LARA MATTANA FERST

TALES FROM A GENOME HEAVILY AFFECTED BY RIP: UNRAVELING *Tc1/Mariner* AND MITE TRANSPOSABLE ELEMENTS IN *Colletotrichum lindemuthianum*

Dissertation submitted to the Agriculture Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Marisa Vieira de Queiroz

Co-adviser: Mateus Ferreira Santana

VIÇOSA – MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus

T
F399t
2023

Ferst, Lara Mattana, 1999-
Tales from a genome heavily affected by RIP : unraveling *Tc1/mariner* and MITE transposable elements in *Colletotrichum lindemuthianum* / Lara Mattana Ferst. - Viçosa, MG, 2023.
1 dissertação eletrônica (73 f.): il. (algumas color.).
Texto em inglês.
Inclui apêndices.
Orientador: Marisa Viera de Queiroz
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Microbiologia, 2023.
Referências bibliográficas: .
DOI: https://doi.org/10.47328/ufvbbt.2023.617
Modo de acesso: World Wide Web.

1. *Colletotrichum lindemuthianum*; 2. Elementos de DNA transponíveis; 3. Genoma; I. Queiroz, Marisa Viera de II. Universidade Federal de Viçosa.. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola III. Título

CDD 22. ed. 579.657

Bibliotecário(a) responsável: ALICE REGINA PINTO PIRES CRB-6/2523

LARA MATTANA FERST

TALES FROM A GENOME HEAVILY AFFECTED BY RIP: UNRAVELING *Tc1/Mariner* AND MITE TRANSPOSABLE ELEMENTS IN *Colletotrichum lindemuthianum*

Dissertation submitted to the Agriculture Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: July 21, 2023.

Assent:

Documento assinado digitalmente
gov.br
LARA MATTANA FERST
Data: 05/10/2023 12:14:13-0300
Verifique em https://validar.iti.gov.br

Lara Mattana Ferst
Author

Documento assinado digitalmente
gov.br
MARISA VIEIRA DE QUEIROZ
Data: 05/10/2023 14:25:54-0300
Verifique em https://validar.iti.gov.br

Marisa Vieira de Queiroz
Adviser

ACKNOWLEDGMENTS

To my parents, my brother, and my entire family, for everything, always. I owe everything I am to you.

To my advisor Marisa, who embraced the challenge and opened the doors of her laboratory, introducing me to the wonderful world of molecular biology of fungi.

To my co-advisor Mateus, for all the patience and everything he taught me.

To Leandro for accepting being a part of the evaluation committee and contributing to the construction of this work.

To all the public employees who worked tirelessly to provide various services and maintain the excellent infrastructure of UFV.

To my colleagues from the Laboratory of Microorganism Molecular Genetics and the Laboratory of Microbial Ecology, for all the coffee-fueled heated discussions and for the partnership.

To the many friends who were my emotional support during my stay in Viçosa: Ross, Jaque, Yam, Raissa, Gabi, Gui, João, Hirlanda, and many others. You are the family I chose.

To Hugo, for showing me that love can be stronger than distance and for always being present even when far away.

To all the investments made in education, science, and technology, which are so important and often neglected, allowing me to study at a public, free, and excellent university.

I thank the funding institutions FAPEMIG, CNPq, and CAPES. This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Finally, I express my gratitude to all those who fight for the valorization of Brazilian research and education.

"In the ever-changing tapestry of life, evolution weaves its threads through the intricate patterns of genome structure, where transposable elements dance as agents of transformation."

(Unknown author)

ABSTRACT

FERST, Lara Mattana, M.Sc., Universidade Federal de Viçosa, July, 2023. **Tales from a genome heavily affected by RIP: unraveling *Tc1/mariner* and MITE transposable elements in *Colletotrichum lindemuthianum***. Adviser: Marisa Vieira de Queiroz. Co-adviser: Mateus Ferreira Santana.

The *Colletotrichum* genus comprises fungal pathogens that inflict severe diseases on a wide range of hosts, including economically important plants. *Colletotrichum lindemuthianum* is the causative agent of anthracnose in common beans (*Phaseolus vulgaris*), leading to significant production and quality losses in this essential legume crop. Transposable elements are mobile genetic units found across all life domains that play a significant role in genomic plasticity, which is particularly relevant for phytopathogenic fungi such as *C. lindemuthianum*. These elements are classified into two distinct classes based on their transposition mode: Class I transposons rely on an RNA intermediate to transpose, while Class II transposons transpose directly as DNA. Additionally, transposons can be divided as autonomous and non-autonomous elements, with the latter depending on other transposable elements for their mobility due to the absence of essential functional components. The *Tc1/mariner* is a widely distributed Class II transposon superfamily. Miniature-inverted repeat elements (MITEs) are non-autonomous elements derived from Class II elements like the *Tc1/mariner* family. Considering the crucial role of transposons in generating genetic variability, this study aimed to identify and characterize *Tc1/mariner* and MITE transposable elements within the genome of *C. lindemuthianum*, using an *in-silico* approach. A total of 615 sequences related to *Tc1/mariner* elements were identified, which represented 0.78% of the genome. Among them, 536 copies were considered degenerated due to nearly unrecognizable terminal inverted repeats. Only one family of complete elements was found, and a derived MITE family was identified. Both the parental and derived families displayed a strong tendency to be found inserted within gene promoter regions. Moreover, two other MITE families were characterized, but the elements that originated them could not be identified. The elements from both families were primarily located within transposon-enriched regions. All complete elements showed putative transposase ORFs interrupted with multiple stop codons, suggesting that there are no active

Tc1/mariner elements in *C. lindemuthianum*. Our analysis revealed that these elements were heavily affected by repeat-induced point mutations (RIP). Furthermore, we identified copies of both methyltransferases involved in RIP, which further supports that this mechanism or a RIP-like mechanism is active in *C. lindemuthianum*. The insights gained from this study contribute to a better understanding of the *Tc1/mariner* and MITE landscape as well as defense mechanisms against transposon proliferation of *C. lindemuthianum*, providing a foundation for further investigations into the genetic variability and transposable element exaptation of this economically important plant pathogen.

Keywords: Class II transposons. Non-autonomous transposons. Repetitive DNA. Genome defense.

RESUMO

FERST, Lara Mattana, M.Sc., Universidade Federal de Viçosa, julho de 2023. **Contos de um genoma fortemente afetado por RIP: desvendando elementos transponíveis *Tc1/mariner* e MITE em *Colletotrichum lindemuthianum*.** Orientadora: Marisa Vieira de Queiroz. Coorientador: Mateus Ferreira Santana.

O gênero *Colletotrichum* compreende patógenos fúngicos que causam doenças graves em uma ampla variedade de hospedeiros, incluindo plantas economicamente importantes. *Colletotrichum lindemuthianum* é o agente causador da antracnose no feijão comum (*Phaseolus vulgaris*), resultando em perdas significativas de produção e qualidade nas culturas dessa leguminosa. Elementos transponíveis são unidades genéticas móveis que desempenham um papel significativo na plasticidade genômica, o que é especialmente relevante para fungos fitopatogênicos como *C. lindemuthianum*. Esses elementos são classificados em duas classes distintas com base em seu modo de transposição: os transposons da Classe I dependem de um intermediário de RNA, enquanto os transposons da Classe II transpõem diretamente como DNA. Além disso, os transposons podem ser divididos em elementos autônomos e não autônomos, sendo que estes últimos dependem de outros elementos transponíveis para sua mobilidade devido à ausência de componentes funcionais essenciais. A *Tc1/mariner* é uma superfamília de transposons de Classe II amplamente distribuída. MITEs (do inglês: *Miniature-Inverted Repeats)* são elementos não autônomos derivados de elementos da Classe II, como a família *Tc1/mariner*. Considerando o papel crucial dos transposons na geração de variabilidade genética, este estudo teve como objetivo identificar e caracterizar elementos transponíveis *Tc1/mariner* e MITEs no genoma de *C. lindemuthianum*, utilizando uma abordagem *in-silico*. Um total de 615 elementos foi identificado, representando 0,78% do genoma. Dentre eles, 536 cópias foram consideradas degeneradas devido às repetições terminais invertidas irreconhecíveis. Apenas uma família de elementos completos foi encontrada, da qual uma família MITE derivada foi identificada. Ambas as famílias parental e derivada mostraram uma forte tendência de inserção nas regiões promotoras de genes. Além disso, outras duas famílias de MITE foram caracterizadas, mas não foi possível identificar os elementos que as originaram. Os elementos de ambas as famílias estavam principalmente

localizados em regiões enriquecidas de transposons. Todos os elementos completos apresentaram ORFs de transposase putativas interrompidas por múltiplos códons de parada, sugerindo a ausência de elementos ativos *Tc1/mariner* em *C. lindemuthianum*. Nossa análise revelou que esses elementos foram fortemente afetados por mutações pontuais induzidas por repetição (RIP). Além disso, identificamos cópias de ambas as metiltransferases envolvidas no RIP, o que indicam ainda mais que esse mecanismo ou um mecanismo semelhante a RIP está ativo em *C. lindemuthianum*. Os resultados obtidos neste estudo contribuem para uma melhor compreensão da paisagem *Tc1/mariner* e MITE, bem como dos mecanismos de defesa contra a proliferação elementos transponíveis de *C. lindemuthianum*, fornecendo uma base para futuras investigações sobre a variabilidade genética e exaptação de elementos transponíveis deste fitopatógeno economicamente importante.

Palavras-chave: Transposons da Classe II. Transposons não-autônomos. DNA repetitivo. Defesa do genoma.

SUMMARY

1. INTRODUCTION

Colletotrichum species are fungal pathogens that cause severe diseases on a wide range of hosts, including economically important plants. The significance of this genus is further reinforced by its historical recognition as an exceptional system for studying the molecular and cellular mechanisms underlying the establishment of the pathogen-host interface. *Colletotrichum lindemuthianum* is the causative agent of anthracnose in common beans (*Phaseolus vulgaris* L.), leading to substantial losses in both production and grain quality.

Transposable elements (TEs), which are mobile genetic units found across all life domains, play a significant role in genomic plasticity. TEs can induce genomic changes, as they can disrupt genes, promote alternative transcripts, gene duplication, large genomic rearrangements, and can interfere with gene regulation. This feature is especially important in fungal plant pathogens, that rely on dynamic genomes to co-evolve with their hosts. Eukaryotic transposons are classified into two distinct classes based on their transposition mechanisms. Class I transposons rely on an RNA intermediate to transpose, while Class II transposons, also known as DNA transposons, do not require an RNA intermediate for transposition. TEs can be further divided into autonomous and non-autonomous elements, which lack the necessary functional elements for their transposition and are therefore dependent on other elements for their mobility. A prominent group of the DNA transposons is the *Tc1/mariner* superfamily, which is widely distributed among eukaryotes. Miniature-inverted repeat elements (MITEs) are non-autonomous elements derived from class II elements such as those in the *Tc1/mariner* family. TEs proliferation is often accompanied by fungal defense mechanisms, such as the repeat-induced point mutations (RIP), which can ultimately lead to TE inactivation.

Considering the important role of transposons in generating genetic variability in fungal plant pathogens, the goal of this study was to use an *in-silico* approach to identify and characterize *Tc1/mariner* and MITE TEs within the genome of *C. lindemuthianum*. In addition, the presence of mutations resulting from the action of RIP or a RIP-like mechanism in these elements and the presence of genes coding for proteins essential for RIP activity in the genome of *C. lindemuthianum* were investigated.

2. LITERATURE REVIEW

2.2 *Colletotrichum lindemuthianum*: the causal agent of the common bean anthracnose

Representatives of the *Colletotrichum* genus are plant pathogenic fungal species causing serious diseases in a wide range of hosts, including economically important plants (CANNON *et al.*, 2012; NABI *et al.*, 2022). Given its agricultural and scientific relevance, the genus was elected as the eighth most relevant phytopathogenic fungi genera (DEAN *et al.*, 2012). Furthermore, the genus relevance is strengthened by the fact that, historically, *Colletotrichum* species and their hosts have been considered exceptional systems for studying the molecular and cellular basis of the establishment of the pathogen-host interface (PERFECT *et al.*, 1999).

Colletotrichum lindemuthianum (Sacc. & Magnus) Briosi & Cavara is the etiological agent of anthracnose in common bean (*Phaseolus vulgari*s L.), resulting in large losses in production and quality of this legume grain (PADDER *et al.*, 2017). It belongs to the order Glomerellales, within the subclass Hypocreomycetidae and the Ascomycota phylum. Within the genus, *C. lindemuthianum* is part of the *Colletotrichum orbiculare* species complex alongside relevant pathogens of weeds and agricultural crops (DAMM *et al.*, 2013). *Colletotrichum lindemuthianum* presents a remarkably high intra-specific variability, which limits the long-term use of anthracnose-resistant strains in common bean cultivation (NABI *et al.*, 2022). To date, 298 distinct races belonging to *C. lindemuthianum* are reported worldwide, of which 89 have been registered in Brazil. This makes Brazil the country harboring the most significant variability within the species (NUNES *et al.*, 2021; PAULINO *et al.*, 2022). Therefore, it is of extreme importance to identify the different factors from which this diversity arises.

Colletotrichum lindemuthianum host, the common bean, is one of the most relevant legume crops utilized for human consumption. Its seeds and pods have a complete nutritional composition, presenting a high protein content combined with carbohydrates, fibers, and minerals (DELFINI *et al.*, 2020; DIDINGER *et al.*, 2022; PATHANIA; SHARMA; SHARMA, 2014). In 2019, the global area for common beans was equivalent to 33.1 million hectares and its production was 28.9 million tons, as

reported by FAO. The importance of this grain legume is even more accentuated in developing countries such as Brazil, which is the second largest dry bean producer accounting for approximately 10% of the world's production (FAOSTAT, 2021). Many different biotic factors act as stressors in the common bean chain of production, among which fungal pathogens, such as *C. lindemuthianum*, are particularly relevant, especially in tropical and subtropical areas. Common bean anthracnose exhibits classical symptoms that include deep, shrunken rust-colored lesions found on bean pods, which later turn into sunken cankers with black ring borders (PASTOR CORRALES; TU, 1989). These cankers make the seeds unsuitable for consumers, rendering them unmarketable and causing substantial economic loss. As the disease progresses, it can ultimately lead to plant death (PASTOR CORRALES; TU, 1989).

Colletotrichum lindemuthianum has a hemibiotrophic infectious cycle characterized by two distinct phases, corresponding to biotrophic and necrotrophic nutrition modes. The infection process begins with conidial adhesion and germination followed by appressoria development, penetration of the plant's epidermis and biotrophic hyphae development (O'CONNELL; BAILEY; RICHMOND, 1985). The biotrophic phase is transient and asymptomatic, during which the fungus spreads to adjacent living cells. Then, the necrotrophic phase begins with secretion of enzymes and toxins responsible for causing plant cell death. New conidia are then formed in the necrotic tissue, completing the infection cycle (O'CONNELL; BAILEY; RICHMOND, 1985).

The sexual stage of *C. lindemuthianum* was first described in 1913 in laboratory settings by Shear and Wood as *Glomerella lindemuthiana* (PADDER *et al.*, 2017). The teleomorph of the species was rediscovered in the seventies and renamed as *Glomerella cingulata* f. sp. *phaseoli* (KIMATI; GALLI, 1970). *In vitro* studies have demonstrated that *C. lindemuthianum* has an unusual mating system, where both parental isolates present MAT1-2 idiomorph. However, sexual reproduction of this pathogen seems to be rare in the field (RODRÍGUEZ-GUERRA *et al.*, 2005), and it is likely that other factors influence genomic plasticity, such as the transposable element landscape of the fungus.

2.2 From concept to classification of transposable elements

Barbara McClintock's discovery of transposable elements (TEs) in the mid-20th century revolutionized the field of genetics (ALMOJIL *et al.*, 2021). Prior to her groundbreaking work, genetic information encoding for a specific trait was believed to be arranged in a fixed manner. However, McClintock's observation of mutable loci that were responsible for the diverse variegated kernel color in maize challenged this conventional theory (MCCLINTOCK, 1956), though it initially received little attention. Nowadays, it is widely recognized that the presence of potentially mobile genetic elements, such as the transposon system that McClintock identified, is common rather than exceptional (WELLS; FESCHOTTE, 2020).

Transposons are mobile genetic units that can be found throughout all domains of life (BOURGEOIS; BOISSINOT, 2019). Initially considered to be “junk DNA” or “parasitic elements”, TEs are now recognized as significant contributors to genome structure and evolution (WELLS; FESCHOTTE, 2020). This status is also valid for fungal genomes, among which TE content ranges from ~3.3% in *Saccharomyces cerevisiae* (CARR; BENSASSON; BERGMAN, 2012) to ~93% in *Phakopsora pachyrhizi* (GUPTA *et al.*, 2023).

David Finnegan proposed in 1989 the first TE systematic organization system in which eukaryotic TEs are divided into two distinct classes based on their transposition mechanism: Class I transposons (RNA or retrotransposons) and Class II (DNA transposons). Later on, many other classification systems were proposed to subdivide these classes into a unified way, among which we will follow Wicker et al., (2007) eukaryotic TEs classification proposal. Each of these two major groups can be further divided into subclasses and orders based primarily on their replication and/or integration mechanism. Then, orders can be divided into superfamilies, grouped based on enzyme organization, non-coding domains or target site duplications (TSD) size, and families, which are grouped according to identity.

Class I TEs require an RNA intermediate that is converted into DNA through a TE-encoded reverse transcriptase to be integrated into a new genomic region (WICKER *et al.*, 2007). It generates a new copy in every replication cycle while the original copy stays in place, and, therefore, often constitutes a significant part of the genome size expansion of eukaryotes (ELLIOTT; GREGORY, 2015). They are divided into five orders according to their insertion mechanisms and organization:

long terminal repeats (LTR) retrotransposons, dictyostelium intermediate repeat sequence (DIRS)-like elements, Penelope-like elements (PLEs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) (WICKER *et al.*, 2007).

Class II TEs do not require an RNA intermediate to transpose, hence the DNA transposon name, and are characterized by terminal inverted repeats (TIR) that act as targets for transposase recognition. Depending on the family, these TIRs can be flanked by TSD, generated in the integration process. This class can be divided into two subclasses (I and II). TEs from Subclass I are often called “cut-and-paste” transposons due to their conservative mechanism of transposition (double-strand transposition). While this transposition process itself does not generate a new TE copy, members of this subclass can duplicate during host DNA replication, when transposing from replicated to un-replicated chromatids. They can also increase their copy number through homologous recombination to repair double-strand DNA breaks left at their excision site, regenerating the TE at the point where it excised (WELLS; FESCHOTTE, 2020). Cut-and-paste transposons are divided into two orders based on the recombinase used for transposition. Members of the TIR order are considered the most widespread and diverse TEs and their transposase is possibly the oldest and most ubiquitous protein on Earth (AZIZ; BREITBART; EDWARDS, 2010). These cut-and-paste TEs are further divided into ten different superfamilies (*hAT*, *Mutator*, *Merlin*, *Transib*, *P-element*, *PiggyBac*, *PIF/Harbinger*, *CACTA*, *Tc1/mariner*, and *Pogo*) based on TIR sequence similarity and length of TSD (GAO *et al.*, 2020; WICKER *et al.*, 2007). TEs belonging to the Crypton order transpose via a tyrosine recombinase. On the other hand, members of the order TIR code for a DDE/D transposase (WICKER *et al.*, 2007).

Finally, Subclass II, also known as the “copy-and-paste” TEs, is divided into two orders. Helitron order members replicate through a rolling circle mechanism and do not present TIRs (KAPITONOV; JURKA, 2007), while Polintons/Mavericks are large (15–20 kb), complex elements that can code for up to twenty different ORFs, and are flanked by long terminal repeats (FESCHOTTE; PRITHAM, 2007). They replicate via a single-strand break and direct synthesis of a DNA copy (WICKER *et al.*, 2007).

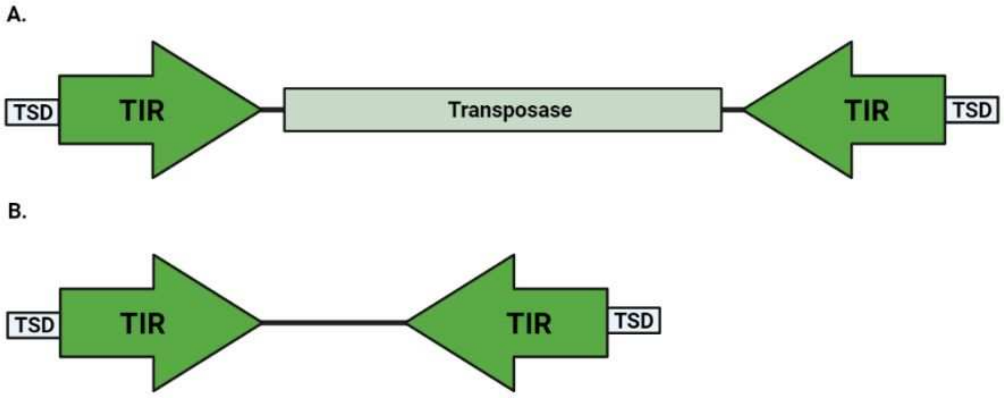
Regardless of class, TEs can be further characterized into autonomous or non-autonomous elements. DNA repair errors can result in different internal deletions and create non-autonomous truncated or degenerated TE copies. Nevertheless, transposition relies on terminal repeats only, so even copies without a functional transposase still can be mobilized by proteins coded by autonomous intact elements based on the similarity of their TIRs (WICKER *et al.*, 2007). They can be derived from class II elements (MITEs- Miniature Inverted-repeat Terminal Elements); from non-LTR retrotransposons (SINEs-short interspersed nuclear element), and from the LTR retrotransposons (TRIMs-Terminal-repeat Retrotransposon In Miniature and LARDs-Large Retrotransposon Derivates).

2.3 *Tc1/Mariner* superfamily and miniature Inverted-repeat elements

Among the cut-and-paste TEs, the Tc1/*mariner* superfamily is known to be one of the most widely distributed mobile elements in eukaryotes, which indicates a very ancient origin (CAPY *et al.*, 1996; PLASTERK, 1996).

Member of this superfamily have a fairly simple structure consisting of a single ORF coding for a transposase enclosed by a pair of TIRs (Fig. 1A). Tc1/*mariner* elements transposition begins with the cohesive cleavage of the 5'-end of both repeats by two transposases, forming a DNA-transposase complex. The transposases then recognize specific DNA motifs elsewhere, often TA nucleotides, and promote the element integration at the new site (Fig. 2). The repair process of the double-strand cleavage during the TE integration creates TSD of 2 to 8 pb (WICKER *et al.*, 2007). Members of this superfamily transpose via a DDE/D transposase, named after their catalytic domain, a motif of three residues: two aspartic acids (Ds) and either another aspartic acid (D) or a glutamic acid (E). The triad's three-dimensional structure facilitates nucleophilic reactions that cleave DNA by creating a catalytic site that coordinates two divalent metal ions (YUAN; WESSLER, 2011). The transposase also holds helix-turn-helix (HTH) DNA binding motifs, capable of recognizing the TIRs (PIETROKOVSKI; HENIKOFF, 1997).

Figure 1 – (A) Structure of TEs in the *Tc1/mariner* superfamily and (B) non-autonomous elements derived from Class II TEs, known as MITES (Miniature Inverted-repeat Terminal Elements). TIR: terminal inverted repeats. TSD: target site



duplication.

Source: the author.

Tc1/mariner superfamily is named after the first elements of its kind identified in the nematode *Caenorhabditis elegans* (Transposon *C. elegans* number 1, *Tc1*) (EMMONS *et al.*, 1983) and the fly *Drosophila mauritiana* (*mariner*) (JACOBSON; MEDHORA; HARTL, 1986). In 1990, the bacterial insertion sequence *IS630* from the *Shigella* genome was associated with *Tc1* due to their common TA TSD that results from transposition (TENZEN; MATSUTANI; OHTSUBO, 1990). Initially believed to belong to distinct groups, *Tc1*, *mariner*, and *IS630* were subsequently grouped together as the *IS630/Tc1*/*mariner* (ITm) group based on their similar mode of transposition involving a DNA intermediate, shared TA target site, and transposase sequence homology (ROBERTSON, 1995). Recently, the *pogo* elements, formerly considered part of the *Tc1/mariner* superfamily, have been separated into a different superfamily of its own (GAO *et al.*, 2020).

Lorrain et al. (2021) identified that among 26 different isolates belonging to the *Zymoseptoria* complex genes present in only a subset of lineages were specifically associated with TIR and MITE elements, which could indicate their association with structural changes in the genomes. They also found that these elements were older compared to retrotransposons (LORRAIN *et al.*, 2021).

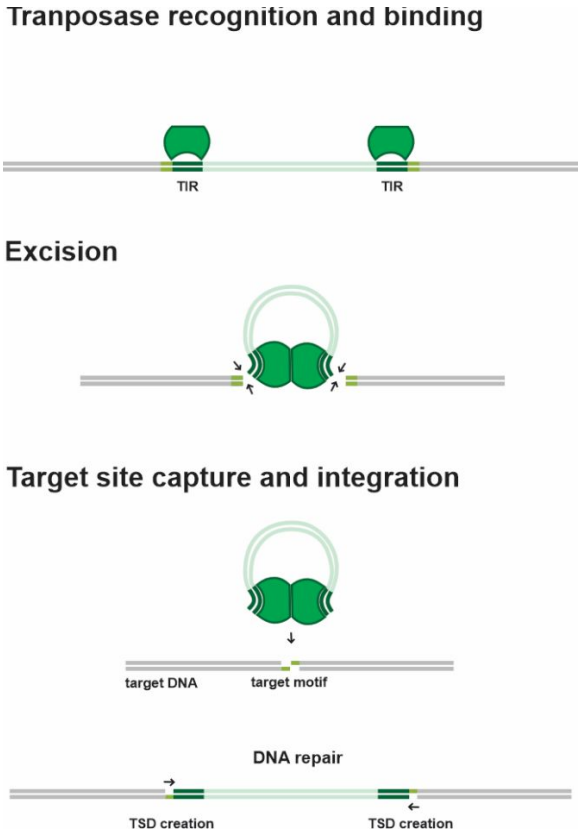
TEs belonging to the *Tc1/mariner* superfamily may play a significant role in driving genomic changes under stressful conditions. In a study conducted by Chen et al. (2015), the *Mftc1* transposon was utilized as an indicator to investigate the effects of fungal azoxystrobin fungicide exposure on transposon movement in *Monilinia fructicola*. This particular TE belongs to the *Tc1/mariner* superfamily and spans 1500 bp, featuring 28 bp terminal inverted repeats and a 4 bp site duplication, and was monitored over the duration of 12 weeks. The results revealed that after the mycelia was continuously exposed to azoxystrobin treatment an altered banding pattern of *Mftc1* arose in Southern blot analysis, a pattern change that was absent in control treatments (CHEN *et al.*, 2015). This finding suggests that prolonged exposure to sublethal fungicide doses can stimulate increased transposition of TEs such as *Mftc1.* Interestingly, the *Mftc1* transposon was discovered inserted in the upstream region of the P450 14α-demethylase (*MfCYP51*) gene, which is associated with fungicide resistance, in mycelia that had been exposed to sublethal doses of azoxystrobin (CHEN *et al.*, 2015). The presence of *Mftc1* transposon in the upstream region of the *MfCYP51* gene had been previously observed in another strain isolated by the same research group (LUO *et al.*, 2008). However, no noticeable impact on *MfCYP51* gene expression was observed in both cases (CHEN *et al.*, 2015; LUO *et al.*, 2008).

Another example that illustrates the increased transposition of a *Tc1/mariner* element under environmental stress can be observed in the *Aspergillus oryzae* fungus. *Crawler* is an active *Tc1/mariner* TE that span 1290 pb, is flanked by 28 bp TIRs, and 2 bp TSD (OGASAWARA *et al.*, 2009). The researchers have successfully shown that the putative transposase encoded by this element becomes upregulated in response to heat shock and exposure to $CuSO_4$. To further validate this finding, the transposon trapping technique was employed using the *crnA* gene (which encodes a nitrate permease) and the *niaD* gene (which encodes a nitrate reductase) as traps to recover and confirm new insertions (OGASAWARA *et al.*, 2009). Furthermore, other stressors such as low temperatures, oxidative stress, and UV irradiation have also been demonstrated to induce transposition of *Tc1/mariner* TEs (BOUVET *et al.*, 2008; CARR *et al.*, 2010; OGASAWARA *et al.*, 2009).

The *Tc1/mariner* superfamily has gained recognition because several of its elements have been employed as genetic tools (SANDOVAL-VILLEGAS *et al.*,

2021). A notable example is the Sleeping Beauty (SB) system, which is a synthetic transposon constructed from inactive fish *Tc1*-like sequences. SB is extensively used in genetic engineering for gene delivery and functional genomics in many different hosts (IVICS *et al.*, 1997; NARAYANAVARI *et al.*, 2017). *Fot1*, a fungal transposon *pogo*-like, is employed as a tagging system to study gene expression regulation in fungi (DESCHAMPS *et al.*, 1999). The *impala Tc1/mariner* element from *Fusarium oxysporum*, induced by low temperatures, is one of the most often employed elements for insertional mutagenesis in fungi (LANGIN; CAPY; DABOUSSI, 1995; VILLALBA *et al.*, 2001; LÓPEZ-BERGES *et al.*, 2009; DING *et al.*, 2022). Furthermore, *Minos*, a *Tc1*-like TE isolated from the fly *Drosophila hydei*, has been employed for genetic engineering in many different organisms, including fungi (EVANGELINOS *et al.*, 2015).

Figure 2 – *Tc1/mariner* mechanism of transposition. The terminal inverted repeats (TIRs) sequences are recognized by two transposases that join with each other and promote double-strand breaks, leading to the element excision. The complex formed by the DNA and transposases recognize specific DNA motifs (often TAs), promote the double-strand break of the target DNA, and insert the transposon, creating target site duplication (TSD).



Source: adapted from WALTER (2015).

TEs contain sequences that primarily serve to regulate the transcription of genes related to transposition. Nevertheless, mounting evidence suggests that TEs can also function as conserved regulatory sequences, providing enhancers, promoters, and insulators to the host genome (CHUONG; ELDE; FESCHOTTE, 2017). Specifically, certain *Tc1/mariner* elements have been shown to possess a "blurry" promoter capability, enabling them to promote transcription activity across diverse cell types such as bacteria, humans, yeast, and insects (PALAZZO *et al.*, 2019). As a result, members of this superfamily offer great potential to be employed as tools to the development of a universal expression vector that could be applied to multiple different organisms (PALAZZO; MARSANO, 2021).

MITEs are non-autonomous elements (Fig. 1B) derived from class II elements such as *Tc1/mariner* that usually range from 50 to 800 bp (FATTASH *et al.*, 2013; PEGLER *et al.*, 2023). In many different organisms, MITEs have been found to be closer to genes than other TE types (OKI *et al.*, 2008; SANTIAGO *et al.*, 2002; STALDER *et al.*, 2023). In fact, MITEs are so often associated to coding sequences of genes that researchers have developed a way to identify novel effector gene candidates through a MITE proximity criterion in *F. oxysporum* (SCHMIDT *et al.*,

2013; VAN DAM; REP, 2017). MITEs also present higher GC content and tend to be less afflicted by defense mechanisms such as RIP, likely due to their small size (OGGENFUSS; CROLL, 2023; STALDER *et al.*, 2023).

Non-autonomous elements such as MITEs have long been recognized to be associated with promoter regions and have been shown to influence gene expression (WESSLER; BUREAU; WHITE, 1995). Particularly in phytopathogenic fungi, multiple small non-autonomous TE insertions in promoter regions have been identified and associated with the development of fungicide resistance phenotypes. For instance, the *Mona* element in the *M. fructicola* has been found to enhance *MfCYP51* gene expression, thereby conferring resistance to sterol demethylation inhibitors (DMI) fungicides (CHEN *et al.*, 2017; LUO; SCHNABEL, 2008). Furthermore, a 169 bp TE named *EL 3,1,2* insertion with transcriptional enhancer properties upstream the *ViCYP51* gene was found to be correlated to difenoconazole (a DMI fungicide) resistance in *Venturia inaequalis* (SCHNABEL; JONES, 2001; VILLANI *et al.*, 2016). A similar pattern was observed in the pathogen *Blumeriella jaapii*, where the researchers analyzed 59 different DMI resistant isolates and found that truncated TEs with variable size were inserted within the upstream region of *BjCYP51* gene in all instances (MA *et al.*, 2006).

Within the citrus pathogen *Penicillium digitatum*, a 199 bp MITE-like element known as PdMLE1 was identified within the promoter region of the *PdCYP51B* gene, where it acts as a driver for gene over-expression. This leads to the acquisition of resistance against DMI fungicides (SUN *et al.*, 2011). Multiple copies of PdMLE1 element are found strategically located within regulatory and coding regions of the genome. To validate its robust promoter activity, a fusion construct of PdMLE1 with the 5' portion of the GFP gene was introduced into *P. digitatum*, resulting in significant expression of the green fluorescence protein. The authors were also able to identify a 20 bp core sequence in the element necessary for its promoter activity, likely due to the recruitment of transcription factors. This experimental evidence underscores the strong promoter functionality of the PdMLE1 element, marking the first time the promoter activity of a TE driving fungicide resistance in fungi was experientially characterized (SUN *et al.*, 2013b).

2.4 Genome guardians: fungal strategies against the proliferation of repetitive sequences

The proliferation of TEs is often accompanied by fungal defense mechanisms, such as meiotic silencing by unpaired DNA (MSUD) (SHIU *et al.*, 2001), methylation induced pre meiotically (MIP) (ROSSIGNOL; FAUGERON, 1994), Quelling (RNA interference) (ROMANO; MACINO, 1992), and repeat-induced mutations (RIP) (SELKER, 2002; SELKER; STEVENS, 1987).

Among them, two consist of RNA silencing pathways: MSUD and quelling. The MSUD was first identified in *Neurospora crassa* and takes place during meiosis. This reversible phenomenon mediates the silencing of any unpaired DNA together with all their homologs DNA (SHIU *et al.*, 2001). Alternately, the quelling mechanism operates during asexual development (DANG *et al.*, 2011). This mechanism was also discovered in *N. crassa* and characterized by the temporary suppression of gene expression of repeated sequences (ROMANO; MACINO, 1992). This process takes place during the vegetative growth phase and affects transgenes and endogenous genes (COGONI *et al.*, 1996).

A different example of fungal defense against repetitive sequence proliferation is the MIP mechanism. MIP was first identified in *Ascobolus immersus* (GOYON; FAUGERON, 1989) and occurs during sexual development. It consists of the recognition of DNA repeats in tandem during the sexual development in the premeiotic phase and induces reversible C5 methylation of cytosines within these regions, which are kept in the vegetative mycelium leading to gene silencing (ROSSIGNOL; FAUGERON, 1994). *De novo* methylation in MIP context is dependent on the MASC1 protein (*masc1*; MTase from *Ascobolus* 1) (MALAGNAC *et al.*, 1997).

Another defense mechanism, RIP, was first identified in *N. crassa* (SELKER *et al.*, 1987) but it was later found, either experimentally or through bioinformatic analysis, to be widespread among many other fungal species (JOHN CLUTTERBUCK, 2011). Fungi with active RIP are capable of recognizing closely positioned duplication as short as ~150 base pairs or widely separated repeats as short as ~400 base pairs (GLADYSHEV; KLECKNER, 2017) and trigger irreversible C:G to T:A transitions within these sequences (SELKER, 2002).

The RIP molecular mechanism still has not been completely uncovered. In the model system *N. crassa*, the mechanism relies on the RID (RIP deficient 1, coded by the *RID1* gene; FREITAG *et al.*, 2002) and DIM2 (decrease in DNA methylation-2, coded by the *dim-2* gene; KOUZMINOVA; SELKER, 2001) proteins, both members Dnmt1 protein superfamily. While the RID mediated RIP targeting restrictively repetitive sequences during the premeiotic phase of the sexual cycle, in *N. crassa,* the DIM-2 protein is not limited to the sexual cycle and duplicated sequences, also targeting adjacent non-repetitive regions during the vegetative cycle (GLADYSHEV; KLECKNER, 2017). After the C5 cytosine methylation, cytosine can mutate into thymine through the activity of a DNA deaminase activated in the premeiotic phase (GLADYSHEV; KLECKNER, 2017). In addition, a set of cofactors seem to be involved in the RIP pathway of this fungus, such as DIM-5, DIM-7, DIM-9, CUL4, DDB1, and HP1 (Heterochromatin Protein 1) (GESSAMAN; SELKER, 2017; LEWIS *et al.*, 2010). RIP machinery often targets CpA dinucleotide sites (which gives rise to TA-rich islands), but it can also target CpT and CpG as alternative dinucleotide sites (AMSELEM; LEBRUN; QUESNEVILLE, 2015; MIN *et al.*, 2020). It has been proposed that the nucleotide bias in fungal species is related to the presence of specific Dnmt1 subfamilies (AMSELEM; LEBRUN; QUESNEVILLE, 2015). Moreover, RIP is a potential source of genetic variation, as it can escape into neighboring sequences and increase the evolutionary rate of gene placed in TE-rich regions (FUDAL *et al.*, 2009; MÖLLER; STUKENBROCK, 2017).

Strains of the wheat pathogen *Zymoseptoria tritici* with non-functional copies of a *dim2* orthologue and, consequently, low levels of 5mC were isolated (DHILLON *et al.*, 2010; MÖLLER *et al.*, 2021). However, this feature is not widespread throughout the entire species, as some strains still retained a functional copy of this gene (MÖLLER et al., 2021). When comparing strains with or without functional *dim2* alleles in a one-year evolution experiment, it was possible to observe that when *dim2* is present the nucleotide composition of TEs changed, with a higher frequency of C to T transition in CpA sites. The phenomenon was observed during vegetative growth, as the cells in the experiment where dividing exclusively through mitosis, and is strong evidence of a mitotic version of RIP DIM2-dependent that still needs further investigation (MÖLLER *et al.*, 2021).

In a comprehensive study conducted by Lorrain et al. (2021) on 26 different isolates belonging to the *Zymoseptoria* species complex, it was discovered that the extent of the RIP signature varied significantly among different strains. RIP affected between 14.6% to 34.5% of the entire genomes of *Z. tritici*. Interestingly, the degree of RIP was found to be correlated with the geographic origin of the isolate. They found that isolates farther from the evolutionary center of origin exhibited lower levels of RIP. These findings correlate with the loss of *dim2*, as the geographical origin of the *Z. tritici* strains that retained functional copies of *dim2* also coincide with the evolutionary center of origin of the species (LORRAIN *et al.*, 2021; MÖLLER *et al.*, 2021).

Wyk et al. (2021) analyzed a set of 58 representatives of the Ascomycota in order to identify to what extent the RIP can act within this phylum. The researchers found that RIP was present in the majority of lineages, with the exception of the Taphrinomycotina and Saccharomycotina classes. Notably, RIP is exceptionally widespread within the Pezizomycotina subphylum, particularly in the Dothideomycetes and Sordariomycetes classes. Among the species analyzed, *Pyrenophora teres* stood out for having the highest coverage of repeat content (57.5%) and being the most affected by RIP (VAN WYK *et al.*, 2021). The *RID* and *dim-2* genes were present in most of the lineages with active RIP, and within the five cofactors searched (DIM-5, DIM-7, DIM-9, CUL4, DDB1), most lineages had two or more of the five genes.

2.5 Transposable elements in fungal phytopathogens

TEs play a crucial role in generating genomic plasticity, which is an essential aspect of evolutionary biology and is critical for the persistence of species in changing environments. This plasticity enables organisms to adapt to new environmental cues and exploit novel niches. Plant pathogens, in particular, rely on this capability as they coevolve with their hosts, facing selective pressure to enhance colonization effectiveness while evading host detection (MAT RAZALI; CHEAH; NADARAJAH, 2019). Dynamic genomes are often exhibited by fungal pathogens, with significant variations in both size and composition, even among closely related species or different strains from the same species. These variations predominantly arise from genomic rearrangements and disparities in repetitive DNA content (MÖLLER; STUKENBROCK, 2017).

TEs contribute to the genomic plasticity of phytopathogens in several ways. Directly, TEs can disrupt genes by integrating into coding sequences, promote alternative transcripts formation by integrating into introns, promote trans-duplication of genes, and interfere with gene regulation by inserting in proximity to genes (ALMOJIL *et al.*, 2021; MAT RAZALI; CHEAH; NADARAJAH, 2019; SCHMITZ; BROSIUS, 2011).

Indirectly, TEs can facilitate ectopic homologous recombination (recombination of homologous sequences located in non-homologous positions in the genome) leading to chromosomal rearrangements, such as deletions, inversions, and transversions (LORRAIN *et al.*, 2021; TSUSHIMA *et al.*, 2019). In fact, machine learning predictions identified TEs as the probable main cause to sequence rearrangements in the pangenome of the wheat pathogen *Z. tritici* (BADET *et al.*, 2021).

The genomic impact caused by TEs may have a beneficial impact on phytopathogens, leading to the generation of novel genetic variations and functional diversity. This can enhance their ability to infect and colonize host plants, evade plant defense mechanisms, and adapt to different environments. Overall, TEs contribute to the evolutionary potential of phytopathogens and their ability to adapt to changing host-pathogen dynamics (MAT RAZALI; CHEAH; NADARAJAH, 2019; MÖLLER; STUKENBROCK, 2017).

Although the presence of active TEs in plant pathogen genomes may increase their evolutionary rate and facilitate adaptation to new hosts and environments, the effects of TEs on genomic dynamics at the population level remain largely unexplored. Active TEs might also provide non-adaptive or even harmful changes (MAT RAZALI; CHEAH; NADARAJAH, 2019). Fouché et al. (2022) refer to the trade-off between genomic novelties brought by the derepression of TEs and the potential harmful changes that it may cause to the fungus as the ‘devil’s bargain’.

One possible way to mitigate the deleterious effects while keeping the advantages of the higher evolutionary rate due to TE activity is through the compartmentalization of the genome. In the two-speed genome architecture, highly variable TE-rich regions containing virulence-related genes such as effectors and secondary metabolic clusters are physically separated from TE-poor conserved segments containing housekeeping genes (DONG; RAFFAELE; KAMOUN, 2015). It

is still unclear whether this bipartite compartmentalization is a result of relaxed selection against TE activity in these regions or positive selection favoring the co-localization of virulence genes and TEs (TORRES; THOMMA; SEIDL, 2021). However, not all phytopathogens follow this model, as some species have evenly distributed repeated elements throughout the entire genome (FRANTZESKAKIS *et al.*, 2018).

2.6 *Colletotrichum* genomics and the current TE landscape of the genus

Due to the rapid decline in the cost of next-generation genome sequencing and great advances in sequence technology, there has been a significant growth in publicly accessible genomic data, a trend that can also be observed for the *Colletotrichum* genus (TSUSHIMA; SHIRASU, 2022). In 2012, O'Connell et al. reported for the first time the complete sequencing of two *Colletotrichum* spp. genomes. The genomic resource of *Colletotrichum graminicola* and *Colletotrichum higginsianum* set the dawn for the initial comparative plant-pathogen interactions from a genomic perspective, quickly followed by the genomic data release of *Colletotrichum graminicola* and *Colletotricum fructicola* (GAN *et al.*, 2013; O'CONNELL *et al.*, 2012). Since then, given the major agricultural, economical, and scientific relevance of *Colletotrichum* fungi, over a hundred *Colletotrichum* spp. genome assemblies have been released (TSUSHIMA; SHIRASU, 2022). However, there still exists a notable gap of information when it comes to gene annotation (CARBÚ *et al.*, 2020).

The availability of genomic data has supported the study of important features of evolutionary and functional genomics regarding *Colletotrichum* spp. It became possible to uncover a broader view of pathogenicity-related factors, such as effector genes repertoire (DE QUEIROZ *et al.*, 2019), secondary metabolite clusters (GAN *et al.*, 2013), and CAzyme profile (GAN *et al.*, 2016). Additionally, RNA-seq analyses became feasible due to comprehensive gene annotations availability, which provides information about the location of known genes within the genome, their function, and other related details that can help identifying differentially expressed genes and splicing events, for instance (GAN *et al.*, 2013; O'CONNELL *et al.*, 2012; SCHLIEBNER *et al.*, 2014).

Nonetheless, the first genomes available sequenced through second-generation sequencing technologies presented the limitation of small reads often leading to small orphan contigs, gaps, and misoriented sequences. This limits our understanding of highly repetitive genomic regions and present a challenge to the accurate mapping of TEs in the genome (ALKAN; SAJJADIAN; EICHLER, 2011; SHAHID; SLOTKIN, 2020). The development and increasing accessibility of third generation sequencing technologies revolutionized the detection of repetitive sequences in the genome (SHAHID; SLOTKIN, 2020).

The first *Colletotrichum* chromosome level genome assembly released was of *C. higginsianum,* with a combination of optical mapping and PacBio long-read sequencing (DALLERY *et al.*, 2017; ZAMPOUNIS *et al.*, 2016). The coverage by TEs went from 1.2% in the first genome assembly to 7% in the latter. This allowed for a comprehensive analysis of repeat content, RIP analysis, and to relate how TEs are spatially distributed in the genome (DALLERY *et al.*, 2017). Moreover, two minichromosomes highly enriched with TEs and with low gene density were identified, of which one is a dispensable chromosome necessary for regular disease manifestation (DALLERY *et al.*, 2017; PLAUMANN *et al.*, 2018). Genome comparison of another *C. higginsianum* strain with the chromosome scale assembly allowed the identification of intra-specific genomic variations, revealing strain-specific regions and effector candidate sets, besides large-scale rearrangements, which are often correlated with TEs presence (TSUSHIMA *et al.*, 2019).

Prior to the popularization of genome sequencing technologies, few TEs have been identified within the *Colletotrichum* genus (CROUCH *et al.*, 2008; DOS SANTOS *et al.*, 2012; HE *et al.*, 1996; ZHU; OUDEMANS, 2000). Among the identified transposons, *Cgt1* (*Colletotrichum gloeosporoides* transposon 1) stands out as the first transposon discovered in *Colletotrichum* spp. It was identified in the anthracnose causal agent in *Stylosanthes* sp., and belongs to the non-LTR class I transposon in the LINE order (HE *et al.*, 1996). Another transposon, known as *Cgret* (*C. gloeosporoides* retrotransposon), was also isolated from *C. gloeosporioides* and is classified as a class I transposon. However, it belongs to the *Gypsy* superfamily from the LTR order, and was found in strains infecting cranberry. Differently than the *Cgt1* transposon, deemed potentially inactive, the *Cgret* elements is structurally integral and possibly a functional TE (ZHU; OUDEMANS, 2000). Five different TEs

were identified in *Colletotrichum cereale*, the anthracnose agent in turfgrasses: two class II TEs (*Collect1*, *C. cereale* transposon 1; *Collect2*, *C. cereale* transposon 2), two LTR retrotransposons (*Ccret1*, *C. cereale* retrotransposon 1; *Ccret2*, *C. cereale* retrotransposon 2) and one non-LTR retrotransposon (*Ccret3*, *C. cereale* retrotransposon 3) (CROUCH *et al.*, 2008). The authors found multiple stop codons interrupting 21 out of the 35 TE copies identified, which combined with the strong A + T bias observed is consistent with a RIP-like pattern (CROUCH *et al.*, 2008). Lastly, *RetroCl1* (Retroelement *C. lindemuthianum* 1) was identified in a *C. lindemuthianum* strain and corresponds to the TRIM family of non-autonomous elements derived from LTR retrotransposon, with intact terminal direct repeats (DOS SANTOS *et al.*, 2012).

With higher resolution and more complete genome assemblies available, it has become evident that TEs constitute a significant portion of *Colletotrichum* genomes. Genome size of sequenced *Colletotrichum* species range from 44.15 Mb (*Colletotrichum truncatum* KLC.C-4) to 109.66 Mb (*Colletotrichum trifolii* MAFF 305078). Genome size expansion in the genus is positively correlated to the coverage of repeat content and negatively correlated to GC content (CHEN *et al.*, 2022). Species belonging to the *Colletotrichum orbiculare* species complex present larger genome size and repeat content when comparing to other species complex in the genus, contrasting with a lower GC content (CHEN *et al.*, 2022; GAN *et al.*, 2019). Regardless, information about repeat content, especially TE composition, from the *C. orbiculare* species complex is still very limited.

Colletotrichum species present a big variation in reported TE content: 3.63% – 5.63% in *C. fructicola,* 4.31% in *Colletotrichum scovillei*, 4.56% in *Colletotrichum lentis*, 4.89% in *C. truncatum*, 5.41% in *Colletotrichum orchidophilum*; 6.01% in *C. higginsianum*, 7.64% *Colletotrichum siamense*, 9.54% in *Colletotrichum chlorophyte*, 21.96% in *Colletotrichum lupini,* 25% in *Colletotrichum tanaceti*, 25.88% in *Colletotrichum graminicola*, and 44.88% in *C. orbiculare* (BECERRA *et al.*, 2023; BHADAURIA *et al.*, 2019; LELWALA *et al.*, 2019; RAO *et al.*, 2018). The most abundant elements in *Colletotrichum* spp. genomes are class I TEs from *Gypsy* superfamily and class II TEs from *Tc1/mariner*, with exception of *C. orbiculare* and *C. siamense*, in which elements belonging to the Class I superfamily *Copia* are in higher number (GAN *et al.*, 2021; RAO *et al.*, 2018).

Bioinformatic analysis have shown that many *Colletotrichum* species present RIP evidence in their genomes and favor CpA dinucleotide mutations (BRAGA *et al.*, 2014; LELWALA *et al.*, 2019; RAO *et al.*, 2018). However, this defense mechanism effectiveness among species within the genus varies in different levels. RIP is considered strong in *C. orbiculare*, *C. graminicola*, and *C. tanaceti* (BECERRA *et al.*, 2023; GAN *et al.*, 2013; LELWALA *et al.*, 2019; RAO *et al.*, 2018), while it is considered weak in *C. higginsianum*, *C. cereale*, *C. truncatum* (CROUCH *et al.*, 2008; DALLERY *et al.*, 2017; RAO *et al.*, 2018).

As an effort to identify and characterize *Tc1/mariner* transposons in the genome of *C. graminicola,* Braga et al. (2014) found no potentially active TEs due to a very strong RIP signature and multiple stop codons in the coding region of putative transposases, suggesting a severe RIP in this species. Even though sexual reproduction is considered rare in its natural habitat, a protein homologous to the cytosine methyltransferase protein RID protein from *N. crassa* was identified in *C. graminicola* (BRAGA *et al.*, 2014). *RID* and *Dim-2* genes were also found in the asexual pathogen *C. higginsianum* (DALLERY *et al.*, 2017), regardless of its weak RIP signature, and in the *C. tanaceti* genome, which is strongly affected by RIP (LELWALA *et al.*, 2019). Alternatively, in *C. truncatum*, which does not present a strong RIP, no RID-like protein was found, instead two genes harboring cytosine-specific methyltransferase domains homologous do *Dim-2* and RAD8 were found (RAO *et al.*, 2018). A RIP-like pattern was also identified in *C. cereale* (CROUCH *et al.*, 2008), but the presence of *RID* or *Dim-2*-like sequences still has not been investigated in this species. Because *C. cereale* is recognized as an asexual pathogen, it is still uncertain whether the RIP signature observed is reminiscent of an ancestral state capable of sexual reproduction or the fungus has a cryptic sexual cycle still elusive to science (CROUCH *et al.*, 2008).

It is believed that genomes affected by RIP-like mutations may display features of compartmentalized genomic distribution (RAO *et al.*, 2018). TE-rich gene sparse and higher A-T content sections represent a strong RIP and contrast with regions dense in genes, with higher G-C content, and fewer TEs. Often, TE-enriched areas are spatially associated with genes encoding effectors candidates as well as genes in secondary metabolism clusters, as seen in *C. higginsianum*, *C. tanaceti*, *C. truncatum*, *C. graminicola* (BECERRA *et al.*, 2023; DALLERY *et al.*, 2017; LELWALA

et al., 2019; RAO *et al.*, 2018; TAGA *et al.*, 2015). It has been shown that RIP might escape into neighboring sequences near repetitive elements (FUDAL *et al.*, 2009), which could increase the evolutionary rate of virulence-related genes placed in TE-rich zones and possibly provide fitness advantages (DA SILVA *et al.*, 2020; MÖLLER; STUKENBROCK, 2017). Queiroz et al. (2019) found signatures consistent with RIP in the repertoire of putative effectors of *C. lindemuthianum*. This suggests that RIP mechanism may play an important role in the evolution of these proteins in *C. lindemuthianum*, even though the RIP machinery still hasn't been investigated in this species.

The spatial proximity of putative effector genes and repeated sequences is often found within minichromosomes. These small chromosomes are also called dispensable or accessory chromosomes, as they are not required for vegetative growth and can vary in size and number among different strains of the same species (PLAUMANN; KOCH, 2020). They are usually TE dense, gene-sparse, harbor low G-C content, and may carry pathogenicity-related genes (DALLERY *et al.*, 2017). Minichromosomes can be found in many *Colletotrichum* spp., such as in *C. fructicola*, *C. higginsianum*, *C. lentis*, *C. gloeosporioides, C. lindemuthianum, Colletotrichum acutatum, Colletotrichum kahawae,* and *C. graminicola* (BECERRA *et al.*, 2023; BHADAURIA *et al.*, 2019; DALLERY *et al.*, 2017; GAN *et al.*, 2021; GARRIDO *et al.*, 2009; MASEL; IRWIN; MANNERS, 1993; O'SULLIVAN *et al.*, 1998; PIRES *et al.*, 2016; TAGA *et al.*, 2015). Curiously, the role played in the virulence expression by minichromosomes as demonstrated in *C. higginsianum* and *C. lentis* (BHADAURIA *et al.*, 2019; PLAUMANN *et al.*, 2018) is not translatable to all species, as Becerra et al. (2023) found no pathogenicity-related genes in the two minichromosomes found in a *C. graminicola* strain.

The *Colletotrichum* genus comprises more than 600 different species (CARBÚ *et al.*, 2020). Despite significant advances in recent years, many species within this genus still lack available genome sequences. Therefore, generating new genome sequences for unstudied species and sequencing multiple strains within the same species, along with characterizing their TE landscape, will contribute to a deeper understanding of the influence of these elements on genome plasticity and evolution. Additionally, this research will enable investigations into TE-mediated genetic changes linked to virulence and host adaptation, shedding light on the molecular

mechanisms underlying these traits. By unraveling the role of TEs in the genome dynamics of *Colletotrichum* species, scientists can gain a more comprehensive understanding of the evolutionary processes and genetic mechanisms that drive the diversification and pathogenicity of this fungal genus.

Taking into consideration the important role of transposons in generating genetic variability in fungal plant pathogens, the goal of this study was to use an *in-silico* approach to identify and characterize *Tc1/mariner* and MITE transposable elements within the genome of *C. lindemuthianum*. In addition, the presence of mutations resulting from the action of RIP or RIP-like mechanism in these elements and the presence of genes coding for proteins essential for RIP activity were investigated in the genome of *C. lindemuthianum*.

3. METHODS

3.1 Mining and characterization of *Tc1/Mariner* TEs in the *C. lindemuthianum* genome

The genomic sequencing of the isolate A2 2-3 of the race 89 of *C. lindemuthianum* was performed by the Laboratório de Genética Molecular de Microorganismos of the Universidade Federal de Viçosa (Silva et al. unpublished) with PacBio RSII (Pacific Biosciences, Menlo Park, CA, United States) technology. A hybrid approach was utilized for genome assembly, with reads generated by a previous sequencing in HiSeq 2500 Illumina platform (QUEIROZ *et al*., 2017) together with the PacBio data. The resulting assembly has 100,48 Mpb within 124

scaffolds, with a GC content of 37.25%. The program RepeatMasker (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org) was used to screen and classify repetitive sequences in the genome of *C. lindemuthianum*. This software tool detects interspersed repeats by comparison of genomic sequences of interest with the open database Dfam 3.0 (http://www.dfam.org). Inverted Repeats Finder (WARBURTON *et al*., 2004) was employed to locate Terminal Inverted Repeats (TIRs). The prediction of open reading frames (ORFs) within the putative transposase coding regions was performed using Expasy (http://expasy.org/) and WebAUGUSTUS (https://bioinf.uni-greifswald.de/augustus/submission) (HOFF; STANKE, 2013).

To analyze if the putative TEs show homology to transposases from other species, a BLASTX (www.ncbi.nlm.nih.gov/BLAST) alignment was conducted using the NCBI (National Center for Biotechnology Information) Non-redundant protein sequence database. TEs were prospected for the presence of DDE/D and HTH domains by submission into the Conserved Domain Database (CDD) (LU *et al.*, 2020). The Target Site Duplications (TSD) of the TEs were identified by directly searching the regions flanking the TEs.

Based on the analysis, the resulting sequences were classified into four categories: potentially active elements, complete sequences, MITEs, and degenerated elements. Complete elements were identified by conserved Terminal Inverted Repeats (TIRs) and Target Site Duplications (TSDs), but with putative transposase ORFs interrupted by stop codons. On the other hand, potentially active elements represent complete elements with intact TIRs, TSD, and ORFs that are characteristic of the *Tc1/Mariner* superfamily, indicating their potential for activity. Miniature Inverted repeat TEs (MITEs) present conserved TIRs and TSDs but lack a transposase encoding sequence. Finally, degenerated copies present interrupted ORFs and unrecognized TIRs. After the TE classification, a BLASTn alignment was conducted with the two most contiguous genomes of the *C. orbiculare* species complex, *C. orbiculare* and *C. trifolii* (GAN et al., 2019; GenBank accession numbers: AMCV02000007.1 and GCA_004367215.1, respectively), to verify if the TEs species-specific.

The classification system proposed by Wicker et al. (2007) was utilized to define families in this study. According to this system, families consist of TEs that exhibit more than 80% identity in their internal transposase domains, TIRs or both, in at least 80% of the sequences. In our analysis, we considered the alignment of the entire length of the elements to define these families. Within a given family of MITEs, elements were categorized as different subfamilies if they displayed sufficient TIR sequence homology (>80%) but distinct size. The complete elements were named according to Kapitonov and Jurka (2009) and MITE families and subfamilies were denominated with a MITE prefix. In our analysis, *pogo*-like elements were considered to belong to the *Tc1/mariner* superfamily. Sequence alignments were performed with MAFFT version 7 (KATOH; ROZEWICKI; YAMADA, 2019).

Following the identification of intact TEs, a subsequent analysis was performed by a BLASTX (www.ncbi.nlm.nih.gov/BLAST) search of approximately 5,000 bp upstream and downstream of each TE against the RefSeq_protein (Reference Sequence Protein) database aiming to identify protein-coding sequences in close proximity to the TEs. To achieve that, a cut-off criterion of an E-value of less than 1×10^{-5} and an identity of greater than 50% were adopted. Functional annotation of ORFs located near TEs was assessed with EggNOG v5.0 (HUERTA-CEPAS *et al.*, 2019). The *C. lindemuthianum* transcripts database (Silva et al. unpublished) was also used as reference. The RNA from *C. lindemuthianum* grown in GPYECH medium was sequenced with Illumina HiSeq platform. Sequences were mapped against the genome with Geneious 2023 (https://www.geneious.com).

3.2 RIP analysis

Alignments that included sequences belonging to the same family, exhibiting full coverage and an identity surpassing 80%, were considered and subsequently subjected to RIPCAL (HANE; OLIVER, 2008) for the computation of RIP index and RIP substrate depletion. The RIP index (TpA/ApT) is a straightforward metric to infer the frequency of RIP products (TpA) while compensating for false positives stemming from regions rich in ApT. TpA/ApT values above 0.89 suggest a strong RIP

response. The (CpA + TpG)/ (ApC + GpT) index assesses the reduction of RIP targets CpA and TpG. A (CpA + TpG)/ (ApC + GpT) score lower that 1.03 indicates the occurrence of RIP (HANE; OLIVER, 2008).

The RIPper online tool available at https://theripper.hawk.rocks/ (WYK *et al.*, 2019) was used to calculate genome-wide RIP index. To achieve that, default parameter was used of 1 kb sliding windows and 500 bp step size. Sequences with substrate index value > 0.75, product index value ≤ 1.1, and composite index value > 0.01 were considered to be RIP-affected (WYK *et al.*, 2019).

3.3 Phylogenetic analysis and protein characterization

To identify DIM-2 and RID homologs, DNA and protein sequences with sufficient similarity (E-value < 1×10^{-5}) to those described in *N. crassa* (GenBank accession numbers: AAM27408.1 for RID and AAK49954.1 for DIM2) and *C. higginsianum* (Genbank accession numbers: XP_018164533.1 for the RID homolog and XP_018164533.1 for DIM2 homolog) were identified using the NCBI non-redundant BLAST database.

The multiple sequence alignments were performed with MUSCLE (EDGAR, 2004). The best suitable evolutionary model was estimated with ModelFinder (KALYAANAMOORTHY *et al.*, 2017) implemented in IQ-TREE. Maximum Likehood analysis was performed with IQ-TREE multicore version 1.6.11 with bootstrap values of 1000 replicates for estimation of branch support (NGUYEN *et al.*, 2015). The phylogenetic tree was visualized and edited with the Interactive Tree Of Life (iTOL) software (LETUNIC; BORK, 2021). Taxa and GenBank accession number for the sequences used in this study are available in the appendix A. Protein domains and families were predicted using InterProScan (PAYSAN-LAFOSSE *et al.*, 2023).

4. RESULTS

The RepeatMasker analysis revealed that approximately 53.71% of the total DNA sequence of *C. lindemuthianum* is covered by repeats. Among these repeats, the putative Class II elements constitute 2.08% of the total repeat content and 1.12% of the genome. This indicates an expressive lower coverage of DNA transposons compared to putative retroelements, which account for 43.14% of the DNA and 80.32% of the total repeat content. Furthermore, a significant portion (17.59%) of the total repeat content was identified as unclassified elements. This underscores the extent of yet-to-be-explored elements, emphasizing the need for further investigation and classification.

4.1 *Tc1/mariner* and MITE elements

A total of 705 repeats potentially belonging to the *Tc1/mariner* superfamily were initially identified. However, after accounting for sequence overlap, this number was reduced to 615 elements, accounting for 0.78% of the total DNA and 1.41% of the total repeat content. Each of these elements underwent thorough scrutiny to identify putative transposase Open Reading Frames (ORFs), Terminal Inverted Repeats (TIRs), and Target Site Duplications (TSDs). Among the identified elements, 536 of copies exhibited highly interrupted TIRs and putative transposase coding regions with numerous stop codons, corresponding to the majority of the elements. As a result, these copies were considered degenerated due to significant disruptions in their structural and functional components and were no longer considered in the following analysis. The average GC content of all elements was calculated to be 24.54%, which is notably lower than the overall GC content of the *C. lindemuthianum* genome at 37.25%. In contrast, MITEs exhibited a slightly higher GC content at 29.37%.

A total of eighteen copies of complete elements were discovered, all of which belonged to the same family named *Mariner-1_CL* (Fig. 3). However, upon further analysis, all of these elements contained putative transposase ORFs significantly disrupted, with more than twenty stop codons present in each element and no

conserved domains. As a result, none of these elements were deemed potentially active, which was further confirmed by the analysis of *C. lindemuthianum* transcriptome database. The elements within this family are 1992 bp long, exhibit TAs as insertion sites, and possess Terminal Inverted Repeats (TIRs) that are 95 nucleotides in length.

To identify potential homolog proteins to the interrupted coding region belonging to this family, a BLASTX alignment was performed, which revealed a single uncharacterized protein (GenBank accession number: KJZ70102.1) from *Hirsutella minnesotensis* that showed very little coverage (24%) and low identity (32%) to the query. The protein contains a DDE superfamily endonuclease domain (IPR004875) and a Tc5 transposase DNA-binding domain (PF03221), potentially being a transposase from *Tc1/mariner* superfamily. However, the *Mariner1-CL* family was considered a new *Tc1/mariner* family, as it does not present sequence identity to any known TE family.

Figure 3 – *Tc1/mariner* and MITE families found in *C. lindemuthianum*. Target site duplication (TSD) are highlighted in red. Asterisks represent polymorphic nucleotides among the individuals.

Element	TSD AND TIR	Size (bp)	TIR size (bp)	Copies
Mariner1-CL	TACCCGTGCCCGCCCCCACTTTAATTGGAACTCGAAGACT CGTCACCCACTAAAGTGGGTGCTGAACACCCATATGTCCA TAACACCCATATGTCCA	1992	95	18
MITE1.1-CL	TACCCGTGCCCGCCCCCACTTTAATTGGAACTCGAAGACT CGTCACCCACTAAAGTGGGTGCTGAACACCCATATGTCCA	328	78	6
MITE1.2-CL	TACCCGTGCCCGCCCCCACTTTAATTGGAACTCGAAGACT CGTCACCCACTAAAGTGGGTGCTGAACACCCATATGTCCA TA	344	80	3
MITE1.3-CL	TACCCGTGCCCGCCCCCACTTTAATTGGAACTCGAAGACT CGTCACCCACTAAAGTGGGTGCTGAACACCCATATGTCCA TAACACCCATATGTCCA	914	95	2
MITE1.4-CL	TACCCGTGCCCGCCCCCACTTTAATTGGAACTCGAAGACT CGTCACCCAC	103	48	1
MITE1.5-CL	TACCCGTGCCCGCCCCCACTTTA	73	21	2
MITE1.6-CL	TACCCGTGCCCGCCCCCACTTTA	63	21	1
MITE2_CL	TAACGTGGTTGATAAGCGAGCTGCCCAGACAGGCGAGCT GCCTAGGCCTTATCACACTTTAAA	435	61	24
MITE3_CL	TAACGTGGTTGATAAGCGAGCTGCCCAGACAAGCGAGCT GCCCAGGCCTTATCAC	435	53	22

Source: the author.

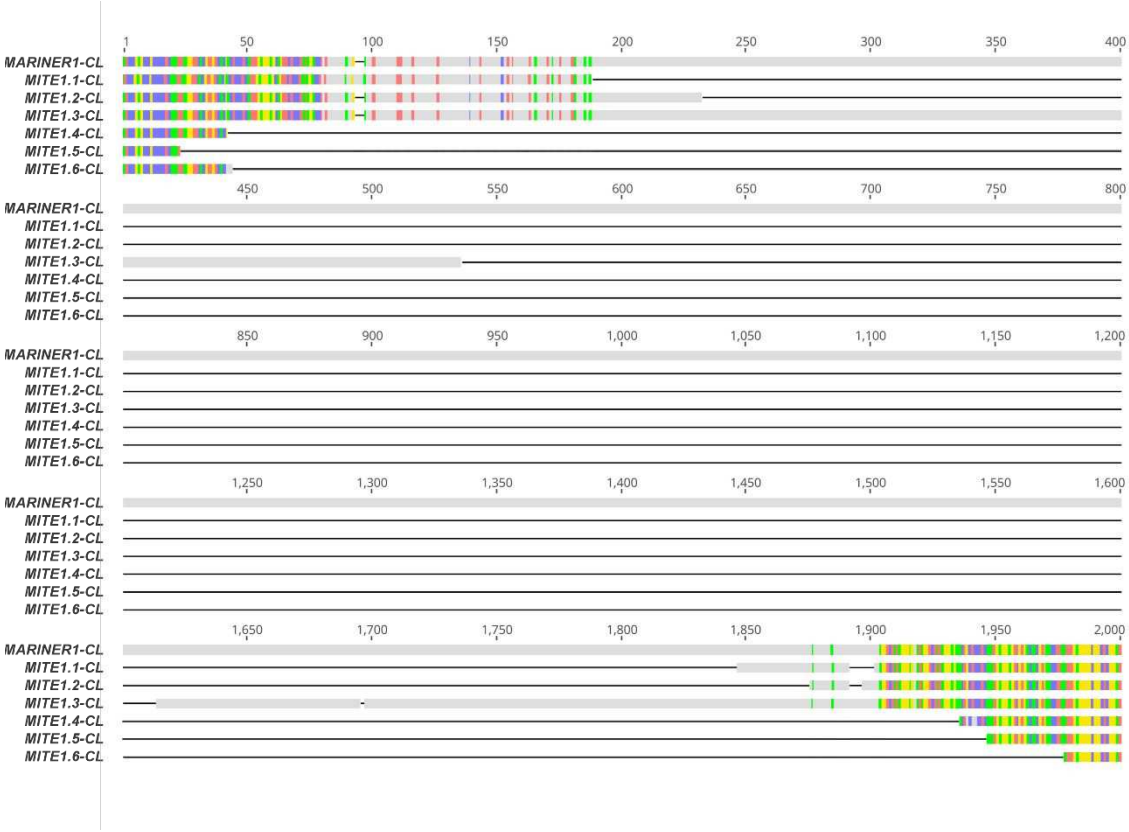
It was also performed a BLASTn alignment using the two most contiguous genomes of the *C. orbiculare* species complex, *C. orbiculare* and *C. trifolii* (GAN et al., 2019) (GenBank accession numbers: AMCV02000007.1 and GCA_004367215.1, respectively). The BLASTn search of *Mariner1-CL* against the *C. orbiculare* genome resulted in 12 hits with a query cover of over 50%. Among them, four hits exhibited a coverage above 97% and near intact TIRs, and therefore were considered complete. The identity of these hits varied from 82% to 85.3%. However, both Expasy and WebAugustus searches failed to identify any intact putative transposase ORFs, indicating that these hits are also non-autonomous. Regarding the *C. trifolii* genome, the BLASTn search of *Mariner1-CL* yielded 9 hits with a query cover above 50%. Four of these hits had complete coverage (100%) and were deemed complete. The identity among these 9 elements ranged from 80.1% to 82.6%. However, like in the copies from the *C. orbiculare* genome, all of them represented interrupted ORFs, indicating that they are not active. The *Mariner1-CL* family is not limited to *C. lindemuthianum* and seems to be widespread among the *C. orbiculare* species complex analysed.

Additionally, sixty copies of MITEs were identified and divided into three distinct families. The first family, named *MITE1-CL*, consists of fifteen copies derived from the *Mariner1-CL* family with different size and coverage (Fig. 4). As a result, they were all categorized as six different subfamilies within the same family (Fig. 3). They also exhibited TAs as target site duplication and present TIRs of variable length. The six subfamilies were established based on sequence similarity (>80%) and TIR length (Fig.3).

The first subfamily, *MITE1.1-CL* (Appendix B) accounts six copies of 328 bp with TIRs of 78 bp of length. The subfamily *MITE1.2-CL* (Appendix C) covers three copies of 344 bp, with 80 bp TIRs. *MITE1.3* harbors two copies of 914 bp, and the longest TIRs among the family with 95 bp. *MITE1.4* has a single copy in the genome, spanning 103 bp and TIRs with 48 bp. The subfamily *MITE1.5* is composed of two copies 73 bp and 21 bp TIRs. Lastly, the *MITE1.6-CL* subfamily has a single copy, with a total length of 63 bp and 21 bp TIRs.

We performed BLASTn alignment against the *C. trifolii* and *C. orbiculare* genomes with the elements with the highest GC content of each *MITE1* subfamilies. The results deemed complete (query cover >97%) together with the *MITE1* subfamilies representative sequences were aligned and redundant hits were excluded. Our search for elements belonging to the *MITE1* family yielded four hits from *C. trifolii* genome, ranging from 71 bp to 380 bp, and four hits within *C. orbiculare* genome, ranging from 69 bp to 333 bp. Even though they presented conserved TIRs and TAs as insertion site, size and internal content differences prevented them to be classified into any of the described subfamilies described for *C. lindemuthianum.* Interestingly, all except one of the hits were placed near coding regions (< 2000 bp from an ORF). One hit in *C. orbiculare* genome (Scaffold 6; 2.513.324 – 2.513.399) was inserted into a predicted ORF coding for an uncharacterized protein (Cob_v003704). Though a homologous sequence is present in *C. lindemuthianum* genome, no TE insertion was found within the ORF.

Figure 4 – Multiple alignment of consensus sequences of *MARINER1-CL* and all derivatives MITE subfamilies. Colors represent conserved nucleotides (A = pink; C = purple; G = yellow; and T = green), while unconserved sites are shown in grey. Lines represent alignment gaps.



Source: the author

It was not possible to identify the elements which originated the remaining MITE groups and they were separated into two separate families based on sequence identity (Fig. 4). The first family, *MITE2-CL*, consists of twenty-four copies, ranging from 435 bp in length. These copies possess TIRs that are sixty-one bp long and exhibit direct repeats of TAs. The second family, MITE3, comprises twenty-two copies, also measuring 435 bp in length. These copies contain direct TA repeats and possess TIRs that are fifty-three nucleotides long. The BLASTn search of the elements with the highest GC content of both families against the genomes of *C. orbiculare* and *C. trifolii* yielded no hits.

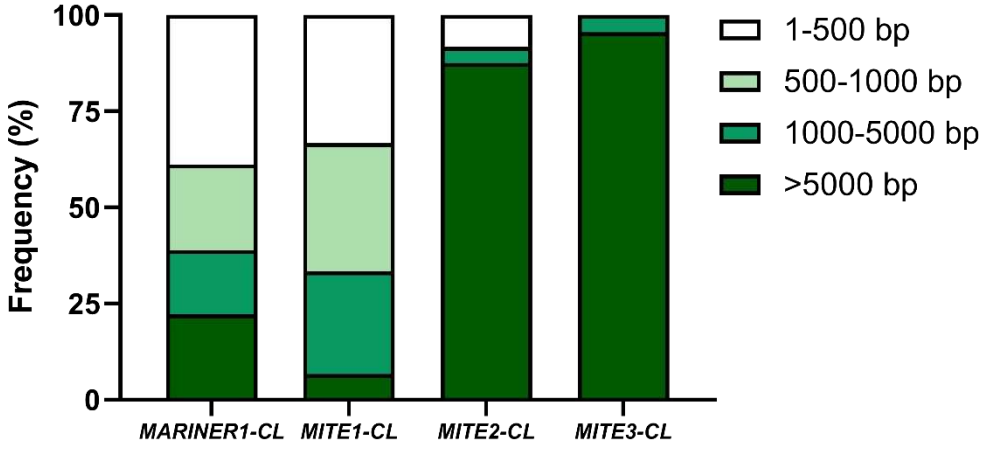
4.2 Transposable elements integration sites

In order to identify protein coding sequences near TEs, a BLASTX alignment was carried out in regions upstream and downstream of each element (Appendix D). Out of the eighteen intact *Mariner1-CL* elements, fourteen were found to be in close proximity to predicted protein coding sequences, which accounts for approximately

78% of the elements. Eleven of these elements were located within a distance of 1000 bp from the nearest ORF. Notably, ten of these elements were situated within putative promoter regions. We didn't find these elements to be enriched nearby a particular a gene group. Nonetheless, notable among the frequently observed protein functions are three proteins linked to translation, ribosomal organization, and biogenesis, as well as two proteins associated with carbohydrate transportation and metabolism (data not shown). Additionally, three proteins play a role in the regulation of transcription.

Similar to the elements from which they originated, the elements belonging to the MITE.1 family demonstrated a propensity to be placed near coding regions, with fourteen out of fifteen elements being in close proximity to predicted genes, accounting for 93% of the sequences (Fig. 5). Notably, eleven of these elements were inserted within gene promoter regions. Once again, genes encoding proteins with predominant functions near the TEs were not identified. However, among the frequently observed protein functions, amino acid transport and metabolism stood out with four proteins (data not shown).

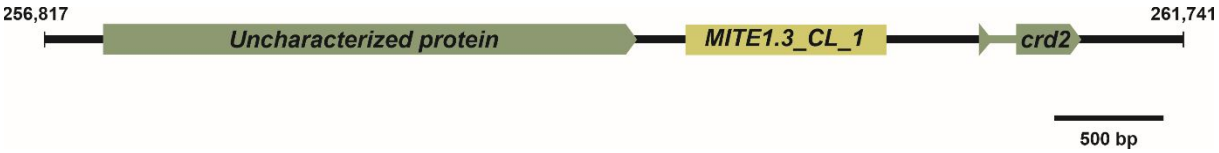
Figure 5 – Frequency of TE insertions into the nearest *C. lindemuthianum* ORF based on the distance range. The frequency is determined by the relative portion of TEs located within specific distance ranges (1–500, 500–1000, >1000 bp) in comparison the total number of TEs.



Source: the author.

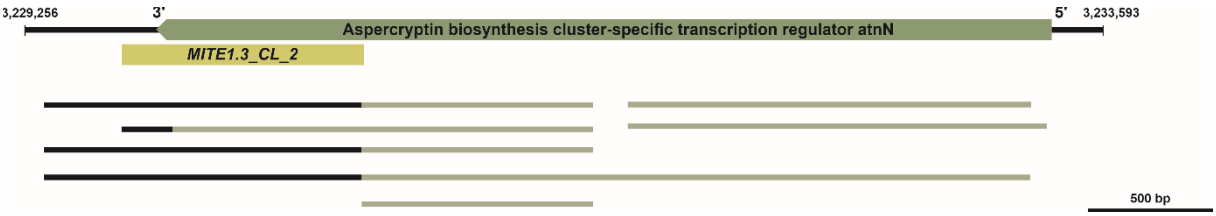
Interestingly, an element belonging to the *MITE1.3-CL* subfamily, named *MITE1.3-CL_1*, was found to be inserted 346 bp upstream of a predicted ORF (Fig. 4). This ORF encodes a 111-amino acid protein that is a homolog to a copper resistance crd2 metallothionein protein (GenBank accession number: XP_018159318; Fig. 6). A homologous gene search was performed in *C. orbiculare* and *C. trifolii* genome (Cob_v010221 and CTRI78_V000038), but no MITE was found within the promoter regions. Furthermore, 750 bp of the element *MITE1.3-CL_2* overlaid the predicted 3' of the gene coding for an Aspercryptin biosynthesis cluster-specific transcription regulator atnN (scaffold 5: 3.229.574 – 3.230.491), which regulates the production of the lipoprotein aspercryptin (HENKE *et al.*, 2016). RNA-seq data from *C. lindemuthianum* shown that this gene produced alternative transcripts. Interestingly, one of the transcripts starts 2048 bp into the gene, has 1699 bp and covers 769 bp from the element (Fig. 7). It shows no hits with any protein, and is likely a new transcript created due to the element insertion. However, 160 bp of the 3' end of the transcript represents mismatches, suggesting a sequencing or assembling error that possibly created a chimeric transcript.

Figure 6 – The *MITE1.3-CL_1* element is inserted within the putative promoter region of the gene responsible for encoding the Copper resistance protein crd2.



Source: the author.

Figure 7 – Organization of the Aspercryptin biosynthesis cluster-specific transcription regulator atnN gene and its alternative transcripts. 750 bp of the element *MITE1.3-CL_2* overlaid the predicted 3' of the gene. Black marks on transcripts represent sequence mismatches, likely representing chimeric transcripts.



Source: the author.

Contrary to the *Mariner1-Cl* and *MITE1-CL* families, the TEs belonging to both the *MITE2-CL* and *MITE3-CL* families were found much less frequently in close proximity to coding sequences. Among the twenty-four copies of *MITE2-CL*, only three were found near open reading frames (ORFs), accounting for 12% of the elements (Fig. 5). Similarly, out of the twenty-two copies of *MITE3-CL*, only one was located adjacent to predicted genes, accounting for approximately 4% of the copies. It is likely that most of the remaining copies of these two families are situated in regions with a high density of transposable elements, as indicated by BLASTx hits. Although the BLASTx alignment of the upstream and downstream regions of the elements revealed hits with reverse transcriptases, no intact ORF was found, as they were interrupted by multiple stop codons, similar to the *Tc1/mariner* elements that were studied.

4.3 RIP analysis

Genome-wide analysis using the RIPper online tool estimated that 54.77% of the total DNA content of *C. lindemuthianum* is affected by RIP. To analyze to which extent RIP defense affected the eighteen elements belonging to the *Mariner1-CL* family, RIPCAL analysis was conducted using *Mariner1-CL_9* as the reference sequence, chosen for its highest GC content among all the elements. Consistent with RIP signatures, transitions were found to be more frequent than transversions, with the most common mutations being C to T and G to A. The most frequent dinucleotide target was CpA. The calculated RIP index (TpA/ApT) was 1.29, indicating RIP activity. Additionally, the RIP substrate depletion index (CpA + TpG/ ApC + GpT) was 0.85, further supporting the presence of active RIP. It is important to note that despite using the highest GC-containing element as a reference, our analysis did not identify any potentially active transposable elements. Moreover, the reference element itself

likely harbors numerous mutations. Therefore, it is possible that these values are underestimated.

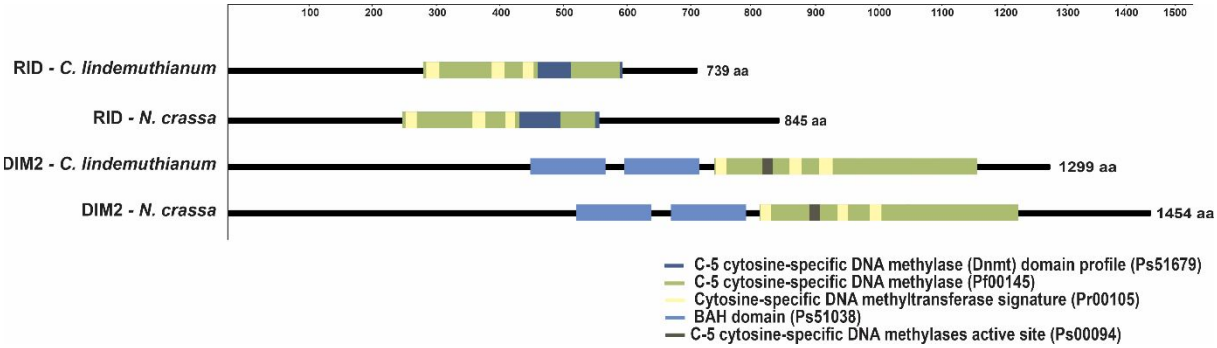
RIP analysis was also conducted with *MITE1-CL* representatives. For that, we submitted to RIPCAL an alignment with the six sequences from *MITE1.1-CL* subfamily, as it is the subfamily with the highest number of copies. The RIP index (TpA/ApT) was 1.11 and the substrate depletion index (CpA + TpG/ ApC + GpT) was 0.92, indication a weaker RIP pressure when comparing to the *Mariner1-CL* family. C to T and G to A transitions were the most common mutations, with CpA as the most frequent dinucleotide target.

Finally, using *MITE2-CL_4* and *MITE3-CL_6* as reference sequences with the highest GC content, RIPCAL analysis of the twenty-four members of the *MITE2-CL* family and the twenty-two members of *MITE3-CL* also indicated an active RIP. The analysis revealed RIP index (TpA/ApT) of 2 and 1.56 and substrate depletion index (CpA + TpG/ ApC + GpT) of 0.62 and 0.52, respectively. Transitions were more abundant than transversions in both families, with C to T and G to A mutations being more frequent. Additionally, CpA dinucleotide was the preferred target for mutation.

4.4 RID and DIM2 structural and phylogenetic analysis

Using BLASTn search associated with structural and phylogenetic analysis, we were able to identify putative homologues of the two genes coding for methyltransferases involved in RIP in other fungi in the genome of *C. lindemuthianum.* The first one, namely the *ClRID* gene (Clind89_008448-T1), codes for a protein with fairly similar organization to *N. crassa* RID protein, as all domains found in *N. crassa* are also present in the protein from *C. lindemuthianum*, characterizing a C-5 cytosine-specific DNA methyltransferase (Fig. 8). ClRID has 783 amino acids, slightly smaller than the *N. crassa* RID protein which has 845 amino acids, with distinctive N- and C-terminal parts. The two proteins show an identity of 53%, with most of the similarities contained within the C-5 cytosine-specific DNA methylase (Dnmt) domain (Fig. 9).

Figure 8 – Comparison of the structural domains present in *N. crassa* RID and DIM2, and their corresponding homologs in *C. lindemuthianum*.



Source: the author.

The second methyltransferase gene identified, named *ClDIM2* (Clind89_008059-T1), codes for a protein with 1299 amino acids, slightly smaller than to *N. crassa* DIM2, which spans 1454 amino acids. Protein domain annotation revealed that the organization of *N. crassa* DIM2 and its corresponding homolog in *C. lindemuthianum* is relatively similar, with both proteins bearing two BAH domains and a common C-5 cytosine-specific DNA methylase domains (Fig. 10). ClDIM2 has 35% identity to *N. crassa* DIM2, with most of the conserved regions placed within these domains (Fig. 8). Furthermore, both proteins count on a C-5 cytosine-specific DNA methylase active site, which is a conserved regions that contains a dipeptide Proline-Cysteine. The cysteine residue is involved with a covalent bond formation with DNA intermediate to methyl catalytic transfer (CHEN *et al.*, 1991).

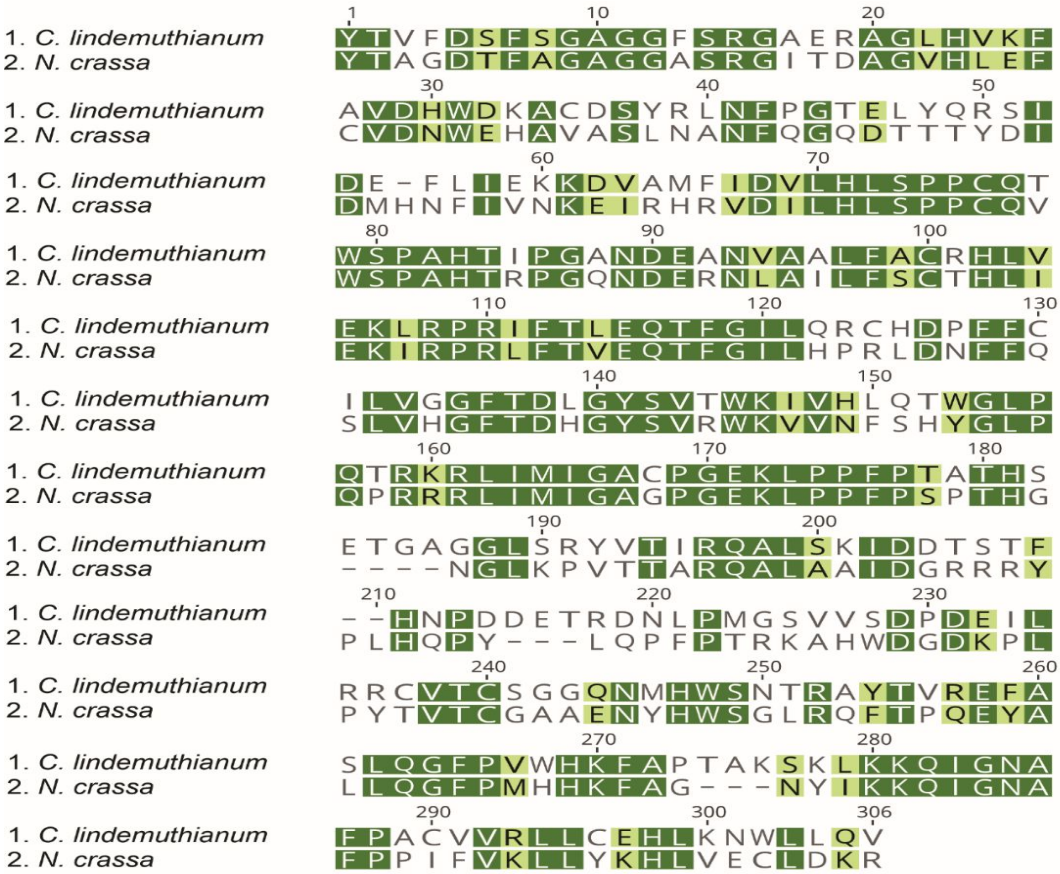
To provide additional confirmation of the placement of ClRID and ClDIM2, phylogenetic analysis was conducted (Fig. 11). This analysis encompassed RID/Masc1 and DIM2 sequences from fifty-two species, including homologs from thirty-three *Colletotrichum* species. Additionally, eleven Dnmt1/Masc2 protein sequences were included to ensure separation, and the sequence of *Schizosaccharomyces pombe* Dnmt2 was used as an outgroup to root the tree. Furthermore, the domain composition of each protein was incorporated into the analysis.

The resulting phylogenetic tree displayed a clear separation of three clades. The first clade consisted of proteins associated with *N. crassa* RID and *Ascobolus*

imersus Masc1. The second clade, in turn, was further subdivided into two subgroups. The first subgroup contained proteins related to *N. crassa* DIM2. The second subgroup encompassed *Ascobolus imersus* Masc2 and the remaining Dnmt1 proteins, which can be distinguished by the possession of the replication foci domain (RFD). The third group accounts for the Dnmt1/Masc2 proteins.

Our tree topology supports the existence of RID and DIM2 homologs in all *Colletotrichum* species analyzed. The structure of RID exhibited overall homogeneity among the different species, except for RID in *Trichoderma harzianum, Tricoderma simmondsii*, and *Trichoderma gamsii*. These three species displayed a noticeably shorter sequence consisting of only 319 amino acid residues. Despite the reduced length, these sequences retained the complete methyltransferase domain. In contrast, the DIM2 protein exhibited greater variation in its domain composition, with each sequence displaying either one or two BAH domains. Interestingly, this variation was observed even within the genus *Colletotrichum*, indicating that the presence of one or two BAH domains is not necessarily correlated with the phylogenetic relationship of the species. Notably, the DIM2 homolog sequence from *C. orbiculare* is much shorter than the remaining, with only 505 residues, and lacks any BAH domains. However, whether *C. orbiculare* DIM2 retains the capability of *de novo* methylation remains to be elucidated.

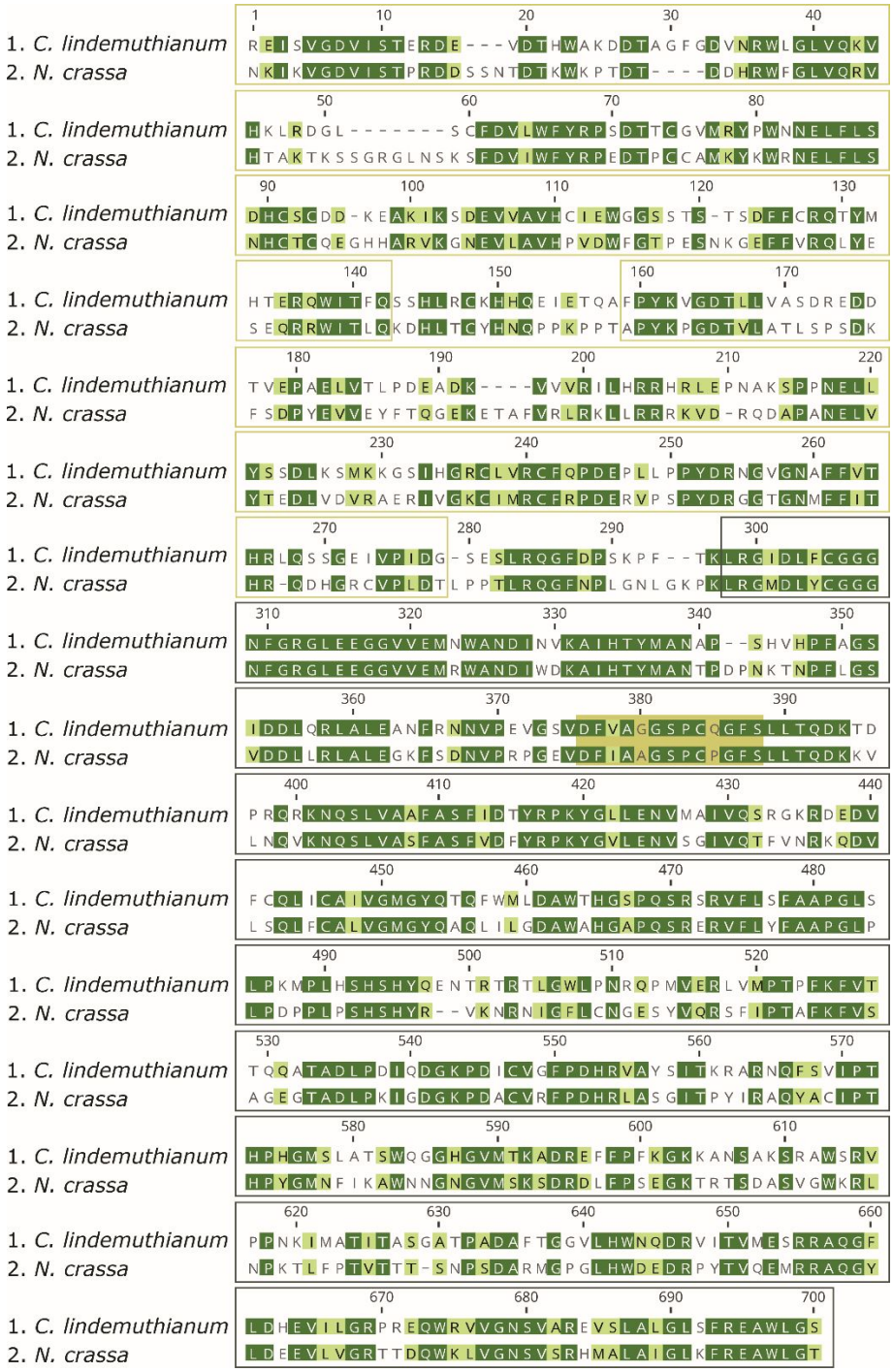
Figure 9 – Comparison of C-5 cytosine-specific DNA methylase (Dnmt) domain between *N. crassa* RID (GenBank accession number: XP_011392925.1) and its putative homolog in *C. lindemuthianum*. Dark green highlights indicate identical amino acids, while light green highlights represent conserved substitutions.



Source: the author.

Figure 10 – Comparison of DMT and BAH domains between *N. crassa* DIM2 (GenBank accession number: AAK49954.1) and its homolog in *C. lindemuthianum*. Dark green highlights indicate identical amino acids, while light green highlights represent conserved substitutions. Hollow yellow boxes denote the BAH domains,

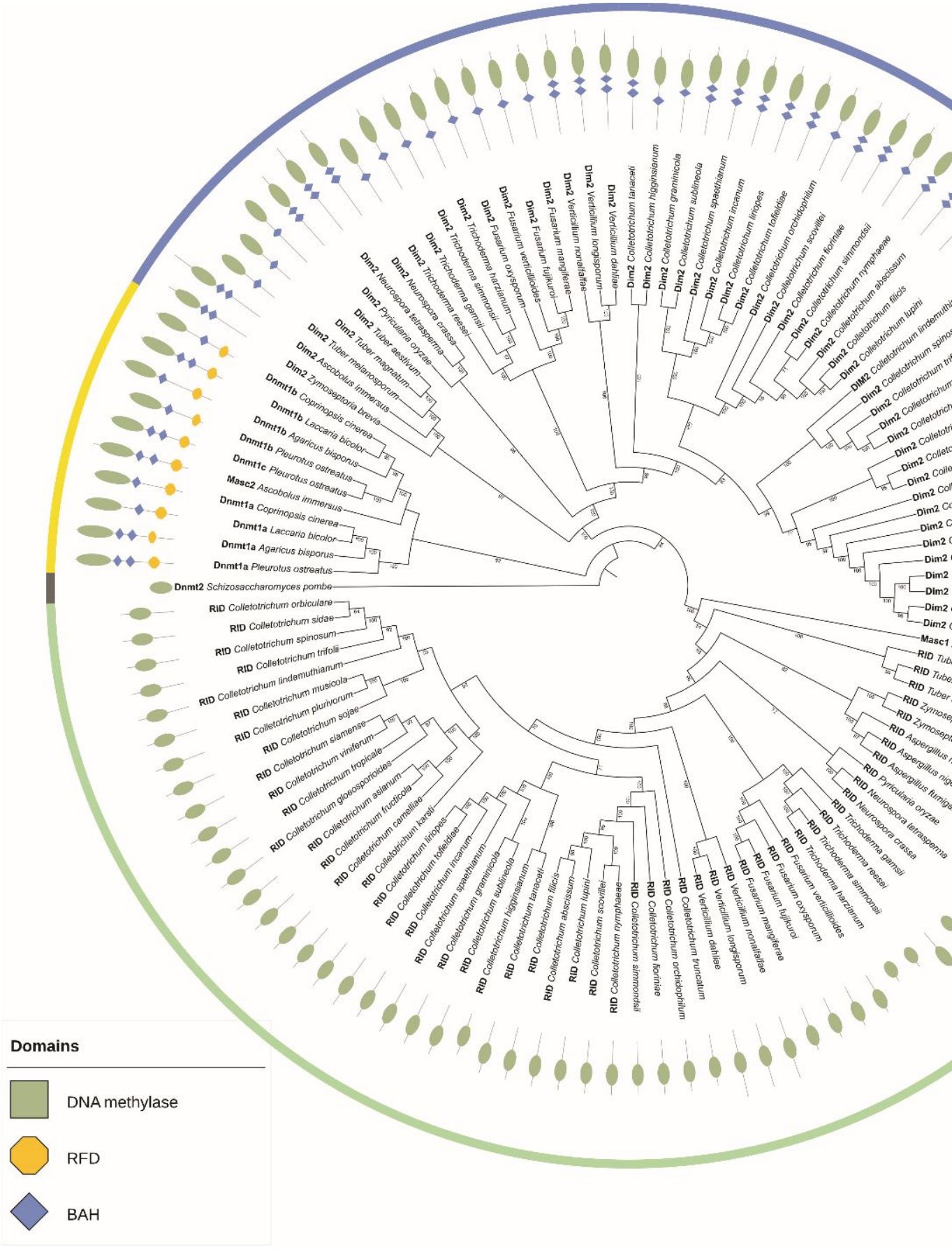
and the DMT domain is marked with a hollow grey box. The C-5 cytosine-specific DNA methylase active site is denotated as a filled yellow box.



Source: the author

Figure 11 – Phylogenetic relationships based on Maximum Likelihood and Domain Architecture Analysis of 118 fungal DNA methyltransferase proteins. Multiple sequence alignments were performed with MUSCLE and the tree was constructed

with IQ-TREE with bootstrap values of 1000 replicates. Bootstrap Support Values above 50 are displayed above the branches. DIM2 proteins are indicated by a blue stripe, RID/Masc1 proteins by a green stripe, and Dnmt1/Masc2 proteins by a yellow stripe. The sequence of *S. pombe* Dnmt2 was utilized as an outgroup.



Source: the author

5. DISCUSSION

The total repeat content in *C. lindemuthianum*, at 53.17%, is the highest among all *Colletotrichum* genome sequenced. However, it is comparable to *C. orbiculare*, which has a content of 44.88% (RAO *et al.*, 2018), which is still considerably smaller than the genome of *C. lindemuthianum*. Notably, *C. orbiculare* has the highest TE content described for *Colletotrichum* species other than *C. lindemuthianum*. This similarity in repeat content suggests a potential trend within the *C. orbiculare* species complex. Other species in the complex, such as *C. trifolii* (109 Mb), *C. spinosum* (82.73 Mb), and *C. sidae* (85.83 Mb) (GAN *et al.*, 2019) have some of the largest genomes available for the genus, which may also be associated with significant TE content, as Chen et al. (2022) found a clear positive correlation between genome size expansion and total coverage of repeat content within the *Colletotrichum* genus. However, the mobilome of these species has not yet been characterized, and further investigation is needed to confirm that.

Although the *Tc1/mariner* superfamily of Class II elements constitutes 70% of all DNA transposons in *C. lindemuthianum*, their presence accounts for only 1.41% of the total repeat content, significantly lower compared to other TE superfamilies in Class I TEs. In contrast, *C. orbiculare* has 0.02% of total Class II elements and negligible *Tc1/mariner* content (RAO *et al.*, 2018). However, the *Tc1/mariner* content in *C. lindemuthianum* is similar to that of *C. truncatum* (0.82%) and *C. chlorophyti* (2.62%). This finding diverges from the pattern observed in other *Colletotrichum* species, where *Tc1/mariner* element coverage ranges from 8.87% in *C. orchidophilum* to 27.12% in *C. higginsianum* (RAO *et al.*, 2018).

The great number of degenerated copies among the *Tc1/mariner* elements in *C. lindemuthianum*, with highly interrupted TIRs and putative transposase coding regions with numerous stop codons, suggests that they are likely relics of old TE bursts followed by strong host defense responses and TE degeneration (BOURGEOIS; BOISSINOT, 2019). Among the complete copies found, all of them belong to same family, named *Mariner1-CL*, none of them had intact transposase ORFs and were deemed inactive. Braga et al. (2014), while investigating *Tc1/mariner* elements in *C. graminicola*, also solely found non-autonomous TEs, with all putative transposases ORFs punctuated by mutations with several stop codons and strong

RIP signature. However, it was possible to identify conserved DDE motifs typical of transposases in many copies (BRAGA *et al.*, 2014), which was not possible in any of the copies of *Mariner1-CL* due to the high number of mutations. It is important to note that even inactive TEs can trigger genomic changes, as repeated copies can promote recombination events and result in large scale deletions, duplications, and inversions (BOURQUE *et al.*, 2018).

BLAST searches confirms that *Mariner1-CL* is a previously undescribed TE family, although it is not exclusive to *C. lindemuthianum.* Complete copies of *Mariner1-CL* are present in the genomes of two other species from the *C. orbiculare* species complex, *C. trifolii* and *C. orbiculare*. This indicates that this element first invasion predates the speciation events that separated these species. However, none of the copies found coded for functional transposases, making the *Mariner1-CL* elements non-autonomous in all three species analyzed. Nevertheless, copy number and TE insertion sites vary among the species, suggesting that this element was still active at the time the species diverged.

A non-autonomous MITE family derived from the *Mariner1-CL* elements found in the *C. lindemuthianum* genome was named *MITE1-CL*. These copies exhibited variations in TIR length and identity, and were categorized into six different subfamilies based on these characteristics. Similar to the *Mariner1-CL* family, *MITE1-CL* copies are present in the genomes of both *C. orbiculare* and *C. trifolli.* Although the TIRs identity and TSD were preserved among the copies, the size and internal content differences rendered them incompatible with the described subfamilies for *C. lindemuthianum*. Consequently, it can be hypothesized that the deletion events that originated the different elements within *MITE1-CL* family happened in a common ancestor and underwent subsequent modifications in each of the species, or, alternatively, occurred independently in each species.

Interestingly, both *Mariner1-CL* and *MITE1-CL* families shared a strong tendency to locate near predicted genes, as 78% of *Mariner1-CL* and 93% *MITE1-CL* copies were found in close proximity to ORFs, which could indicate a possible role in transcription regulation (FESCHOTTE; PRITHAM, 2007). This trend was also observed in the copies from the other two species within the *C. orbiculare* complex. Numerous Class II TEs exhibit a clear preference for insertion into or near coding

sequences (FESCHOTTE; PRITHAM, 2007). For instance, P elements from *Drosophila* show a preference to insert into ORFs (SPRADLING *et al.*, 1995) and *Mutator* elements from maize crops preferentially integrate into the 5′ end of coding regions (JIANG *et al.*, 2011). Furthermore, MITE elements have also shown a tendency to inserting adjacent to coding regions (FATTASH *et al.*, 2013; OKI *et al.*, 2008). For instance, in *Botrytis cinerea*, 60% of *Scatter* MITE copies were found within 500 bp upstream or downstream to ORFs (DENG *et al.*, 2013). Similarly, in rice plants, the *mPing* MITE demonstrated a proximity to gene coding regions in 70% of insertions. Interestingly, in this particular case, the TE placement tendency is a result of targeting bias and not the result of selective pressures (NAITO *et al.*, 2006). The placement tendency observed in both *Marine1-CL* and *MITE1-CL* is a shared feature of the parental complete element and the derived non-autonomous family. It is uncertain whether there is a selective force favoring the maintenance of these elements near coding sequences for potential adaptative advantages or if the transposition machinery showed a bias towards integration into common motifs found in the vicinity of ORFs.

One of the copies belonging to the *MITE1.3-CL* subfamily is placed within the promoter sequence of a *crd2* homolog. In *Candida albicans*, where it was first described, the *cdr2* (copper resistance determinant 2) gene codes for a metallothionein protein, which is a small cysteine-rich protein that acts as a copper chelator. Knockout *C. albicans* strains with disrupted *crd2* genes present reduced growth rate in higher copper concentrations (RIGGLE; KUMAMOTO, 2000). In fact, copper is an important trace element relevant in fungal growth, development, and pathogenicity. However, in high concentrations, this cation is harmful to cells. Copper metallothioneins are capable of sequestrating this cation and avoid intracellular toxicity (GARCÍA-SANTAMARINA; THIELE, 2015). Copper homeostasis is so important for plant pathogens that copper-based chemicals are effectively used as a control strategy in crops (LA TORRE; IOVINO; CARADONIA, 2018). As MITEs have the potential to interfere with nearby gene expression, acting as promoters, enhancers, or silencers (CHEN *et al.*, 2017; KUANG *et al.*, 2009; SUN *et al.*, 2013a), it is possible that *MITE1.3-CL* insertion upstream the *crd2* gene is somehow influencing the gene transcription and affecting *C. lindemuthianum* copper tolerance. However, further investigation is necessary to confirm this hypothesis.

Furthermore, 750 bp of the element *MITE1.3-CL_2* overlaid the predicted 3' of the gene coding for an Aspercryptin biosynthesis cluster-specific transcription regulator atnN. In *Aspergillus nidulans*, AtnN is a transcription factor that regulates the biosynthetic gene cluster culminating in the production of the lipoprotein aspercryptin (HENKE *et al.*, 2016). Notably, among the alternative transcripts produced by this gene, one seems to cover 769 bp from the MITE. However, the 3' end of the transcript represents mismatches with the genomic sequence, suggesting a sequencing or assembling error possibly created a chimeric transcript. This could be confirmed employing RNA-seq approach.

Opposing to the *Mariner1-CL* and *MITE1-CL* families, *MITE2-CL* and *MITE3-CL* do not share the tendency of being in close proximity to coding sequences, as only 12% and 4% of the copies of each family, respectively, occupy such position. The parental copies for neither of the families was identified. It is possible that the full-length elements that originated these families have been eliminated or never became fixed in the population (FESCHOTTE; PRITHAM, 2007). Furthermore, no similar copies were found in *C. trifolii* and *C. orbiculare* genomes, suggesting that this families may be unique to *C. lindemuthianum*.

Genome-wide analysis revealed that 54.77% of the total DNA content of *C. lindemuthianum* is affected by RIP. Among the genomes previously analyzed using the RIPper tool, *Pyrenophora teres* has the highest RIP coverage reported at 29.53% (VAN WYK et al., 2021). This makes *C. lindemuthianum* the species with the strongest RIP signature among all fungi analyzed with the same tool. Other *Colletotrichum* species also exhibit a large part of their genome affected by RIP, such as *C. graminicola* and *C. lupini,* with 27.41% and 24.08% of their genomes RIP-affected. This seems to be positively correlated to repeat elements coverage, which is 25.88% in *C. graminicola* and 21.96% for *C. lupini* (BECERRA *et al.*, 2023). In turn *Colletotrichum* spp. with lower repeat coverage such as *C. fructicola* with 1.48% and *C. higginsianum* with 7.24%, show a much weaker RIP signature with 3.62% and 7.24%, respectively (BECERRA *et al.*, 2023). Given the much higher repeat coverage in *C. lindemuthianum*, it is not surprising that the RIP percentage is significantly higher as well. According to Muszewska et al. (2017), there is a positive correlation between defense mechanisms and TEs abundance. They argue that organisms with defense against TE proliferation possess a larger amount of inactive

TEs, which can be recruited for various functions without the risks associated with active TEs' excision and insertion (MUSZEWSKA et al., 2017). A highly effective RIP mechanism can impair gene duplications, which is an important form of evolution (ARAMAYO; SELKER, 2013). However, RIP can also be a source of genetic variation, as RIP might escape into sequences neighboring repetitive elements (FUDAL *et al.*, 2009). Indeed, De Queiroz et al. (2019) reported that most of nucleotide variations within effector genes between two different strains of *C. lindemuthianum* consisted in C to T and G to A transitions, indicating that RIP might be a source of variability in this pathogen.

The genome-wide RIP analysis that revealed a strong RIP signature is in line with our findings that there are no conserved *Tc1/mariner* transposases ORFs in *C. lindemuthianum*, probably due to RIP mutations. To explore this even further, we analyzed RIP signatures for each described family individually. Our analysis confirmed that most common mutations in all families are C to T and G to A transitions, consistent with RIP activity. The CpA nucleotide bias observed explain the high frequency of stop codons found within the transposase coding regions, as CpA to TpA increases the probability of introducing TAA and TAG within the sequence (MONTIEL; LEE; ARCHER, 2006).

The RIP activity is further supported by the calculated RIP index and RIP substrate depletion which are consistent with RIP. *Mariner1-CL* elements seem to have suffered a stronger by RIP pressure than the derived MITE family, evidenced by a higher RIP index and lower substrate depletion index in the complete elements. This could indicate that the truncated copies might have been affected milder by RIP due to their compact size (PEREIRA *et al.*, 2021).

Surprisingly, the MITE families *MITE2-CL* and *MITE3-CL* have shown the strongest RIP pressure among the four families analyzed. This could be related to their genomic placement, as most copies of both families are nested within TE-rich regions, which is not true for *Mariner1-CL* and *MITE1-CL* families. Furthermore, considering that all *Mariner1-CL* presented highly interrupted ORFs and high AT content and RIPCAL analysis are based on the copy harboring the highest GC content, it is likely that have RIP was underestimated in this family.

Homologues of the two genes coding for methyltransferases involved in RIP, RID and DIM2, were identified. Both proteins show a similar domain organization to the *N. crassa* methyltransferases, suggesting they may have conserved function. Additionally, a phylogenetic analysis covering RID and DIM2 sequences from fifty-two different species was performed, incorporating the overall domain composition of each protein. The resulting tree confirms the presence of RID and DIM2 homologs in all *Colletotrichum* spp. included. RID proteins shown domain composition similarity among the different species. Curiously, three *Trichoderma* species, *T. harzianum, T. simmondsii*, and *T. gamsii*, presented a considerably shorter sequence compared to *Trichoderma reseei* RID homolog, but retained the complete methyltransferase domain. RIP has been experimentally confirmed in *Trichoderma reseei*, as well as RID relevance in this species sexual development (LI *et al.*, 2017; LI; CHEN; WANG, 2018). However, further investigation is needed to understand if the truncated proteins belonging to the three species plays the same role than in *T. reseei*.

Conversely, DIM2 proteins displayed greater variation in its domain composition, with each sequence harboring either one or two BAH domains. This variation was observed regardless of species placement in the tree, which suggests that the presence of one or two BAH domains is not necessarily correlated with the phylogenetic relationship of the species, and domain lost/gain might have happened multiple times over the course of evolution. Strikingly, the putative DIM2 from *C. orbiculare* appears to be truncated, as it is much smaller than DIM2 from other *Colleotrichum* species and lacks BAH domains. Non-functional copies of DIM2 homologues have been reported previously for a *Z. tritici* strain and resulted in almost complete loss of cytosine methylation capacity, although the species still retained the ability to developed normally (DHILLON *et al.*, 2010; MÖLLER *et al.*, 2021). However, whether *C. orbiculare* DIM2 retains the capability of *de novo* methylation remains to be elucidated.

6. CONCLUSIONS

This study sheds light on the *Tc1/mariner* and MITE landscape of *C. lindemuthianum*, a pathogenic fungus responsible for anthracnose in common beans. The presence of degenerated elements and the absence of active *Tc1/mariner* elements in *C. lindemuthianum* suggests a strong defense mechanism against repetitive sequences, as indicated by interrupted putative transposase ORFs by multiple stop codons. The finding of RIP signatures and the presence of methyltransferases associated with RIP support the active role of this genome defense mechanism in *C. lindemuthianum*. The discovery of transposon insertions within gene promoter regions highlights the potential impact of transposable elements on the regulation of gene expression in *C. lindemuthianum*. The insights gained from this study contribute to a better understanding of the repetitive landscape and defense mechanisms against TE proliferation of *C. lindemuthianum*, providing a foundation for further investigations into the genetic variability and TE exaptation of this economically important plant pathogen.

REFERENCES

ALKAN, C.; SAJJADIAN, S.; EICHLER, E. E. Limitations of next-generation genome sequence assembly. **Nature Methods**, v. 8, n. 1, p. 61–65, 2011.

ALMOJIL, D. *et al.* The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes. **Genes**, v. 12, n. 6, p. 918, 2021.

AMSELEM, J.; LEBRUN, M.-H.; QUESNEVILLE, H. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. **BMC Genomics**, v. 16, n. 1, p. 141, 2015.

ARAMAYO, R.; SELKER, E. U. *Neurospora crassa*, a Model System for Epigenetics Research. **Cold Spring Harbor Perspectives in Biology**, v. 5, n. 10, p. a017921, 2013.

AZIZ, R. K.; BREITBART, M.; EDWARDS, R. A. Transposases are the most abundant, most ubiquitous genes in nature. **Nucleic Acids Research**, v. 38, n. 13, p. 4207–4217, 2010.

BADET, T. *et al.* Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. **Nature Communications**, v. 12, n. 1, p. 3551, 2021.

BECERRA, S. *et al.* Chromosome-level analysis of the *Colletotrichum graminicola* genome reveals the unique characteristics of core and minichromosomes. **Frontiers in Microbiology**, v. 14, n. 760, 2023.

BHADAURIA, V. *et al.* Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen *Colletotrichum lentis*. **New Phytologist**, v. 221, n. 1, p. 431–445, 2019.

BOURGEOIS, Y.; BOISSINOT, S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. **Genes**, v. 10, n. 6, p. 419, 2019.

BOURQUE, G. *et al.* Ten things you should know about transposable elements. **Genome Biology**, v. 19, n. 1, p. 199, 2018.

BOUVET, G. F. *et al.* Stress-induced mobility of OPHIO1 and OPHIO2, DNA transposons of the Dutch elm disease fungi. **Fungal Genetics and Biology**, v. 45, n. 4, p. 565–578, 2008.

BRAGA, R. M. *et al.* Transposable elements belonging to the Tc1-Mariner superfamily are heavily mutated in *Colletotrichum graminicola*. **Mycologia**, v. 106, n. 4, p. 629–641, 2014.

CANNON, P. F. *et al. Colletotrichum* – current status and future directions. **Studies in Mycology**, v. 73, n. 1, p. 181–213, 2012.

CAPY, P. *et al.* Relationships Between Transposable Elements Based Upon the Integrase-Transposase Domains: Is There a Common Ancestor? **Journal of Molecular Evolution**, v. 42, p. 359–368, 1996.

CARBÚ, M. *et al.* Recent approaches on the genomic analysis of the phytopathogenic fungus *Colletotrichum* spp. **Phytochemistry Reviews**, v. 19, n. 3, p. 589–601, 2020.

CARR, P. D. *et al.* The Transposon impala Is Activated by Low Temperatures: Use of a Controlled Transposition System to Identify Genes Critical for Viability of *Aspergillus fumigatus*. **Eukaryotic Cell**, v. 9, n. 3, p. 438–448, 2010.

CARR, M.; BENSASSON, D.; BERGMAN, C. M. Evolutionary Genomics of Transposable Elements in *Saccharomyces cerevisiae*. **PLOS ONE**, v. 7, n. 11, 2012.

CHEN, Y. *et al.* Comparative genomics provides new insights into the evolution of Colletotrichum. **Mycosphere**, v. 13, n. 2, p. 134–187, 2022.

CHEN, L. *et al.* Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. **Biochemistry**, v. 30, n. 46, p. 11018–11025, 1991.

CHEN, S. *et al.* Function of the genetic element 'Mona' associated with fungicide resistance in *Monilinia fructicola*. **Molecular Plant Pathology**, v. 18, n. 1, p. 90–97, 2017.

CHEN, F. *et al.* Fungicide-induced transposon movement in *Monilinia fructicola*. **Fungal Genetics and Biology**, v. 85, p. 38–44, 2015.

CHUONG, E. B.; ELDE, N. C.; FESCHOTTE, C. Regulatory activities of transposable elements: from conflicts to benefits. **Nature Reviews Genetics**, v. 18, n. 2, p. 71–86, 2017.

COGONI, C. *et al.* Transgene silencing of the *al-1* gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. **The EMBO Journal**, v. 15, n. 12, p. 3153–3163, 1996.

CROUCH, J. A. *et al.* The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale:* Reconciling sex, recombination and homoplasy in an ''asexual" pathogen. **Fungal Genetics and Biology**, v. 45, n. 3, p. 190–206, 2008.

DA SILVA, L. L. *et al. Colletotrichum*: species complexes, lifestyle, and peculiarities of some sources of genetic variability. **Applied Microbiology and Biotechnology**, v. 104, n. 5, p. 1891–1904, 2020.

DALLERY, J.-F. *et al.* Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. **BMC Genomics**, v. 18, n. 1, p. 667, 2017.

DAMM, U. *et al.* The *Colletotrichum orbiculare* species complex: Important pathogens of field crops and weeds. **Fungal Diversity**, v. 61, n. 1, p. 29–59, 2013.

DANG, Y. *et al.* RNA Interference in Fungi: Pathways, Functions, and Applications. **Eukaryotic Cell**, v. 10, n. 9, p. 1148–1155, 2011.

DE QUEIROZ, C. B. *et al.* The repertoire of effector candidates in *Colletotrichum lindemuthianum* reveals important information about *Colletotrichum* genus lifestyle. **Applied Microbiology and Biotechnology**, v. 103, n. 5, p. 2295–2309, 2019.

DEAN, R. *et al.* The Top 10 fungal pathogens in molecular plant pathology. **Molecular Plant Pathology**, v. 13, n. 4, p. 414–430, 2012.

DELFINI, J. *et al.* Diversity of nutritional content in seeds of Brazilian common bean germplasm. **PLOS ONE**, v. 15, n. 9, 2020.

DENG, H. *et al.* Scatter: a novel family of miniature inverted-repeat transposable elements in the fungus *Botrytis cinerea*. **Journal of Basic Microbiology**, v. 53, n. 10, p. 815–822, 2013.

DESCHAMPS, F. *et al.* Specific expression of the *Fusarium* transposon Fot1 and effects on target gene transcription. **Molecular Microbiology**, v. 31, n. 5, p. 1373–1383, 1999.

DHILLON, B. *et al.* Accidental Amplification and Inactivation of a Methyltransferase Gene Eliminates Cytosine Methylation in *Mycosphaerella graminicola*. **Genetics**, v. 186, n. 1, p. 67–77, 2010.

DIDINGER, C. *et al.* Nutrition and Human Health Benefits of Dry Beans and Other Pulses. *In*: DRY BEANS AND PULSES: John Wiley & Sons, Ltd, 2022. p. 481–504.

DING, L. *et al.* Transposon insertion mutation of Antarctic psychrotrophic fungus for red pigment production adaptive to normal temperature. **Journal of Industrial Microbiology and Biotechnology**, v. 49, n. 1, 2022.

DONG, S.; RAFFAELE, S.; KAMOUN, S. The two-speed genomes of filamentous pathogens: waltz with plants. **Current Opinion in Genetics & Development**, v. 35, Genomes and evolution, p. 57–65, 2015.

DOS SANTOS, L. V. *et al.* Development of new molecular markers for the *Colletotrichum* genus using RetroCl1 sequences. **World Journal of Microbiology and Biotechnology**, v. 28, n. 3, p. 1087–1095, 2012.

EDGAR, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics**, v. 5, n. 1, 113, 2004.

ELLIOTT, T. A.; GREGORY, T. R. Do larger genomes contain more diverse transposable elements? **BMC Evolutionary Biology**, v. 15, n. 1, p. 69, 2015.

EMMONS, S. W. *et al.* Evidence for a transposon in *Caenorhabditis elegans*. **Cell**, [*s. l.*], v. 32, n. 1, p. 55–65, 1983.

EVANGELINOS, M. *et al.* Minos as a novel Tc1/mariner-type transposable element for functional genomic analysis in *Aspergillus nidulans*. **Fungal Genetics and Biology**, v. 81, p. 1–11, 2015.

FAOSTAT.
Available in: https://www.fao.org/faostat/en/#rankings/countries_by_commodity. Access: May 9, 2023.

FATTASH, I. *et al.* Miniature inverted-repeat transposable elements: discovery, distribution, and activity. **Genome**, v. 56, n. 9, p. 475–486, 2013.

FESCHOTTE, C.; PRITHAM, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. **Annual review of genetics**, v. 41, p. 331–368, 2007.

FRANTZESKAKIS, L. *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**, v. 19, n. 1, p. 381, 2018.

FREITAG, M. *et al.* A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. **Proceedings of the National Academy of Sciences**, v. 99, n. 13, p. 8802–8807, 2002.

FUDAL, I. *et al.* Repeat-Induced Point Mutation (RIP) as an Alternative Mechanism of Evolution Toward Virulence in *Leptosphaeria maculans*. **Molecular Plant-Microbe Interactions**, v. 22, n. 8, p. 932–941, 2009.

GAN, P. *et al.* Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. **New Phytologist**, v. 197, n. 4, p. 1236–1249, 2013.

GAN, P. *et al.* Genome Sequence Resources for Four Phytopathogenic Fungi from the *Colletotrichum orbiculare* Species Complex. **Molecular Plant-Microbe Interactions**, v. 32, n. 9, p. 1088–1090, 2019.

GAN, P. *et al.* Genus-Wide Comparative Genome Analyses of *Colletotrichum* Species Reveal Specific Gene Family Losses and Gains during Adaptation to Specific Infection Lifestyles. **Genome Biology and Evolution**, v. 8, n. 5, p. 1467–1481, 2016.

GAN, P. *et al.* Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic *Colletotrichum* fungi. **Environmental Microbiology**, v. 23, n. 10, p. 6004–6018, 2021.

GAO, B. *et al.* Evolution of pogo, a separate superfamily of IS630-Tc1-mariner transposons, revealing recurrent domestication events in vertebrates. **Mobile DNA**, v. 11, n. 1, p. 25, 2020.

GARCÍA-SANTAMARINA, S.; THIELE, D. J. Copper at the Fungal Pathogen-Host Axis. **Journal of Biological Chemistry**, v. 290, n. 31, p. 18945–18953, 2015.

GARRIDO, C. *et al.* Phylogenetic relationships and genome organization of *Colletotrichum acutatum causing* anthracnose in strawberry. **European Journal of Plant Pathology**, v. 125, n. 3, p. 397–411, 2009.

GESSAMAN, J. D.; SELKER, E. U. Induction of H3K9me3 and DNA methylation by tethered heterochromatin factors in *Neurospora crassa*. **Proceedings of the National Academy of Sciences**, v. 114, n. 45, 2017.

GLADYSHEV, E.; KLECKNER, N. Recombination-independent recognition of DNA homology for repeat-induced point mutation. **Current Genetics**, v. 63, n. 3, p. 389–400, 2017.

GOYON, C.; FAUGERON, G. Targeted transformation of Ascobolus immersus and de novo methylation of the resulting duplicated DNA sequences. **Molecular and Cellular Biology**, v. 9, n. 7, p. 2818–2827, 1989.

GUPTA, Y. K. *et al.* Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. **Nature Communications**, v. 14, n. 1, p. 1835, 2023.

HANE, J. K.; OLIVER, R. P. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. **BMC Bioinformatics**, v. 9, n. 1, p. 478, 2008.

HE, C. *et al.* CgT1: a non-LTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. **Molecular & general genetics: MGG**, v. 252, n. 3, p. 320–331, 1996.

HENKE, M. T. *et al.* New Aspercryptins, Lipopeptide Natural Products, Revealed by HDAC Inhibition in *Aspergillus nidulans*. **ACS chemical biology**, v. 11, n. 8, p. 2117–2123, 2016.

HOFF, K. J.; STANKE, M. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. **Nucleic Acids Research**, v. 41, n. W1, 2013.

HUERTA-CEPAS, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. **Nucleic Acids Research**, v. 47, n. Database issue, 2019.

IVICS, Z. *et al.* Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. **Cell**, v. 91, n. 4, p. 501–510, 1997.

JACOBSON, J. W.; MEDHORA, M. M.; HARTL, D. L. Molecular structure of a somatically unstable transposable element in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 22, p. 8684–8688, 1986.

JIANG, N. *et al.* Pack-Mutator–like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition. **Proceedings of the National Academy of Sciences**, v. 108, n. 4, p. 1537–1542, 2011.

JOHN CLUTTERBUCK, A. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. **Fungal Genetics and Biology**, v. 48, n. 3, p. 306–326, 2011.

KALYAANAMOORTHY, S. *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. **Nature Methods**, v. 14, n. 6, p. 587–589, 2017.

KAPITONOV, V. V.; JURKA, J. Helitrons on a roll: eukaryotic rolling-circle transposons. **Trends in Genetics**, v. 23, n. 10, p. 521–529, 2007.

KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1160–1166, 2019.

KIMATI, H.; GALLI, F. *G. cingulata* f.sp. *phaseoli*, the ascogenous state of the causal agent of Bean anthracnose. **Anais da Escola Superior de Agricultura "'Luiz de Queiroz'" (Brazil)**, v. 27, 1970.

KOUZMINOVA, E.; SELKER, E. U. dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. **The EMBO journal**, v. 20, n. 15, p. 4309–4323, 2001.

KUANG, H. *et al.* Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New functional implications for MITEs. **Genome Research**, v. 19, n. 1, p. 42–56, 2009.

LA TORRE, A.; IOVINO, V.; CARADONIA, F. Copper in plant protection: current situation and prospects. **Phytopathologia Mediterranea**, v. 57, n. 2, p. 201–236, 2018.

LANGIN, T.; CAPY, P.; DABOUSSI, M.-J. The transposable element impala, a fungal member of the Tc1-mariner superfamily. **Molecular and General Genetics MGG**, v. 246, n. 1, p. 19–28, 1995.

LATUNDE-DADA, A. O. *Colletotrichum*: tales of forcible entry, stealth, transient confinement and breakout. **Molecular Plant Pathology**, v. 2, n. 4, p. 187–198, 2001.

LELWALA, R. V. *et al.* Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in *Colletotrichum tanaceti*. **PLOS ONE**, v. 14, n. 5, 2019.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. **Nucleic Acids Research**, v. 49, n. W1, 2021.

LEWIS, Z. A. *et al.* Identification of DIM-7, a protein required to target the DIM-5 H3 methyltransferase to chromatin. **Proceedings of the National Academy of Sciences**, v. 107, n. 18, p. 8310–8315, 2010.

LI, W.-C. *et al. Trichoderma reesei* complete genome sequence, repeat-induced point mutation, and partitioning of CAZyme gene clusters. **Biotechnology for Biofuels**, v. 10, n. 1, p. 170, 2017.

LI, W.-C.; CHEN, C.-L.; WANG, T.-F. Repeat-induced point (RIP) mutation in the industrial workhorse fungus *Trichoderma reesei*. **Applied Microbiology and Biotechnology**, v. 102, 2018.

LÓPEZ-BERGES, M. S. *et al.* Identification of virulence genes in *Fusarium oxysporum* f. sp. *lycopersici* by large-scale transposon tagging. **Molecular Plant Pathology**, v. 10, n. 1, p. 95–107, 2009.

LORRAIN, C. *et al.* Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. **G3 Genes|Genomes|Genetics**, v. 11, n. 4, 2021.

LU, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. **Nucleic Acids Research**, v. 48, n. D1, 2020.

LUO, C.-X. *et al.* Occurrence and Detection of the DMI Resistance-Associated Genetic Element 'Mona' in *Monilinia fructicola*. **Plant Disease**, v. 92, n. 7, p. 1099–1103, 2008.

LUO, C.-X.; SCHNABEL, G. The cytochrome P450 lanosterol 14alpha-demethylase gene is a demethylation inhibitor fungicide resistance determinant in *Monilinia fructicola* field isolates from Georgia. **Applied and Environmental Microbiology**, v. 74, n. 2, p. 359–366, 2008.

MA, Z. *et al.* Overexpression of the 14α-Demethylase Target Gene (CYP51) Mediates Fungicide Resistance in *Blumeriella jaapii*. **Applied and Environmental Microbiology**, v. 72, n. 4, p. 2581–2585, 2006.

MALAGNAC, F. *et al.* A Gene Essential for De Novo Methylation and Development in *Ascobolus* Reveals a Novel Type of Eukaryotic DNA Methyltransferase Structure. **Cell**, v. 91, n. 2, p. 281–290, 1997.

MASEL, A. M.; IRWIN, K, John A. G.; MANNERS, J. M. DNA addition or deletion is associated with a major karyotype polymorphism in the fungal phytopathogen *Colletotrichum gloeosporioides.* **Molecular and General Genetics MGG**, v. 237, n. 1, p. 73–80, 1993.

MAT RAZALI, N.; CHEAH, B. H.; NADARAJAH, K. Transposable Elements Adaptive Role in Genome Plasticity, Pathogenicity and Evolution in Fungal Phytopathogens. **International Journal of Molecular Sciences**, v. 20, n. 14, p. 3597, 2019.

MCCLINTOCK, B. Controlling Elements and the Gene. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 21, p. 197–216, 1956.

MIN, B. *et al.* Unusual genome expansion and transcription suppression in ectomycorrhizal *Tricholoma matsutake* by insertions of transposable elements. **PLOS ONE**, v. 15, n. 1, 2020.

MÖLLER, M. *et al.* Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. **PLOS Genetics**, v. 17, n. 3, 2021.

MÖLLER, M.; STUKENBROCK, E. H. Evolution and genome architecture in fungal plant pathogens. **Nature Reviews Microbiology**, v. 15, n. 12, p. 756–771, 2017.

MONTIEL, M. D.; LEE, H. A.; ARCHER, D. B. Evidence of RIP (repeat-induced point mutation) in transposase sequences of *Aspergillus oryzae*. **Fungal genetics and biology: FG & B**, v. 43, n. 6, p. 439–445, 2006.

NABI, A. *et al. Phaseolus vulgaris-Colletotrichum lindemuthianum* Pathosystem in the Post-Genomic Era: An Update. **Current Microbiology**, v. 79, n. 2, p. 36, 2022.

NAITO, K. *et al.* Dramatic amplification of a rice transposable element during recent domestication. **Proceedings of the National Academy of Sciences**, v. 103, n. 47, p. 17620–17625, 2006.

NARAYANAVARI, S. A. *et al.* Sleeping Beauty transposition: from biology to applications. **Critical Reviews in Biochemistry and Molecular Biology**, v. 52, n. 1, p. 18–44, 2017.

NGUYEN, L.-T. *et al.* IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. **Molecular Biology and Evolution**, v. 32, n. 1, p. 268–274, 2015.

NUNES, M. P. B. A. *et al.* Relationship of *Colletotrichum lindemuthianum* races and resistance loci in the Phaseolus vulgaris L. genome. **Crop Science**, v. 61, n. 6, p. 3877–3893, 2021.

O'CONNELL, R. J. *et al.* Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. **Nature Genetics**, v. 44, n. 9, p. 1060–1065, 2012.

O'CONNELL, R. J.; BAILEY, J. A.; RICHMOND, D. V. Cytology and physiology of infection of Phaseolus vulgaris by *Colletotrichum lindemuthianum*. **Physiological Plant Pathology**, v. 27, n. 1, p. 75–98, 1985.

OGASAWARA, H. *et al.* Crawler, a novel Tc1/mariner-type transposable element in *Aspergillus oryzae* transposes under stress conditions. **Fungal Genetics and Biology**, v. 46, n. 6, p. 441–449, 2009.

OGGENFUSS, U.; CROLL, D. Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. **PLOS Pathogens**, v. 19, n. 2, 2023.

OKI, N. *et al.* A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*. **Genes & Genetic Systems**, v. 83, n. 4, p. 321–329, 2008.

O'SULLIVAN, D. *et al.* Variation in genome organization of the plant pathogenic fungus *Colletotrichum lindemuthianum*. **Current Genetics**, v. 33, n. 4, p. 291–298, 1998.

PADDER, B. A. *et al. Colletotrichum Lindemuthianum,* the Causal Agent of Bean Anthracnose. **Journal of Plant Pathology**, v. 99, n. 2, p. 317–330, 2017.

PALAZZO, A. *et al.* Transcriptionally promiscuous "blurry" promoters in *Tc1/mariner* transposons allow transcription in distantly related genomes. **Mobile DNA**, v. 10, n. 1, p. 13, 2019.

PALAZZO, A.; MARSANO, R. M. Transposable elements: a jump toward the future of expression vectors. **Critical Reviews in Biotechnology**, v. 41, n. 5, p. 792–808, 2021.

PASTOR CORRALES, M. A.; TU, J. C. Anthracnose. *In*: International Center for Tropical Agriculture, 1989.

PATHANIA, A.; SHARMA, S.; SHARMA, P. Common Bean. *In*: Broadening the Genetic Base of Grain Legumes. 2014.

PAULINO, P. P. S. *et al.* Occurrence of anthracnose pathogen races and resistance genes in common bean across 30 years in Brazil. **Agronomy Science and Biotechnology**, v. 8, p. 1–21, 2022.

PAYSAN-LAFOSSE, T. *et al.* InterPro in 2022. **Nucleic Acids Research**, v. 51, n. D1, 2023.

PEGLER, J. L. *et al.* Miniature Inverted-Repeat Transposable Elements: Small DNA Transposons That Have Contributed to Plant MICRORNA Gene Evolution. **Plants**, v. 12, n. 5, p. 1101, 2023.

PEREIRA, D. *et al.* Population genomics of transposable element activation in the highly repressive genome of an agricultural pathogen. **Microbial Genomics**, v. 7, n. 8, 2021.

PERFECT, S. E. *et al.* Colletotrichum: A Model Genus for Studies on Pathology and Fungal–Plant Interactions. **Fungal Genetics and Biology**, v. 27, n. 2, p. 186–198, 1999.

PIETROKOVSKI, S.; HENIKOFF, S. A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. **Molecular & general genetics: MGG**, v. 254, n. 6, p. 689–695, 1997.

PIRES, A. S. *et al.* Cytogenomic characterization of *Colletotrichum kahawae*, the causal agent of coffee berry disease, reveals diversity in minichromosome profiles and genome size expansion. **Plant Pathology**, v. 65, n. 6, p. 968–977, 2016.

PLASTERK, R. H. A. The Tc1/mariner Transposon Family. *In*: SAEDLER, H.; GIERL, A. (org.). **Transposable Elements**. Berlin, Heidelberg: Springer, 1996. (Current Topics in Microbiology and Immunology). p. 125–143.

PLAUMANN, P.-L. *et al.* A Dispensable Chromosome Is Required for Virulence in the Hemibiotrophic Plant Pathogen *Colletotrichum higginsianum*. **Frontiers in Microbiology**, v. 9, 2018.

PLAUMANN, P.-L.; KOCH, C. The Many Questions about Mini Chromosomes in *Colletotrichum* spp. **Plants**, v. 9, n. 5, p. 641, 2020.

RAO, S. *et al.* The Landscape of Repetitive Elements in the Refined Genome of Chilli Anthracnose Fungus *Colletotrichum truncatum*. **Frontiers in Microbiology**, v. 9, 2018.

RIGGLE, P. J.; KUMAMOTO, C. A. Role of a *Candida albicans* P1-Type ATPase in Resistance to Copper and Silver Ion Toxicity. **Journal of Bacteriology**, v. 182, n. 17, p. 4899–4905, 2000.

ROBERTSON, H. M. The Tc1-mariner superfamily of transposons in animals. **Journal of Insect Physiology**, v. 41, n. 2, p. 99–105, 1995.

RODRÍGUEZ-GUERRA, R. *et al.* Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. **Mycologia**, v. 97, n. 4, p. 793–803, 2005.

ROMANO, N.; MACINO, G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. **Molecular Microbiology**, v. 6, n. 22, p. 3343–3353, 1992.

ROSADA, L. J. *et al.* Parasexuality in Race 65 *Colletotrichum lindemuthianum* Isolates. **Journal of Eukaryotic Microbiology**, v. 57, n. 4, p. 383–384, 2010.

ROSSIGNOL, J.-L.; FAUGERON, G. Gene inactivation triggered by recognition between DNA repeats. **Experientia**, v. 50, n. 3, p. 307–317, 1994.

SANDOVAL-VILLEGAS, N. *et al.* Contemporary Transposon Tools: A Review and Guide through Mechanisms and Applications of Sleeping Beauty, piggyBac and Tol2 for Genome Engineering. **International Journal of Molecular Sciences**, v. 22, n. 10, p. 5084, 2021.

SANTIAGO, N. *et al.* Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis thaliana*. **Molecular Biology and Evolution**, v. 19, n. 12, p. 2285–2293, 2002.

SCHLIEBNER, I. *et al.* New gene models and alternative splicing in the maize pathogen *Colletotrichum graminicola* revealed by RNA-Seq analysis. **BMC Genomics**, v. 15, n. 1, p. 842, 2014.

SCHMIDT, S. M. *et al.* MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. **BMC genomics**, v. 14, p. 119, 2013.

SCHMITZ, J.; BROSIUS, J. Exonization of transposed elements: A challenge and opportunity for evolution. **Biochimie**, v. 93, n. 11, p. 1928–1934, 2011.

SCHNABEL, G.; JONES, A. L. The 14alpha-Demethylasse (CYP51A1) Gene is Overexpressed in *Venturia inaequalis* Strains Resistant to Myclobutanil. **Phytopathology**, v. 91, n. 1, p. 102–110, 2001.

SELKER, E. U. 15 - Repeat-Induced Gene Silencing in Fungi. *In*: DUNLAP, J. C.; WU, C. ting. **Advances in Genetics**: Academic Press, 2002. (Homology Effects). v. 46, p. 439–450.

SELKER, E. U. *et al.* Rearrangement of duplicated DNA in specialized cells of Neurospora. **Cell**, v. 51, n. 5, p. 741–752, 1987.

SELKER, E. U.; STEVENS, J. N. Signal for DNA Methylation Associated with Tandem Duplication in *Neurospora crassa*. **Molecular and Cellular Biology**, v. 7, n. 3, p. 1032–1038, 1987.

SHAHID, S.; SLOTKIN, R. K. The current revolution in transposable element biology enabled by long reads. **Current Opinion in Plant Biology**, v. 54, Genome studies and molecular genetics, p. 49–56, 2020.

SHIU, P. K. T. *et al.* Meiotic Silencing by Unpaired DNA. **Cell**, v. 107, n. 7, p. 905–916, 2001.

SPRADLING, A. C. *et al.* Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 24, p. 10824–10830, 1995.

STALDER, L. *et al.* The population genetics of adaptation through copy number variation in a fungal plant pathogen. **Molecular Ecology**, v. 32, n. 10, p. 2443–2460, 2023.

SUN, X. *et al.* PdCYP51B, a new putative sterol 14α-demethylase gene of *Penicillium digitatum* involved in resistance to imazalil and other fungicides inhibiting ergosterol synthesis. **Applied Microbiology and Biotechnology**, v. 91, n. 4, p. 1107–1119, 2011.

SUN, X. *et al.* PdMLE1, a specific and active transposon acts as a promoter and confers *Penicillium digitatum* with DMI resistance. **Environmental Microbiology Reports**, v. 5, n. 1, p. 135–142, 2013a.

SUN, X. *et al.* Genome-wide investigation into DNA elements and ABC transporters involved in imazalil resistance in *Penicillium digitatum*. **FEMS Microbiology Letters**, v. 348, n. 1, p. 11–18, 2013b.

TAGA, M. *et al.* Cytological analyses of the karyotypes and chromosomes of three *Colletotrichum* species, *C. orbiculare*, *C. graminicola* and *C. higginsianum*. **Fungal Genetics and Biology**, v. 82, p. 238–250, 2015.

TENZEN, T.; MATSUTANI, S.; OHTSUBO, E. Site-specific transposition of insertion sequence IS630. **Journal of Bacteriology**, v. 172, n. 7, p. 3830–3836, 1990.

TORRES, D. E.; THOMMA, B. P. H. J.; SEIDL, M. F. Transposable Elements Contribute to Genome Dynamics and Gene Expression Variation in the Fungal Plant Pathogen *Verticillium dahliae*. **Genome Biology and Evolution**, v. 13, n. 7, p. evab135, 2021.

TSUSHIMA, A. *et al.* Genomic Plasticity Mediated by Transposable Elements in the Plant Pathogenic Fungus *Colletotrichum higginsianum*. **Genome Biology and Evolution**, v. 11, n. 5, p. 1487–1500, 2019.

TSUSHIMA, A.; SHIRASU, K. Genomic resources of *Colletotrichum* fungi: development and application. **Journal of General Plant Pathology**, v. 88, n. 6, p. 349–357, 2022.

VAN DAM, P.; REP, M. The Distribution of Miniature Impala Elements and SIX Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. **Journal of Molecular Evolution**, v. 85, n. 1–2, p. 14–25, 2017.

VAN WYK, S. *et al.* Genome-Wide Analyses of Repeat-Induced Point Mutations in the Ascomycota. **Frontiers in Microbiology**, v. 11, 2021.

VILLANI, S. M. *et al.* Overexpression of the *CYP51A1* Gene and Repeated Elements are Associated with Differential Sensitivity to DMI Fungicides in *Venturia inaequalis*. **Phytopathology**, v. 106, n. 6, p. 562–571, 2016.

WALTER, M. **Transposon regulation upon dynamic loss of DNA methylation**. 2015. Thesis (PhD in Biology), Université Pierre et Marie Curie, Paris, France, 2015.

WARBURTON, P. E. *et al.* Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. **Genome Research**, v. 14, n. 10A, p. 1861–1869, 2004.

WELLS, J. N.; FESCHOTTE, C. A Field Guide to Eukaryotic Transposable Elements. **Annual Review of Genetics**, v. 54, n. 1, p. 539–561, 2020.

WESSLER, S. R.; BUREAU, T. E.; WHITE, S. E. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. **Current Opinion in Genetics & Development**, v. 5, n. 6, p. 814–821, 1995.

WICKER, T. *et al.* A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973–982, 2007.

WYK, S. van *et al.* The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. **PeerJ**, v. 7, 2019.

YUAN, Y.-W.; WESSLER, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. **Proceedings of the National Academy of Sciences**, v. 108, n. 19, p. 7884–7889, 2011.

ZAMPOUNIS, A. *et al.* Genome Sequence and Annotation of Colletotrichum higginsianum, a Causal Agent of Crucifer Anthracnose Disease. **Genome Announcements**, v. 4, n. 4, 2016.

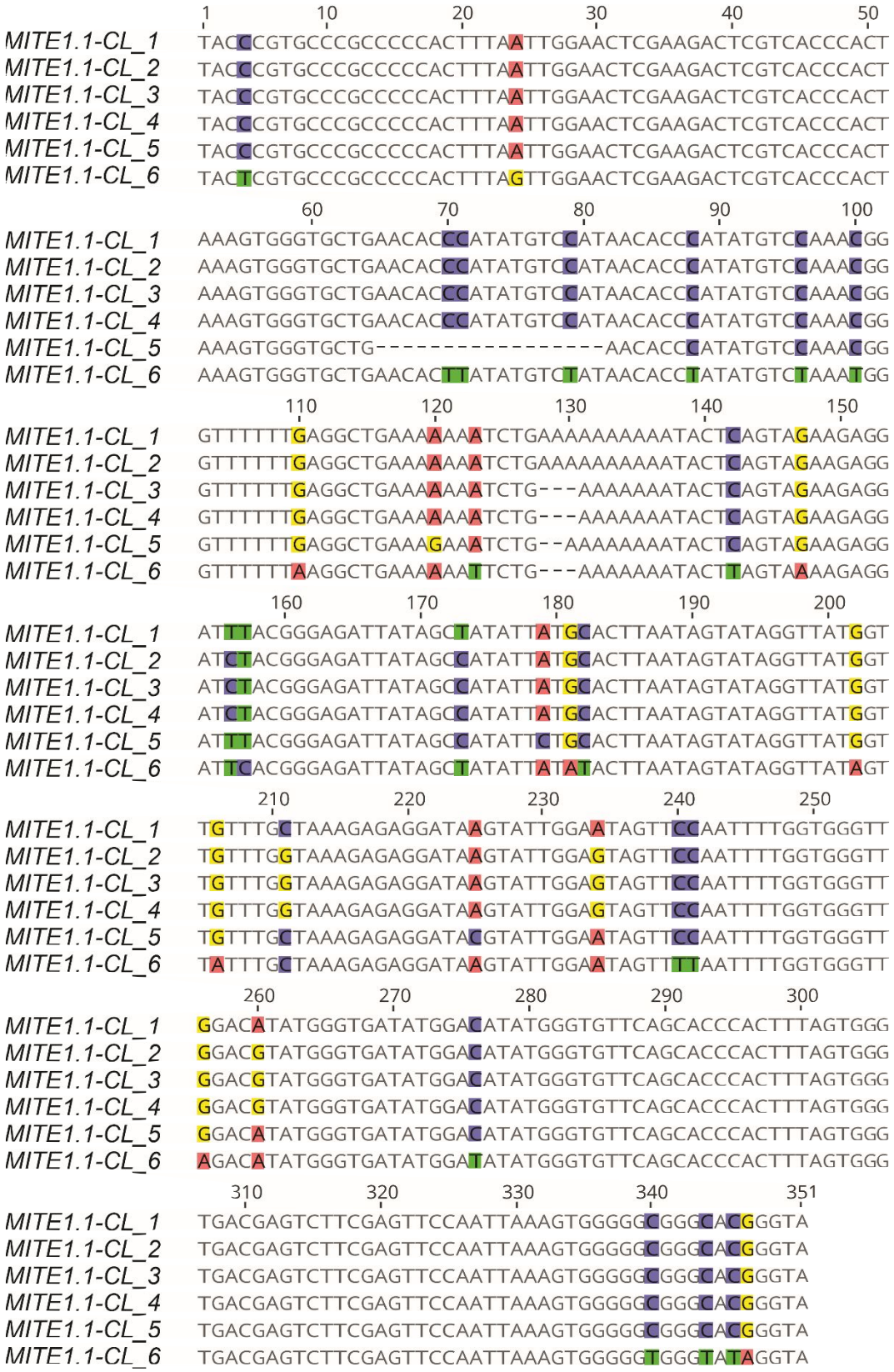
ZHU, P.; OUDEMANS, P. V. A long terminal repeat retrotransposon *Cgret* from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. **Current Genetics**, v. 38, n. 5, p. 241–247, 2000.

Appendix A – Taxa and GenBank accession numbers for sequences used in this study.

Species	**RID/MASC1**	**DIM2**	**Dnmt1/MASC2a**	**Dnmt1b**	**Dnmt1c**	**Dnmt2**
Neurospora crassa	XP_011392925.1	AAK49954.1	-	-	-	-
Ascobolus immersus	AAC49849.1	RPA84378.1	AAC03766.1	-	-	-
Coprinopsis cinerea	-	-	XP_001829400.2	XP_001833175.2	-	-
Laccaria bicolor	-	-	XP_001873517.1	XP_001875435.1	-	-
Agaricus bisporus	-	-	XP_006458036.1	XP_006453977.1	-	-
Pleurotus ostreatus	-	-	KDQ33030.1	KDQ30406.1	KDQ31353.1	
Colletotrichum lindemuthianum	?	?	-	-	-	-
Colletotrichum higginsianum	XP_018157144.1	XP_018164533.1	-	-	-	-
Colletotrichum graminicola	XP_008090605.1	XP_008095792.1	-	-	-	-
Colletotrichum truncatum	XP_036589023.1	XP_036579617.1	-	-	-	-
Colletotrichum trifolii	TDZ60654.1	TDZ37985.1	-	-	-	-
Colletotrichum spinosum	TDZ34105.1	TDZ29949.1	-	-	-	-
Colletotrichum tanaceti	KAJ0165390.1	KAJ0167397.1	-	-	-	-
Colletotrichum orchidophilum	XP_022480581.1	XP_022481547.1	-	-	-	-
Colletotrichum fioriniae	XP_053047814.1	EXF74904.1	-	-	-	-
Colletotrichum musicola	KAF6844540.1	KAF6833979.1	-	-	-	-
Colletotrichum orbiculare	TDZ21265.1	TDZ21777.1	-	-	-	-
Colletotrichum sidae	TEA20388.1	TEA11878.1	-	-	-	-
Colletotrichum sojae	KAF6799487.1	KAF6815586.1	-	-	-	-
Colletotrichum plurivorum	KAF6835883.1	KAF6834828.1	-	-	-	-
Colletotrichum karsti	XP_038748878.1	XP_038743037.1	-	-	-	-
Colletotrichum tropicale	KAF4820807.1	KAF4822175.1	-	-	-	-
Colletotrichum siamense	KAF4871104.1	KAF4806978.1	-	-	-	-
Colletotrichum gloeosporioides	XP_045256529.1	EQB47784.1	-	-	-	-
Colletotrichum camelliae	KAH0430935.1	KAH0441906.1	-	-	-	-
Colletotrichum asianum	KAF0318980.1	KAF0328797.1	-	-	-	-
Colletotrichum incanum	KZL86356.1	OHW99548.1	-	-	-	-
Colletotrichum scovillei	XP_035333537.1	KAG7042851.1	-	-	-	-
Colletotrichum sublineola	KDN71417.1	KDN71636.1	-	-	-	-
Colletotrichum fructicola	XP_031877683.1	KAF4483729.1	-	-	-	-

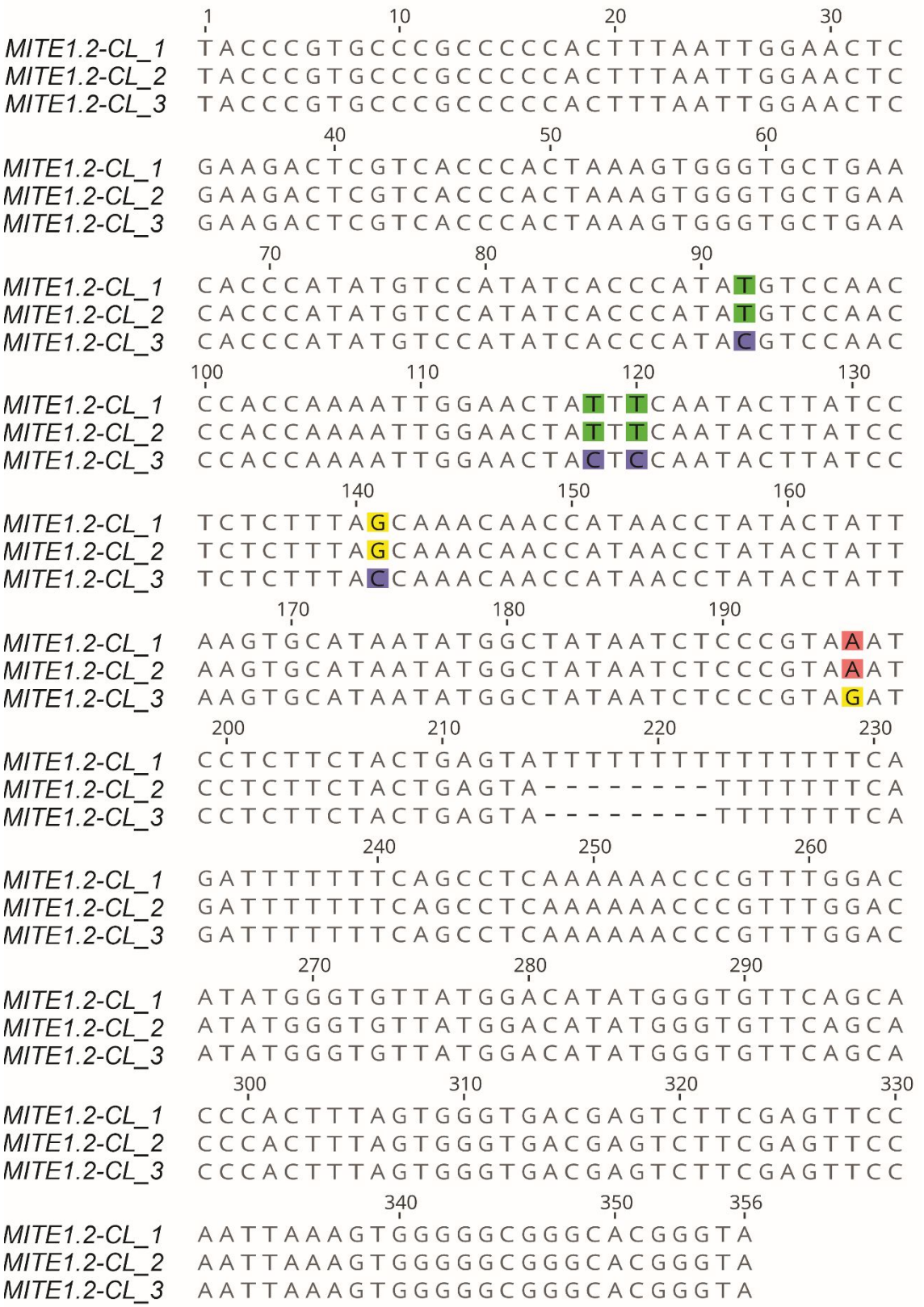
Species	RID/MASC1	DIM2	Dnmt1/MASC2a	Dnmt1b	Dnmt1c	Dnmt2
Colletotrichum simmondsii	KXH31686.1	KXH50054.1	-	-	-	-
Colletotrichum filicis	KAI3532994.1	KAI3541129.1	-	-	-	-
Colletotrichum lupini	XP_049145007.1	XP_049142404.1	-	-	-	-
Colletotrichum nymphaeae	KXH46116.1	KXH64228.1	-	-	-	-
Colletotrichum abscissum	KAI3540457.1	KAI3545951.1	-	-	-	-
Colletotrichum liriopes	GJC82750.1	GJC77729.1	-	-	-	-
Colletotrichum tofieldiae	GKT96222.1	GKT55587.1	-	-	-	-
Colletotrichum spaethianum	XP_049124832.1	XP_049130814.1	-	-	-	-
Colletotrichum viniferum	KAF4929374.1	KAF4911010.1	-	-	-	-
Verticillium dahliae	XP_009656107.1	RBQ98597.1	-	-	-	-
Verticillium longisporum	KAG7120512.1	KAG7112761.1	-	-	-	-
Verticillium nonalfalfae	XP_028490376.1	XP_028498554.1	-	-	-	-
Fusarium mangiferae	XP_041680684.1	XP_041689532.1	-	-	-	-
Fusarium verticillioides	XP_018745179.1	XP_018759005.1	-	-	-	-
Fusarium fujikuroi	KLP14596.1	SCO51736.1	-	-	-	-
Fusarium oxysporum	KAH7227828.1	KAJ4105733.1	-	-	-	-
Trichoderma harzianum	KKO99120.1	KKP05280.1	-	-	-	-
Trichoderma simmonsii	QYS93156.1	QYS96178.1	-	-	-	-
Trichoderma gamsii	PNP39283.1	XP_018657421.1	-	-	-	-
Trichoderma reesei	AEM66210.1	XP_006964860.1	-	-	-	-
Pyricularia oryzae	KAH8838335.1	KAI7923655.1	-	-	-	-
Neurospora tetrasperma	AAM27410.1	XP_009853512.1	-	-	-	-
Aspergillus fumigatus	KAH1495790.1	-	-	-	-	-
Aspergillus niger	GKZ82800.1	-	-	-	-	-
Aspergillus nidulans	XP_664242.1	-	-	-	-	-
Zymoseptoria brevis	KJY02274.1	KJX92855.1	-	-	-	-
Zymoseptoria tritici	SMY24142.1	-	-	-	-	-
Tuber melanosporum	XP_002842459.1	XP_002837027.1	-	-	-	-
Tuber aestivum	CUS09459.1	CUS10452.1	-	-	-	-
Tuber magnatum	PWW78466.1	PWW80586.1	-	-	-	-
Schizosaccharomyces pombe	-	-	-	-	-	NP_595687.1

Appendix B – Alignment of copies belonging to the *MITE1.1-CL* family.



Source: the author.

Appendix C – Alignment of copies belonging to the *MITE1.2-CL* family.


1 10 20 30
MITE1.2-CL_1 TACCCGTGCCCGCCCCCACTTTAATTGGAACTC
MITE1.2-CL_2 TACCCGTGCCCGCCCCCACTTTAATTGGAACTC
MITE1.2-CL_3 TACCCGTGCCCGCCCCCACTTTAATTGGAACTC
40 50 60
MITE1.2-CL_1 GAAGACTCGTCACCCACTAAAGTGGGTGCTGAA
MITE1.2-CL_2 GAAGACTCGTCACCCACTAAAGTGGGTGCTGAA
MITE1.2-CL_3 GAAGACTCGTCACCCACTAAAGTGGGTGCTGAA
70 80 90
MITE1.2-CL_1 CACCCATATGTCCATATCACCCATATGTCCAAC
MITE1.2-CL_2 CACCCATATGTCCATATCACCCATATGTCCAAC
MITE1.2-CL_3 CACCCATATGTCCATATCACCCATACGTCCAAC
100 110 120 130
MITE1.2-CL_1 CCACCAAAATTGGAACTATTTCAATACTTATCC
MITE1.2-CL_2 CCACCAAAATTGGAACTATTTCAATACTTATCC
MITE1.2-CL_3 CCACCAAAATTGGAACTACTCCAATACTTATCC
140 150 160
MITE1.2-CL_1 TCTCTTTAGCAAACAACCATAACCTATACTATT
MITE1.2-CL_2 TCTCTTTAGCAAACAACCATAACCTATACTATT
MITE1.2-CL_3 TCTCTTTACCAAACAACCATAACCTATACTATT
170 180 190
MITE1.2-CL_1 AAGTGCATAATATGGCTATAATCTCCCGTAAAT
MITE1.2-CL_2 AAGTGCATAATATGGCTATAATCTCCCGTAAAT
MITE1.2-CL_3 AAGTGCATAATATGGCTATAATCTCCCGTAGAT
200 210 220 230
MITE1.2-CL_1 CCTCTTCTACTGAGTATTTTTTTTTTTTTTCA
MITE1.2-CL_2 CCTCTTCTACTGAGTA--------TTTTTTTCA
MITE1.2-CL_3 CCTCTTCTACTGAGTA--------TTTTTTTCA
240 250 260
MITE1.2-CL_1 GATTTTTTTTCAGCCTCAAAAAACCCGTTTGGAC
MITE1.2-CL_2 GATTTTTTTTCAGCCTCAAAAAACCCGTTTGGAC
MITE1.2-CL_3 GATTTTTTTTCAGCCTCAAAAAACCCGTTTGGAC
270 280 290
MITE1.2-CL_1 ATATGGGTGTTATGGACATATGGGTGTTCAGCA
MITE1.2-CL_2 ATATGGGTGTTATGGACATATGGGTGTTCAGCA
MITE1.2-CL_3 ATATGGGTGTTATGGACATATGGGTGTTCAGCA
300 310 320 330
MITE1.2-CL_1 CCCACTTTAGTGGGTGACGAGTCTTCGAGTTCC
MITE1.2-CL_2 CCCACTTTAGTGGGTGACGAGTCTTCGAGTTCC
MITE1.2-CL_3 CCCACTTTAGTGGGTGACGAGTCTTCGAGTTCC
340 350 356
MITE1.2-CL_1 AATTAAAGTGGGGGCGGGCACGGGTA
MITE1.2-CL_2 AATTAAAGTGGGGGCGGGCACGGGTA
MITE1.2-CL_3 AATTAAAGTGGGGGCGGGCACGGGTA

Appendix D – Elements nearby genes within the *Mariner1-CL*, *MITE1-CL*, *MITE2-CL*, and *MITE3-CL* families.

Element	Genomic position	Gene upstream the element	Distance from transposon	GenBank Reference	Gene downstream the element	GenBank Reference	Distance from transposon
Mariner1-CL_1	Scaffold 59: 35.999-35.993	Pseudouridine-metabolizing bifunctional protein	2992 bp	TDZ17802.1	U3 small nucleolar RNA-associated protein 18	TDZ17799.1	528 bp
Mariner1-CL_2	Scaffold 43: 264.712-266.706	54S ribosomal protein yml6	146 bp	TDZ34617.1	54S ribosomal protein L35	TEA22241.1	77 bp
Mariner1-CL_3	Scaffold 3: 1.147.257-1.149.251	-	-	-	Choline-phosphate cytidylyltransferase	TDZ13251.1	4812 bp
Mariner1-CL_4	Scaffold 5: 4.030.357-4.032.351	Putative sulfate transporter	424 bp	TDZ49656.1	Galacturan 1,4-alpha-galacturonidase C	TDZ20086.1	1847 bp
Mariner1-CL_5	Scaffold 1: 6.637.741-6.637.740	Carnitine O-acetyltransferase	291 bp	TDZ25017.1	-	-	-
Mariner1-CL_6	Scaffold 9: 2.409.679-2.411.674	-	-	-	Grainyhead-like protein 2-like protein	TDZ28542.1	3794 bp
Mariner1-CL_7	Scaffold 8: 2.065.757-2.067.750	-	-	-	Sucrose utilization protein SUC1	TEA15704.1	148 bp
Mariner1-CL_8	Scaffold 17: 914.770-916.764	Uracil-regulated protein 1	629 bp	TDZ23273.1	-	-	-
Mariner1-CL_9	Scaffold 61: 222.532-224.526	Arca-like protein	120 bp	KAF0324368.1	Scyllo-inositol 2-dehydrogenase	TEA16378.1	1104 bp
Mariner1-CL_10	Scaffold 5: 826.577-828.571	Uncharacterized protein	152 bp	XP_036588841.1	ATP-dependent 6-phosphofructokinase	TDZ15051.1	279 bp
Mariner1-CL_11	Scaffold 1: 6.637.741-6.637.740	AdoMet-dependent rRNA methyltransferase SPB1	1953 bp	KAF4921029.1	-	-	-
Mariner1-CL_12	Scaffold 20: 3.848.357-3.850.347	Dihydroxyacetone synthase	929 bp	TDZ31404.1	Transcription activator AMTR1	TDZ18009.1	775 bp
Mariner1-CL_13	Scaffold 26: 326.669-328.537	-	-	-	Folliculin	TQN69034.1	999 bp

Element	Genomic position	Gene upstream the element	Distance from transposon	GenBank Reference	Gene downstream the element	GenBank Reference	Distance from transposon
Mariner1-CL_14	Scaffold 43: 89.679-91.673	Cutinase transcription factor 1 beta	296 bp	TDZ31244.1	-	-	-
MITE1.1-CL_1	Scaffold 49: 89.679-91.673	-	-	-	Mitogen activated protein kinase	AAD50496.1	143 bp
MITE1.1-CL_2	Scaffold 4: 663.692-664.040	Exportin-T	962 bp	TDZ24371.1	GPI anchored serine-rich protein	KZL68891.1	4009 bp
MITE1.1-CL_3	Scaffold 11: 1.745.262-1.745.610	Uncharacterized protein	3964 bp	TDZ40738.1	-	-	-
MITE1.1-CL_4	Scaffold 2: 2.154.711-2.155.061	Homocysteine synthase	244 bp	TDZ46062.1	Uncharacterized protein	TDZ25070.1	1323 bp
MITE1.1-CL_1	Scaffold 14: 361.170-361.520	Uncharacterized protein	992 bp	DZ65113.1	Galactose oxidase	TDZ25686.1	2174 bp
MITE1.2-CL_2	Scaffold 35: 180.559-180.906	Intermediate cleaving peptidase 55	732 bp	TDZ24550.1	Hsk1-interacting molecule 1	TDZ30326.1	279 bp
MITE1.2-CL_3	Scaffold 14: 1.078.385-1.078.740	Arrestin domain-containing protein	3292 bp	XP_036588422.1	Zinc transporter	TDZ58581.1	1106 bp
MITE1.2-CL_4	Scaffold 21: 1.126.804-1.127.151	proteophosphoglycan ppg4	1466 bp	KAF6838901.1	1-Phosphatidylinositol 4,5-bisphosphate phosphodiesterase 1	TDZ21071.1	83 bp
MITE1.3-CL_1	Scaffold 27: 259.717-260.634	Uncharacterized protein	244 bp	TDZ36238.1	Copper resistance protein crd2	XP_018159318.1	346 bp
MITE1.3-CL_2	Scaffold 5: 3.229.574-3.230.491	Uncharacterized protein	154 bp	TDZ48403.1	Aspercryptin biosynthesis cluster-specific transcription regulator atnN	TDZ37324.1	0**
MITE1.4-CL_1	Scaffold 41: 479.082-479.188	Guanine nucleotide exchange factor LTE1	3270 bp	TEA18202.1	Vesicular-fusion protein sec18	TDZ39626.1	1943 bp
MITE1.5-CL_1	Scaffold 41: 215.647-215.723	BC10 family protein	3583 bp	OHX00344.1	Protein MNN4	TDZ22576.1	2286 bp

*Part of the transposon overlaid the predicted mRNA (750 bp from the 5' end)

Element	Genomic position	Gene upstream the element	Distance from transposon	GenBank Reference	Gene downstream the element	GenBank Reference	Distance from transposon
MITE1.5-CL_2	Scaffold 31: 677.736-677.797	Non-reducing end alpha-L-arabinofuranosidase	1614 bp	TDZ68042.1	Itaconate transport protein	TDZ27911.1	992 bp
MITE1.6-CL_1	Scaffold 1: 2.066.584-2.066.650	MutS protein-like protein 1	833 bp	TEA11277.1	Uncharacterized protein	TDZ73285.1	1924 bp
MITE2-CL_1	Scaffold 5: 64.267-64.707	Chromosome segregation in meiosis protein 3	1555 bp	TDZ13840.1	-	-	-
MITE2-CL_2	Scaffold 23: 895.817-896.255	Uncharacterized protein	847 bp	KAF6784891.1	2,2-Dialkylglycine decarboxylase	TEA14939.1	962 bp
MITE2-CL_3	Scaffold 1: 7.730.643-7.731.083	Urease accessory protein	214 bp	XP_036583082.1	-	-	-
MITE3-CL_1	Scaffold 14: 871.405-871.844	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase	818 bp	TDZ61066.1	-	-	-