et de m'encourager même quand je râlais que "rien ne marche, c'est nul, t'façon j'arrête".

Un énorme merci à **ma famille, mes parents**, pour votre soutien même si je ne peux pas encore soigner la vigne.

Cette thèse clôt 10 années d'études supérieures, au cours desquelles j'ai eu la chance d'avoir toujours eu beaucoup de soutien et d'accompagnement dans les plus beaux et les pires moments, de passer par un nombre incalculable d'assos, de sièges, de soirées, de responsabilités plus ou moins agréables, de pin's, de fromage, etc.

J'ai toujours apprécié apprendre et découvrir, mais je tenais à remercier **Mme Chapat,** professeure de SVT, qui m'a donné l'envie d'être un jour chercheuse.

Pour tous ceux qui m'ont accompagnée sur un bout de chemin, que ce soit à son début, à sa fin, ou tout le long de ce projet, **MERCI**.

I was so determined. I said ‘I will show you even in this situation’

~ Michele Mouton

Glossary

1KFG: *1000 Fungal Genomes Project*

Ac: *Activator*

AGO: *Argonaute*

AP: *Aspartic proteinase*

Avr: *Avirulence*

CNNs: *Convolutional Neural Networks*

DCL: *Dicer-like*

DIRS: *Dictyostelium Intermediate Repeat Sequence*

DRM2: *Domains Rearranged Methyltransferase 2*

Ds: *Dissociation*

DUF: *Domain of Unknown Function*

ETI: *Effector-triggered immunity*

ETS: *Effector-triggered susceptibility*

ERVs: *Endogenous retroviruses*

GAG: *Capside protein*

GOLD: *Genome Online Database*

HGT: *Horizontal Gene Transfer*

HTT: *Horizontal Transfer of Transposable Elements*

INT: *Integrase*

IFN-γ: *Interferon-gamma*

LINEs: *Long Interspersed Nuclear Elements*

MChelper: *Manual Curator Helper*

ORFs: *Open Reading Frames*

PAMPs: *Pathogen-associated molecular patterns*

piRNA: *PIWI-interacting RNA*

PLEs: *Penelope-like elements*

PTI: *PAMP-triggered immunity*

R: *Resistance*

RdDM: *RNA-directed DNA methylation*

RH: *Reverse transcriptase RNase H*

RIP: *Repeat-Induced Point mutation*

RISCs: *RNA-induced silencing complexes*

SF: *Superfamily*

SINEs: *Short Interspersed Nuclear Elements*

siRNAs: *Small interfering RNAs*

SMRT: *Single Molecule Real Time sequencing technology*

T2T: *Telomere to telomeres*

TE: *Transposable element*

TSD: *Target site duplication*

TSS: *Transcription start site*

Figures and supplementals figures

Introduction

Figure 1: Classification of transposable elements (from Wicker et al., 2007)

Figure 2: Examples of important features of transposable elements and their impacts on eukaryotic genomes (from Bourque et al., 2018)

Figure 3: HTT events in the fungal kingdom (from Romeijn et al., 2025)

Figure 4: Examples of rust diseases

Figure 5: Life cycle of a typical heteroecious and macrocyclic rust fungus (from Lorrain et al., 2019).

Figure 6: List of genomic resources available for the subphylum Pucciniomycotina on the Joint Genome Institute MycoCosm portal (as of July 12th, 2025)

Chapter 1

Figure 1: The phylogenetic tree and transposable element content in the Pucciniomycotina

Figure 2: Composition of transposable elements in Pucciniomycotina genomes

Figure 3: Transposable element insertion ages

Figure S1: Association between genome size and gene/TE content

Figure S2: Improvement of TE annotations through manual curation

Figure S3: Shared and specific TE families between the Pucciniomycotina species

Figure S4: Number of LTR-Gypsy elements accumulated through time

Figure S5: TE controls involved in methylations are detected in all species

Figure S6: TE controls involved in RNA-induced silencing complex are detected for all genomes

Figure S7: Contraction and Expansion of multigene families in Pucciniomycotina

Figure S8: Contraction of Multigene Families in Pucciniales

Figure S9: Significant contraction and expansion of multigene families in Pucciniomycotina

Chapter 2 part 1

Figure 1: Comparison of *Melampsora larici-populina* 98AG31 genome assemblies

Figure 2: Genome size and heterozygosity estimation from HiFi read k-mer spectra

Figure 3: Whole-genome dot plots of *Melampsora larici-populina* 98AG31 haplotype-phased assemblies and comparison to the genome assembly version 2

Figure 4: Whole-genome dot plots of *Melampsora allii-populina* 12AY07 haplotype-phased assemblies and comparison with *Melampsora lini* haplotype C
Figure 5: Annotation and distribution of transposable elements (TEs) in the genomes of *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07

Chapter 2 part 2

Figure 6: Genome size and heterozygosity estimation from HiFi read k-mer spectra
Figure 7: Whole-genome dot plots of *Melampsora larici-populina* 08KE26 haplotype-phased assemblies and comparison to reference isolate 98AG31 genome version 3

Chapter 3

Figure 1: Genome-wide Hi-C contact maps
Figure 2: Transposable elements (TEs) conservation and insertion ages among Melampsoraceae
Figure 3: Repeat content in centromeric and telomeric regions of *Melampsora allii-populina* 12AY07 haplotypes
Figure 4: Comparative synteny analysis between three Melampsoraceae species
Figure S1: Repeat content in centromeric and telomeric regions of *Melampsora larici-populina* 98AG31 haplotypes
Figure S2: Repeat content in centromeric and telomeric regions of *Melampsora lini* CH5 haplotypes
Figure S3: Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 0
Figure S4: Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 1
Figure S5: Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 0
Figure S6: Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 1
Figure S7: Distribution of LINE transposable elements along the scaffolds of *Melampsora lini C*
Figure S8: Distribution of LINE transposable elements along the scaffolds of *Melampsora lini H*

Table List

Chapter 1

Table S1: General information of selected genomes

Table S2: TE annotations at the Class level for the Pucciniomycotina genomes

Table S3: TE annotations at the Order level for the Pucciniomycotina genomes

Table S4: TE annotations at the Superfamily level for the Pucciniomycotina genomes

Table S5: Average of the TE sequence identity at the Class level

Table S6: Average of the TE sequence identity at the Order level

Table S7: Shared TEs within the three Melampsoraceae genomes

Table S8: Shared TEs within the five Pucciniaceae genomes

Table S9: RIP signature analysis

Table S10: Orthogroups summary metrics of the selected Pucciniomycotina genomes

Table S11: Orthogroups in contraction of the selected Pucciniomycotina genomes

Chapter 2 part1

Table 1: Assembly metrics of *M. larici-populina* genome version 3

Table 2: Assembly metrics of *M. allii-populina* genome version 2

Table 3: Assembly metrics of *M. larici-populina* 98AG31 and 08KE26

Table S1: Mating type genes homologues

Table S2: Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Class level

Table S3: Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Order level

Table S4: Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Superfamily (SF)

Chapter 2 part 2

Table S5: Mating type genes homologues

Table S6: Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Class level

Table S7: Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Order level

Table S8: Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Superfamily level

Chapter 3

Table 1: Synteny statistics of the three Melampsoraceae genomes

Table 2: Transposable elements density

Table S1: Hi-C matrix contact coverage per scaffold for each Melampsora spp. genome

Table S2: Transposable elements shared within the three Melampsoraceae at Class, Order, and Superfamily level

Table S3: Specific transposable elements within the three Melampsoraceae at Class, Order, and Superfamily level

Table S4: Centromeric regions annotations

Index

Remerciements **3**
Introduction **1**
1) Transposable elements: drivers of genomic diversity and complexity 3
2) Diversity and impacts of TEs in fungi 18
3) The Pucciniales: an order of fungal plant pathogens with complex life cycles and large genomes 27
4) Context and Objectives of the Thesis 40
Chapter I
Ancestral and recent bursts of transposition shaped the massive genomes of plant pathogenic rust fungi **43**
Abstract 46
BACKGROUND 48
RESULTS 52
DISCUSSION 64
CONCLUSION 69
METHODS 69
Supplemental figures 91
Chapter II **101**
BACKGROUND 102
Chapter II - Part 1
Haplotype-phased assemblies of the two poplar rust fungi Melampsora larici-populina and Melampsora allii-populina **103**
INTRODUCTION 104
MATERIALS & METHODS 106
RESULTS 112
DISCUSSION 123
FIGURE LEGENDS 127
Chapter II - Part 2
Haplotype resolved and complete assembly of an asexual lineage of the poplar rust fungus Melampsora larici-populina reveals large structural rearrangements at the mating type locus **130**
INTRODUCTION 131
MATERIALS & METHODS 132
RESULTS AND DISCUSSION 135
FIGURE LEGENDS 142
Chapter III
Impact of transposable elements on the genome organization of three Melampsoraceae species **144**
BACKGROUND 145

INTRODUCTION 146
MATERIALS & METHODS 148
RESULTS 150
DISCUSSION 161
FIGURE LEGENDS 165
Supplemental figures **170**
General Discussion **179**
1) Context and research questions for the study of transposable elements dynamics in rust fungal genomes 180
2) The role of transposable elements in advancing the understanding of Pucciniales genome evolution 182
3) The role of transposable elements in advancing the understanding of Pucciniales biology. 184
4) The contribution of phased genome assemblies in advancing the understanding of Pucciniales biology 186
5) Conclusion and Perspectives 189
Bibliography 192
Résumé étendu en francais **230**
APPENDICES **241**

Introduction

1) Transposable elements: drivers of genomic diversity and complexity

1.1) The discovery of transposable elements

Transposable elements (TEs) are mobile DNA sequences present in nearly all eukaryotic and prokaryotic species (Bourque et al., 2018). They constitute at least 45% of the human genome (Lander et al., 2001), and in certain plant species such as maize, they can account for up to 80% of the genome (Chen et al., 2023). TEs have been studied in a wide range of model organisms, such as *Drosophila melanogaster*, *Arabidopsis thaliana*, *Nicotiana tabacum* and *Neurospora crassa* (Bargues & Lerat, 2017; Kinsey & Helber, 1989; Quesneville, 2020; Zan et al., 2025). They were first discovered in the 1940s by Dr. Barbara McClintock during her investigation of pigment variability in mutant maize kernels (McClintock, 1951). Initially, this study explored the genetic basis of maize kernels' mosaic colour phenotypes and their unstable heritability. She identified that the two genetic elements responsible for this phenotype, called Dissociation (Ds) and Activator (Ac), could change their position on the chromosome. She demonstrated that these elements could insert into pigment-associated genes and suppress their expression, or excise from these loci and restore gene activity, thereby producing the characteristic spotted or various patterns in maize kernels. At this time, the genome was viewed as static and fixed in organisms with no plasticity that contain only gene sequences and

non-coding regions (Chomet & Martienssen, 2017). This discovery of moving genetic elements was the starting point for a paradigm shift in understanding genome stability, and Barbara McClintock was later honoured with the Nobel Prize in Physiology or Medicine in 1983 (Chomet & Martienssen, 2017). Since McClintock's pioneering work, a remarkable diversity of TEs has been discovered throughout the tree of life, highlighting their pervasive presence and evolutionary significance in genomes.

1.2) Classification and diversity of TEs

To organize the diversity of TEs observed across genomes, a unified classification system was proposed by Wicker et al. (Figure 1) (Wicker et al., 2007). This system categorizes TEs based on their transposition mechanisms, structural features, and enzymatic machinery, and establishes three hierarchical levels: Class, Order, and Superfamily (SF). At the highest level, TEs are divided into two major Classes, reflecting their distinct modes of transposition. Class I elements transpose via a "copy-and-paste" mechanism, where the TE is first transcribed into RNA, then reverse-transcribed into DNA, and inserted at a new site. In contrast, Class II elements use a "cut-and-paste" mechanism, where the DNA element itself is excised and relocated without an RNA intermediate. Within each Class, TEs are further divided into Orders, which capture differences in their specific insertion mechanisms, and then into SF, defined by shared structural and sequence features. Many TEs are flanked by repeated sequences and generate characteristic target site duplications (TSDs) upon insertion (Figure 1)(Wicker et al., 2007). Some TEs encode their own proteins essential for mobility, enabling autonomous transposition, whereas others, called non-autonomous TEs, lack such capacity and depend on the transposition machinery provided by autonomous TEs for their mobility.

The Class I (also called retrotransposons) includes the five Orders LTR, DIRS, PLE, LINE, and SINE (Wicker et al., 2007). LTRs are particularly abundant in plant genomes, frequently

observed in fungi, and present in humans, although they are inactive in the human genome (Malik & Eickbush, 2001). LTRs are typically large elements that contain multiple open reading frames (ORFs) encoding the structural and enzymatic proteins required for transposition, notably the Capsid protein (GAG) analogous to virus particles, an aspartic proteinase (AP), a reverse transcriptase RNase H (RH), and an integrase (INT) (Figure 1). The LTRs are divided into different Superfamilies, such as the Gypsy, Copia, Bel-Pao or ERV elements, depending on their ORF composition and Order. The Long Interspersed Nuclear Elements (LINEs) are also autonomous retrotransposons distinguished by a 3′ polyadenylated tail and a unique replication mechanism called target-primed reverse transcription. The LINE Order contains five Superfamily that are divided into 17 different clades. This TE Order is the only active one in the human genome (Hancks & Kazazian, 2012). For instance, the genome contains approximately 500,000 LINE-1 copies, which account for up to 20% of the genome size, of which over 99.9% are inactive (Chénais, 2022). The LINE also contains the Superfamily Jockey that has been described in *D. melanogaster.* They are mostly present in the telomeric region of distant species (McCullers & Steiniger, 2017). Penelope-like elements (PLEs), which encode a reverse transcriptase, are primarily found in plants, but they have also been identified in *D. melanogaster* (Evgen'ev et. al 1997). The Class I elements also comprise non-autonomous elements, including the Short Interspersed Nuclear Elements (SINEs), which depend on the replication machinery of LINEs for mobilization. Although SINEs possess an internal RNA polymerase III promoter, they do not encode any proteins (Wicker et al., 2007)(Figure 1). The most prominent SINE in the human genome is the *Alu* element, which exists in approximately 500,000 copies and can influence gene expression by affecting transcriptional regulation or transcript stability (van Bree et al., 2022). Finally, the *Dictyostelium* Intermediate Repeat Sequence (DIRS) elements encode a tyrosine recombinase and differ from other Class I elements by lacking the integrase ORF, which means they do not have TSD (Wicker et al., 2007).

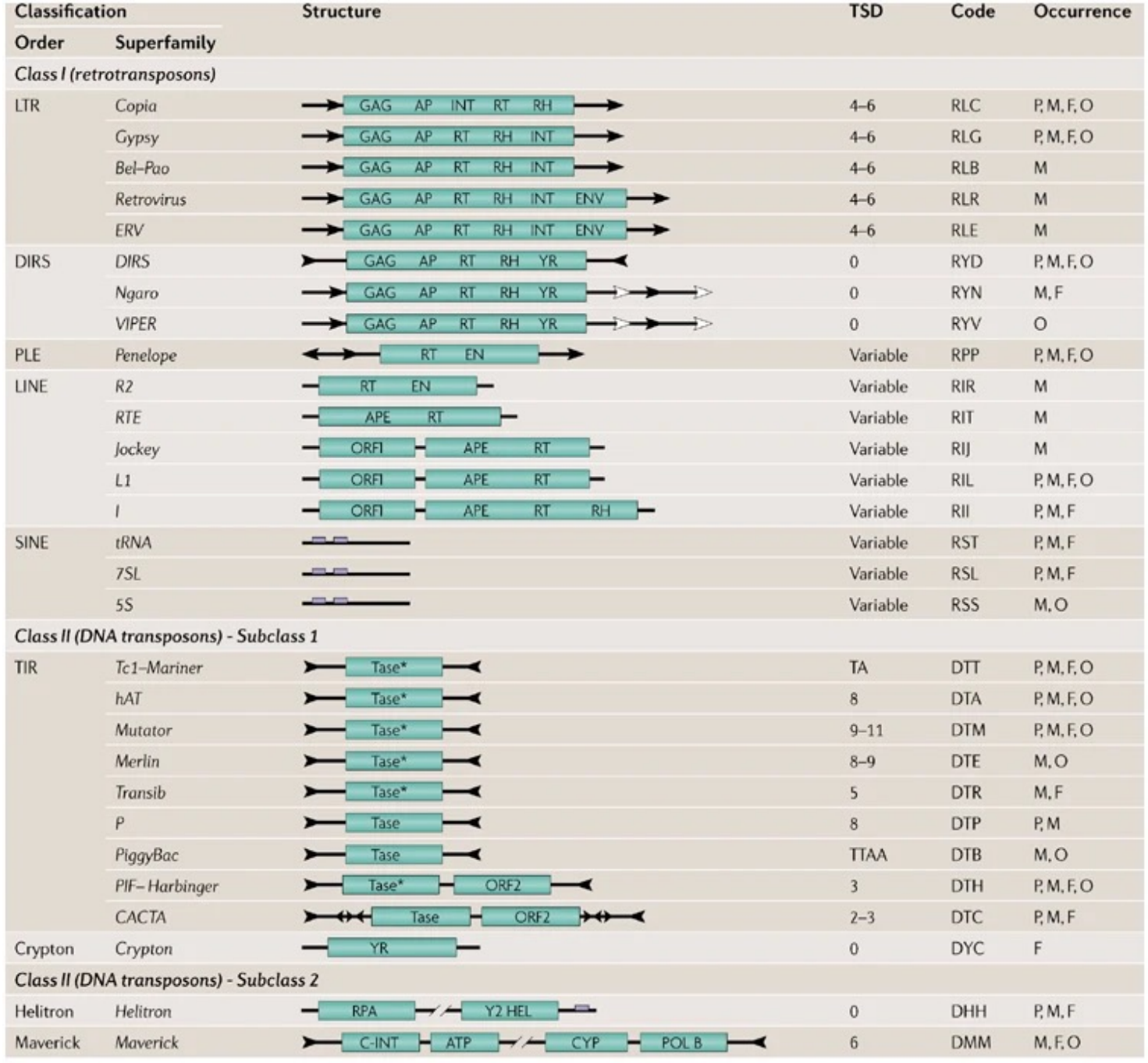


Figure 1**: Classification of transposable elements** (from Wicker et al., 2007):
The TEs are classified in Class I or Class II. The first column describes the TE Order, and the second column represents the categorization of TEs in Superfamily. The third column represents the structure of each TE Superfamily within their type of repeats (long, inverted), the amount of coding region and their annotation. The fourth column describes the TSD for each Superfamily, with its size or pattern. The fifth column is the unified code present in most of the databases and the tool currently used. The last column describes the detection of the TE in the different phylogeny kingdoms

The Class II comprises the four Orders TIR, MITE, Crypton, Helitron, and Maverick. The TIR is one of the biggest Order with nine described Superfamilies that are different in their TSD size and the type of TIR repeat element that flanks the sequences (Figure 1). The TIR transpose itself during chromosomal replication and possesses an intronless transposase gene. This Order notably contains the Superfamily hAT corresponding to the Ac/Ds described by Dr. McClintock (McClintock, 1951). The hAT Superfamily is represented in most eukaryotic species and has probably proliferated in this wide range of hosts with horizontal transfers (Atkinson, 2015). MITE elements are small and non-autonomous TEs of twelve to a few hundred nucleotides. They contain TIR repeats but are directly flanked by the TSD and do not possess any ORF (Lu et al., 2012). They are well studied in rice (*Oryza sativa*) as they have completely invaded their genomes and are associated with most of their genes (58.2% of the genes are linked with such TEs)(Lu et al., 2012). Crypton elements contain a tyrosine recombinase but no RT domain. Their proliferation is probably done by a DNA intermediate through a circular molecule and does not generate TSDs (Wicker et al., 2007). They were first described in fungal genomes, but have since been described in other eukaryotes as animals or oomycetes (Kojima & Jurka, 2011). Helitrons do not have a TSD, nor a RNAse-H-like catalytic domain in the transposase, but they possess a RepHel pattern and a DNA Helicase domain for replication. They also transpose themselves through rolling-circle mechanisms. They have been identified in *Caenorhabditis elegans* and *A. thaliana* (Garrigues et al., 2019; Hollister & Gaut, 2007). Finally, Maverick elements are large TEs, probably autonomous, possess TIRs and contain from four to up to eleven proteins. They are larger compared to other Class II elements with up to 20kb length. As Helitron, they do not contain RT, leading them to the non-autonomous side of Class II TEs (Wicker et al., 2007). They do not seem to be present in mammals or in land plants and seem to have a close evolutionary link with viruses (Barreat & Katzourakis, 2021). This link

with viruses is mainly based on the fact that their structures are similar to those found in viruses such as an adenoviral-like protease and retroviral-like integrase (Pritham et al., 2007).

1.3) Impacts of transposable elements in the host genome

As mobile DNA, TEs can induce both structural and functional genomic modifications. In the genome, even if a TE is inactive, it could remain structurally intact and capable of notably influencing the genome architecture (Bourque et al., 2018; Fouché et al., 2022). Indeed, they can modify the compaction of DNA if they are targeted by TE control mechanisms that repress them, inducing modification on expression with their own regulatory sequences (Bourque et al., 2018)(Figure 2).

Notably, TEs can contribute to large-scale genome rearrangements such as duplications, deletions, and inversions, often through recombination between homologous TE sequences located at different genomic sites (Bourque et al., 2018) (Figure 2). In addition, even TE remnants can still mediate recombination events (Derbyshire et al., 2025; Rizzon et al., 2002). Functionally, TE insertions can modulate the expression of neighbouring genes by inserting into or near regulatory regions, such as promoters or transcription factor binding sites, disrupting or rewiring regulatory landscapes (Figure 2). In humans, TEs are notorious for inducing diseases such as cancer, autoimmune diseases or neurological diseases (Chénais, 2022). Indeed, L1 TE that belongs to the LINE Order and is still active in human genomes is linked with multiple impacts on the cancer cells. A study of colorectal cancer in 22 patients has shown that the insertion of L1 disrupts the expression of a tumour repressor gene, which in turns, leads to a proliferation of abnormal cells (Ardeljan et al., 2020). In addition, TEs can influence gene expression through transcriptional read-through, where RNA polymerase II typically initiates transcription at the TE and fails to terminate properly, continuing into downstream genomic regions. This can lead to overexpression or misexpression of downstream genes and the

production of extended transcripts. Alternatively, TEs can activate cryptic promoters within their own sequences, initiating transcription independently and often generating new or ectopic transcripts that may be chimeric, antisense, or otherwise misregulated (Bourques et al, 2018) (Figure 2). For instance, in blood oranges, the insertion of a Copia-like element near the *Ruby* gene modulates its overexpression in response to low temperatures, leading to the colored phenotype of the fruits (Butelli et al., 2012). The transcription start is settled at the 3' of the Copia-like instead of the original transcription factor. When low temperature stress occurs, the TE expression is activated at the transcription start (TSS) of the Copia-like instead of the original TSS of the Ruby gene, changing its expression. This example also illustrates that environmental stress can activate TEs, triggering their transcription and potentially affecting nearby genes. Additionally, TE insertions within intronic regions may interfere with splicing, potentially resulting in gene disruption or the formation of new exons, a process referred to as exonization, subsequently susceptible to acquiring a new biological function, through exaptation (Schrader & Schmitz, 2019) (Figure 2). These modifications can serve as raw material for evolution, influencing phenotypic diversity and species adaptation (Bourque et al., 2018)(Figure 2). This exonization event has been described in plants such as *A. thaliana* and *O. sativa,* with an increasing number of isoforms and protein variants. Indeed, the D elements possess intronic regions that could be detected as alternative isoforms during the splicing, leading to new exons and potentially new functions (Liu & Charng, 2012).

While the vast majority of TE insertions are likely neutral (e.g., insertions within introns that do not affect gene expression or nested insertions within other TEs that do not compromise genome function), certain TE-mediated variations can play a pivotal role in adaptation. One iconic example of the impact on TEs in adaptation is the case of the peppered moth (*Biston betularia*) in the United Kingdom, which shifted from a white to a dark phenotype during the Industrial Revolution (Hof et al., 2016). The insertion of a DNA transposon in the first intron of the *cortex* gene altered pigmentation, leading to an adaptive phenotype that increased survival

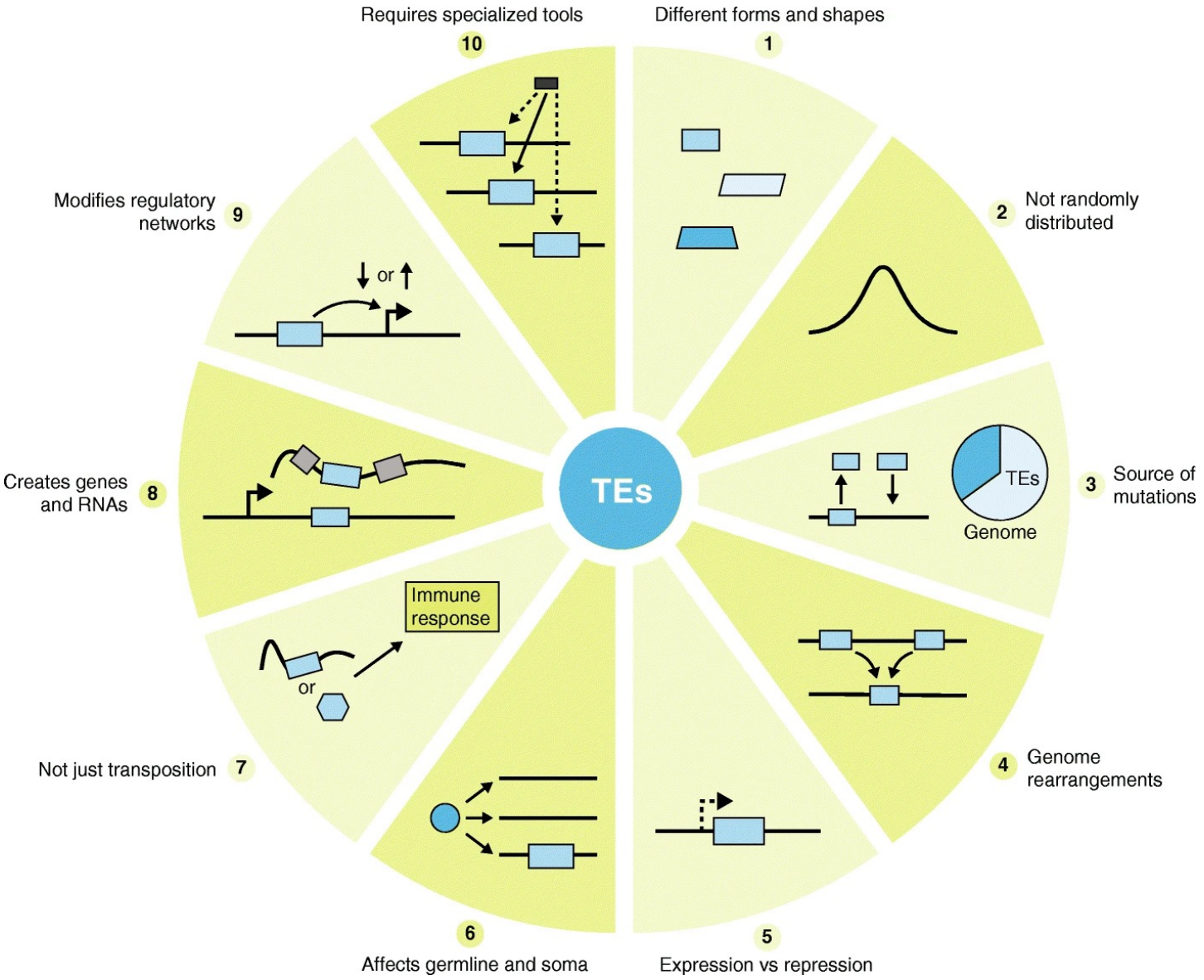


Figure 2: **Examples of important features of transposable elements and their impacts on eukaryotic genomes** (from Bourque et al., 2018). TEs vary in transposition mechanisms, size and content (1), their insertion loci and dynamics are not random (2), they can induce mutations (3), their insertions can lead to rearrangements such as insertions or deletions (4), they can repress gene expression (5), their transposition can affect both germline and somatic cells (6), they can modify immune responses (7), they can create new genes with new functions as well as RNAs (8), they can modify regulatory networks through insertion into regulatory loci (9), they require specific tools for their detection and annotation (10).

in polluted environments (Hof et al., 2016). This insertion leads to an increased production of the *cortex* genes with different isoforms. One of the isoforms is linked with higher melanization concentration.

Finally, TEs can give rise to novel genes through processes of domestication or co-option, whereby TE sequences are repurposed for new cellular functions (Cosby et al., 2019). In this case, a TE or part of a TE becomes co-opted by the host genome and integrated into existing gene regulatory networks or protein-coding regions (Cosby et al., 2019). Notable examples of TE co-option include LTRs, which have been shown to act as alternative promoters or enhancers regulating immune response genes in humans, such as those involved in interferon-gamma (IFN-γ) signalling (Chuong et al., 2016). Additionally, endogenous retroviruses (ERVs) have been domesticated to generate *syncytin* genes, which play a critical role in trophoblast fusion during placental development (Cornelis et al., 2015; Pastuzyn et al., 2018). Furthermore, primate-specific *Alu* elements have been co-opted as exonized sequences or regulatory elements, contributing to gene expression diversity and regulation in humans (Lavi & Carmel, 2018).

Since their first discovery, TEs have undergone a profound paradigm shift, from being dismissed as “junk” or selfish DNA to key components of the genome. TEs are now understood as sources of genetic variation and important drivers of evolution, involved in reshaping gene regulation, creating novel genes, and facilitating rapid adaptive changes.

1.4) Control of TEs in eukaryotic genomes

To preserve genome integrity potentially altered by TE insertions, eukaryotes have evolved multilayered defence systems encompassing epigenetic, transcriptional, post-transcriptional, and genome-level mechanisms. The TE control can operate at the epigenetic level through histone modifications and DNA methylation, which together establish transcriptionally repressive

chromatin states (Bewick et al., 2019; Fouché et al., 2022; Haggerty et al., 2021; Lippman et al., 2004; Nai et al., 2021). In plants, mammals and fungi, TEs are typically marked by cytosine methylation, predominantly at CG sites in mammals and at CG, CHG, and CHH sites in plants, resulting in durable transcriptional silencing (Kumar et al., 2018; Liu & Zhao, 2023; Vanyushin & Ashapkin, 2011). This process is mediated by DNA methyltransferases such as MET1, CMT3, and DRM2 in plants, and DNMT1, DNMT3A, and DNMT3B in mammals and DNMT1, DNMT5, DIM2 and RID in fungi (Bewick et al., 2019). Histone modifications such as H3K9me2/3 and H3K27me3 play key roles in recruiting chromatin-binding proteins such as HP1 to establish and maintain heterochromatin (Jamieson et al., 2016). In *A. thaliana*, for example, the SUVH4/KYP methyltransferase deposits H3K9me2 at TE loci, facilitating DNA methylation via a reinforcing loop (Hung et al., 2023). Indeed, the TE is first detected by SUVH4/KYP, inducing H3K9me2 patterns; this pattern is then recognized by CMT3. Upon this recognition, CMT3 adds more CHG methylation, and this methylation is once again targeted by SUVH4/KYP, creating this reinforcing loop of methylation (Hung et al., 2023). Similarly, in mammals, the zinc-finger transcription factors KRAB-ZFPs recognize TE sequences and recruit KAP1 and the SETDB1 methyltransferase to add H3K9me3, ensuring stable silencing of retroelements (Matsushima et al., 2025).

In plants, the RNA-directed DNA methylation (RdDM) pathway has been described as a plant-specific epigenetic mechanism that plays a crucial role in the transcriptional silencing of TEs and other repetitive sequences. RdDM is initiated when double-stranded or aberrant TE-derived RNAs are processed by Dicer-like (DCL) enzymes into 24-nucleotide small interfering RNAs (siRNAs). These siRNAs are incorporated into Argonaute (AGO) proteins, forming RNA-induced silencing complexes (RISCs), which guide the complex to complementary TE sequences in the genome. Once recruited to the TE loci, the RdDM machinery directs de novo DNA methyltransferases, such as Domains Rearranged Methyltransferase 2 (DRM2), to catalyze cytosine methylation in all sequence contexts (CG, CHG, and CHH) at the TE regions.

This methylation establishes a heterochromatic state, characterized by condensed chromatin and repressive histone modifications, which suppresses TE transcription and mobility (Zilberman et al., 2004). In animals, the PIWI-interacting RNA (piRNA) pathway is essential for silencing TEs in the germline (Charmant et al., 2024). It relies on PIWI proteins and their associated small RNAs (piRNAs), which are derived from TE-rich genomic clusters. These piRNA–PIWI complexes silence TEs both post-transcriptionally, by degrading TE transcripts in the cytoplasm, and transcriptionally, by promoting heterochromatin formation and repressive histone marks at TE loci in the nucleus, as shown in *D. melanogaster* (Saito et al., 2006).

Several fungal species employ an additional genome defence mechanism known as Repeat-Induced Point mutation (RIP), which introduces C to T transitions in duplicated sequences during meiosis (Cambareri et al., 1989; Gladyshev, 2017; Vyas & Kasbekar, 2005). RIP effectively mutates TE copies beyond functionality before they can transpose. This strategy prevents future TE proliferation that could lead to genome disruptions (Gladyshev, 2017). This process was first described in *N. crassa* and has since been reported in other filamentous fungi, including *Leptosphaeria maculans* and *Zymoseptoria tritici* (Freitag et al., 2002; Möller et al., 2021; Van de Wouw et al., 2019). RIP is particularly effective against newly acquired or recently amplified TE families and leaves a distinctive mutational signature in fungal genomes (Cambareri et al., 1989; Freitag et al., 2002; Van de Wouw et al., 2019).

Finally, TEs can contribute to their own inactivation by inserting into other TEs, thereby disrupting their sequence and mobility. For example, in repeat-rich genomes such as in maize, the vast majority of LTR retrotransposons are nested, their sequence frequently interrupted by the insertion of other TEs, leading to fragmentation and loss of transpositional activity. This passive mechanism of inactivation contributes to the observed high abundance of truncated TE copies (Gao et al., 2012).

The dynamic interplay between TE activity and host genome defence mechanisms shapes genome evolution, gene regulation, and structural diversity in eukaryotes. The high number of

species harbouring highly-repeated genomes where TEs have accumulated to extraordinary levels illustrates that, while TE regulation mitigates their potentially deleterious effects, it does not always prevent extensive TE expansion. Thus, high TE content can remain compatible with organismal viability and function as observed in pine trees (Wang et al., 2020). Therefore, understanding TE-host genome regulation dynamics is crucial, as they may highlight the trade-off between genomic variability and stability.

1.5) TE detection and annotation challenges

Annotating the location, identity, and diversity of TEs within a genome assembly is a critical first step in genome analysis and TE studies. As genomic data continues to accumulate at an unprecedented rate, there is a growing demand for reproducible, high-throughput TE annotation strategies that balance automation with accuracy. Accurately identifying and annotating TEs is inherently difficult, due to their remarkable diversity, rapid evolution, and the fact that many are still poorly characterized. This task is further complicated by host-induced mutational processes that degrade TE sequences over time, erasing key features needed for reliable classification. In addition, many eukaryotic genomes contain multigene families and other genes that originated from TEs, which can be mistakenly annotated as TEs rather than as host genes (Bourque et al., 2018). As a result, the full spectrum of TE diversity within eukaryotic genomes is likely still largely underestimated.

Numerous approaches and tools have been developed to address different aspects of the challenge of TE annotation. Many methods tackle specific features of specific TEs types. For example, *LTRharvest* and *LTR_FINDER* are frequently used to detect long terminal repeat retrotransposons (Ellinghaus et al., 2008; Xu & Wang, 2007), while *MITE-Hunter* focuses on MITEs (Han & Wessler, 2010). These tools rely on characteristic sequence motifs or structural

patterns, which make them effective for well-preserved elements but less reliable for older or highly degenerated copies (Goubert et al., 2022).

Thus far, the most commonly used approach for TE annotation in newly assembled genomes involves an automated two-step process: first, the de novo identification of TE families - combining repeat detection with clustering methods to generate a TE library (e.g. RepeatModeler, RepeatModeler2 or TEdenovo) (Flutre et al., 2011; Flynn et al., 2020), followed by an annotation step using the *de novo* generated TE library (e.g. RepeatMasker or TEannot) (Quesneville et al., 2005; Tarailo-Graovac & Chen, 2009). This two-step method provides a broad overview of repeat content and is relatively easy to apply across multiple genomes. Multiple studies aim to integrate different tools and annotation layers within a single pipeline, such as EDTA (Ou et al., 2019), PiRATE (Berthelier et al., 2018), and EarlGrey (Baril et al., 2022), or REPET (Flutre et al., 2011), which combine structural, homology-based, and de novo approaches. REPET is particularly well-suited for large genomes and has been successfully applied in several plant and fungal genome projects (Flutre et al., 2011; Hoede et al., 2014; Jamilloux et al., 2017; Quesneville et al., 2003, 2005).

In parallel, new TE annotation frameworks based on machine learning have emerged. The pipeline DeepTE uses convolutional neural networks (CNNs) trained on large TE datasets to classify them into Superfamilies and clades based on sequence features alone, independent of structural signatures (Yan et al., 2020). Similarly, FasTE employs neural networks for fast and scalable TE classification in large genomes, offering competitive accuracy with significant gains in runtime performance, notably with overlooked fish species (Bell et al., 2022). These approaches show promise for detecting ancient or degraded TE copies that may escape detection by traditional tools. However, such approaches need particularly accurate training datasets to continue to grow and for the model interpretability to improve.

Despite recent methodological advances, some aspects of TE biology remain difficult to resolve with automated annotation pipelines and where TE manual curation remains needed. For instance, nested TEs, in which one or several elements are inserted within another, complicate both detection and classification (Gao et al., 2012). Similarly, domesticated TEs that have been co-opted by the host genome for novel functions are often highly diverged and lack canonical TE features, making them hard to recognize (Jangam et al., 2017). Finally, degraded or ancient TEs, whose sequences have accumulated mutations or have been fragmented over evolutionary time, may escape detection altogether using standard approaches. In a recent attempt to help with manual curation, Orozco-Arias and collaborators developed Manual Curator Helper (MChelper) (Orozco-Arias et al., 2024). This approach reduces redundancy, extends and validates consensus sequences, filters false positives, and classifies previously unclassified elements using clustering, homology, and structural criteria. In addition, it generates supporting outputs such as alignment visualizations and metrics on TE length and copy number, which assist the curator in evaluating and curating sequences flagged as incomplete. This level of scrutiny enables high-confidence annotation of both structure and boundaries, which is crucial for downstream evolutionary or functional analyses (Goubert et al., 2022). Nonetheless, manual curation is time-consuming, requires extensive expertise in TE biology, and is not scalable for large comparative studies involving multiple genomes. It is therefore generally restricted to model organisms or reference-quality assemblies where exhaustive repeat annotation is required (Baril et al., 2022; Goubert et al., 2022).

Accurate TE detection and annotation often rely on the quality of TE library databases. The available databases integrate various data sources, including manually curated, experimentally validated, and *de novo* identified TEs across a broad range of taxa. However, taxonomic representation within these resources often reflects research biases, with mammals and plants being largely covered compared to other groups like fungi. For instance, Repbase is a TE

consensus database that also includes satellite and microsatellites and multicopy genes of reference organisms like human, zebra fish or mice (Bao et al., 2015; Jurka et al., 2005). Dfam is a database for TE detected by a hidden Markov model that covers TE in a smaller number of species (Hubley et al., 2016). Finally, RepetDB is mainly composed of plant and fungal TEs (Amselem et al., 2019). To address the fragmentation of resources and facilitate TE research, the online platform TEHub provides a centralized repertoire of annotation tools, databases, and best practices for TE detection and annotation (Elliott et al., 2021).

2) Diversity and impacts of TEs in fungi

2.1) Study of TEs in Fungal genomes

The kingdom Fungi is estimated to comprise between 2.2 and 3.8 million species (Hawksworth & Lücking, 2017), representing an extraordinary diversity of ecological roles and lifestyles. Yet, only a small fraction of this diversity has been explored at the genomic level. As of 2024, more than 1,500 fungal genomes have been sequenced and released through public databases such as MycoCosm (Grigoriev et al., 2014) and the Genome Online Database (GOLD) (Mukherjee et al., 2021), though the number continues to grow each year. One of the most ambitious efforts in this regard, the *1000 Fungal Genomes Project* (1KFG), launched by the Joint Genome Institute (JGI), aimed to sequence at least two representative species from each fungal family, providing a broad taxonomic coverage of the kingdom (Grigoriev et al., 2014). However, taxonomic representation remains uneven, with some phyla, such as Aphidiomycota and Entomophthoromycota, still lacking reference genomes, likely due to their complex life cycles and difficulty of cultivation (Li et al., 2021). This heterogeneity in taxonomic representation highlights both the progress made in fungal genomics and the challenges that remain.

This increasing availability of high-quality genomes across the kingdom Fungi also revealed the remarkable heterogeneity in fungal genome sizes and contents. Fungal genomes range from compact genomes, such as *Saccharomyces cerevisiae* (12.1 Mb; Goffeau et al., 1996), the first fungal genome sequence that was published, to much larger genome sizes exceeding 1 Gb, as seen in rust fungi like *Phakopsora pachyrhizi* (Gupta et al., 2023) and *Austropuccinia psidii* (Tobias et al., 2021) or in the genome of *Entomorphthora muscae,* an ascomycete fungal (insect pathogens) (Stajich et al., 2024). Larger genome sizes are expected in the Dikarya according to various reports of cytological estimations (Egertová & Sochor, 2017; Hill et al., 2021; Ramos et al., 2015). Such variation is often associated with differences in TE content and proliferation. For

example, in *P. pachyrhizi*, the massive genome expansion correlates with extensive TE accumulation (Edwards et al., 2022; Stajich et al., 2024; Gupta et al., 2023), while in other species with more compact genomes, TEs occupy a much smaller fraction (Castanera et al., 2016). In addition, the availability of multiple genome sequences from different strains within a species has enabled new studies investigating the impact of TEs on genome structure and evolution. For example, the study of 15 *Microbotryum lychnidis-dioicae* species deciphered the evolution of TEs within this species and its impacts (Duhamel et al., 2023). This study notably shows that the accumulation of TE in this species is probably due to a lack of recombination. Moreover, the TE proliferation occurs and is inserted more quickly in the mating-type region rather than the rest of the genome.

2.2) TE amount is variable in the fungal kingdom

Fungal genomes exhibit a wide range of strategies for managing the activity of TEs, reflecting the balance between suppressing potential deleterious effects and the potential to generate genetic diversity. Based on TE abundance, organisation, and host regulatory capacity, fungal genomes can be positioned along a continuum from TE-resistant to TE-tolerant.

At one extreme, TE-resistant fungal genomes maintain low TE content through highly effective host defences. In general, TE-resistant species tend to have compact, gene-dense genomes with very low repeat content. Any new TE insertions in these genomes are likely to disrupt coding sequences and are eliminated by strong purifying selection, further restricting TE proliferation (Fouché et al. 2022). In specific filamentous ascomycetes, it is mainly attributed to the activity of the RIP mechanism to mutate duplicated DNA sequences (Fouché et al, 2022). In *Neurospora crassa*, for example, RIP is highly efficient at degrading most TE copies (Cambareri et al., 1989). However, low TE content is not limited to fungi that exhibit an ability to perform RIP. Species such as *Saccharomyces cerevisiae*, which lacks the machinery for RIP, also harbour

very few TEs (approximately 3% of the genome) (Carr et al., 2012; Maxwell, 2020), suggesting that alternative mechanisms, such as RNA silencing and heterochromatin-mediated repression, may also successfully limit TE proliferation.

At the opposite end of the spectrum stand TE-tolerant genomes in which TEs have proliferated extensively, often resulting in genome size expansions (Fouché et al. 2022). These fungi tolerate a heavy repeat burden, with TEs constituting a substantial fraction of their DNA and imposing relatively little immediate fitness cost. Notably, many obligate biotrophic and basidiomycete pathogens fall into this category. For example, rust fungi or powdery mildew possess some of the largest known fungal genomes with TEs composing up to 90% of the sequence (Gupta et al., 2022; Spanu et al., 2010). It remains unclear what has enabled TEs to accumulate over evolutionary timescales; however, new insertions tend to occur in pre-existing repetitive regions, forming nested TE insertions and minimizing disruption to essential genes. As a result, TE-tolerant species carry massive “mobilomes”, of which most copies appear inactive, exemplifying TE-driven genome inflation (Gupta et al, 2023). This extreme tolerance demonstrates that some fungi can persist with extensive TE content. In such genomes, TEs potentially increase genomic plasticity, providing a reservoir of evolutionary innovation, albeit at the cost of maintaining large, repeat-rich genomes.

Between the TE-resistant and TE-tolerant genomes, we can consider a range of intermediate situations with genomes that could be considered as TE-semi-tolerant, which contain a moderate TE content and exhibit genomic adaptations to mitigate their impact. A hallmark of some of these genomes is their compartmentalization into so-called “two-speed” regions with conserved, gene-rich and repeat-poor compartments interspersed with dynamic, gene-poor and TE-rich compartments. In these species, TEs are not entirely eliminated but are largely confined to specific genomic niches, often AT-rich regions, where they are less likely to disrupt essential genes (Dong et al., 2015). Repeat-rich compartments frequently harbour genes involved in

host–pathogen interactions or environmental adaptation, genes potentially under diversifying selection that could benefit from the variability generated by nearby TEs. This compartmentalized organisation, observed in several plant-pathogenic ascomycetes, exemplifies a partial tolerance of TEs. In the wheat pathogen *Z. tritici*, for instance, approximately 16–24% of the genome is composed of TEs, which cluster in repeat-rich accessory regions and even chromosomes, also enriched in pathogenicity-related gene candidates (Lorrain et al., 2021; Oggenfuss & Croll, 2023). Such genomes exemplify that allowing moderate TE activity could provide adaptive variation while spatially segregating TEs to protect core genomic regions.

2.3) TEs are major drivers of adaptation in fungi

2.3.1) TEs impact fungal adaptation to the host and the environment

Growing evidence depicts TEs as powerful drivers of genomic and phenotypic variation in fungi, enabling rapid adaptation to environmental changes and to the host. Through structural modifications, regulatory rewiring, and the diversification of key genes, TEs generate genetic variability that underlies many adaptive traits, including stress tolerance, virulence, and resistance to antifungal compounds as described by the following examples.

In *Z. tritici*, a TE insertion upstream of the *Zmr1* gene, which controls melanin production, attenuates its expression and reduces melanin levels. This reduction leads to a trade-off between protection against environmental stress and growth rate (Krishnan et al., 2018). Similarly, an LTR retrotransposon inserted upstream of the multidrug transporter gene *MFS1* in *Z. tritici* increases its expression, conferring resistance to fungicides (Omrane et al., 2017). A comparable mechanism occurs in *Botrytis cinerea*, in which a retroelement-like sequence inserted in the *mfsM2* promoter also results in multidrug resistance (Omrane et al., 2017). *Paecilomyces variotii* is a fungal species which is present in a wide range of environments as

the food industry, human infection or even in the water environment and between plant cells (endophyte). In *P. variotii*, TE-mediated adaptation has been documented in response to heavy-metal stress (Urquhart et al., 2020). This species harbours a large TE carrying an entire cluster of heavy-metal resistance genes, enabling survival in highly toxic environments. Together, these examples demonstrate how TEs can swiftly modulate gene expression and contribute to adaptive responses to both environmental and anthropogenic pressures.

TEs also play a central role in the diversification of effector genes (i.e., genes encoding proteins that interfere with host immunity and physiology to the benefit of the pathogen), notably avirulence (Avr) genes that have been particularly studied in various fungal species because they are the target of the host immune system. Avr genes are under strong selection to evade recognition by host resistance genes, and TE insertions facilitate this by disrupting coding sequences, altering transcription, or even generating novel effector variants. For example, in *Magnaporthe oryzae*, the DNA transposon *POT2* inserted into the *AVR-Pib* gene disables recognition by the host immune system and allows for establishment of a successful infection (Shirke et al., 2016). A similar consequence of the loss of avirulence is observed when a Mg-SINE retrotransposon inserts into the AvrPi9 locus. This insertion also results in virulence toward rice cultivars carrying the corresponding resistance gene (Wu et al., 2015). Though the specific transposon classes and insertion sites differ, the functional outcome is analogous. Indeed, in both cases, TE insertions encompass the host recognition and facilitate pathogen infection.

TE insertions in promoter regions can also modulate effector expression: in *M. oryzae*, a TE upstream of *AVR-Pita* downregulates its expression, enhancing pathogen virulence (Khang et al., 2008). Remarkably, in some cases, TEs themselves have been domesticated into functional effectors, as illustrated by the LINE-derived *AvrK1* gene in *Blumeria graminis* f. sp. *hordei* (Amselem et al., 2015). In animal-associated and opportunistic fungal pathogens, TEs also

contribute to adaptation to the host. In the chytrid *Batrachochytrium salamandrivorans*, massive TE proliferation has expanded its genome and duplicated virulence genes, enabling colonization of the hosts (Wacker et al., 2023). In addition, genes that are highly expressed during infection or show signatures of positive selection are predominantly located in the TE-dense compartments in this system. Beyond host-pathogen interactions, TEs also promote adaptation in mutualistic fungi. In the arbuscular mycorrhizal fungus *Rhizophagus irregularis*, TE activity contributes to the expansion of gene families involved in nutrient exchange and signalling (Dallaire et al., 2021). Intriguingly, *R. irregularis* shows developmentally regulated transposon expression: during spore development, numerous TEs are transcriptionally activated. This dynamic TE activity suggests that even without conventional meiosis, the fungus generates new genomic variants via transposition. Other symbiotic fungi show similar patterns of TE-mediated gene expansions associated with a mutualist lifestyle. For instance, comparative genomics of ectomycorrhizal mushrooms has revealed bursts of TE proliferation coinciding with shifts to symbiosis, contributing to expanded genomes and new gene repertoires for host interaction (Martin et al., 2010; Muszewska et al., 2019).

These examples illustrate that TEs, through their capacity to restructure genomes, modulate gene expression, and diversify key gene families, act as engines of fungal adaptation. By supplying the genetic variation required to respond to changing environments and host defences, TEs enable fungi to thrive across a remarkable range of ecological niches.

2.3.2) TEs are drivers of horizontal transfers between fungi

Horizontal Gene Transfer (HGT) refers to the non-vertical acquisition of genetic material between distinct species, bypassing classical inheritance (Keeling & Palmer, 2008). While this phenomenon is well-documented in prokaryotes, growing evidence indicates that it also plays a

role in eukaryotic evolution, including fungi (Romeijn et al., 2025; Van Etten & Bhattacharya, 2020). In fungi, TEs are thought to facilitate HGT by acting as genetic shuttles, transferring not only their sequences but also flanking genes (Keeling & Palmer, 2008). A new type of TEs named Starship has recently been described in fungal species and is involved in HGT between fungal species (Gluck-Thaler et al., 2022). Specifically, starship elements contribute to the spread of adaptive genes, in fungal pathogens, for instance (Bucknell et al., 2023). These elements are a Class of large cargo transposons which typically contain a regulatory gene, the "captain", encoding a putative tyrosine recombinase protein, DUF3435 domain (DUF, Domain of Unknown Function). The cargo region encompasses genes usually associated with virulence, stress response, metabolism, or climate adaptation (Gluck-Thaler et al., 2022). The presence of a tyrosine recombinase gene could form an extrachromosomal circular intermediate, suggesting a mechanism of transposition *via* a recombinase-mediated mechanism, although this remains hypothetical (Gluck-Thaler et al., 2022).

In pathogenic fungi, several examples have emerged supporting the role of starships in HGT. The most notable example involves the pathogenicity-related gene *ToxA*, which has been identified in three unrelated wheat pathogens: *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, and *Bipolaris sorokiniana* (McDonald et al. 2019). *ToxA* encodes a virulence protein that contributes to severe disease symptoms, and its sequence resides close to a highly conserved region of ~62kbp, strongly suggesting acquisition through HGT. Specifically, *ToxA* is flanked by two hAT transposons that together form a composite transposon termed ToxhAT. This mobile unit, embedded within the Sanctuary and Horizon starship cargo regions of 170-196 kbp, was shown to mediate the horizontal transfer of *ToxA* among multiple wheat pathogens (Buknell et al 2024). In *Z. tritici*, a Starship element named Swordfish carries the *Sir* locus, encoding an NAD-dependent deacetylase (sirtuin), which is among the strongest genome-wide association loci for climate adaptation in this species (Tralamazza et al., 2024). If horizontal transfer was not

demonstrated here, the duplication of the Sir locus in the population was shown to follow the geographical dispersal of *Z. tritici* (Tralamazza et al., 2024). To date, starship elements have been identified exclusively in Ascomycetes and appear to be absent from Basidiomycetes (Gluck-Thaler et al., 2022).

Beyond these notable examples of starship-mediated gene transfers, recent large-scale surveys have revealed that horizontal transfer of transposable elements (HTT) is much more pervasive in fungi than previously appreciated (Romeijn et al., 2025) (Figure 3). A systematic analysis of 1,348 fungal genomes uncovered over 5,500 independent HTT events, in some species contributing up to 70% of their total TE content (Romeijn et al., 2025). These transfers are particularly frequent among taxa with gene-sparse, repeat-rich genome compartments and in species lacking known fungal-specific TE defence mechanisms, suggesting that both ecological and genomic factors influence the propensity for HTT. Together, these findings illuminate the dynamic role of TEs as vehicles of horizontal transfer in fungi, shaping genome evolution by disseminating both their own sequences and, potentially, adaptive host genes across species boundaries.

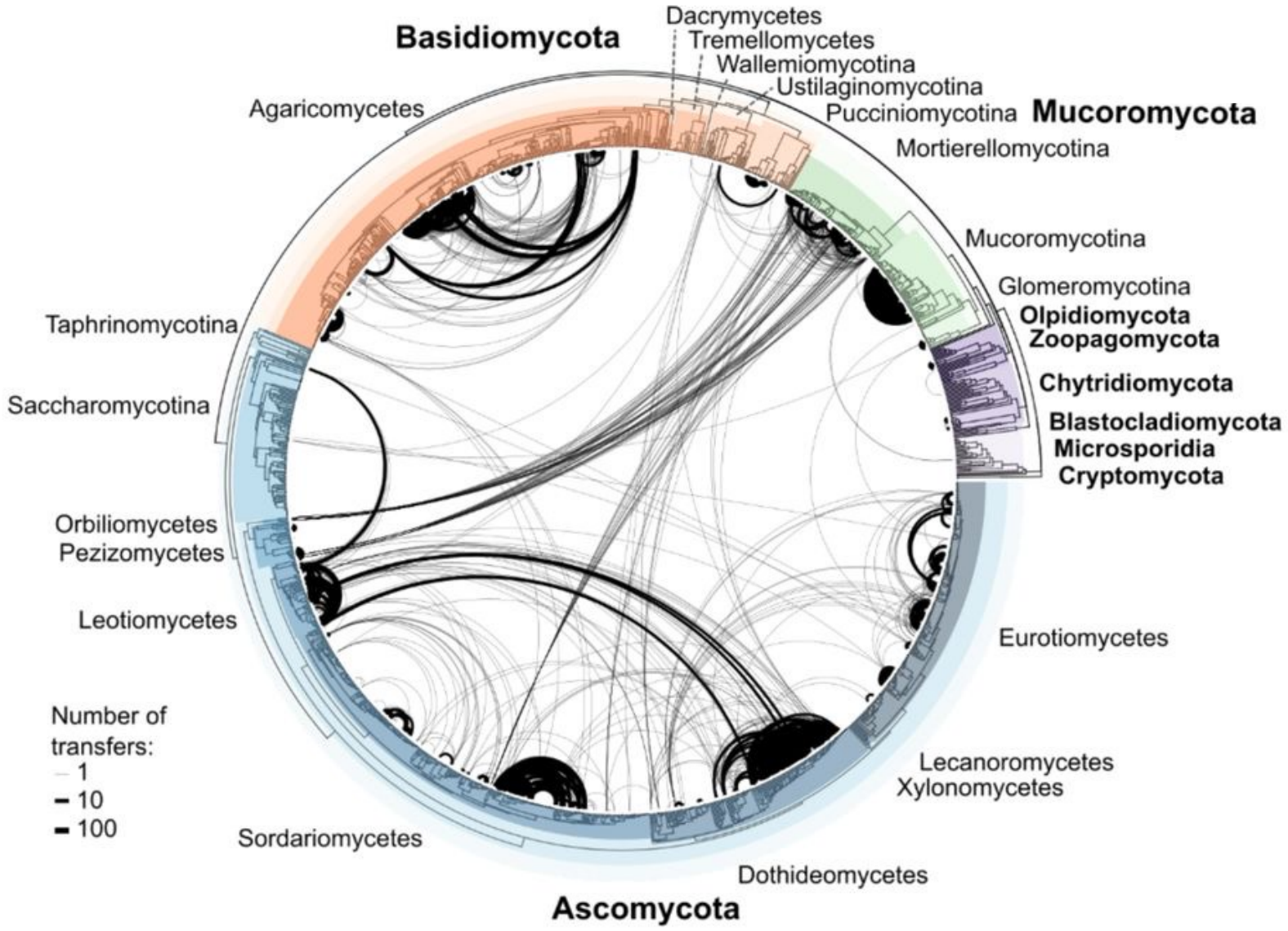


Figure 3: **HTT events in the fungal kingdom (from Romeijn et al., 2025)**. The 1,348 available genomes of the fungal kingdom are represented in the circle plot branch. Each subphylum is described, and each phylum corresponds to a colour: Ascomycota in blue, Basidiomycota in orange, Mucoromycota in green, and other basal phyla of the kingdom fungi in purple. The HTT are represented by solid lines, and the thickness of the line corresponds to the number of transfers. The Mucoromycota seem to share HTT with Ascomycota, while Basidiomycota HTT are more prevalent with other species of the same phylum.

3) The Pucciniales: an order of fungal plant pathogens with complex life cycles and large genomes

3.1) Taxonomy and diversity of Pucciniales and their impacts on agroecosystems

The fungal kingdom contains millions of species with diverse lifestyles, including decomposers, mutualistics and pathogens. The rust fungi (or Pucciniales) represent the largest order of plant pathogens, grouping more than 8,000 species (Aime et al., 2018). The Pucciniales belong to the phylum *Basidiomycota*, subphylum Pucciniomycotina, and are characterized by their complex life cycles and biotrophic status, *i.e.*, strict dependence on living host tissue. Phylogenetic studies suggest that the *Pucciniales* order originated around 175 million years ago (Aime et al., 2014; Aime & McTaggart, 2021). This order is remarkably diverse, encompassing 125 genera across 11 families, including *Mikronegeriaceae*, *Coleosporiaceae*, *Melampsoraceae*, *Phakopsoraceae*, *Phragmidiaceae*, *Pileolariaceae*, *Pucciniaceae*, *Pucciniastraceae*, *Raveneliaceae*, *Sphaerophragmiaceae*, and *Uropyxidaceae*. The *Melampsoraceae*, which contains notably the poplar rust and the flax rust fungi, is estimated to have diverged between 85 and 91 million years ago, whereas *Pucciniaceae*, containing species causing rust diseases in several crops, such as the wheat rust or the barley rust fungi, diverged 44 million years ago (Aime et al., 2017).

Throughout history, species of the Pucciniales have exerted a profound influence on human societies, primarily through recurrent epidemics that have ravaged key crops and significantly reshaped farming systems across diverse geographic regions. Rust fungi can infect a broad spectrum of vascular plants, including ferns, trees, and major crops such as wheat or soybean (Figure 4). These pathogens have posed a persistent threat to food security since the dawn of

agriculture (Gregory, 1948; Henningsen et al., 2025; Lorrain et al., 2019). The severity of rust fungi epidemics on food production was historically so significant that the ancient Romans dedicated a festival (called Robigalia) held annually on April 25th, to honor Robigus, the deity invoked to protect wheat fields from disease (Kislev, 1982). The repercussions of rust diseases have continued through time, with outbreaks often resulting in significant agricultural losses and broader societal impacts. One notable example is the British shift toward tea consumption originating from a severe rust epidemic. Initially engaged in coffee production, British planters were compelled to transition after severe epidemics of coffee leaf rust, caused by *Hemileia vastatrix*, drastically reduced coffee yields (Henningsen et al., 2025; Lorrain et al., 2019; Orozco-Arias et al., 2022). To date, *H. vastatrix* continues to pose a global threat to coffee production, contributing to yield losses exceeding 35% annually (Orozco-Arias et al., 2022; Talhinhas et al., 2017). Rust diseases also affect forestry; for instance, *Melampsora larici-populina*, a causal agent of poplar rust, inflicts substantial damage on commercial poplar plantations (Rubin, 2008)(Figure 4). *Austropuccinia psidii*, known as the myrtle rust fungus, is closely surveyed as it strongly impacts eucalypt and closely related species in native forest ecosystems in Australia (Pegg et al., 2017). This species also has an important impact on eucalypt in South American forests (Pegg et al., 2017). Collectively, multiple examples highlight the continuous and evolving threat posed by rust fungi to both agriculture and forestry, necessitating sustained research and management efforts to understand their biology and mitigate their global impact.

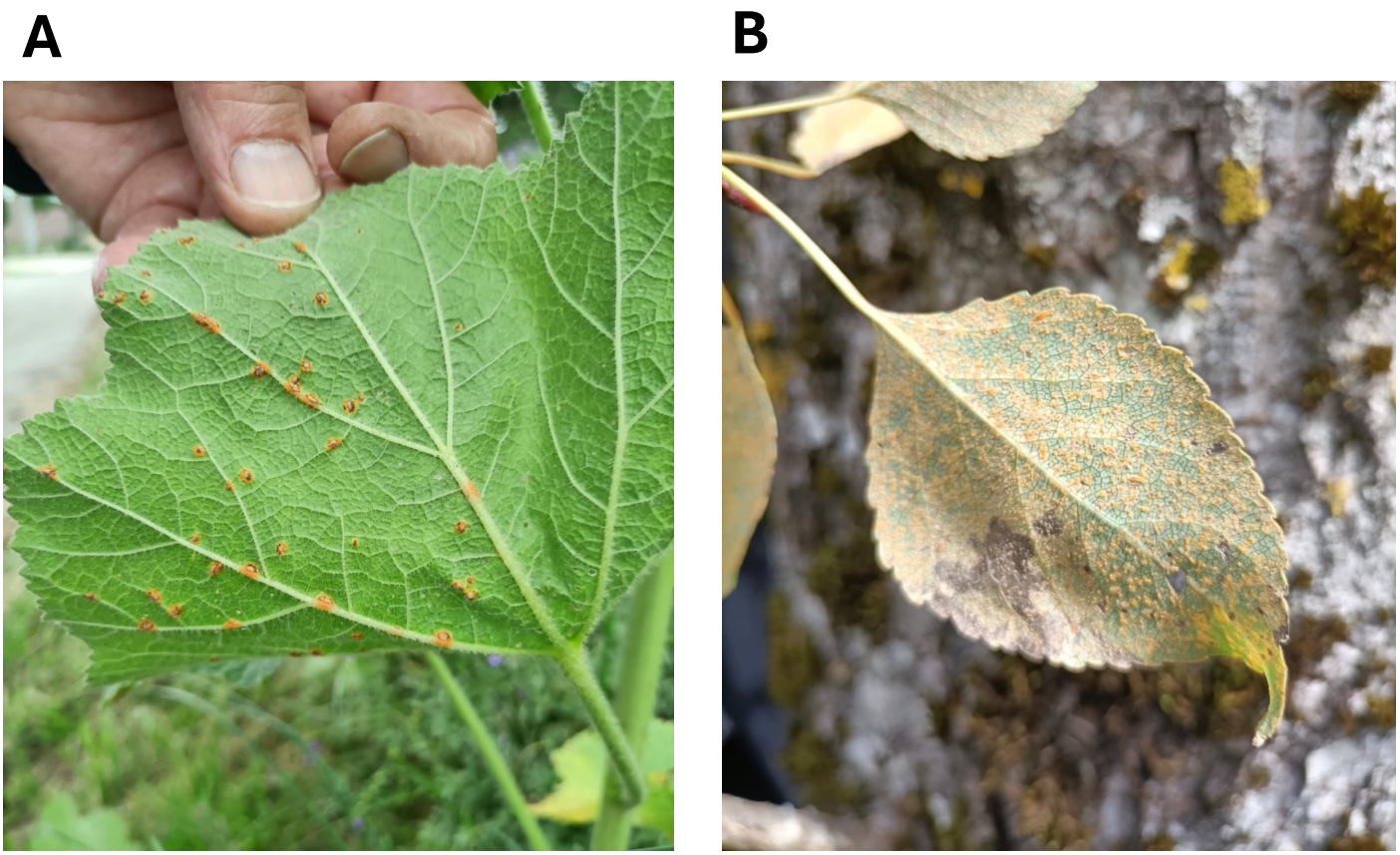


Figure 4: **Examples of rust diseases** on *Alcea rosea* (*Puccinia malvacearum*) in Isere and on poplar leaves (*Melampsora spp.*) in Poland (personal pictures taken in 2024). On each picture, the orange pustules represent masses of urediniospores produced by the respective rust species on the back of the leaves.

3.2) The Pucciniales are obligate biotrophs with complex life cycles

3.2.1) Pucciniales exhibit various life cycles

As obligate biotrophs, Pucciniales must infect a living host to fully complete their life cycle. Due to this biotrophic nature, they cannot be maintained *in vitro* on an axenic medium. The life cycles of Pucciniales are among the most complex in the kingdom Fungi, exhibiting a remarkable diversity across species (Lorrain et al., 2019). Some rust fungi are heteroecious, requiring two distinct and often phylogenetically unrelated host species to complete their sexual and asexual stages (Figure 5). In contrast, autoecious species are capable of completing their entire life cycle on a single host species. Host specificity also varies: while most rust fungi infect only a specific host species or a narrow range of hosts, others, such as the soybean rust fungus *P. pachyrhizi*, have an extensive host range encompassing hundreds of legume plants in the Fabaceae (Harmon et al., 2006). Many rust fungi exhibit highly elaborated life cycles that can involve up to five distinct spore forms: pycniospores, aeciospores, urediniospores, teliospores, and basidiospores. In some cases, these five spore stages are all produced on a single host species, characterizing these fungi as both macrocyclic and autoecious, as for *Melampsora lini*, the flax rust pathogen (Figure 5) (Lorrain et al., 2019). By contrast of the five distinct spore types, reduced cycles are described as demicyclic (4 spore types), hemicyclic (2 spore types) or microcylic (2 spore types). Interestingly, for some economically important rust fungi, such as *P. pachyrhizi* and *H. vastatrix*, full life cycles remain partially unresolved, as we have no knowledge of existence of a possible alternating host for sexual reproduction (Hossain et al., 2024; Talhinhas et al., 2017).

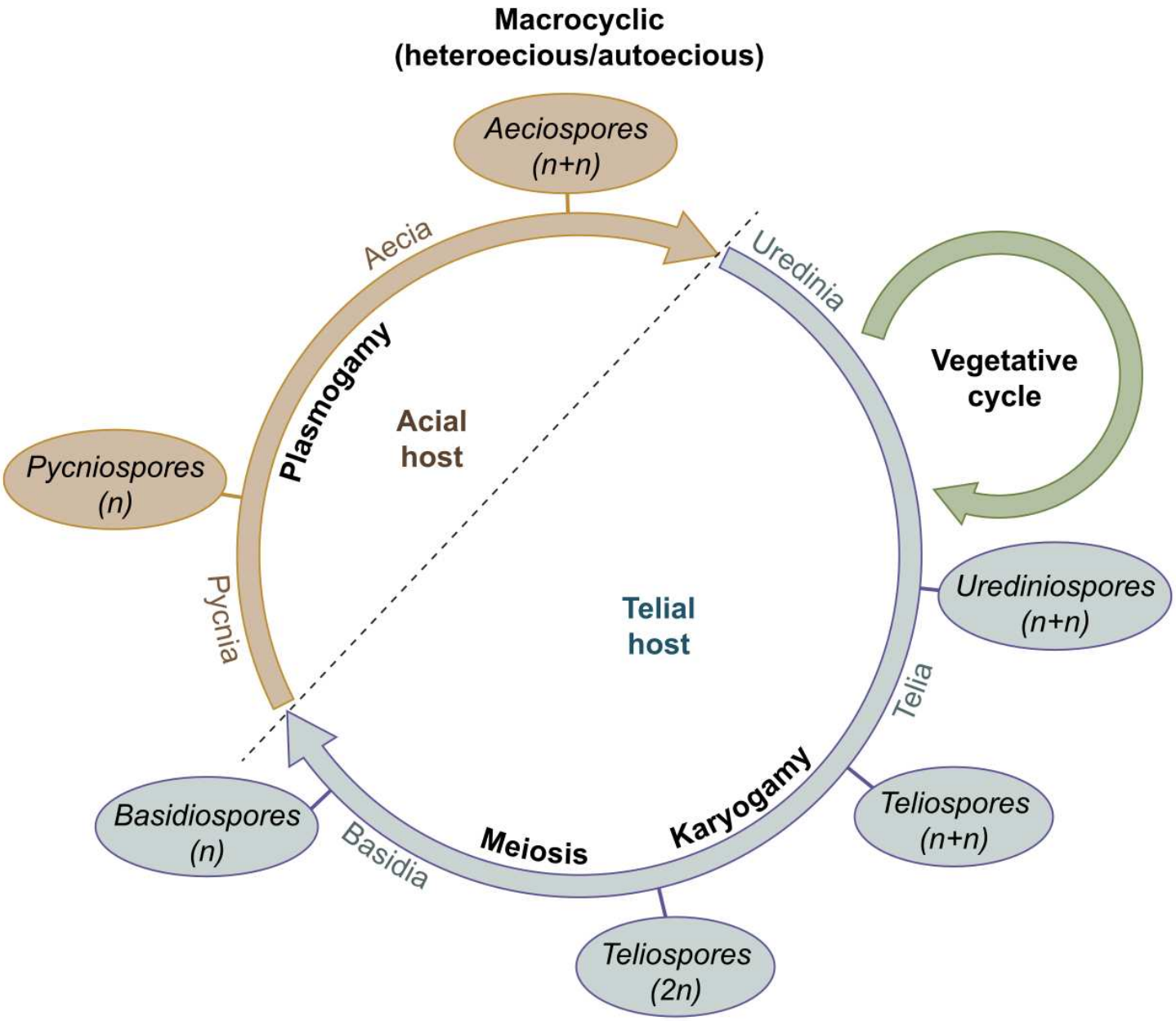


Figure 5: **The life cycle of a typical heteroecious and macrocyclic rust fungus** (from Lorrain et al., 2019). This schematic representation of the typical heteroecious and macrocyclic life cycle shared by many rust species. The first host (referred to as telial host or sporothallus host) is represented in grey, the second host (referred to as acial host or gametothallus host) is represented in brown. The vegetative cycle i.e. the clonal propagation of urediniospores, is represented in green. Each spore type is represented in an oval shape, the haploid forms are marked by (n), the dikaryotic stage by (n+n) and the diploid stage by (2n). Some crucial biological stages and events are indicated between the spores along the cycle. Meiosis occurs in late teliospores (2n) and leads to the production of basidia. The haploid basidiospores (n) then infect an alternate (or aecial) host, where specialized structures like protoaecia and pycnia develop. Within pycnia, haploid pycniospores (n) and haploid receptive hyphae are formed, and exchange of pycniospores enables sexual reproduction via plasmogamy on receptive hyphae with compatible mating types. This fertilization leads to the development of dikaryotic aecia, which produce wind-dispersed aeciospores (n+n). These aeciospores infect the primary (or telial) host, initiating cycles of production of urediniospores (i.e. vegetative propagation throughout the host growing season). The last phase is most often the cause of epidemics. As the season ends, uredinia transition into telia, which overwinter in the host tissues. Before the dormant phase, karyogamy occurs within teliospores, restoring diploidy and completing the cycle.

3.2.2) Interactions between Pucciniales and their plant hosts

Plants have evolved a multilayered immune system to perceive and defend against microbial attacks. The current paradigm in plant immunity provided under the form of the so-called “zig-zag” model was established by Jones and Dangl in 2006. Although the model is debated in some details by the scientific community, is a simple general scheme describing events occurring at the molecular level between a pathogen and its host plant and leading to resistance when the immune response is effective or to disease when pathogen infection is effective (Jones and Dangl, 2006; Dodds and Rathjen, 2010; Dangl et al. 2013). A first barrier relies on the recognition of conserved pathogen-associated molecular patterns (PAMPs) by host surface receptors called pattern recognition receptors. This recognition activates the PAMP-triggered immunity (PTI), which restricts early colonization (Jones and Dangl, 2006; Dodds and Rathjen, 2010). However, pathogens can suppress PTI by secreting effector proteins that interfere with plant immune signalling pathways and the host physiology to establish a compatible interaction that leads to infection and disease; a process termed effector-triggered susceptibility (ETS) (Mentlak et al., 2012). There is a second layer of defense, known as effector-triggered immunity (ETI), during which pathogen effectors are recognized by intracellular receptors encoded by plant resistance genes. Effectors recognized by resistance proteins are also called avirulence proteins, since the product of the pathogen avrirulence gene leads to recognition by the host. This principle was first described by Harold H. Flor in the twentieth century at the genetic level in the so-called gene-for-gene model, which states that for a dominant resistance gene in the plant, there is a corresponding dominant avirulence gene in the pathogen that determines its recognition (Flor, 1971; Dodds, 2024). When an effector is recognized by a resistance protein, a strong immune reaction is activated, often leading to localized cell death (i.e., hypersensitive response) which arrest the infection by the pathogen. In the PTI-ETS-ETI “zig-zag model”, when

a pathogen effector escape recognition by resistance genes in the course of evolution, e.g., through mutations, the pathogen resumes ETS. Rust fungi, as obligate biotrophs, are under constant selective pressure in this molecular arms race. They must suppress host immunity while ensuring that the host remains alive to complete their life cycle. This dynamic interplay between Avr effectors and plant resistance genes has profoundly shaped both pathogen adaptation and host resistance evolution. It is to note that the gene-for-gene model was established by Harold H. Flor in the flax-flax rust fungus model interaction and that this pathosystem has provided a series of advances in the genetic characterization of effector genes and in their functional analysis (Flor, 1971; Dodds, 2024). The last two decades have been marked by particularly strong efforts from the scientific community to advance the understanding of the mechanisms of molecular interactions between plants and their pathogens for the benefit of a durable agriculture (Dodds, 2024). Such an effort has led to the discovery and the prediction of many effectors and/or avriulence genes in many pathosystems and this field has been particularly active in the study of rust fungi (Petre et al. 2014).

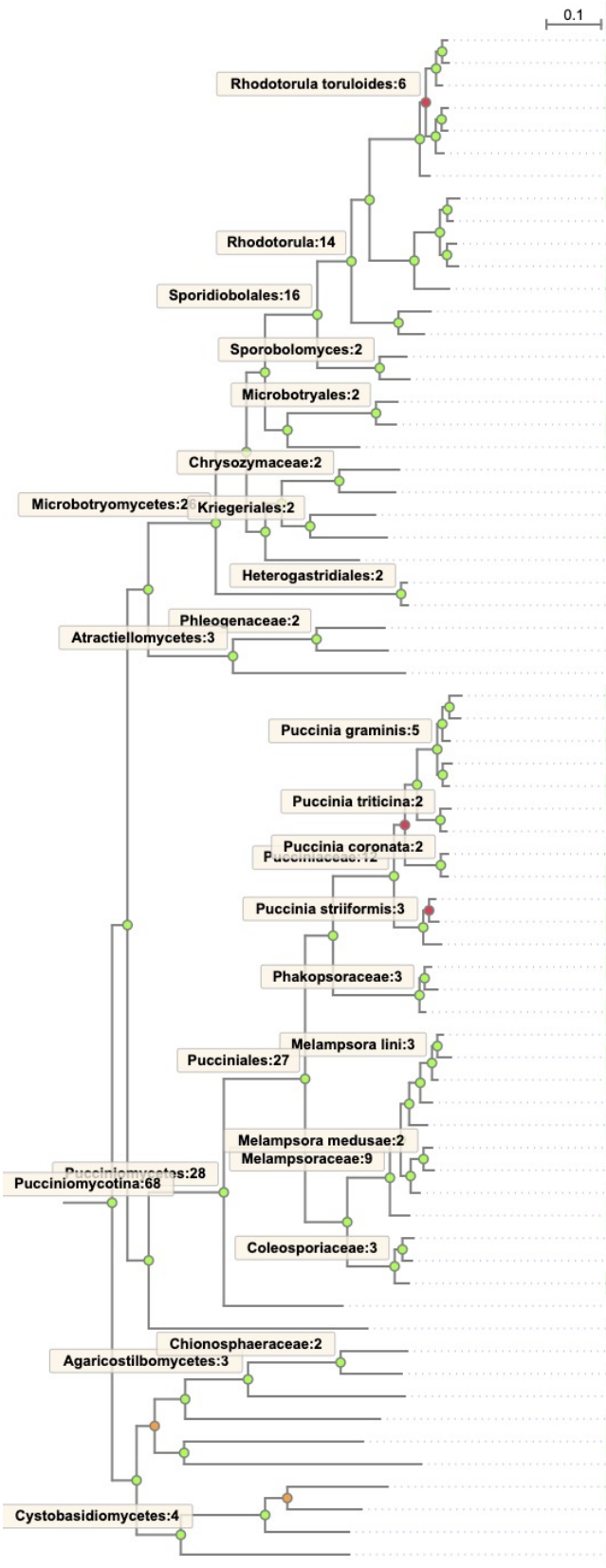
3.3) Rust fungi in the genomic era

Advances in sequencing technologies and progress in genomics have dramatically changed our knowledge of rust fungi biology (Lorrain et al., 2019; Petre & Duplessis, 2022). Although rust fungi have been the focus of biologists for a very long time, especially in plant pathology, progress at the molecular level has been exponential since their genome's availability (Henningsen et al., 2025). Genome analysis in rust fungi is particularly challenging due to their obligate biotrophic nature, and also due to the complexity of their genomes, which are large in size, highly repetitive, and demanding in term of sequencing, assembly and analysis (Gupta et al., 2023; Henningsen et al., 2025; Lorrain et al., 2019; Tobias et al., 2021). The complexity of the rust genome architecture has only been highlighted with the improvement of sequencing

technologies. Indeed, some genomic studies conducted with rust fungi have highlighted how the repetitive nature of the genome impeded proper assembly when using short-read technologies (Cantu et al., 2011; Loehrer et al., 2014). Accessibility to affordable high-throughput long-read sequencing platforms has drastically improved the quality of assemblies of rust fungal genomes, moving from highly fragmented scaffolding to telomere-to-telomere haplotype-phased reconstructed references (Duan et al., 2022; Henningsen et al., 2025; Liang et al., 2023; Schwessinger et al., 2018; Wang et al., 2024). The latest genomic resources provide foundational insights into gene content, genome size variation, and structural organization, laying the groundwork for more detailed comparative, functional, and evolutionary analyses in rust fungi.

The first Pucciniales genomes (*M. larici-populina* and *P. graminis* f. sp. *tritici*) were generated with the Sanger sequencing technology (Duplessis et al., 2011). Despite its accuracy, Sanger sequencing is limited by low throughput and high costs in terms of manpower, reagents and time. The second-generation sequencing platforms, such as sequencing-by-synthesis (Illumina) and pyrosequencing (454), revolutionized the field of genomics by massively increasing read sequencing outputs and lowering per-base costs. However, these short-read technologies were not adapted to assemble highly repetitive genomes such as those of rust fungi, leading to highly fragmented assemblies (Cantu et al. 2011; Cuomo et al., 2017). Third-generation sequencing technologies have advanced rust fungal genome sequencing by enabling the generation of long reads, often exceeding 10 kb, which improve assembly contiguity and structural resolution (Scarano et al., 2024). Technologies like single-molecule real-time sequencing technology (SMRT, PacBio) and nanopore sequencing technology (Oxford Nanopore Technologies, ONT) allow for accurate, real-time sequencing of complex, repeat-rich genomes. These recent technologies are optimal to overcome the challenges posed by the large size and the repetitive nature of rust genomes. Rust species are mostly in an asexual stage during their life cycle, with

spores in a dikaryotic form, meaning that both parental nuclei coexist in the same spore (n+n). The use of third-generation sequencing technologies does not allow for the complete resolution of the haplotypes from both nuclei in the generated sequences. As a result, even if it is possible to discriminate between fragments of primary and secondary haplotypes with PacBio or ONT technologies, it is not possible to reconstruct each haplotype entirely separately. The integration of chromatin conformation information through chromosome conformation capture technologies (Hi-C) has recently addressed previous limitations by enabling the assignment of assembled sequences to respective nuclear genomes (Duan et al., 2022). Indeed, the HiC-sequencing technology relies on the detection of the interaction between physically closed sequences after fixing DNA molecules in their 3D conformation in the nucleus (Ramírez et al., 2017). Multiple bioinformatic tools are released to assemble and analyze such data (e.g., hifiasm Cheng et al., 2021). In 2021, the first haplotype-resolved genome assembly of *P. graminis* f. sp. *tritici* was achieved by combining long-read and Hi-C sequencing (Henningsen 2021), establishing a new standard for generating high-quality, haplotype-phased reference genomes in rust fungi. Later, Nuclear Phaser, released in 2022, is a tool based on gene bins separation combined with HiC information to separate haplotype (Duan et al., 2022). This tool has been used to resolve the latest Pucciniales genomes as *M. lini* (Sperschneider et al., 2025). Moreover, in addition to providing a fully resolved haplotype, the use of HiC-map also provides evidence of centromere areas and 3D structural conformations. With these improvements of both bioinformatics tools and sequencing technologies, the future generations of rust genomes will belong to a new area of genomics studies.



Name	Assembly(bp)	Genes	Published
Rhodosporidium toruloides NP11	20,223,942	8,140	Zhu Z et al., 2012
Rhodosporidium toruloides IFO0559_1	21,148,033	8,165	Zhang S et al., 2016
Rhodosporidium toruloides ATCC26217 v1.0	20,209,575	8,226	Mondo SJ et al., 2025
Rhodosporidium glutinis ATCC 204091 v1.0	20,478,880	7,374	Paul D et al., 2014
Rhodosporidium toruloides IFO0880 v4.0	20,683,845	8,490	Coradetti ST et al., 2018
Rhodosporidium toruloides IFO1236_1	20,452,897	7,479	Zhang S et al., 2016
Rhodotorula glutinis v1.0	20,088,400	8,277	
Rhodotorula sp. CCFEE 5036	19,072,598	6,434	Coleine C et al., 2020
Rhodotorula mucilaginosa ATCC58901 v1.0	19,873,570	7,810	
Rhodotorula mucilaginosa J31 v1.0	20,111,900	7,934	
Rhodotorula sp. JG-1b	19,393,393	6,680	Goordial J et al., 2016
Rhodotorula taiwanensis MD1149	19,568,495	7,014	Tkavc R et al., 2017
Rhodotorula graminis strain WP1 v1.1	21,013,998	7,283	Firrincieli A et al., 2015
Rhodosporidium diobovatum 08-225 v1.0	21,142,622	7,974	
Sporidiobolus pararoseus UCDFST 68-350 v1.0	21,101,591	8,378	
Sporobolomyces roseus v1.0	21,172,659	5,536	
Microbotryum intermedium 1389 BM 12 12	23,461,035	8,142	Branco S et al., 2017
Microbotryum lychnidis-dioicae p1A1 Lamole	26,138,885	7,819	Perlin MH et al., 2015
Leucosporidiella creatinivora UCDFST 62-1032 v1.0	26,329,764	9,854	Mondo SJ et al., 2017
Slooffia pilatii Y-17178 v1.0	34,802,152	12,822	
Microbotryomycetes incertae sedis isolate P6C30 v1.0	33,036,925	11,616	
Kriegeriales sp. isolate P6C61 v1.0	75,288,537	15,297	
Meredithblackwellia eburnea MCA 4105 v1.0	30,684,950	11,135	Mondo SJ et al., 2025
Microbotryomycetes sp. P2C70 v1.0	25,493,598	9,956	
Heterogastridium pycnidioideum ATCC MYA-4631 v1.0	17,579,447	7,658	
Hyalopycnis blepharistoma ATCC 48560 v1.0	17,929,094	7,872	
Phleogena faginea OMC 1709 v1.0	61,418,375	13,022	
Saccosoma farinaceum IVC104 v1.0	142,433,378	20,245	
Atractiellales rhizophila PMI 95 v2.0	51,470,471	17,606	
Puccinia graminis f. sp. tritici Ug99 haplotype C	86,818,685	18,536	Li F et al., 2019
Puccinia graminis f. sp. tritici v2.0	88,644,628	15,979	Duplessis S et al., 2011
Puccinia graminis f. sp. tritici 21-0 haplotype B	89,387,476	18,843	Li F et al., 2019
Puccinia graminis f. sp. tritici 21-0 haplotype A	88,783,524	19,164	Li F et al., 2019
Puccinia graminis f. sp. tritici Ug99 haplotype A	91,841,532	19,772	Li F et al., 2019
Puccinia triticina 1-1 BBBD Race 1	135,344,771	15,685	Cuomo CA et al., 2017
Puccinia triticina Pt76	253,664,742	31,930	
Puccinia coronata avenae 12NC29	166,280,221	28,270	Nazareno ES et al., 2018
Puccinia coronata avenae 12SD80	150,467,806	26,323	Nazareno ES et al., 2018
Puccinia striiformis f. sp. tritici 104 E137 A-	156,834,097	30,249	Schwessinger B et al., 2018
Puccinia striiformis f. sp. tritici PST-78 v1.0	117,311,690	20,482	Cuomo CA et al., 2017
Puccinia striiformis f. sp. tritici PST-130	64,782,816	18,021	Cantu D et al., 2011
Phakopsora pachyrhizi MT2006 v1.0	1,057,430,280	19,618	Gupta YK et al., 2023
Phakopsora pachyrhizi UFV02 v2.1	1,273,655,505	22,467	Gupta YK et al., 2023
Phakopsora pachyrhizi K8108 v2.0	1,082,905,572	18,216	Gupta YK et al., 2023
Melampsora lini CH5 haplotype C v2.0	239,950,640	17,015	Sperschneider J et al., 2025
Melampsora lini CH5	189,516,653	16,335	Nemri A et al., 2014
Melampsora lini CH5 haplotype H v2.0	242,316,735	16,654	Sperschneider J et al., 2025
Melampsora allii-populina 12AY07 v1.0	335,730,080	23,089	
Melampsora larici-populina v2.0	109,877,997	19,550	Persoons A et al., 2022
Melampsora medusae f. sp. tremuloidae Mmt05Ida529 v1.0	145,187,885	22,850	
Melampsora medusae f. sp. Deltoidae Mmd05TRE539 v1.0	139,726,837	20,491	
Melampsora x columbiana Clatskanie v1.0	184,873,583	37,633	
Melampsora americana R15-033-03 v1.0	112,350,849	15,984	Crowell CR et al., 2022
Cronartium comandrae Tree 4 v2.0	153,180,348	13,579	
Cronartium ribicola Cypress_4 v1.0	211,630,673	14,868	
Cronartium quercuum f. sp. fusiforme G11 v1.0	76,567,842	13,903	Pendleton AL et al., 2014
Septobasidium sp. PNB30-8B v1.0	27,514,649	10,893	
Naiadella fluitans ATCC 64713 v1.0	52,667,713	13,270	
Chionosphaera apobasidialis 52639 v1.0	21,830,827	9,875	
Chionosphaera cunicullicola CBS10063 v1.0	14,160,267	7,187	
Agaricostilbum hyphaenes ATCC MYA-4628 v1.0	17,894,169	7,746	
Sporobolomyces linderae CBS 7893 v1.0	18,244,034	9,062	
Mixia osmundae IAM 14324 v1.0	13,634,488	6,903	Toome M et al., 2014
Tritirachium sp. CBS 265.96 v1.0	18,890,683	7,254	
Cystobasidium minutum MCA 4210 v1.0	20,824,934	7,814	Camiolo S et al., 2019
Sporobolomyces gracilis NRRL Y-5504 v1.0	19,998,122	8,038	
Erythrobasidium hasegawianum ATCC 9536 v1.0	40,686,788	9,521	
Naohidea sebacea CBS 8477 (P95) v1.0	20,431,960	10,410	

Figure 6: **List of genomic resources available for the subphylum Pucciniomycotina on the Joint Genome Institute MycoCosm portal** (as of July 12th, 2025). This figure represents a phylogenetic tree of the subphylum Pucciniomycotina as presented on the MycoCosm portal from the Joint Genome Institute (JGI). This portal from the JGI is publicly available, and genomic resources from published genomes (green colour) can be retrieved after acknowledged login; other unpublished genomes can be accessed upon acknowledgement from the genome project investigators. The first column reports the species names (strain and version of the genome). The second column provides the size of the genome assembly in base pairs (bp). The third column represents the number of genes annotated in the corresponding genome. The fourth column indicates the literature related to the genome. Reported genomes were sequenced with various sequencing strategies and technologies depending on related projects. The MycoCosm website regroups projects entirely covered by the JGI, as well as external projects (i.e. personal initiatives). The quality of the available genomes can greatly vary in terms of assembly and annotation. Some species are represented by several strains and/or several haplotypes from the same dikaryon.The phylogeny was established by the JGI, and a detailed explanation of the method used is available on the MycoCosm.

3.4) Characteristics of the rust fungal genomes: large dikaryotic genomes with conserved chromosome numbers

The MycoCosm database currently contains 2,754 genomes of the Kingdom Fungi, of which 697 belong to the Phylum Basidiomycota, 68 to the subphylum Pucciniomycotina and 27 to the order Pucciniales (as of July 12th, 2025). Basidiomycetes (except Pucciniales) available in the MycoCosm exhibit a genome size between 6.6 and 176.7 Mb (Aime et al., 2017; Grigoriev et al., 2014) (Figure 6). In the subphylum Pucciniomycotina, the available genomes in MycoCosm, ranges from 13.6 Mb for *Mixia osmundae* to 1.2 Gb for *P. pachyrizi,* with a mean size of 120 Mb (Gupta et al., 2022; Toome et al., 2014) (figure 6). The available genomes of Pucciniales in Mycocosm range from 60 Mb to 1.2 Gb, whereas other Pucciniomycotina species have an average genome size of approximately 36 Mb (Figure 6). The size of genomes can greatly vary between families in the order Pucciniales (e.g. between *P. pachyrhizi* and species in the Pucciniaceae), as well as between genera within the same taxonomical family (e.g. *M. larici-populina* and *M. lini* in the Melampsoraceae).

Regarding their genome structure, the chromosomal analysis during meiosis in the flax rust *M. lini* revealed two sets of 18 chromosomes (n= 18) (Boehm & Bushnell, 1992). More recently, haplotype-phased assemblies of several Pucciniales species (*Melampsora* and *Puccinia* spp.) confirmed the presence of two haploid chromosome sets in their dikaryotic forms (Liang et al., 2022; Schwessinger et al., 2018; Sperschneider et al., 2025a; Wang et al., 2024). Comparisons between haplotypes within species have then been carried out to assess their degree of similarity. Collinearity analyses, based on SNP detection and structural comparisons, showed a range of outcomes from highly conserved to divergent haplotypes (Henningsen et al., 2025). Across *Pucciniaceae* species such as *P. graminis* f. sp. *tritici*, *P. striiformis* f. sp. *tritici*, and *P. coronata* f. sp. *avenae*, no large chromosomal rearrangements were observed. By contrast,

comparisons between isolates of the same species revealed much higher synteny. For example, haplotypes from *P. graminis* isolates 99KS76A and Ug99 were collinear to 96.7% and 97.4%, respectively, with only a few local rearrangements, such as on chromosome 8 (Guo et al., 2022). Yet, conservation was not uniform across haplotypes. Haplotype “A,” shared between strains 20-1 and Ug99, was nearly identical (>99.7% identity). In contrast, haplotypes “B” and “C” of the same strains diverged strongly, with only 76.2% identity (Li et al., 2019). This divergence between haplotypes has provided key insights into the evolution of rust fungi and the emergence of new epidemics, such as recent wheat rust outbreaks (Li et al., 2019). Overall, Pucciniales display both conserved chromosome numbers and macrosyntenic architecture, while maintaining strong haplotype divergence and lineage-specific rearrangements. The availability of new haplotype phased genomes resolution is a starting point for further research and better understanding their biology and their pathogenicity mechanisms.

3.5) The gene annotation of rust genomes shows typical features of biotrophic fungi and an enrichment in small secreted protein-encoding genes

The obligate biotrophic lifestyle of rust fungi have profoundly influenced the composition and evolution of their gene repertoires. Rust fungal genomes typically encode between 15,000 and 20,000 genes (Aime et al. 2017), while in the other order of Pucciniomycotina, the number varied between 5,536 in *Sporobolomyces roseus* and 20,245 in *Saccosoma farinaceum* (unpublished data). Nonetheless, some contractions of gene families are also known in Pucciniales (Duplessis et al., 2011). For example, gene families involved in nitrogen transport and reduction and in cell wall degradation, affiliated to CAZymes (carbohydrate-active enzymes)

are commonly reduced (Aime et al., 2017, Duplessis et al., 2011; Guerillot et al., 2023; Petre & Duplessis, 2022). Another striking feature is the large proportion of genes with no functional annotation, a pattern that highlights the originality of rust biology and how much remains to be unraveled (Aime et al., 2017; Guerillot et al., 2023). Despite the contractions, some gene families, notably those encoding secreted proteins and oligopeptide membrane transporters, are expanded, reflecting adaptations to host interaction and nutrient acquisition (Duplessis et al., 2011; Guerillot et al., 2023; Petre & Duplessis, 2022). The large gene catalogues in rust fungi genomes are correlated with extensive lineage-specific expansions, inducing multigene families (Duplessis et al. 2011; Guerillot et al., 2023). Comparative studies revealed that rust fungi harbor some of the largest secretomes among plant pathogens, enriched in small secreted proteins (SSPs) that are considered prime candidates for effectors (Petre et al. 2014; Sperschneider et al., 2015; Gupta et al., 2022).

Fungal effectors are most often small, secreted proteins deployed during infection to manipulate host physiology and suppress immune responses (Presti et al., 2015). Despite their importance, effector genes are notoriously difficult to annotate due to their general features: they are typically short, rich in cysteine residues, and highly specific to each fungal species (Petre et al., 2016; Petre & Kamoun, 2014). The lack of homology between established *bona fide* effector genes complicates their identification. To overcome these challenges, various approaches have been developed, including computational prediction tools such as EffectorP, which screen for potential effector candidates for experimental validation (Sperschneider et al., 2018). Detection strategies generally combine multiple criteria, including the detection of secretion signals, subcellular localization, cysteine content, and expression profiling during infection, especially in specialized infection structures, to improve the accuracy of candidate effector identification (Sperschneider et al., 2017). Secretome prediction performed from genome annotation in fungi have revealed clear expansion in rust fungi, highlighting a strong potential for interaction with

their host plants (Lo Presti et al., 2015). Within rust fungi secretomes, thousands of SSPs were predicted across Pucciniales, but only a handful of Avr genes have been molecularly characterized to date (Petre et al. 2014; Lorrain et al. 2019; Duplessis et al., 2021). In *Puccinia graminis* f. sp. *tritici*, for example, AvrSr50, AvrSr27, and AvrSr35 have been identified and shown to interact with corresponding wheat resistance genes, illustrating the gene-for-gene model and the evolutionary arms race between host and pathogen (Salcedo et al., 2017; Sperschneider et al., 2025a). Similarly, in *Melampsora lini, AvrL567*, *AvrP123*, and *AvrM* genes are typical examples of *Avr* genes described in the flax-flax rust model pathosystem (Dodds et al., 2004; Ellis et al., 2007; Duplessis et al., 2021). Recently, several new Avr loci (e.g. *AvrM3* and *AvrN*) were identified with copy number variation between the two haplotypes of *M. lini* (Sperschneider et al., 2025). Using a haplotype-phased genome assembly, the authors were able to resolve these loci within large, gene-rich regions previously inaccessible due to assembly limitations. They also highlighted that some *Avr* candidates were missed or misannotated by automated pipelines, underlining the importance of manual curation in effector gene discovery. Notably, both *AvrM3* and *AvrN* encode unusually large proteins, challenging the traditional focus on small secreted proteins in effector prediction. This work demonstrates how haplotype-resolved, chromosome-scale assemblies are essential to understand the evolution of effectors and *Avr* genes, their genomic context, and their diversification under host selection pressure.

4) Context and Objectives of the Thesis

The Tree-Microbe Interactions Research Department (UMR IAM) investigates the various interactions that occur in forest ecosystems, intending to understand how these interactions contribute to ecosystem sustainability. The UMR IAM project spans a broad spectrum of themes, from molecular mechanisms of tree-fungal symbioses and pathogenic infections, to

comparative genomics studies aimed at deciphering the evolution of trophic modes in fungi and tree immune responses, and up to larger-scale analyses of microbial ecology in forest systems. A key area of interest involves studying fungal virulence factors, which can improve our understanding of pathogenic fungi and support the development of control strategies within these ecosystems.

Building on current knowledge of rust fungi genome biology, this PhD thesis aims to deepen our understanding of genome structure and evolution in *Pucciniales* through a detailed investigation of TEs and their genomic impact. The work combines expert TE annotation across Pucciniales species, the generation of haplotype-phased genome assemblies for *M. larici-populina* and *M. allii-populina*, and an in-depth comparative analysis of three Melampsoraceae genomes. This comparative framework integrates genome architecture, TE comparison at different levels (subphylum, species, haplotype and chromosomes), to provide new insights into the evolutionary forces shaping rust fungal genomes.

The first part of my thesis involved expert annotation of TEs across 12 Pucciniomycotina genomes to establish a reliable TE reference library. This work is presented as a published article in Chapter I, entitled *Ancestral and recent bursts of transposition shaped the massive genomes of plant pathogenic rust fungi*. We revealed multiple waves of TE invasions shaping Pucciniales genome expansions, highlighting both ancient and recent proliferation events predominantly driven by LTR-Gypsy elements.

The second part focused on generating haplotype-phased genome assemblies for *M. larici-populina* and *M. allii-populina*. These high-quality genome assemblies are reported in Chapter II.

The third part initiates the comparative genomic analysis of the phased genomes of *M. larici-populina*, *M. allii-populina*, and *M. lini* (Chapter III), encompassing genome organization

and TE annotation. This axis further explores the three-dimensional genome architecture of the three species, with a focus on chromosome structure and TE dynamics among three closely related species.

Chapter I

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

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

In 2011, *Melampsora larici-populina* was one of the first two rust species to be sequenced, paving the way for the sequencing of other Pucciniales. After the sequencing of other families in the Pucciniales, such as the Pucciniaceae and Phakopsoraceae, it became clear that TEs had massively invaded the genomes of Pucciniales. Strikingly, the genomes of other species belonging to other orders in the Pucciniomycotina subphylum (to which Pucciniales belong) did not exhibit such an expansion. In the last five years, the number of long-read sequencing-based genomes available for Pucciniales has increased. Beyond their high TE content, genomes of rust fungi also present expanded gene catalogues compared to other species represented in the subphylum Pucciniomycotina. These observations led to open questions regarding the potential impact of TE proliferation on gene numbers, since the mobility of TEs can trigger structural rearrangements and gene duplications, leading to the formation of multigene families. These questions were the starting point for setting up a comparative genomic analysis between Pucciniales and other Pucciniomycotina with the aim of better understanding the TE composition in these genomes. The main question in this project was: how and when did TE proliferation happen in the genomes of Pucciniales? To answer this, we proposed hypotheses regarding the timing of proliferation and investigated potential TE control mechanisms across genomes. This work was published as a research article in *BMC Genomics* (doi: https://doi.org/10.1186/s12864-025-11726-3), on which I am the first author and main contributor to data analysis. The rust fungal TE library annotated and curated in this study was used as a basis for the genomic studies at the core of Chapters 2 and 3 of the thesis.

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 12 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 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]. These two major classes are further organized into Orders and Superfamilies (SFs) based on their sequence architecture [19]. Within each class, elements are grouped into Families that share phylogenetic ancestry. Both classes include autonomous elements, which encode the enzymatic machinery necessary for their own mobilization, and non-autonomous elements, which depend on the transposition machinery provided by autonomous counterparts. In Class I TEs, autonomous elements include long terminal repeat (LTR) retrotransposons, which transpose via an RNA intermediate and are flanked by direct terminal repeats, and long interspersed nuclear elements (LINEs), non-LTR retrotransposons that encode reverse transcriptase. Additional representatives of Class I include Penelope-like elements (PLEs), that exhibit reverse transcriptase evolutionarily related to telomerase, and DIRS (Dictyostelium Intermediate Repeat Sequence) elements, which utilize a tyrosine recombinase for genomic integration, setting them apart from canonical retrotransposons. In Class II TEs, autonomous elements are predominantly represented by terminal inverted repeat (TIR) transposons, which mobilize by a transposase, and Helitrons, which propagate through a rolling-circle replication mechanism. Other Class II elements include Cryptons, unique fungal-specific DNA transposons

that also use a tyrosine recombinase for insertion and were first identified in fungal genomes, and Maverick elements, exceptionally large DNA transposons that encode multiple proteins, including a DNA polymerase and an integrase, with notable similarities to viral proteins. Conversely, non-autonomous elements such as Miniature Inverted-repeat Transposable Elements (MITEs), Short Interspersed Nuclear Elements (SINEs), Terminal-repeat Retrotransposons in Miniature (TRIMs), and Large Retrotransposon Derivatives (LARDs), lack the capacity to encode the proteins necessary for independent transposition and instead rely entirely on the enzymatic functions provided by autonomous elements [19]. If the elements described above represent the principal Orders and SFs within the unified classification of TEs, additional TE types, such as Academ and Starships, have also been recently identified, and novel TE types continue to be discovered as genome sequencing expands [20, 21].

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 [22]. 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, 23]. There is also evidence that fungal genomes have co-opted some TEs [24]. For instance, the effector AvrK1 in the TE-rich *B. graminis* f. sp. *hordei* genome is derived directly from a LINE retrotransposon [25]. 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 [26]. 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 [27, 28]. TEs are frequently associated with heterochromatin histone marks and cytosine DNA methylation (5mC) [29]. 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 [27, 29, 30]. To date, RIP has not been functionally characterized in Basidiomycetes, although RIP-like mutational signatures have been reported in the genomes of some smut fungi [31]. Notably, the canonical RIP-associated DNA methyltransferases, DIM2 and RID, are absent from the gene repertoires of multiple Basidiomycete species, including the Pucciniales [32]. 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 (RdRP), which amplifies siRNAs to sustain the silencing response [33]. This pathway targets TE transcripts for degradation or chromatin modifications, thereby repressing TE expression and preventing transposition [34]. 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 [27, 29, 33].

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 12 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 12 Pucciniomycotina species with reference genomes comprising eight 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 nine 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 26 Mb (*Microbotryum lychnidis-dioicae*) (Table S1; Fig. 1B). Assemblies included either primary scaffolds from unphased builds or fully separated haplotype-phased scaffolds. BUSCO (basidiomycota_odb10) completeness was uniformly high (≤ 10 % missing for all, except *Leucosporidiella creatinivora* at 10.6 %), and most showed low duplication (≤ 10 % duplicated BUSCOs). However, the three Pucciniales assemblies *M. allii-populina* (22 %),

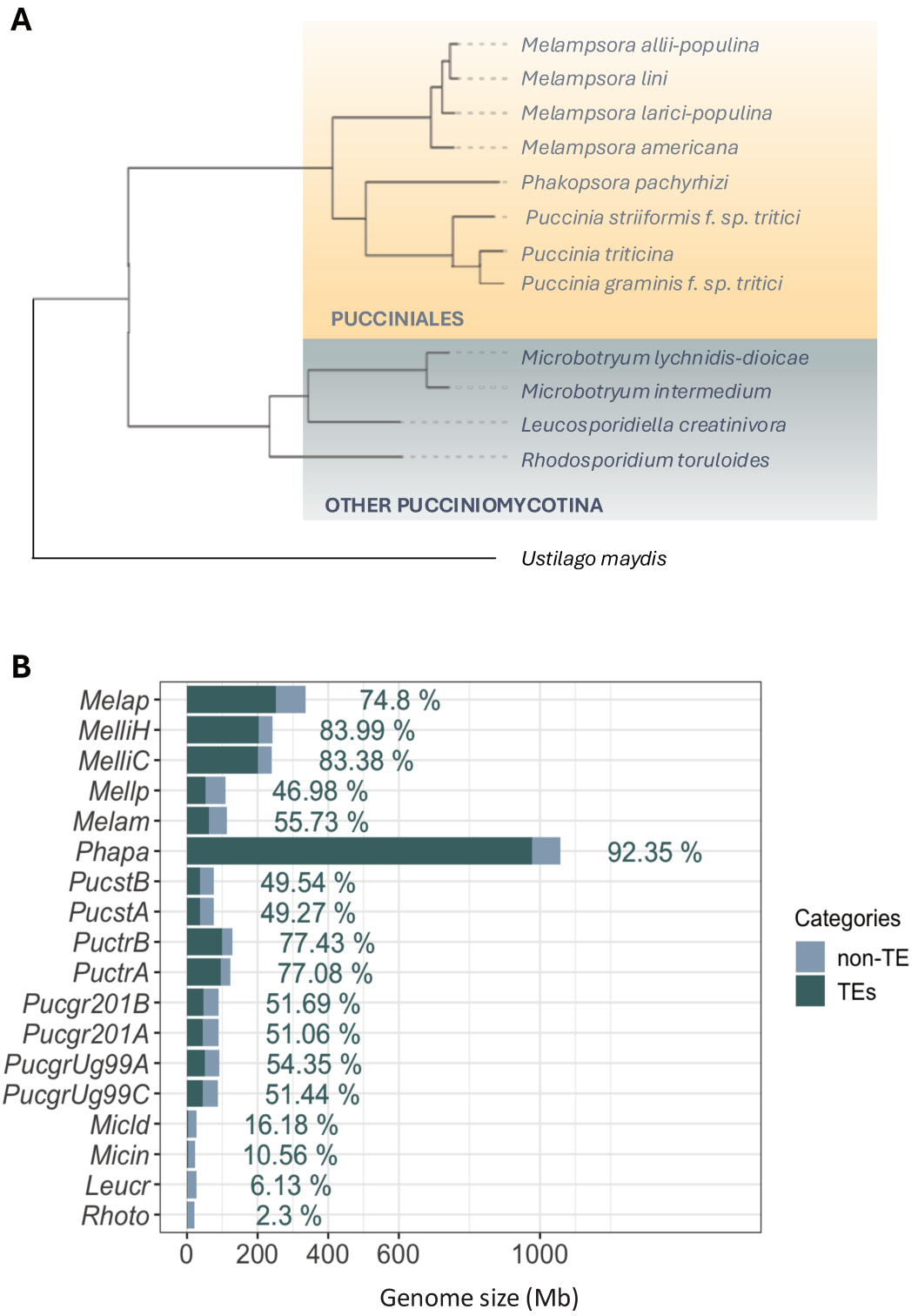


Figure 1: **The phylogenetic tree and transposable element content in the Pucciniomycotina**. A) phylogenetic tree of the 12 selected Pucciniomycotina species based on a single-copy orthogroup from Orthofinder v2.4.0 and visualized using iTOL (https://itol.embl.de/). The Pucciniales are represented in 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), PucgtrUg99 (Puccinia graminis f. sp. tritici Ug99), Pucgtr201 (Puccinia graminis f. sp. tritici 20-1), Micld (Microbotryum lychnidis-dioicae), Micin (Microbotryum intermedium), Leucr (Leucosporidiella creatinivora), Rhoto (Rhodosporidium toruloides).

M. americana (44 %) and *P. pachyrhizi* (36 %), exhibited elevated BUSCO duplication, consistent with uncollapsed haplotypes in these dikaryotic, repeat-rich genomes. To mitigate potential bias for comparisons, phylogenetic analyses were made on single-copy orthogroups and performed TE comparisons on a global genomic scale.

We first compared gene and TE contents and their contribution to the genome inflation of Pucciniales. Pucciniales genomes harbor substantially more genes ranging from 15,984 genes in *Melampsora americana* to 19,618 genes in *P. pachyrhizi,* compared to the other Pucciniomycotina species with 7,819 genes in *M. lychnidis-dioicae* to 9,854 in *L. creatinivora* (Table S1). Although gene coverage showed a slight but significant positive relationship with genome size ($r^2 = 0.42$, $p < 0.05$), TE content appeared as the primary driver of genome inflation in Pucciniales, correlating very strongly with genome size ($r^2 = 0.91$, $p < 0.001$; FigS1). In addition, the TE coverage in Pucciniales genomes ranges from 47% to 92.3% in *M. larici-populina* and *P. pachyrhizi*, respectively (Fig. 1B; Table S2). 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 83.4-84% of the 239.9-242.3 Mb genome of *Melampsora lini C* and *M. lini H,* respectively. In the Pucciniaceae, TEs covered from 49.3-49.5% of the 75.6-75.9 Mb genome of *Puccinia striiformis* f. sp. *tritici* A and B*,* respectively, to 77.1-77.4% of the 123.9-121.6 Mb genome of *Puccinia triticina* A and B*,* respectively (Fig. 1B). In comparison, the TE coverage in the other Pucciniomycotina ranges between 2.3% in *R. toruloides* and 16.2% in *M. lychnidis-dioicae*. Finally, haplotype-phased Pucciniales assemblies display nearly identical TE proportions between phases, further validating our global-TE comparison approach. Together, these results highlight that TE accumulation, more than gene number, underlies the extensive genome expansion in Pucciniales relative to other Pucciniomycotina.

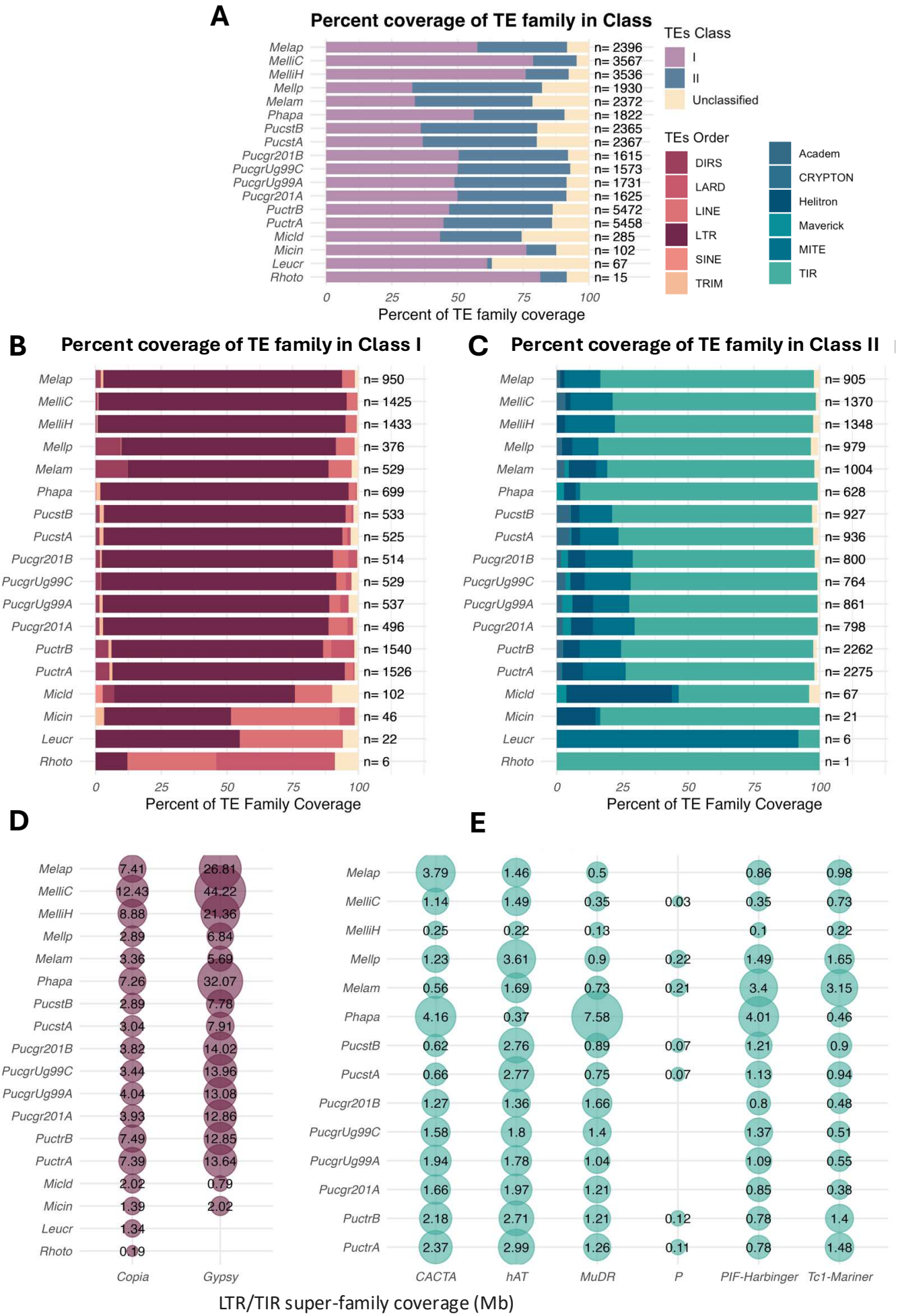


Figure 2: **Composition of transposable elements in Pucciniomycotina genomes.** The stacked bar plots represent the percentage of transposable elements (TEs) annotated at the A) Class level, B) Class I Order level, and C) Class II Order level for each Pucciniomycotina genome. The n values indicate the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

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 39, 834 TE consensus sequences, constituting the Pucciniomycotina TE library. TE annotation before curation detected 20, 958 (52.6%) classified TE consensus sequences, 16,811 (42.2%) unclassified TE consensus sequences, and 2,065 (5.2%) TE consensus sequences with ambiguous annotations (or "conflicts"), which occur when the classifier cannot assign a TE to a single category due to nested or overlapping elements (Fig. S2). After the first round of manual curation, we reduced the number of ambiguous annotations to 651 (1.6%) TE consensus sequences but could not resolve unclassified annotations. Therefore, we used MCHelper [35] and improved the annotation of ~16% (2, 700) of previously unclassified TE consensus sequences (Fig. S2). 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 ambiguous annotations. Our final curated Pucciniomycotina TE library comprised 28,868 (72.5%) classified TE consensus sequences and 10,966 (27.5%) 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 51.7% (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 eight Pucciniales species.

The TE Orders content of the Pucciniales genomes predominantly comprises LTR retrotransposons and TIR DNA transposons. LTRs represent 12.5% of the genome in *M. larici-populina* (13.1 Mb) to 63.0% in *M. lini* haplotype C (151.1 Mb), respectively (Fig. 2B-C; Table S3). At the LTR Superfamily level, we mostly found Copia and Gypsy elements, representing 2.9% of genome (3.2 Mb of LTR-Copia in *M. larici-populina*) to 12.7% (30.5 Mb in *M. lini* haplotype C) and 5.5% (6.2 Mb in *M. americana*) to 48.2% (115.6 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 3.3% of the genome of *M. intermedium* (0.77 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 Superfamily 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 analyzed genome-wide TE divergence profiles using Kimura 2-parameter (K2P) substitution rates. In this framework, low K2P values signify recent TE insertions with minimal divergence from the consensus, whereas higher values denote ancient insertions indicative of long-term genomic retention and mutation accumulation. Our comparative landscape analysis across 12 Pucciniomycotina species revealed species-specific TE activity patterns (Fig. 3A–P). Among non-Pucciniales species, TE proliferation appears minimal. For instance, *R. toruloides* and *L. creatinivorum* exhibit low recent TE activity, with less than 3% of their genomic content composed of insertions with K2P < 0.05. Conversely, *M. intermedium* and *M. lychnidis-dioicae* demonstrate moderate levels of recent activity, with 8.5% and 10.5% of their genomes, respectively, reflecting young TE insertions (Fig. 3A–D). Within Pucciniales, two distinct TE proliferation patterns emerge. Species from the Melampsoraceae family show evidence of dual TE expansion waves with recent bursts (K2P < 0.05) accompanied by older insertions peaking between 0.4 and 0.6 K2P (Fig. 3E–I). Notably, *M. lini* presents a prominent accumulation of ancient LTR elements, indicating substantial historical activity. In contrast, members of the Pucciniaceae exhibit a predominance of recent insertions with a marked peak at K2P < 0.05 and a flatter distribution at higher divergence values, suggesting reduced retention of older TE sequences. In addition, 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

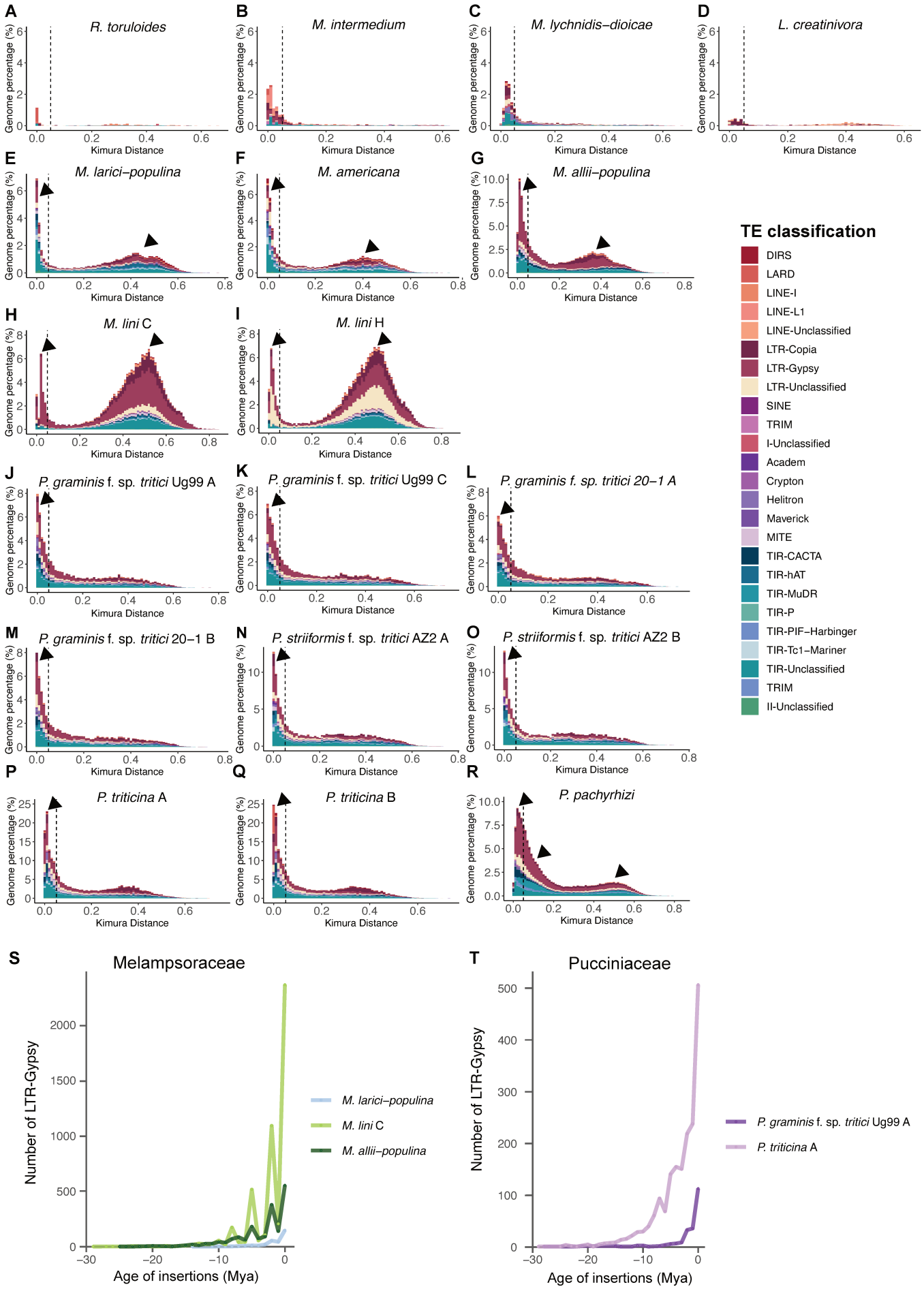


Figure 3: **Transposable element insertion ages.** A-R Kimura distances across the TE landscapes of Pucciniomycotina genomes. Each distribution shows the proportion of the genome (y-axis) composed of copies exhibiting varying degrees of divergence (x-axis). Arrows indicate TE mobilization events. S-T) Number of LTR-Gypsy elements accumulated through time (million years ago - Mya), based on insertion age for the S) Melampsoraceae species and T) the Pucciniaceae species.

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 80.6% (*R. toruloides*) to 90.6% (*M. lychnidis dioicae*). We observed 13.0--66.7% divergent (*M. lychnidis-dioicae* and *R. toruloides*, respectively), 33.3-76.5% intermediate (*R. toruloides* and *M. intermedium*, respectively), and 0.0-16.5% conserved (*R. toruloides* and *M. lychnidis-dioicae*, respectively) TE consensus sequences in the other Pucciniomycotina (Table S5). Within the Pucciniales, TE consensus sequences similarity profiles revealed 30.0-66.2% divergent (*P. pachyrhizi* and *M. larici-populina*, respectively), 27.4-65.6% intermediate (*M. americana* and *P. pachyrhizi*, respectively), and 2.7-7% conserved TE consensus sequences (*M. allii-populina* and *P. triticina*, respectively) (Table S6). Among the Pucciniales genomes, we observed that LTRs and TIRs compose most of each relative TE age category (Table S6). LTR elements represent 17.4-33.6% of the divergent TEs (in *M. americana* and *M. allii-populina*, respectively), while TIR elements represent 26.1-38.0% of the divergent TEs (in *P. pachyrhizi* and *M. larici-populina*, respectively). In the conserved TEs, LTR elements represent 7.2-34.4% (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) (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.

Based on sequence similarity among all TE consensus sequences, we 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 Pucciniales species. However, we identified

three clusters that are specific to the Melampsoraceae and one that are specific to the Pucciniaceae (Fig. S3). 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 one LTR cluster (Fig. S3; 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. S3). 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. S3). 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. S4). 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 (Fig. S4). 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. 3S-T; Fig. S4). 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. 3S; Fig. S4). The second profile, characteristic of Phakopsoraceae and Pucciniaceae species, demonstrates a more gradual accumulation of LTRs over time (Fig. 3S-T; Fig. S4). 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 [27]. Using the composite RIP index in TE sequences and in genome-wide 1-kb sliding windows, we quantified RIP-like signatures in TE family consensus sequences in the Pucciniales and the other Pucciniomycotina genomes. In all Pucciniomycotina genomes we detected less than 0.5% of the genomes containing RIP-like signatures (Table S9). 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. S5).

Finally, we investigated the presence or absence of genes involved in the RNAi pathway, namely Argonaute, Dicer, and RNA-dependent RNA polymerase RdRP. 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. S6). Similarly, we identified 17-60 (and *M. lini*) RdRP-like genes in the Pucciniales and 12-15 RdRP-like genes in the other Pucciniomycotina genomes (Fig. S6). 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. S5). 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 43,847 orthogroups (Table S10). Of these, 3067 core orthogroups (17,1758 genes) were shared across all genomes, including *U. maydis*, while 16, 585 orthogroups (85,786 genes) were specific to Pucciniomycotina (Table S10). We observed substantial variation in species-specific orthogroups ranging from 68 in *M. intermedium* to 1310 in *P. triticina.* We identified 23 orthogroups in contraction and 380 orthogroups in expansion at the separation of the Pucciniales and the other Pucciniomycotina (Fig S7). The orthogroups in contraction in the Pucciniales comprise a total of 483 genes, of which 19.5% (94 genes) are of unknown functions (Table S11; Fig. S8). The five most represented functions within the remaining orthogroups in

contraction in the Pucciniales, include 62 genes related to “Transcription”, 52 genes related to “Signal transduction mechanisms”, 52 genes with function related to “Amino acid transport and metabolism”, 25 genes related to “Posttranslational modification, protein turnover, chaperones”, 23 genes related to ““Secondary metabolites biosynthesis, transport and catabolism” and 21 genes related to “Chromatin structure and dynamics”(Fig S8; Table S11). However, we identified only 22 orthogroups with a significant accelerated gene gain/loss between the Pucciniales and the other Pucciniomycotina genomes and all the 22 orthogroups are in expansion in the Pucciniales (p-value < 0.05; Fig S9). These results show that Pucciniales underwent extensive gene losses compared to other Pucciniomycotina species. 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 within the 'chromatin structure and dynamics' category or other, as yet unidentified, functions.

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 12 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 ancient 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 *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 one (LTR-Copia) 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 [36]. Likewise, in large, complex plant genomes such as wheat and maize, retrotransposons are retained across extended timescales [37, 38]. 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 are a common feature in other TE-rich fungal genomes, such as those of

B. graminis and *Entomophthoraceae* spp. [39, 40]. 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 [41]. 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 [40]. 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 [36, 42]. 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 [29, 43]. In addition, we found the DNA methyltransferases DNMT1 and DNMT5 in all 12 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 [29]. In addition, Bewick et al. (2019) [29] 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 12 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. A comparable phenomenon occurs in the arbuscular mycorrhizal fungus *Gigaspora margarita*, where multiple AGO-like copies lack conserved PIWI catalytic residues and may have diverged to non-canonical roles or become pseudogenes [44]. 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 [45]. In *Rhizophagus irregularis*, several AGO loci localize adjacent to TE insertions and transcriptional profiling has shown that AGO genes are highly expressed in spores but downregulated during plant colonization, coincident with transient TE activation [46, 47]. 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 [48]. 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 ambiguous 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 12 Pucciniomycotina species 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*), four genomes from the family Pucciniaceae (*P. graminis* f. sp. *tritici Ug99 and 21-0*, *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 selected genomes with either haplotype-phased assemblies, long-read assemblies, or assemblies based on genetic maps. We assessed genome assembly completeness using BUSCO v5.4.7 with the basidiomycota_odb10 marker set, which provided a measure of the quality 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 “conflicts” ambiguous TE annotations (i.e. the classifier cannot assign a TE to a single category) 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 [35] 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 all 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 [49]. 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. The curated library of TE consensus sequences has been deposited in the Dfam database (https://dfam.org/home).

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) [50]. 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%. In addition, we estimated the sequence divergence of TE copies relative to their consensus sequences, we computed Kimura 2-parameter distances. For each annotated TE copy, the aligned region was extracted and compared to its corresponding consensus sequence. K2P distances were calculated based on the frequency of transitions and transversions and correcting for multiple substitutions following the approach described by Baril et al. (2024)[51] and adapted to our TE annotations (https://github.com/Raistrawby/TE_Rust/tree/main). Briefly, for each TE copy, the corresponding genomic sequence was retrieved based on genomic coordinates and strand orientation. Each extracted sequence was then pairwise aligned to its cognate consensus sequence using

matcher [52] and the Kimura 2-parameter distance [53] was computed from the alignment. TE copies with incomplete or failed alignments were excluded from downstream analyses. The resulting Kimura distances were used as a proxy for the relative age of TE insertions.

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 [54] 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 [55] to calculate Kimura's P distances [53]. 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 [53].

Gene family annotations and evolution

We predicted orthologs between the Pucciniomycotina genomes and the reference *U. maydis* genome [25], 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)[56] on the proteomes. We chose one haplotypic proteome for species with haplotype-resolved annotation. 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 genome) and infer the tree with *U. maydis* as an outgroup with Raxml

and 500 bootstraps [57]. 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 [58].

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

We used EggNOG-mapper v2.1.12 [59] 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 Pucciniomycotina genomes, we extracted the genes with corresponding PFAM domains. For DNA methyltransferases, we selected genes with PF00145 (C-5 cytosine-specific DNA methylase) and used orthogroups and BLASTp searches to differentiate DNMT classes. 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 [27] and GC content in TE family sequences of all Pucciniomycotina genomes, as described in Feurtey et al.

(2023)[60]. 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 [43]. In addition, we performed a 1kb genome-wide sliding window RIP-like signature detection using TheRIPper [61].

Data availability

The accession numbers and links to publicly available genome assemblies and gene annotations used in this study can be retrieved from Table S1. The TE annotations and gene functional predictions generated in this study, are openly available at https://doi.org/10.5281/zenodo.14616359.

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.

Funding Declaration

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 Ambizione 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. Kojima KK. AcademH, a lineage of Academ DNA transposons encoding helicase found in animals and fungi. Mob DNA. 2020;11:15.

21. Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, et al. Giant Starship Elements Mobilize Accessory Genes in Fungal Genomes. Mol Biol Evol. 2022;39:msac109.

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

23. 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.

24. 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.

25. 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.

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

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

28. 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.

29. 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.

30. 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.

31. Duhamel M, Hood ME, Rodríguez de la Vega RC, Giraud T. Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi. Nat Commun. 2023;14:5692.

32. 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.

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

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

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

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

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

47. 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;31:2290–302.

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

49. 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.

50. 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.

51. Baril T, Galbraith J, Hayward A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. Mol Biol Evol. 2024;41:msae068.

52. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet TIG. 2000;16:276–7.

53. 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.

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

55. 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.

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

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

58. 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.

59. 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.

60. 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.

61. 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.

62. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, et al. Widespread adenine N6-methylation of active genes in fungi. Nat Genet. 2017;49:964–8.

63. Crowell CR, Wilkerson DG, Beckauri M, Cala AR, McMullen PW, Mondo S, et al. The Melampsora americana Population on Salix purpurea in the Great Lakes Region Is Highly Diverse with a Contributory Influence of Clonality. Phytopathology. 2022;112:907–16.

64. Persoons A, Maupetit A, Louet C, Andrieux A, Lipzen A, Barry KW, et al. Genomic Signatures of a Major Adaptive Event in the Pathogenic Fungus Melampsora larici-populina. Genome Biol Evol. 2022;14:evab279.

65. Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, et al. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc Natl Acad Sci U S A. 2017;114:7067–72.

66. Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, et al. Sex and parasites: genomic and transcriptomic analysis of Microbotryum lychnidis-dioicae, the biotrophic and plant-castrating anther smut fungus. BMC Genomics. 2015;16:461.

67. Gupta YK, Marcelino-Guimarães FC, Lorrain C, Farmer A, Haridas S, Ferreira EGC, et al. Major proliferation of transposable elements shaped the genome of the soybean rust pathogen Phakopsora pachyrhizi. Nat Commun. 2023;14:1835.

68. Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, et al. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. Nat Commun. 2019;10:5068.

69. Wang J, Xu Y, Peng Y, Wang Y, Kang Z, Zhao J. A fully haplotype-resolved and nearly gap-free genome assembly of wheat stripe rust fungus. Sci Data. 2024;11:508.

70. Holden S, Bakkeren G, Hubensky J, Bamrah R, Abbasi M, Qutob D, et al. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. BMC Biol. 2023;21:233.

71. Coradetti ST, Pinel D, Geiselman GM, Ito M, Mondo SJ, Reilly MC, et al. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. eLife. 2018;7:e32110.

72. SPERSCHNEIDER J, CHEN J, ANDERSON C, MORIN E, ZHANG X, LEWIS D, HENNINGSEN E, GRIGORIEV IV, RATHJEN J, JONES DA, DUPLESSIS S, DODDS PN. A chromosome-scale genome assembly of the flax rust fungus reveals the two unusually large effector proteins, AvrM3 and AvrN. submitted.

FIGURES LEGENDS

Fig. 1: **The phylogenetic tree and transposable element content in the Pucciniomycotina.** A) phylogenetic tree of the 12 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*), PucgtrUg99 (*Puccinia graminis* f. sp. *tritici* Ug99), Pucgtr201 (*Puccinia graminis* f. sp. *tritici* 20-1), Micld (*Microbotryum lychnidis-dioicae*), Micin (*Microbotryum intermedium*), Leucr (*Leucosporidiella 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 Superfamily (SF) level for E) Class I elements (purple) and F) Class II elements (blue).

Fig. 3: **Transposable element insertion ages.** A-R Kimura distances across the TE landscapes of Pucciniomycotina genomes. Each distribution shows the proportion of the

genome (y-axis) composed of copies exhibiting varying degrees of divergence (x-axis). Arrows indicate TE mobilization events. S-T) Number of LTR-Gypsy elements accumulated through time (million years ago - Mya), based on insertion age for the S) Melampsoraceae species and T) the Pucciniaceae species.

Fig. S1: **Association between genome size and gene/TE content.** Genome sizes (Mb) were plotted against genes or TE coverage for each Pucciniomycotina species. The lines represent the linear regression.

Fig. S2: **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. S3: **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. S4: **Number of LTR-Gypsy elements accumulated through time** (million years ago - Mya), based on insertion age for each genome.

Fig. S5: **TE controls involved in methylations are detected in all species**. The number of genes involved in TE control corresponds to DNMT1 and DNMT5.

Fig. S6: **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. S7: **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. S8: **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. S9: **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** [62–72].

Table S2: **TE annotations at the Class level for the 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 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 Superfamily level for the 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**. Genome-wide RIP signature detection using TheRIPper [61]. 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)[60].

Table S10: **Orthogroups summary metrics of the selected Pucciniomycotina genomes.** Genomes 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*), PucgtrUg99 (*Puccinia graminis* f. sp. *tritici* Ug99), Micld (*Microbotryum lychnidis-dioicae*), Micin (*Microbotryum intermedium*), Leucr (*Leucosporidiella creatinivora*), Rhoto (*Rhodosporidium toruloides*).

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

Supplemental figures

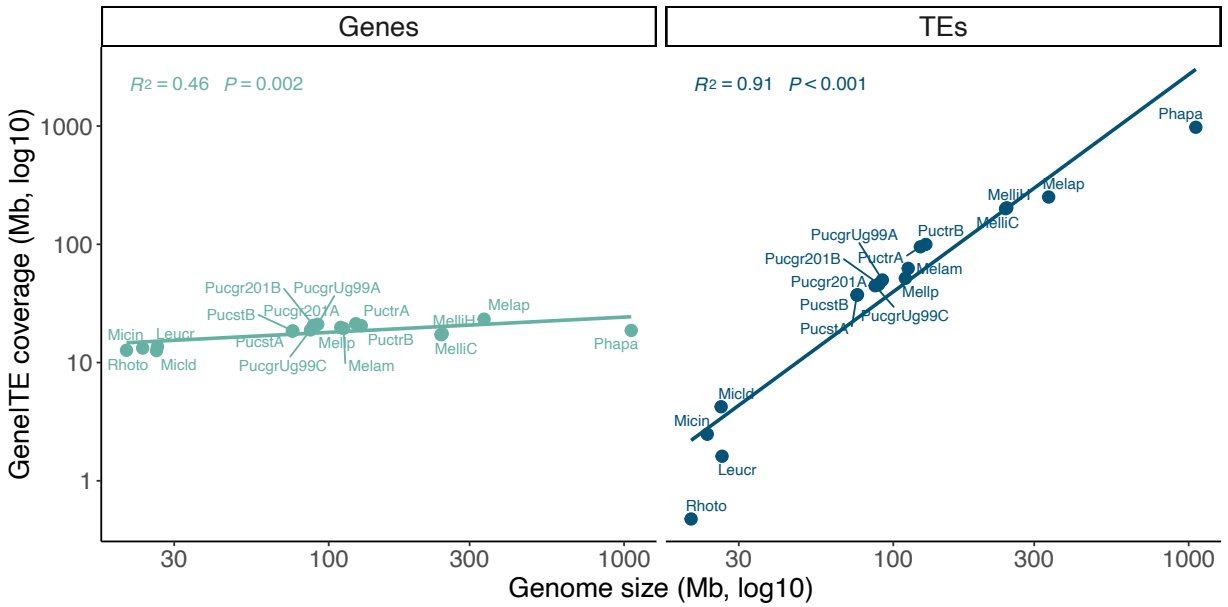


Figure S1: **Association between genome size and gene/TE content.** Genome sizes (Mb) were plotted against genes or TE coverage for each Pucciniomycotina species. The lines represent the linear regression.

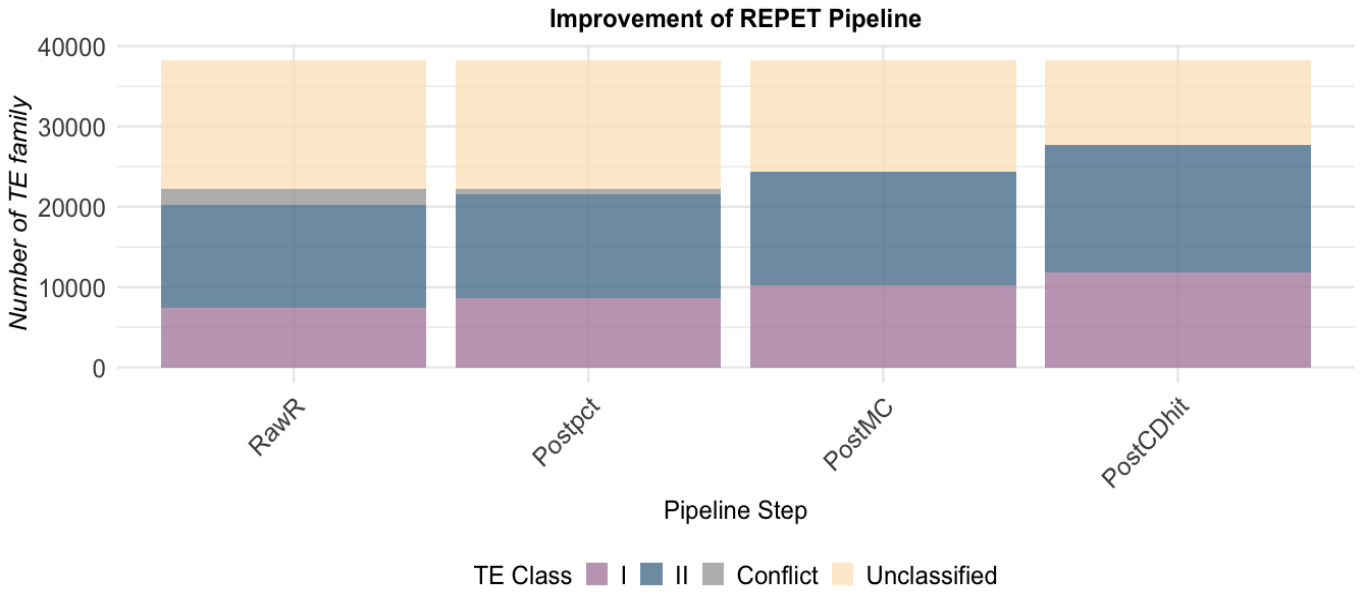


Figure S2: **Improvement of TE annotations through manual curation**. The proportion of annotated TE families after REPET detection and annotation (RawR), after the first round of curation (Postpct), after the second round of curation with MCHelper (postMC), after a third round of curation with Pucciniomycotina-guided clustering (PostCDhit).

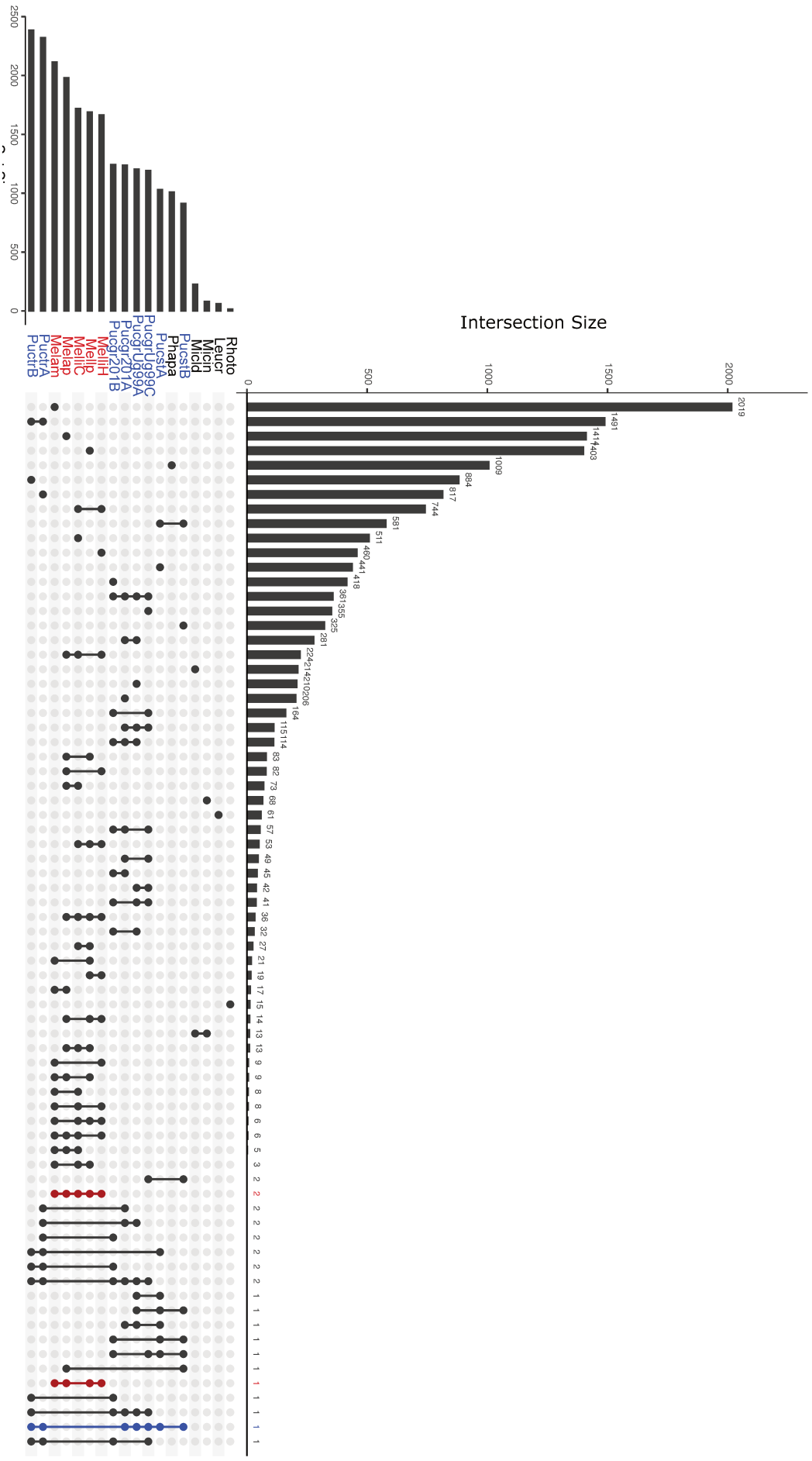


Figure S3: **Shared and specific TE families between the Pucciniomycotina species.** Numbers of TE families shared among the 14 genomes based on CDHIT clustering. The bars on the left represent the total number of genes per species. The clusters shared among Pucciniaceae are colored in blue and the clusters shared among Melampsoraceae are colored in red.

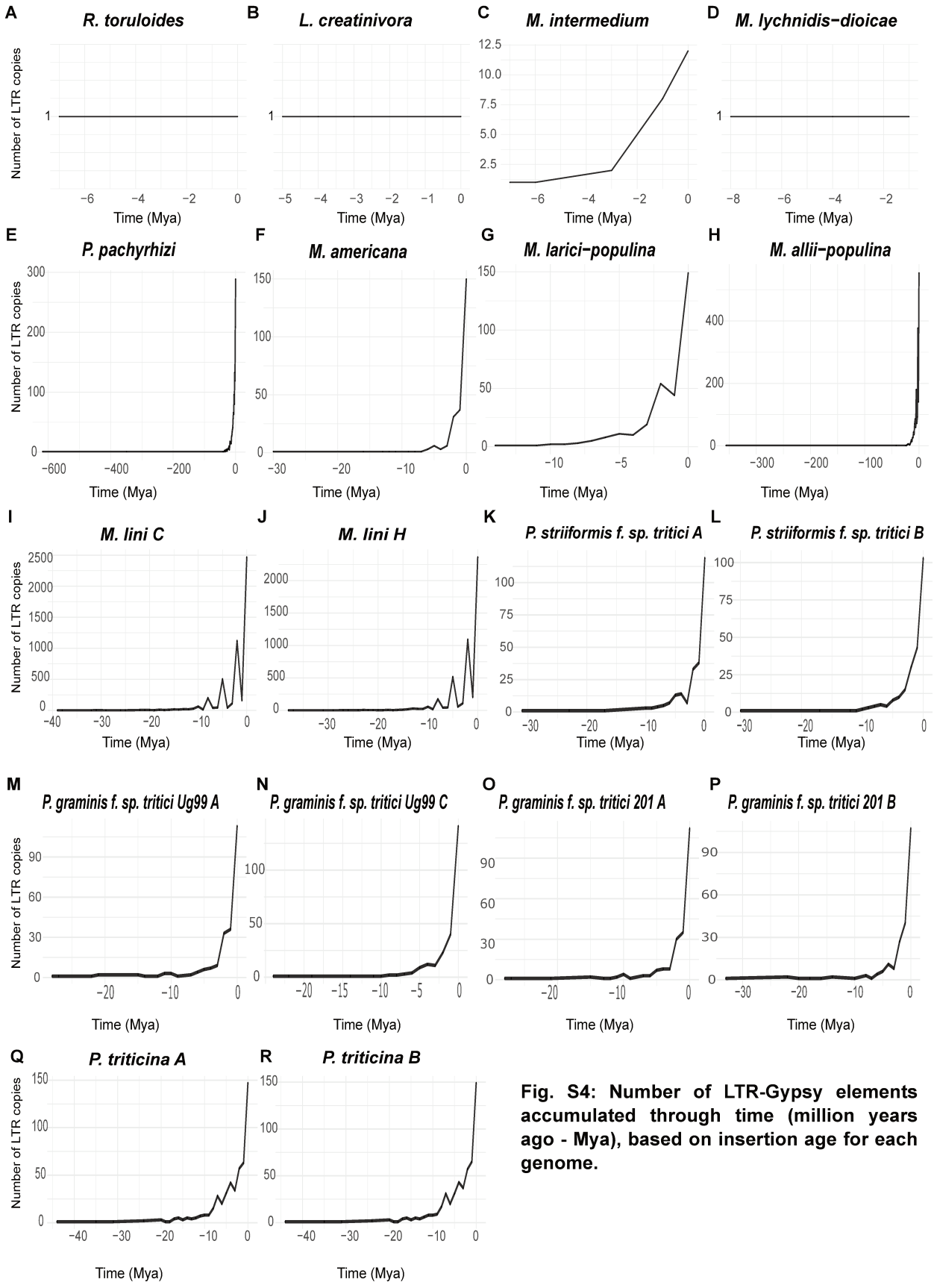


Fig. S4: Number of LTR-Gypsy elements accumulated through time (million years ago - Mya), based on insertion age for each genome.

Figure S4: **Number of LTR-Gypsy elements accumulated through time** (million years ago - Mya), based on insertion age for each genome.

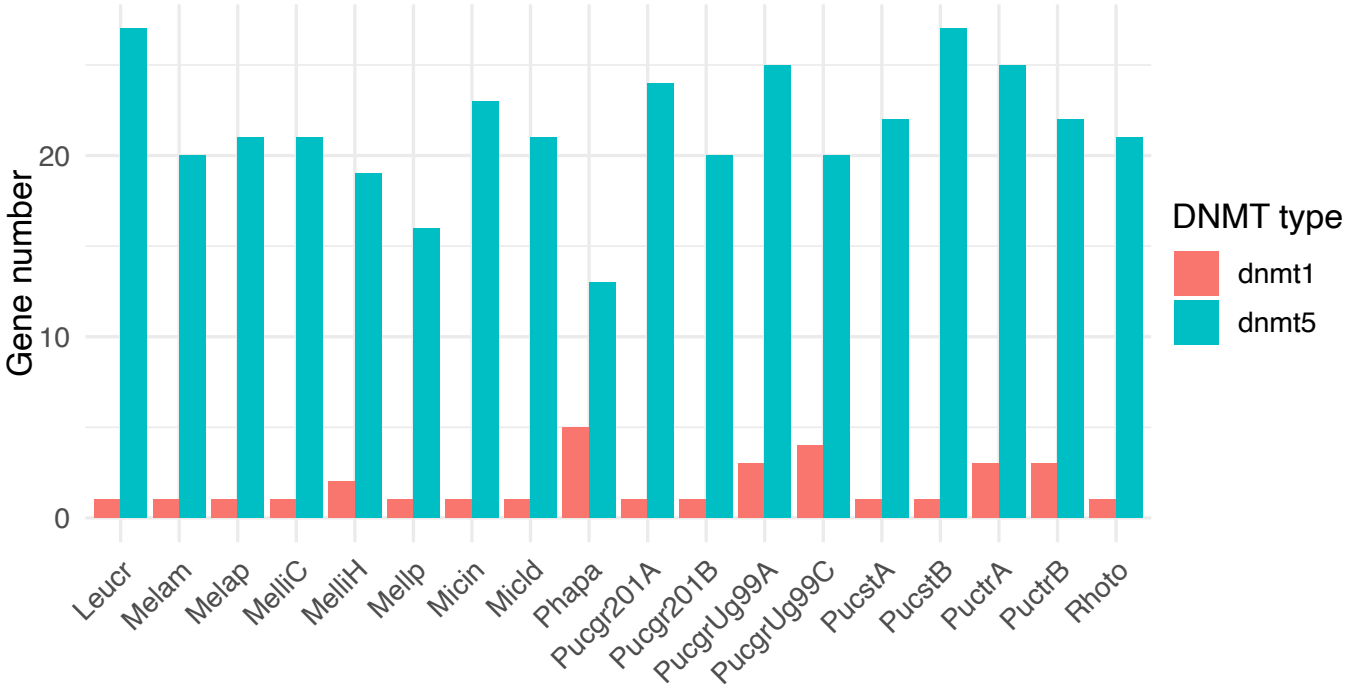


Figure S5: **TE controls involved in methylations are detected in all species**. The number of genes involved in TE control corresponds to DNMT1 and DNMT5.

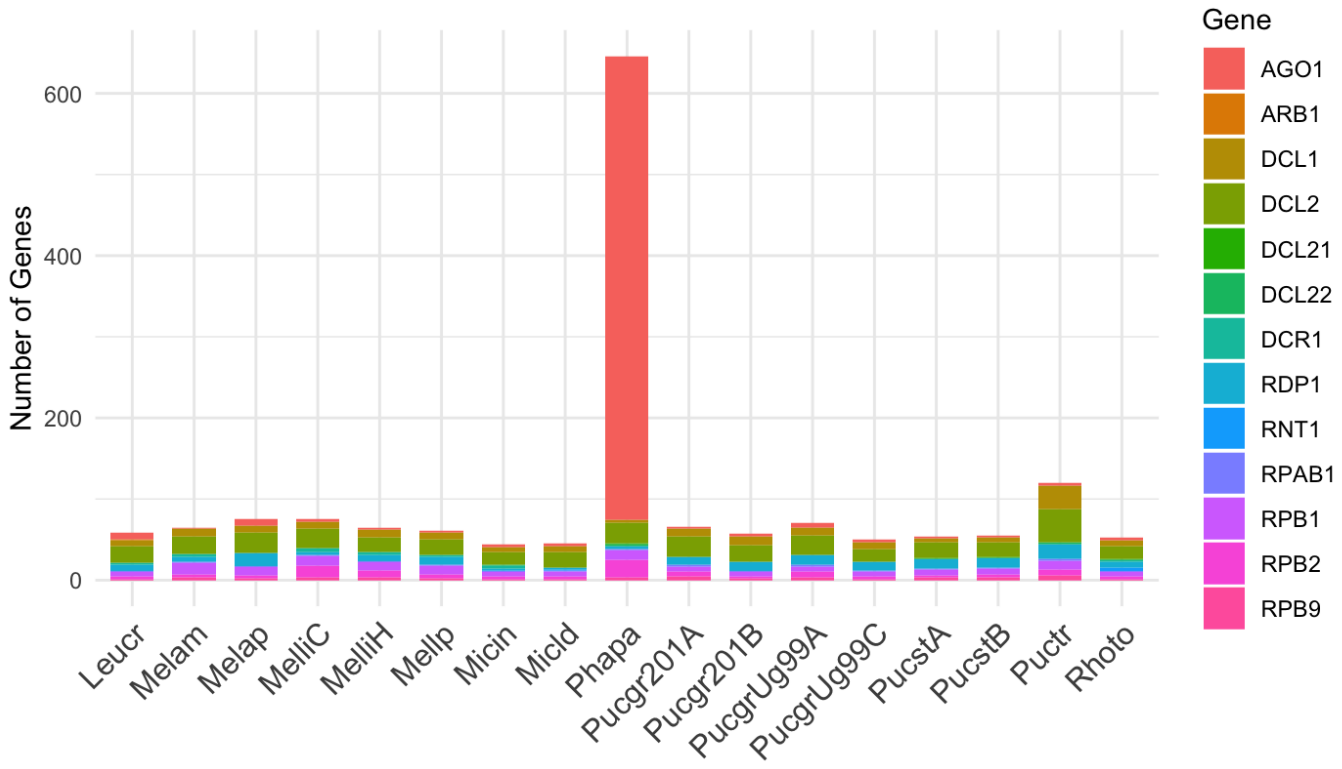


Figure S6: **TE controls involved in RNA-induced silencing complex are detected for all genomes.** The number of genes involved in TE control corresponds to the RNAi pathway: Dicer-like (DCL; DCR), RNA-dependent RNA polymerase (RDP; RNT; RPAB; RPB), Argonaute (AGO; ARB) for each genome.

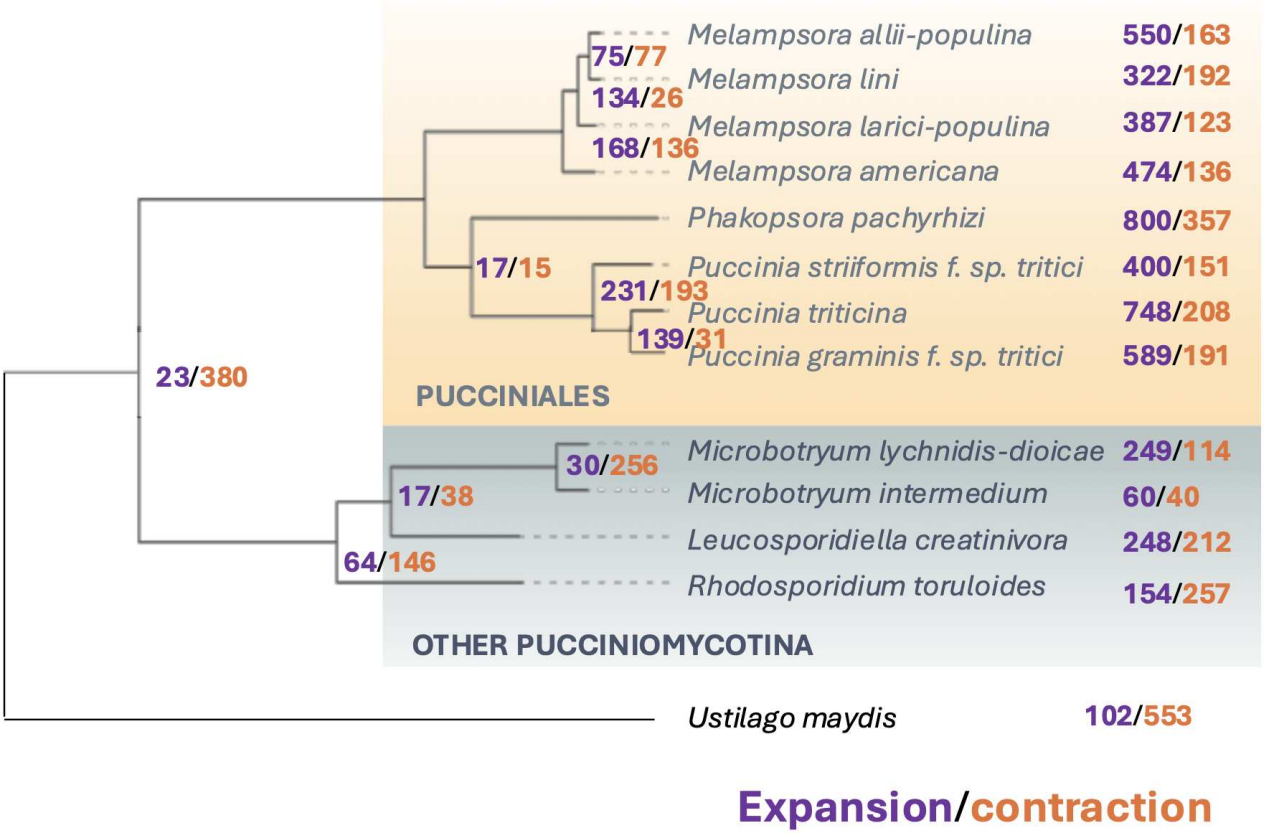


Figure S7: **Contraction and Expansion of multigene families in Pucciniomycotina.**The phylogenetic tree of the selected Pucciniomycotina species based on single-copy orthogroups from Orthofinder v2.4.0 and infer with Raxml (500 bootstraps) and visualized using iTOL, rooted with *Ustilago maydis.* Numbers in purple correspond to the number of orthogroups in contractions at each branch or node and yellow at expansions.

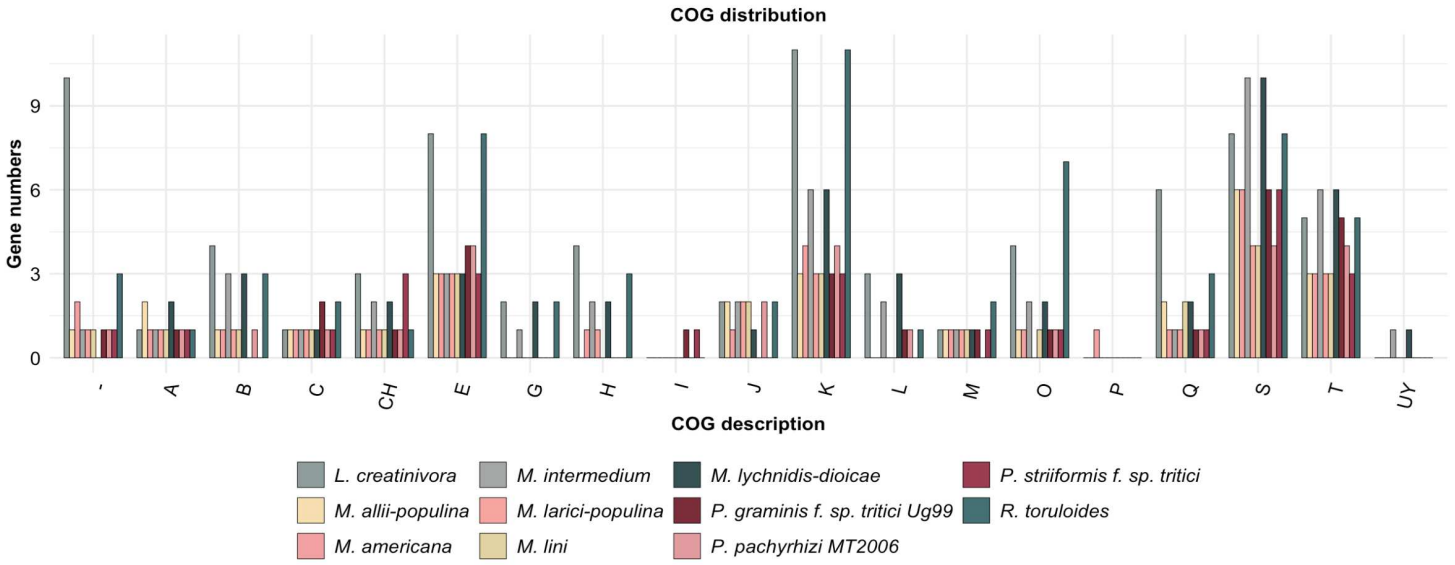


Figure S8: **Contraction of Multigene Families in Pucciniales.** Functional annotation of multigene families in contraction in Pucciniales. Genes are re-annotated by EggNOG-mapper v2.1.12. A threshold is set to kept only species with at least 20 proteins per COG category. COG categories correspond to: "-": 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.

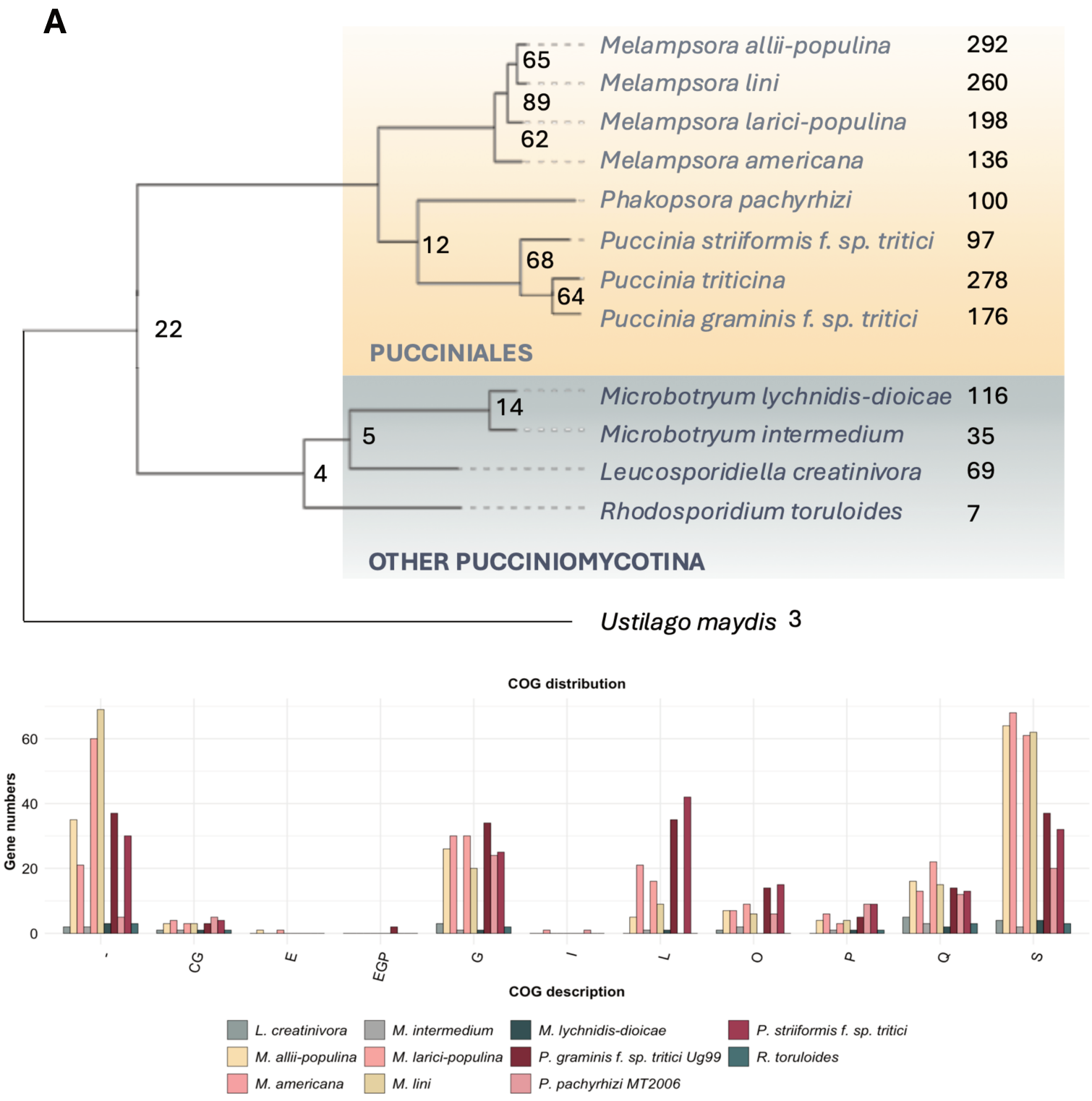


Figure S9: **Significant contraction and expansion of multigene families in Pucciniomycotina.** A) The phylogenetic tree of the selected Pucciniomycotina species based on a single-copy orthogroup from Orthofinder v2.4.0 and visualized using iTOL, rooted with *Ustilago maydis.* Numbers in black correspond 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.

Chapter II

BACKGROUND

Recent advances in long-read sequencing technologies and chromatin conformation capture approaches (HiC) have enabled the production of fully phased, telomere-to-telomere (T2T) genome assemblies in rust fungi. Cereal rust fungi were primary targets for the resolution of such T2T genomes in the order Pucciniales. More recently, the T2T genome of the model rust fungus *M. lini* was produced, opening interesting perspectives for the production of genomes of greater quality for poplar rust fungi. In this chapter, we present chromosome-scale, haplotype-phased genome assemblies and TEs annotations for two *Melampsora* species pathogenic on poplars. In Part 1 of this chapter, we describe the new T2T assemblies for *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07, which were both generated using PacBio HiFi long-read sequencing and Hi-C. These assemblies offer the first high-quality, phased genomes for poplar rust fungi and provide critical resources for comparative analyses of genome architecture and TE dynamics. In Part 2 of this chapter, we report the assembly and TE annotation of *M. larici-populina* isolate 08KE26, which belongs to an asexual lineage under study in a transversal project of the UMR IAM and in the frame of Ammar Abdalrahem's thesis project (supervision of Drs. Fabien Halkett, Pascal Frey, and Sébastien Duplessis). The collaborative work reported in this part of the chapter will be the basis of a joint article with co-authorship between the two PhD students. I only present here the part of the work that I personally realized. It should be noted that at the time of thesis writing, gene annotations were still ongoing for these genomes at the US DoE JGI, and related metrics could not be included in the chapter.

Chapter II - Part 1
Haplotype-phased assemblies of the two poplar rust fungi *Melampsora larici-populina* and *Melampsora allii-populina*

INTRODUCTION

Rust fungi have remarkable biological and genomic complexity, particularly in their life cycles and chromosomal organization (Henningsen et al., 2025). The rust fungi are part of the Pucciniales order, which represents the largest group of obligate biotrophic fungal pathogens, containing more than 8,000 species (Aime et al., 2018). The Pucciniales can infect a wide range of plants, including important crop species (e.g. wheat, soybean, oat) (Henningsen et al., 2025). Many rust fungal species follow a heteroecious and macrocyclic life cycle, alternating between two taxonomically distinct host plants and producing multiple spore types (Lorrain et al., 2019). This cycles are composed of asexual clonal replications and a sexual stage during which mating-types genes allow reproduction. This sexual recombination is determined within STE and HD genes encoding respectively pheromone receptors and homeodomain transcription factors (Luo et al., 2024). Additionally, rust fungi maintain a dikaryotic stage for the majority of their life cycle, in which two haploid nuclei coexist within the same cell but remain physically separated (Henningsen et al., 2025). Unlike diploid organisms, for which homologous chromosomes co-exist in a single nucleus, rust fungal genomes display distinct haplotypes into these two nuclei. This genomic arrangement, along with a high repeat content (Duplessis et al., 2011) has historically complicated efforts to generate fully phased genome assemblies at the chromosome level with the previous available technologies. However, advances in sequencing technologies have transformed our ability to reconstruct complex fungal genomes with high precision. Long-read sequencing platforms enabled the generation of long reads that span repetitive regions, significantly improving genome contiguity and scaffolding. Furthermore, Hi-C chromatin conformation capture techniques provide valuable chromosome-scale contact maps, facilitating haplotype phasing and structural genome resolution (Edwards et al., 2022; Henningsen et al., 2022, 2024; F. Li et al., 2019; Liang et al., 2023; Sperschneider et al., 2025a; Tobias et al., 2021; Wang et al., 2024; Wu et al., 2021; Xia et al., 2022). The generation of

phased and complete rust fungal genome assemblies, i.e. telomere-to-telomere (T2T) genomes, marks a critical advance, enabling in-depth exploration of structural variations, transposable elements, and key genomic features associated with host adaptation (Corre et al., 2025; Henningsen et al., 2024; Li et al., 2019).

The poplar rust fungi *Melampsora larici-populina* and *Melampsora allii-populina* are damaging pathogens of poplar trees, with major implications for forestry, bioenergy production, and ecosystem stability (Frey et al., 2005). These two species share a common plant host, poplar, at the telial stage during their clonal propagation as sporothallus but they differ in their sexual reproduction phase, as gametothallus, cycling on larch (*Larix spp.*) and Alliaceae species, respectively (Hacquard et al., 2011; Vialle et al., 2011). Earlier efforts to sequence the genome of *M. larici-populina* provided foundational insights into rust fungal biology, revealing extensive expansions in effector gene families, high genome plasticity, and coordinated expression of candidate secreted effector proteins (CSEPs) across developmental stages and host interactions (Duplessis et al., 2011; Lorrain et al., 2018). In addition, the generation of a high-density genetic map for *M. larici-populina* provided important insights into recombination patterns, forming a critical reference for validating and scaffolding genome assembly into 18 linkage groups (Pernaci et al., 2014). However, this previous assembly of 110Mb represented a collapsed version of the two haploid genomes, limiting the resolution of inter-haplotype diversity, and structural variants in *M. larici-populina*. For *M. allii-populina*, an initial draft genome assembly of 335.7Mb was generated using long-read sequencing in the frame of the 1000 fungal genome project at the Joint Genome Institute (Grigoriev et al., 2014). Although this assembly was gapless, it remained fragmented and unphased (Duplessis et al., unpublished).

Here, we present the first haplotype-phased and near chromosome-scale genome assemblies for *M. larici-populina* and *M. allii-populina*, generated by combining PacBio long-read

sequencing and Hi-C scaffolding. These assemblies resolve both nuclear genomes of the dikaryotic state and enable comprehensive haplotype-resolved gene and TE annotations. By providing high-resolution references for two closely related but ecologically distinct rust fungi, our work lays the foundation for comparative and functional genomics of rust pathogens and to advance the understanding of biotrophic adaptation in plant-fungal interactions.

MATERIALS & METHODS

Fungal material

Urediniospores of *M. larici-populina* strain 98AG31 and *M. allii-populina* strain 12AY07 were multiplied and collected from poplar leaves as previously described in (Duplessis et al., 2011). Leaf infections and material for RNA extraction and sequencing used for gene prediction were performed as described in (Duplessis et al., 2011).

Pacbio HiFi DNA and Hi-C sequencing

High molecular weight DNA extractions were performed from urediniospores for *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07, as described in Persoons et al. (2014). Briefly, 50-100 mg of urediniospores were ground with liquid nitrogen-cooled mortar and pestle, and DNA was extracted using a CTAB buffer followed by phenol-chloroform purification (Persoons et al., 2014). RNA was digested with RNaseA, and DNA was precipitated with isopropanol, washed with ethanol, and resuspended in Tris-EDTA. The resulting purified DNA samples were sent for library preparation and sequencing using PacBio Revio platform (Novogene Europe, Germany).

Urediniospores of *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 were prepared for Hi-C analysis using the PhaseGenomics protocol (Seattle, CA, USA). The samples were

crosslinked with formaldehyde according to the manufacturer's instructions to preserve chromatin structure. Hi-C library preparation was carried out using the Proximo Hi-C Kit (Fungal), which incorporates a four-enzyme digestion strategy specifically optimized for fungal genomes with high AT content by PhaseGenomics (Seattle, Washington, USA). The resulting libraries were sequenced on an Illumina NovaSeq X platform using paired-end sequencing to generate 50M of 75 bp reads.

Genome assembly and scaffolding

Genome assemblies for *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 were performed using the procedure described in Sperschneider et al. (2023). HiFi reads were assembled with hifiasm v0.19.9-r616 with Hi-C integration mode and default parameters (25.3 Gb HiFi reads and 65M Hi-C reads for *M. larici-populina* 98AG31; 39.7Gb HiFi reads and 55M Hi-C reads for *M. allii-populina* 12AY07) (Cheng et al 2021). We estimated the contigs coverage by re-aligning HiFi reads against the diploid assemblies using minimap2 (v2.24) (Li et al., 2018), and extracted contigs coverage using bbmap package (v 38.18), (Bushnell 2014). Contigs with less than 7X and 2X coverage were filtered from *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 assemblies, respectively. The two haplotype assemblies from hifiasm were combined into a diploid genome assembly, and mitochondrial and contaminant contigs were removed based on BLAST similarity searches (v2.12)(Altschul et al. 1990). Sequences with BLAST hits assigned to taxa outside the Pucciniales order and mitochondrial contigs were identified based on high coverage and low GC content, then confirmed via BLASTn against the NCBI RefSeq mitochondrial genome database (≥90% identity; DUST masking parameters enabled; https://github.com/JanaSperschneider/GenomeAssemblyTools/tree/master/ContaminantScreening). For each haplotype, the most represented contigs ranging from 70 to 90 kb in length were extracted as putative mitochondrial contigs. We assessed the completeness of each draft assembly using BUSCO v5 (Simão et al. 2015).

To phase the dikaryotic genome assemblies, we used NuclearPhaser (Duan et al., 2022), a graph-based pipeline that assigns contigs to haplotypes based on Hi-C chromatin contact data and gene homology (Duan et al., 2022). Briefly, NuclearPhaser identifies homologous contigs using gene and BUSCO duplication patterns to build scaffold bins, then uses Hi-C contact graphs to assign these bins to haplotypes through community detection. Contigs with inconsistent Hi-C signals are flagged as potential phase switches, which can be manually corrected to improve phasing accuracy. We generated gene hit tables using Biokanga v4.4.2 (Whan et al., 2019) and BUSCO hit tables using BUSCO v5 (Simão et al. 2015). Hi-C reads were mapped to each diploid assembly of *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 using bwa-mem2 v 0.7.17 (Vasimuddin et al., 2019). Hi-C interaction matrices were constructed using the *hicBuildMatrix* function from the HiCExplorer (Ramírez et al., 2018a) suite with a 100 kb bin size and corrected with iterative correction and eigenvector decomposition (ICE) via *hicCorrectMatrix* (Ramírez et al., 2018a). The HiCExplorer matrices were converted into *ginteractions* format using *hicConvertFormat*. Phase switches were corrected before running the NuclearPhase pipeline for a second round to validate the phase switches and scaffolding the assemblies.

For *M. larici-populina* 98AG31, the scaffolding of the fully phased assembly was finalized using Ragtag v2.1.0 (Alonge & Leonard, 2021) *scaffold* function with default parameters, using the *M. larici-populina* 98AG31 version 2 as reference (Persoons et al. 2022). Scaffolds were re-ordered following *M. lini haplotype C* haplotype-phased assembly (Sperschneider et al., 2025a). For *M. allii-populina* 12AY07 scaffolding, we constructed a *de novo* genome diploid assembly using the assembly from the first round of NuclearPhaser with the pipeline HapHiC (Zeng et al. 2024). For this, we re-mapped the Hi-C reads against the phased switch NuclearPhaser assembly with bwa-mem2 and ran the HapHiC pipeline with default parameters and expected number of

chromosomes set at 36. We then compared the resulting HapHic diploid assembly to our previous assembly to separate the two haploid genomes with whole-genome alignments dot plots generated with D-genies (Cabanettes and Klopp 2018). We used alignments with *M. lini haplotype C* haplotype-phased assembly to re-order scaffolds of *M. allii-populina*. We corrected the remaining mis-assemblies using whole-genome alignments between the two haplotypes and used Ragtag *correct* to break mis-assemblies, followed by Ragtag *scaffold* to correct them (Alonge & Leonard, 2021). Finally, we verified potential remaining misplaced contigs by running NuclearPhaser again to correct the last phase switch with the new Hi-C contact map.

We searched for telomeric repeats on contigs with FindTelomeres.py (Sperschneider 2025). We detected the putative centromeres by visual inspection of the Hi-C contact maps for each haplotype using the HiCExplorer suite (Ramírez et al., 2018a)**.** The step-by-step workflow for *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 assemblies is available here: https://github.com/Raistrawby/MrT2T. The hypotheticals centromeres area are detected by visual inspections.

Genome alignments and synteny visualization

To assess structural collinearity and divergence between haplotypes, we performed pairwise whole-genome alignments using Minimap2 v2.1 (Li, 2018) with the -x asm5 parameter, which is optimized for aligning closely related assemblies. Alignments were generated between haplotype 0 and haplotype 1 for each species. The resulting alignments were output in PAF format and processed using the paftools.js stat utility from Minimap2 to extract summary statistics, including the percentage of aligned bases between haplotypes. For synteny visualization, we used D-GENIES (Cabanettes & Klopp, 2018) to generate dot plots. Dot plots

were used to visually inspect the overall genome collinearity and detect large-scale structural rearrangements.

Transposable elements annotations and mating-type genes detection

The detailed TE annotation and classification of both haplotypes for *M. larici-populina* and *M. allii-populina,* was performed using REPETv3.0 (Quesneville et al. 2005) and a curated TE library, as described in Corre et al. (2025). To detect the mating type genes, we retrieved the mating type genes annotated in *M. larici-populina v2* and used BLAST searches to detect them in the different haplotypes (Altschul et al. 1990).

Table 1: Genome information of *M. larici-populina* v3 versus *M. larici-populina* v2

	M. larici-populina v2	***M. larici-populina H0***	***M. larici-populina H1***	***M. larici-populina unphased contigs****
Scaffolds	511 scaffolds	18 scaffolds	18 scaffolds	-
Assembly Size	109.88Mb	100.80 Mb	102.64 Mb	4.04 Mb
Contigs	4520	32	39	44
scaffold N/L50	8/5.82 Mbp	8/5.74 Mbp	8/5.42 Mbp	11/100.57 Mbp
scaffold N/L90	18/4.01 Mbp	16/4.15 Mbp	16/4.17 Mbp	33/38.92 Kbp
Max scaffold length	8.57 Mbp	8.24 Mbp	8.67 Mbp	-
Gaps	5.83%	0.001%	0.001%	
%GC	41.0%	40.9%	40.9%	38.9%
Telomeres	-	30	28	3
Centromeres	-	18	18	-
BUSCO complete	90.3%	90.5%	90.6%	0.1%
BUSCO duplicated	3.1%	0.5%	0.5%	0%
Aligned bases H0 vs H1	-	95788054	-	-
Aligned percentage H0 vs H1	-	93.32%	-	-
Number of SNPs H0 vs H1	-	196960	-	-
Genes	19550			
%TEs	46.98%	58.9 %	59.52%	5.71%

* contig that were not attributed to either haplotype

RESULTS

Complete assemblies for *M. larici-populina* and *M. allii-populina*

We generated high-quality, haplotype-phased, near chromosome-scale genome assemblies for *M. larici-populina* 98AG31 and for *M. allii-populina* 12AY07 using long-read sequencing and Hi-C scaffolding. For *M. larici-populina*, the new assembly comprises 18 chromosomes per haplotype, each representing a near-complete nuclear genome, with a total assembly size of 202.72 Mb (100.88 Mb for haplotype H0 and 102.64 Mb for haplotype H1; Table 1) and 4.04 Mb of unplaced contigs. The N50 for each haplotype assembly is 5.74-5.42 Mb, similar to the previous version genome, which consisted of 511 scaffolds including 18 linkage groups and an N50 of 5.82 Mb (Table 1; Figure 1). The improved assembly also closed numerous gaps compared to the previous version, with 5.83% of gaps in *M. larici-populina* 98AG31 version 2 and 0-001% gaps in *M. larici-populina* 98AG31 version 3 H0 and H1, respectively (Table 1; Figure 1). We detected 18 centromeres and 61 telomeres, indicating that 24 scaffolds were assembled at the chromosome level and 31 showed at least one telomere. Similarly, the *M. allii-populina* 12AY07 full assembly size is 415.95 Mb with 224.03-219.39 Mb per haplotype and 8.53 Mb of unplaced contigs (Table 2). In contrast, the previous version 1 assembly for *M. allii-populina* was fragmented, containing 3637 scaffolds over 335.73Mb.

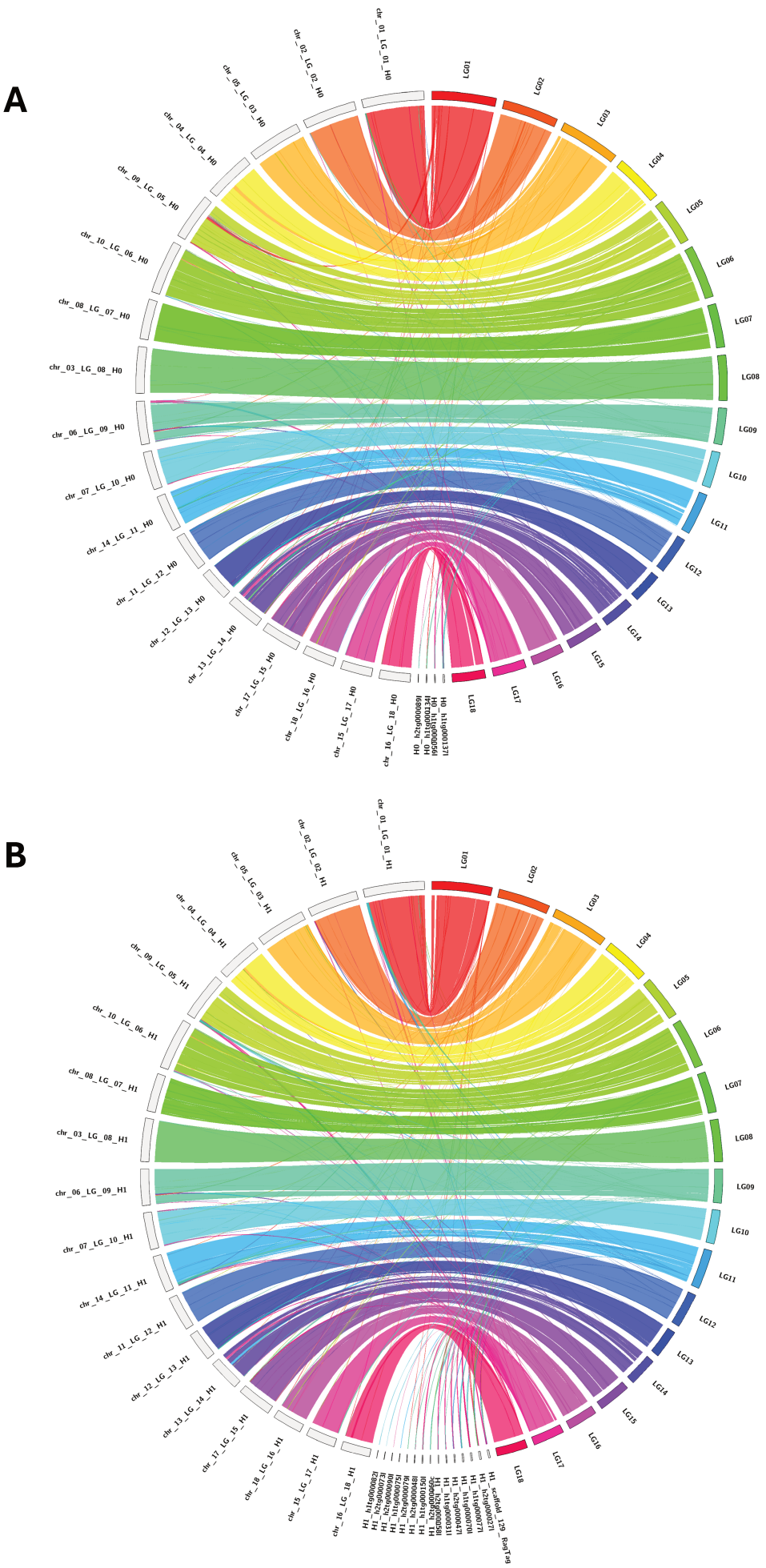


Figure 1. Comparison of *Melampsora larici-populina* 98AG31 genome assemblies. Circos plots showing synteny between the *M. larici-populina* v2 reference assembly and A) the v3 haplotype 0 assembly and B) the v3 haplotype 1 assembly. The plots were generated using Circos with the following parameters: window size = 50 kb, minimum gap size = 5 kb, minimum quality score = 20, scaffold break threshold = 50 kb gap, and a minimum link size = 10 kb (Krywinski et al., 2009)

Table 2: *M. allii-populina* v2 genome information versus *M. allii-populina* v1

	M. allii-populina v1	***M. allii-populina H0***	***M. allii-populina H1***	***M. allii-populina unphased contigs****
Scaffolds	3637 scaffolds	18 scaffolds	18 scaffolds	-
Assembly Size	335.73Mb	224.03Mb	219.39Mb	8.53 Mb
Contigs	3637	33	46	116
scaffold N/L50	440/191.40 Kbp	8/13.31 Mbp	8/12.91 Mbp	30/79.11 Kbp
scaffold N/L90	2188/35.5 Kbp	16/9.29 Mbp	16/9.11 Mbp	92/41.98 Kbp
Max scaffold length	1.29 Mbp	17.52 Mbp	16.99 Mbp	312.62 Kbp
Gaps	0%	0%	0.001%	0%
%GC	40.0%	40.0%	40.0%	37.8%
Telomeres	-	26	24	2
Centromeres	-	18	18	-
BUSCO complete	89.7%	90.7%	86.1%	0.3 %
BUSCO duplicated	22.1%	1.1%	1.1%	0.1 %
Aligned bases H0 vs H1	-	192817739	-	-
Aligned percentage H0 vs H1	-	87.9%	-	-
Number of SNPs H0 vs H1	-	906156	-	-
Genes	23089			
%TEs	74.80%	82.27%	82%	2.7%

*contigs that were not attributed to either haplotype

The N50 for each haplotype assembly is 13.31-12.91 Mb for H0 and H1, respectively, representing a dramatic increase in contiguity compared to the version 1 genome, which consisted of an N50 of only 0.19 Mb (Table 2). At least one telomeric repeat was identified for 17 and 17 scaffold ends in H0 and H1, respectively, supporting the near T2T level of this assembly. Both genome assemblies exhibit high completeness, with BUSCO scores of 90.5% and 88.4% for *M. larici-populina* and *M. allii-populina*, respectively, based on the Basidiomycota reference set (Table 1 & 2). In addition, we found 0.5% and 1.1% of BUSCO duplicated levels in *M. larici-populina* v3 and *M. allii-populina* v2, compared to 3.1% and 22.1% in the previous versions of *M. larici-populina* v2 and *M. allii-populina* v1, respectively. This reduction of duplication levels supports the haplotype resolution of these new assemblies. We have generated two fully phased and near T2T reference genome assemblies of two poplar rust species.

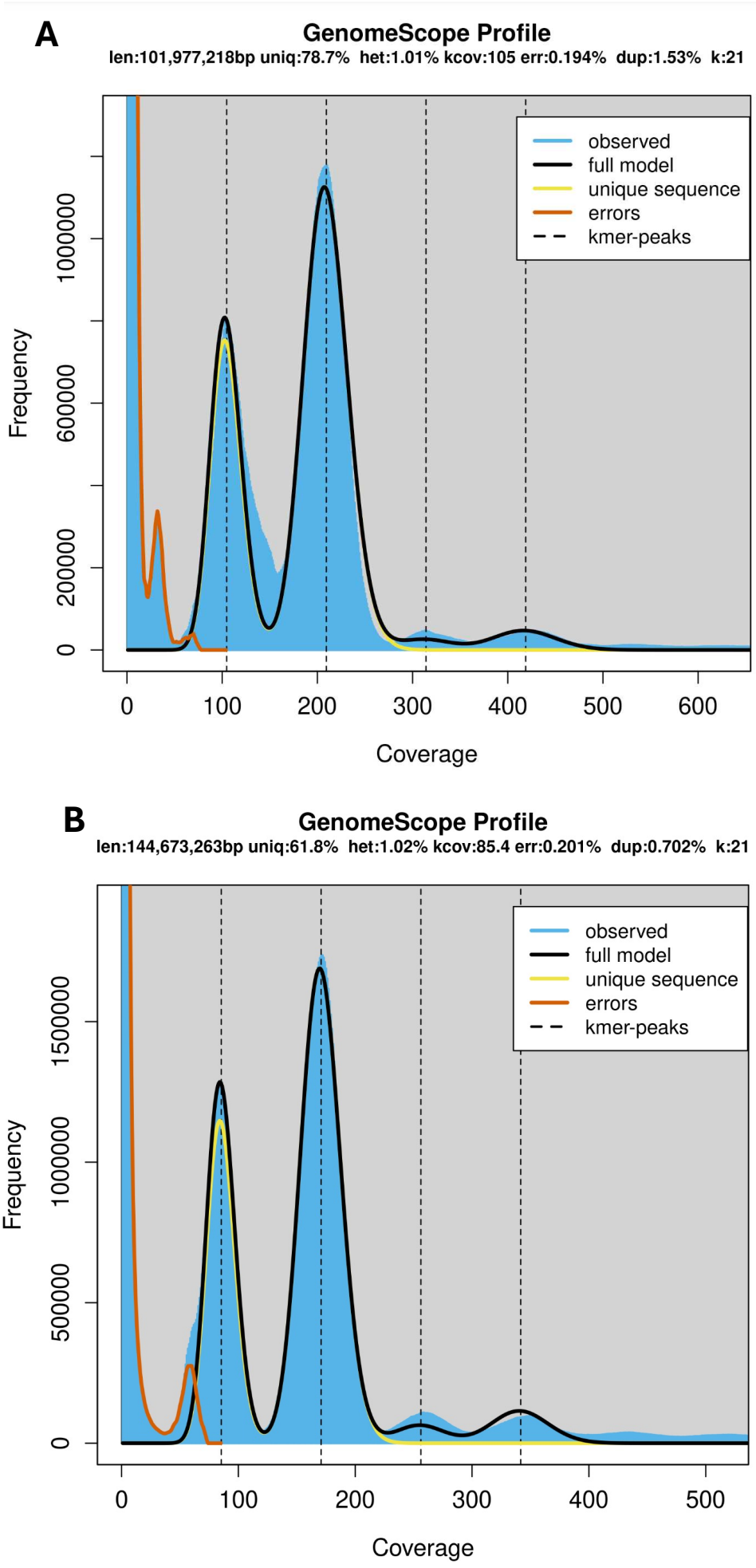


Figure 2. Genome size and heterozygosity estimation from HiFi read k-mer spectra.
GenomeScope 21-mer profiles of HiFi sequencing reads from A) *M. larici-populina* 98AG31 with an estimated genome heterozygosity (het) of 1.01%, and B) *M. allii-populina* 12AY07 with an estimated genome heterozygosity of 1.02%. The distinct 1:2 ratio of the k-mer peaks supports the diploid nature of both species.

The dikaryotic genomes of *M. larici-populina* 98AG31 and *M. allii-populina 12AY07* display similar levels of heterozygosity and collinearity

The dikaryotic genomes of *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 exhibit both high heterozygosity and strong collinearity between haplotypes, consistent with the long-term maintenance of divergent nuclei in rust fungi. We assessed the heterozygosity levels of raw reads using a 21-mer-based approach. Heterozygosity was estimated at 1.02% for *M. larici-populina* and 1.01% for *M. allii-populina* (Figure 2), indicating substantial allelic divergence between the two dikaryotic haplotypes.

We used D-genies dot plots to detect the collinearity and the rearrangement between each haplotype and within the previous versions (Figure 3 & 4). Whole-genome alignments revealed that 93.3% of the haplotype 0 sequence aligned to haplotype 1 in *M. larici-populina*, indicating a high level of genome conservation between haplotypes (Table 1). In contrast, only 87.2% of the sequence aligned between haplotypes in *M. allii-populina*, reflecting greater structural divergence within this species (Table 2). Despite these differences, both species exhibited overall high collinearity between haplotypes, with no evidence of large-scale structural rearrangements (Figure 3C & 4C). The only major non-syntenic region identified in both species spanned approximately ~1.5-2 Mb on scaffold 9, where no alignment was detected between haplotypes (Figure 3D & 4D). This region corresponds to the mating-type locus containing the PR genes, previously characterized in *M. lini* (Sperschneider et al., 2025a).

To further explore the mating-type loci in poplar rusts, we identified both mating-type genes in our assemblies, including the pheromone receptor (PR) *STE3* and pheromone *mfa* genes, and the homeodomain (HD) locus, which harbours the divergently transcribed transcription factors *bE* and *bW*, consistent with the known mating-type organization in rust fungi (Cuomo et al. 2017).

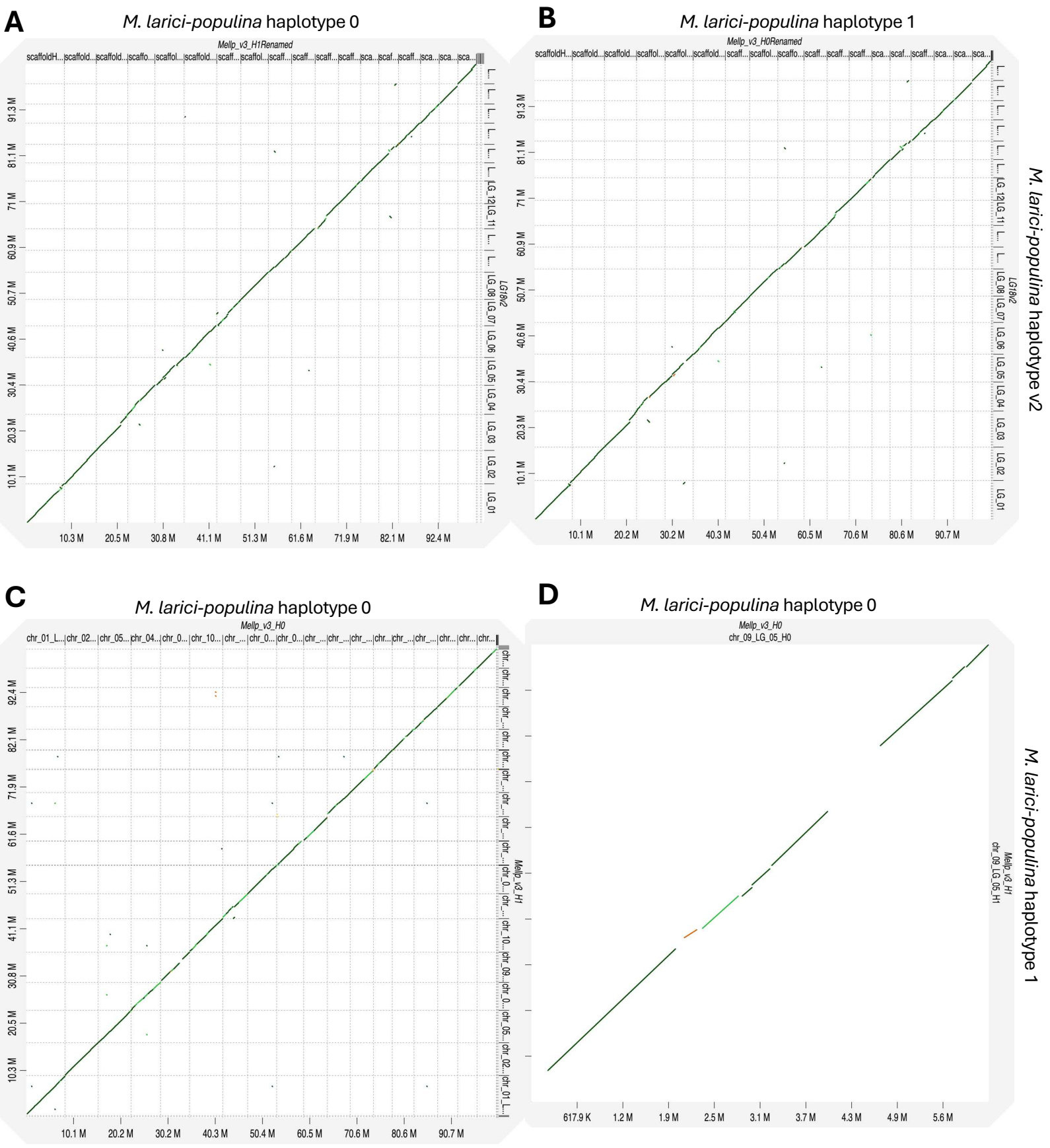


Figure 3. Whole-genome dot plots of *Melampsora larici-populina* 98AG31 haplotype-phased assemblies and comparison to the genome assembly version 2. A) Synteny dot plot of the newly assembled haplotype 0 (x-axis) against the previous reference genome version of *M. larici-populina* v2 (y-axis). B) Synteny dot plot of the newly assembled haplotype 1 (x-axis) against the *M. larici-populina* v2 (y-axis). C) Synteny dot plot between haplotype 0 (x-axis) and haplotype 1 of *M. larici-populina* (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~1.5-1.7 Mb non-syntenic region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*.

We used the annotated STE and HD genes from M. larici-populina v2 and M. lini CH5 to search for corresponding loci in the new assemblies M. larici-populina v3 and M. allii-populina v2 (Table S1). We located the STE3.2 genes on the scaffold 9 of M. larici-populina H1 and on the scaffold 9 of M. allii-populina H0, while the STE3.2.3 and STE3.2.4 genes were found on the scaffold 9 of M. larici-populina H0 and on the scaffold 9 of M. allii-populina H1. Similarity search placed the STE3.2.1 genes on scaffold 1 in both haplotypes of both species. Similarly, we identified the HD loci on scaffold 4 in both haplotypes of both species (Table S1). Together, our results indicate that the MAT loci of both STE3 and HD are conserved within the different poplar rust.

Improvement of the TE annotations in the haplotype-phased assemblies

TEs were annotated in each haplotype of the two species and classified at the Class, Order, and Superfamily (SF) levels (Figure 5; Tables S2, S3 & S4). The phased assemblies revealed differences in the overall TE genome coverage relative to previously published unphased references. The TE content in *M. larici-populina* v2 represented 47% of the genomes, whereas 55.7% and 56.3% of TEs were accounted for in haplotype 0 and haplotype 1, respectively (Figure 5A; Table 1). This indicates that the improvement in genome continuity (5% gaps in version 2 versus 0.001% gaps in the haplotype-phased assemblies) allowed for a more comprehensive recovery of repetitive elements, capturing TE sequences that were previously collapsed or fragmented in the earlier assembly. However, we observed that the *M. larici-populina* genome still displayed slightly more Class II DNA transposons than Class I retrotransposons, with 29.7-28.7% Class II and 22.3-24.0% Class I in haplotype 0 and 1, respectively, whereas we found 15.4% and 23.2% in the earlier genome version (Figure 5B; Table S2). Similarly, the *M. allii-populina* haplotype-phased assembly displayed 82.3-82.0% of TEs in haplotypes 0 and 1, respectively, whereas the previous version accounted for 74.8% of

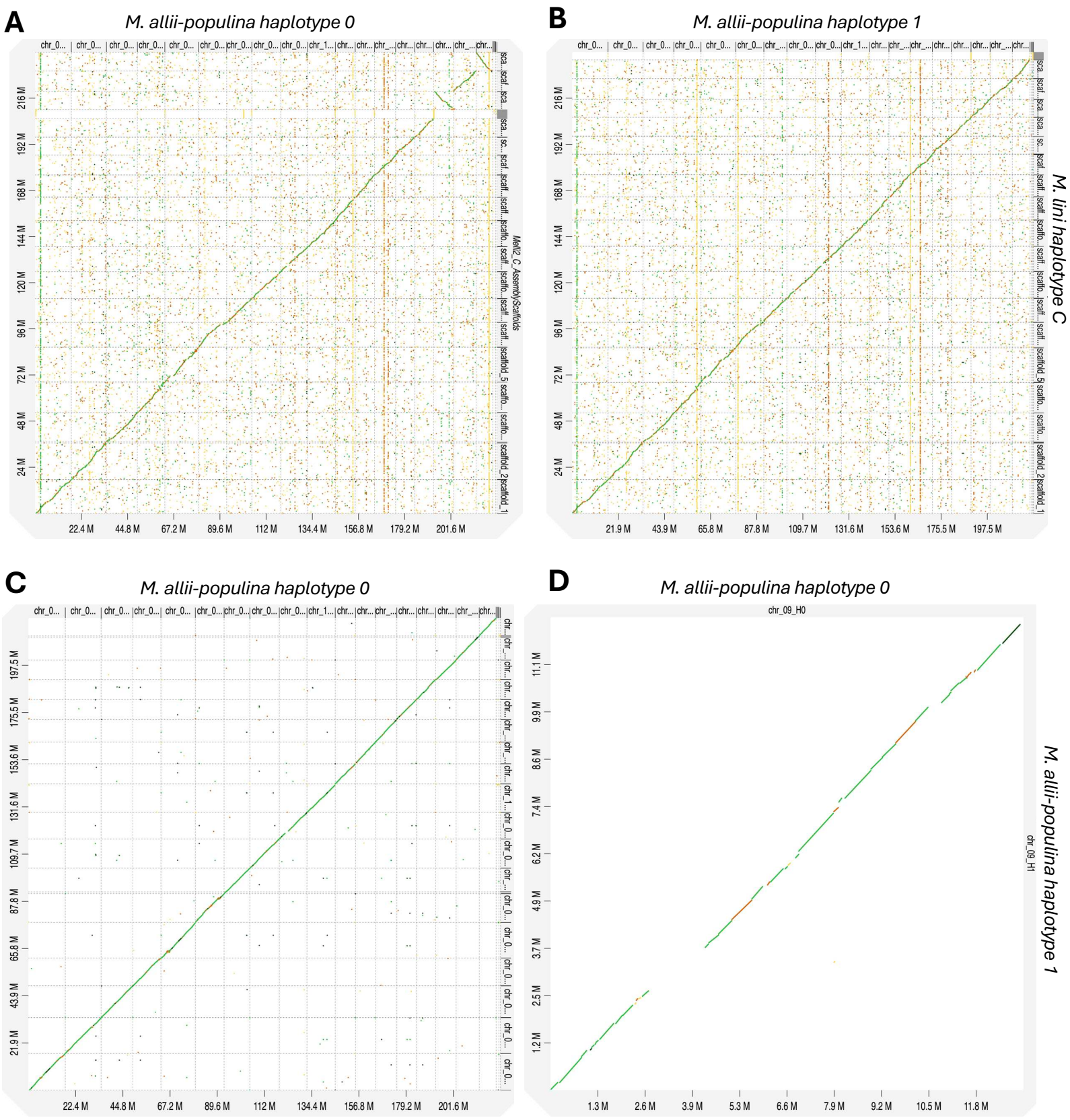


Figure 4. Whole-genome dot plots of *Melampsora allii-populina* 12AY07 haplotype-phased assemblies and comparison with *Melampsora lini* haplotype C. A) Synteny dot plot of the *M. allii-populina* haplotype 0 (x-axis) against the reference genome version of *M. lini* haplotype C (y-axis). B) Synteny dot plot of the *M. allii-populina* haplotype 1 (x-axis) against the *M. lini* haplotype C (y-axis). C) Synteny dot plot between *M. allii-populina* haplotype 0 (x-axis) and *M. allii-populina* haplotype 1 (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between *M. allii-populina* haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~2 Mb non-syntenic region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*.

TEs (Figure 5A; Table 2). Class I elements covered 49.5-49.2% of haplotypes 0 and 1, respectively; and Class II elements represented 27.8-27.9% of haplotypes 0 and 1, respectively (Figure 5B; Table S2). At the Order level, LTR retrotransposons dominated the TE landscapes of *M. allii-populina* (43.5% of LTRs in haplotype 0 and 43.9% in haplotype 1), whereas *M. larici-populina* displayed a higher proportion of TIR DNA transposons (Figure 5C & D; Table S3). In the latter, TIR elements accounted for 23.9% in haplotype 0 and 22.9% in haplotype 1, similar to the coverage found in the previous version and remained the most abundant Order in the phased genome (Figure 5C & D; Table S3). At the Superfamily level, Gypsy-elements remained the most prevalent in *M. allii-populina* (30.0% and 30.3% of the genome in haplotype 0 and 1, respectively), whereas in *M. larici-populina*, Gypsy elements constituted the second most abundant category after TIR elements (with 8.86% in Gypsy in both haplotypes; Figure 5E & F; Table S4). Taken together, these findings confirmed that haplotype-resolved assemblies offered a significantly enhanced representation of the TE landscape in the two poplar rust species. The improved genome continuity enabled the recovery of TE sequences that were previously fragmented or collapsed, thereby allowing for a more complete annotation of repetitive elements.

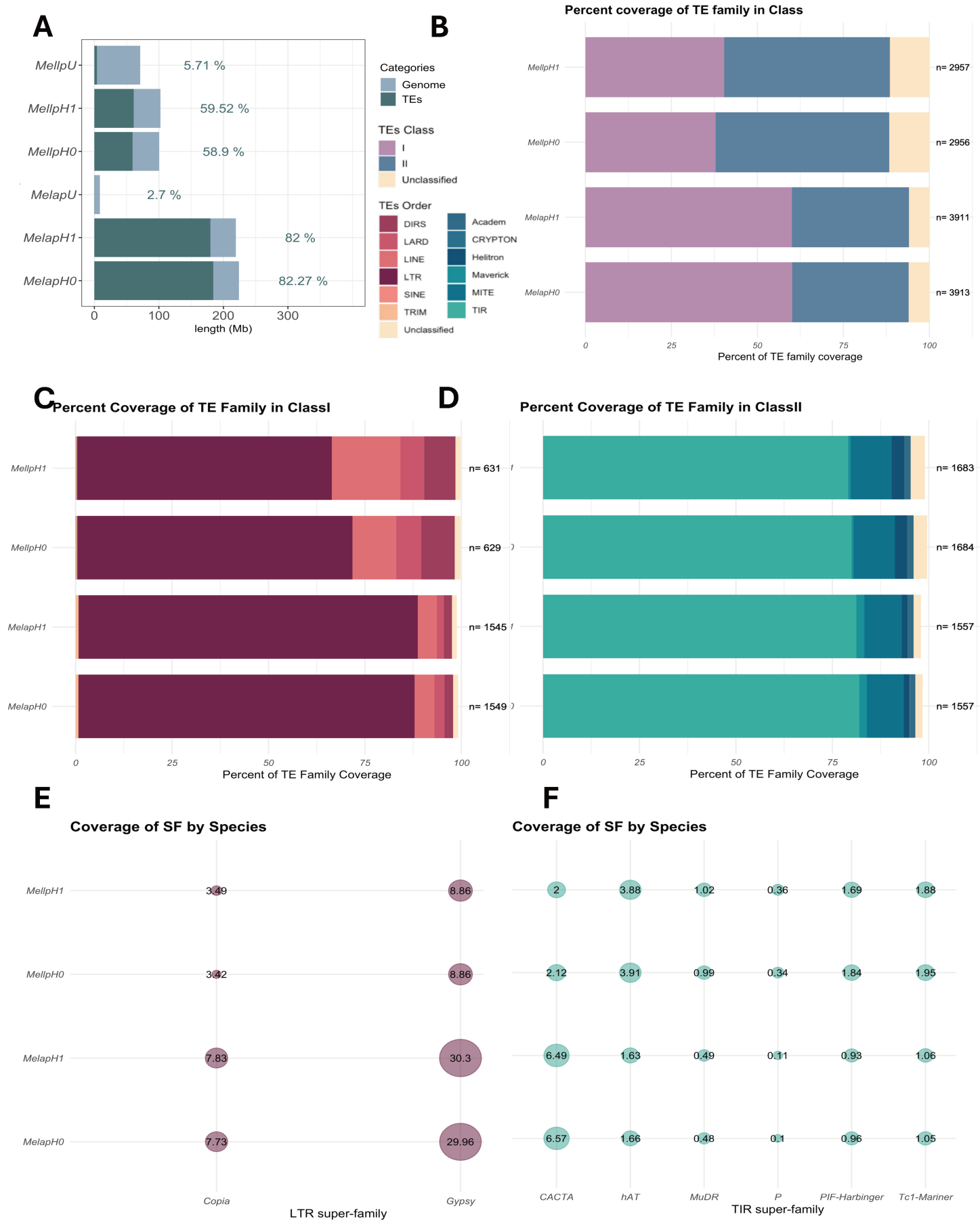


Figure 5. Annotation and distribution of transposable elements (TEs) in the genomes of *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07. A) Genome coverage of TEs (dark blue) relative to the whole genome assembly (light blue) in *M. larici-populina* unphased contigs (MellpU), *M. larici-populina* haplotype 0 (MellpH0), *M. larici-populina* haplotype 1 (MellpH1), *M. allii-populina* unphased contigs (MelapU), *M. allii-populina* haplotype 0 (MelapH0), and *M. allii-populina* haplotype 1 (MelapH1). B) Distribution of TE classes across species, expressed as percentage of total TEs (number of TE families indicated as *n*). C) Distribution of TE Orders within Class I elements, expressed as percentage of total Class I TEs (families *n*). D) Distribution of TE Orders within Class II elements, expressed as percentage of total Class II TEs (families *n*). E) Coverage (in Mb) of long terminal repeat (LTR) Superfamilies and F) terminal inverted repeats (TIR) elements Superfamilies across *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 haplotypes.

DISCUSSION

A T2T haplotype-phased genome assembly provides a crucial foundation for understanding genome structure, functional gene content, and the evolutionary dynamics of repetitive elements. In obligate biotrophic fungi with dikaryotic life stages such as rust fungi, fully phased assemblies are essential to resolve haplotypic divergence, structural rearrangements, and resolution of mating-type loci. Here, we present the first fully phased, near T2T genome assemblies for two species of the *Melampsora* genus, *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07. The haploid genome sizes of these species are 100-102 Mbp and 219-224 Mbp, respectively, consistent with earlier genome size estimates and comparative metrics within Pucciniales (Corre et al., 2025; Duplessis et al., 2011; Sperschneider et al., 2025a). Previous collapsed assemblies were highly fragmented (e.g., >500 scaffolds in *M. larici-populina v2*, N50: 5.8 Mb) or incomplete (e.g., *M. allii-populina v1* had N50: 191 Kbp, and over 3,600 scaffolds), reflecting the limitations of assembling large, repeat-rich fungal genomes even with long-read technologies. Using the combination of long-read sequencing and Hi-C scaffolding, we resolved 18 near-complete chromosomes per haplotype for each species, with 11-13 scaffolds in *M. larici-populina* and 10-6 scaffolds in *M. allii-populina* capped by telomeres at both ends and all scaffolds were marked by centromeres. The enhanced continuity of these assemblies not only allowed us to recover a more complete gene set (BUSCO scores >88%) and repeat landscapes (up to 82% TE content in *M. allii-populina*), but also enabled the accurate localization of mating-type loci and analysis of haplotype collinearity. These assemblies represent a substantial step forward in *Melampsora* genomics and contribute to the growing foundation of haplotype-resolved rust fungal genomes, enabling more accurate investigations of haplotype evolution in genome architecture.

These high-contiguity, fully haplotype-phased assemblies for *M. larici-populina* and *M. allii-populina* represent a valuable addition to the expanding set of chromosome-scale genome references in rust fungi. The success of our assemblies highlights the robustness and versatility of the current framework combining PacBio HiFi and Hi-C-based nuclear phasing (Henningsen et al., 2022), which has been successfully applied across multiple Pucciniales species including *M. lini* (Henningsen et al., 2025; Sperschneider et al., 2025a). Our results confirm that this integrated strategy is well-suited for resolving the large, heterozygous, and repeat-rich rust fungal genomes. Both assemblies resolve 18 chromosomes per haplotype, with telomeric and centromeric features clearly identified, in line with the conserved karyotype of n = 18 chromosomes observed across most rust fungi (Henningsen et al., 2025). In addition, we report ~1% heterozygosity between haplotypes in both species and detect high levels of inter-haplotype collinearity with only a few structural differences. A single large divergent region (~1-2 Mb) on scaffold 9 was associated with the mating-type loci, similar to observations in *Puccinia graminis* f. sp. *tritici* and *M. lini*, where mating loci often occur within rearranged, haplotype-specific genomic regions (Sperschneider et al., 2025; Cuomo et al., 2017). By fully resolving the haplotypes of two poplar rust pathogens, our assemblies revealed conserved genomic features of *M. larici-populina* and *M. allii-populina*, including chromosome number, large-scale inter-haplotype synteny, and mating-type architecture, thereby reinforcing the emerging model of genome organization within the rust fungi.

The resolution of mating-type loci in both *M. larici-populina* and *M. allii-populina* assemblies supports the presence of a tetrapolar mating system, with physically unlinked PR (STE genes) and HD (bE/bW genes) loci located on scaffolds 9 and 4, respectively. This genomic organization is consistent with previous studies across rust fungi (Cuomo et al., 2017; Sperschneider et al., 2025), and our assemblies further clarify the arrangement and haplotype-specific positioning of these loci. Notably, whereas cereal rusts such as *Puccinia*

triticina and *P. graminis* f. sp*. tritici* typically encode a single STE3.2 gene per haplotype (Cuomo et al., 2017), a recent study in *M. lini* revealed a duplication of this gene on one haplotype (*STE3.2.2* and *STE3.2.4*) and suggested the existence of recombination suppression spanning a ~2 Mb region surrounding the mating-type loci (Sperschneider et al., 2025). In the poplar rust genomes, the mating-type loci also appear to fall within structurally divergent haplotype regions with minimal alignment between haplotypes, raising the possibility of similar recombination suppression mechanisms. Whether such divergence reflects functional diversification of mating compatibility or structural barriers to recombination remains to be investigated. As proposed for *M. lini*, additional population-scale phased assemblies will be required to test whether mating-type structure in *M. larici-populina* and *M. allii-populina* conforms to a classic tetrapolar system or represents a more complex configuration (Lawrence, 1980).

The increased contiguity of genome assemblies also improves annotation of TEs, which play a dominant role in shaping rust genomes (Corre et al., 2025). TE content was 25% higher in *M. larici-populina* and 10% higher in *M. allii-populina* in the new haplotype-phased assemblies compared to previous versions, supporting the observation that collapsed assemblies underestimate repetitive regions (Henningsen et al., 2025; Edwards et al., 2022). Despite such differences between genome assemblies, the overall tendencies in TE composition remained consistent between assembly versions for each strain, indicating that while total TE content was underestimated previously, the dominant TE classes were already correctly identified. The improved continuity and haplotype resolution in our assemblies enabled the recovery of previously collapsed or fragmented elements, resulting in more accurate estimates of repeat abundance. In *M. allii-populina*, Class I LTR retrotransposons, particularly Gypsy elements, remained the most abundant, while *M. larici-populina* continued to show an enrichment of Class II TIR DNA transposons. These refined annotations align with TE compositions reported in other

rust fungi and illustrate the value of long-read, chromosome-scale assemblies in capturing complex repetitive landscapes (Zheng et al., 2021; Sperschneider et al., 2025).

FIGURE LEGENDS

Figure 1. Comparison of *Melampsora larici-populina* 98AG31 genome assemblies. Circos plots showing synteny between the *M. larici-populina* v2 reference assembly and A) the v3 haplotype 0 assembly and B) the v3 haplotype 1 assembly. The plots were generated using Circos with the following parameters: window size = 50 kb, minimum gap size = 5 kb, minimum quality score = 20, scaffold break threshold = 50 kb gap, and a minimum link size = 10 kb (Krzywinski et al., 2009).

Figure 2. Genome size and heterozygosity estimation from HiFi read k-mer spectra. GenomeScope 21-mer profiles of HiFi sequencing reads from A) *M. larici-populina* 98AG31 with an estimated genome heterozygosity (het) of 1.01%, and B) *M. allii-populina* 12AY07 with an estimated genome heterozygosity of 1.02%. The distinct 1:2 ratio of the k-mer peaks supports the diploid nature of both species.

Figure 3. Whole-genome dot plots of *Melampsora larici-populina* 98AG31 haplotype-phased assemblies and comparison to the genome assembly version 2. A) Synteny dot plot of the newly assembled haplotype 0 (x-axis) against the previous reference genome version of *M. larici-populina* v2 (y-axis). B) Synteny dot plot of the newly assembled haplotype 1 (x-axis) against the *M. larici-populina* v2 (y-axis). C) Synteny dot plot between haplotype 0 (x-axis) and haplotype 1 of *M. larici-populina* (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~1.5-1.7 Mb non-syntenic region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*.

Figure 4. Whole-genome dot plots of *Melampsora allii-populina* 12AY07 haplotype-phased assemblies and comparison with *Melampsora lini* haplotype C. A) Synteny dot plot of the *M. allii-populina* haplotype 0 (x-axis) against the reference genome version of *M. lini* haplotype C (y-axis). B) Synteny dot plot of the *M. allii-populina* haplotype 1 (x-axis) against the *M. lini* haplotype C (y-axis). C) Synteny dot plot between *M. allii-populina* haplotype 0 (x-axis) and *M. allii-populina* haplotype 1 (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between *M. allii-populina* haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~2 Mb non-syntenic region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*.

Figure 5. Annotation and distribution of transposable elements (TEs) in the genomes of *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07. A) Genome coverage of TEs (dark blue) relative to the whole genome assembly (light blue) in *M. larici-populina* unphased contigs (MellpU), *M. larici-populina* haplotype 0 (MellpH0), *M. larici-populina* haplotype 1 (MellpH1), *M. allii-populina* unphased contigs (MelapU), *M. allii-populina* haplotype 0 (MelapH0), and *M. allii-populina* haplotype 1 (MelapH1). B) Distribution of TE classes across species, expressed as percentage of total TEs (number of TE families indicated as *n*). C) Distribution of TE Orders within Class I elements, expressed as percentage of total Class I TEs (families *n*). D) Distribution of TE Orders within Class II elements, expressed as percentage of total Class II TEs (families *n*). E) Coverage (in Mb) of long terminal repeat (LTR) Superfamilies and F) terminal inverted repeats (TIR) elements Superfamilies across *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 haplotypes.

Table S1. **Mating type genes homologues**

MAT genes retrieved through blast searches against the genomes of *M. larici-populina 98AG31 haplotype 0, M. larici-populina 98AG31 haplotype 1, M. allii-populina haplotype 0, M. allii-populina haplotype 1, M. lini haplotype C, M. lini haplotype H.*

Table S2. **Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Class level**. 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. **Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Order level**. 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. **Transposable elements (TEs) annotations for *Melampsora larici-populina* 98AG31 and *Melampsora allii-populina* 12AY07 genomes at the TE Superfamily (SF) level**. SFAnn corresponds to 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

Chapter II - Part 2

Haplotype resolved and complete assembly of an asexual lineage of the poplar rust fungus *Melampsora larici-populina* reveals large structural rearrangements at the mating type locus

INTRODUCTION

The poplar rust fungus *Melampsora larici-populina* exhibits a complex heteroecious life cycle, alternating between larch (*Larix* spp.) as the sexual host and poplar (*Populus* spp.) as the asexual host, where repeated clonal cycles occur throughout the growing season (Frey et al., 2005). *M. larici-populina* populations in Europe contain both sexually and asexually reproducing lineages coexisting in the same environment (Xhaard et al., 2011). While sexual lineages complete a complete heteroecious cycle, certain genotypes persist exclusively through clonal reproduction on poplars, apparently bypassing the larch-dependent sexual phase (Abdalrahem et al., 2025). These long-term asexual lineages are intriguing from an evolutionary perspective as they offer natural experiments for studying the genomic consequences of mitotic propagation, such as mutation accumulation, structural instability, and altered transposable element (TE) dynamics (Hartfield & Keightley, 2012; Lynch et al., 1993; Billiard et al., 2012). Yet despite their ecological relevance and persistence over decades, the genomic basis of long-term clonal propagation in poplar rust fungi remains poorly characterized, largely due to the absence of high-quality phased genome assemblies from clonal isolates.

The recent development of long-read sequencing combined with Hi-C scaffolding and phasing has enabled chromosome-scale genome assemblies in rust fungi (Li et al., 2019; Wu et al., 2022; Sperschneider et al., 2025; this Thesis, Chapter II Part 1). Here, we present a near chromosome-level, haplotype-resolved genome assembly of *M. larici-populina* 08KE26, an isolate representing a long-term asexual lineage prevalent in the south of France (Abdalrahem et al., 2025 ; Abdalrahem, 2025). Using PacBio HiFi sequencing and Hi-C contact mapping, we generated a highly contiguous genome with fully phased haplotypes. This assembly provides new insights into the structural dynamics of clonal rust genomes and complements existing resources such as the reference genome of the sexual isolate 98AG31. Our analyses focused

on three aspects: (1) genome structure and completeness; (2) haplotype collinearity and mating-type locus architecture; and (3) TE dynamics. Together, these data offer new insights into the evolutionary genomics of asexual rust fungi and provide a foundation for future studies of reproductive mode transitions in this poplar rust fungus.

MATERIALS & METHODS

DNA extraction, Pacbio HiFi and Hi-C sequencing

We multiplied and collected *M. larici-populina* 08KE26 urediniospores from poplar leaves as previously described in (Duplessis et al., 2011). High molecular weight DNA isolation was performed from *M. larici-populina* 08KE26 urediniospores, as described in Persoons et al. (2014). In summary, after grinding the snap-frozen spores with mortar and pestle, DNA extraction was carried out using a CTAB-based buffer, followed by purification with phenol-chloroform. We used RNaseA treatment to remove RNA contaminations and the purified DNA was subsequently submitted to Novogene Europe (Germany) for library construction and sequencing using the PacBio Revio platform. For Hi-C sequencing, the urediniospores of *M. larici-populina* 08KE26 were crosslinked with formaldehyde and Hi-C library preparation was carried out using the Proximo Hi-C Kit (Fungal) (Phase Genomics, Seattle, CA, USA). The resulting libraries were sequenced on an Illumina NovaSeq X platform using paired-end sequencing to generate 150M of 75 bp reads.

Genome assembly, scaffolding and transposable elements annotations

M. larici-populina 08KE26 PacBio HiFi reads were assembled with hifiasm v0.19.9-r616 (Cheng et al 2021) with Hi-C integration mode and default parameters (31.70 Gb HiFi reads and 150 M Hi-C reads). We filtered mitochondrial and contaminant contigs based on BLAST similarity searches, excluding any contigs that matched non-Pucciniales species (v2.12) (Altschul et al.,

1990). We also filtered contigs supported by low coverage (less than 33 X) using the bbmap package (v 38.18) (Bushnell, 2014), after re-aligning HiFi reads against the diploid assemblies using minimap2 (v 2.1) (H. Li, 2018). Assembly completeness was assessed with BUSCO v5 (Waterhouse et al., 2018). We separated the phased haplotypes using the NuclearPhaser pipeline (Duan et al., 2022) using the *M. larici-populina* 98AG31 genome version 3 to scaffold and correct the resulting phased assemblies with RagTag (Alonge et al., 2019; Alonge & Leonard, 2021). We performed two rounds of NuclearPhaser followed by scaffolding (cf. Chapter 2 part 1). The Hi-C contact matrices were generated using the HiCExplorer suite with a bin size of 100kb for the diploid genome and 10kb for haploid genomes (Ramírez et al., 2018b). We searched for telomeric repeats on contigs with FindTelomeres.py (Sperschneider et al., 2025b). We detected the putative centromeres by generating haplotype-specific Hi-C contact maps using the HiCExplorer suite (Ramírez et al., 2018b). In a subsequent step, transposable elements were annotated using REPET v3.0 (Quesneville et al. 2005) alongside *M. larici-populina* 98AG31 curated TE library, as described in this Thesis, Chapter II Part 1.

Table3: *M. larici-populina* 08KE26 genome information versus *M. larici-populina* 98AG31 v3

	Mlp98AG31 H0	***Mlp98AG31 H1***	***Mlp98AG31 unphased contigs***	***Mlp08KE26 H0***	***Mlp08KE26 H1***	***Mlp08KE26 unphased contigs***
Scaffolds	18 scaffolds	18 scaffolds	-	18 scaffolds	18 scaffolds	-
Assembly Size	100.80 Mb	102.64 Mb	4.04 Mb	101.18 Mb	101.79 Mb	19. 68 Mb
Contigs	32	39	44	112	81	364
scaffold N/L50	8/5.74 Mb	8/5.42 Mb	11/100.57 Mb	8/5.544 Mb	8/5.38 Mb	108/57.26 Kb
scaffold N/L90	16/4.15 Mb	16/4.17 Mb	33/38.92Kb	17/4 Mb	17/4.05 Mb	294/32.16 Kb
Max scaffold length	8.24 Mb	8.67 Mb	-	7.83 Mb	7.88 Mb	-
Gaps	0.001%	0.001%	-	0.006%	0.002%	-
%GC	40.9%	40.9%	38.9%	40.9%	40.9%	38.3%
Telomeres	30	28	3	24	22	
BUSCO complete	90.5%	90.6%	0.1%	89.0%	90.4%	10.3%
BUSCO duplicated	0.5%	0.5%	0%	1.9%	1.9%	0.4%
Aligned bases H0 vs H1	95788054	-	-	98925323	-	-
Aligned percentage H0 vs H1	93.32%	-	-	97.2%	-	-
Number of SNPs H0 vs H1	196960	-	-	90675	-	-
%TEs	58.9%	59.5%	5.7%	59.3%	59.8%	2.1%

RESULTS AND DISCUSSION

Genome assembly of *M. larici-populina* isolate 08KE26

The genome of *M. larici-populina* isolate 08KE26 was assembled into two haplotype-resolved assemblies, each comprising 18 scaffolds, with total sizes of 101.18 Mb for haplotype 0 (H0) and 101.79 Mb for haplotype 1 (H1), respectively, and a set of unphased contigs totalling 19.68 Mb (Table 3). These values are comparable to the haplotype genome sizes obtained for the reference isolate 98AG31 (100.8 Mb for H0 and 102.64 Mb for H1). The contiguity of the 08KE26 haplotypes are also comparable to 98AG31 haplotype-phased genome, with scaffold N50 values of 5.54 Mb (H0) and 5.38 Mb (H1) for 08KE26, and 5.74 Mb for H0 and 5.42 Mb for H1 in 98AG31. We identified 18 putative centromeres in each haplotype based on analysis of contact maps, while we found 24 and 22 telomeres for H0 and H1 respectively (Table 3).

The maximal scaffold lengths were also comparable between isolates, with the longest reaching 7.8-7.9 Mb in 08KE26 H0 and H1, respectively, compared to the 8.2-8.7 Mb in 98AG31 H0 and H1, respectively (Table 3). The 08KE26 haplotypes are highly complete based on BUSCO assessments, with 89.0% (H0) and 90.4% (H1) of genes identified as complete, which is slightly lower but comparable to 98AG31 (90.5% and 90.6%). The proportion of duplicated BUSCO genes was low in both 08KE26 haplotypes (1.9%), suggesting minimal redundancy. Additionally, the assemblies showed very low levels of gaps (0.006% and 0.002% for H0 and H1, respectively), and a consistent GC content of 40.9%, identical to 98AG31. In contrast to the phased haplotypes, the unphased contigs (19.68 Mb) from 08KE26 were substantially more fragmented (364 contigs), and demonstrated lower completeness (10.2% BUSCO complete), and a lower GC content (38.3%). However, this fraction only represents a minor portion of the total assembly. Together, these results highlight the high contiguity, completeness, and

structural accuracy of the *M. larici-populina* 08KE26 genome assembly, with quality metrics comparable to the *M. larici-populina* 98AG31 reference.

The nearly identical haplotype genome sizes between the asexual isolate 08KE26 and the sexual isolate 98AG31 suggested that clonal propagation in this lineage has not been accompanied by a significant genome expansion. A similar pattern is observed in *Puccinia triticina*, in which long-term clonal lineages maintain genome sizes comparable across isolates and instead showed genomic variation primarily through nuclear exchange events (Sperschneider et al., 2023). This contrasted with *Puccinia striiformis* f. sp. *tritici*, in which comparative analyses of sexual and asexual strains revealed that approximately 50 years of clonal propagation led to a substantial increase in genome size (Schwessinger et al., 2020). These observations showed that genome size stability during clonal evolution is not uncommon in rust fungi and likely reflects lineage-specific dynamics, including mechanisms of genetic variation outside of sexual recombination.

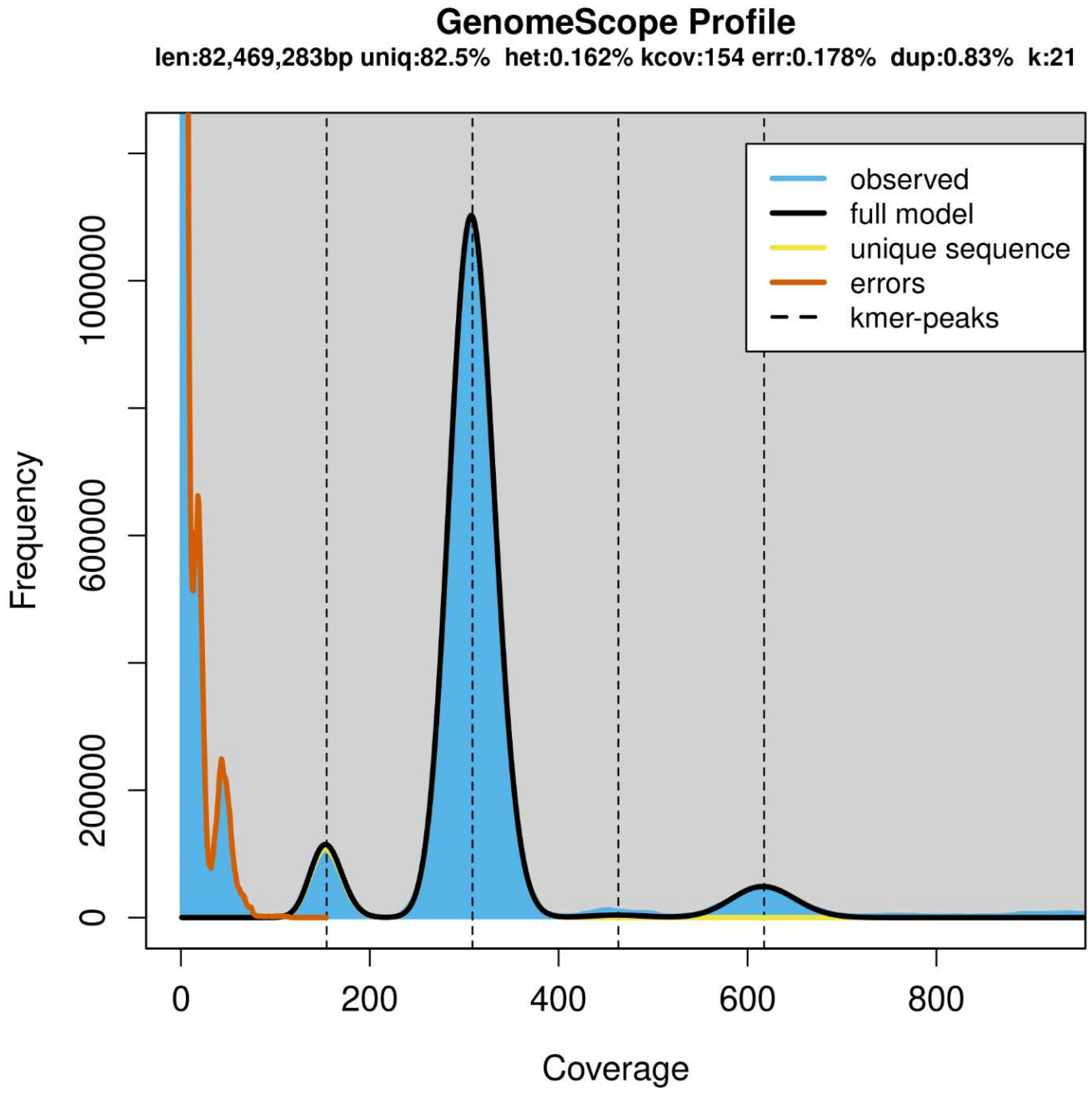


Figure 6. Genome size and heterozygosity estimation from HiFi read k-mer spectra.
GenomeScope 21-mer profiles of HiFi sequencing reads from *M. larici-populina* isolate 08KE26 with an estimated genome heterozygosity (het) of 0.162%, The distinct 1:2 ratio of the k-mer peaks supports the diploid nature this species.

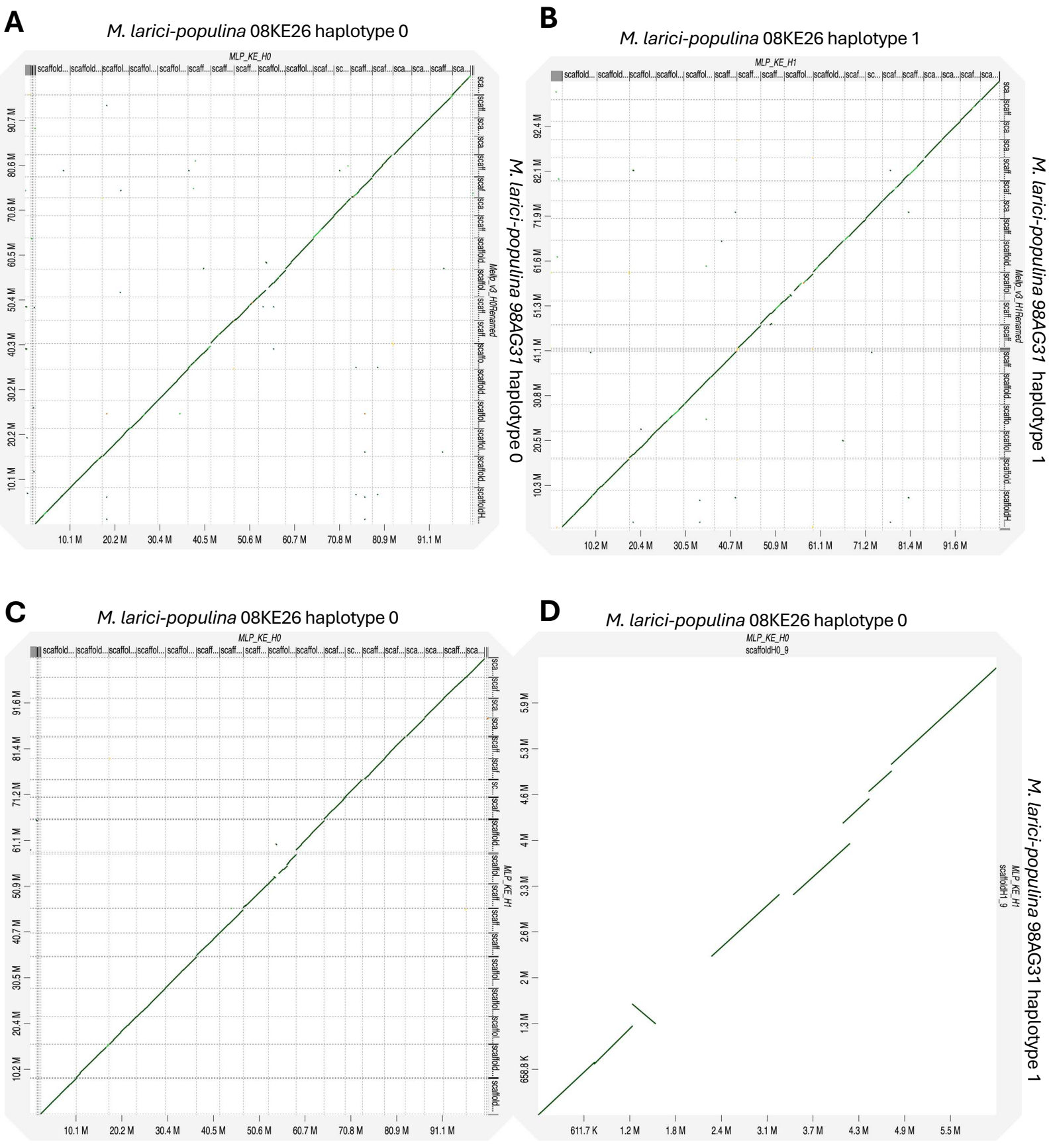


Figure 7. Whole-genome dot plots of *Melampsora larici-populina* 08KE26 haplotype-phased assemblies and comparison to reference isolate 98AG31 genome version 3. A) Synteny dot plot of the newly assembled isolate 08KE26 haplotype 0 (x-axis) against the reference genome version of *M. larici-populina* isolate 98AG31 v3 (y-axis). B) Synteny dot plot of the newly assembled isolate 08KE26 haplotype 1 (x-axis) against the *M. larici-populina* isolate 98AG31 v3 (y-axis). C) Synteny dot plot between haplotype 0 (x-axis) and haplotype 1 of *M. larici-populina* isolate 08KE26 (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between 08KE26 haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~1 Mb non-syntenic and inverted region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*

M. larici-populina isolate 08KE26 exhibits a large inversion at the PR mating-type locus region

Heterozygosity estimated from PacBio reads using a 21-mer frequency profile with GenomeScope (Vurture et al., 2017), was substantially lower in 08KE26 (0.17%) compared to 98AG31 (1.01%), indicating a markedly reduced allelic divergence between haplotypes in 08KE26 (Figure 6). This over 5-fold reduction in allelic divergence between the haplotypes suggests a prolonged history of clonal reproduction, likely leading to haplotype homogenization over time. Consistent with this, whole-genome alignments showed 97.2% sequence conservation between haplotypes in 08KE26, compared to 93.3% in 98AG31, further supporting a reduced haplotypic divergence in the clonal lineage relative to its sexual counterpart (Table 3).

In addition, the whole-genome alignments between the two haplotypes of *M. larici-populina* isolate 08KE26 also revealed strong collinearity across the genome, with the notable exception of a ~1 Mb region on scaffold 9 (Figure 7). This pattern is consistent with previous observations in the isolate 98AG31 and in related rust fungi such as *M. allii-populina* and *M. lini*, all exhibiting disrupted alignments between haplotypes in this region. However, 08KE26 displayed an additional structural rearrangement within this locus, characterized by a large inversion of 307.4 kb in haplotype H0 and 280.3 kb in haplotype H1, followed by substantial alignment gaps spanning 753.3 kb and 691.2 kb in H0 and H1, respectively (Figure 7 C & D). This highly non-syntenic region on scaffold 9 coincides with the pheromone receptor (PR) mating-type locus, which includes key STE3 pheromone receptor genes (Table S5). BLAST searches identified distinct repertoires of STE3 paralogs in the two haplotypes of the isolate 08KE26: STE3.2.2 was found in haplotype H0 though located on an unplaced contig, while STE3.2.3 and STE3.2.4 were located in haplotype H1 on the scaffold 9, approximately 130 kb and 326 kb

downstream of the inversion, respectively. The placement of these genes within a large non-syntenic block and adjacent to an inversion, suggested that extensive structural variation may have underlied haplotype-specific organization of mating-type alleles (Table S5).

The organization of STE3 pheromone receptor genes in *M. larici-populina* 08KE26 mirrors that of the sexual isolate 98AG31, however, 08KE26 exhibits a large inversion upstream of the STE genes, which is absent in 98AG31. Although the inversion does not disrupt the coding sequences themselves, its proximity to the PR locus suggests it could influence gene regulation or recombination dynamics. Comparable structural rearrangements near mating-type loci have been implicated in recombination suppression and sexual system evolution in other fungi. In *Cryphonectria parasitica*, a ~1 Mb inversion flanking the MAT locus was identified to define a large non-recombining region (Hartmann et al., 2025). In contrast, studies in *Microbotryum* anther-smut fungi revealed diverse consequences of mating-type-linked structural variations such as stepwise recombination suppression, gene loss at HD loci, and transitions from bifactorial to unifactorial mating-type determination (Lucotte et al., 2025). The discovery of a large inversion unique to the clonal *M. larici-populina* isolate 08KE26, while absent in the sexual isolate 98AG31, raises the possibility that such structural rearrangements may contribute to the loss of sexual reproduction in clonal lineages potentially through inducing changes in mating-type genes regulation at the pre-mating stage (Hartmann et al., 2021). The availability of a haplotype-resolved reference genome now provides a critical foundation for investigating this hypothesis further.

Transposable elements annotation of *M. larici-populina* 08KE26

TE annotations of *M. larici-populina* 08KE26 revealed that TEs comprise 59.4% and 59.8% of the genome in H0 and H1, respectively (Table S6). These proportions are highly similar to those in the sexual reference isolate 98AG31, which contains 58.9-59.6% TE content in H0 and H1,

respectively. The TE composition in 08KE26 was also consistent across *M. larici-populina* strains, with Class I elements contributing 21.6-22.5% and Class II elements accounting for 31.1-31.4% of each haplotype of 08KE26 (Table S6). Among TE Orders, TIR elements were the most abundant, representing 24.6-24.8% of each haplotype, followed by LTR elements at 15.5-15.8%. Within the LTRs, Gypsy elements were predominant, covering 8.4-8.6% of each haplotype (Tables S7 & S8).

These results highlight that the clonally reproducing *M. larici-populina* strain 08KE26 exhibits nearly identical TE content and composition to its sexually reproducing counterpart 98AG31 suggests that extended clonal propagation in this lineage has not resulted in TE accumulation. Such TE stability contrasts with observations in *P. striiformis* f. sp. *tritici*, in which 50 years of clonal evolution led to an increase of the TE content (Schwessinger et al., 2020). It instead supports findings from experimental and theoretical studies showing that clonal reproduction can limit TE spread due to restricted opportunities for transposition into new genomic backgrounds and the action of host defence mechanisms (Bast et al., 2019). For example, in yeast, asexual lineages evolved higher TE excision rates and showed reduced TE loads over time. Similar dynamics may occur in *M. larici-populina*, where purifying selection or epigenetic silencing could constrain TE proliferation despite the absence of sexual recombination.

FIGURE LEGENDS

Figure 6. Genome size and heterozygosity estimation from HiFi read k-mer spectra. GenomeScope 21-mer profiles of HiFi sequencing reads from *M. larici-populina* isolate 08KE26 with an estimated genome heterozygosity (het) of 0.162%, The distinct 1:2 ratio of the k-mer peaks supports the diploid nature this species.

Figure 7. Whole-genome dot plots of *Melampsora larici-populina* 08KE26 haplotype-phased assemblies and comparison to reference isolate 98AG31 genome version 3. A) Synteny dot plot of the newly assembled isolate 08KE26 haplotype 0 (x-axis) against the reference genome version of *M. larici-populina* isolate 98AG31 v3 (y-axis). B) Synteny dot plot of the newly assembled isolate 08KE26 haplotype 1 (x-axis) against the *M. larici-populina* isolate 98AG31 v3 (y-axis). C) Synteny dot plot between haplotype 0 (x-axis) and haplotype 1 of *M. larici-populina* isolate 08KE26 (y-axis), showing high collinearity with limited structural variation. D) Close-up view of scaffold 9 synteny dot plot between 08KE26 haplotypes 0 (x-axis) and 1 (y-axis), highlighting a ~1 Mb non-syntenic and inverted region. This region corresponds to the mating-type locus containing the pheromone receptor genes, previously characterized in *M. lini*.

Table S5. **Mating type genes homologues**

MAT genes retrieved through blast homology searches against the genomes of *M. larici-populina* 08KE26 haplotype 0 and haplotype 1.

Table S6. **Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Class level.** ClassAnn corresponds to the Class annotation, nbTE corresponds to the number of TE families regarding the

Class,ncovP represents percent of genome coverage and nbcopie represents the number of TE copies regarding the Class.

Table S7. **Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Order level.** OrderAnn corresponds to the Class annotation, nbTE corresponds to the number of TE families regarding the Order, ncovP represents the percent of genome coverage and nbcopie represents the number of TE copies regarding the Order.

Table S8. **Transposable elements annotations for *Melampsora larici-populina* 08KE26 and *Melampsora larici-populina* 98AG31 genomes at the Superfamily level.** SFAnn corresponds to the Superfamily annotation, nbTE corresponds to the number of TE families regarding the Superfamily, ncovP represents the percent of genome coverage and nbcopie represents the number of TE copies regarding the Superfamily.

Chapter III

Impact of transposable elements on the genome organization of three Melampsoraceae species

BACKGROUND

In Chapter II, we have presented the fully phased, near chromosome-scale, genome assemblies for two *Melampsora* species by integrating long-read PacBio sequencing with Hi-C data. These assemblies provide near-complete haplotype resolution, enabling an in-depth view of the genome organization that was not accessible with previous collapsed assemblies. By combining these new reference genomes with the recent published assembly of the genome of the flax rust fungus *Melampsora lini* (Sperchneider et al. 2025), we can investigate the genome organization of three closely related species, from large-scale three-dimensional conformation to the fine-scale influence of repetitive elements on Melampsoraceae genome organization. This investigation completes a long-time goal of the UMR IAM to explore genome organization in rust fungi exhibiting various life cycles in the same taxonomical family, in collaboration with the CSIRO in Australia (Jana Sperschneider and Peter Dodds) and the US DoE Joint Genome Institute (Igor Grigoriev). Indeed, the poplar rust fungi *M. larici-populina* and *M. allii-populina* are both heteroecious species infecting poplar trees and infecting distinct alternating host species (i.e. *Larix* spp. and Alliaceae, respectively), whereas the flax rust fungus *M. lini* is an autoecious species only infecting flax. The three species are all macrocyclic, meaning they all form five distinct spore types along their life-cycle. The combinations of similarities and differences in their life cycles are a unique opportunity to explore the molecular and genomic mechanisms underlying unique features of rust fungi biology.

INTRODUCTION

The spatial conformation of eukaryotic genomes plays a fundamental role in genome function, influencing chromosomal dynamics, gene regulation, and genome stability (Bonev & Cavalli, 2016). Advances in chromosome conformation capture technologies, particularly Hi-C (Lieberman-Aiden et al., 2009), have enabled the reconstruction of chromatin folding at the genome scale, revealing that chromosomes often occupy either discrete territories or adopt polarized Rabl-like configurations (Hoencamp et al. 2021). In fungi, including in Pucciniales species, Hi-C analyses have shown that many species display Rabl-like organization, characterized by centromere clustering (Glavincheska and Lorrain 2025; Torres et al. 2023; Xia et al. 2022). For example, the wheat stripe rust *Puccinia striiformis* f. sp. *tritici* exhibits this centromere clustering at one pole of the nucleus and telomere alignment at the opposite pole (Xia et al. 2022). This polarized arrangement produces distinctive centromere-centromere interaction patterns in Hi-C contact maps, providing a basis for detecting the position of centromeres and further investigating their sequence composition in rust fungi.

Centromeres and telomeres are fundamental chromosomal features that define chromosome structure and ensure genome stability (Osterhage and Friedman 2009). Centromeres mediate the proper segregation of chromosomes during cell division, whereas telomeres cap chromosome ends, protecting them from degradation and preventing end-to-end fusions (Osterhage and Friedman 2009). In many fungi, both centromeric and subtelomeric regions are highly enriched in repetitive DNA, particularly transposable elements (TEs), which can contribute to the establishment of heterochromatin and influence nuclear architecture (Dutta et al. 2025, Seidl et al. 2022). Hi-C-based centromere detection has shown that the centromeres of *Puccinia graminis* f. sp *tritici* are large, spanning approximately 250 kb, and are enriched in young LTR-Gypsy elements (Sperschneider et al. 2021). Interestingly, in *M. lini*, centromeres

are less repetitive than the surrounding genome and do not show enrichment for a specific TE family (Sperschneider et al. 2025a), highlighting potential variation in centromere architecture among rust fungi. Similar to centromeres, multiple studies have also documented strong associations between telomeres and TEs in fungal genomes. In *Saccharomyces cerevisiae*, the Ty5 retrotransposon preferentially integrates into subtelomeric regions linking TE activity to telomere organization (Gai and Voytas 1998). In fungal pathogens such as *Magnaporthe oryzae*, telomere-targeted non-LTR retrotransposons (MoTeRs) insert directly into telomeric repeats, destabilizing chromosome ends and promoting large-scale rearrangements via break-induced replication (Starnes et al. 2012). Leveraging the haplotype-resolved assemblies of *M. larici-populina* and *M. allii-populina*, we can now systematically compare the TE composition of centromeric and subtelomeric regions across Melampsoraceae species to test whether these elements show conserved or specific enrichment patterns.

In plant pathogenic fungi, synteny analyses often reveal large-scale rearrangements such as inversions, translocations, and chromosome fusions, that are associated with TE-rich regions (Faino et al., 2016; Plissonneau et al., 2018). The resolution of haplotype-phased genomes in rust fungi has shown that some species are highly syntenic between haplotypes with no major rearrangements, while others present extensive rearrangements between haplotypes (cf. Chapter II) (Henningsen et al. 2022, 2025; Li et al. 2019). For example, *Puccinia coronata* f. sp. *avenae* isolates derived from sexual populations display a high degree of chromosome-level synteny and no evidence of large translocation. In contrast, some haplotypes from *Puccinia graminis* f. sp. *tritici* and *Puccinia triticina* isolates representing long-term clonal lineages contain extensive translocations between chromosomes. Despite these interchromosomal translocations, all of the available haplotype-resolved rust fungi genome assemblies present 18 chromosomes per haplotype (n + n = 36) with a globally conserved macrosynteny (i.e., the order of genes on homologous chromosomes) with few large inversions, despite size variation (Duan

et al. 2022; Henningsen et al. 2024, 2025; Liang et al., 2023.; Schwessinger et al. 2018; Sperschneider et al. 2025b; Wang et al. 2024).

In this chapter, we investigated the genome architecture of three species in the taxonomical family Melampsoraceae at the large-scale nuclear organization. Using Hi-C data, we reconstructed and compared nuclear conformation, characterized centromeres and telomeres, and assessed potential TE enrichment in these regions. We then performed synteny analyses to identify chromosomal rearrangements between Melampsoraceae genomes and to test the potential role of TEs in structural variation. Together, these analyses revealed how repetitive elements shape genome organization stability.

MATERIALS & METHODS

Genomic, Hi-C data processing, binning, and visualization

The genome assembly and Hi-C data of *M.lini* CH5 were retrieved from (Sperschneider et al., 2025). Both genomes and Hi-C data of *M. larici-populina* 98AG31 and *M. allii-populina* 12AY07 were generated in Chapter II. To perform a Hi-C contact map, the Hi-C reads were first cleaned with Trimmomatic v0.39 (Bolger et al. 2014) in paired-end mode to remove adaptors and low-quality reads. The parameters were settled to a minimum length of 50 nucleotides and nucleotide extremities were cleaned and removed if they had a score of less than (15 LEADING:15 TRAILING:15). The reads were mapped to each haplotype with BWA v 2.2.1 (Jung and Han 2022), and converted to BAM with samtools v 1.21 (Danecek et al. 2021). The Hi-C map of each haplotype was built as previously described in Chapter II. Briefly, the Hi-C map was built with Hicexplorer v3.7.2 (Ramírez et al. 2017), the restriction enzymes were detected and split (hicFindRestSite), the HiC matrices were built with a bin size of 10kb (hicBuildMatrix) before a correction with Iterative Correction and Eigenvector decomposition,

before conversion of the h5 files generated previously into ginteraction files. Finally, the HiC-map visualization is done with hicPlotMatrix. We estimated the coverage of each Hi-C contact map at 10kb and 50kb bins using a homemade script available at https://github.com/lorraincecile/Hi-C_Ztritici_IPO323 (Glavincheska and Lorrain 2025).

Centromeres and telomeres detection

Putative centromeres were detected with a homemade tool. For each chromosome, we estimated the inter-chromosomal contacts across all scaffolds using a 250kb sliding-window approach. We then compared the local inter-chromosomal contacts to the average inter-chromosomal contacts per scaffold to detect the region with a peak of inter-chromosomal contacts value corresponding to the putative centromeres positions. We verified each predicted centromere position by visual inspections of inter-chromosomal Hi-C contact maps. Finally, each centromere was classified as: validated, potentially close and not detected, depending on the manual confirmation. Telomeres were detected as in Chapter II, FindTelomeres.py was used to detect the telomere signatures on both sides of each scaffold (Sperschneider et al. 2025a).

Synteny analysis and visualization

Both haplotypes were used to estimate the mutation divergence between genomes using MAFFT v 7.527 (Rozewicki et al. 2019). We used 0.14 % as the mutation rate (based on the highest mutation rate within the species) for the synteny analysis between haplotypes and species. The haplotypes were aligned using a k-mer-based tool for macrosynteny analysis called nt-synt v1.0.0 (Coombe et al., 2025). First, k-mers present in each haplotype were identified and filtered to retain only those shared between haplotypes. Among these, single-copy shared k-mers were conserved and used as the basis for building a minimizer graph that captures links between adjacent k-mers. Linear paths in this graph were retained as initial

synteny blocks, while unresolved regions were carried over to the next iteration using a smaller window size to identify additional synteny blocks. Finally, all collinear blocks were merged if located within a defined proximity range. The visualisation was made with nt-viz v1.0.0 (Coombe et al. 2025), gaps smaller than 40kb were ignored, and synteny blocks of at least 40kb bases were shown.

TE annotations and comparisons

TE detection and annotation were performed using REPET v3 (Quesneville et al. 2005), followed by manual curation, as described in Corre et al. (2025). A TE family library was independently generated for each haplotype using the TEdenovo pipeline. The TEannot pipeline was then run separately on each haplotype, including both scaffolded chromosomes and unscaffolded contigs. TE annotations were further refined using MCHelper, and redundant elements were clustered using CD-HIT to consolidate TE families. The age of each TE consensus sequence was estimated by the Kimura 2-parameter method as described by Corre et al. (2025). CD-HIT v4.8.1 tool is used to detect the shared and specific TE families between haplotypes and species (Li & Godzik, 2006).

RESULTS

The three-dimensional genome architecture of Melampsoraceae haplotype-resolved genomes reveals a conserved Rabl-like organization

To investigate the 3D genome conformation among the three Melamsporaceae species, we used the Hi-C data generated for genome assemblies (cf. Chapter II). We tested the lowest resolution we could reach to at least 80% of the bins covered by 1000 contacts, to explore the global genome conformation for each haplotype (Rao et al. 2014). For *M. larici-populina*

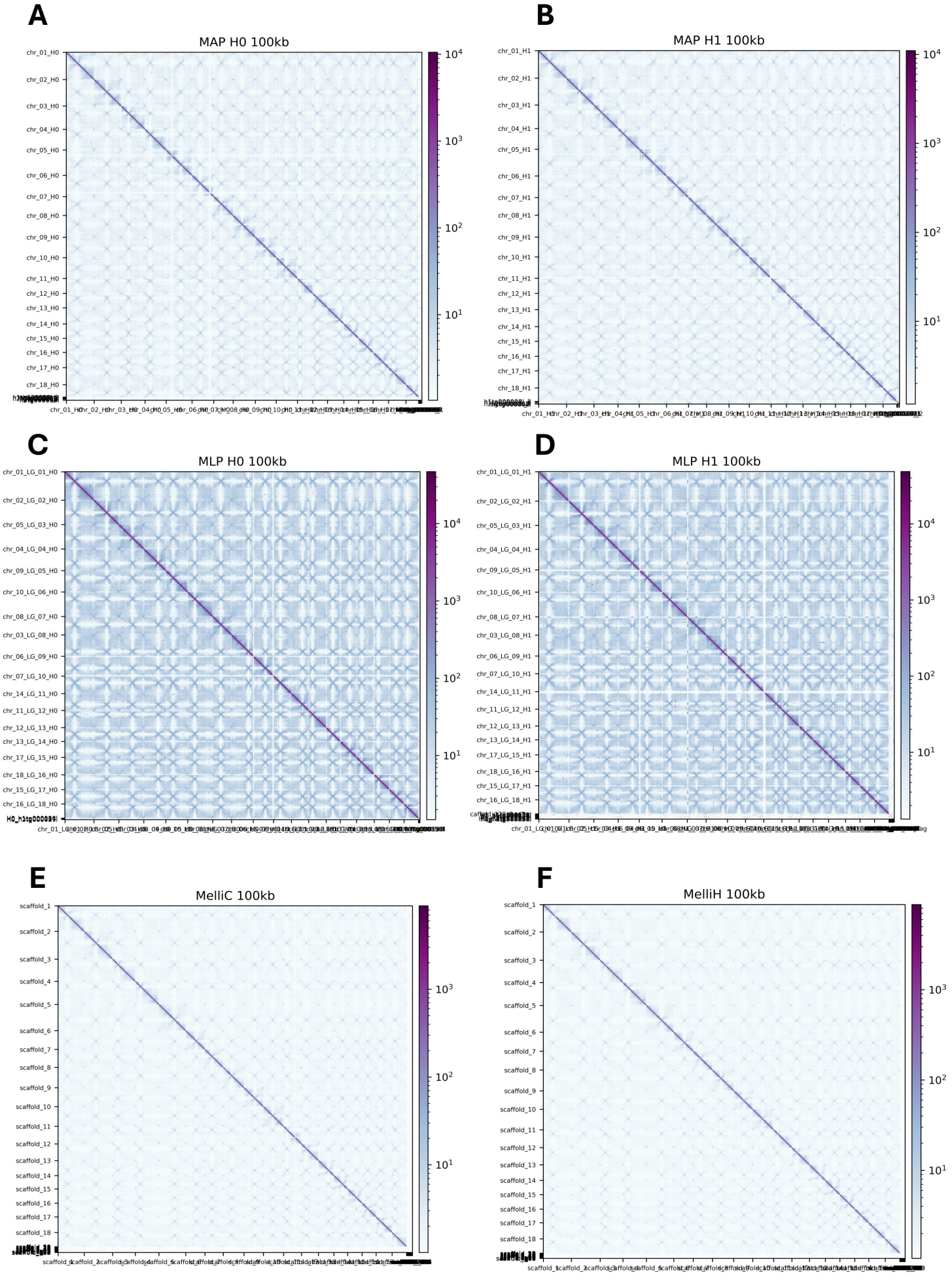


Figure 1. **Genome-wide Hi-C contact maps.** Heatmaps of normalized Hi-C contact frequency at 100 kb bin resolution for: A) *M. allii-populina* haplotype 0, B) *M. allii-populina* haplotype 1, C) *M. larici-populina* haplotype 0, D) *M. larici-populina* haplotype 1, E) *M. lini* haplotype C, and F) *M. lini* haplotype H. The dark purple diagonal indicates intra-chromosomal interactions, while the bow-tie-shaped cross corresponds to the centromeric region.

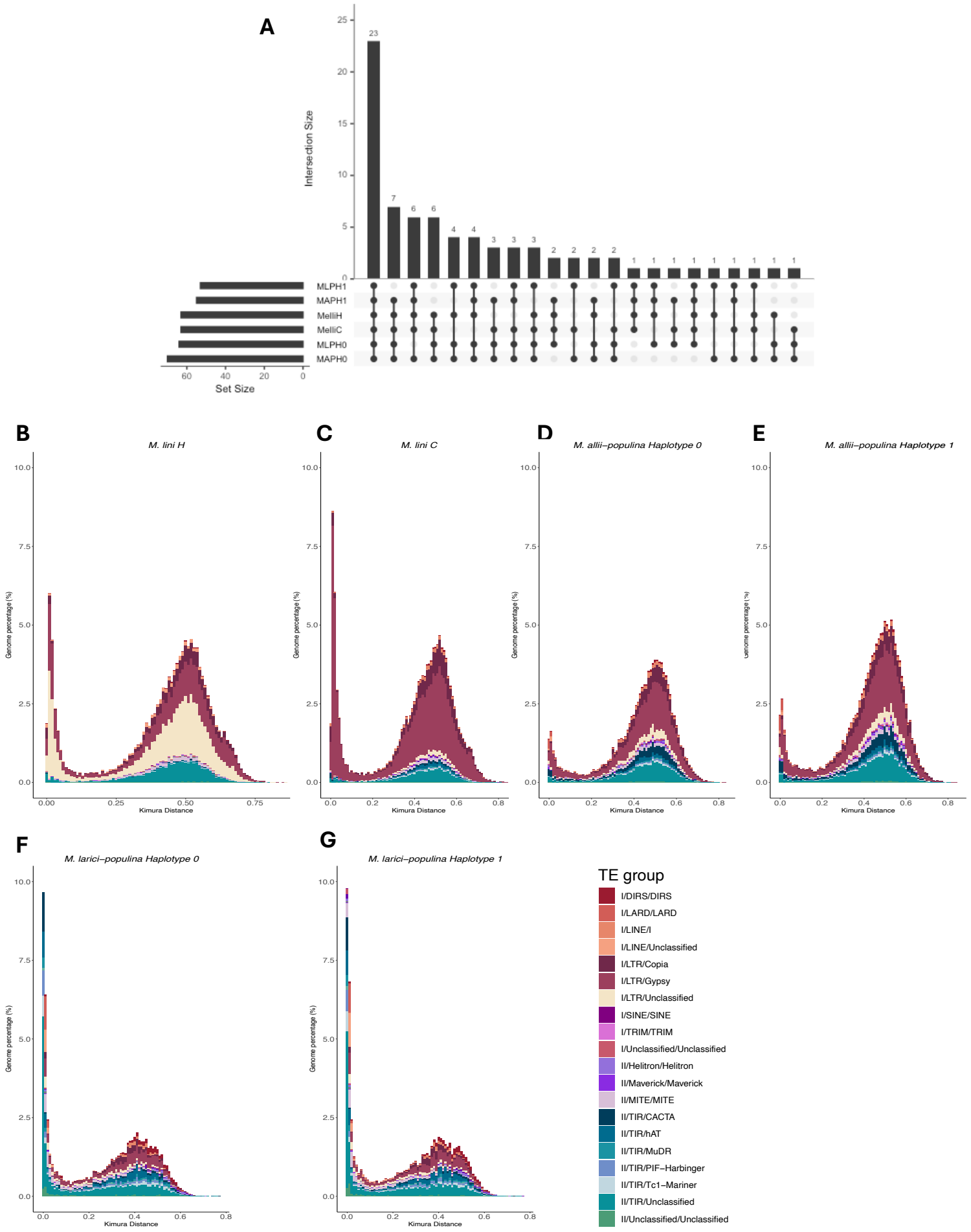


Figure 2. **Transposable elements (TEs) conservation and insertion ages among Melampsoraceae.** A) Shared TE families between the six *Melampsora* haplotypes. Numbers of TE families shared among the three Melampsoraceae genomes based on CDHIT clustering. B-G) Kimura distances across the TE landscapes of the Melampsoraceae genomes of B) *M. lini* haplotype C, C) *M. lini* haplotype H, D) *M. allii-populina* haplotype 0, E) *M. allii-populina* haplotype 1, F) *M. larici-populina* haplotype 0, and G) *M. larici-populina* haplotype 1. Each distribution shows the proportion of the genome (y-axis) composed of copies exhibiting varying degrees of divergence (x-axis).

haplotypes, at 10kb resolution, all of the 18 chromosomes were covered by at least 1000 contacts in more than 90% of the bins (Table S1). For *M. allii-populina* and *M. lini*, we could only

reach such coverage for 50kb and 80% bins, respectively (Table S1). Each haploid genome Hi-C contact matrix was characterized by a predominance of local chromatin interactions, evident as a strong signal along the diagonal of the genome-wide contact maps (Figure 1). We identified an enrichment of the strongest interchromosomal contacts with bowtie-like shapes corresponding to the putative centromeric regions (Figure 1). The overall contact patterns are consistent with clustering of centromeric regions, depicting a Rabl-like chromatin conformation observed across all haplotypes and all three species.

Comparative TE content across Melampsoraceae haplotype-resolved genomes

To further explore the impact of TEs on genome architecture in the three Melampsoraceae species, we performed a refined comparative analysis using the new haplotype-resolved genome assemblies. We clustered TE families to identify elements shared between species and those that are species- or haplotype-specific. This analysis of TE families revealed 64 shared TE family clusters among the three Melampsoraceae species (Figure 2A). Among these shared TE family clusters, we found 17 LTR-Gypsy clusters and 2 TIR-CACTA clusters (Table S2). In contrast, each species harboured thousands of species-specific TE families: *M. lini* (2,663 families), *M. larici-populina* (2,323 families), and *M. allii-populina* (2,258 families), while no haplotype-specific TE families were detected within species (Table S3). Species-specific families were dominated by LTR retrotransposons in *M. allii-populina* (14.7% and 13.4% genome coverage in haplotypes 0 and 1, respectively) and in *M. lini* (43.1% in haplotype C and 42.4% in haplotype H). In contrast, *M. larici-populina* species-specific TEs were enriched in Class II

elements (21.5% and 20.6% genome coverage), particularly TIRs (17.4% and 16.7% in haplotypes 0 and 1, respectively).

Kimura 2-parameter (K2P) divergence landscapes, calculated between individual TE copies and their consensus sequences, revealed broadly similar age distributions between haplotypes within species but marked differences between Melampsoraceae species (Figure 2B-2G). In *M. allii-populina* and *M. lini*, two distinct proliferation waves were detected: an older peak at K2P 0.4-0.6, covering up to ~5% of the genome, and a more recent peak near K2P 0.0, covering ~2.5% of the genome. Both events were dominated by LTR retrotransposons and TIR DNA transposons. In *M. larici-populina*, two waves were also observed, but with contrasting intensities: a smaller, older peak at K2P 0.3-0.5 (<2.5% genome coverage) and a strong recent peak at K2P 0.0 (>9% coverage), driven primarily by TIR elements (Figure 2 F,G). These results confirmed our previous analyses performed with the haplotype-collapsed versions of the assemblies (cf. Chapter 1). We further explored potential haplotype-level differences as we previously found in *M. lini* (cf. Chapter I). In *M. lini*, haplotype C exhibited a recent LTR-Gypsy expansion absent in haplotype H, whereas haplotype H displayed distinct PIF-Harbinger and CACTA peaks at K2P 0.5 that were not detected in haplotype C (Figure 2 B,C). In *M. allii-populina*, haplotype 1 showed slightly more proliferation of old LTRs at K2P 0.5 (>5% genome coverage) compared to haplotype 0 (4%) and a unique LARD burst absent from haplotype 0 (Figure 2 D,E). In *M. larici-populina*, we found that haplotype 0 had a greater TIR expansion, with 9.41% of genome coverage at K2P 0.0 *versus* 8.62% in haplotype 1. This TIR belongs to CACTA and PIF-harbinger with 1.26%, 1.04% of genome coverage for CACTA in respectively haplotype 0 and 1 and 0.07% and 0.81% for PIF-Harbinger (Figure 2 F,G). Overall, the haplotype-resolved TE dynamics largely confirm the patterns identified using collapsed genome assemblies, while providing finer resolution on species- and haplotype-level variation.

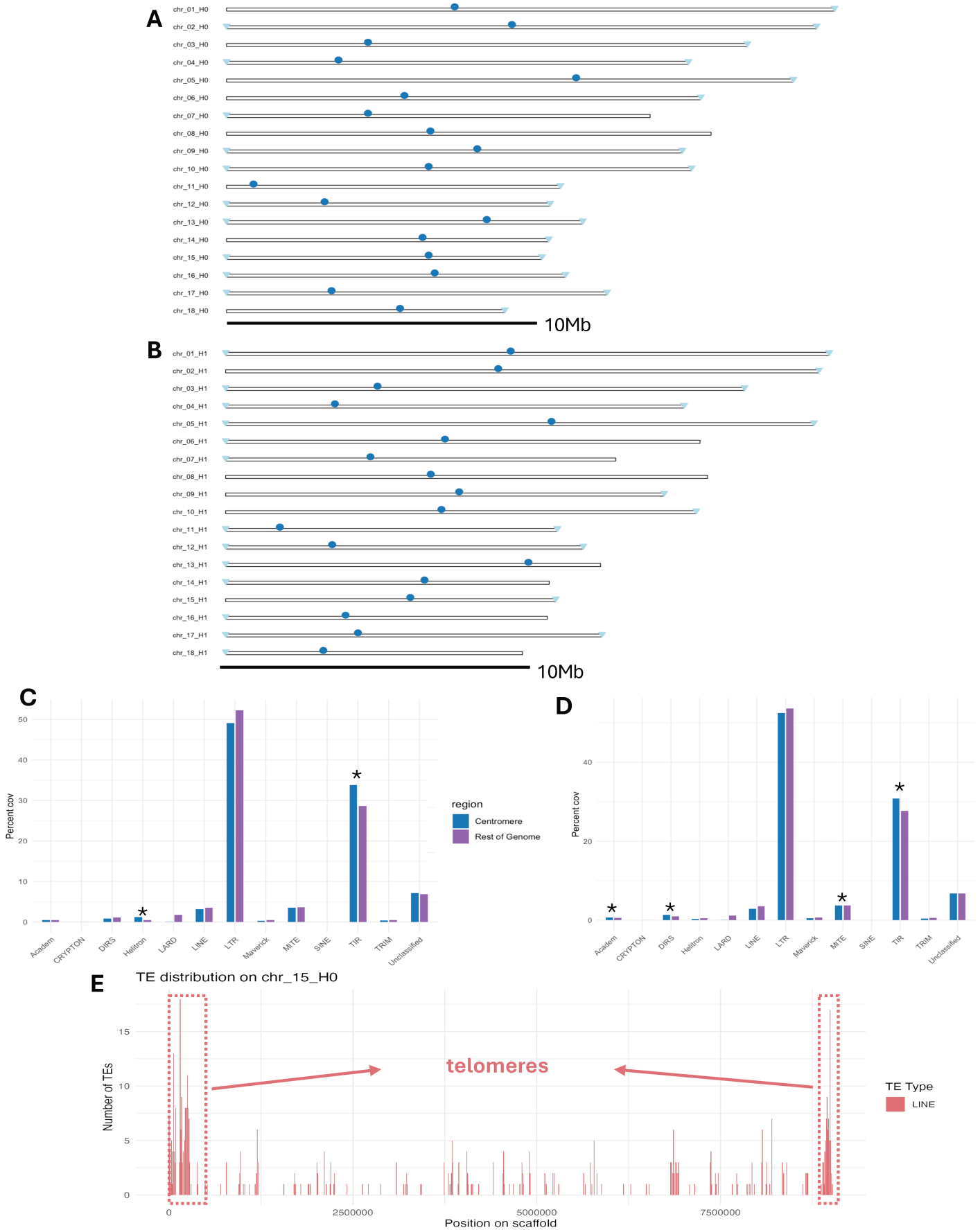


Figure 3: **Repeat content in centromeric and telomeric regions of *Melampsora allii-populina* 12AY07 haplotypes.** A) Karyotype of haplotype 0. Light blue triangles indicate detected telomeres, and dark blue circles indicate centromere positions. B) Karyotype of haplotype 1. Light blue triangles indicate detected telomeres, and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. allii-populina* haplotype 0, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. allii-populina* haplotype 1, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). E) TE LINE coverage along chromosome 15 of *M. allii-populina* haplotype 0. Telomeric ends are enriched in LINE elements.

Centromeric and telomeric regions exhibit conserved TE signatures across Melampsoraceae genomes

We next examined whether TE content patterns extend to functionally distinct chromosomal regions, focusing on centromeric and telomeric regions across all six haplotypes. We mapped the centromeric regions for all six haplotypes, based on interchromosomal Hi-C contact ratios.

We could annotate 18 and 16 centromeres in *M. larici-populina* haplotype 0 and haplotype 1, respectively using the inter-chromosomal contact ratios (Table S4). The remaining centromeres were recovered by visual inspection of the contact maps. For *M. allii-populina*, we only annotated 14 centromere coordinates for each haplotype using the inter-chromosomal contact ratios, due to the lower resolution of the Hi-C matrices at 10kb for this species (Table S4). However, for *M. larici-populina*, 18 chromosomes were detected by visual inspection in both haplotypes. The 18 centromeres of *M. lini* were retrieved from Sperschneider et al. (2025) annotations. We further explored the genomic features of the centromeres within the Melampsoraceae. In *M. larici-populina* and *M. allii-populina,* we found average TE densities of 58% and 88% within the centromeres respectively, whereas the rest of the genome averaged at 61% and 90% respectively, which indicated no enrichment in TEs in centromeric regions (Figure 3 C,D; Figure S1 C,D; Figure S2 C,D; Table S5). However, we observed a significant enrichment of TIR elements in the centromeric regions of all six haplotypes of *M. allii-populina*, *M. larici-populina* and *M. lini,* compared to non-centromeric regions (Figure 3C & 3D, Figure S1C & S1D, Figure S2C & S2D; chi-squared test, BH-adjusted $p < 0.05$). In *M. allii-populina* haplotype 1, additional enrichment was detected for MITE, DIRS, and Academ elements (Figure 3C & 3D; chi-squared test, BH-adjusted $p < 0.05$). In *M. larici-populina*, Helitrons and DIRS were also enriched in the centromeric regions of both haplotypes, while Academ elements were specific to haplotype 1 (Figure S1C & S1D; chi-squared test, BH-adjusted $p < 0.05$). In *M. lini*,

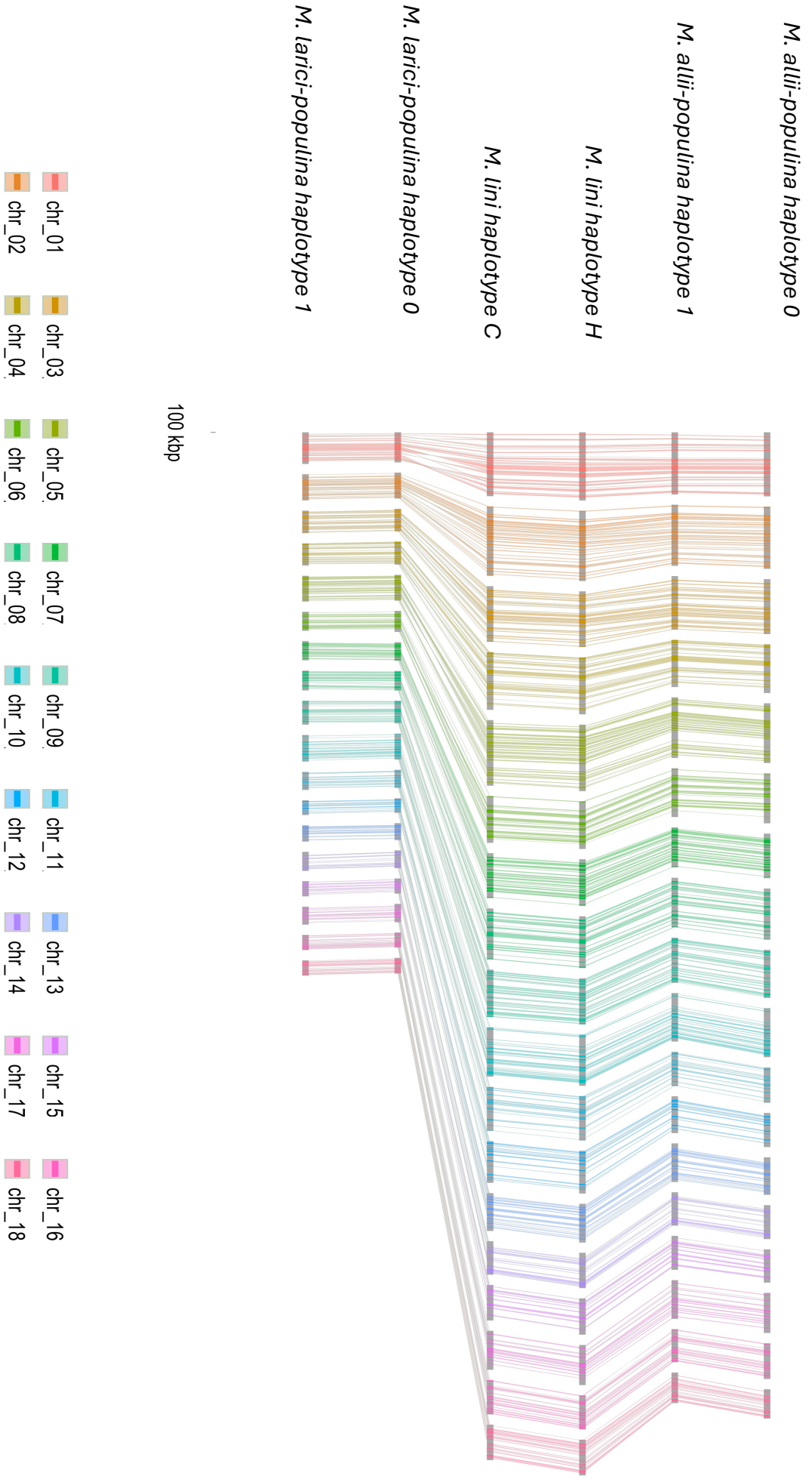


Figure 4: **Comparative synteny analysis between three Melampsoraceae species**
Homologous genomic regions are indicated by colored blocks, revealing conserved macrosynteny as well as lineage-specific rearrangements with a 40kb window (gaps smaller than 40kb are removed).

MITE elements were enriched in centromeric regions of both haplotypes, whereas Helitron, DIRS, and Academ elements were specific to haplotype C, and TRIM and LARD elements were enriched in the centromeric regions of haplotype H (Figure S2C & S2D; chi-squared test, BH-adjusted $p < 0.05$). These results indicate that DNA transposons and particularly TIR elements represent important components of centromeric regions in the Melampsoraceae genomes. Finally, we detected hotspots of LINE elements in the subtelomeric region of 10 and 11 scaffolds in *M. allii-populina* haplotype 0 and haplotype 1, respectively (Figure 3E, Figure S 3,4). These hotspots were also detected in 12 and 15 scaffolds of *M. larici-populina* haplotype 0 and 1, but they did not appear in *M. lini* (Figure S 5,6,7,8). The LINEs that are not detected in *M. lini* could have been disrupted over time and maintained in the other species. This maintenance of LINE could be linked with TE protection of telomere disruptions over time.

A chromosome-scale synteny is conserved in the family Melampsoraceae

We next examined whether TEs are also associated with large-scale structural variation. To this end, we conducted a synteny analysis across the six haplotype-resolved genomes to assess the conservation and rearrangement of chromosomal segments and to evaluate the contribution of TEs to synteny breakpoints and syntenic block reorientations. Even though the haploid genome sizes varied substantially among species, from 100-102 Mb in *M. larici-populina* to 224-219 Mb in *M. allii-populina* and 233-237 Mb in *M. lini*, we identified 718 syntenic blocks representing a total length of ~76.7 Mb per haplotype that are shared by all three species, with an average length of 106.8 kb (median 77.5 kb; Table 1). Among the global chromosome number conserved at 18 per haplotype, six chromosomes (3, 4, 6, 13, 14, and 15) maintained identical block order between species, indicating regions of high structural stability (Figure 4). We still identified substantial synteny block order and orientation variation between species, with 26 instances of order inconsistency and 224 orientation differences, in addition to 440 insertions or deletions of

Table 1: Synteny statistics	
Number_blocks	718
Number_blocks_all_asm	718
Average_coverage (%)	42,26
Average_coverage_all_asm (%)	42,26
Coverage_min_genome_size (%)	49,57
Average_length (bp)	106822,69
Median_length (bp)	77512,25
Total_length (synteny_blocks per number of genomes)	76698694
NG50_length (bp)	0
N50_length (pb)	159826
order inconsistency	26
orientation differences	224
insertions or deletion	440

Table 2: TE density on regions (centromere and synteny break) vs the rest of the genome

	centromere region	no-centromeric region	synteny region	no-synteny region
M. larici-populina haplotype 0	0,55	0,61	0,61	0,60
M. larici-populina haplotype 1	0,60	0,61	0,61	0,64
M. allii-populina haplotype 0	0,89	0,88	0,89	0,92
M. allii-populina haplotype 1	0,91	0,89	0,89	0,92
M. lini haplotype C	0,86	0,91	0,91	0,92
M. lini haplotype H	0,80	0,88	0,88	0,86

syntenic sequence. We found that these structural differences were mostly in *M. larici-populina* compared to the two other closer species, *M. allii-populina* and *M. lini*. Notably, we found four large inversions on chromosome 1 in *M. allii-populina*. However, examination of TE content at synteny breakpoints does not reveal a higher density in the synteny break regions than in the rest of the genome. Indeed, *M. larici-populina* and *M. allii-populina* have, respectively, 0.61% and 0.89% TE coverage within the syntenic region and 0.62% and 0.92% TE density in the non-syntenic region (Table 2). These results indicate the overall conservation of synteny across Melampsoraceae genomes without apparent direct association with TE proliferation, but deeper investigation of synteny breaks and gene content will be needed to better understand the structural differences among Melampsoraceae.

DISCUSSION

Here, we initiated a comparative analysis of genome architecture across three *Melampsora* species using haplotype-phased assemblies and Hi-C contact maps. Our aim was to begin exploring the structural organization of rust fungal genomes and assess the contribution of TEs to their genome architecture and evolution. By integrating 3D genome conformation with repeat annotations and preliminary comparative genomics, we uncovered conserved features among the Melampsoraceae genomes, such as Rabl-like nuclear organization and centromere composition, but we also identified species-specific differences in TE composition and structural variation. While many of these observations remain exploratory, they provide a valuable framework for understanding how TEs and genome architecture intersect in the evolution of *Melampsora* genomes.

Our comparative Hi-C analysis of three Melampsoraceae species offers novel insights into the global organization of fungal genomes, revealing a conserved chromosomal architecture. Despite species-specific differences in contact map resolution, all three genomes exhibited

broadly similar folding patterns, characterized by dominant intra-chromosomal interactions. Notably, we did not detect checkerboard-like patterns in the Hi-C matrices that would suggest A/B compartmentalization. The absence of such bipartite segregation is consistent with previous observations in other fungi (Glavincheska and Lorrain 2025; Torres et al. 2023, 2024), where spatial compartments typical of mammalian genomes are largely lacking (Rao et al. 2014; Szalay et al. 2024). To date, the only fungal species shown to exhibit clear A/B compartmentalization is the arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Yildirir et al. 2022). In contrast to the lack of compartments, all Hi-C contact maps displayed prominent interchromosomal interactions localized to centromeric regions, often appearing as distinct “bowtie”-shaped patterns (Sperschneider et al. 2025a). This signature is consistent with a Rabl-like chromosome conformation, wherein centromeres from multiple chromosomes cluster in close spatial proximity (Cremer et al. 1982). Such centromere clustering, previously documented in other Pucciniales, such as *P. striiformis f. sp. tritici,* and in other filamentous fungi, supports that the Rabl-like conformation is an evolutionarily conserved feature of fungal nuclear architecture (Glavincheska and Lorrain 2025; Torres et al. 2023, 2024; Xia et al. 2022).

Our comparative analysis of TE distribution in three haplotype-phased *Melampsora* spp. genomes reveals both species-specific variation and conserved patterns of local TE accumulation, in agreement with our previous findings across Pucciniales species (Corre et al. 2025). The overall TE landscape across *M. larici-populina*, *M. allii-populina*, and *M. lini* is shaped by distinct dominant TE families, with *M. allii-populina* and *M. lini* enriched in LTR Gypsy retrotransposons, and *M. larici-populina* featuring more abundant TIR DNA transposons. Despite these interspecific differences in TE composition, a conserved chromosomal signature emerges at centromeric and telomeric regions that display distinct and consistent TE profiles across all three species. The *Melampsoraceae* centromeres are enriched in Class II DNA transposons, particularly TIR elements that are significantly overrepresented in centromeric

regions relative to the genome-wide TE distribution in all six haplotypes. This contrasts with the centromeres of other rust fungi, such as *P. graminis* f. sp. *tritici,* where LTR retrotransposons are enriched within centromeric regions (Sperschneider et al. 2021). While we did not find a significant enrichment in repeat coverage within centromeres compared to the rest of the *Melampsora* genomes, the *P. graminis* f. sp. *tritici* centromeres were shown to be highly repetitive, hypermethylated, and divergent even between haplotypes, suggesting a role for epigenetic regulation in maintaining centromere function and stability (Sperschneider et al. 2021). The divergence in centromeric TE composition among rust species highlights the evolutionary plasticity of centromere sequences, even among closely related species. On the other hand, the presence of specific TEs in centromeres of three *Melampsora* spp. genomes suggests these regions may also participate in epigenetic and genome stability mechanisms.

In addition to centromeres, we observed that some subtelomeric regions in *M. larici-populina* and *M. allii-populina* show a hotspot of Class I non-LTR retrotransposons, particularly LINE elements, but not in *M. lini*. Across all six haplotypes analyzed, the terminal segments of specific scaffolds exhibited hotspots of LINEs, in the two poplar rusts compared to the rest of the genome. However, in *M. lini,* this feature was not observed. These LINE hotspots may suggest that chromosome ends could constitute a permissive environment for the retention, or may even represent a preferential insertion hotspot of LINE retrotransposons. Subtelomeric regions in fungi are well-known as hotspots for repeat accumulation and structural variation, playing roles in both genome defense and adaptive evolution. For instance, similar patterns of telomeric TEs have been observed in other fungal species, such as *M. oryzae,* where non-LTR retrotransposons known as MoTeRs insert into telomeric repeat arrays and can destabilize chromosome ends, potentially contributing to chromosomal rearrangements and genome evolution (Starnes et al. 2012). In other model species, such as *Drosophila melanogaster*, telomeres are maintained entirely by specialized non-LTR retrotransposons that substitute

canonical telomeric repeats (Pardue and DeBaryshe 2011). The functional and evolutionary implications of TE hotspots in the subtelomeric regions of the two poplar rust fungi remain to be elucidated. Interestingly, despite the closer phylogenetic relationship between *M. lini* and *M. allii-populina*, only the two poplar rusts exhibit similar subtelomeric TE accumulation. This suggests that lineage-specific genome dynamics may underlie the observed differences in TE local landscape rather than shared evolutionary history.

Whereas all three *Melampsora* genomes share a conserved chromosome number (i.e., 18 chromosomes per haplotype), we observed limited structural divergence, primarily in *M. larici-populina*, which exhibits four large inversions on scaffold 1 relative to *M. allii-populina* and *M. lini*. These rearrangements suggest a degree of species-specific genome evolution in *M. larici-populina*, the most phylogenetically divergent of the three species (Aime and McTaggart 2021; Corre et al. 2025). However, a comparison of syntenic and non-syntenic regions revealed no significant difference in TE content, indicating that large-scale chromosomal rearrangements in the Melampsoraceae are not clearly associated with TE accumulation, but further synteny analyses with Melamsporaceae orthologous genes will be needed to explore this tendency further. The lack of TE association with synteny breaks, contrasts with other fungal systems, such as *Zymoseptoria tritici*, where TEs are often enriched at structural breakpoints and implicated in genome instability (Lorrain et al. 2021). In TE-rich genomes as the rust fungi, the influence of TEs on genome architecture may therefore be more localized, acting primarily at the scale of chromatin organization and fine-scale genome structure rather than driving large interspecific rearrangements. For instance, at the microsyntenic level, Sperschneider et al. (2025) reported that one of the six known *Avr* genes in *M. lini*, *AvrM*, is embedded within a structurally complex region. The avirulent haplotype (H) contains seven *AvrM* paralogs spanning approximately 285 kb, whereas the virulent haplotype (C) carries only a single *avrM* gene. This gene family resides in a highly repetitive genomic context, flanked by a DNA/hAT-Ac transposon

and LINE/TAD1 elements, suggesting that local TE insertions may facilitate gene duplication, deletion, or diversification, particularly in regions relevant to host-pathogen interactions (Sperschneider et al. 2025a). Ongoing gene annotation efforts will be essential to further investigate how TEs contribute to functional genome evolution and to uncover additional loci shaped by TE-mediated dynamics in poplar rust genomes.

FIGURE LEGENDS

Figure 1. **Genome-wide Hi-C contact maps.** Heatmaps of normalized Hi-C contact frequency at 100 kb bin resolution for: A) *M. allii-populina* haplotype 0, B) *M. allii-populina* haplotype 1, C) *M. larici-populina* haplotype 0, D) *M. larici-populina* haplotype 1, E) *M. lini* haplotype C, and F) *M. lini* haplotype H. The dark purple diagonal indicates intra-chromosomal interactions, while the bow-tie-shaped cross corresponds to the centromeric region.

Figure 2. **Transposable elements (TEs) conservation and insertion ages among Melampsoraceae.** A) Shared TE families between the six *Melampsora* haplotypes. Numbers of TE families shared among the three Melampsoraceae genomes based on CDHIT clustering. B-G) Kimura distances across the TE landscapes of the Melampsoraceae genomes of B) *M. lini* haplotype C, C) *M. lini* haplotype H, D) *M. allii-populina* haplotype 0, E) *M. allii-populina* haplotype 1, F) *M. larici-populina* haplotype 0, and G) *M. larici-populina* haplotype 1. Each distribution shows the proportion of the genome (y-axis) composed of copies exhibiting varying degrees of divergence (x-axis).

Figure 3: **Repeat content in centromeric and telomeric regions of *Melampsora allii-populina* 12AY07 haplotypes.** A) Karyotype of haplotype 0. Light blue triangles indicate

detected telomeres, and dark blue circles indicate centromere positions. B) Karyotype of haplotype 1. Light blue triangles indicate detected telomeres, and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. allii-populina* haplotype 0, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. allii-populina* haplotype 1, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). E) TE LINE coverage along chromosome 15 of *M. allii-populina* haplotype 0. Telomeric ends are enriched in LINE elements.

Figure 4: **Comparative synteny analysis between three Melampsoraceae species**
Homologous genomic regions are indicated by colored blocks, revealing conserved macrosynteny as well as lineage-specific rearrangements with a 40kb window (gaps smaller than 40kb are removed).

Figure S1: **Repeat content in centromeric and telomeric regions of *Melampsora larici-populina* 98AG31 haplotypes.** A) Karyotype of haplotype 0. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. B) Karyotype of haplotype 1. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. larici-populina* haplotype 0, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. larici-populina* haplotype 1, separated between non-centromeric regions

(purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$).

Figure S2: **Repeat content in centromeric and telomeric regions of *Melampsora lini* CH5 haplotypes.** A) Karyotype of haplotype C. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. B) Karyotype of haplotype H. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. lini* haplotype C, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. lini* haplotype H, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$).

Figure S3: **Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 0**

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

Figure S4: **Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 1**

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

Figure S5: **Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 0**

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

Figure S6: **Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 1**

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

Figure S7: **Distribution of LINE transposable elements along the scaffolds of *Melampsora lini C***

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold.

Figure S8: **Distribution of LINE transposable elements along the scaffolds of *Melampsora lini H***

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold.

Table 1: **Synteny statistics of the three Melampsoraceae genomes.**

Table 2: **Transposable elements density** in centromeric and no-centromeric regions, in syntenic and no-syntenic regions within the Melampsoraceae genomes.

Table S1: **Hi-C matrix contact coverage per scaffold for each Melampsora spp. genome.**

Table S2: **Transposable elements shared within the three Melampsoraceae at Class, Order, and Superfamily level**. The number of TE family is represented by nbTE, the genome covered by TE families is in the column ncovP and the number of copies in nbcopie.

Table S3: **Specific transposable elements within the three Melampsoraceae at Class, Order, and Superfamily level.** The number of TE family is represented by nbTE, the genome covered by TE families is in the column ncovP and the number of copies in nbcopie.

Table S4: **Centromeric regions annotations.** Centromeres area detected based on inter- vs intra-chromosomal contact ratio, and the “manually” selected centromere are detected by visual Hi-C contact map inspection.

Supplemental figures

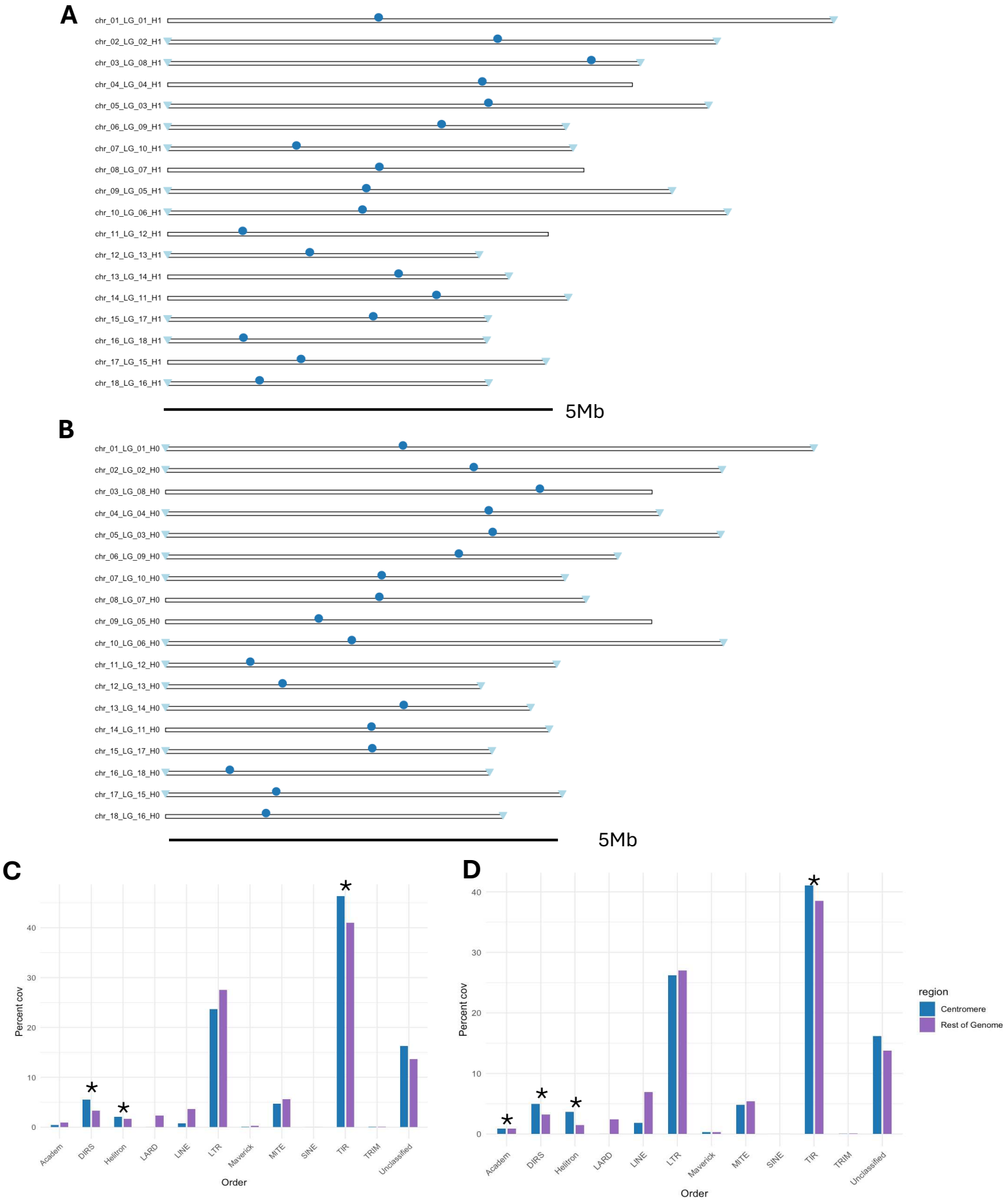


Figure S1: **Repeat content in centromeric and telomeric regions of *Melampsora larici-populina* 98AG31 haplotypes.** A) Karyotype of haplotype 0. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. B) Karyotype of haplotype 1. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. larici-populina* haplotype 0, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. larici-populina* haplotype 1, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$).

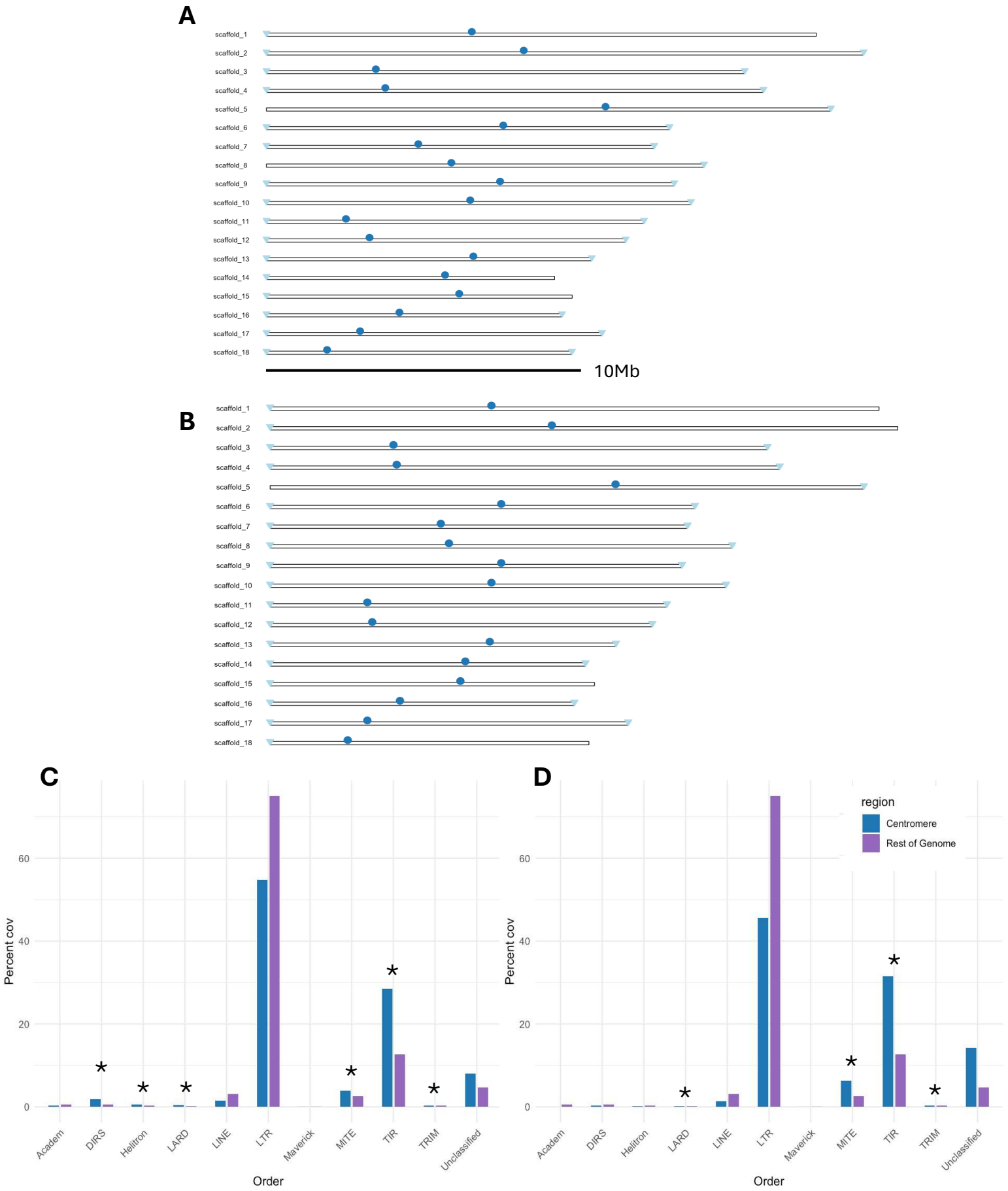


Figure S2: **Repeat content in centromeric and telomeric regions of *Melampsora lini* CH5 haplotypes.** A) Karyotype of haplotype C. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. B) Karyotype of haplotype H. Light blue triangles indicate detected telomeres and dark blue circles indicate centromere positions. C) Percentage coverage of TE density at the Order level in *M. lini* haplotype C, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$). D) Percentage coverage of TE density at the Order level in *M. lini* haplotype H, separated between non-centromeric regions (purple) and centromeric regions (blue). Asterisks indicate TE Orders significantly enriched in centromeric regions (chi-squared test, BH-adjusted $p < 0.05$).

Figure S3: **Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 0**

The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

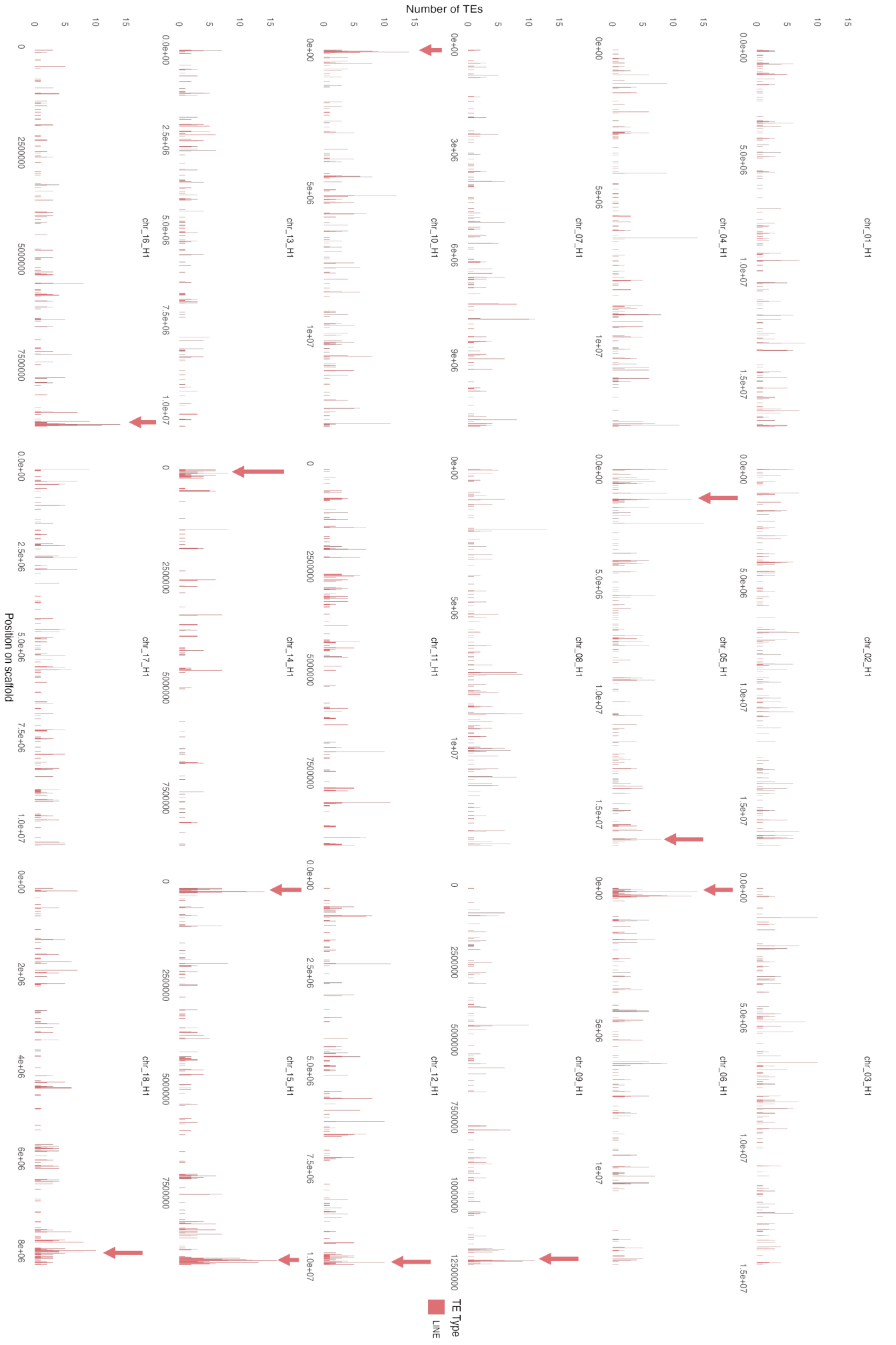


Figure S4: **Distribution of LINE transposable elements along the scaffolds of *Melampsora allii-populina* haplotype 1**
The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

Figure S5: **Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 0**
The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

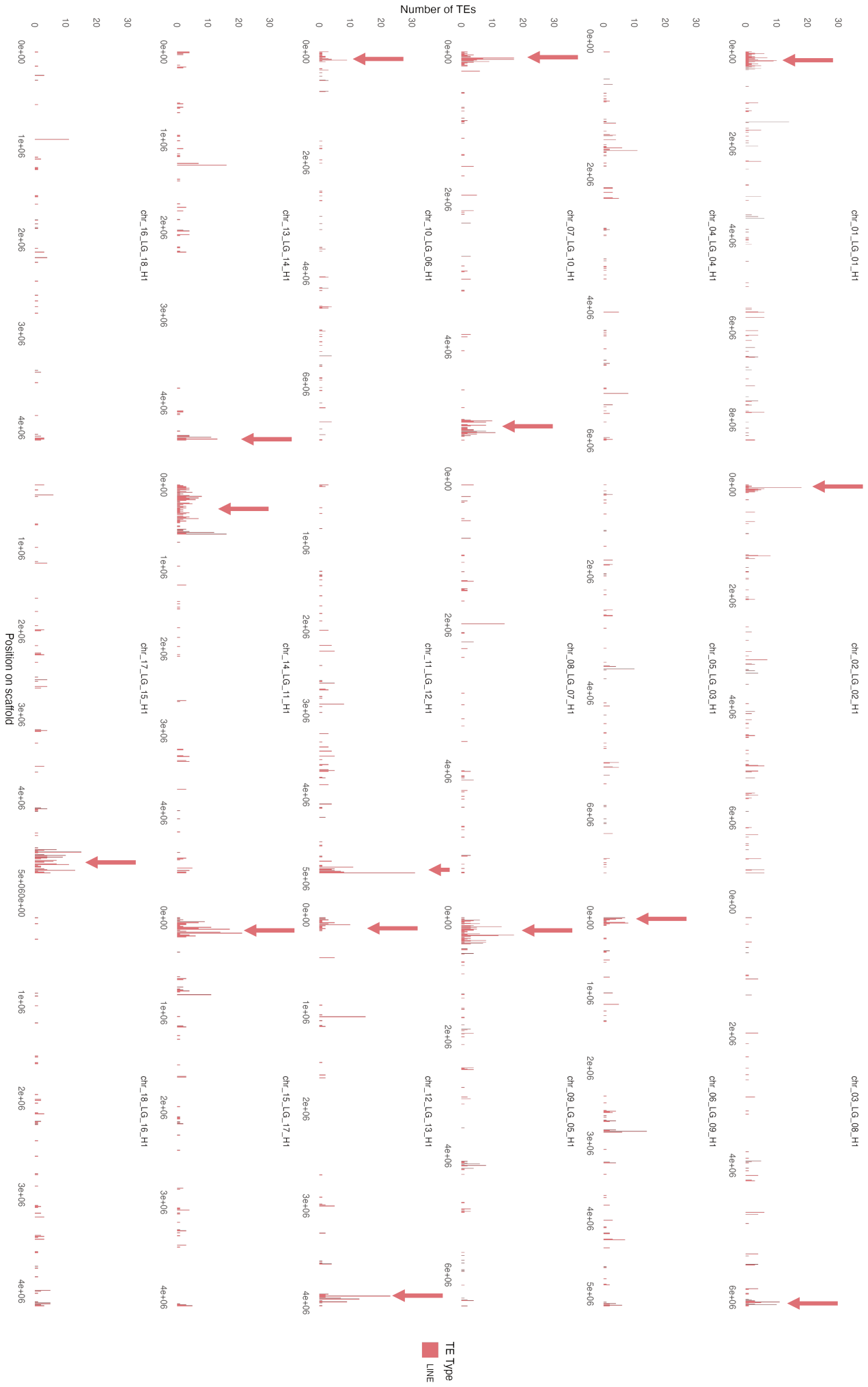


Figure S6: **Distribution of LINE transposable elements along the scaffolds of *Melampsora larici-populina* haplotype 1**
The number of LINE elements was quantified in 10-kb sliding windows across each scaffold. Pink arrows indicate regions of hotspots of LINE insertions.

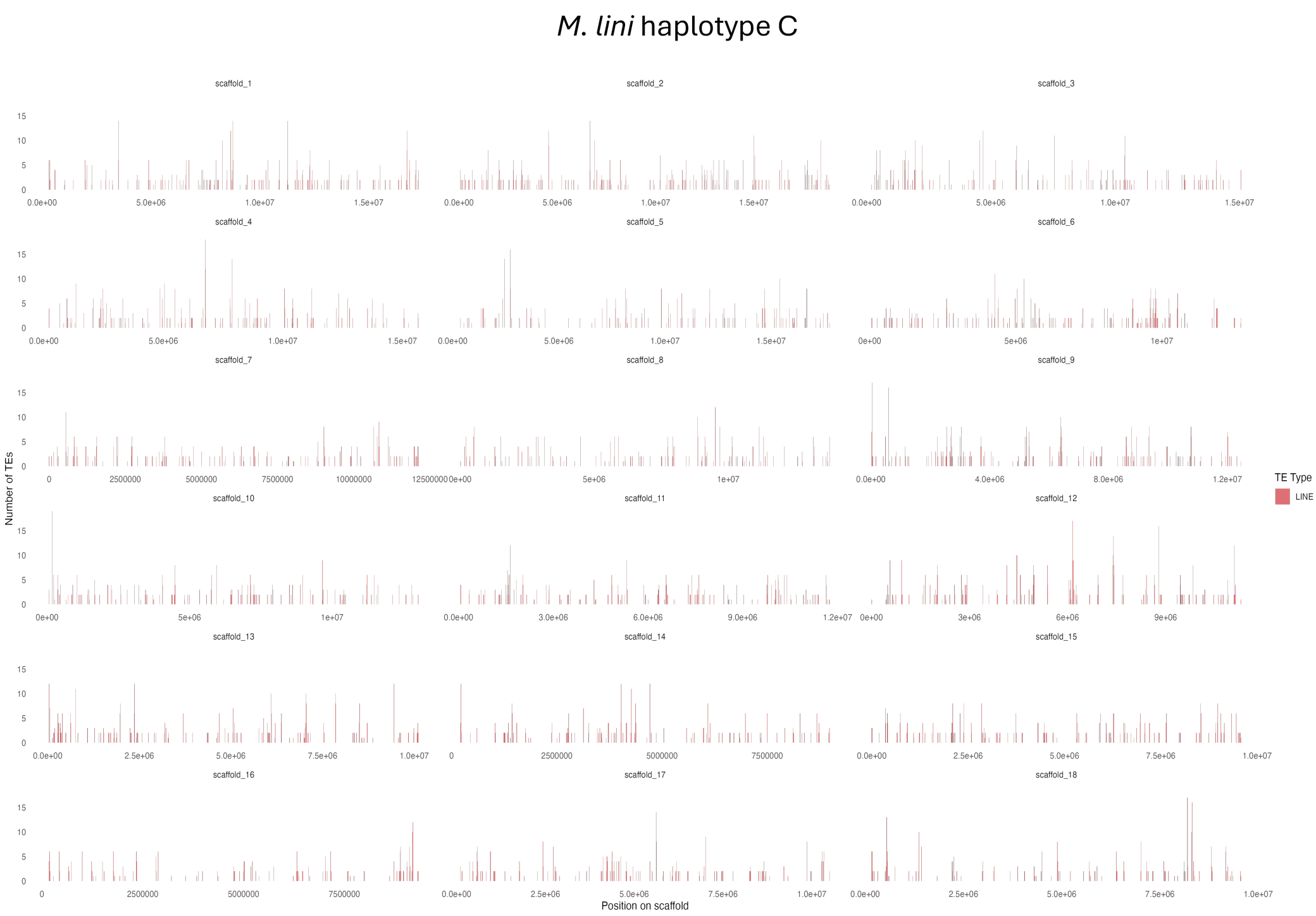


Figure S8: **Distribution of LINE transposable elements along the scaffolds of *Melampsora lini* C**
The number of LINE elements was quantified in 10-kb sliding windows across each scaffold.

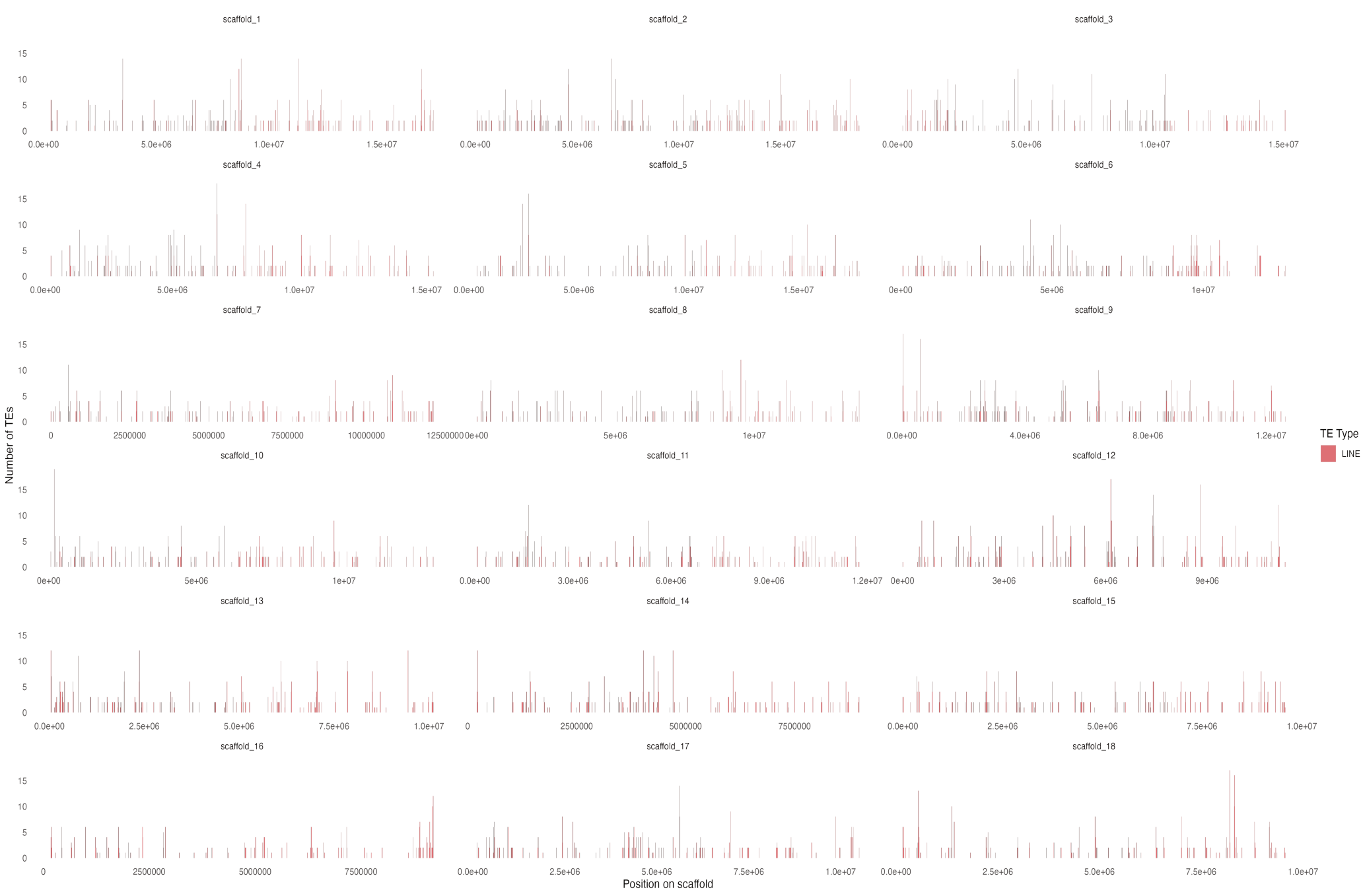


Figure S7: **Distribution of LINE transposable elements along the scaffolds of *Melampsora lini H***
The number of LINE elements was quantified in 10-kb sliding windows across each scaffold.

General Discussion

1) Context and research questions for the study of transposable elements dynamics in rust fungal genomes

The research project presented in this thesis was conducted within the broader framework of the *Ecogenomics of Interactions* team of the Tree-Microbe Interactions Department (UMR IAM) and in the transversal project *Mechanisms and Evolution of the Poplar-Rust Interaction* of the UMR IAM, bridging the three teams of the unit. In particular, the *Ecogenomics* team has developed extensive expertise in fungal genomics, gene annotation and transcriptomics with strong collaboration with the US DoE JGI; whereas in the transversal project, scientists from the different teams bring together their diverse expertise in phytopathology, epidemiology, genomics, population genetics, transcriptomics and biochemistry to study a model forest pathosystem established between poplars and rust fungi. In the transversal project, the genomics angles are performed with strong international collaborations (ETH Zurich, Switzerland; CSIRO Canberra, Australia; US DOE JGI, California, USA). As part of the transversal project, this PhD thesis was established in collaboration between the UMR IAM and the Department of Environmental Systems Science of the ETH Zurich, with strong connections with Dr. Jana Sperschneider at the CSIRO. Over the past decade, collaborative efforts within the UMR IAM have led to the generation of reference genomes for several *Melampsora* species, the identification of candidate effector genes, and transcriptomic analyses of infection stages on both telial and aecial hosts. Yet, despite these advances, a fundamental understanding of the genomic architecture and evolutionary forces shaping rust fungal genomes remained elusive, primarily because of the technical limitations inherent to sequencing and assembling these large and highly repetitive dikaryotic genomes. At the time this thesis was initiated, a growing number of rust genomes were being published, revealing some of the largest fungal genomes known to date, often exceeding several hundred megabases in size. These large genome sizes were

consistently associated with a massive expansion of TEs, in some cases accounting for over 80% of total genome content (Edwards et al. 2022; Gupta et al. 2023; Liang et al. 2023). However, TE annotation in individually analyzed genomes remained either superficial or resulted from different annotation tools, which typically restricted comparisons to global percentage estimates and broad classifications and with little attention given to lineage-specific dynamics or structural impacts on genome organization.

The research questions addressed in this thesis arose directly from these limitations, motivated by a central observation: rust fungal genomes are not only large but structurally complex, and this complexity appears to be driven by TE proliferation. Also, almost all fungi belonging to the taxonomical subphylum Pucciniomycotina, outside of the order Pucciniales, exhibit compact genomes of a lesser repetitive nature. The project was structured around three major axes: (i) a comparative analysis of TE content, diversity, and distribution across available genomes in the order Pucciniales using a standardized annotation pipeline; (ii) the generation and analysis of fully phased and complete genome assemblies for two *Melampsora* species, *M. larici-populina* and *M. allii-populina*, using long-read and Hi-C sequencing; and (iii) the exploration of three-dimensional genome architecture to understand how TEs have shaped genome organization in the taxonomical family Melampsoraceae. These complementary approaches aimed to provide a coherent picture of how TEs interact with genome architecture at multiple scales from sequence to structure, with the long-term goal to determine whether TEs could have impacted the evolution of pathogenicity-related genes in rust genomes. Taken together, the results presented in this thesis contribute significantly to the understanding of genome evolution in rust fungi. Throughout this thesis, we show that TE diversity is both conserved and dynamic across rust lineages, with implications for genome plasticity, gene family expansion, and chromosomal organization. In addition, we highlight that high-quality, haplotype-phased genome assemblies are essential to capture the full complexity of dikaryotic fungal genomes' organization. Finally, by integrating genome structure and nuclear architecture, this work opens

new perspectives on the interplay between TEs, chromatin domains, and the evolutionary stability of large, repeat-rich fungal genomes.

In the following sections, we discuss in more detail the implications of these findings, beginning with the evolution and impact of TEs and their role in shaping genome function, before turning to their broader significance in the context of fungal adaptation and host-pathogen interactions, and finally the technical and conceptual advances in rust genome assembly and structural analysis.

2) The role of transposable elements in advancing the understanding of Pucciniales genome evolution

In the current paradigm, TEs are now recognized as important drivers of genome evolution and innovation, far from the outdated view of them as “junk DNA” (Bourque et al. 2018; Feschotte 2008). In the context of fungal biology, TEs play critical roles in shaping genome size, driving chromosomal rearrangements, and modulating gene regulation and chromatin architecture (Grandaubert et al. 2014; Lorrain et al. 2021; Torres et al. 2021). This thesis contributes to this expanding body of knowledge by elucidating the evolutionary dynamics of TEs in Pucciniales. Through a combination of standardized TE annotation across diverse rust species (Chapter 1) and structural-genomic analyses using fully resolved haplotype assemblies and Hi-C data (Chapter 3), this thesis provides a coherent framework to interpret how TE activity has shaped the architecture and evolutionary trajectory of rust fungal genomes. In Chapter 1, we implemented a systematic annotation pipeline for TE detection, which enabled robust comparative analyses across 12 Pucciniales and related Pucciniomycotina species. These analyses revealed that high TE content is a widespread and ancient characteristic of the order Pucciniales, contrary to other species of the Pucciniomycotina subphylum. Upon further investigation of the composition and evolutionary dynamics of the Pucciniales mobilomes, we

demonstrate that TE accumulation in these genomes reflects both the persistence of ancient elements (i.e. some predating the divergence of Pucciniomycotina) and multiple lineage-specific bursts of activity, particularly involving LTR-Gypsy retrotransposons and TIR DNA transposons. These observations reinforce a model in which TE-driven genome expansion in rust fungi is the result of sustained, recurrent activity rather than a singular proliferation event. In addition, the massive TE accumulation in Pucciniales occurs despite the apparent retention of canonical TE silencing pathways. As shown in Chapter 1, rust fungi possess conserved RNA interference (RNAi) machinery and DNA methyltransferases, suggesting that TE control mechanisms remain operational. This apparent paradox points toward a regulatory equilibrium in which TE activity is not entirely suppressed but may be spatially or temporally modulated. Such a balance likely plays a critical role in shaping genome stability in rust fungal genomes. This is similar to observations in other repeat-rich eukaryotic systems, such as vertebrates and plants, where TEs are epigenetically regulated despite occupying large proportions of the genome (Law and Jacobsen 2010).

The generation of fully phased, telomere-to-telomere haplotype assemblies in Chapter 2 enabled, for the first time in poplar rust fungi, a high-resolution analysis of the spatial distribution of TEs along entire chromosomes. While we showed that global TE density is relatively homogeneous across most chromosomes, a more detailed Order-specific analysis revealed non-random accumulation patterns of a few TE Orders. Specifically, TIR elements are enriched at pericentromeric loci, and LINE retrotransposons seem to preferentially accumulate at subtelomeric ends. These contrasting distribution patterns suggest that distinct TE Orders may be differentially retained in specific chromosomal compartments, possibly TEs contributing to the establishment or maintenance of centromeres and telomeres. Such an organization is consistent with growing evidence in other rust fungi that TEs can contribute to the establishment or maintenance of chromosomal landmarks, including centromeres and telomeres. In both *P.*

striiformis f. sp. *tritici* and *P. graminis* f. sp. *tritici,* centromeres are found to be enriched in LTR retrotransposons and show considerable inter-haplotype sequence divergence, underscoring the dynamic and heterogeneous nature of these chromosomal domains in rust fungi (Sperschneider et al. 2021; Tam et al. 2025). In our study, while gene annotations were not yet available for the newly assembled *Melampsora* haplotypes, the identification of TE enrichment at putative centromeric and subtelomeric regions supports the hypothesis that TEs contribute to defining structural features of the fungal genome. By leveraging curated TE annotation and high-resolution chromosomal reconstructions, the work presented here lays the foundation for future studies exploring how TE-gene interactions contribute to genome regulation and adaptive potential in rust fungi and other filamentous plant pathogens.

3) The role of transposable elements in advancing the understanding of Pucciniales biology

Beyond their role in shaping genome architecture, transposable elements also provide important insights into the broader biology of rust fungi. One interesting finding of this work is that large TE-rich genomes appear to be a common feature across multiple plant-associated biotrophic species. The high TE content observed in Pucciniales mirrors similar expansions found in other obligate fungal pathogens such as *B. graminis*, as well as in beneficial fungal symbionts, including arbuscular and ectomycorrhizal fungi (Looney et al., 2022; Martin et al., 2010; Parlange et al., 2011). This recurrent association between TE expansion and biotrophic or symbiotic lifestyles raises the possibility that TE-rich genomes reflect a broader evolutionary strategy, in which increased genomic plasticity may enable adaptation to specialized ecological niches and long-term host associations. The comparative analysis across rust species presented in Chapter 1 supports the hypothesis that a dynamic TE landscape could contribute to the diversification of gene families in rust fungi, including genes involved in host interaction. Indeed, although gene annotations are not yet available for the newly assembled *Melampsora*

genomes, Sperschneider and collaborators already observed that some haplotype-specific expansions of virulence genes in *M. lini* are associated with the vicinity of TEs (Sperschneider et al. 2025). Once gene predictions become available for the new telomere-to-telomere phased assemblies of *M. larici-populina* and *M. allii-populina*, future work will permit to assess in detail whether multigene families, and particularly those encoding effectors, associate with specific TE-proliferation events. In addition to investigating the structural association of TEs with effector gene families, another promising direction is to explore how TE activity might intersect with gene regulation during host colonization. The life cycle of Pucciniales is among the most complex in the fungal kingdom, with multiple spore types, alternation of sexual and asexual reproductions on two different host plants, and establishment of strictly biotrophic interactions with their hosts (Aime et al., 2018). Such a biological complexity demands a high degree of regulatory control and plasticity, particularly at the host interface, where rust fungi must fine-tune gene expression in response to distinct plant tissues and immune responses (Duplessis et al., 2021). In the soybean rust fungus *P. pachyrhizi*, a recent study has shown that TE expression is induced during host colonization and occurs in parallel with expression of effector genes (Gupta et al. 2023). This co-expression pattern suggests that TEs may influence transcriptional dynamics during infection. Whether similar patterns exist in *Melampsora* species remains unknown, but once full gene annotations become available, it will be possible to analyse infection-stage transcriptomes and investigate potential co-expression patterns between TEs and pathogenicity-related genes in these two poplar rust species.

We provided an additional layer of insights into TE dynamics in Pucciniales biology by the comparative analysis of two haplotype-resolved genomes of *M. larici-populina*, representing a long-term asexual lineage (08KE26) and the sexual reference strain (98AG31) in collaboration with Ammar Abdalrahem, a PhD student who did his doctoral thesis project on the topic of reproductive modes in *M. larici-populina*, concomitantly with the present thesis, in the frame of

the UMR IAM transversal project. A widely held assumption in fungal genome evolution is that clonal reproduction, by reducing opportunities for purifying selection and recombination, might allow for unchecked TE accumulation (Gupta et al. 2023). Surprisingly, our results challenge this view as both strains exhibit remarkably similar levels of TE content, despite the contrasting reproductive histories (Abdalrahem et al. 2025; Abdalrahem 2025). This finding suggests that clonal propagation alone may not be sufficient to drive large-scale TE proliferation in rust fungi, at least over short evolutionary timescales. It also implies that control mechanisms, such as TE repression via epigenetic silencing, may effectively limit TE accumulation even in the absence of sexual recombination. However, additional data across multiple clonal and sexual isolates will be needed to generalize this observation.

Taken together, these findings underscore that TE are not merely drivers of genome expansion in rust fungi but are tightly linked with their biology. TE-rich genomes appear to be a recurrent feature of obligate biotrophic lifestyles, potentially enabling long-term host adaptation through increased genome plasticity. Furthermore, the observation that TE load remains stable across strains with divergent reproductive histories highlights the likely role of regulatory and structural constraints in modulating TE dynamics in Pucciniales. As such, TEs provide a powerful lens through which it is possible to investigate the evolutionary trajectories, host interaction strategies, and regulatory innovations of rust fungi. Future studies integrating gene annotations and infection-related transcriptomes will be essential to fully decipher the functional interplay between TEs and the rust fungal lifestyle.

4) The contribution of phased genome assemblies in advancing the understanding of Pucciniales biology

The study of rust fungi has long been constrained by the biological complexity of these organisms: they are obligate biotrophs with complex life cycles, recalcitrant to *in vitro* culture and genetic manipulation, and capable of infecting multiple host species across distinct life

stages. For decades, these limitations hampered our ability to interrogate their molecular biology and evolutionary dynamics. The advent of high-throughput sequencing technologies, and more recently of long-read, chromatin-conformation capture (Hi-C), and haplotype-phasing strategies, has ushered in a new era in rust fungal genomics (Henningsen et al. 2025). This thesis capitalizes on these technological advances and contributes several foundational resources and conceptual frameworks for the study of rust fungi.

From a methodological perspective, the telomere-to-telomere, haplotype-resolved assemblies of *M. larici-populina* and *M. allii-populina* produced in this work (Chapter 2) constitute a substantial advancement over previous genomic references for rust fungi. Earlier assemblies were more fragmented, collapsed in repeat-rich regions, and unphased, thereby obscuring key features of genome organization and impeding evolutionary comparisons in dikaryotic organisms. The use of PacBio HiFi reads in conjunction with Hi-C scaffolding has enabled the resolution of both haplotypes into complete chromosomal sequences, thereby allowing the identification of centromeres and telomeres, structural rearrangements, and inter-haplotype variation with unprecedented resolution. These assemblies now serve as foundational references for future comparative, functional, and population genomic studies across the order Pucciniales. Notably, the generation of fully phased haplotype-resolved genome assemblies also provides a critical foundation for investigating nuclear dynamics in dikaryotic rust fungi. In the taxonomical family Pucciniaceae, recent genomic studies have demonstrated that non-sexual processes such as somatic nuclear exchange (i.e. also referred to as somatic hybridization) can occur in clonal lineages and significantly contribute to genetic diversity (Henningsen et al. 2024; Li et al. 2019; Sperschneider et al. 2023). For instance, *P. coronata* f. sp. *avenae* populations were shown to undergo intercontinental nuclear exchange events, resulting in new combinations of haplotypes and substantial shifts in fitness and virulence (Henningsen et al. 2024). Similarly, in *P. triticina*, nuclear-phased assemblies revealed that newly emerged lineages arose through the

reassortment of haplotypes between long-standing clonal lineages, suggesting that nuclear shuffling rather than mutation or recombination may be the predominant source of genotypic innovation in these species (Sperschneider et al. 2023). The presence of similar processes in the family Melampsoraceae remains unknown. However, the high-quality phased genomes generated in this thesis establish the necessary genomic resolution to explore this question. Notably, Drs. Fabien Halkett and Pascal Frey (INRAE IAM) have already generated long-read and short-read sequencing data across natural populations of *M. larici-populina*, creating an opportunity to apply haplotype-resolved phylogenomic approaches to test for nuclear exchange or somatic hybridization within this species in the frame of the UMR IAM transversal project.

From a biological perspective, the high-resolution assemblies generated in this thesis have revealed a notable feature of rust fungal genome organization: all telomere-to-telomere assemblies currently available for Pucciniales, including those produced in this work, consistently report 18 chromosomes per haploid genome, suggesting that chromosome number is highly conserved across the order (Henningsen et al. 2025). In the newly assembled poplar rust genomes as well as in *M. lini*, the only structurally divergent region between haplotypes maps to the vicinity of the pheromone receptor locus (Chapter 2)(Sperschneider et al. 2025). In this region, 1-2 Mb fail to align between homologous chromosomes, a pattern that is particularly striking in *M. larici-populina* clonal isolate 08KE26, where the divergent region is additionally associated with a large chromosomal inversion. While gene annotation is currently unavailable for these assemblies, the positional vicinity between the large inversion and the mating-related pheromone receptor genes locus raises the possibility that transposable elements may have contributed to structural instability in this region. These observations are congruent with the hypothesis that TEs mediate local rearrangements, which, in turn, may influence the evolution of mating-type loci, recombination suppression, or nuclear compatibility (Duhamel et al., 2022, 2022; Lucotte et al., 2025). This hypothesis is currently being investigated in an asexual lineage

of *M. larici-populina* population in the context of the doctoral project of Ammar Abdalrahem and more generally in the UMR IAM transversal project (Abdalrahem 2025).

5) Conclusion and Perspectives

In 2011, *M. larici-populina* and *P. graminis* f. sp. *tritici* were the first genomes sequenced in the order Pucciniales, opening a new horizon for exploring the peculiar features of rust fungi. Today, reference genomes are available for just over a dozen genomes, which remains a tiny fraction of the >8,000 species that comprise the Pucciniales order. In particular, this thesis sets out to investigate a central question in rust fungal genomics: how do rust fungi, which are organisms that exhibit obligate biotrophy and seemingly reduced functional autonomy, harbour some of the largest and most structurally complex genomes - and predicted gene complements - in the fungal kingdom? This work contributes to this growing body of work by focusing on the architecture, composition, and regulation of rust fungal genomes, with a particular emphasis on the role of TEs as central agents of genome evolution.

First, by using a comparative approach applied to two closely related *Melampsora* species, we provide important insights into the genomic specificities of Pucciniales. The detailed, curated annotation of TEs across high-quality, chromosome-scale genome assemblies allowed the identification of both conserved and lineage-specific TE families. These annotations revealed the persistence of TE proliferation over evolutionary time and highlighted the occurrence of lineage-specific bursts that may have contributed to local genomic rearrangements. The second major outcome of this thesis is the generation of telomere-to-telomere, haplotype-resolved assemblies for *M. larici-populina* and *M. allii-populina*. These new genome assemblies confirmed that all *Melampsora* genomes examined contain 18 chromosomes, a number consistent with all telomere-to-telomere assemblies published so far in Pucciniales. This suggests a strong chromosomal conservation across the order. However, chromosomal

conservation does not imply stasis: structural variation between haplotypes, including inversions, chromosomal fusions, and genome size expansion, was observed and opens new avenues for research on genome evolution in dikaryotic fungi. In addition, they revealed a consistent pattern of haplotype divergence localized near the mating-related PR locus, including a large inversion in the genome of an isolate belonging to an asexual lineage in the *M. larici-populina* population. This suggests that TEs may have contributed to chromosomal instability or recombination suppression in this region, which may impact mating compatibility. While these results offer a robust foundation, several avenues for future research now emerge.

The phased near T2T genome assemblies of two poplar rust fungi enable a wide range of downstream functional and comparative studies. The top priority will be the integration of gene annotations for the newly assembled *Melampsora* spp. genomes which are currently running at the Joint Genome Institute and that should be released soon. In addition, we will refine the discovery of putative effector genes in collaboration with Dr. Jana Sperschneider at the CSIRO, thanks to a new annotation pipeline she has developed based on deep-learning specifically trained for effector genes (EffectorGeneP, J. Sperschneider Personal communication). The gene prediction by different annotation pipelines will enable us to compare methods and pinpoint which model is the best suited for rust fungal genome annotations. Accurate gene annotation will enable systematic analyses of whether TEs have contributed to structural variation near virulence genes, effector clusters, or mating-type loci. Such efforts should include a focus on haplotype-specific gene presence/absence and the potential role of TEs in shaping gene expression dynamics during host colonization. In addition, the availability of new genome resources paves the way for population-level investigations of TE polymorphisms and genome structural diversity in *Melampsora* species. With short- and long-reads datasets already available from collaborative projects, it will be possible to assess whether nuclear exchanges occur in Melampsoraceae, as observed in Pucciniaceae. The ability to trace haplotype-level

diversity across isolates will offer key insights into the role of clonal propagation, sexual recombination, and possible somatic hybridization in shaping genome dynamics over time. Transcriptomic data generated from infection time-courses in *M. larici-populina* and *M. allii-populina* will be re-analyzed in light of the new genome assemblies and annotation. This will allow for the detection of coordinated expression patterns between TEs and nearby genes, including candidate effectors. While such co-expression does not imply causality, it could suggest a role for TEs in influencing or opportunistic cohabitation in regulatory landscapes. Finally, this work highlights the importance of high-quality TE annotations as well as phased assemblies for functional and evolutionary genomics in dikaryotic fungi. To ensure comparability across future studies, it will be essential for the community to establish shared annotation pipelines, quality metrics, and standards. A concerted effort to improve TE classification in fungal system datasets will provide a common ground for large-scale comparative studies.

Bibliography

Abdalrahem A. (2025) Evolutionary and functional investigations of reproductive mode polymorphism in the poplar rust fungus. Thèse de doctorat de l'Université de Lorraine. Defence expected on October 22nd, 2025.

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

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.

Aime, M. C., & McTaggart, A. R. (2021). A higher-rank classification for rust fungi, with notes on genera. Fungal Systematics and Evolution, 7(1), 21–47. https://doi.org/10.3114/fuse.2021.07.02

Aime, M. C., Bell, C. D., & Wilson, A. W. (2018). Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. Studies in Mycology, 89, 143–152. https://doi.org/10.1016/j.simyco.2018.02.002

Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). Chapter Seven—Phylogenetics and Phylogenomics of Rust Fungi. In J. P. Townsend & Z. Wang (Eds.), Advances in Genetics (Vol. 100, pp. 267–307). Academic Press. https://doi.org/10.1016/bs.adgen.2017.09.011

Aime, M.C., Toome, M., McLaughlin, D.J. (2014). 10 Pucciniomycotina. In: McLaughlin, D., Spatafora, J. (eds) Systematics and Evolution. The Mycota, vol 7A. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55318-9_10

Alonge, M., & Leonard, A. (2021). malonge/RagTag: RagTag v2.1.0 [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.5634263

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biology, 20(1), 224. https://doi.org/10.1186/s13059-019-1829-6

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

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.

Amselem, J., Cornut, G., Choisne, N., Alaux, M., Alfama-Depauw, F., Jamilloux, V., Maumus, F., Letellier, T., Luyten, I., Pommier, C., Adam-Blondon, A.-F., & Quesneville, H. (2019). RepetDB: A unified resource for transposable element references. Mobile DNA, 10(1), 6. https://doi.org/10.1186/s13100-019-0150-y

Ardeljan D, Steranka JP, Liu C, et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. Nat Struct Mol Biol. 2020;27(2):168-178. doi:10.1038/s41594-020-0372-1

Atkinson PW. hAT Transposable Elements. Microbiol Spectr. 2015;3(4):10.1128/microbiolspec.MDNA3-0054-2014. doi:10.1128/microbiolspec.MDNA3-0054-2014

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, 6(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Bargues, N., & Lerat, E. (2017). Evolutionary history of LTR-retrotransposons among 20 Drosophila species. Mobile DNA, 8, 7. https://doi.org/10.1186/s13100-017-0090-3

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

Baril, T., Imrie, R. M., & Hayward, A. (2022). Earl Grey: A fully automated user-friendly transposable element annotation and analysis pipeline (p. 2022.06.30.498289). bioRxiv. https://doi.org/10.1101/2022.06.30.498289

Barreat JGN, Katzourakis A. Phylogenomics of the Maverick Virus-Like Mobile Genetic Elements of Vertebrates. Mol Biol Evol. 2021;38(5):1731-1743. doi:10.1093/molbev/msaa291

Bast, J., Jaron, K. S., Schuseil, D., Roze, D., & Schwander, T. (2019). Asexual reproduction reduces transposable element load in experimental yeast populations. eLife, 8, e48548. https://doi.org/10.7554/eLife.48548

Bell EA, Butler CL, Oliveira C, Marburger S, Yant L, Taylor MI. Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines. Mol Ecol Resour. 2022;22(2):823-833. doi:10.1111/1755-0998.13489

Berthelier, J., Casse, N., Daccord, N., Jamilloux, V., Saint-Jean, B., & Carrier, G. (2018). A transposable element annotation pipeline and expression analysis reveal potentially active

elements in the microalga Tisochrysis lutea. BMC Genomics, 19(1), 378. https://doi.org/10.1186/s12864-018-4763-1

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.

Billiard S, López-Villavicencio M, Hood ME, Giraud T. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. J Evol Biol. 2012;25(6):1020-1038. doi:10.1111/j.1420-9101.2012.02495.x

Boehm, E. W. A., & Bushnell, W. R. (1992). An ultrastructural pachytene karyotype for Melampsora lini. Phytopathology, 82(11), 1212–1218. https://doi.org/10.1094/Phyto-82-1212

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. Genome Biology, 19(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, et al. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc Natl Acad Sci U S A. 2017;114:7067–72.

Bucknell, A. H., & McDonald, M. C. (2023). That's no moon, it's a Starship: Giant transposons driving fungal horizontal gene transfer. Molecular Microbiology. https://doi.org/10.1111/mmi.15118

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner (No. LBNL-7065E). Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US). https://www.osti.gov/biblio/1241166

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., & Martin, C. (2012). Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges[W][OA]. The Plant Cell, 24(3), 1242–1255. https://doi.org/10.1105/tpc.111.095232

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ, 6, e4958. https://doi.org/10.7717/peerj.4958

Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in Neurospora. Science. 1989;244(4912):1571-1575. doi:10.1126/science.2544994

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.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., Jurka, J., Michelmore, R. W., & Dubcovsky, J. (2011). Next generation sequencing provides rapid access to the genome of Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust. PLoS ONE, 6(8), e24230. https://doi.org/10.1371/journal.pone.0024230

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

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

Carr M, Bensasson D, Bergman CM. Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. PLoS One. 2012;7(11):e50978. doi:10.1371/journal.pone.0050978

Casacuberta, E., & Pardue, M.-L. (2005). HeT-A AND TART, TWO DROSOPHILA RETROTRANSPOSONS WITH A BONA FIDE ROLE IN CHROMOSOME STRUCTURE FOR MORE THAN 60 MILLION YEARS. Cytogenetic and Genome Research, 110(0), 152–159. https://doi.org/10.1159/000084947

Castanera R, López-Varas L, Borgognone A, et al. Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet. 2016;12(6):e1006108. Published 2016 Jun 13. doi:10.1371/journal.pgen.1006108

Charmant, O., Gruchota, J., Arnaiz, O., Nowak, K. P., Moisan, N., Zangarelli, C., Bétermier, M., Anielska-Mazur, A., Legros, V., Chevreux, G., Nowak, J. K., & Duharcourt, S. (2024). The PIWI-interacting protein Gtsf1 controls the selective degradation of small RNAs in Paramecium. Nucleic Acids Research, gkae1055. https://doi.org/10.1093/nar/gkae1055

Chen J, Wang Z, Tan K, et al. A complete telomere-to-telomere assembly of the maize genome. Nat Genet. 2023;55(7):1221-1231. doi:10.1038/s41588-023-01419-6

Chénais, B. (2022). Transposable Elements and Human Diseases: Mechanisms and Implication in the Response to Environmental Pollutants. International Journal of Molecular Sciences, 23(5), 2551. https://doi.org/10.3390/ijms23052551

Cheng, H., Concepcion, G.T., Feng, X. et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 18, 170–175 (2021). https://doi.org/10.1038/s41592-020-01056-5

Chomet, P., & Martienssen, R. (2017). Barbara McClintock's Final Years as Nobelist and Mentor: A Memoir. Cell, 170(6), 1049–1054. https://doi.org/10.1016/j.cell.2017.08.040

Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083-1087. doi:10.1126/science.aad5497

Coombe, L., Warren, R. L., and Birol, I. 2025. ntSynt-viz: Visualizing synteny patterns across multiple genomes. 2025.01.15.633221. https://doi.org/10.1101/2025.01.15.633221.

Coradetti ST, Pinel D, Geiselman GM, Ito M, Mondo SJ, Reilly MC, et al. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. eLife. 2018;7:e32110.

Cornelis G, Vernochet C, Carradec Q, et al. Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. Proc Natl Acad Sci U S A. 2015;112(5):E487-E496. doi:10.1073/pnas.1417000112

Corre, E., Morin, E., Duplessis, S., & Lorrain, C. (2025). Ancestral and recent bursts of transposition shaped the massive genomes of plant pathogenic rust fungi. BMC Genomics, 26(1), 627. https://doi.org/10.1186/s12864-025-11726-3

Cosby RL, Chang NC, Feschotte C. Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 2019;33(17-18):1098-1116. doi:10.1101/gad.327312.119

Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V., and Zorn, C. 1982. Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. Hum Genet 60:46–56. https://doi.org/10.1007/BF00281263.

Crowell CR, Wilkerson DG, Beckauri M, Cala AR, McMullen PW, Mondo S, et al. The Melampsora americana Population on Salix purpurea in the Great Lakes Region Is Highly Diverse with a Contributory Influence of Clonality. Phytopathology. 2022;112:907–16.

Cuomo, C. A., Bakkeren, G., Khalil, H. B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J. M., Levin, J. Z., Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B., … Fellers, J. P. (2017). Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. G3 Genes|Genomes|Genetics, 7(2), 361–376. https://doi.org/10.1534/g3.116.032797

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;31:2290–302.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10. https://doi.org/10.1093/gigascience/giab008.

Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. Science, 341(6147), 746–751.https://doi.org/10.1126/science.1236011

Derbyshire, M.C., Newman, T.E., Khentry, Y. et al. Recombination and transposition drive genomic structural variation potentially impacting life history traits in a host-generalist fungal plant pathogen. BMC Biol 23, 110 (2025). https://doi.org/10.1186/s12915-025-02179-x

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.

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: Towards an integrated view of plant–pathogen interactions. Nature Reviews Genetics, 11(8), 539–548. https://doi.org/10.1038/nrg2812

Dodds PN, Chen J, Outram MA. Pathogen perception and signaling in plant immunity. Plant Cell. 2024;36(5):1465-1481. doi:10.1093/plcell/koae020

Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev. 2015;35:57-65. doi:10.1016/j.gde.2015.09.001

Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., Lewis, D., Mago, R., Upadhyaya, N. M., Rathjen, J. P., Stone, E. A., Schwessinger, B., Figueroa, M., Dodds, P. N., Periyannan, S., & Sperschneider, J. (2022). Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. Genome Biology, 23, 84. https://doi.org/10.1186/s13059-022-02658-2

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

Duhamel M, Carpentier F, Begerow D, Hood ME, Rodríguez de la Vega RC, Giraud T. Onset and stepwise extensions of recombination suppression are common in mating-type chromosomes of Microbotryum anther-smut fungi. J Evol Biol. 2022;35(12):1619-1634. doi:10.1111/jeb.13991

Duhamel M, Hood ME, Rodríguez de la Vega RC, Giraud T. Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi. Nat Commun. 2023;14:5692.

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.

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., … Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proceedings of the National Academy of Sciences of the United States of America, 108(22), 9166–9171. https://doi.org/10.1073/pnas.1019315108

Dutta, S., Bhat, K., Aggarwal, R., and Sanyal, K. 2025. Fungi as models of centromere innovation: from DNA sequence to 3-dimensional arrangement. Chromosome Res 33:18. https://doi.org/10.1007/s10577-025-09775-1.

Edwards, R. J., Dong, C., Park, R. F., & Tobias, P. A. (2022). A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, Austropuccinia psidii (p. 2022.04.22.489119). bioRxiv. https://doi.org/10.1101/2022.04.22.489119

Egertová, Z., & Sochor, M. (2017). The largest fungal genome discovered in Jafnea semitosta. Plant Systematics and Evolution, 303(7), 981–986. http://www.jstor.org/stable/44853684

Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics, 9(1), 18. https://doi.org/10.1186/1471-2105-9-18

Elliott, T. A., Heitkam, T., Hubley, R., Quesneville, H., Suh, A., Wheeler, T. J., & The TE Hub Consortium. (2021). TE Hub: A community-oriented space for sharing and connecting tools, data, resources, and methods for transposable element annotation. Mobile DNA, 12(1), 16. https://doi.org/10.1186/s13100-021-00244-0

Ellis JG, Dodds PN, Lawrence GJ. Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. Annu Rev Phytopathol. 2007;45:289-306. doi:10.1146/annurev.phyto.45.062806.094331

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

Evgen'ev MB, Zelentsova H, Shostak N, et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A. 1997;94(1):196-201. doi:10.1073/pnas.94.1.196

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.

Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 2007;41:331-368. doi:10.1146/annurev.genet.40.110405.090448

Flor, H. H. (1971). Current status of the gene-for-gene concept. Annual Review of Phytopathology, 9, 275–296. https://doi.org/10.1146/annurev.py.09.090171.001423

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering Transposable Element Diversification in De Novo Annotation Approaches. PLOS ONE, 6(1), e16526. https://doi.org/10.1371/journal.pone.0016526

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences, 117(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

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.

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

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.

Freitag M, Williams RL, Kothe GO, Selker EU. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in Neurospora crassa. Proc Natl Acad Sci U S A. 2002;99(13):8802-8807. doi:10.1073/pnas.132212899

Frey, P., Gérard, P., Feau, N., Husson, C., & Pinon, J. (2005). Variability and population biology of Melampsora rusts on poplars. In Rust diseases of willow and poplar (pp. 63–72). https://doi.org/10.1079/9780851999999.0063

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.

Gai, X., and Voytas, D. F. 1998. A Single Amino Acid Change in the Yeast Retrotransposon Ty5 Abolishes Targeting to Silent Chromatin. Molecular Cell 1:1051–1055. https://doi.org/10.1016/S1097-2765(00)80105-7.

Gao C, Xiao M, Ren X, et al. Characterization and functional annotation of nested transposable elements in eukaryotic genomes. Genomics. 2012;100(4):222-230. doi:10.1016/j.ygeno.2012.07.004

Garrigues JM, Tsu BV, Daugherty MD, Pasquinelli AE. Diversification of the Caenorhabditis heat shock response by Helitron transposable elements. Elife. 2019;8:e51139. Published 2019 Dec 11. doi:10.7554/eLife.51139

Gladyshev, E. (2017). Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. Microbiology Spectrum, 5(4). https://doi.org/10.1128/microbiolspec.FUNK-0042-2017

Glavincheska, I., and Lorrain, C. 2025. Three-dimensional genome architecture connects chromatin structure and function in a major wheat pathogen. 2025.05.13.653796. https://doi.org/10.1101/2025.05.13.653796.

Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, et al. Giant Starship Elements Mobilize Accessory Genes in Fungal Genomes. Mol Biol Evol. 2022;39:msac109.

Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. Science. 1996;274(5287):546-567. doi:10.1126/science.274.5287.546

Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV. A beginner's guide to manual curation of transposable elements. Mob DNA. 2022;13(1):7. Published 2022 Mar 30. doi:10.1186/s13100-021-00259-7

Grandaubert J, Lowe RG, Soyer JL, et al. Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species

complex of fungal pathogens. BMC Genomics. 2014;15:891. Published 2014 Oct 12. doi:10.1186/1471-2164-15-891

Gregory, P. H. (1948). The Nature and Prevention of Cereal Rusts as Exemplified in the Leaf Rust of Wheat. Nature, 161(4085), 220–221. https://doi.org/10.1038/161220a0

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., & Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Research, 42(Database issue), D699–D704. https://doi.org/10.1093/nar/gkt1183

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.

Gupta, Y. K., Marcelino-Guimarães, F. C., Lorrain, C., Farmer, A., Haridas, S., Ferreira, E. G. C., Lopes-Caitar, V. S., Oliveira, L. S., Morin, E., Widdison, S., Cameron, C., Inoue, Y., Thor, K., Robinson, K., Drula, E., Henrissat, B., LaButti, K., Bini, A. M. R., Paget, E., … Esse, H. P. van. (2022). The soybean rust pathogen Phakopsora pachyrhizi displays transposable element proliferation that correlates with broad host-range adaptation on legumes (p. 2022.06.13.495685). bioRxiv. https://doi.org/10.1101/2022.06.13.495685

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The poplar-poplar rust interaction: Insights from genomics and transcriptomics. Journal of Pathogens, 2011, 716041. https://doi.org/10.4061/2011/716041

Haggerty C, Kretzmer H, Riemenschneider C, et al. Dnmt1 has de novo activity targeted to transposable elements. Nat Struct Mol Biol. 2021;28(7):594-603. doi:10.1038/s41594-021-00603-8

Han, Y., & Wessler, S. R. (2010). MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Research, 38(22), e199. https://doi.org/10.1093/nar/gkq862

Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev. 2012;22(3):191-203. doi:10.1016/j.gde.2012.02.006

Harmon, C. L., Harmon, P. F., Mueller, T. A., Marois, J. J., & Hartman, G. L. (2006). First Report of Phakopsora pachyrhizi Telia on Kudzu in the United States. Plant Disease, 90(3), 380–380. https://doi.org/10.1094/PD-90-0380C

Hartfield M, Keightley PD. Current hypotheses for the evolution of sex and recombination. Integr Zool. 2012;7(2):192-209. doi:10.1111/j.1749-4877.2012.00284.x

Hartmann, F. E., Duhamel, M., Carpentier, F., Hood, M. E., Foulongne-Oriol, M., Silar, P., Malagnac, F., Grognet, P., & Giraud, T. (2021). Recombination suppression and evolutionary strata around mating-type loci in fungi: Documenting patterns and understanding evolutionary and mechanistic causes. The New Phytologist, 229(5), 2470–2491. https://doi.org/10.1111/nph.17039

Hartmann, F. E., Rodríguez de la Vega, R. C., Demené, A., Badet, T., Vernadet, J.-P., Rougemont, Q., Labat, A., Snirc, A., Stauber, L., Croll, D., Prospero, S., Dutech, C., & Giraud, T. (2025). An Inversion Polymorphism Under Balancing Selection, Involving Giant Mobile Elements, in an Invasive Fungal Pathogen. Molecular Biology and Evolution, 42(2), msaf026. https://doi.org/10.1093/molbev/msaf026

Hawksworth DL, Lücking R. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. Microbiol Spectr. 2017;5(4):10.1128/microbiolspec.funk-0052-2016. doi:10.1128/microbiolspec.FUNK-0052-2016

Henningsen, E. C., Hewitt, T., Dugyala, S., Nazareno, E. S., Gilbert, E., Li, F., Kianian, S. F., Steffenson, B. J., Dodds, P. N., Sperschneider, J., and Figueroa, M. 2022. 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 Genes|Genomes|Genetics 12:jkac149. https://doi.org/10.1093/g3journal/jkac149.

Henningsen, E. C., Langlands-Perry, C., Stone, E. A., Dodds, P. N., Figueroa, M., and Sperschneider, J. 2025. Genome Biology of Rust Fungi. Annu Rev Phytopathol. https://doi.org/10.1146/annurev-phyto-121823-085551.

Henningsen, E. C., Lewis, D., Nazareno, E. S., Mangelson, H., Sanchez, M., Langford, K., Huang, Y.-F., Steffenson, B. J., Boesen, B., Kianian, S. F., Liachko, I., Stone, E., Dodds, P. N., Sperschneider, J., & Figueroa, M. (2024). A high-resolution haplotype collection uncovers somatic hybridization, recombination and intercontinental movement in oat crown rust. PLoS Genetics, 20(11), e1011493. https://doi.org/10.1371/journal.pgen.1011493

Hill, R., Leitch, I. J., & Gaya, E. (2021). Targeting Ascomycota genomes: What and how big? Fungal Biology Reviews, 36, 52–59. https://doi.org/10.1016/j.fbr.2021.03.003

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., & Quesneville, H. (2014). PASTEC: An Automatic Transposable Element Classification Tool. PLOS ONE, 9(5), e91929. https://doi.org/10.1371/journal.pone.0091929

Hoencamp, C., Dudchenko, O., Elbatsh, A. M. O., Brahmachari, S., Raaijmakers, J. A., van Schaik, T., Sedeño Cacciatore, Á., Contessoto, V. G., van Heesbeen, R. G. H. P., van den Broek, B., Mhaskar, A. N., Teunissen, H., St Hilaire, B. G., Weisz, D., Omer, A. D., Pham, M., Colaric, Z., Yang, Z., Rao, S. S. P., Mitra, N., Lui, C., Yao, W., Khan, R., Moroz, L. L., Kohn, A., St. Leger, J., Mena, A., Holcroft, K., Gambetta, M. C., Lim, F., Farley, E., Stein, N., Haddad, A.,

Chauss, D., Mutlu, A. S., Wang, M. C., Young, N. D., Hildebrandt, E., Cheng, H. H., Knight, C. J., Burnham, T. L. U., Hovel, K. A., Beel, A. J., Mattei, P.-J., Kornberg, R. D., Warren, W. C., Cary, G., Gómez-Skarmeta, J. L., Hinman, V., Lindblad-Toh, K., Di Palma, F., Maeshima, K., Multani, A. S., Pathak, S., Nel-Themaat, L., Behringer, R. R., Kaur, P., Medema, R. H., van Steensel, B., de Wit, E., Onuchic, J. N., Di Pierro, M., Lieberman Aiden, E., and Rowland, B. D. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science 372:984–989. https://doi.org/10.1126/science.abe2218.

Holden S, Bakkeren G, Hubensky J, Bamrah R, Abbasi M, Qutob D, et al. Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. BMC Biol. 2023;21:233.

Hollister JD, Gaut BS. Population and evolutionary dynamics of Helitron transposable elements in Arabidopsis thaliana. Mol Biol Evol. 2007;24(11):2515-2524. doi:10.1093/molbev/msm197

Hossain, M. M., Sultana, F., Yesmin, L., Rubayet, M. T., Abdullah, H. M., Siddique, S. S., Bhuiyan, M. A. B., & Yamanaka, N. (2024). Understanding Phakopsora pachyrhizi in soybean: Comprehensive insights, threats, and interventions from the Asian perspective. Frontiers in Microbiology, 14. https://doi.org/10.3389/fmicb.2023.1304205

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. Nucleic Acids Research, 44(Database issue), D81–D89. https://doi.org/10.1093/nar/gkv1272

Hung FY, Feng YR, Hsin KT, et al. Arabidopsis histone H3 lysine 9 methyltransferases KYP/SUVH5/6 are involved in leaf development by interacting with AS1-AS2 to repress KNAT1 and KNAT2. Commun Biol. 2023;6(1):219. Published 2023 Feb 24. doi:10.1038/s42003-023-04607-6

Jamilloux, V., Daron, J., Choulet, F., & Quesneville, H. (2017). De Novo Annotation of Transposable Elements: Tackling the Fat Genome Issue. Proceedings of the IEEE, 105(3), 474–481. Proceedings of the IEEE. https://doi.org/10.1109/JPROC.2016.2590833

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. Nature, 444(7117), 323–329. https://doi.org/10.1038/nature05286

Jamieson K, Wiles ET, McNaught KJ, et al. Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. Genome Res. 2016;26(1):97-107. doi:10.1101/gr.194555.115

Jangam D, Feschotte C, Betrán E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. 2017;33(11):817-831. doi:10.1016/j.tig.2017.07.011

Jones JD, Dangl JL. The plant immune system. Nature. 2006;444(7117):323-329. doi:10.1038/nature05286

Jung, Y., and Han, D. 2022. BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics 38:2404–2413. https://doi.org/10.1093/bioinformatics/btac137.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research, 110(1–4), 462–467. https://doi.org/10.1159/000084979

Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008;9(8):605-618. doi:10.1038/nrg2386

Khang CH, Park SY, Lee YH, Valent B, Kang S. Genome organization and evolution of the AVR-Pita avirulence gene family in the Magnaporthe grisea species complex. Mol Plant Microbe Interact. 2008;21(5):658-670. doi:10.1094/MPMI-21-5-0658

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.

Kinsey, J. A., & Helber, J. (1989). Isolation of a transposable element from Neurospora crassa. Proceedings of the National Academy of Sciences of the United States of America, 86(6), 1929–1933. https://doi.org/10.1073/pnas.86.6.1929

Kislev ME. Stem rust of wheat 3300 years old found in Israel. Science. 1982;216(4549):993-994. doi:10.1126/science.216.4549.993

Kojima KK. AcademH, a lineage of Academ DNA transposons encoding helicase found in animals and fungi. Mob DNA. 2020;11:15.

Kojima KK, Jurka J. Crypton transposons: identification of new diverse families and ancient domestication events. Mob DNA. 2011;2(1):12. Published 2011 Oct 19. doi:10.1186/1759-8753-2-12

Krishnan P, Meile L, Plissonneau C, et al. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biol. 2018;16(1):78. Published 2018 Jul 16. doi:10.1186/s12915-018-0543-2

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research. https://doi.org/10.1101/gr.092759.109

Kumar S, Karmakar BC, Nagarajan D, Mukhopadhyay AK, Morgan RD, Rao DN. N4-cytosine DNA methylation regulates transcription and pathogenesis in Helicobacter pylori. Nucleic Acids Res. 2018;46(7):3429-3445. doi:10.1093/nar/gky126

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921. doi:10.1038/35057062

Lavi E, Carmel L. Alu exaptation enriches the human transcriptome by introducing new gene ends. RNA Biol. 2018;15(6):715-725. doi:10.1080/15476286.2018.1429880

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11(3):204-220. doi:10.1038/nrg2719

Lawrence, G. J. (1980). Multiple mating-type specificities in the flax rust Melampsora lini. Science, 209(4455), 501–503. https://doi.org/10.1126/science.209.4455.501

Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, et al. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. Nat Commun. 2019;10:5068.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-1659. doi:10.1093/bioinformatics/btl158

Li Y, Steenwyk JL, Chang Y, et al. A genome-scale phylogeny of the kingdom Fungi. Curr Biol. 2021;31(8):1653-1665.e5. doi:10.1016/j.cub.2021.01.074

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.

Liang, J., Li, Y., Dodds, P. N., Figueroa, M., Sperschneider, J., Han, S., Tsui, C. K. M., Zhang, K., Li, L., Ma, Z., & Cai, L. (2023). Haplotype-phased and chromosome-level genome assembly of Puccinia polysora, a giga-scale fungal pathogen causing southern corn rust. Molecular Ecology Resources, 23(3), 601–620. https://doi.org/10.1111/1755-0998.13739

Lippman Z, Gendrel AV, Black M, et al. Role of transposable elements in heterochromatin and epigenetic control. Nature. 2004;430(6998):471-476. doi:10.1038/nature02651

Liu LY, Charng YC. Genome-wide survey of ds exonization to enrich transcriptomes and proteomes in plants. Evol Bioinform Online. 2012;8:575-587. doi:10.4137/EBO.S10324

Liu B, Zhao M. How transposable elements are recognized and epigenetically silenced in plants?. Curr Opin Plant Biol. 2023;75:102428. doi:10.1016/j.pbi.2023.102428

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., & Kahmann, R. (2015). Fungal effectors and plant susceptibility. Annual Review of Plant Biology, 66, 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Loehrer M, Vogel A, Huettel B, et al. On the current status of Phakopsora pachyrhizi genome sequencing. Front Plant Sci. 2014;5:377. Published 2014 Aug 27. doi:10.3389/fpls.2014.00377

Looney B, Miyauchi S, Morin E, et al. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. New Phytol. 2022;233(5):2294-2309. doi:10.1111/nph.17892

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.

Lorrain, C., Feurtey, A., Möller, M., Haueisen, J., and Stukenbrock, E. 2021. Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. G3 Genes|Genomes|Genetics 11:jkab068. https://doi.org/10.1093/g3journal/jkab068.

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., & Duplessis, S. (2019). Advances in understanding obligate biotrophy in rust fungi. New Phytologist, 222(3), 1190–1206. https://doi.org/10.1111/nph.15641

Lorrain, C., Marchal, C., Hacquard, S., Delaruelle, C., Pétrowski, J., Petre, B., Hecker, A., Frey, P., & Duplessis, S. (2018). The Rust Fungus Melampsora larici-populina Expresses a Conserved Genetic Program and Distinct Sets of Secreted Protein Genes During Infection of Its Two Host Plants, Larch and Poplar. Molecular Plant-Microbe Interactions: MPMI, 31(7), 695–706. https://doi.org/10.1094/MPMI-12-17-0319-R

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol Biol Evol. 2012;29(3):1005-1017. doi:10.1093/molbev/msr282

Lucotte, E. A., Jay, P., Rougemont, Q., Boyer, L., Cornille, A., Snirc, A., Labat, A., Chahine, E., Duhamel, M., Namias, A., Gendelman, J., Ma, W.-J., Hayes, R. K., Baruri, S., Ham, J. P., Perlin, M. H., Hood, M. E., Rodríguez de la Vega, R. C., & Giraud, T. (2025). Repeated loss of function at HD mating-type genes and of recombination in anther-smut fungi. Nature Communications, 16(1), 4962. https://doi.org/10.1038/s41467-025-60222-5

Luo, Z., McTaggart, A., & Schwessinger, B. (2024). Genome biology and evolution of mating-type loci in four cereal rust fungi. PLOS Genetics, 20(3), e1011207. https://doi.org/10.1371/journal.pgen.1011207

Lynch M, Bürger R, Butcher D, Gabriel W. The mutational meltdown in asexual populations. J Hered. 1993;84(5):339-344. doi:10.1093/oxfordjournals.jhered.a111354

Malik, H. S., & Eickbush, T. H. (2001). Phylogenetic Analysis of Ribonuclease H Domains Suggests a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses. Genome Research, 11(7), 1187–1197. https://doi.org/10.1101/gr.185101

Martin, F., Kohler, A., Murat, C. et al. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464, 1033–1038 (2010). https://doi.org/10.1038/nature08867

Matsushima W, Duc J, Sheppard S, Pulver C, Grun D, Offner S, Raclot C, Planet E, Trono D. Zinc-finger proteins with a co-opted capsid domain anchor nucleosomes over transposon sequences. bioRxiv. 2025;2025.03.03.638093. doi:10.1101/2025.03.03.638093

Maxwell PH. Diverse transposable element landscapes in pathogenic and nonpathogenic yeast models: the value of a comparative perspective. Mob DNA. 2020;11:16. Published 2020 Apr 21. doi:10.1186/s13100-020-00215-x

McClintock, B. (1951). Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology, 16, 13–47. https://doi.org/10.1101/sqb.1951.016.01.004

McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A., & Solomon, P. S. (2019). Transposon-Mediated Horizontal Transfer of the Host-Specific

Virulence Protein ToxA between Three Fungal Wheat Pathogens. mBio, 10(5), e01515-19. https://doi.org/10.1128/mBio.01515-19

McCullers TJ, Steiniger M. Transposable elements in Drosophila. Mob Genet Elements. 2017;7(3):1-18. Published 2017 Apr 19. doi:10.1080/2159256X.2017.1318201

Mentlak TA, Kombrink A, Shinya T, et al. Effector-mediated suppression of chitin-triggered immunity by magnaporthe oryzae is necessary for rice blast disease. Plant Cell. 2012;24(1):322-335. doi:10.1105/tpc.111.092957

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.

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.

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

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.

Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, et al. Widespread adenine N6-methylation of active genes in fungi. Nat Genet. 2017;49:964–8.

Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine Database (GOLD) v.8: overview and updates. Nucleic Acids Res. 2021;49(D1):D723-D733. doi:10.1093/nar/gkaa983

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., & Ginalski, K. (2019). Transposable elements contribute to fungal genes and impact fungal lifestyle. Scientific Reports, 9(1), Article 1. https://doi.org/10.1038/s41598-019-40965-0

Nai YS, Huang YC, Yen MR, Chen PY. Diversity of Fungal DNA Methyltransferases and Their Association With DNA Methylation Patterns. Front Microbiol. 2021;11:616922. Published 2021 Jan 22. doi:10.3389/fmicb.2020.616922

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

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.

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

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.

Omrane S, Audéon C, Ignace A, et al. Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen Zymoseptoria tritici. mSphere. 2017;2(5):e00393-17. Published 2017 Oct 25. doi:10.1128/mSphere.00393-17

Orozco-Arias, S., Candamil, M. S., Jaimes, P. A., Cristancho, M., Tabares-Soto, R., & Guyot, R. (2022). Composition and Diversity of LTR Retrotransposons in the Coffee Leaf Rust Genome (Hemileia vastatrix). Agronomy-Basel, 12(7), 1665. https://doi.org/10.3390/agronomy12071665

Orozco-Arias, S., Sierra, P., Durbin, R., & González, J. (2024). MCHelper automatically curates transposable element libraries across eukaryotic species (p. 2023.10.17.562682). bioRxiv. https://doi.org/10.1101/2023.10.17.562682

Osterhage, J. L., and Friedman, K. L. 2009. Chromosome End Maintenance by Telomerase*. Journal of Biological Chemistry 284:16061–16065. https://doi.org/10.1074/jbc.R900011200.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biology, 20(1), 275. https://doi.org/10.1186/s13059-019-1905-y

Pardue, M.-L., and DeBaryshe, P. G. 2011. Retrotransposons that maintain chromosome ends. Proceedings of the National Academy of Sciences 108:20317–20324. https://doi.org/10.1073/pnas.1100278108.

Parlange, F., Oberhaensli, S., Breen, J., et al. (2011). A major invasion of transposable elements accounts for the large size of the Blumeria graminis f. sp. tritici genome. Functional & Integrative Genomics, 11, 671–677. https://doi.org/10.1007/s10142-011-0240-5

Pastuzyn ED, Day CE, Kearns RB, et al. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell. 2018;172(1-2):275-288.e18. doi:10.1016/j.cell.2017.12.024

Pegg, G., Taylor, T., Entwistle, P., Guymer, G., Giblin, F., & Carnegie, A. (2017). Impact of Austropuccinia psidii (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. PLoS ONE, 12(11), e0188058. https://doi.org/10.1371/journal.pone.0188058

Petre, B., Joly, D. L., & Duplessis, S. (2014). Effector proteins of rust fungi. Frontiers in Plant Science, 5, 416. https://doi.org/10.3389/fpls.2014.00416

Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, et al. Sex and parasites: genomic and transcriptomic analysis of Microbotryum lychnidis-dioicae, the biotrophic and plant-castrating anther smut fungus. BMC Genomics. 2015;16:461.

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S., & Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: The construction of a linkage genetic map of the poplar rust fungus Melampsora larici-populina. Frontiers in Plant Science, 5, 454. https://doi.org/10.3389/fpls.2014.00454

Persoons A, Maupetit A, Louet C, Andrieux A, Lipzen A, Barry KW, et al. Genomic Signatures of a Major Adaptive Event in the Pathogenic Fungus Melampsora larici-populina. Genome Biol Evol. 2022;14:evab279.

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Silva, C. D., Poulain, J., Delaruelle, C., Martin, F., Rouhier, N., Kohler, A., & Duplessis, S. (2012). RNA-Seq of Early-Infected Poplar Leaves by the Rust Pathogen Melampsora larici-populina Uncovers PtSultr3;5, a Fungal-Induced Host Sulfate Transporter. PLOS ONE, 7(8), e44408. https://doi.org/10.1371/journal.pone.0044408

Petre B, Joly DL, Duplessis S. Effector proteins of rust fungi. Front Plant Sci. 2014;5:416. Published 2014 Aug 20. doi:10.3389/fpls.2014.00416

Petre B, Duplessis S. A decade after the first Pucciniales genomes: A bibliometric snapshot of (post) genomics studies in three model rust fungi. Front Microbiol. 2022;13:989580. Published 2022 Sep 14. doi:10.3389/fmicb.2022.989580

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.

Quesneville, H. (2020). Twenty years of transposable element analysis in the Arabidopsis thaliana genome. Mobile DNA, 11(1), 28. https://doi.org/10.1186/s13100-020-00223-x

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLOS Computational Biology, 1(2), e22. https://doi.org/10.1371/journal.pcbi.0010022

Quesneville, H., Nouaud, D., & Anxolabéhère, D. (2003). Detection of New Transposable Element Families in Drosophila melanogaster and Anopheles gambiae Genomes. Journal of Molecular Evolution, 57(1), S50–S59. https://doi.org/10.1007/s00239-003-0007-2

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

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A., & Manke, T. (2018a). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications, 9(1), 189. https://doi.org/10.1038/s41467-017-02525-w

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.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., and Aiden, E. L. 2014. A 3D map of the

human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet TIG. 2000;16:276–7.

Rizzon C, Marais G, Gouy M, Biémont C. Recombination rate and the distribution of transposable elements in the Drosophila melanogaster genome. Genome Res. 2002;12(3):400-407. doi:10.1101/gr.210802

Romeijn, J., Bañales Belaunde, I., & Seidl, M. F. (2025). Extensive horizontal transfer of transposable elements shape fungal mobilomes. bioRxiv. https://doi.org/10.1101/2025.06.16.659975

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Research 47:W5–W10. https://doi.org/10.1093/nar/gkz342.

Rubin EM. Genomics of cellulosic biofuels. Nature. 2008;454(7206):841-845. doi:10.1038/nature07190

Saito K, Nishida KM, Mori T, et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 2006;20(16):2214-2222. doi:10.1101/gad.1454806

Salcedo A, Rutter W, Wang S, et al. Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. Science. 2017;358(6370):1604-1606. doi:10.1126/science.aao7294

Scarano C, Veneruso I, De Simone RR, Di Bonito G, Secondino A, D'Argenio V. The Third-Generation Sequencing Challenge: Novel Insights for the Omic Sciences. Biomolecules. 2024;14(5):568. Published 2024 May 10. doi:10.3390/biom14050568

Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. Molecular Ecology, 28(6), 1537–1549. https://doi.org/10.1111/mec.14794

Schwessinger, B., Chen, Y.-J., Tien, R., Vogt, J. K., Sperschneider, J., Nagar, R., McMullan, M., Sicheritz-Ponten, T., Sørensen, C. K., Hovmøller, M. S., Rathjen, J. P., & Justesen, A. F. (2020). Distinct Life Histories Impact Dikaryotic Genome Evolution in the Rust Fungus Puccinia striiformis Causing Stripe Rust in Wheat. Genome Biology and Evolution, 12(5), 597–617. https://doi.org/10.1093/gbe/evaa071

Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., Dodds, P. N., Figueroa, M., Park, R. F., and Rathjen, J. P. 2018. A Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus Puccinia striiformis f. sp. tritici Reveals High Interhaplotype Diversity. mBio 9:e02275-17. https://doi.org/10.1128/mBio.02275-17.

Shattock, R. C., & Preece, T. F. (2000). Tranzschel revisited: Modern studies of the relatedness of different rust fungi confirm his law. Mycologist, 14(3), 113–117. https://doi.org/10.1016/S0269-915X(00)80086-5

Shirke MD, Mahesh HB, Gowda M. Genome-Wide Comparison of Magnaporthe Species Reveals a Host-Specific Pattern of Secretory Proteins and Transposable Elements. PLoS One. 2016;11(9):e0162458. Published 2016 Sep 22. doi:10.1371/journal.pone.0162458

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.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-3212. doi:10.1093/bioinformatics/btv351

Spanu PD, Abbott JC, Amselem J, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 2010;330(6010):1543-1546. doi:10.1126/science.1194573

Sperschneider, J., Chen, J., Anderson, C., Morin, E., Zhang, X., Lewis, D., Henningsen, E., Grigoriev, I. V., Rathjen, J. P., Jones, D. A., Duplessis, S., and Dodds, P. N. 2025b. A Chromosome-Scale Genome Assembly of the Flax Rust Fungus Reveals the Two Unusually Large Effector Proteins, AvrM3 and AvrN. Mol Plant Microbe Interact. https://doi.org/10.1094/MPMI-04-25-0047-R.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Molecular Plant Pathology, 19(9), 2094–2110. https://doi.org/10.1111/mpp.12682

Sperschneider, J., Hewitt, T., Lewis, D. C., Periyannan, S., Milgate, A. W., Hickey, L. T., Mago, R., Dodds, P. N., & Figueroa, M. (2023). Nuclear exchange generates population diversity in the wheat leaf rust pathogen Puccinia triticina. Nature Microbiology, 8(11), 2130–2141. https://doi.org/10.1038/s41564-023-01494-9

Sperschneider, J., Jones, A. W., Nasim, J., Xu, B., Jacques, S., Zhong, C., Upadhyaya, N. M., Mago, R., Hu, Y., Figueroa, M., Singh, K. B., Stone, E. A., Schwessinger, B., Wang, M.-B.,

Taylor, J. M., and Dodds, P. N. 2021. The stem rust fungus Puccinia graminis f. sp. tritici induces centromeric small RNAs during late infection that are associated with genome-wide DNA methylation. BMC Biology 19:203. https://doi.org/10.1186/s12915-021-01123-z.

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.

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

Starnes, J. H., Thornbury, D. W., Novikova, O. S., Rehmeyer, C. J., and Farman, M. L. 2012. Telomere-targeted retrotransposons in the rice blast fungus Magnaporthe oryzae: agents of telomere instability. Genetics 191:389–406. https://doi.org/10.1534/genetics.111.137950.

Szalay, M.-F., Majchrzycka, B., Jerković, I., Cavalli, G., and Ibrahim, D. M. 2024. Evolution and function of chromatin domains across the tree of life. Nat Struct Mol Biol 1–14. https://doi.org/10.1038/s41594-024-01427-y.

Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., Tavares, S., Pereira, A. P., Azinheira, H. G., Guerra‐Guimarães, L., Várzea, V., & Silva, M. do C. (2017). The coffee leaf rust pathogen Hemileia vastatrix: One and a half centuries around the tropics. Molecular Plant Pathology, 18(8), 1039–1051. https://doi.org/10.1111/mpp.12512

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics, Chapter 4, 4.10.1-4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

Talhinhas P, Batista D, Diniz I, et al. The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. Mol Plant Pathol. 2017;18(8):1039-1051. doi:10.1111/mpp.12512

Tam R, Möller M, Luo R, et al. Long-read genomics reveal extensive nuclear-specific evolution and allele-specific expression in a dikaryotic fungus. Genome Res. 2025;35(6):1364-1376. Published 2025 Jun 2. doi:10.1101/gr.280359.124

Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., Jones, A., Lou, Z., Zhang, P., Sandhu, K., Smith, G. R., Tibbits, J., Chagné, D., & Park, R. F. (2021). Austropuccinia psidii, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 (Bethesda, Md.), 11(3), jkaa015. https://doi.org/10.1093/g3journal/jkaa015

Torres, D. E., Kramer, H. M., Tracanna, V., Fiorin, G. L., Cook, D. E., Seidl, M. F., and Thomma, B. P. H. J. 2024. Implications of the three-dimensional chromatin organization for genome evolution in a fungal plant pathogen. Nat Commun 15:1701. https://doi.org/10.1038/s41467-024-45884-x.

Torres, D. E., Reckard, A. T., Klocko, A. D., and Seidl, M. F. 2023. Nuclear genome organization in fungi: from gene folding to Rabl chromosomes. FEMS Microbiology Reviews 47:fuad021. https://doi.org/10.1093/femsre/fuad021.

Torres DE, Thomma BPHJ, Seidl MF. Transposable Elements Contribute to Genome Dynamics and Gene Expression Variation in the Fungal Plant Pathogen Verticillium dahliae. Genome Biol Evol. 2021;13(7):evab135. doi:10.1093/gbe/evab135

Tralamazza, S.M., Gluck-Thaler, E., Feurtey, A. et al. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. Nat Commun 15, 5728 (2024). https://doi.org/10.1038/s41467-024-49913-7

Urquhart, A. S., Chong, N. F., Yang, Y., & Idnurm, A. (2020). Eukaryotic transposable elements as “cargo carriers”: The forging of metal resistance in the fungus Paecilomyces variotii. bioRxiv. https://doi.org/10.1101/2020.03.06.981548

van Bree, E. J., Guimarães, R. L. F. P., Lundberg, M., Blujdea, E. R., Rosenkrantz, J. L., White, F. T. G., Poppinga, J., Ferrer-Raventós, P., Schneider, A.-F. E., Clayton, I., Haussler, D., Reinders, M. J. T., Holstege, H., Ewing, A. D., Moses, C., & Jacobs, F. M. J. (2022). A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. Genome Research, 32(4), 656–670. https://doi.org/10.1101/gr.275515.121

Van Etten J, Bhattacharya D. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much?. Trends Genet. 2020;36(12):915-925. doi:10.1016/j.tig.2020.08.006

Van de Wouw AP, Elliott CE, Popa KM, Idnurm A. Analysis of Repeat Induced Point (RIP) Mutations in Leptosphaeria maculans Indicates Variability in the RIP Process Between Fungal Species. Genetics. 2019;211(1):89-104. doi:10.1534/genetics.118.301712

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.

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.

Van't Hof AE, Campagne P, Rigden DJ, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-105. doi:10.1038/nature17951

Vanyushin BF, Ashapkin VV. DNA methylation in higher plants: past, present and future. Biochim Biophys Acta. 2011;1809(8):360-368. doi:10.1016/j.bbagrm.2011.04.006

Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 314–324. https://doi.org/10.1109/IPDPS.2019.00041

Vialle, A., Frey, P., Hambleton, S., Bernier, L., & Hamelin, R. C. (2011). Poplar rust systematics and refinement of Melampsora species delineation. FUNGAL DIVERSITY, 50(1), 227–248. https://doi.org/10.1007/s13225-011-0129-6

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics, 33(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Vyas M, Kasbekar DP. Collateral damage: spread of repeat-induced point mutation from a duplicated DNA sequence into an adjoining single-copy gene in Neurospora crassa. J Biosci. 2005;30(1):15-20. doi:10.1007/BF02705146

Wacker T, Helmstetter N, Wilson D, Fisher MC, Studholme DJ, Farrer RA. Two-speed genome evolution drives pathogenicity in fungal pathogens of animals. Proc Natl Acad Sci U S A. 2023;120(2):e2212633120. doi:10.1073/pnas.2212633120

Wang J, Lu N, Yi F, Xiao Y. Identification of Transposable Elements in Conifer and Their Potential Application in Breeding. Evol Bioinform Online. 2020;16:1176934320930263. Published 2020 Jun 15. doi:10.1177/1176934320930263

Wang, J., Xu, Y., Peng, Y., Wang, Y., Kang, Z., and Zhao, J. 2024. A fully haplotype-resolved and nearly gap-free genome assembly of wheat stripe rust fungus. Sci Data 11:508. https://doi.org/10.1038/s41597-024-03361-6.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution, 35(3), 543–548. https://doi.org/10.1093/molbev/msx319

Whan, A., Suchecki, R., & Stephen, S. (2019). csiro-crop-informatics/biokanga: V4.4.2 [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2649581

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.

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.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics, 8(12), Article 12. https://doi.org/10.1038/nrg2165

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.

Wu J, Kou Y, Bao J, et al. Comparative genomics identifies the Magnaporthe oryzae avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. New Phytol. 2015;206(4):1463-1475. doi:10.1111/nph.13310

Wu, J. Q., Song, L., Ding, Y., Dong, C., Hasan, M., & Park, R. F. (2021). A Chromosome-Scale Assembly of the Wheat Leaf Rust Pathogen Puccinia triticina Provides Insights Into Structural Variations and Genetic Relationships With Haplotype Resolution. Frontiers in Microbiology, 12, 704253. https://doi.org/10.3389/fmicb.2021.704253

Xhaard C, Fabre B, Andrieux A, et al. The genetic structure of the plant pathogenic fungus Melampsora larici-populina on its wild host is extensively impacted by host domestication. Mol Ecol. 2011;20(13):2739-2755. doi:10.1111/j.1365-294X.2011.05138.x

Xia, C., Huang, L., Huang, J., Zhang, H., Huang, Y., Benhamed, M., Wang, M., Chen, X., Zhang, M., Liu, T., & Chen, W. (2022). Folding Features and Dynamics of 3D Genome Architecture in Plant Fungal Pathogens. Microbiology Spectrum, 10(6), e0260822. https://doi.org/10.1128/spectrum.02608-22

Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research, 35(Web Server issue), W265–W268. https://doi.org/10.1093/nar/gkm286

Yan, H., Bombarely, A., & Li, S. (2020). DeepTE: A computational method for de novo classification of transposons with convolutional neural network. Bioinformatics, 36(15), 4269–4275. https://doi.org/10.1093/bioinformatics/btaa519

Yildirir, G., Sperschneider, J., Malar C, M., Chen, E. C. H., Iwasaki, W., Cornell, C., and Corradi, N. 2022. Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytologist 233:1097–1107. https://doi.org/10.1111/nph.17842.

Zan, Y., Chen, S., Ren, M., Liu, G., Liu, Y., Han, Y., Dong, Y., Zhang, Y., Si, H., Liu, Z., Liu, D., Zhang, X., Tong, Y., Li, Y., Jiang, C., Wen, L., Xiao, Z., Sun, Y., Geng, R., … Yang, A. (2025). The genome and GeneBank genomics of allotetraploid Nicotiana tabacum provide insights into genome evolution and complex trait regulation. Nature Genetics, 57(4), 986–996. https://doi.org/10.1038/s41588-025-02126-0

Zeng X, Yi Z, Zhang X, et al. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat Plants. 2024;10(8):1184-1200. doi:10.1038/s41477-024-01755-3

Zilberman D, Henikoff S. Silencing of transposons in plant genomes: kick them when they're down. Genome Biol. 2004;5(12):249. doi:10.1186/gb-2004-5-12-249

Résumé étendu en francais

Les éléments transposables : moteurs de diversité et de plasticité génomique

Les éléments transposables (ETs), ou séquences mobiles d'ADN, sont omniprésents dans les génomes procaryotes et eucaryotes. Depuis leur découverte par Barbara McClintock dans les années 1940 à partir d'expériences sur le maïs (McClintock, 1951), ces éléments ont radicalement transformé la compréhension de la dynamique génomique. Initialement considérés comme du "*junk DNA*", les ETs sont désormais reconnus comme des agents structurants du génome et des moteurs d'évolution. Ils représentent plus de 45 % du génome humain (Lander et al., 2001) et jusqu'à 80 % dans certaines plantes comme le maïs. Les ETs sont classés en deux grandes classes selon leur mécanisme de transposition : les rétrotransposons (classe I), qui utilisent une étape intermédiaire ARN ("copier-coller"), et les transposons ADN (classe II), qui fonctionnent selon un mécanisme de type "couper-coller". Chaque classe se subdivise en ordres et Superfamilles, caractérisés par leurs séquences répétées, enzymes codantes, ou motifs structuraux (Wicker et al., 2007).

Les ETs sont à l'origine de nombreuses modifications structurales du génome, telles que des duplications, inversions, insertions ou délétions (Bourque et al., 2018). Ils peuvent altérer l'expression des gènes adjacents par activation de promoteurs cryptiques, interférence avec les séquences régulatrices, ou exonisation d'introns (Bourque et al., 2018). Ces phénomènes contribuent à l'apparition de nouveaux transcrits, voire de nouvelles fonctions protéiques. Par exemple, l'insertion d'un ET chez l'orange sanguine module l'expression d'un gène pigmentaire en réponse au froid (Butelli et al., 2012), ou encore l'insertion d'un transposon dans le gène *cortex* du papillon de nuit *Biston betularia*, entraîne une coloration sombre avantageuse en milieu industriel (Hof et al., 2016). Les ETs peuvent même être cooptés par le génome hôte pour générer de nouvelles fonctions : ainsi, les LTRs peuvent agir comme promoteurs

d'interférons, et certains rétrovirus endogènes comme les syncytines sont essentiels au développement placentaire (Chuong et al., 2016 ; Cornelis et al., 2015).

La régulation de la prolifération des ETs repose sur des mécanismes de silencing épigénétiques, transcriptionnels et post-transcriptionnels. Chez les plantes et animaux, la méthylation de l'ADN (ex. CG, CHG, CHH) et les modifications d'histones (H3K9me2/3) participent au maintien de l'hétérochromatine autour des ETs (Bewick et al., 2019; Fouché et al., 2022; Haggerty et al., 2021; Lippman et al., 2004; Nai et al., 2021). Des systèmes comme le RNA-directed DNA methylation (RdDM) chez les plantes ou la voie piRNA chez les animaux garantissent un contrôle efficace de leur activité (Charmant et al., 2024; Saito et al., 2006). Chez les champignons, la mutation induite par répétition (RIP) constitue un mécanisme de défense unique, induisant des transitions C→T dans les séquences répétées (Cambareri et al., 1989; Gladyshev, 2017; Vyas & Kasbekar, 2005). Ce processus, observé chez *Neurospora crassa*, rend les copies TE inactives avant même leur mobilisation(Cambareri et al., 1989; Freitag et al., 2002). Néanmoins, dans les génomes riches en répétitions (comme le maïs), les TE inactivés s'accumulent sous forme fragmentée, illustrant les limites de ces mécanismes.

Les ETs jouent un rôle clé dans la dynamique évolutive des champignons dont la taille des génomes est extrêmement variable, allant de quelques mégabases à plus d'un gigabase, souvent corrélée à l'accumulation d'ETs (e.g. *Phakopsora pachyrhizi*). Les ETs participent à l'adaptation rapide par divers mécanismes. Certaines espèces, comme *Zymoseptoria tritici* ou *Botrytis cinerea*, présentent des insertions d'ETs modulant directement l'expression de gènes de virulence ou de résistance aux fongicides (Omrane et al., 2017; Porquier et al., 2021). Les ETs participent aussi à la diversification des effecteurs fongiques, en altérant leur structure ou leur expression. Chez les champignons symbiotiques, des ETs sont impliqués dans l'expansion de familles de gènes liées aux échanges nutritionnels, comme chez *Rhizophagus irregularis*

(Morin et al., 2019). Enfin, certains ETs facilitent les transferts horizontaux de gènes entre espèces fongiques. Les éléments Starship, récemment identifiés, transportent des gènes adaptatifs et sont associés à des régions virulentes (e.g. ToxA)(McDonald et al., 2019). Une étude récente révèle plus de 5 500 événements de transfert horizontal d'ETs entre champignons, contribuant parfois à 70 % du contenu en ETs d'un génome donné (Romeijn et al., 2025).

Les Pucciniales : un modèle d'étude des ETs chez les pathogènes biotrophes

L'ordre des Pucciniales représente le groupe le plus vaste de champignons phytopathogènes du règne fongique, avec plus de 8000 espèces identifiées (Aime et al., 2018). Les Pucciniales sont responsables de maladies appelées rouilles, affectant de nombreuses plantes d'intérêt agronomique comme le blé (par exemple, *Puccinia graminis* f. sp. *tritici*), le café (*Hemileia vastatrix*), ou encore l'eucalyptus (*Austropuccinia psidii*). Les Pucciniales sont des biotrophes obligatoires, c'est-à-dire qu'ils dépendent de tissus vivants pour compléter leur cycle de vie. Celui-ci est complexe et hautement spécialisé, pouvant comporter jusqu'à cinq types de spores distincts. De nombreuses espèces sont également hétéroïques, alternant entre deux hôtes phylogénétiquement distincts pour accomplir les différentes phases de reproduction sexuelle et asexuée (Lorrain et al., 2019).

Sur le plan génomique, les Pucciniales présentent une architecture particulièrement singulière. Leurs génomes, de taille exceptionnelle pour des champignons (pouvant excéder 1,2 Gb), se caractérisent par une forte proportion en ETs (jusqu'à 90 % du contenu total), une structure dikaryotique, et une divergence inter-haplotypique pouvant être très marquée (Henningsen et al., 2025). Les avancées récentes dans les technologies de séquençage dites de longues séquences (PacBio HiFi, ONT), couplées aux approches de phasage par capture de

conformation (Hi-C), ont permis d'obtenir des assemblages haplotype-résolus de haute qualité pour plusieurs espèces modèles, telles que *Melampsora lini* ou *Puccinia triticina* (Li et al., 2023; Sperschneider et al., 2025). Ces données génomiques offrent un cadre robuste pour explorer les effets des éléments transposables sur la dynamique évolutive des gènes codant des effecteurs de pathogénie, l'organisation chromosomique et les mécanismes d'adaptation à l'hôte. De plus, une caractéristique remarquable de ces génomes réside dans l'expansion de certaines familles de gènes, notamment codant pour des protéines sécrétées, souvent associées à des évènements de transposition (Duplessis et al., 2011). Ces protéines jouent un rôle central dans la manipulation des réponses immunitaires végétales et participent au contournement des barrières de défense de l'hôte. Ainsi, les Pucciniales s'imposent comme un modèle de choix pour l'analyse des relations entre éléments transposables, plasticité génomique et pathogénicité chez les champignons biotrophes.

Objectifs de la thèse

Cette thèse s'inscrit dans une démarche de recherche fondamentale visant à élucider les mécanismes évolutifs ayant contribué à la structuration et à la diversification des génomes au sein des champignons de l'ordre des Pucciniales, et plus particulièrement ceux de la famille des Melampsoraceae. Le contraste entre une biotrophie obligatoire, supposant une forte dépendance à l'hôte, et des génomes exceptionnellement vastes et riches en séquences répétées, mais aussi en gènes, soulève des interrogations majeures sur les processus évolutifs à l'origine de cette architecture génomique singulière. L'objectif principal de cette thèse est ainsi de décrypter le rôle des ETs dans l'organisation, la dynamique et l'évolution des génomes au sein de l'ordre des Pucciniales, en mobilisant une approche intégrative de génomique comparative. Ce travail vise, in fine, à mieux comprendre comment ces éléments mobiles ont pu

contribuer à l'innovation génétique, à l'adaptation parasitaire, et à l'émergence de fonctions associées à la virulence chez les Melampsoraceae.

Chapitre 1 : Expansion des éléments transposables à l'échelle des Pucciniales

Ce premier chapitre examine le rôle déterminant des ETs dans l'expansion et l'évolution des génomes des Pucciniales. L'annotation et la curation manuelle de mobilomes dans 12 génomes représentatifs du sous-embranchement des Pucciniomycotina, dont huit dans l'ordre des Pucciniales, révèlent que la couverture en ETs varie de 47 à 92 % chez ces derniers, contre 2 à 36 % chez les autres espèces du Pucciniomycotina. La corrélation étroite entre taille du génome et contenu en ETs indique que ces derniers constituent le principal facteur d'inflation génomique, bien plus que l'augmentation du nombre de gènes. L'analyse des familles d'ETs met en évidence la prépondérance des rétrotransposons à LTR (notamment Gypsy et Copia) et des transposons à répétitions inversées terminales TIR, responsables majeurs des expansions observées. Les profils de divergence de séquences (indices K2P) et les estimations d'âge moléculaire montrent la coexistence d'insertions anciennes, parfois antérieures à la divergence des Pucciniomycotina (~176 Ma), et d'invasions récentes, survenues principalement au cours des 50 derniers millions d'années, avec des expansions particulièrement marquées chez certaines espèces (par exemple, *M. lini*). Ces résultats soutiennent l'hypothèse d'une dynamique combinant rétention d'éléments ancestraux et proliférations spécifiques à certaines espèces ou lignées. La comparaison interspécifique révèle l'existence de familles d'ETs propres aux différentes familles de Pucciniales, ainsi qu'une forte proportion de familles spécifiques aux espèces et même aux haplotypes, attestant de l'activité persistante et diversifiée des ETs. Toutefois, aucune déficience claire dans les mécanismes connus de régulation des ETs (mutations RIP, méthylation de l'ADN, voie de l'interférence par ARN) n'a été détectée,

suggérant que la tolérance à l'accumulation des ETs constitue une caractéristique intrinsèque des génomes des Pucciniales.

En conclusion, ce chapitre montre que la prolifération massive de ETs, en particulier des rétrotransposons Gypsy, constitue une signature ancestrale des Pucciniales et une caractéristique fondamentale de leur architecture génomique. Ces génomes tolérants aux ETs apparaissent ainsi façonnés par une succession de vagues anciennes et récentes d'expansion, fournissant un cadre pour comprendre leur plasticité et leur succès évolutif en tant que pathogènes biotrophes.

Chapitre 2 : Assemblages haplotype-résolus et annotation des génomes de Melampsoraceae

Dans ce chapitre, nous présentons l'assemblage et l'annotation répétitive de trois génomes d'espèces de Pucciniales appartenant à la famille des Melampsoraceae : deux souches de *Melampsora larici-populina* (98AG31 et 08KE26) et une souche de *Melampsora allii-populina* (14AY07). Ces espèces constituent des modèles pertinents pour explorer l'architecture génomique des rouilles hétéroïques, qui alternent entre deux hôtes, et possèdent des génomes exceptionnellement vastes et riches en éléments répétitifs. Dans un premier temps, nous avons généré des assemblages de haute qualité de *M. larici-populina* 98AG31 et *M. allii-populina* 14AY07 à partir de longues séquences PacBio HiFi, combinées à des données Hi-C. Ces assemblages présentent une résolution haploïde complète (haplotype-phasée), allant de télomère à télomère, et révèlent 18 chromosomes par haplotype, un nombre cohérent avec d'autres espèces de Pucciniales. L'annotation des ETs a été réalisée de manière homogène entre haplotypes, confirmant une abondance massive d'ETs (jusqu'à 80 % du contenu

génomique), dominée par les éléments LTR, en particulier les Gypsy pour *M. allii-populina* 14AY07 et des TIR pour *M. larici-populina* 98AG31.

Dans un second temps, nous avons assemblé le génome d'une souche clonale de *M. larici-populina* (08KE26). Cette souche appartient à une lignée asexuée récemment mise à jour dans les populations de *M. larici-populina* collectées en France dans l'UMR IAM sur les dernières décennies. La comparaison entre les souches 08KE26 et 98AG31 met en évidence un niveau élevé de colinéarité entre haplotypes et entre souches, suggérant une stabilité structurale du génome malgré les différences de mode de reproduction (clonal *vs.* recombiné). Une seule région présente une divergence structurelle notable entre haplotypes : elle est localisée à proximité d'un locus contenant un gène codant un récepteur de phéromone (PR), associé à la reproduction chez les champignons, et coïncide avec une inversion chromosomique dans la souche clonale. La comparaison des deux souches de *M. larici-populina* n'a pas révélé de différence majeure dans le contenu ou la composition en ETs, remettant en question l'hypothèse selon laquelle la reproduction clonale favoriserait l'accumulation incontrôlée de ETs.

Ces assemblages haplotype-phasés, réalisés à l'échelle chromosomique, constituent une ressource précieuse pour la communauté scientifique. Ils permettront d'approfondir l'étude de l'architecture génomique, de la plasticité structurale et des effets potentiels des ETs sur l'expression et la variation des gènes. Toutefois, l'absence d'annotation génique au moment de la finalisation de cette thèse limite les interprétations fonctionnelles, en particulier quant aux interactions entre ETs et gènes.

Chapitre 3 : Organisation du génome et rôle des éléments transposables chez les Melampsoraceae

Ce chapitre explore l'organisation du génome de trois espèces de la famille des Melampsoraceae (*M. larici-populina*, *M. allii-populina* et *M. lini*), en s'appuyant sur des assemblages de génomes haplotypiques et des données de conformation tridimensionnelle issues de la technologie Hi-C. En combinant ces approches, l'étude vise à caractériser la structure nucléaire, à identifier les centromères et télomères, et à analyser le rôle des ETs dans l'architecture et l'évolution des génomes de ces champignons pathogènes des plantes.

Les analyses de conformation nucléaire révèlent une organisation spatiale conservée de type Rabl dans les trois espèces, caractérisée par un regroupement des centromères observable sous forme de motifs en "nœud papillon" sur les cartes de contact Hi-C. L'analyse fine des ETs confirme les résultats du Chapitre 1, avec une diversité d'ETs entre les espèces. Bien que 64 familles d'ETs soient partagées entre les trois espèces, chacune en possède des milliers qui leur sont spécifiques. Les génomes de *M. lini* et *M. allii-populina* présentent une prédominance de rétrotransposons de type LTR-Gypsy, tandis que celui de *M. larici-populina* est enrichi en transposons à répétitions terminales inversées TIR. Deux vagues majeures de prolifération des ETs, ancienne et récente, sont identifiées dans chaque génome. Ces vagues varient en intensité et composition selon les haplotypes, soulignant une dynamique évolutive active au sein des espèces. Les centromères, présentent une signature spécifique marquée par une forte présence de transposons de classe II, en particulier des TIRs, dans toutes les espèces. D'autres familles telles que les Helitrons, MITEs ou DIRS montrent également des enrichissements haplotype-spécifiques. Quant aux télomères, seuls *M. larici-populina* et *M. allii-populina* présentent des points chauds d'enrichissement en LINEs dans les régions

subtélomériques, absents chez *M. lini*. Cela suggère un rôle potentiel des LINEs dans la stabilité chromosomique, peut-être via la protection des extrémités contre les réarrangements.

L'analyse de la synténie révèle une conservation générale du nombre et de l'ordre des chromosomes (18 par haplotype), bien que des inversions spécifiques, notamment chez *M. larici-populina*, témoignent de réarrangements ponctuels. Cependant, aucune corrélation claire n'a été établie entre les régions de rupture de synténie et une accumulation locale d'ETs, ce qui contraste avec d'autres champignons pathogènes. Cela suggère que chez les Melampsoraceae, l'impact des ETs se manifeste davantage à l'échelle locale, notamment dans les régions fonctionnelles comme les centromères, plutôt qu'à l'échelle des réarrangements chromosomiques majeurs.

Ce chapitre ouvre la voie à des investigations plus poussées sur l'implication des ETs dans la diversification fonctionnelle des gènes liés aux interactions hôte-pathogène et à l'adaptation évolutive des espèces fongiques.

Conclusion

À travers une combinaison d'analyses comparatives, d'assemblages haplotypiques complets et de données de conformation génomique tridimensionnelles, cette thèse met en lumière l'importance des ETs dans l'expansion et l'organisation des génomes des Pucciniales, ainsi que leur possible implication dans la régulation des gènes liés à la pathogénicité. Les ETs, longtemps considérés comme de l'ADN "inutile", sont ici montrés comme des moteurs d'innovation génomique, contribuant à la plasticité et à l'architecture des génomes des rouilles. Cette thèse révèle que ces ETs persistent depuis des temps anciens tout en connaissant des phases d'activité spécifiques à certaines lignées. De plus, les nouveaux génomes assemblés

ont permis d'observer des motifs de répartition non aléatoire des ETs (centromères, télomères). Enfin, cette étude suggère que l'expansion des ETs n'est pas nécessairement corrélée à la reproduction clonale, questionnant certaines hypothèses sur leur accumulation. Les résultats offrent une base solide pour des recherches futures sur la régulation, l'expression génique et l'adaptation des Pucciniales, tout en ouvrant la voie à l'exploration de processus comme l'échange nucléaire somatique dans les lignées clonales.

APPENDICES

APPENDIX 1 : *Curriculum vitae*

APPENDIX 2: 2023 Poster JOBIM (Journée ouvertes en Biologie, Informatique et Mathématiques)

APPENDIX 3: 2024 Poster ICTE (International Congress on Transposable Elements)



EMMA CORRE

PHD STUDENT IN BIOINFORMATICS

ORCID:0000-0002-2985-8903

After graduating with a Master degree in Bioinformatics, Modelling and Statistics at the University of Rouen, I am currently pursuing a PhD at INRAE. My main research focuses on understanding the evolution of rust fungi through the study of transposable elements. I am a responsible, dynamic and committed researcher, able to adapt quickly to various research fields. My goal is to collaborate with leading experts in evolutionary genomics, acquire new skills, and share my expertise internationally.

PERSONAL INFORMATION

28 years old
Driving Licence

INFORMATION

532 Route de l'église
38440 Villeneuve de Marc
FRANCE
correemma@yahoo.com
+33 06 74 27 50 40
/emma-corre-9b3726115

IT SKILLS

language:
Python, R, Markdown
Versionning:
Conda, Git, Docker
Computing cluster:
Slurm

LANGUAGE SKILLS

French:
Native language
English:
Cambridge test B2 (2022)
CLES B1 & B2 (2019)
TOEIC 810 (2018)

INTERESTS

Music: Guitar, Bass, Ukulele
Other: Trek

EDUCATION

2025 **Current position - PhD Student in Bioinformatics**
INRAE - Lorraine University

2022 **Master degree in Bioinformatics Modeling and Statistics**
Rouen Normandie University

2019 **Bachelor degree in Cellular Biology and Physiology**
Clermont Auvergne University

2018 **Technological degree in Bioinformatics**
Clermont Auvergne University

PROFESSIONAL EXPERIENCES- PROJECTS

PhD Student: Invasion of Pucciniales genomes by transposable elements: How transposition events shape the evolution and the genomic architecture in rust fungi

NOV 2022 - OCT 2025
UMR 1136 IAM - INRAE NANCY, FRANCE (3 VISITS AT ETH ZURICH, SWITZERLAND - 4 MONTHS)
Supervisors: Dr. S. Duplessis, Dr. C. Lorrain

TE detection and annotation in obligate biotrophic fungi, **Genome assembly telomere to telomere** of dicaryotic species**, transcriptomics, genomic comparison, gene contraction and expansion analysis**. Development of Python and R scripts to analyse high-throughput data. Usage of Docker, snakemake pipelines. Opportunity to present my works at national and international conferences. Involving myself in laboratory life (PhD Student representative) as well as in grant demands.

Apprenticeship: identification of secreted peptides involved in the control of development in *Solanum nigrum*

SEPT 2020 - AUG 2022
HORTICULTURE AND SEED RESEARCH INSTITUTE, ANGERS, FRANCE
Supervisors: IA S. Pelletier, Dr. S. Aubourg

Construction of a pipeline for ***de novo* adventitious transcriptome** and selection of **secretory peptides**. Search for **analogous peptides** between *Arabidopsis thaliana* and *Solanum nigrum* via a reference organism, *Solanum lycopersicum*. Developing Python scripts for handling data files

Internship: Development of a comparative genomics resource for prasinoviruses

MAI 2020 - AOUT 2021
OBSERVATOIRE OCÉANOGRAPHIQUE, BANYULS-SUR-MER, FRANCE
Supervisors: Dr. G. Piganeau, V. Daric, Dr K. Vandepoele.

Genomic analyses of prasinoviruses, as well as the establishment of a **sequence annotation pipeline** to describe the diversity and **evolution** of these parasites. Search for homologs, study of the genome via **syntenies of genes and collinearities,** study of **horizontal transfers** of genes between algae and viruses.

Internship: Pipeline investigation for a MinION sequencer

MAY 2018 - AUG 2018
BEE PATHOLOGY AND Q FEVER UNIT - ANSES, SOPHIA-ANTIPOLIS, FRANCE
Supervisors: Dr. V. Duquesne, Dr. K. Sidi-Boumedine, Dr. F. Sircoulomb.

Implementation of a **MinION sequencer** and bioinformatic analysis. Construction of the **WGS libraries** of *Coxiella burnetii* and amplicons of *Tropilaelaps spp*, as well as **assembly and alignment pipeline**.

PUBLICATIONS

2025:

Corre E, Morin E, Duplessis S, Lorrain C, "Ancestral and recent bursts of transposition shaped the massive genomes of plant pathogenic rust fungi."
BMC Genomics - doi: https://doi.org/10.1186/s12864-025-11726-3
also available as a BioRxiv - doi: https://doi.org/10.1101/2025.01.10.632365

SCIENTIFIC COMMUNICATION & CONFERENCE ORGANIZATION

The presenting author is underlined

2025:

Corre E, Morin E, Duplessis S, Lorrain C. Telomere to telomere Assembly of diploid fungal pathogen in high repeat content, *RIP Symposium, Neuchatel, Switzerland, July 2025. Selected Oral presentation*

Duplessis S, Lorrain C, Morin E, **Corre E**, Abdalrahem A, Halkett F, Frey P. Unravelling the complex life cycle of rust fungi through genomics. *12ème colloque de la Société Française de Phytopathologie, SFP general conference. Bordeaux, France, May 2025. Selected Oral presentation*

Corre E, Morin E, Duplessis S, Lorrain C. A comparative analysis of transposable elements in the subphylum Pucciniomycotina reveals a massive retrotransposon-driven expansion in the genomes of rust fungi.
17th European Conference on Fungal Genetics (ECFG17), Dublin, Ireland, March 2025. Selected oral presentation in the Genomes and multiomics session.

2024:

Corre E, Morin E, Lorrain C, Duplessis S. A comparative analysis of transposable elements in the subphylum Pucciniomycotina reveals a retrotransposon-driven expansion in the massive genomes of rust fungi. *International Congress on Transposable elements (ICTE2024). Saint Malo, France, April 2024. Selected poster communication.*

Lorrain C, Glavincheska I, **Corre E**, Duplessis S. Coevolution between transposable elements and fungal plant pathogen genomes shapes the genome architecture, plasticity, and adaptation. *International Congress on Transposable elements (ICTE2024). Saint Malo, France, April 2024. Selected poster communication.*

Corre E, Morin E, Lorrain C, Duplessis S. A comparative analysis of transposable elements in the subphylum Pucciniomycotina reveals a retrotransposon-driven genome expansion in rust fungi. *Journées Jean Chevaugeon 2024, 14èmes rencontres de Phytopathologie et Mycologie, Aussois, France, January 2024. Selected oral communication in the Genomes, functional and evolutionary epigenomics session.*

Pint Of Sciences (Organisation team)
Invited Seminar - ETH

2023:

Corre E, Morin E, Lorrain C, Duplessis S. Genomics landscape shaped by transposable elements through rust pathogens history. *JOBIM 2023, Edition multisite, France, June 2023. Selected poster communication.*

Corre E, Morin E, Lorrain C, Duplessis S. Transposable elements lead to genome landscapes differentiation in the Pucciniomycotina subdivision. *Doc and Post-Doc Day of UMR 1136 IAM Tree-Microbe Interactions Department, INRAE/Université de Lorraine, Nancy, France. Selected oral communication.*

Corre E, Morin E, Lorrain C, Duplessis S. Transposable elements within Pucciniales. *SIRENA doctoral school, Université de Lorraine, Nancy, Selected oral communication.*

Regional Finalist - MT180's - Ma thèse en 180 secondes (science communication competition)
Organisation team - JOBIM
Organisation team - Doc and Post-Doc Day
Organisation team - Bake your PhD contest

2022:

Corre E, Gilbart V, Vesselle M, Lanselle D, Lahaye M, Lorthiois M, Gspann M, Dauchel H Green-BIM: a study to make young bioinformaticians aware of the carbon footprint of bioinformatics. *JOBIM 2022, Edition Rennes, France, June 2022. Selected poster communication*

Corre E, Mallegol P, Hanteville S, Tran Van Canh L, Ahmadi-Afzadi M, Celton JM, Renou JP, Aubourg S, Pelletier S Hybrid assembly of a de novo transcriptome from Solanum nigrum. *JOBIM 2022, Edition Rennes, France, June 2022. Selected poster communication*

Guinchard S, Galati M, Grentzinger V, El Khiari S, Bussell X, Von Grafenstein K, **Corre E** JeBif - Association for the Young Bioinformaticians of France. *JOBIM 2022, Edition Rennes, France, June 2022. Selected poster communication*

Corre E, Guinchard S, Bioinformatics student organisation JEBIF. *BigDay 2022, Rennes, France, 2022. Selected poster communication*

2021:

Vaginay A, Galati M, Grentzinger V, El Khiari S, Bussell X, **Corre E** JeBif - Association for the Young Bioinformaticians of France. *JOBIM 2021, Edition Online, France, June 2021. Selected poster communication*

Organisation team - ISCB Young Bioinformatics Day (French representative)

SUPERVISION

2024: RNA-Seq transcriptomic analysis of poplar leaf infection by the pathogenic fungus *Melampsora larici-populina*: search for genes associated with immunity in poplar and infection in the fungus.
Co-supervision of J. Mourot Master of Microbiology, M2 student internship (6 months), Université de Lorraine, Nancy, France (co-supervisor: S. Duplessis)

FELLOWSHIPS

2025: **DrEAM** (Doctor, Explore and Achieve More by the French I-SITE programme Lorraine Université d'Excellence): 2872 euros for mobility and collaboration with ETH Zurich. This project aimed to generate a new reference genome for the poplar rust fungus.
DESSE (Europe Direction): Financial support of 2500 euros for mobility and collaboration with ETH Zurich. This project aimed to generate a new reference genome for the poplar rust fungus.
SFP (French Society for Phytopathology): 600 euros for the 17th European Conference on Fungal Genetics (ECFG17), Dublin, Ireland, March 2025

2024: **SFP** (French Society for Phytopathology): 200 euros for the International Congress on Transposable elements (ICTE2024). Saint Malo, France, April 2024.
PRETE (PopRust-EvolTE) Setting definitive reference genomes for poplar rust fungi to improve our understanding of the impact of Transposable Elements (TEs) on evolutionary processes related to virulence. Labex ARBRE 2024 call Graines d'Artemis: Financial support of 10k euros for long-read sequencing and HiC; support for 2 months visit at ETH Zurich.

2023: **Labex Arbre** (French local research programmes): 800 euros for collaborations
Agreenium - EIR-a: 2 weeks of seminars

MEMBERSHIPS AND SCIENTIFIC COMMITTEES

2022 - to date: Elected PhD students representative elected member of the Scientific Council of UMR 1136 IAM Tree-Microbe Interactions Department, INRAE/Université de Lorraine, Nancy, France

2020 - 2022: Elected student representative, the University Council at Université de Rouen Normandie
Elected student representative, FTLV Committee (lifelong learning) at Université de Rouen Normandie
Elected student representative, CRIC Committee (international relations and cooperation) at Université de Rouen Normandie
Elected student representative, Culture Committee at Université de Rouen Normandie
Vice Chair, JEBIF (International Society for Computational Biology's Regional Student Group)
Committee member of ISCB (International Society for Computational Biology's)

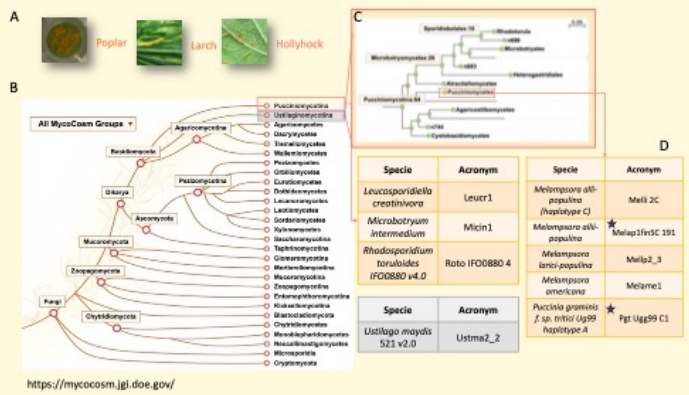
2017 - 2018: Elected student representative, UFR (Sciences and technologies department) at Université Clermont Auvergne
Elected student representative, disciplinary Council at Université Clermont Auvergne

Evolutionary history of transposable elements in the genomic landscape of rust pathogens

Emma CORRE[1], Emmanuelle MORIN[1], Cécile LORRAIN[2], Sébastien DUPLESSIS[1]
1 Université de Lorraine, INRAE, UMR IAM, 54280 Champenoux, France
2 Plant Pathology Group, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

Pucciniomycotina represents the largest subdivision in the kingdom Fungi with more than 8000 species [1]. Pucciniomycotina are obligate biotrophs infecting a wide panel of plants (soybean, wheat, coffee) causing rust diseases [2]. The order Pucciniales in this subphylum differs from other Pucciniomycotina, e.g. exhibiting larger genomes. The main hypothesis for such a difference is the presence of mobile elements called Transposable Elements (TEs) in their genomics landscape. TEs have a capacity for transposition– they can insert themselves in their hosting genome [3]. The increase of genomic data available for species falling in the Pucciniomycotina (including haplophased telomere-to-telomere fully resolved genomes) makes possible to elaborate a genomic comparative study of TEs in rust fungi.

Background: the subphylum Pucciniomycotina



https://mycocosm.jgi.doe.gov/

Fig 1: The subphylum Puccinomycotina.
A) Picture of rust diseases in various plants. B) Fungal phylogeny. C) Zoom on the subphylum Pucciniomycotina. D) Pucciniomycotina reference species and outgroup from the subphylum Ustilagomycotina; haplotype-phased genomes are marked by a red star.

Research questions

- How do transposable elements shape the genomic landscape of rust fungi compared to other species in the Pucciniomycotina?
- What is the evolutionary history of TEs invasion in rust fungal genomes?

Approach: TE analyses lead to phylogenetic understanding

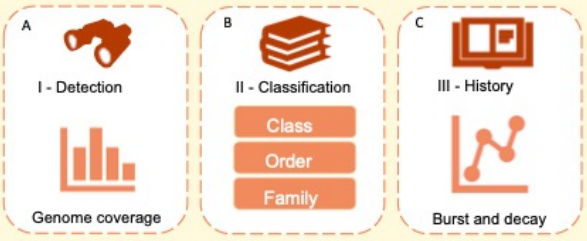


Fig 2: Description of three approaches to trace TE information in rust genomes.
A) Detection of TEs with REPET TEdenovo. B) Classification with REPET TEannot and home-made script. C) Analyses of TEs conservation and LTR insertion age in the family *Melampsoraceae*.

Genome size expansion is associated with TE content

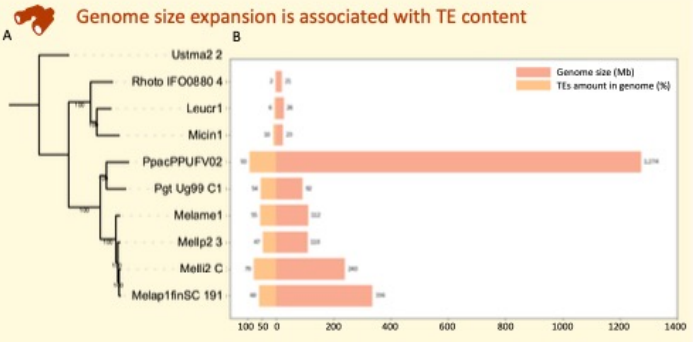


Fig 3: Genome size and TE coverage in the Pucciniomycotina.
A) Phylogeny based on single copy orthologues. B) Orange bars represent percent TE covering in the genome, salmon bars represents genome length in Mb [4]

Various genomic landscapes within the Pucciniomycotina

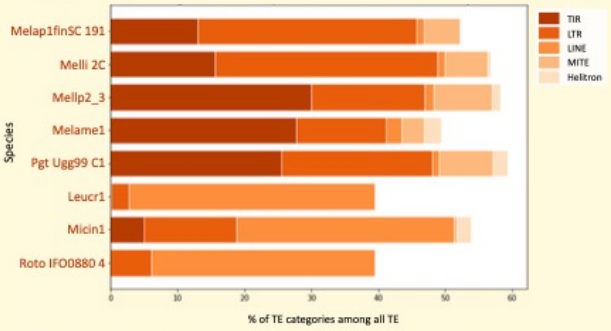


Fig 4: Proportion of TE categories among all TEs detected in the genomes of Pucciniomycotina
TE class I are composed of TIR, LTR and LINE. TE class II are composed of MITE, Helitron

TE conservation profiles are diverse in the Pucciniomycotina

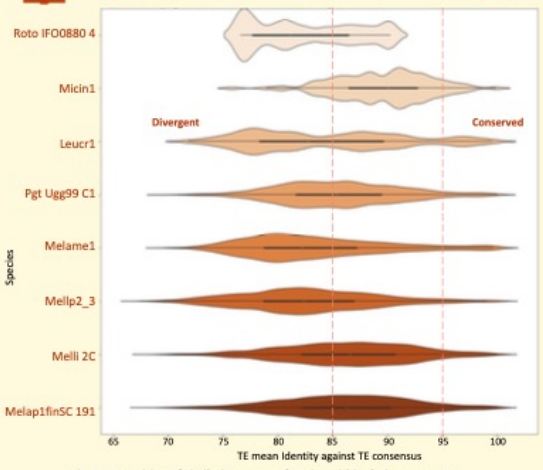


Fig 5: Repartition of similarity means of copies within their consensus
Copies with more than 95% of similarity against their consensus are considered conserved through history, whereas copies with less than 85% of similarity against their consensus are considered as degraded through evolution.

LTR transposition bursts in the family *Melampsoraceae*

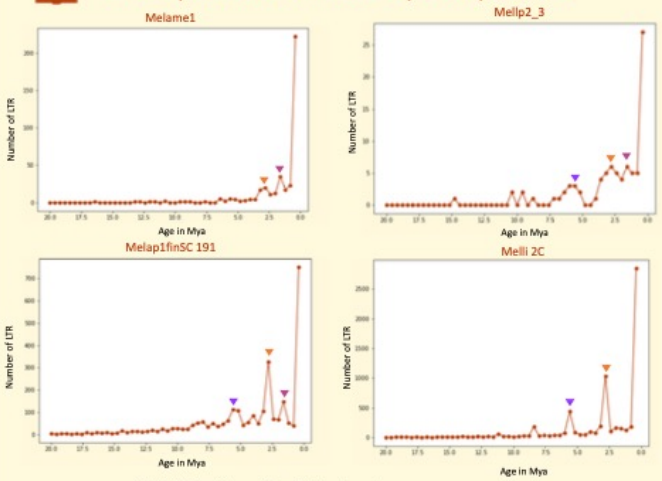


Fig 6: LTR insertion age from 20 Mya to our day.
Similar bursts of LTR transposition are highlighted by same colour flags

TAKE HOME MESSAGES: species from the subphylum Pucciniomycotina show different TE landscapes in their evolutionary history

- TEs proliferate more in Pucciniales than in other Pucciniomycotina (respectively, with a maximum of coverage of 93 % and 10%) increasing the genome size
- LINE copies are more abundant in other Pucciniomycotina species than in Pucciniomycetes, whereas MITEs are more abundant in the Pucciniales
- TE copies exhibit different level of conservation respective to their consensus, indicative of different transposition evolutionary trajectories
- Analysis of LTR insertion age in the *Melampsoraceae* shows similar events of transposition bursts across the evolution of the taxonomical family

References
[1]: Aime *et al. 2021 Fungal Systematics and Evolution.*; [2]: Petre & Duplessis 2022 *Frontiers Microbiol.*; [3]: Fouché *et .al* 2022 *Science Direct.*; [4]: Gupta *et al.* 2023 *Nature comms.*

Aknowledgment
Thanks to A. Lebreton, F. Maumus, A. Feurtey, P. Frey for their help on both biological and bioinformatic topics

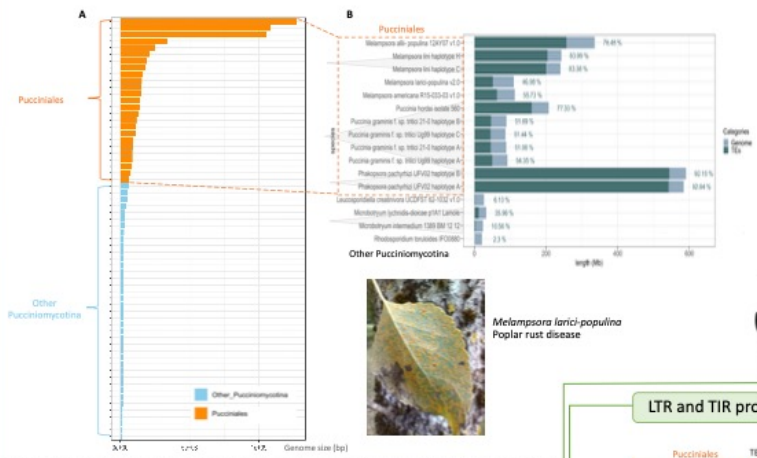
Exploring the TE landscapes in a unique order of fungal plant pathogens



Emma CORRE[1], Emmanuelle MORIN[1], Cécile LORRAIN[2], Sébastien DUPLESSIS[1]
1 Université de Lorraine, INRAE, UMR IAM, 54280 Champenoux, France
2 Plant Pathology Group, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland
e-mail: emma.corre@inrae.fr

Rust diseases pose significant challenges to agronomy, affecting staple food crops such as wheat, soybeans, and barley annually. These plant diseases are caused by obligate biotrophic fungi belonging to the order Pucciniales, representing both a historical and ongoing threat to food security. Pucciniales stands as the largest order of plant pathogens within the fungal kingdom, comprising over 8000 species [1]. Distinguished by their larger genome sizes and extensive gene catalogues compared to other orders within the same subphylum (Pucciniomycotina), Pucciniales exhibit a unique genomic landscape. Recent advancements in genome sequencing have enabled comparative analyses of the atypical genomic features and mobilome of Pucciniales [2]. Notably, transposable elements (TEs), often overlooked or basically described in genome studies, may hold key insights into the evolutionary history of these pathogens and contribute to a deeper understanding of their biology [3][4].

Genome size expansions are driven by TEs in Pucciniales

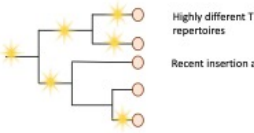


Melampsora larici-populina
Poplar rust disease

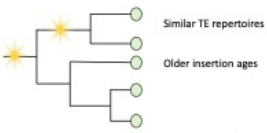
Figure 1. Rust fungi: a phylogenetic group of plant pathogens with large genomes and high TE content. A) Genome sizes (bp) of species in the subphylum Pucciniomycotina. Pucciniales species (i.e. rust fungi) are represented in orange and the other Pucciniomycotina are in blue; B) TE genome coverage in 12 species selected in the Pucciniomycotina

TE invasion in Pucciniales: ancestral genomic trait or independent bursts?

1) The proliferation of TEs in the genomes of Pucciniales occurred independently throughout evolution



2) Proliferation of TEs is an ancestral genomic trait of Pucciniales



1) How does the TE content differ between Pucciniales and other Pucciniomycotina?
2) What type of TEs had proliferated the most in Pucciniales?
3) Does TE proliferation occurred recently in Pucciniales?

TE annotation and insertion age estimations

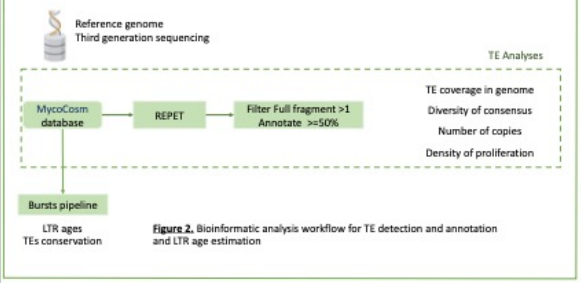


Figure 2. Bioinformatic analysis workflow for TE detection and annotation and LTR age estimation

LTR and TIR proliferate in Pucciniales

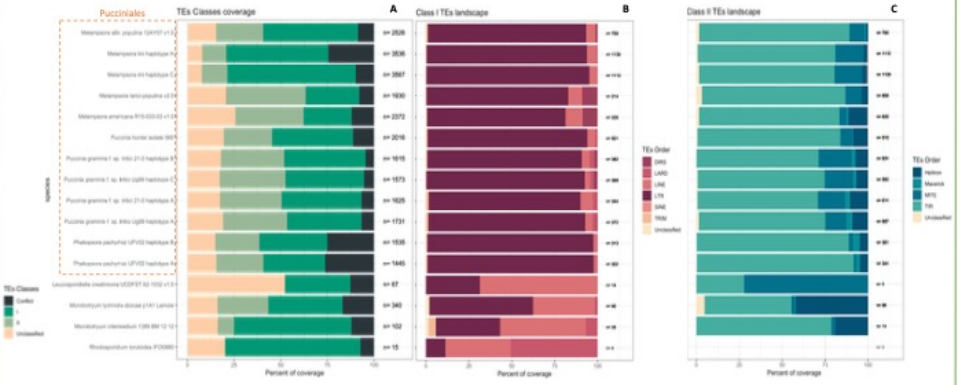


Figure 3. Coverage of TE at different TE-classification levels for the 12 species selected in the Pucciniomycotina
A) TE classes; B) class I orders and C) class II orders

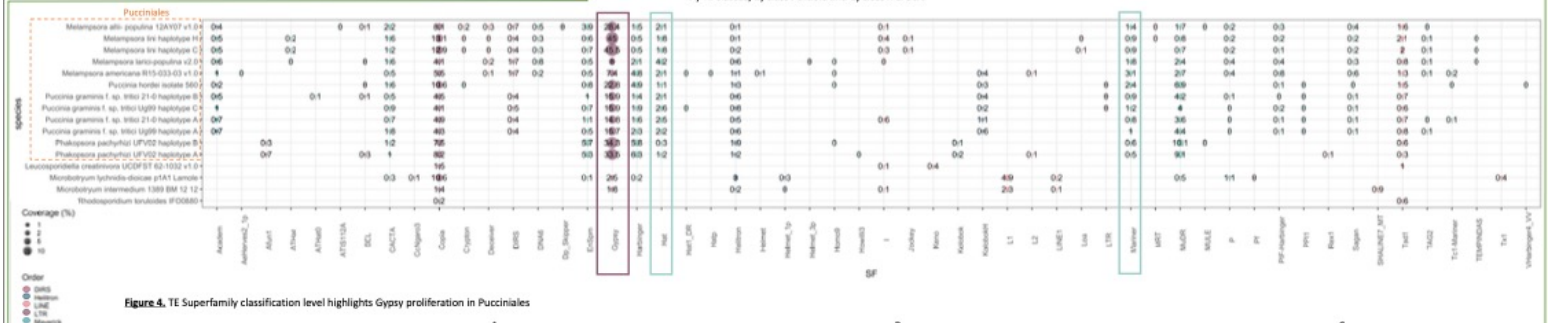


Figure 4. TE Superfamily classification level highlights Gypsy proliferation in Pucciniales

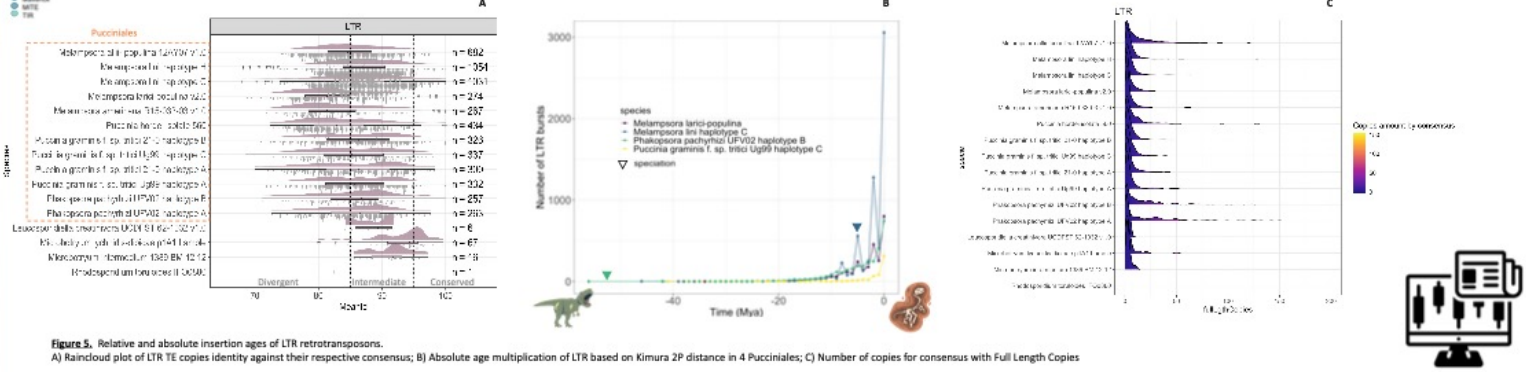


Figure 5. Relative and absolute insertion ages of LTR retrotransposons.
A) Raincloud plot of LTR TE copies identity against their respective consensus; B) Absolute age multiplication of LTR based on Kimura 2P distance in 4 Pucciniales; C) Number of copies for consensus with Full Length Copies

Take home messages: Pucciniales not only depict higher TE content compared to other Pucciniomycotina, but also different TE repertoires. Genome expansions in Pucciniales are driven by the proliferation of LTR retrotransposons, mainly Gypsy elements. Both relative and absolute insertion ages of LTRs in Pucciniales revealed a complex invasion history with ancient proliferation and more recent species-specific bursts.

Acknowledgments: We thank Lucas Auer and Annie Lebreton for their help in Bioinformatics. J. Sperschneider is acknowledged for the access to the *Melampsora lini* haplotype-phased genome.

References: [1]: Duplessis *et al. 2021 Annu. Rev. Phytopathol.*; [2]:Liang *et al.* 2022 Mol. Ecol. Resour.; [3]: Oggenfuss *et al. 2021 eLife;* [4]: Möller and Stukenbrock *et al. 2017 Nat. Rev. Microbiol.*

Résumé

Les agents phytopathogènes causant les maladies de rouilles (Basidiomycètes, Pucciniomycotina, Pucciniales) possèdent parmi les génomes les plus larges et complexes du règne fongique, notamment marqués par la prolifération massive d'éléments transposables (ETs). Cette thèse explore l'impact évolutif des ETs sur l'architecture et la biologie des génomes des Pucciniales, à travers une approche intégrative combinant génomique comparative, assemblages haplotypiques de haute qualité et analyses de l'architecture 3D du génome. Nous avons d'abord mis en place un pipeline d'annotation standardisé appliqué à 12 espèces de Pucciniales et apparentées aux Pucciniomycotina, révélant ainsi que les génomes riches en ETs sont une caractéristique ancienne et répandue dans cet ordre. Le paysage ET traduit à la fois la persistance d'éléments ancestraux et des vagues de transposition spécifiques à certaines espèces, en particulier impliquant des rétrotransposons LTR et des transposons à ADN. Malgré la conservation des gènes impliqués dans les mécanismes de contrôle des ETs, l'accumulation de ETs apparaît dynamique et récurrente. Nous avons ensuite généré des assemblages chromosomiques phasés pour deux rouilles du peuplier, *Melampsora larici-populina* et *Melampsora allii-populina*, en combinant séquençage PacBio et Hi-C. Ces nouveaux génomes de référence ont permis d'identifier 18 chromosomes par génome haploïde et révélé une distribution non-aléatoire des ETs dans les régions centromériques et subtélomériques. La comparaison entre une souche clonale et une souche non clonale de *M. larici-populina*, montre des contenus en ETs similaires, remettant en question l'hypothèse selon laquelle la reproduction clonale favorise une accumulation incontrôlée de ETs. Ces travaux démontrent que les ETs ne sont pas de simples éléments passifs des génomes, mais des acteurs centraux de l'architecture, de la stabilité et potentiellement de la régulation des génomes des Pucciniales. Ces travaux de thèse soutiennent l'idée que les génomes riches en ET offrent potentiellement une plasticité accrue favorisant l'adaptation à un mode de vie biotrophe. Les ressources génomiques produites dans cette thèse ouvrent la voie à des futures études intégrant épigénomique, génomique des populations et transcriptomique chez les *Melampsora* spp. et d'autres rouilles.

Abstract

Rust fungi (Basidiomycetes, Pucciniomycotina, Pucciniales) possess some of the largest and most complex genomes in the fungal kingdom, driven in large part by the massive proliferation of transposable elements (TEs). This thesis investigates the evolutionary impact of TEs on rust fungal genome architecture and biology, using a combination of comparative genomics, high-quality haplotype-resolved assemblies, and 3D genome architecture analysis. We first implemented a standardized TE annotation pipeline across 12 Pucciniales and related species, revealing that large TE-rich genomes are an ancient and signature feature of the order. The TE landscape reflects both persistent ancestral elements and lineage-specific bursts of activity, notably involving TIR and LTR-Gypsy elements. These results suggest that TE accumulation in Pucciniales is a dynamic and recurrent process, despite the presence of conserved elements participating in silencing mechanisms. To explore genome structure at high resolution, we generated haplotype-phased assemblies for two poplar rust species, *Melampsora larici-populina* and *Melampsora allii-populina*, using long-read and Hi-C sequencing. These assemblies, among the most complete for rust fungi to date, confirmed that both species harbour 18 chromosomes and uncovered non-random TE distribution in centromeric and subtelomeric regions. We also compared *M. larici-populina* strains from contrasting reproductive modes and found similar TE content between a long-term clonal lineage and a sexual strain, challenging the idea that clonal propagation leads to unchecked TE accumulation. Together, this work demonstrates that TEs are not only drivers of genome expansion in rust fungi but also key contributors to genome structure, stability, and possibly gene regulation. Our findings support a model in which TE-rich genomes potentially provide increased plasticity, enabling adaptation to complex life cycles and host interactions in biotrophic fungi. The resources and insights developed here pave the way for future studies integrating epigenomics, population genomics, and transcriptomics in *Melampsora* spp. and other rust fungi.