Potential effect of retrotransposons as determinant factors in the genetic variability and virulence of *Hemileia vastatrix*

Alexandre Miguel Ribeiro Laureano

Mestrado em Biologia dos Recursos Vegetais

Versão definitiva, 2024

Presidente

Doutora Maria Leonor Morais Mota Cecílio, Professora auxiliar do(a) Instituto Superior de Agronomia.

Vogais

Doutora Maria Manuela Antunes Gomes da Silva, Professora auxiliar do(a) Instituto Superior de Agronomia.

Doutora Dora Cristina Vicente Batista Lyon de Castro, Professora auxiliar do(a) Instituto Superior de Agronomia, Orientadora.

Dissertação orientada por:

Dora Cristina Vicente Batista Lyon de Castro, PhD

Pedro Manuel Vieira Talhinhas, PhD

AGRADECIMENTOS

Este projeto surgiu a partir de um desafio que me foi lançado, com perspetivas inovadoras e ambiciosas, o qual aceitei sem pensar duas vezes. Se voltasse atrás, repetiria a mesma escolha e a mesma experiência. Foi uma etapa marcante da minha vida, onde adquiri muitas coisas importantes que levo para o resto da minha vida. Este trabalho é o culminar de muitas horas de esforço e dedicação, bem como da resiliência de quem me orientou e guiou durante os anos em que estive no Centro de Investigação das Ferrugens do Cafeeiro (CIFC) em Oeiras, ao abrigo do projeto PATHOmics (PTDC/ASP-PLA/29189/2017, LISBOA-01-0145-FEDER-029189).

Primeiramente, um obrigado aos meus orientadores, Doutora Dora Batista e Doutor Pedro Talhinhas, que não me deixaram desistir e sempre me transmitiram uma palavra de confiança e de crença. A vós, um agradecimento do fundo do coração, porque nunca me esquecerei de quem acreditou em mim enquanto fui estudante, para fazer o meu caminho até me tornar na pessoa que sou hoje.

Agradeço também a toda a equipa do CIFC, local onde fui extremamente bem recebido e que vou levar sempre no coração. Quero assim agradecer de forma mais particular à Doutora Maria do Céu, Doutora Andreia Loureiro, Engenheiro Vitor Várzea, Doutora Ana Paula Pereira, Doutora Helena Azinheira, Doutora Leonor Guerra-Guimarães e não posso esquecer nem deixar de agradecer aos restantes funcionários do CIFC, pela disponibilidade, incentivo, sugestões, e pelo excelente ambiente de trabalho que me proporcionaram. Em especial, agradeço à Inês Diniz por ter tido paciência, disponibilidade e pelo apoio técnico prestado no desenvolvimento deste trabalho.

Aos amigos do curso de Biologia e da AgriculTUNA que perduram no tempo (sabeis quem sois), tornaram este percurso muito mais bonito com toda a amizade e companheirismo digno dos valores que o Instituto Superior de Agronomia representa. Obrigado por ainda estarem do meu lado nestes anos todos! Aos meus pais e irmã, que têm feito o vosso melhor para me apoiar no meu trajeto, por acreditarem nas minhas escolhas e deixarem-me voar, por serem aquele abraço e serem casa, obrigado! A ti Joana Filipa Matias, namorada, melhor amiga e companheira, que todos os dias me demonstras com atitudes o que é estar ao lado de alguém que tem um bom coração, és uma pessoa mesmo bonita e tornas a vida mais leve por seres quem és, obrigado do fundo do meu coração!

RESUMO

A ferrugem do cafeeiro é uma doença devastadora causada por *Hemileia vastatrix* Berkeley & Broome, levando a enormes perdas na produção de café. *Hemileia vastatrix* é um patógeno biotrófico obrigatório da família Pucciniales, sendo o ciclo de vida urediniospórico o único conhecido em campo. As estratégias de melhoramento do cafeeiro têm-se tornado cada vez mais uma alternativa viável para mitigar a ferrugem do cafeeiro, embora a capacidade adaptativa do fungo venha superando a durabilidade dos genótipos resistentes de cafeeiro. Existem mais de 55 raças/patótipos de *Hemileia vastatrix*, portanto é de grande importância compreender como a variação genética é promovida, bem como os mecanismos reguladores dos perfis diferenciados de virulência do patógeno. Os retrotransposões estão associados a variações estruturais do genoma e à regulação da expressão génica em agentes patogénicos de plantas. Estas caraterísticas podem estar associadas à sua capacidade de evoluir rapidamente em virulência. Neste trabalho, a estimativa do tamanho do genoma e a variabilidade genética, no que diz respeito à sequência de nucleótidos e ao número de cópias, de três retrotransposões selecionados (*R190*, *R1057* e *R2407*), bem como os perfis de expressão relativa ao longo do processo de infeção, foram investigados entre patótipos contrastantes de *Hemileia vastatrix* para avaliar a sua associação putativa com perfis de virulência. O tamanho do genoma medido (789 Mbp) e a sua variação intra-específica (166 Mbp) sugerem uma forte dinâmica genómica dos episódios de expansão/contração. Além disso, a análise das sequências de DNA dos retrotransposões e a construção de redes de haplótipos de junção mediana revelaram um elevado nível de polimorfismo e divergência, tanto entre isolados como dentro do mesmo isolado. A atividade aparentemente notável e a taxa de mutação dos retrotransposões estudados podem representar um mecanismo determinante para a geração de diversidade genética num agente patogénico supostamente assexuado como H. vastatrix. Além disso, a análise de qPCR de duas interações compatíveis (Hv1427; raça II [v5] e Hv70; raça XXIV [v2,4,5]), ao longo do processo de infeção (interação compatível com *Coffea arabica* var. Caturra 19/1 [SH5] e H152/3 [SH2,4,5], respetivamente).

Palavras-chave: cafeeiro; ferrugem alaranjada; variabilidade genética; elementos transponíveis; perfis de expressão.

ABSTRACT

Coffee leaf rust is a devastating disease caused by *Hemileia vastatrix* Berkeley & Broome, leading to huge losses in coffee production. *Hemileia vastatrix* is an obligate biotrophic pathogen from the Pucciniales family, with the urediniosporic life cycle as the only one known in field. Coffee plant breeding strategies have increasingly become a viable alternative for mitigating coffee rust, although the adaptative capacity of the fungus have been overcoming the duration of the coffee restistant genotipes. There are over 55 races/pathotypes of *Hemileia vastatrix*, therefore it is of great importance to comprehend how genetic variation is promoted, as well as the regulator mechanisms of the differentiated pathogen virulence profiles. Retrotransposons are associated with genome structural variations and gene expression regulation in plant pathogens. These features may be associated to their capacity to rapidly evolve virulence. In this work, the genome size estimation and the genetic variability, concerning nucleotide sequence and copy number, of three selected retrotransposons (*R190*, *R1057* and *R2407*), as well as the relative expression profiles along the course of the infection process, were investigated among *Hemileia vastatrix* contrasting pathotypes to assess their putative association with virulence profiles. The measured genome size (789 Mbp) and its intraspecific variation (166 Mbp) suggests a strong genomic dynamic of expansion/contraction episodes. Moreover, the analysis of the retrotransposons DNA sequences and the construction of *median-joining* haplotype networks revealed a high level of polymorphism and divergence, both between isolates and within the same isolate. The apparent remarkable activity and mutational rate of the retrotransposons studied may represent a determining mechanism for the generation of genetic diversity in a supposedly asexual pathogen such as *H. vastatrix*. Moreover, the qPCR analysis of two compatible interactions (Hv1427; race II [v5] and Hv70; race XXIV [v2,4,5]), throughout the infection process (compatible interaction with *Coffea arabica* var. Caturra 19/1 [SH5] and H152/3 [SH2,4,5], respectively).

Key-words: coffee plant; orange rust; genetic variability; transposable elements; expression profiles.

RESUMO ALARGADO

A ferrugem do cafeeiro é uma doença devastadora causada pelo fungo *Hemileia vastatrix* Berkeley & Broome, levando a epidemias recorrentes e enormes perdas na produção de café arábica a nível mundial. Atualmente, a doença encontra-se presente em todas as regiões cafeicultoras do mundo. O fungo *Hemileia vastatrix* é um agente patogénico biotrófico obrigatório da família Pucciniales, em que o ciclo de vida urediniospórico (assexual) é o único conhecido na Natureza. As estratégias de melhoramento genético do cafeeiro constituem a alternativa mais viável para mitigar a ferrugem do cafeeiro. No entanto, a durabilidade da resistência de genótipos melhorados de cafeeiro é constantemente desafiada pela elevada adaptibilidade do fungo. Nesse sentido, e considerando que existem mais de 55 raças/patótipos de *H. vastatrix*, é de grande importância compreender como a variação genética é promovida, bem como os mecanismos envolvidos na diferenciação dos perfis de virulência do patógeno, de forma a contribuir para o delineamento de medidas eficazes de controlo da doença. Estudos anteriores de sequenciação do genoma, transcriptómica e genómica populacional revelaram que o genoma de *H. vastatrix* apresenta um conteúdo muito elevado de elementos transponíveis (ca. 80%), bem como evidências de que os retrotransposões podem ter um papel adaptativo. Os retrotransposões, inseridos na categoria dos elementos repetitivos ou transponíveis do genoma, estão associados a variações estruturais do genoma e regulação da expressão génica em fitopatógenos. Estas características podem estar associadas à sua capacidade de evoluir rapidamente em termos de virulência e de diversidade genética.

O âmbito deste projecto centrou-se na investigação do papel putativo dos retrotransposões na diversidade genética e virulência da ferrugem do cafeeiro, através da análise da variação do tamanho do genoma entre múltiplos isolados de interesse, bem como da análise da variabilidade genómica de três retrotransposões selecionados, a nível da sequência nucleotídica e do número de cópias, e dos seus perfis de expressão relativa entre patótipos contrastantes. Numa primeira fase do trabalho, o tamanho do genoma de 21 isolados de *Hemileia vastatrix* foi estimado por citometria de fluxo, utilizando-se fungo existente em folhas infetadas, após inoculação artificial. As estimativas obtidas neste estudo colocam o tamanho médio do genoma de *H. vastatrix* em 789 Mbp. No entanto, foi possivel detetar adicionalmente diferenças significativas entre os isolados, traduzindo-se numa variação intra-específica de 166 Mbp, três vezes superior ao tamanho médio de genoma no reino Fungi (60 Mbp), o que sugere uma forte dinâmica genómica de episódios de expansão/contração. Para avaliar uma putativa relação causal com a actividade de retrotransposões, foram estudados três

retrotransposões LTR (*R190, R1057* e *R2407*), selecionados a partir de dados anotados de transcritoma, obtido em três fases de infeção/diferenciação de *H. vastatrix* e alinhados com loci de SNPs diagnóstico diferenciadores de linhagens genéticas de *H. vastatrix*. Com base na procura de homologia destas sequências contra uma base de dados específica de elementos transponíveis em Pucciniales (PucciDB), *R190* foi identificado como pertencendo à superfamília Copia (linhagem Mapi), enquanto *R1057* e *R2407* revelaram ser elementos não-autónomos TRIM e LARD, respetivamente. O número de cópias por genoma para cada retrotransposão foi estimado por PCR quantitativo em tempo real (qPCR). Os retroelementos *R190* e *R2407* revelaram-se os elementos mais proliferativos no genoma, com o número de cópias estimado a superar valores na ordem das 3000 cópias por genoma. Curiosamente, *R2407*, embora seja um elemento não-autónomo cuja mobilização depende da maquinaria de transposição de outros elementos autónomos, apresentou o maior número de cópias detetado neste estudo. No entanto, embora se tenha observado uma geral extensa actividade dos retrotransposões estudados no genoma de *H. vastatrix*, não se verificou qualquer correlação directa entre o tamanho do genoma e o número de cópias. A variabilidade genética das cópias de *R190, R1057* e *R2407* foi também investigada em 9-10 isolados de *H. vastatrix* com patótipos contrastantes por clonagem e sequenciação. A análise das sequências e a construção de redes de haplótipos *median-joining* para cada retrotransposão revelou um elevado nível de polimorfismo e divergência, quer entre isolados, quer dentro de um mesmo isolado. De facto, para cada colónia de *E. coli* sequenciada e correspondente ao mesmo isolado, obteve-se um haplótipo diferente, o que leva a presumir que a diversidade de haplótipos poderá ser bastante superior à encontrada. Em concordância com os dados obtidos para o número de cópias, as redes de haplótipos de *R190* e *R2407* apresentam um padrão de diversificação muito complexo, correlacionando um maior número de cópias com um maior nível mutacional, mas não aparente relação com isolado/patótipo. Em particular, o padrão de haplótipos observado para *R190* sugere uma rápida divergência e expansão do retroelemento, detectando-se diferenças de mais do que 30 SNPs e indels até 84 pb em cópias do mesmo isolado. A aparente notável actividade e taxa mutacional dos retrotransposões estudados poderá representar um mecanismo determinante para a geração de diversidade genética num patógenio supostamente assexuado como é *H. vastatrix*. Por fim, os perfis de expressão dos retroelementos em estudo foram analisados em dois isolados de *H. vastatrix* com patótipos distintos (Hv1427; raça II [v5] e Hv70; raça XXIV [v2,4,5]), ao longo do processo de infeção (interação compatível com *C. arabica* var. Caturra 19/1 [S_H5] e H152/3 [SH2,4,5], respetivamente), por qPCR como uma primeira abordagem para investigar o seu papel funcional e possivel relação com virulência.

Foram colhidas amostras de folhas inoculadas 1, 2, 3, 4, 6 e 8 dias após inoculação (dai) para Hv70, e 1, 2, 3, 4, 8 e 11 dai para Hv1427, correspondendo a pontos chave no processo de infeção. Para a normalização dos dados de qPCR, foram utilizados três genes de referência [proteína ribossomal 40S (*40S_Rib*), glyceraldehyde-3-phosphate dehydrogenase (*GADPH*), e subunidade III do complexo cytochrome c oxidase (*Cyt III*)] para *H. vastatrix*, previamente estabelecidos como os mais estáveis *in planta.* A análise da expressão relativa, realizada em comparação com o tempo 1 dai, revelou perfis de expressão bem distintos e significativos entre os dois isolados, mas similares entre os diferentes retrotransposões. Enquanto para Hv1427, todos os retrotransposões apresentaram sobreexpressão, em comparação com 1 dai, ao longo do processo de infeção, para Hv70 foi detetada uma geral diminuição de expressão dos retrotransposões, particularmente nas fases mais tardias da infeção. Estes resultados sugerem uma resposta específica dos retrotransposões à infeção em função do genótipo do isolado. Uma vez que a ativação da expressão de retroelemetos é normalmente espoletada em resposta a stresses e associada à regulação de genes fisicamente próximos, poderia avançar-se a hipótese de que os retrotransposões estudados terão um papel mais relevante na virulência de Hv1427 do que de Hv70. No isolado Hv1427, todos os retrotransposões são altamente expressos ao longo da infeção, em comparação com 1 dai, atingindo valores de fold change extremamente elevados no *R2407,* com um forte pico de activação aos 8 dai. Estes dados são extremamente promissores para uma possível contribuição destes elementos na diferenciação da virulência de *H. vastatrix*.

Este trabalho apresenta pela primeira vez dados detalhados sobre a variabilidade genómica e expressão de três LTR-retrotransposões em diferentes isolados de *H. vastatrix*, levantando a possibilidade do seu envolvimento na promoção de diversidade genética e na diferenciação de virulência específica. São necessários estudos adicionais para melhor esclarecer e compreender o papel dos retrotransposões na evolução de *H. vastatrix*. O aprofundamento destes temas será essencial para gerar conhecimento importante para o controlo da ferrugem.

Palavras-chave: *Coffea arabica*; ferrugem alaranjada; variabilidade genética; elementos transponíveis; perfis de expressão.

Table of Contents

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

AFLPs – Amplified Fragment Length Polymorphisms

CBD – Coffee Berry Disease

cDNA – complementary DNA

CIFC – Centro de Investigação das Ferrugens do Cafeeiro

CLR – Coffee Leaf Rust

Ct – threshold cycle

dai – days after inoculation

DNA – Deoxyribonucleic Acid

EST – Expressed Sequence Tag

ETI – Effector-Triggered Immunity

GADPH – glyceraldehyde 3-phosphate dehydrogenase

HDT – Híbrido de Timor (Timor hybrid)

HMC – Haustorial Mother Cell

HR – Hypersensitive Reaction

ICO – International Coffee Organization

IPTG – Isopropyl β-D-1-thiogalactopyranoside

ISA – Instituto Superior de Agronomia

ITS – Internal Transcribed Spacer

JA – Jasmonic Acid

LB – Luria-Bertani

LARD – Large retrotransposon derivatives

LTR – Long-Terminal Repeat

mRNA – messenger RNA

PAMPs – Pathogen-Associated Molecular Patterns

PCD – Programmed Cell Death

PCR – Polymerase Chain Reaction

PRR – Pattern Recognition Receptors

PTI – PAMP-Triggered Immunity

qPCR – quantitative real-time PCR

RAPDs – Random Amplified Polymorphic DNAs

RFLPs - Restriction Fragment Length Polymorphism

RIP – Repeat-Induced Point Mutation

RNA – Ribonucleic acid

ROS – Reactive Oxygen Species

SA – Salicylic Acid

SNPs – Single Nucleotide Polymorphisms

iRNA – interfering RNA

SOD – Superoxide Dismutase

TEs – Transposable Elements

TRIM – Terminal repeat retrotransposons in miniature

WFO – World Flora Online

1 INTRODUCTION

1.1 *Coffea* spp.

1.1.1 Historical context, taxonomy, and general characteristics

The Arabica coffee beans are considered the primordial source of coffee beverage for human consumption, dating back to 1000 B.C. by the Oromo tribe, from the Kingdom of Kaffa, Ethiopia (Bekele, 2010). The further migration from Ethiopia to Yemen and Arabia in the 7th century led to the current designation of Arabica coffee (World Coffee Research, 2024).

Coffea spp. belong to the Rubiaceae family (class Magnoliopsida, order Rubiales), whose fruits are typically coffee berries (Charrier and Berthaud, 1985). The first reports of the taxon *Coffea* date back to 1735-1737 by Carl Linnaeus, but only eighteen years after he described botanically the Arabic coffee tree, and named it *Coffea arabica* L. (Krug and Carvalho, 1951). However, earlier in 1713, Antoine de Jussieu had described a single coffee plant, from a botanic garden in Amsterdam, as *Jasminum arabicum* due to the flower's odour similar to jasmine (Charrier and Berthaud, 1985).

The *Coffea* L. genus is very diversified and consists of multiple types of woody perennial plants, going from small trees to large trees that can reach 5 meters in height (Bremer, 1996). Chevalier (1947) classified *Coffea*, dividing it into four sections (*Argocoffea*, *Paracoffea*, *Mascarocoffea*, and *Eucoffea*) and subdividing the section *Eucoffea* into three subsections (*Eucoffea*, *Malanocoffea*, and *Mozambicoffea*) (Davis, 2003). The two commercially traded coffee species, *Coffea arabica* and *Coffea canephora*, are included in the *Eucoffea* subsection (Carvalho *et al.*, 1969; Charrier and Berthaud, 1985). In 1961, Conagin and Mendes (1961) divided the genus *Coffea* into two groups, based on chromosome number: tetraploid species with 2n=44 chromosomes (*C. arabica*) and diploid species with 2n=22 chromosomes (*C. canephora*, *C. dewerei*, *C. congensis*, etc.). The latest classification method for *Coffea* divides it into two subgenera: *Eucoffea* (caffeine-containing species distributed along East, West and Central Africa) and *Mascaracoffea* (species with absence of caffeine native from Madagascar, the Mauritius and La Reunión archipelago, and Comores islands) (Charrier and Berthaud, 1985; Geletu, 2006). According to WFO (World Flora Online) Plant List, *Coffea* has 132 subordinate taxa (WFO, 2023). *Coffea arabica* is the only autogamous self-fertile species within the genus, being the rest self-incompatible (Krug and Carvalho, 1951; Fernandez *et al.*, 2004).

Every species of coffee is African native (Gichimu and Omondi, 2010), but only two (*Coffea arabica* and *Coffea canephora*) are used for large-scale production in tropical and subtropical countries in Africa, South & Central America, Asia and Oceania. *Coffea arabica* grows naturally at high altitudes in Ethiopia, south-east Sudan, and northern Kenya (Figure 1), ranging in altitude from 1200 to 2000 meters, with average annual temperature optimal for growth around 18-21°C. It is described as a low genetic diversity species (Prakash *et al.*, 2004). It is also the most important species for the coffee industry (Davis *et al.*, 2006), being firstly reported as a beverage product in the 14th century in Yemen, expanding among Middle Eastern countries in the 15th century, as well as India and Sri Lanka in the 17th century. The European trade market promoted the spreading of Arabica coffee to Asia and South America (Waller *et al.,* 2007).

Coffea canephora extends its yield areas from Central tropical and subtropical African equatorial forests, occurring at lower altitudes compared to Arabica coffee (Waller *et al.*, 2007). The two main areas of diversity of *C. canephora* are reported to be forests in the Gulf of Guinea countries and in the broader Congo basin area (Figure 1) (Charrier and Eskes, 2004). Commonly known as robusta coffee, this species only started to be commercially cultivated in the 20th century (Waller *et al.*, 2007).

The knowledge of the specific environments of coffee populations in each ecosystem is essential to develop *in vivo* collections and perform studies for agricultural and genetic traits of those populations (Charrier and Berthaud, 1985).

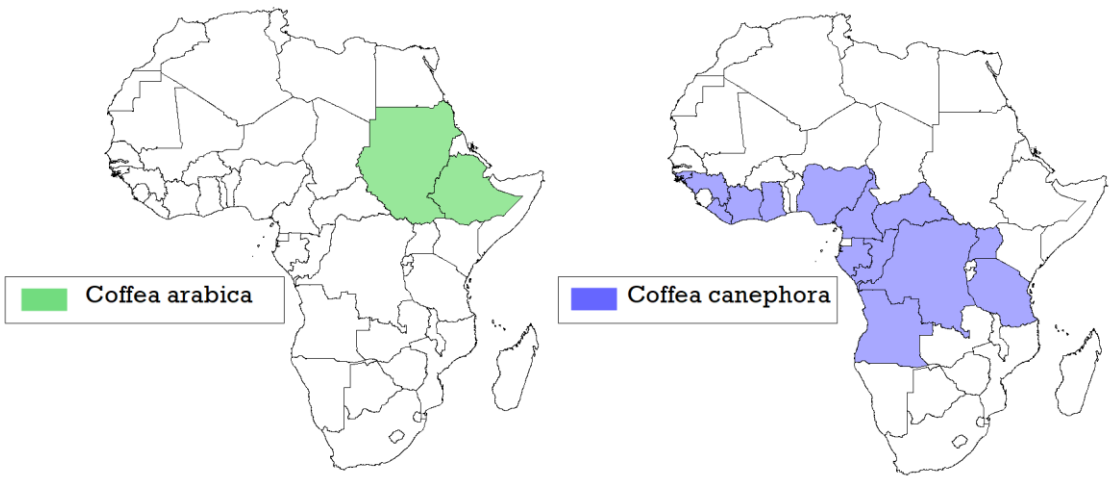


Figure 1 - *Coffea arabica* and *Coffea canephora* native growing areas (retrieved from https://colostate.pressbooks.pub/cropwildrelatives/chapter/case-study-coffee-wild-species-and-cultivars/)

1.1.2 Worldwide production and economic trademarks

Coffee is a major consumed product exported by more than 60 countries all over the tropics, being essential in the life of 100 million people who work in cultivating, harvesting, roasting and industrial processes (Waller *et al.,* 2007), making it a cash crop. According to ICO (International Coffee Organization) monthly reports, from 2017 to 2023, the world coffee consumption has been an average of ~10 million ton per year, being Brazil the biggest exporter with around 3 million ton produced (ICO, 2023). Its economic value is estimated to be around 70 billion US$ (Hoffman, 2018). This data shows the economic importance of this industry in the producing countries such as Brazil, India, Colombia, Vietnam, Thailand, Honduras, Philippines, Ethiopia, Kenya, Madagascar, Tanzania, etc. (ICO, 2023) and enhances the constant efforts to achieve better agricultural traits of coffee plant varieties (Etienne *et al.*, 2002). *Coffea arabica* represents a major part of the world coffee production, with a percentage value around 60%, as *C. canephora* completes the remaining 40% (Talhinhas *et al.*, 2017; Silva *et al.*, 2022; ICO, 2023).

Coffee-growing is mostly a small-holders' business, holding 10 million hectares among tropical producing countries (Waller *et al.*, 2007 and references therein). Natural disasters, civil and military armed conflicts, political instability and market economy brought market price fluctuations and socio-economic disturbances, affecting particularly the coffee growers (Waller *et al.*, 2007). Climate changes increased susceptibility to biotic stresses (pests and pathogens), and consequent episodes of abiotic stresses, causing major crop damages. So, it is essential to assess better crop practices and define strategies to control pests and diseases in the field (Davis *et al.*, 2012; Liebig *et al.*, 2016; Cerda *et al.*, 2017).

1.1.3 Coffee pests and diseases

Coffea spp. have climatic and ecologic requirements that need to be maintained within an acceptable range to avoid physiologic alterations and crop losses – the exposure of coffee leaves to constant temperatures, higher than 30°C, promotes the formation of chlorosis and stem tumours; under medium temperatures of 18°C occurs decreased fruit growth (Davis *et al.*, 2012 and references therein). Moreover, the occurrence of pests and diseases in the field are the major factors in crop losses. They can affect coffee plantations in two diverse ways: primary losses and secondary losses. Primary losses are the ones deriving from the plant injuries in that year; secondary losses are consequence of previous pests and diseases injuries (Zadoks and Schein, 1979; Cerda *et al.*, 2017). The complex interactions of perennial crops production with meteorological parameters along the development stages, and also the constant episodes of biotic

stresses, tend to complicate the quantification of crop losses. Quantifying yield losses and understanding its driven factors is fundamental to evaluate the efficiency of crop protection practices and applied sustainable systems for pests management, leading to better decisions in the future regarding pests and disease control in field (Cerda *et al.*, 2017 and references therein).

The current market pressure to increase worldwide coffee production, within the required standard quality levels, is in conflict with decreasing prices, plus market fluctuations. The farmers involved in coffee production have been facing reductions in their economic returns, being under financial conditions that jeopardize proper pest management with pesticides and/or fungicides. Table 1 catalogues the coffee pests and diseases that affect mostly the commercial coffee species (Liebig *et al.*, 2016).

Coffee Berry Disease (CBD) is an African endemic disease caused by *Colletotrichum kahawae* J.M. Waller & P.D. Bridge (phylum Ascomycota, class Sordariomycetes, order Glomerellales) (Silva *et al.*, 2006; van der Vossen and Walyaro, 2009; Loureiro *et al.*, 2012). In situations where the disease is not controlled, yield losses can reach up to 75%. CBD can be controlled with the application of fungicides, even though yield losses still reach 30%. *Colletotrichum kahawae* attacks green coffee berries, under humid conditions. The berries turn black, wrinkled, and decayed, with a hard skin – anthracnose – making them unusable for producers (Silva *et al.*, 2006; Hindorf and Omondi, 2011). CBD has been only reported in *Coffea arabica* in field conditions, as *Coffea canephora* and *Coffea liberica* only develop berry lesions if inoculated under specific conditions. Chemical control by fungicides and organic protectants is established as essential in the majority of producer countries, considered to be effective against CBD (Silva *et al.*, 2006 and references therein).

Coffee Leaf Rust (CLR), caused by the fungus *Hemileia vastatrix,* is the most important disease of *C. arabica*, affecting fruit quality and yield quantity. Nonetheless, *H. vastatrix* can infect all known cultivated *Coffea* species around the world even though at different levels of severity (McCook, 2006). In the next subsection, both the disease and the causal agent will be reviewed, being the pathogen *H. vastatrix* the main focus of this dissertation.

Table 1 - List of coffee pests and diseases and respective causal agents (adapted from Liebig *et al.*, 2016).

Coffee pests	Causal agent
Coffee Berry Borer	*Hypothenemus hampei* (Ferrari, 1867) (Order: Coleoptera)
White Coffee Stem Borer	*Monochamus leuconotus* (F.P. Pascoe, 1869) (Order: Coleoptera)
Green Scales	*Coccus* viridis (Green, 1889) (Order: Hemiptera)
Root Mealybug	*Geococcus coffeae* (Green, 1933) (Order: Hemiptera)
Coffee Leaf Miner	*Leucoptera coffeella* (Guérin-Méneville, 1842) (Order: Lepidoptera)
Antestia bugs	*Antestiopsis* spp. (Leston, 1952) (Order: Hemiptera)
Coffee diseases	**Causal agent**
Brown Eye Spot Disease	*Cercospora coffeicola* Berkeley & Cooke
Coffee Bark Diseases	*Fusarium stilboides* Wollenw
Coffee Wilt Disease	*Fusarium xylarioides* Steyaert
Coffee Berry Disease	*Colletotrichum kahawae* J.M. Waller & Bridge
Coffee Leaf Rust	*Hemileia vastatrix* Berkeley & Broome

1.2 *Hemileia vastatrix* and Coffee Leaf Rust

1.2.1 Disease description and economic impacts

Coffee leaf rust (CLR), also known as orange rust, is a devastating disease caused by *Hemileia vastatrix* Berkeley & Broome (phylum: Basidiomycota; order: Pucciniales), an obligate biotrophic fungus of *Coffea* spp. (Bettencourt and Rodrigues Jr, 1988; McCook, 2006), that promotes major yield losses to coffee crops in producer countries all around the globe (Bettencourt and Rodrigues Jr, 1988; McCook, 2006; Talhinhas *et al.*, 2017; Silva *et al.*, 2022). The first CLR epidemic outbreaks occurred in 1869 in Ceylon (now Sri-Lanka) where the most impacting coffee burst occurred, erasing all coffee cultivation (Silva *et al.*, 2006), and then in Southern India in 1870 (McCook, 2006). During the 19th century, it was also responsible for ruinous impacts in the regions of Java, Indian Union, Philippines and Madagascar (Noronha-Wagner and Bettencourt, 1967).

The 21th century has had a resurgence of epidemic outbreaks in the Americas, known as the 'Big Rust', affecting countries such as Colombia, Peru, Ecuador and the Mesoamerica & Caribbean Islands (Avelino *et al.*, 2015; McCook and Vandermeer, 2015). Recently, in 2020, CLR was first reported in Hawaii (Keith *et al.*, 2022; Keith *et al.*, 2023), the last major coffee producing region of the world that was free of CLR. Also, in November 2021 signs of CLR were reported in Florida (Urbina and Aime, 2024), and more recently, in Fyfa district, Saudi Arabia (Alhudaib and Ismail, 2024), the first documentation of CLR in these countries.

The disease impact in susceptible cultivars can cause up to 40% of production losses, when the meteorologic conditions are propitious to the disease development. The devastating impact of CLR has severe socioeconomic consequences for the agricultural sector, affecting millions of families involved in coffee production (Avelino *et al.*, 2015). There's an estimation of losses between 1 and 2 billion US$ per year in Arabica coffee production, due to CLR (Silva *et al.*, 2022; ICO, 2024).

Rust disease manifestations depend on the host-pathogen interactions - if the host is resistant, the most common reaction is the formation of small chloroses (flt, small chlorotic fleks associated with punctiform tumefactions) in the leaf abaxial surface; if the host is susceptible, it occurs the formation of orange uredosporic sori, leading to necroses, leaf abscission and dieback of branches. The most severe symptoms can be observed as full plant defoliation (Figure 2) (Silva *et al.*, 2006; Guzzo *et al.* 2009; Talhinhas *et al.*, 2017). Disease harshness can be dependent of several factors, such as humidity, host genotype and plant nutritional deficit (Avelino *et al.*, 2015).

Figure 2 - Coffee leaf rust symptoms and sign (retrieved from Talhinhas *et al.,* 2017): (A) Chlorotic spots and uredosporic sori on the lower leaf surface. (B) Severe defoliation in plants as consequence of disease, contrasting with resistant plants in the background.

1.2.2 Taxonomy and Phylogeny

Hemileia genus belongs to the phylum Basidiomycota, class Pucciniomycetes, order Pucciniales. It is constituted by 42 species from tropical and subtropical regions of Africa and Asia. *Hemileia* can be distinguished from other rust fungi genera for having unique morphological features: suprastomatal bouquet-shaped sori; ovoid to reniform uredospores with a smooth ventral side and a delicately to coarsely echinulated convex dorsal side; and angular-globose to very irregular teliospores (Ritschel, 2005). *Hemileia* genus is phylogenetically settled at the base of the *Mikronegeriineae* suborder in the *Zaghouaniaceae* family, supported by molecular and phylogenetic analyses (Aime and McTaggart, 2021). The *Mikronegeriineae* suborder represents the most ancestral lineages of rust fungi, that diverged at c. 98 million years ago. It is assumed that ancestral rust species evolved as tropical short-cycle species, having angiosperms as their hosts, defying the previous theory where ancestral rusts are harboured by phylogenetically ancestral hosts (McTaggart *et al.*, 2016).

1.2.3 Life cycle and Infection process

Hemileia vastatrix is an obligate biotroph, depending on the living host to complete the life cycle (Heath *et al.*, 1997). It is a hemicyclic fungus that produces uredospores, teliospores and basidiospores. In the field, only the asexual phase of the life cycle occurs, producing uredospores that infect the leaves and cause the disease (Silva *et al.*, 2022). While, pycniospores and aeciospores remain unnoted (Talhinhas *et al.*, 2017). Uredospores are dikaryotic thin-walled spores majorly propagated through rain, with ability to infect *de novo* when the environmental conditions are propitious (Bock, 1962; Chinnappa and Sreenivasan, 1968). Teliospores are occasionally generated *in situ*, inducing the production of basidiospores from a 4-celled hyphal filament called epibasidium. Basidiospores, representing the sexual phase, cannot infect coffee, even though no other host has been found (Kushalappa *et al.*, 1989). Therefore, uredospores are the only functional propagules known and the only reported means of propagation for *H. vastatrix*.

For a successful infection process, uredospores should be able to adhere to the leaf surface and germinate – the presence of water is crucial for germination, having maximum efficiency at 24°C. The adherence of uredospores to the leaf surface is a key step in the cycle, as it promotes the appressoria differentiation and functionality, avoiding fungal dislocation in leaf (Braun *et al.*, 1994). After germination, uredospores form an appressorium, penetrating through the stoma and forming a penetration hypha (Figure 3A-C).

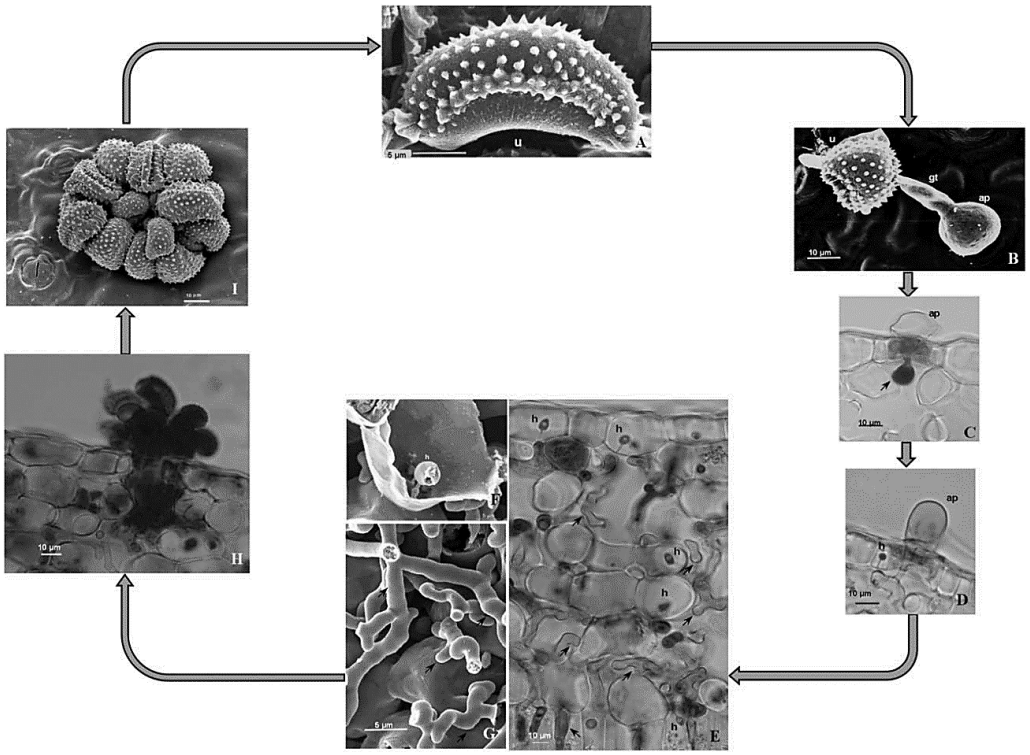


Figure 3 - *Hemileia vastatrix* infection process. (A) Urediniospore (u), scanning electron microscopy (SEM). (B) Germinated urediniospore (u) with germ tube (gt) and appressorium (ap) over stomata on the lower surface of the coffee leaf, 17 h after inoculation (hai), SEM. (C) Appressorium (ap) over stomata and penetration hypha (arrow), 24 hai, light microscopy (LM). (D) Appressorium (ap) over stomata and intercellular hypha with an haustorium (h) within a subsidiary cell, 48 hai, LM. (E) Intercellular hyphae (arrows) and haustoria (h) within epidermal and mesophyll cells, 20 days after inoculation (dai), LM. (F) Haustorium (h) within a spongy parenchyma cell, 20 dai, SEM. (G) Intercellular hyphae (arrows) in the spongy parenchyma, 20 dai, SEM. (H) Urediniosporic sorum protruding through the stomata in a bouquet shape, 21 dai, LM. (I) Urediniosporic sorum, 21 dai, SEM. (retrieved from Talhinhas *et al.* 2017).

After that, the hypha forms two specialized lateral branches that anchor in the chamber. Those branches will be differentiated in haustorial mother cell (HMC) and consequently, each HMC forms a haustorium which will infect the stomatal subsidiary cells (Figure 3D). The development of more hyphae will be continuous as HMCs generate more haustoria in parenchyma, as well as in upper epidermis (Figure 3 E-G). From this stage on, lesions on leaf surface will be macroscopically visible, depending on hyphae development and external conditions. The hyphae invade the substomatal cavities promoting protosori

production (Kushalappa *et al.*, 1989; Talhinhas *et al.*, 2017). Within three weeks since the beginning of infection, sporogenous cells are produced from the protosori, around one to seven cells per protosorus, with the ability to emerge through the stomatal opening in a bouquet shape. This bouquet will form an orange spore mass, usually recognized as a typical sign of CLR (Figure 3H) (McCain *et al.*, 1984; Talhinhas *et al.*, 2017).

1.2.4 Disease control

The most used method to constrain CLR in the field is the employment of fungicides. However, the chemical approach not only does not eradicate the disease, but it also entails enormous costs to farmers. The application of phytochemical products is not environmentally beneficial, and the applying techniques require machinery not accessible to small farmers (Talhinhas *et al.*, 2017). Therefore, breeding is the main sustainable approach to overcome CLR, aiming to obtain resistant coffee cultivars. In collaboration with many universities and research centers around the globe (China, Brazil, France, Colombia, Kenya, Ethiopia, Mexico, etc.), the pioneering work carried out in Centro de Investigação das Ferrugens do Cafeeiro (CIFC, Oeiras) led to the development of resistant coffee genotypes, which supported several different breeding programs and originated over 90% of the resistant coffee cultivars worldwide (Rodrigues Jr *et al.*, 1975; Bettencourt and Rodrigues Jr, 1988; Talhinhas *et al.*, 2017). The most notorious source of resistance used in breeding was Timor hybrid (Híbrido de Timor - HDT), a spontaneous hybrid between *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta), found in Timor Leste in the 1940's, being resistant to all known *H. vastatrix* races at the time (Silva *et al.*, 2006). HDT has been an important resource in providing resistance genes through cross-breeding, generating reliable hybrids such as Catimor and Sarchimor (Talhinhas *et al.*, 2017; Silva *et al.*, 2022). Since the 1920's, coffee protection & development programs, such as the ones in Balehonnur Coffee Research Station (India Union) and CIFC, had a breeding goal to obtain resistant/tolerant varieties (Noronha-Wagner and Bettencourt, 1967). Although the use of resistant coffee varieties has been successful, breakdown of resistance in many of these improved varieties in several countries has been observed (Silva *et al.*, 2022).

1.2.5 Fungal variability and populational structure

Hemileia vastatrix (Hv) physiological races are associated to isolates with specific and unique combination of virulence genes, as inferred by Flor's gene-to-gene theory, organised by sequential Roman numerals in order of detection (Noronha-Wagner and Bettencourt, 1967). They were first described in 1932, by W. Wilson Mayne, who reported evidence of physiological specialization in *H. vastatrix*, identifying four Hv races (Mayne, 1932). Ensuing the foundation of CIFC by Professor Branquinho d'Oliveira in

1955, many contrasting Hv pathotypes have been correlated with the different coffee resistance gene combinations, as studies still progress to this day on (Talhinhas *et al.*, 2017; Silva *et al.*, 2022). Currently, there are more than 55 *H. vastatrix* races described (Rodrigues Jr *et al.*, 1975; Várzea and Marques, 2005; Talhinhas *et al.*, 2017; Silva *et al.*, 2022). These races have been identified according to the virulence or avirulence reactions they induce when inoculated on a set of 27 coffee differentials: clonal lines of five *C. arabica* selections, 16 tetraploid hybrids of *C. arabica* × *Coffea* spp., and six *Coffea* spp. selections (Bettencourt and Rodrigues Jr, 1988; Várzea and Marques, 2005; Talhinhas *et al.*, 2017). Through the application of standard molecular methods, such as random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs) and sequences of Internal Transcribed Spacer (ITS), *H*. *vastatrix* populations have been studied to assess genetic diversity within and between populations, and population genetic structure. Globally, no correlations were established between pathogen genetic structure and geographical origin, host, or physiological races of the fungal isolates under study (Gouveia *et al.*, 2005; Nunes *et al.*, 2009; Batista *et al.*, 2010; Rozo *et al.*, 2012; Maia *et al.*, 2013; Cabral *et al.*, 2016; Kosaraju *et al.*, 2017; Quispe-Apaza *et al.*, 2017; Santana *et al.*, 2018; Quispe-Apaza *et al.*, 2021; Bekele *et al.*, 2022). For many years, the characterization of Hv's population genetic diversity has been a discrepant subject, with some studies reporting low differentiation levels, where others report high differentiation levels of genetic variability within populations. Another debatable subject is whether the fungus is undergoing sexual reproduction, although population studies have frequently reinforced the interpretation of clonal reproduction in the sampled populations (Ramirez-Camejo *et al.*, 2022), suggesting that other mechanisms may play an important role in promoting genetic diversity through mutations, asexual recombination, or other sources of genetic variation. Commonly, no research project studying *H. vastatrix* populations was able to find populational structure related with *H. vastatrix* race, host type or geographical origin (Talhinhas *et al.*, 2017). Only recently, in 2018, Silva *et al.* allowed a better understanding of the complex phylogenetic relationships of *Hemileia vastatrix* populations. Using restriction site-associated DNA sequencing (RADseq), they were able to find structured Hv populations and ascertain genetic diversity and worldwide population dynamics. The phylogenetic analysis revealed three well-structured divergent lineages (C1-C3), where C1 and C2 grouped rust isolates that infect diploid coffee species (*C*. *canephora*, *C. excelsa*, *C. liberica* and *C. racemosa*), while C3 comprise isolates that infect tetraploid coffee species (*C*. *arabica* and interspecific hybrids, like HDT derivatives). Signals of introgression were found within rust isolates among the C2 and C3 groups. These evidence is well marked by heterozygotic shared alleles in single nucleotide polymorphisms (SNPs) and suggest

an evolution of C3 rusts shaped by hybridization/recombination between diverging lineages (Silva *et al.*, 2018).

Furthermore, in 2022, new data obtained through phylogenetic and clustering analyses of the CIFC rust collection allowed the establishment of a population substructure, within the C3 group, with three differentiated subgroups (SGI, SGII and SGIII), likely following a pattern of local adaptation to coffee hosts (Rodrigues *et al.*, 2022). Rust isolates from SGI are worldwide spread, whereas SGII are mainly African native and SGIII isolates are only found in East Timor. Such structuring seems linked with historical distribution and commercial exchanges of coffee cultivars, as well as the extent of breeding strategies (Rodrigues *et al.*, 2022). Isolates from subgroup SGI present low levels of differentiation, whilst very distant geographically, suggesting rapid evolution and population expansion through exchange of infected coffee material. Isolates from SGII reveal high levels of intraspecific differentiation, possibly associated with wild rust populations as a source of genetic variability, through the inadvertent usage of infected wild germplasm within the range of native populations of *C. arabica*, or naturally dispersing into coffee farms. The SGIII isolates also presents high levels of intraspecific genetic diversity, being constituted by the most virulent rust isolates so far, suggesting adaptation to new host genotypes (Rodrigues *et al.*, 2022), with Timor being the cradle of the tetraploid hybrid Híbrido de Timor (Talhinhas *et al.*, 2017). Moreover, the differentiation between the lineages can be explained by a small set of putatively adaptive loci (SNPs) under the effect of natural selection. These results support the idea that *H. vastatrix*'s evolutionary dynamics are strongly shaped by coffee hosts (Várzea and Marques, 2005; Silva *et al.*, 2018), relying on the genetic variation within a small portion of the genome (29 SNP *loci*), where the associated genes/loci may be strongly involved in the adaptation to new coffee hosts and/or new environments (Rodrigues *et al.*, 2022). Interestingly, Rodrigues *et al.* (2022) found that more than 90% of the few annotated loci are orthologs to genes encoding predicted proteins of the transposable elements structure, namely retrotransposons.

Also, fungi from three rust races (II, VI and XXIV) were analysed and compared at a proteomic level, attempting the establishment *of H. vastatrix* uredospores core proteome. (Guerra-Guimarães *et al.*, 2021). Proteins were functionally annotated as being involved mainly in DNA integration, RNA-dependent DNA biosynthesis, proteolysis, translation, oxidation-reduction process, primary metabolism, nitrogen compound metabolism and macromolecule metabolic processes. Resulting from the comparative analysis, the protein factors that differentiate races and may be involved in pathogenicity are under analysis, representing 5% of the total functionally annotated proteins (94 of 1874) (Silva *et al.*, 2022).

1.3 *Coffea-Hemileia* pathosystem

1.3.1 Pathosystem interactions

Coffea-Hemileia vastatrix interactions follow Flor's gene-for-gene model (Flor, 1942). Noronha-Wagner and Bettencourt (1967) reported the existence of *Hemileia* susceptibility (S_H) genes in coffee that corresponded to the virulence (*v*) genes in *H. vastatrix*. Those genes are the major factors that condition the interactions during infection, either singly or in combination (Rodrigues Jr *et al.*, 1975; Bettencourt *et al.*, 1988). There are S_H1 – S_H9 dominant coffee genes identified, corresponding to v_1-v_9 genes in the pathogen (Bettencourt and Rodrigues Jr, 1988). Nevertheless, there are recent *H. vastatrix* races that infect all HDT genotypes, known for its resistance to all *H. vastatrix* races with virulence genes within v_1-v_9, making those races undefined (Talhinhas *et al.*, 2017). The *H. vastatrix*-coffee pathosystem is based on a dynamic co-evolution, in which new pathotypes with increased virulence profiles have been emerging, capable of overcoming the resistance mechanisms of the host. The continuous resistance breakdown observed in the field has been providing evidence of the high pathogenic variability in *H. vastatrix* (Talhinhas *et al.*, 2017). The genetic diversity and characterization of the virulence genes is still unsolved, due to the absence of *H. vastatrix*'s sexual phase and the lack of correlation between phenotypic and molecular diversity. Moreover, the complex and massive genome size of *H. vastatrix* and the recent discovery of an abundant existence of replicant diploid nuclei (Talhinhas *et al.*, 2023), with high levels of variability among isolates (Tavares *et al.*, 2014), creates difficulties for the development of molecular markers for virulence diagnostic purposes.

1.3.2 Virulence mechanisms and involvement of effectors in pathogenicity

Host resistance to *Hemileia vastatrix* is a critical trait in coffee plant breeding and management. It is often achieved through defense mechanisms that enable coffee plants to resist or tolerate infection (Bettgenhaeuser *et al.*, 2014; Li *et al.*, 2020). *Coffea-H. vastatrix* interactions are mainly driven by effector proteins, or effectors, which are secreted proteins that go from the pathogen to the host cells, in order to trigger the infection process. Effectors can supress or directly manipulate host cellular processes, besides contributing to the creation of a suitable environment for pathogen colonization and proliferation (Lovelace *et al.*, 2023). In response to the fungal effectors, the coffee host plant cells responds through the effector-triggered immunity (ETI) mechanism (De Wit *et al.*, 2009). The ETI is one of the mechanisms of plant immune responses, the other being pathogen-associated molecular pattern-triggered immunity (PTI). Coffee plant cells have plasma membrane receptors to recognize pathogen invasion through the recognition of conserved microbe-associated or pathogen-associated molecular patterns

(MAMPs/PAMPs) and host damage-associated molecular patterns (DAMPs) by pattern recognition receptors (PRRs) that activate PTI (Delplace *et al.*, 2022). Both immune responses promote the accumulation of reactive oxygen species (ROS), like hydrogen peroxide (H_2O_2), changes in cellular ion fluxes, activation of protein kinase cascades, production of stress-related hormones, the reinforcement of cell walls and changes in protein and gene expression (Silva *et al.*, 2002, 2008; Jones and Dangl, 2006; De Wit, 2007; De Wit *et al.*, 2009). In ETI, the plant's immune system has evolved the ability to recognize specific effector proteins as foreign or detrimental. This recognition occurs through the plant's intracellular receptors, known as resistance (R) proteins. Upon recognition of the pathogen's effector, the R proteins trigger a signaling cascade that activates a strong and rapid immune response (Martin, 1999; Fernandez *et al.*, 2004). This response involves the initiation of a programmed cell death (PCD) response known as the hypersensitive response (HR), arresting fungal growth of the biotrophic pathogen (Jones and Dangl, 2006). *Hemileia vastatrix* effectors (Loureiro *et al.* 2014) also have the ability to interfere with host's hormone signaling pathways (Fernandez *et al.*, 2012; Talhinhas *et al.*, 2014). In particular, they may influence the levels of plant hormones, such as salicylic acid (SA), jasmonic acid (JA), and ethylene, which are key players in plant defense responses (Ganesh *et al.*, 2006).

Full knowledge about the fungal effectors may provide valuable information for breeding programs, allowing plant resistance genes selection for host varieties of agronomic interest (Vleeshouwers and Oliver, 2014). A 454-pyrosequencing analysis of *C. arabica* leaves infected by *H. vastatrix* at late stages of infection (21 days) revealed the first Hv contigs with homology to other rust fungi secreted proteins, showing conserved amino-acid domains and conserved patterns of cysteine positions which suggests conserved function during the infection process (Fernandez *et al.*, 2012). Following this work, a 454-pyrosequencing transcriptome analysis of *H. vastatrix* germinating uredospores and appressoria was performed against published *in planta* haustoria-rich data. From the resulting 9234 annotated transcripts, 516 transcripts were predicted to encode secreted proteins potentially involved in appressoria-mediated penetration. This candidate genes are mostly transcribed in pre-penetration fungal structures (Talhinhas *et al.*, 2014).

Evolutionarily, pathogens move towards genetic alterations in effectors to avoid host immunity mechanisms. The effector's coding genes rely within, or proximal, to chromosomic regions with great abundance of transposable elements (TEs) (Rovenich *et al.* 2014; Cook *et al.* 2015). In *Magnaporthe oryzae*, the host-pathogen interaction occurs through recognition of *Avr-Pita* effector by the Pita receptor in rice plant (Orbach *et al.*, 2000). Induction of mutations and/or deletions in *Avr-Pita* effectors, located in a

subtelomeric region flanked by TEs, has shown to benefit the pathogen overcoming the host defensive mechanisms (Orbach *et al.*, 2000; Chang *et al.*, 2008; Dai *et al.*, 2010). Moreover, a TE insertion in the promotor region of *Avr-Pita* genes leads to an increase of pathogenic virulence (Kang *et al.*, 2001).

The ability of *Hemileia vastatrix* to overcome host resistance, linked with its large genome (Tavares *et al.*, 2014) and the predominance of SNPs involved in host adaptation located in retrotransposons (Rodrigues et al, 2022), may lead further studies to focus on genome transposable elements, as they can regulate gene expression (Zemojtel *et al.*, 2009), induce mutations and epigenetic changes (Lorrain *et al.*, 2021), and are in massive number in the *H. vastatrix* genome (Cristancho *et al.*, 2014; Porto *et al.*, 2019; Ángel *et al.*, 2023).

1.4 Retrotransposons in fungi genomes

1.4.1 What are retrotransposons?

Rust fungi (Pucciniales) genomes are constituted by a large percentage of repetitive elements, such as transposable elements (TEs), contributing to their large genome sizes (Porto *et al.*, 2019; Orozco-Arias *et al.*, 2022; Ángel *et al.*, 2023). The transposable elements are divided into two major classes due to various characteristics, particularly the mechanism of transposition, namely retrotransposons (Class I) and transposons (Class II). Transposons are DNA-intermediated, and retrotransposons are RNA-derived (Boeke *et al.*, 1985; Feschotte and Pritham, 2007). Retrotransposons transpose by a replicative “copy and paste” mechanism, whereas transposons mainly move by a conservative “cut and paste" mechanism, even though there are exceptions (Daboussi and Capy, 2003; Muñoz-López and García-Pérez, 2010). These genetic elements have the ability to move around and duplicate themselves within a genome, potentially causing mutations and contributing to genome evolution (Orozco-Arias *et al.*, 2019). Retrotransposons replicate and move within a genome via a process called reverse transcription. Also referred as "jumping genes", they can jump to different locations in the genome, increasing their copy number. This can have both positive and negative effects or, quite frequently, null effects on the host organism (Muszewska *et al.*, 2011; Zhang *et al.*, 2014).

There are two principal types of retrotransposons: the LTR-retrotransposons and the non-LTR retrotransposons. LTR-retrotransposons are constituted by one long-terminal repeat (LTR) region at each 5’ and 3’ ends, a *gag* gene, and a *pol* open reading frame (ORF). The ORF region has several genes such as aspartic protease (AP), reverse transcriptase (RT), DDE integrase (INT) and RNAse H (RH) (Figure 4). The organization

of these genes within the pol region will distinguish between the two retrotransposon superfamilies: *Ty1/Copia* (*Pseudoviridae*) and *Ty3/Gypsy* (*Metaviridae*) (Muszewska *et al.*, 2011).

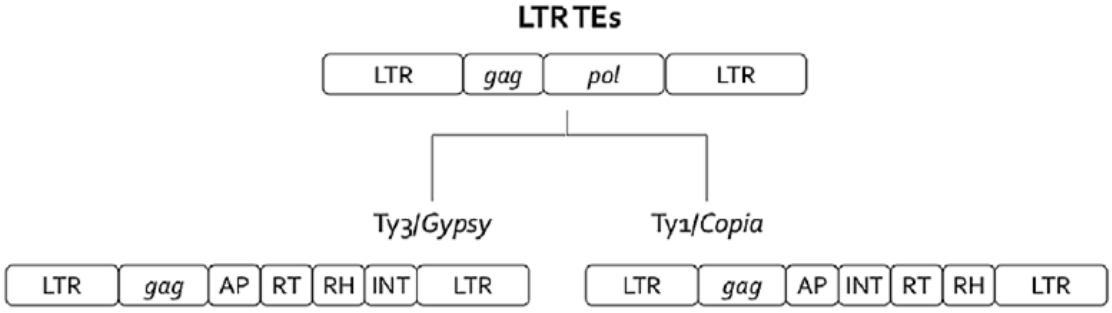


Figure 4 - A schematic representation of LTR transposable elements present in fungal genomes (retrieved and adapted from Muszewska *et al.*, 2011). The following domains are shown: LTR (Long Terminal Repeat), RT (Reverse transcriptase), *gag* (Group specific antigen), AP (Aspartic protease), RH (RNAse H), INT (Integrase) and Chr (Chromodomain).

There are also other groups of long terminal retrotransposons, such as Terminal-repeat retrotransposons in miniature (TRIM) and large retrotransposon derivatives (LARD), with important roles in the genome. TRIM are non-autonomous elements that require mobility-related proteins, encoded by other retrotransposons to transpose, and seem to be involved actively in the restructuring of genomes, affecting the promoter, coding region and intron-exon structure of genes (Witte *et al.,* 2001; Koziol *et al.* 2015). LARD are also non-autonomous retroelements, with both sequence and structure conserved, and are associated to a complementation mechanism wich encode the protein products necessary for replication and integration (Kalendar *et al.,* 2004).

1.4.2 How do they work and what is their role in the genome?

Retrotransposons use reverse transcriptase transposition mechanisms, with a RNA intermediate, to produce a cDNA copy which will be inserted into the host genome. The copy insertion is transcribed via host RNA Polymerase II, leading to the production of a polyadenylated mRNA (Fink *et al.*, 1986; Boeke and Corces, 1989). This mRNA is then translated and originates structural proteins of the *gag* gene and the *pol* enzymatic proteins (McHale *et al.*, 1992). Each transposition event of a retrotransposon produces an additional genomic copy, which explains the contribution of retrotransposons to the increase in repetitive elements in many fungal genomes, especially in rust fungi (Duplessis *et al.*, 2011; Lorrain *et al.*, 2021).

Retrotransposons expansion in fungal genomes have shown to be host-specific, as they are responsive to variable environmental conditions. The INT-RH-RT domain in the *pol* region is genetically more stable than the AP gene, which is much more variable among

diverse LTR-retrotransposons (Muszewska *et al.*, 2011). Comparative studies for conserved regions in genes coding RT, among different retrotransposons, unveil the establishment of dynamic relationships. There is evidence that allows inferring on vertical and horizontal transmission as major factors in the evolution of fungal pathogen retroelements (Hua-Van *et al.*, 2001; Daboussi *et al.*, 2002; Santana *et al.*, 2013; Alzohairy *et al.*, 2014). Retrotransposons can invade new populations and provide sequence dispersal throughout lateral gene transfer. As the insertion of repetitive elements near the promotor region of genes can have partial or total effect, generating a null phenotype by altering/blocking gene transcription, they are considered regulators of gene expression (Daboussi, 1996, 1997; Chadha and Sharma, 2014; Castanera *et al.*, 2016).

1.4.3 Association to pathogen virulence

The proliferation of transposable elements, such as transposons and retrotransposons, have been associated not only with genome expansion, but also with regulation of gene expression (Castanera *et al.,* 2016) and evolution of avirulence effectors (Zhong *et al.,* 2017). For *Austropuccinia psidii*, the methylation of TEs is easily observed, as older TEs present a lower GC content due to a heavy methylation that occurs during the plant-pathogen interactions (Tobias *et al.,* 2021). Retrotransposons were found to be associated with higher nucleotide substitution levels, shaping genome evolution, proliferating and inserting themselves into genomic regions that regulate gene expression, such as promoter regions. This is particularly relevant, as retrotransposons appear to be regulated by stress conditions during plant infection (Muszewska *et al.,* 2019).

Many pathosystems, especially when obligate biotrophic fungi are involved, promote coevolution between host and pathogen's genomes. The coevolution process is guided by biotic interactions, triggering stress events. In fungi genomes, stress events are the key to transcriptionally express most of the inactive TEs copies (Fouché *et al.*, 2020). Various epigenetic mechanisms, such as DNA methylation, histone modifications, and small RNA silencing, can regulate gene expression, also affecting regions involving TEs genes (Huda *et al.* 2010; Freitag, 2017; Borgognone *et al.*, 2018; Lorrain *et al.*, 2021). TEs insertions can affect histone modifications marks. As so, altering histone marks can enable pathogens to adapt under environmental stress episodes, improve pathogenicity levels and regulate secondary metabolites (Soyer *et al.*, 2014; Krishnan *et al.*, 2018; Fouché *et al.*, 2020). DNA methylation is a process mediated by methyltransferases (Lyko, 2018), in which duplicated DNA sequences have cytosines modified to 5-methylcytosines, by adding a methyl group (Slotkin and Martienssen, 2007). This

epigenetic mechanism can affect gene expression, without altering the DNA sequence. Therefore, TEs gene expression can be affected, being this mechanism a regulator of TEs activity (Bewick *et al.*, 2019). Small RNA silencing is another epigenetic regulator mechanism, mediated by interfering RNAs (RNAi) (Luo and Lu, 2017). The process has been associated with TEs regulation, although the molecular evidence is limited. Lorrain *et al.* (2021) proposed further approaches into small RNA-mediated silencing in rust fungi genomes to study TEs regulation, due to their absence of Repeat-Induced Point Mutation (RIP).

1.4.4 First approach to retrotransposons in *Hemileia vastatrix*

The complexity of the genomes of rust species (Pucciniales) has been limiting the sequencing and assembly processes, with only three draft genome assemblies for *H. vastatrix*. The first sequencing approach resulted in an hybrid assembly of 201 Mb (Cristancho *et al.*, 2014), the second one generated a quite fragmented assembly with 543 Mb (Porto *et al.*, 2019) and the third one assembled a highly contiguous genome with 747.98 Mb, where transposable elements covered 81.66% of the obtained genomic scaffolds (Ángel *et al.*, 2023). The prospect of a dikaryotic phased chromosome-level genome assembly for *H. vastatrix* under development (Tobias *et al.,* 2022) offers further chances of better characterizing the genome TE content in a near future.

Orozco-Arias *et al.* (2022) reported the first study to characterise retrotransposon lineages in *Hemileia vastatrix*. Using the available genomic data to date, they aimed to identify, annotate, and understand the contribution of transposable elements (TEs) to the genome size of *H. vastatrix*. In this study, complementary *de novo* approaches (structure-based and homology-based methodologies) were used to construct a first reference library of transposable elements for *H. vastatrix*, as the available genomes to date only represented 25% (Cristancho *et al.*, 2014) and 67% (Porto *et al.*, 2019) of the estimated genome size for *H. vastatrix*. Their results showed that 85.6% of the TEs were mainly LTR-retrotransposons. Based on phylogenetic analyses of the reverse transcriptase (RT) domains, the LTR-retrotransposons were classified into superfamilies (*Copia* and *Gypsy*) and corresponding lineages. Two new *Copia* lineages (Labe and Mapi), and three new *Gypsy* lineages (Soroa, Baco, CO-HUI) were identified. The *Gypsy* lineages CO-HUI and Soroa represent the majority of all *H. vastatrix* TEs and about 37% of the available genome assemblies. Soroa is the lineage with the highest number of insertions in coding regions of the genome (exons and genes), while Mapi appears to be the most involved lineage in co-expression with genes through gene-to-gene interaction (Orozco-Arias *et al.*, 2022).

1.5 Main questions & Objectives

In this context, this study intended to understand the role of retrotransposons in genetic diversity and virulence of *Hemileia vastatrix*, when interacting with its coffee host, focusing on addressing the following questions:

a) Are retrotransposons variable (in sequence and copy number) among the *H. vastatrix* isolates available in CIFC collection?
b) Do those isolates differ in genome size, in a way relatable with retrotransposon diversity?
c) Are these retrotransposons differentially expressed along the course of infection? And are those expression profiles related with virulence profiles?

To answer these questions, the following objectives were established:

a) Assess a previously obtained collection of annotated *H. vastatrix* transcripts (Talhinhas *et al.*, 2014) with retrotransposon homology by in silico analysis, based on BLAST alignments against *H. vastatrix* genomic data (Cristancho *et al.*, 2014; Porto *et al.*, 2019) and diagnostic SNPs data (Silva *et al.*, 2018) to select candidate retrotransposons
b) Analyse candidate retrotransposons genetic diversity through amplification and sequencing, to potentially detect polymorphisms among contrasting pathotypes; Estimate copy number variation through real-time quantitative PCR analysis.
c) Measure genome size for selected *H. vastatrix* isolates, using flow cytometry, by comparison to appropriate DNA standards.
d) Assess expression profiles for the retrotransposons under study, using real-time quantitative PCR, along different time-points during the infection process, matching different life cycle stages of the pathogen, and between different pathotypes, to characterize retrotransposon activity and identify putative causal relations.

2 MATERIALS AND METHODS

2.1 Genome size estimation

2.1.1 Fungal material

The fungal material used in this study is described in Table 2. *Hemileia vastatrix* isolates were obtained from the collection maintained at CIFC/ISA, ULisboa, and inoculated on compatible coffee genotypes for multiplication as part of the routine tasks performed at CIFC. The inoculation was performed by spreading fresh uredospores (1 mg/pair of leaves) on the lower surface of young coffee leaves with a soft camel brush and sprayed with distilled water. After an incubation period of 24 hours in a dark moist chamber, the plants were moved to greenhouse conditions (d'Oliveira, 1954). After sporulation and upon collecting the uredospores, coffee leaves (two pairs of leaves from two different plants for each genotype) were collected and taken to the laboratory in order to perform flow cytometry analyses (Tavares *et al.*, 2014).

Table 2 – *Hemileia vastatrix* isolates used in this study, detailing fungal isolates, virulence profiles, geographical origin, and year of reception at CIFC, as well as the phylogenetic group according to Rodrigues *et al.* (2022).

***Hemileia vastatrix* isolate**	**Virulence genes/Race**	**Origin**	**Year**	**Host species**	**Phylogenetic group**
22	v2,5/I	Tanzania	1953	*C. arabica*	C3_SGI
70*****	v2,4,5/XXIV	Kenya	1954	*C. arabica*	C3_SGI
71*****	v?/IV	Mozambique	1954	*C. racemosa*	C2
92*****	v?/XVIII	São Tomé	1956	*C. liberica*	C2
166	v2,3,5/VIII	India	1957	*C. arabica*	C3_SGI
178*****	v2,3,5/VIII	India	1958	*C. arabica*	C3_SGI
264	v1,4,?/XIX	Central African Republic	1963	*C. excelsa*	N/A
535	v5,6/XXII	Timor	1963	HDT	C3_SGI
741*****	v2,5/I	Sri Lanka	1964	*C. arabica*	C3_SGI
995*****	v1,5/III	India	1968	*C. arabica*	N/A
999*****	v2,4,5,6/XXVIII	India	1968	*Robusta* x *Arabica*	C3_SGI^{int}
1065	v5/II	Brazil	1970	*C. arabica*	C3_SGI
1321	v5,6,7,8,9/XXIX	Timor	1972	HDT	C3_SGI
1427*	v5/II	Kenya	1976	*C. arabica*	C3_SGII
2191	v2,5,6,7,9/XXXVII	India	1992	HDT	C3_SGI
3238	NA	Ethiopia	2008	*C. arabica*	C3_SGII
3302	v1,2,4,5,6,7,8,9,+	India	2009	HDT	C3_SGI
3305*****	v1,2,4,5,6,7,8,9,+	India	2009	HDT	C3_SGI
178a	v2,3,4,5/XIV	CIFC mutant	1960	*C. arabica*	C3_SGI
264a*****	v1,4,6,?/XXVII	CIFC mutant	1968	*C. excelsa*	C1
999a	NA	CIFC mutant	1970	HDT	C3_SGI

***** Rust isolates utilized for genetic diversity analysis.

2.1.2 Flow cytometry

This technique was performed according to Tavares *et al.* (2014) to estimate the genome size among various pathotypes, in comparison with DNA reference standards *Rhamnus alaternus* (2C = 0.68 pg; Carvalho *et al.,* 2018) and *Solanum lycopersicum* (2C = 1.96 pg; Doležel *et al.,* 1992).

Following Galbraith *et al.* (1983) protocol, ~50mg of coffee infected leaf was chopped, with proper razor blades, in a Petri plate with 1mL of Woody Plant Buffer (WPB; 0.2M Tris–HCl, 4mM $MgCl_2$, 1% Triton X-100, 2mM Na_2EDTA, 86mM NaCl, 20mM sodium metabisulfite,1% PVP-10, pH 7.5) allowing the release of the nuclei, as the cell wall and membranes breakdown. The same portion of leaf from plant DNA reference standard was simultaneously chopped to release the nuclei for standardization. The suspension was filtered in 30 µm nylon filter to remove any large residue. Propidium iodide (Sigma-Aldrich) was added to the solution in a 50 µg/mL concentration to stain the DNA, allowing the fluorescence reading to measure nuclei size individually. After incubation at room temperature, the fluorescence of each sample was analysed with a Sysmex CyFlow Space flow cytometer (Sysmex, Norderstedt, GmbH), with a 30mW green laser emitting at 532nm for optimal excitation.

To validate each peak for *H. vastatrix* pathotypes, coffee and reference standard plants, separate analysis were performed to confirm each one. For each sample, five replicate measurements were performed. The results were obtained in four graphics within FloMax software (Sysmex, Norderstedt, GmbH): Fluorescence pulse integral in linear scale (FL); Forward Light Scatter (FSC) *versus* Side light Scatter (SSC) both in logarithmic (log) scale; Time *versus* FL in linear scale; SSC in log scale *versus* FL in linear scale.

Then, with FloMax gating tools, linear regions were outlined in the FL histogram to restrain the nuclei peaks and obtain quantitative statistic data about nuclei number, mean channel position and coefficient of variation (CV). The following formula as used to calculate genome size in mass units (1C in pg for fungi):

$$\frac{\text{Mean G1 fluorescence of sample nuclei}}{\text{Mean G1 fluorescence of reference standard}} \times \text{2C genome size of the reference standard}$$

The conversion of mass values to number of base pairs (bp) was assessed according to the factor 1pg = 978 Mbp (Doležel and Bartoš, 2005).

To have high confidence on the results, the quality of the flow cytometry histograms was evaluated based on the CV of the G1 peaks and the background debris, as well as the

CV of the genome size estimation for each pathotype based on the individual measurements. There were only assessed CV values of DNA peaks below 10% according to criteria established by Bourne *et al.* (2014).

Data measurements were inserted into an Excel sheet to calculate the average genome size to each Hv isolate. Then, a statistical analysis of variance (ANOVA) was performed for all averages with a Tukey HSD (Honest Significant Difference), α=0.05, in STATISTICA 8® (StatSoft Inc.) software. A matrix was obtained, from which were assessed the homogeneous groups among contrasting pathotypes.

2.2 Genetic diversity assessment for the selected retrotransposons

2.2.1 Fungal material

Twenty-one *Hemileia vastatrix* isolates, previously described in Table 2, were used for the genetic diversity assessment study. The fungal material were fresh uredospores collected upon multiplication on coffee leaves (*Coffea arabica* var. Caturra) in the greenhouse as previously described in section 2.1.1, as well as frozen uredospores maintained at -80ºC.

2.2.2 DNA extraction

The protocol used *for H. vastatrix* DNA extraction was developed based on Schwessinger and Rathjen (2017) and Kolmer *et al.* (1995) protocols. About 20mg of Hv spores were transferred to a 1.5mL eppendorf tube together with sterilized sand, and grounded with a sterile micro-pestle, keeping the tube bottom tip drowned in liquid nitrogen. A volume of 500 µl of buffer mix (2.5 buffer A: 2.5 buffer B: 1 buffer C + 2% 2-mercaptoethanol, Annex) plus 15 µl of RNAse A (20mg/ml) (Promega©, Madison, Wisconsin, USA) were added in each eppendorf, mixed through inversion, and incubated at 37°C for 30 minutes. Then, 20 µl of Proteinase K (10mg/ml) was added to each tube and incubated at 65°C for 1 hour. They were post-incubated in ice for 5 minutes to stop enzyme activity. Tubes were vortexed and 500 µl of chloroform/isoamyl alcohol (24:1) (Sigma-Aldrich©, Saint Louis, Missouri, USA) was added in each one. The solutions were mixed by inversion and centrifuged at 13500rpm for 15 minutes. The supernatant was transferred to new 1.5 mL sterilized eppendorf tubes, 15 µl of RNAse A (20 mg/ml) was added and the microtubes were incubated in a warm bath at 37°C for 30 minutes. Afterwards, another 500 µl of chloroform/isoamyl alcohol (24:1) were added to each tube and centrifuged at 13500 rpm for 5 minutes. The supernatant was transferred to new eppendorf tubes, measured with micropipette, and 1 volume of ice-cold isopropanol was added. Upon incubation at -80°C for 30-45 minutes, the samples were centrifuged for 20 minutes at 13500 rpm, and the supernatant was carefully removed. The pellet was

washed with 500 µl of ethanol (70%), centrifuged for 5 minutes at 13500 rpm and the ethanol discarded. A following maximum speed centrifugation was done to remove the ethanol remains. Finally, samples were set to air dry and resuspended in 30 µl of Tris-HCl (10mM; pH 8.5) each.

The DNA samples were quantified by spectrophotometry with Multiskan Sky Microplate Spectrophotometer (ThermoFisher Scientific™, Waltham, Massachusetts, USA) to verify DNA purity, and concentration. DNA integrity and RNA contamination was verified by electrophoresis in 0.8% (w/v) TBE (Tris-borate-EDTA) agarose gel at 85V, loading 1µl of DNA template + 2µl loading buffer (Bromophenol Blue, Xylene Cyanol FF (XCFF); Thermo Scientific 6X DNA Loading Dye diluted to 1X).

The RNA digestion step during the DNA extraction protocol was not enough to fully remove the RNA, therefore another RNA digestion was done. To each sample presenting RNA contamination, 10 µl of RNAse A (20mg/ml) was added and the eppendorfs were incubated in a warm bath at 37°C for 1 hour. To remove the enzyme and precipitate a purified DNA, three different protocols (Supplementary Material) were tested using two samples. Based on the results, the following protocol was selected for purification and applied to the rest of the samples:

PacificBiosciences® "Extracting DNA Using Phenol-Chloroform" protocol

First, in the eppendorfs with previously extracted DNA, 170 µl of Tris-HCl (10mM; pH 8.5) was added to fulfil 200 µl of final volume. Then, 200 µl of phenol/chloroform/isoamyl alcohol (P/C/I) solution (25:24:1; pH 8.0) (Thermo Fisher Scientific™, Waltham, Massachusetts, USA) was added, and the solution was mixed through inversion for 1 minute. Next, the tubes were centrifuged at high speed for 5 minutes, and then 180 µl of the top aqueous solution was transferred to a new sterile eppendorf. Another 200 µl of P/C/I solution were added to the first tubes, mixed by inversion for 1 minute and centrifuged at high speed for 5 minutes. All the top aqueous solution was transferred for the second tube, without collecting any of the P/C/I phase. The volume of the second tube's solution was measured and an equal volume of chloroform/isoamyl alcohol (24:1) solution was added to the tubes. Reiteratively, the solution was mixed by inversion for 1 min and centrifuged at maximum speed for 5 minutes. The top aqueous solution was transferred to another sterile eppendorf tube, without picking any of the chloroform/isoamyl alcohol phase. At this stage, an RNA digestion step was performed to ensure that no RNA remains in the final solution. Therefore, 10 µl of RNase A/T1 Mix (2mg/mL, 5000U/mL) (Thermo Fisher Scientific™, Waltham, Massachusetts, USA) were added to each tube, mixed by pipetting, and incubated between 30 minutes and 1 hour

in a heated bath at 37°C. In the final procedure, ammonium acetate in concentrated solution (8M) was added to a final concentration of 0.75M. Also, 1 µl of glycogen 20 mg/mL (Thermo Fisher Scientific™, Waltham, Massachusetts, USA) was added. The solution was mixed by inversion. A volume of 2.5X of 100% ethanol was added, mixed well by inversion, and centrifuged at top speed for 20 minutes at 4°C. The supernatant was discarded, 300 µl of 80% ethanol was added, the pellet was washed and centrifuged for 15 minutes (4°C) at top speed. This step was repeated, and the pellet was air dried, after discarding the supernatant, and resuspended in 25 µl of Tris-HCl (10mM; pH 8.5).

2.2.3 Selection of candidate retrotransposons and primer design

Candidate retrotransposons were chosen from BLASTn querying of 9234 expressed sequence tags (ESTs) of *H. vastatrix* from three differentiation/infection stages (Talhinhas *et al.*, 2014) against full-genome annotated sequence data (PRJNA912191; PRJNA419278) (Cristancho *et al.*, 2014; Porto *et al.*, 2019) with an e-value threshold of 1e-50 as the cut-off for restricting the alignments to the most significant ones. The candidate retrotransposons were then blasted against diagnostic SNPs for Hv genetic lineages (Silva *et al.*, 2018) to restrict them into 15-20 annotated retrotransposons for *Hemileia vastatrix*. Then, considering the best e-values, bit score and expression profile (Talhinhas *et al.*, 2014), seven candidate retrotransposons were chosen for subsequent analyses (Table 3).

Further annotation of the selected retrotransposons was performed by local BLASTn using BioEdit v7.7.1 against PucciDB, a specific database of transposable elements in the Pucciniales (Orozco-Arias *et al.*, 2022), with an e-value threshold of 1e-35. Primer's design was done with PrimerSelect module from DNASTAR software ™ (Lasergene, version 11.1.0.54) according to conventional PCR parameters: primer size between 17-24bp; melting temperature around 55-65°C; amplicon bigger than 400 bp; GC% content between 40-60% (Table 3). Secondary structures verification was performed with OligoEvaluator™ online tool by Sigma-Aldrich© and Multiple Primer Analyzer™ by Thermo Scientific©.

Table 3 - PCR primers sequences for seven Hemileia vastatrix ESTs (Talhinhas *et al.*, 2014) to assess genetic diversity among contrasting pathotypes.

Hv transcripts	Retro--transposon ID	Primer Sequence (5´- 3´)	Annealing Temperature (°C)
Hv00075	*R75*	F - CAAGTTTTGGCCAATTTCTCT R - CTTGCTGGGTAATTGTGGAT	60
Hv00190	*R190*	F - CTTATACCCCAATTATCTTCCGC R - CCAGTGGAAGATGAAAGAGCC	68
Hv01057	*R1057*	F - GACATCCTCCACGCAATTC R - GGTGCCAATCCCTATCTTTG	68
Hv01207	*R1207*	F - CAAGTTTTGGCCAATTTCTCT R - GCTTCAACCACTTCCTATTCC	64
Hv02407	*R2407*	F - CTCGGAATTAGGGTTAG R - CCTTATCCTCTAAATATGC	62
Hv04547	*R4547*	F - ACCCCTTGAATTTCTGTTGCT R - GATGCCTTTCCCTATATTCAG	60
Hv08122	*R8122*	F - GTATACCCCTTATTCACTG R - CTGAGACCCTTATATTGAG	60

2.2.4 Amplification and sequencing of candidate retrotransposons

PCR reactions were prepared for a final volume of 20 µl with 0.8 µl of each primer (10 µM) + 2 µl DNA template + 0.4 µl dNTPs (10mM) + 3 µl $MgCl_2$ (25mM) + 4 µl GoTaq® Buffer (5x) + 0.2 µl Taq DNA Polymerase (5U/ µl) + 8.8 µl H_2O Milli-Q.

All PCR amplifications were done in a MyCycler™ thermal cycler system by Bio-Rad® (Hercules, California, USA) using the following program: 3 min at 95°C followed by 35 cycles of 1 min at 94°C, 1 min at 60- 68°C, depending on the ideal annealing temperature for each pair of primers (Table 3);and 1 min at 72°C, with a final extension of 7 min at 72°C. PCR products were verified by electrophoresis on a 1% w/v agarose gel stained with GelRed (Biotium™) (Single amplified PCR products were purified with SureClean Plus (Bioline Meridian Bioscience™), while Wizard® SV Gel (Promega) and PCR Clean-Up System (Promega©) was used for multiple product amplification to cut the desired DNA band from the gel, purify it and prepare it for sequencing. PCR purified products were sequenced at STAB VIDA, Lda. (Caparica, Portugal) in both directions. Forward and reverse sequences were assembled into contigs and edited with SeqMan module from DNASTAR® (Lasergene; version 11.1.0.54).

2.2.5 Cloning of retrotransposon fragments (PCR products)

First sequencing results revealed, for all the retrotransposons, many positions with multiples peaks, proving to be necessary an alternative approach to obtain clean retrotransposon sequences. The multiple peaks suggested the existence of many retrotransposon copies with considerable variation. Therefore, to assess retrotransposon copies, a cloning with pGEM®-T Easy Vector System II (Promega) For reasons directly related to the time involved in this method, as well as its cost, we decided to select and clone 3 retrotransposons for various isolates. The purified fragments of candidate retrotransposons *R190*, *R1057*, and *R2407* were individually ligated to pGEM-T Easy vector, according to the manufacturer's instructions, using ~20 ng of the insert for a 3:1 (insert:vector) molar ratio, for 10 different rust isolates (Table 2). Ligation reactions were incubated at 4°C overnight to optimize ligation efficiency. The ligation reactions were briefly centrifuged, and 2 µl of each reaction was transferred to sterile 1.5mL eppendorf tubes on ice. For transformation, *Escherichia coli* JM109 High Efficiency Competent Cells (Promega) were thawed in an ice bath for 5 minutes, lightly flicked, and 50 µl of cells transferred to each ligation reaction tube, gently flicking the mix, and incubating on ice for 20 minutes. After ice incubation, a heat-shock process was done by putting the tubes in a water bath at 42°C, immediately returning the tubes to ice for 2 minutes. Afterwards, 950 µl of SOC (Super Optimal broth with Catabolite repression) growth medium at room temperature was added to each ligation tubes and incubated for 90 minutes at 37°C with shaking (~150rpm). A volume of 100 µl of each transformation culture were plated in Luria-Bertani medium (marca) plates [LB/ampicillin (100µg/mL), IPTG (100mM) and X-Gal (50mg/mL)] replicating each one to have two plates per transformation reaction, incubating plates overnight at 37°C (Figure 5), as the remaining 900 µl of the transformation culture were stored at -20°C in eppendorf tubes. After selection of transformed colonies (white colonies), colony PCR was performed using the universal primers SP6 and T7, under the following conditions: 20 µl reaction mixtures were prepared with 4 µl of GoTaq Flexi buffer (1X; Promega), 2 µl of MgCl2 (2.5 mM; Promega), 0.4 µl dNTPs mix (0.5mM; Promega), 0.8 µl of each forward and reverse primers (0.4 mM), 0.2 µl of GoTaq G2 Flexi DNA Polymerase (5U/ µl; Promega), 2 µl of water where an isolated colony picked with a toothpick was suspended, and water to the final volume. PCR was carried out with an initial denaturation step at 95°C for 1 min, followed by 35 cycles of denaturation at 95°C for 10 s, annealing at 50°C for 30 s and extension at 72°C for 30 s, and one final step of extension at 72°C for 5 min. The visualization of successful plasmid insertion was done by an agarose gel electrophoresis (0.8%) run of the PCR products.

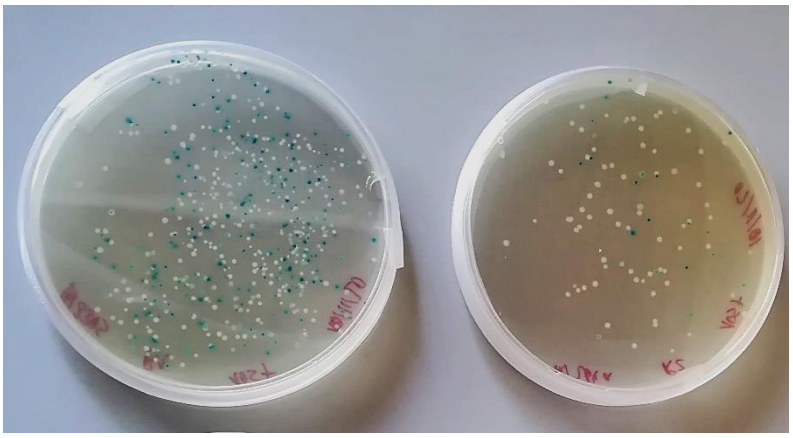


Figure 5 – Example of Luria-Bertani (LB) agar plates with *Escherichia coli* colonies. Successful plasmid insertions are represented as white colonies and non-successful insertions are the blue colonies.

After the agarose gel electrophoresis, to collect successful transformants, the "Wizard® *Plus* SV Minipreps DNA Purification System" was used, according to manufacturer's instructions, to isolate plasmids from the *E. coli* host cells. Three samples of plasmid DNA from each *H. vastatrix* isolate were sent to sequencing with "You Tube It" service by STAB VIDA, Lda. (Caparica, Portugal), using the universal primers T7 and SP6. Sequences were assembled and edited as described in 2.2.4 and deposited in GenBank (accession numbers in Supplementary Table 1).

2.2.6 Sequence Data analysis

Sequences from the Sequencing data were aligned with MAFFT v.7 (Katoh *et al.,* 2022) in AliView v.1.26 (Department of Systematic Biology, Uppsala University, Sweden) (Larsson, 2014). The detection and characterization of polymorphisms, genetic diversity statistics and evaluation of differential segregation of alleles/copies, for the assessment of specific patterns related with pathotypes, was carried out in AliView v.1.26.

The sequencing data such as segregating sites, number of indels, number of haplotypes, haplotype diversity, and nucleotide diversity was obtained, gathering information to generate haplotype networks. The construction of Median-joining Networks (MJN) of haplotypes was done with PopART (Population Analysis with Reticulate Trees) v1.7 software, commonly used to infer and analyse genealogical relationships among populations (Bandelt *et al.*, 1999; Leigh and Bryant, 2015).

2.2.7 Estimation of retrotransposon copy number

Retrotransposons copy number of the selected candidates *R190*, *R1057*, *R2047* were estimated by real-time PCR for all *H. vastatrix* isolates studied, sampling a total of 21 isolates (Table 2). Samples were diluted to the same concentration (~50ng/µl), as measured by Qubit 4 Fluorometer (Thermo Fisher Scientific). For a reliable copy number estimation, a calibration curve was established for each retrotransposon under study, contemplating a series of dilutions, calculated with the online tool "DNA Copy Number and Dilution Calculator" (ThermoFischer Scientific,). Starting with a concentration equivalent to 10^9 copies, calibration points were selected in accordance with the range of copy number of each retrotransposon – *R190*: 100, 50, 25, 10, 5 and 1 copies; *R1057*: 100, 50, 25, 10, 5 and 1 copies; *R2407*: 10000, 5000, 2500, 1000 and 100 copies. Following these calculations, quantitative real-time PCR (qPCR) reactions were performed to obtain the cycle threshold value (Ct) associated to each dilution, i.e., the number of cycles that each dilution requires to cross the fluorescence threshold, to plot Ct against copy number. The more diluted the DNA was, the more cycles, therefore the biggest the Ct value. The number of copies had to be converted in a quantity scale, from where those values were converted in decimal logarithmic values to obtain the calibration curves (Supplementary Table 3).

As qPCR requires small amplicons, new primers were designed within the amplicon sequences obtained from the cloning procedure, based upon the alignment of the sequencing data. Primer's design was done with PrimerSelect module from DNASTAR® (Lasergene, version 11.1.0.54) according to conventional PCR parameters: primer size between 17-24bp; annealing temperature around 55-65°C; amplicon between 80 bp and 200 bp; GC% content between 40-60% (Table 4). The formation of secondary sctrutures was verified with OligoEvaluator™ online tool (Sigma-Aldrich) and Multiple Primer Analyzer (Thermo Scientific©). All qPCR experiments were carried out in 96-well Multiplate (Bio-Rad, Hercules, California, USA) white plates sealed with Optical tape adhesives (Bio-Rad) in an iQ5 real-time Thermal cycler (Bio-Rad). Previous gradient qPCRs (56°C to 64°C) were performed for each gene to determine the optimal annealing temperature. Each 25µl amplification reaction was composed of 12.5µl of iQ™ SYBR® Green Supermix (Bio-Rad,), 0.5µl of each primer (10 µM), 4µl of DNA template and 7.5µl H_2O Milli-Q. The following thermal cycling conditions were used: 10 min at 95°C followed by 45 cycles at 95°C for 15 s and annealing/extension at the respective temperature for each gene for 1 min (Table 4).

Table 4 - Real-time PCR primers for 3 selected transcripts (Talhinhas *et al.* 2014) to estimate retrotransposons copy number and assess expression profiles, and primers for reference genes (Vieira *et al.,* 2011).

Gene ID	Primer sequence	Annealing T°C	Amplicon *(bp)*	PCR Efficiency
R190	F - CTTATACCCCAATTATCTTCCGC R - CTGAGCGGATAGAATGAACAGG	60	227	1.94
R1057	F - CTCTTCTCATCTTGGTTTGTCTCT R - GCATCATTGAAGCCAGTAAAG	60	138	1.91
R2407	F - TCTCGGAATTAGGGTTAGGT R - TTGTGGGAAATGAAAGGTAG	60	84	1.96
GADPH	F- ACTTGGACAGCTACGAC R - CCATACCAGTGAGCTTCC	58	280	1.92
40S_Rib	F - CACACGGAAAGATTGGTACG R - CCTTGAGCGAATCAACGG	60	114	1.97
CytIII	F - AGTAGATATGAGTCCCTGACC R - CACCTTCAGCACTTACATCC	58	173	1.93

To assess the amplification specificity, a melting curve analysis was performed at the end of the PCR run increasing the temperature in a stepwise fashion by 0.5°C every 10s over the range 55 to 95°C. Each set of reactions included a negative control with no template. LinRegPCR version 2013.0 (Rujiter *et al.,* 2014) software was used to determine the primer's amplification efficiencies for each gene of interest. Three biological replicates and two technical replicates were used for each sample, and only technical replicates with a sample variance of CTs below 0.5 were considered for the experimental analysis. To normalize the DNA concentration needed for a single genome amplification, a formula was applied for each isolate, based on its genome measurement by flow cytometry. The resulting values were correspondent to two haploid nuclei, in picograms, so that the proper dilution could be done. After the dilutions done, qPCR reactions were made for all isolates with each retrotransposon, so that the Ct values could be applied to each correspondent calibration curve equation, therefore obtaining the copy number estimations per genome. The estimation values for the retrotransposon copy number were calculated through the least squares method in Excel (Microsoft Office 365), based on the calibration curves for each retrotransposon. The Ct values from the qPCR reactions were set upon the least-squares regression line equations to obtain the copy number estimation values. Dissociation curves and agarose gel electrophoresis were used to analyze non-specific PCR products.

2.3 Expression studies of selected retrotransposons along the infection process

2.3.1 Plant-fungal material

Compatible interactions were established with two *H. vastatrix* isolates inoculated in differential coffee host plants. The isolate Hv70 ($v_{2,4,5}$) was inoculated in H152/3 (S_H 2,4,5) (*C. arabica* hybrid; 32/1 x 110/5) and the isolate Hv1427 (v_5) was inoculated in *C. arabica* var. Caturra 19/1 (S_H 5). The infected coffee leaves were collected in six different time-points post-inoculation: 1 dai (days after inoculation), 2 dai, 3 dai, 4 dai, 6 dai and 8 dai; and 1 dai, 2 dai, 3 dai, 4 dai, 8 dai and 11 dai, for Hv70 and Hv1427 respectively. These time-points are correlated with the development stages of the infection, namely fungal penetration (1 dai), appressorium formation (2 dai), anchor and first haustoria formation (3 dai), >50% infection sites with haustoria (4 dai), colonization (6 to 8 dai, for Hv70 and Hv1427, respectively), and extensive colonization prior to sporogenic hyphae formation – macroscopically identified as beginning of chlorosis before sporulation (8 to 11 dai, for Hv70 and Hv1427, respectively). After collection, the central vein of each leaf was removed using a blade and immediately frozen in liquid nitrogen and conserved at -80°C until used.

2.3.2 RNA extraction and cDNA synthesis

All samples were grounded to powder in sterilized pestles and mortars, previously treated with sodium hydroxide (NaOH, pellets, 98%) to prevent possible contaminations from residues, with liquid nitrogen and stored at -80°C. Total RNA was extracted using Spectrum Plant Total RNA kit (Sigma), following the manufacturer's instructions. RNA concentration and purity of each sample was measured at 260/280 nm and 260/230 nm using a Multiskan Sky Microplate Spectrophotometer (ThermoFisher Scientific™). RNA integrity was verified by electrophoresis in 1.2% (w/v) TAE (Tris-acetate-EDTA) agarose gel at 70V, loading 1µl (~200ng/µl) of RNA template.

RNA samples to be used in qPCR experiments were treated with TURBO DNA-free Kit (Thermo Scientific) to eliminate residual DNA. For this procedure the following steps were performed: 35-40µl of RNA extracted volume, 1µl of Turbo DNAse (2U), 5µl of 10x turbo DNAse buffer (0.1 V) and DEPC-treated water to a final volume of 50 µl were mixed in a tube and incubated at 37°C for 1 hour. Then, 5µl of DNase Inactivation Reagent (0.1V) was added and the mixture was left 5 minutes at room temperature. Samples were then centrifuged for 2 min at 10000 rpm, and the supernatant was transferred to a new tube. To precipitate the RNA, 0.1 vol of NaAc and 2.5 vol of 100% ethanol (-20°C) were added to each sample. Samples were left at -80°C for 45 min and then centrifuged

at 14000 rpm for 30 min at 4°C. Supernatant was discarded and the pellet was washed twice with 750 µl of 70% ethanol. Each sample was then centrifuged at 14000 rpm for 10 min at 4°C. The pellet was resuspended in 14µl of DEPC-treated water.

Complementary DNA (cDNA) was synthesized from the RNA templates using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific™). First, for each sample was added 1µl of Oligo-dT (10 µM), 1000 ng of total RNA, and sterile water to complete 12 µl of total volume. An incubation process was performed to all samples at 70°C for 10 minutes and final holding temperature of 4°C. Then, with all samples on ice, 2 µl of dNTP mix (10 mM), 4 µl of 5X Reaction Buffer, 1 µl of Ribolock RNase inhibitor (20 U U/µl), and 1 µl of Omniscript Reverse Transcriptase (10 U/µl) were added to each sample to a final volume of 20 µl. A second incubation process was done in three steps: 42°C for 60 minutes; 70°C for 10 minutes; and final holding temperature of 4°C. All cDNA samples were stored at -20°C, to further use in qPCR analysis.

2.3.3 Expression analysis by quantitative real-time PCR

The target retrotransposons under study *R190, R1057* and *R2407* were amplified using the specific primers previously designed (Table 4). Reference genes previously established as the most stable for *H. vastatrix in planta* (Vieira *et al.*, 2011) were used: 40S ribosomal protein (*40S_Rib*), glyceraldehyde-3-phosphate dehydrogenase *(GADPH),* and cytochrome c oxidase subunit III (*Cyt III*) (Table 4). The qPCR reactions were performed using an iQ5 real-time thermal cycler (Bio-Rad, Hercules, USA), in 96-well Clear Multiplate PCR Plates (Bio-Rad) sealed with iCycler iQ Optical tape adhesives (Bio-Rad). Previous gradient qPCRs (56°C to 64°C) were performed for each gene to determine the optimal annealing temperature. Each 15 µl reaction volume comprised 4 µl cDNA template, 7.5 µl SsoFast EvaGreen Supermix (Bio-Rad), 0.3 µl of each primer (10 µM), and 2.9 µl of sterile distilled water. The reactions were subjected to an initial denaturation step at 95°C for 10 min followed by 40 cycles at 95°C for 10 s and 58°C/60°C for 30 s, according to Table 4. A melting curve analysis was performed at the end of the PCR runincreasing the temperature in a stepwise fashion by 0.5°C every 10 seconds over the range 60-95°C. Baseline correction was performed with the iQ5 Optical System Software (Bio-Rad), and the corresponding fluorescence values were exported and used to determine Ct and PCR efficiency values with LinRegPCR program.Dissociation curves was used to analyze non-specific PCR products. A no-template reaction was included has control. Three biological replicates and two technical replicates were used for each sample.

The assessment of the correct expression levels was calculated in two steps, using the 1 day after inoculation (1 dai) Ct values as a control for both isolates Hv70 and Hv1427 – for the reference genes, the average Ct values ($Ct_{ref\ genes}$) were calculated for each biological replica, as well as for the spores ($Ct_{control}$). Then, $Ct_{ref\ genes}$ were subtracted to the average value for all 1 dai replicas ($\bar{x}_{controls}$) to obtain Ct variation (ΔCt):

$$\Delta Ct = \bar{x}_{controls} - Ct_{control}$$

and the $2^{-\Delta Ct}$ method, described by Livak and Schmittgen (2001), could be applied. This mathematical formula assumes maximum reaction efficiency, as expressed by the value 2 in the math potency basis. In our study, where the efficiency isn't maximum, the basis will be the calculated primer efficiency (E), being the formula $E^{-\Delta Ct}$ (Hellemans *et al.* 2007). At last, a geometric mean for each replica ($\bar{x}_{replica}$) is calculated:

$$\bar{x}_{replica} = \bar{x}\ (E^{-\Delta Ct}\ \textit{GADPH};\ E^{-\Delta Ct}\ \textit{40S};\ E^{-\Delta Ct}\ \textit{CytIII})$$

For the retrotransposon loci, calculations are similar to those described above, to assess final $E^{-\Delta Ct}$ values. Using the geometric mean values, a fold change value was calculated, representing the accurate gene expression levels considering the controls and the reference genes:

$$\text{Fold change} = \frac{E^{-\Delta Ct}}{\bar{x}\ \text{replica}}$$

Statistical significance ($p \leq 0.05$) of differential gene expression between different compatible Hv-coffee interactions and between different time-points was determined by the non-parametric Mann–Whitney U test using IBM® SPSS® Statistics version 27.0.1.0 (SPSS Inc., USA) software, and by the parametric T-Student in Statistica version 10.0 (StatSoft, GmbH) software.

Relative gene expression (fold change) was calculated according to Hellemans *et al.* (2007), using the 1 dai time-point as a control for both isolates Hv70 and Hv1427, a mean of the the three reference genes and the PCR efficiency for each gene/interation determined by LineReg. Statistical significance ($p \leq 0.05$) of gene expression between the two compatible Hv-coffee interactions and also between different time-points in each compatible interation was determined by the non-parametric Mann–Whitney U test (SPSS® Statistics, version 27.0.1.0, IBM) and the parametric student-*t* test (Statistica version 10.0, StatSoft).

3 RESULTS

3.1 Genome size estimation

Flow cytometry analyses allowed the separation and identification of nuclei of 21 rust isolates, whereby comparison with the internal reference standards (*Rhamnus alaternus* and *Solanum lycopersicum*) permitted the estimation of the genome size for each isolate under study (Table 5). The analyses revealed high levels of genomic variation among the *H. vastatrix* isolates, which ranged between 713 Mbp and 879 Mbp. From all measurements, the average genome size estimation was settled in 786 Mbp, with a 166 Mbp genome size difference between the lowest and highest measurement.

The biggest genome estimations (>800Mbp) contemplate isolates from all phylogenetic groups (C1, C2 and C3), regardless the virulence profile. The isolate Hv3238 (Ethiopia; C3_ SGII) has the smallest genome with 713 Mbp, while the isolate Hv70 (Kenya; subgroup C3_SGI) has the biggest genome with 879 Mbp – both have *C. arabica* (tetraploid) as their host coffee plant species. Isolate Hv999 (India; subgroup C3_SGI^{int}), with signs of introgression from the C2 group (Silva *et al.* 2018), sets as one of the isolates with the smallest genome estimated at 719Mbp, much below the average estimation.

Tukey HSD (honestly significant difference) results revealed various homogenous groups within 95% confidence level ($\alpha = 0.05$). Among the *H. vastatrix* isolates, Hv3238, Hv999 and Hv1321 were outlying in only one homogenous group, as well as Hv70 in another isolated group. Nonetheless, no correlation was found between genome size and pathotype, geographical origin, coffee host or phylogenetic group.

3.2 Selection and characterization of candidate retrotransposons

The BLASTn results identified 201 *H. vastatrix* transcripts in common with genome sequences representing retrotransposons sequences. To narrow down these results and improve the chances of selecting candidates with a role in host adaptation, these transcripts were blasted against diagnostic single nucleotide polymorphisms (SNPs) loci for phylogenetic lineages generated from restriction site-associated DNA sequencing (Silva *et al.*, 2018).

Table 5 - Genome size estimation by flow cytometry for *Hemileia vastatrix* isolates under study. NA: Not Available; HDT: Timor hybrid and derivative coffee genotypes; *as describe in Rodrigues *et al.* (2022): C1, C2, and C3 - genetic lineages; I,II,II - subgroups within the genetic lineage C3; int - signals of introgression; CV - variance coefficient; n – number of samples; a to k – homogenous groups from the Tukey HSD.

Hv isolate	Pathotype /race	Origin	Year	Coffee Host	Phylo genetic group*	(1C) genome size	n	Average CV (%)
22	v2,5/I	Tanzania	1953	*C. arabica*	C3_SGI	834^{ij}	5	8,71
70	v2,4,5/XXIV	Kenya	1954	*C. arabica*	C3_SGI	879^{k}	5	5,56
71	v?/IV	Mozambi-que	1954	*C. racemosa*	C2	824^{hij}	6	2,84
92	v?/XVIII	S.Tomé	1956	*C. liberica*	C2	794^{efghi}	3	NA
166	v2,3,5/VIII	India	1957	*C. arabica*	C3_SGI	777^{cdef}	5	3,95
178	v2,3,5/VIII	India	1958	*C. arabica*	C3_SGI	774^{cdef}	5	2,95
264	v1,4,?/XIX	Central African Republic	1963	*C. excelsa*	NA	788^{efgh}	6	3,64
535	v5,6/XXII	Timor	1963	HDT	C3_SGI	827^{hij}	2	NA
741	v2,5/I	Sri Lanka	1964	*C. arabica*	C3_SGI	758^{bcd}	10	5,44
995	v1,5/III	India	1968	*C. arabica*	NA	783^{cdef}	5	3,09
999	v2,4,5,6/ XXVIII	India	1968	Inter_HIB robustaXarabica	C3_SGI^{int}	719^{a}	5	3,62
1065	v5/II	Brazil	1970	*C. arabica*	C3_SGI	782^{cdef}	8	5,69
1321	v5,6,7,8,9/ XXIX	Timor	1972	HDT	C3_SGI	721^{a}	6	3,89
1427	v5/II	Kenya	1976	*C. arabica*	C3_SGII	751^{abc}	6	4,00
2191	v2,5,6,7,9/ XXXVII	India	1992	HDT	C3_SGI	842^{ik}	4	NA
3238	NA	Ethiopia	2008	*C. arabica*	C3_SGII	713^{a}	2	4,17
3302	V1,2,4,5,6,7, 8,9,+	India	2009	HDT	C3_SGI	789^{efgh}	7	5,33
3305	V1,2,4,5,6,7, 8,9,+	India	2009	HDT	C3_SGI	792^{efgh}	7	4,32
178a	v2,3,4,5/XIV	CIFC mutant	1960	*C. arabica*	C3_SGI	772^{cde}	10	4,66
264a	v1,4,6,?/ XXVII	CIFC mutant	1968	*C. excelsa*	C1	824^{hij}	4	3,80
999a	NA	CIFC mutant	1970	HDT	C3_SGI	771^{cde}	8	4,25

From this analysis, 21 transcripts were obtained common to all variables searching for retrotransposon sequences and containing homology with diagnostic SNP loci (Supplementary Table 1). From those 21 transcripts, 7 of them were selected for this study, based on the best hit parameters. An homology search in NCBI revealed that all 7 EST candidates are an integrant part of *Gypsy*-like lineages of LTR (Long Terminal Repeat) retrotransposons in the genome of *H. vastatrix* (Supplementary Table 2). Further annotation by searching homologies using BLASTn of the selected retrotransposons against PucciDB (Orozco-Arias *et al.*, 2022), a specific database of transposable elements in the Pucciniales, provided a more detailed characterization with high levels

of similarity (93,5 to 100% identity). Within the Gypsy superfamily, homology was found with lineages CO-HUI (sub-family Betania), V_Clade and Soroa (Sub-family JCO). However, in contradiction with the first annotation, one retrotransposon (*R190*) was assigned to the Copia superfamily (lineage Mapi), while three other retrotransposons were classified as non-autonomous elements TRIM (Terminal repeat retrotransposons in miniature) (*R1057* and *R4547*) and LARD (Large retrotransposon derivatives) (*R2407*) (Table 6).

Table 6 - BLASTn Annotation of the selected retrotransposons against PucciDB.

Candidate RT	Description	% Identity	Alignm-ent Length	Nº of Gaps	Nº of Missm-atch	e-Value	Score
R75	Hvastatrix_Sequence93#LTR/SOROA/JCO	98.462	260	4	0	4.54e-129	462
R190	Hvastatrix_Sequence119#LTR/MAPI	98.176	987	14	4	0.0	1720
R1057	Hvastatrix_Sequence1012#LTR/TRIM	93.488	215	11	3	1.75e-84	316
R1207	Hvastatrix_Sequence148#LTR/V_CLADE	96.949	721	20	2	0.0	1208
R2407	Hvastatrix_Sequence298#LTR/LARD	100.0	155	0	0	1.41e-76	287
R4547	Hvastatrix_Sequence1012#LTR/TRIM	94.845	194	8	2	2.97e-80	302
R8122	Hvastatrix_Sequence8#LTR/BETANIA	96.45	169	6	0	5e-74	279

3.3 Long-Terminal Repeat Retrotransposons (LTR-RTs) genetic diversity

To study the genetic diversity of the candidate retrotransposons, an amplicon-based sequencing approach was used. Focusing on three cloned retrotransposons (*R190*, *R1057* and *R2407*), retrotransposons were studied for 9-10 rust isolates (Table 2). Clean sequences were obtained upon cloning for all isolates, as well as multiple isoforms of the retrotransposons under study. For all retrotransposons under study, each haplotype generated from the sequencing data is a single different copy, corresponding to a single cloning event. Sequencing data was analysed for various parameters, such as segregating sites, number of indel mutations (insertions/deletions), haplotype diversity, and nucleotide diversity (Tables 7, 8 and 9).

A high level of genetic variability was found for all retrotransposons, both among and within Hv pathotypes, revealing highly polymorphic singleton haplotypes/copies. The *R190* sequencing data revealed many polymorphic sites among all isolates (211), a considerable difference to the other retrotransposons under study, where the segregating sites were 73 for *R1057* and 100 for *R2407*. The number of indel mutations

is also higher in *R190* data (18 indel), compared to the other retrotransposons (3 indel), among the multiple haplotypes. The same situation occurs for the nucleotide diversity, where the parameter is much higher for *R190* (0.1197) in comparison to *R1057* (0.005619) and *R2407* (0.033491). When analysing copy diversity within isolates, the intraspecific variation is larger for the isolate Hv3305 for both *R190* (78) and *R2407* (30), whereas Hv999 is the isolate with most copy variability for *R1057* (36). Comparing all the copies within each rust isolate, the retrotransposon *R190* shows the highest level of genetic diversity for each rust isolate, having the widest range of segregating sites and also the number of indels.

Table 7 – Genetic diversity statistics for 9 rust isolates concerning *R190* retrotransposon. The sequences were analysed to various parameters to assess genetic diversity among contrasting pathotypes.

R190	**Segregating sites**	**Nr of Indel**	**Nr of haplotypes (excluding indel sites)**	**Haplotype Diversity**	**Nucleotide diversity**
Hv1427	37	6	3	1.0	0.1265
Hv178	49	3	2	1.0	0.0843
Hv264a	63	5	3	1.0	0.1683
Hv3305	78	3	3	1.0	0.1403
Hv70	71	6	3	1.0	0.1761
Hv71	42	3	3	1.0	0.0748
Hv741	34	2	3	1.0	0.0894
Hv92	31	2	3	1.0	0.0870
Hv995	65	9	3	1.0	0.1932
Total	211	18	26	1.0	0.1197

The high genetic variability can be supported by the results of the haplotype networks. Those networks are representative of the multiple divergent haplotypes/copies of each retrotransposon, among the various rust isolates, as well as their probable evolutionary proliferation in the genome. The networks generated are very low structured with a high number of mutational steps. For instance, regarding the *R190* network (Figure 6A), there is no central haplotype representing an ancestor from which the others diverged, but rather quite independent and diversified haplotypes which can differ by more than 30 SNPs and up to 84bp indels. Also, this complex diversification pattern within the same rust isolate suggests a rapid evolution in the genome. A similar pattern is found for *R2407* network (Figure 6C), although the genealogical relationships are less complex than for *R190*. In accordance, the number of mutations differentiating haplotypes are also lower, reaching up to 10 in between copies. In contrast, in the *R1057* network (Figure 6B), all haplotypes seem to descend from a common ancestor, represented by a central haplotype. The network structure is more organized than the previous ones described, showing a direct divergence of single haplotypes from the central one (belonging to

Hv999) and an expanding lineage comprising different copies of Hv999, Hv1427 and Hv741. Also, the genetic distance between copies is shorter as shown by the few mutational steps between different copies (1 to 5).

Table 8 – Genetic diversity statistics for 10 rust isolates concerning *R1057* retrotransposon. The sequences were analysed to various parameters to assess genetic diversity among contrasting pathotypes.

R1057	**Segregating sites**	**Nr of Indel**	**Nr of haplotypes (excluding indel sites)**	**Haplotype Diversity**	**Nucleotide diversity**
Hv1427	12	0	3	1.0	0.006388
Hv178	9	0	3	1.0	0.005226
Hv264a	14	1	3	1.0	0.008130
Hv3305	6	0	2	0.6667	0.003484
Hv70	10	0	3	1.0	0.005807
Hv71	11	0	3	1.0	0.005517
Hv741	12	0	3	1.0	0.006969
Hv92	15	0	3	1.0	0.008711
Hv995	9	0	3	1.0	0.004355
Hv999	36	2	14	0.9750	0.003796
Total	73	3	39	0.9954	0.005619

Table 9 – Sanger sequencing data results for 9 rust isolates concerning *R2407* retrotransposon. The sequences were analysed to various parameters to assess genetic diversity among contrasting pathotypes.

R2407	**Segregating sites**	**Nr of Indel**	**Nr of haplotypes**	**Haplotype Diversity**	**Nucleotide diversity**
Hv1427	22	1	3	1.0	0.038801
Hv178	13	0	3	1.0	0.022928
Hv264a	24	1	3	1.0	0.043210
Hv3305	30	3	3	1.0	0.053509
Hv70	11	2	3	1.0	0.019400
Hv71	13	1	3	1.0	0.022928
Hv741	22	2	3	1.0	0.038801
Hv92	21	0	3	1.0	0.037037
Hv995	18	2	3	1.0	0.032628
Total	100	3	27	1.0	0.033491

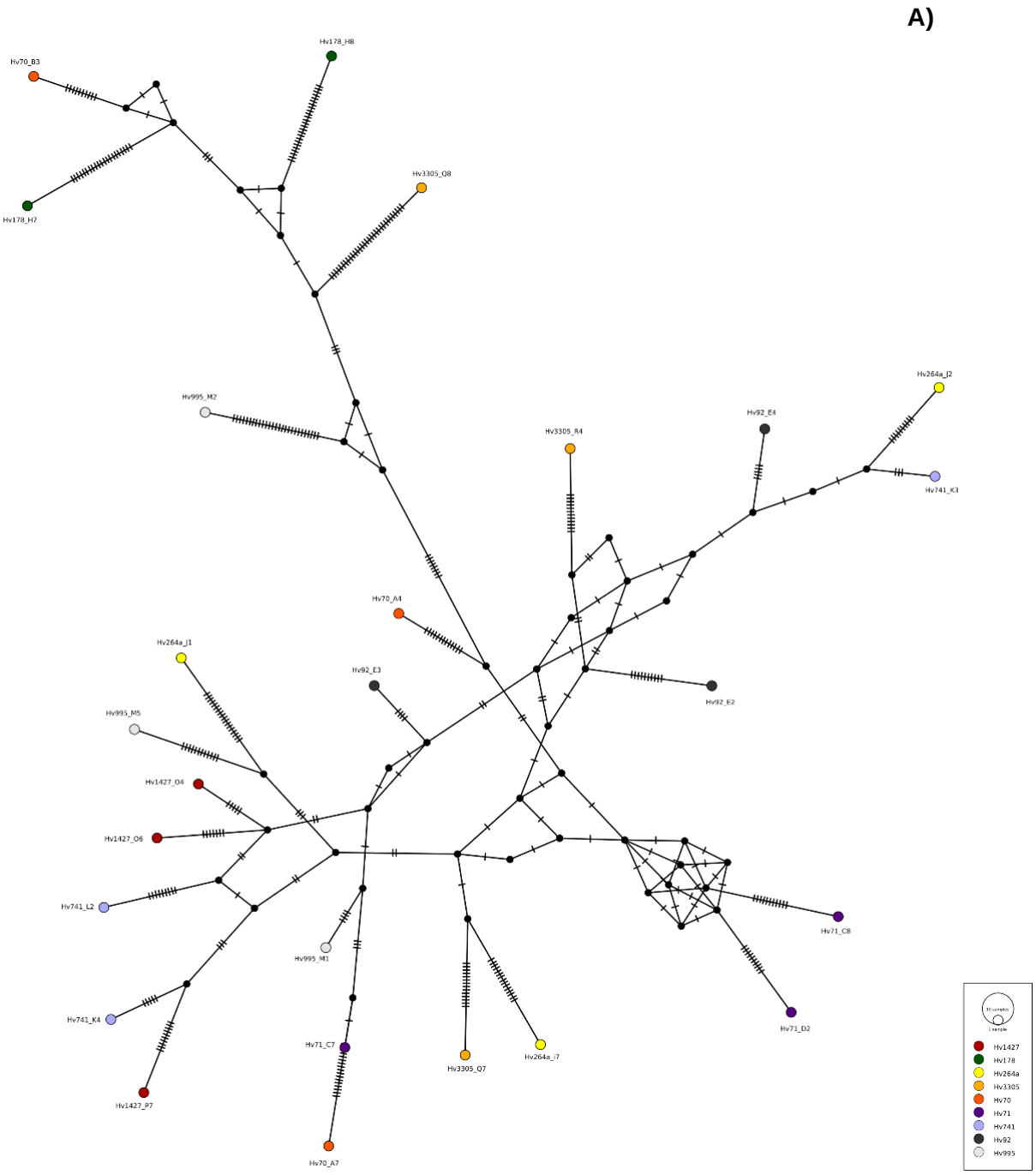


Figure 6 – Median-Joining Network for the retrotransposons *R190* (A), *R1057* (B), and *R2407* (C) for 9-10 *Hemileia vastatrix* isolates. Each coloured circle is representative of an isolate, as seen in the legend, corresponding to a sequenced haplotype. The dark circles are representative of hypothetical haplotypes that diverge from each real haplotype. Evolutionary mutations are represented by the lines in between copies.

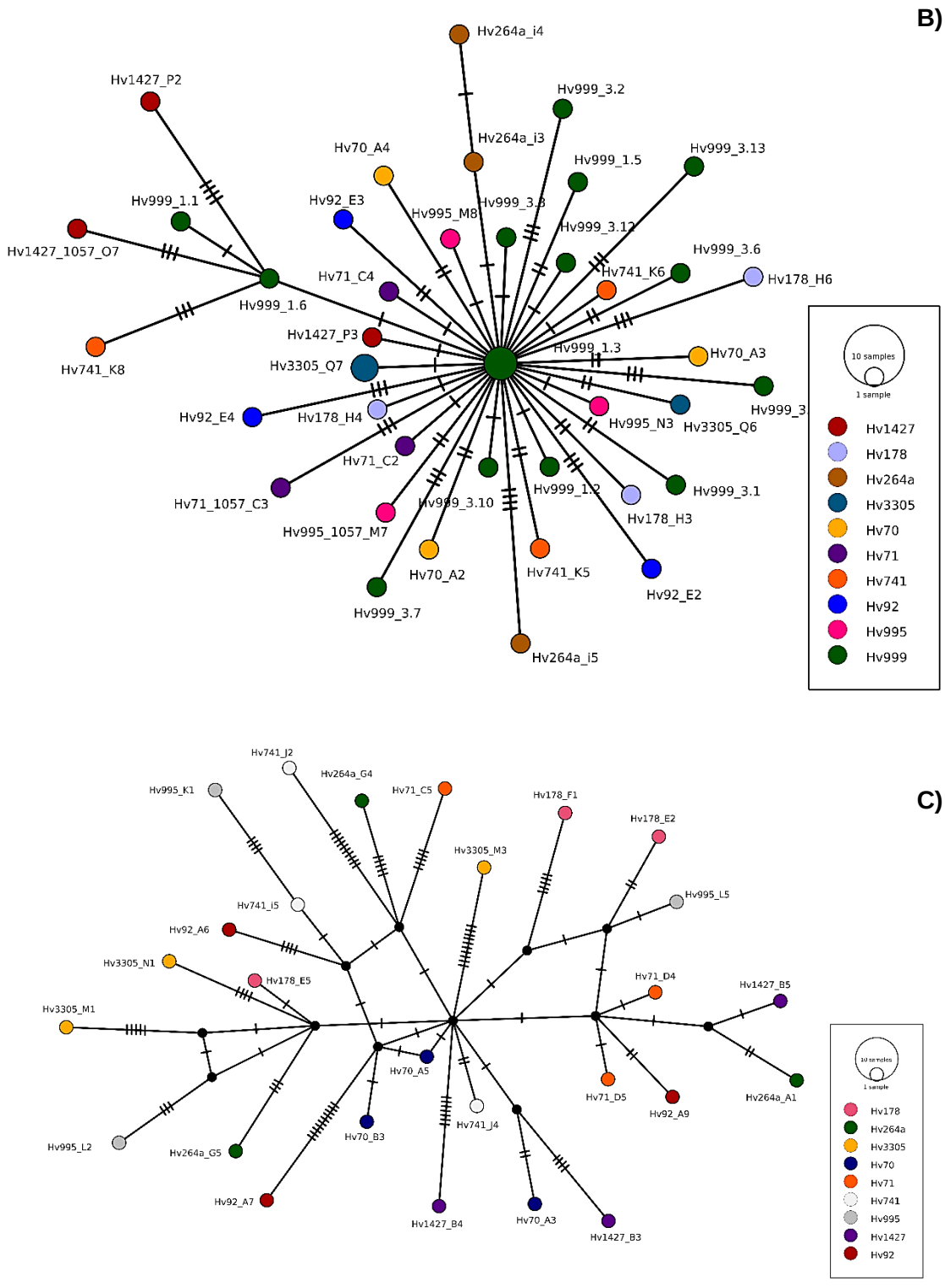


Figure 6 – continuation.

3.4 Retrotransposon copy number estimation per genome

Real-time quantitative PCR (qPCR) was used to estimate the number of copies of 3 retrotransposons (*R190*, *R1057*, *R2407*) per genome in 21 isolates of *Hemileia vastatrix*. Using the same rust isolates for which the genome size was measured by flow cytometry, this approach intended to evaluate the potential relation between the genome size and the number of copies of these transposable elements in the genome.

After the qPCR reactions, a calibration curve was constructed, for each retrotransposon – *R190*, *R1057*, and *R2407* (Supplementary Table 3 and Supplementary Figure 1). For *R1057*, the calibration curves were set up with six Log_{10} values corresponding to 100, 50, 25, 10, 5, and 1 genome copies. The coefficient of determination (R^2; $R^2 \leq 1$) for *R1057* is 0.9982 with a slope-intercept equation $y = -3.6412x + 25.975$. For *R190*, the curve was done with seven Log_{10} values corresponding to 1000, 750, 500, 250, 100, 50, and 25 copies. The R^2 value is 0.9967 with a slope-intercept equation $y = -3.336x + 24.325$. As for *R2407*, the curve was constructed with five Log_{10} values corresponding to 10000, 5000, 2500, 1000, and 100 copies. The R^2 value is 0.9835 with a slope-intercept equation $y = -4.2873x + 26.5$.

Copy number estimation showed that the three retrotransposons present very distinctive ranges of copy number values: *R1057* varies between 1-103 copies per genome; *R190* values vary between 495-3486 copies; and *R2407* values range from 304-4457 copies (Table 10). *R190* and *R2407* seem to be the most proliferative retrotransposons, ascending to over 3000 copies per genome, which is in line with the higher variability within copies found for these retrotransposons.

Contrary to the expected, there was no apparent direct correlation between genome size and retrotransposon copy number. The rust isolate with most estimated copies for *R190* is Hv3305, with 3486 copies, and the same for *R2407* with 4457 copies per genome, in agreement with the genetic diversity data. However, this isolate presents a medium genome size (792 Mbp), but it is interesting to note that this is the most virulent isolate included in our sampling. As for *R1057*, the isolate with most copies is Hv1065 (782 Mbp). On the other hand, the rust isolate with the fewest copies for *R190* is Hv92 (794 Mbp); as for the retrotransposon *R1057* the isolate Hv999 (719 Mbp) presents a single copy per genome; for *R2407* the isolate Hv999a is the one with fewest copies, revealing 304 copies per genome.

Table 10 – Compilation of qPCR results for copy number estimation per genome (1C) for the three retrotransposons under study: *R190*, *R1057*, and *R2407*. Isolates are organized from the lowest to the biggest genome size measurement. Data values are colour coded in gradual shading from green (low) to red (high).

		Retrotransposons		
Hv isolates	**genome size (Mbp)**	***R190***	***R1057***	***R2407***
Hv3238	713	1112	N/A	1143
Hv999	719	800	1	1216
Hv1321	721	1781	3	3377
Hv1427	751	1108	3	578
Hv741	758	873	4	1949
Hv999a	771	623	4	304
Hv178a	772	934	4	1008
Hv178	774	663	2	1109
Hv166	777	932	2	889
Hv1065	782	1650	103	1515
Hv995	783	725	3	1144
Hv264	788	1079	41	758
Hv3302	789	1435	6	649
Hv3305	792	3486	4	4457
Hv92	794	495	10	923
Hv71	824	1008	24	408
Hv264a	824	1771	39	728
Hv535	827	1934	2	1067
Hv22	834	725	N/A	1322
Hv2191	842	1447	28	3460
Hv70	879	1848	2	1486

3.5 Expression profiles of LTR-retrotransposons

Two compatible *Coffea arabica-Hemileia vastatrix* interactions were studied to assess expression profiles of the retrotransposons *R190*, *R1057* and *R2407*, and assess putative relations with virulence. Isolates Hv70 (race XXIV [$v_{2,4,5}$]; host H152/3 [S_H2,4,5]) and Hv1427 (race II [v_5]; host *C. arabica* var. Caturra 19/1 [S_H5]) were inoculated and collected in six different time points post-inoculation. The expression profiles of these retrotransposons were analysed during the pathogen's infection process, to assess whether they were upregulated or downregulated, taking the first day after inoculation as a control (1 dai). Data analysis showed quite distinctive expression profiles for the two *H. vastatrix* isolates. In comparison with 1 dai, all retrotransposons were upregulated in all time points for the Hv1427 compatible interaction, and mostly downregulated in the Hv70 compatible interaction, albeit with differences in expression levels and some differential patterns of expression between isolates and retrotransposons (Figure 7).

Globally, when comparing retrotransposon relative expression values between isolates, Hv1427 shows a much higher level of expression of *R2407* for at all time-points, in relation to *R190* and *R1057*, with a distinct peak of upregulation at 8 dai (Figure 7; Supplementary Figure 3). On the other hand, in Hv70, all retrotransposons presented a similar expression regulation from 4 dai onwards, showing a consistent and stronger downregulation at 4 dai. However, on the early stages of infection, *R2407* (2 dai) and *R1057* (3 dai) were detected as upregulated (Figure 7; Supplementary Figure 2). Also, the statistical analysis from the Mann-Whitney U Test revealed no significant statistical differences, opposing to the Student-*t* test where all time-points revealed to be significantly different when comparing isolates for each retrotransposon under study (Supplementary Table 4). When analyzing expression along the infection process, for the *R190* retrotransposon, the major significant statistically difference of expression (*p*-value = 0.000904; Supplementary Table 4) between Hv70 and Hv1427 occurs at the early stages of infection (2 dai). The peak of relative expression fold-change occurs at 3 dai (15.5 ± 5.4) for Hv1427, a time point that matches with haustoria formation. On the other hand, Hv70 presents almost all time points as downregulated throughout the infection process, except at the 3 dai (1.1 ± 1.0) (Figure 7). A similar pattern of relative expression was found for *R1057* retrotransposon in Hv 70 (Figure 7). For the Hv70 compatible interaction, *R1057* expression has bigger fold change value at 3 dai (1.4 ± 1.3), being the only time-point where the retrotransposon is upregulated. While for the Hv1427 interaction, *R1057* expression is quite steady along time, with the major fold change occurring at 8 dai (10.0 ± 3.0). The lowest relative expression values can be observed at 4 dai for Hv70 (-2.5 ± 0.3) and at 11 dai for Hv1427 (5.8 ± 0.7). Major discrepancies between isolates can be observed at two crucial phases, namely the appressoria formation phase at 2 dai, and the haustoria infection phase at 4 dai, where statistically significant differences are found with p-value < 0.01 (Figure 7, Supplementary Table 4). For the *R2407* retrotransposon, the Hv70 interaction is only upregulated at the appressorium formation phase at 2 dai (1.3 ± 1.1), opposing to Hv1427 compatible interaction where the retrotransposon is upregulated throughout the infection process. The peak expression in fold change value is found at 8 dai for Hv1427 (62.2 ± 1.3) during the colonization phase of the pathogen preceding sporulation, similarly to what happens for *R1057*. Identically to what was detected for the other retrotransposons, the most significant downregulation in Hv 70 occurs during the haustoria infection phase at 4 dai (-2.7 ± 0.2) (Figure 7). In fact, the time points with stronger statistically significant differences between the compatible interactions are at 4 dai (p-value = 0,000005) and 6 / 8 dai (p-value = 0,000003) (Supplementary Table 4).

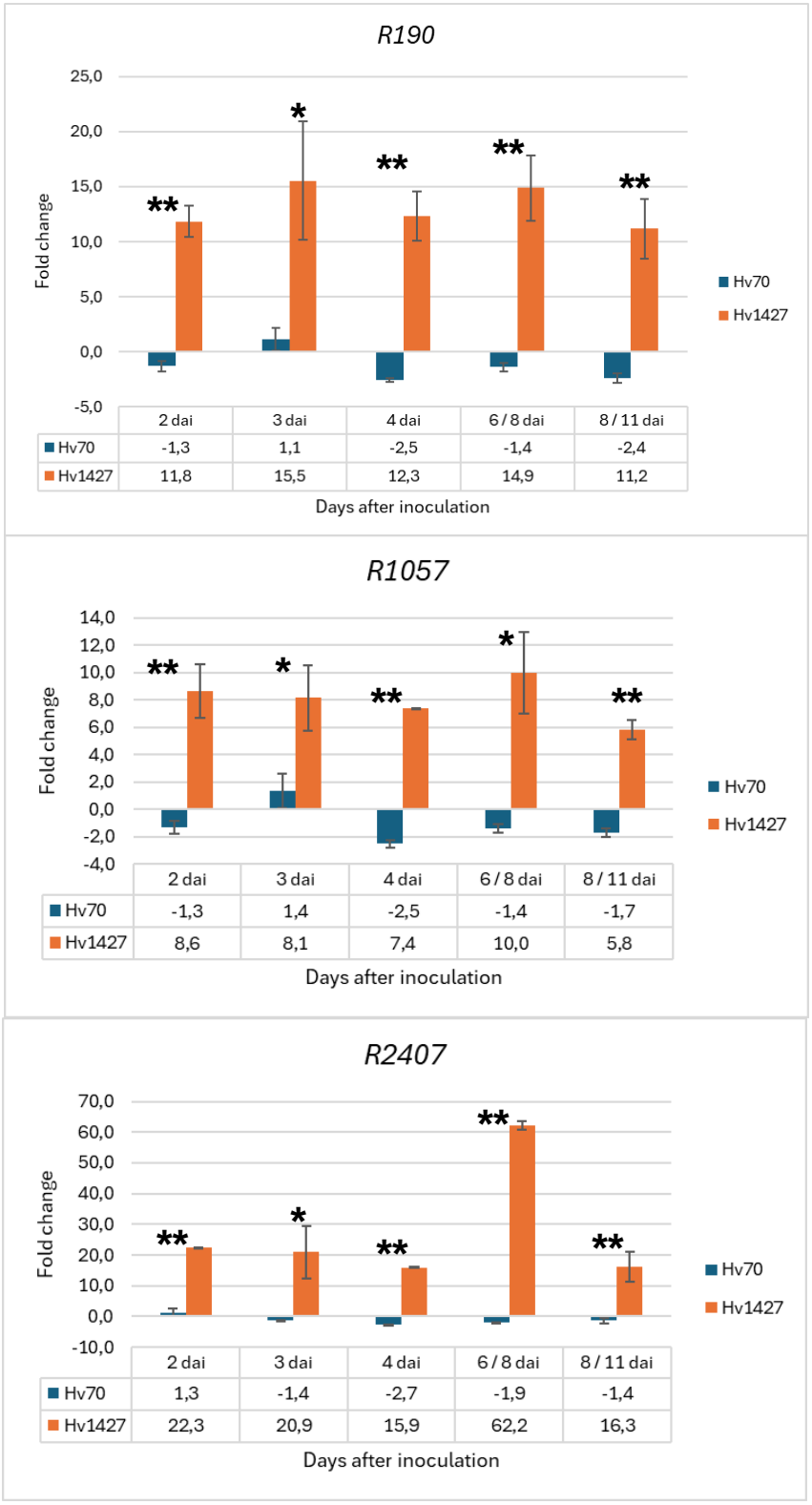


Figure 7 – qPCR expression analysis of three *Hemileia vastatrix* retrotransposons (*R190*, *R1057* and *R2407*). The y-axis represents relative expression calculated as fold change of the retrotransposons under study, as the x-axis is representative of the various post inoculation time-points where samples were collected: 6 / 8 dai -6 dai for Hv70 and 8 dai for Hv1427; 8 / 11 dai – 8 dai for Hv70 and 11 dai for Hv1427. One spot mark * represents statistical difference at $0.01 \leq p < 0.05$; Two spot marks ** represent statistical difference at $p < 0.01$ for the T-Student; The vertical lines represent the standard deviation (SD).

4 DISCUSSION

4.1 *Hemileia vastatrix* genome size variation

Rust fungi (Pucciniales) are considered one of the most important orders among Basidiomycota. This order comprises obligate biotrophic pathogens with complex life cycles and large genome sizes with a repetitive nature (Aime *et al.*, 2017). The average estimation for rust fungi genomes is 380Mbp (Tavares *et al.*, 2014), about six times bigger than the most recently updated average size for fungal genomes - 60Mbp (Talhinhas *et al.*, 2021). These estimations have been made through flow cytometry, a technique that is considered one of the most effective to obtain genome size measurements in a fast and inexpensive way (Talhinhas *et al.*, 2021). In this work, 21 isolates of *Hemileia vastatrix* were analyzed through the employment of flow cytometry, using *Rhamnus alaternus* and *Solanum lycopersicum* as DNA standards, to assess the average size for the species, as well as the intraspecific variability among isolates and contrasting pathotypes.

Considering the studied isolates, our results estimate the average genome size for *H. vastatrix* at 789Mbp. The obtained average value is in accordance with previous studies (Tavares *et al.*, 2014; Ramos *et al.*, 2015). Tavares *et al.* (2014) analyzed 32 samples of 30 rust fungi species, by flow cytometry, to support the evidence of genome expansion among Pucciniales species, where the occurrence of very large genome sizes was observed. The results varied over 10-fold, from 70Mbp to 893Mbp, with an average genome size value of 380.2Mbp. Among the samples under study, *Hemileia vastatrix* was revealed to be one of the species with the largest genome (796.8Mbp). In another study, six rust species such as *Melampsora pulcherrima*, *Puccinia behenis*, *Puccinia cichorii*, *Puccinia pimpinellae*, *Puccinia vincae* and *Uromyces dianthi*, had their genome size estimated by flow cytometry and the results obtained were 215.6Mbp, 182.1Mbp, 334.8Mbp, 321.6Mbp, 566.4Mbp and 418.8Mp, respectively (Talhinhas *et al.*, 2015). In comparison to *H. vastatrix*, all these six rust species have much smaller genomes which can be up to four times smaller, as is the case of *Puccinia behenis* with just 182.1Mbp. Otherwise, the *Bidens pilosa* rust pathogen (*Uromyces bidentis*) presents the largest Basidiomycota genome reported employing flow cytometry with 2489Mbp (Ramos *et al.*, 2015), three times larger than *Hemileia vastatrix* genome.

Rust fungi genomes have been reported, among all plant pathogens, to comprise many repetitive portions of DNA – including microsatellites, transposable elements, and other tandem repeats (Aime *et al.*, 2017). Repetitive elements tend to cover, approximately, 30% to 50% of the whole genome of 11 analysed rust species, depending on the

sequencing approach. However, *H. vastatrix* genome seems to present an even higher proportion of repetitive elements. The first whole genome sequencing of *H. vastatrix* revealed an estimated genome size of 333Mbp with 74.4% of repetitive elements in the genome (Cristancho *et al.*, 2014). The second work to assemble the whole genome of *H. vastatrix* combined two sequencing approaches for race XXXIII (Hv33), determining the genome size as 547Mbp with 82% of repetitive elements, among which 43.6% corresponded to transposable elements, being *Gypsy* retrotransposons the most predominant (Porto *et al.*, 2019). More recently, a whole-genome assembly of *H. vastatrix* (Race I) was made with PacBio HiFi sequencing obtaining a highly contiguous 747.98Mbp genome where transposable elements cover 81.66% of the chromosomal scaffolds (Ángel *et al.*, 2023). Thus, the retrotransposon content could be a major driver for the larger genome size of *H. vastatrix*. According to Spanu (2012), as Pucciniales are mainly obligate biotrophic pathogens, their genomes tend to overcome other fungi in genome size due to retrotransposon proliferation.

In addition to the large genome size, significant differences among rust isolates were detected in our study, reaching about 166Mbp, which suggests dynamic events of genome expansion/contraction. However, when analyzing al the level of individual isolates, genome expansion does not appear to have a direct proportional relation with the number of virulence genes. In this work, two of the isolates with the smallest genome size estimations, Hv999 (v2,4,5,6/XXVIII; 719Mbp) and Hv1321 (v5,6,7,8,9/XXIX; 721Mbp) have higher virulence profiles than the isolate Hv70 (v2,4,5/XXIV; 879Mbp) which has the biggest genome determination. Furthermore, Hv999 was shown to present signals of introgression from a *H. vastatrix* lineage infecting diploid coffee hosts (Silva *et al.*, 2018) and Hv1321 is considered to be one of the most virulent races, apart from Hv3302 and Hv3305, as it infects Híbrido de Timor (HDT) derived varieties (Várzea and Marques, 2005). The evolutionary progression of virulent pathotypes does not appear to be directly related to the increase of genome size.

Nonetheless, multiple studies suggested that polyploidy events and transposable elements proliferation, such as transposons and retrotransposons, could be possibly correlated with intraspecific wide genome variability (Duplessis *et al.*, 2011; Kelkar and Ochman, 2012; Carvalho *et al.*, 2014). The repetitive nature of *H. vastatrix* genome could be associated with these events. Moreover, transcriptomic studies and protein annotation has shown that transposable elements not only are present in the genome, but are expressed (Talhinhas *et al.*, 2014). More and more, previous transcripts annotated as fungal conserved hypothetical proteins are now described as integrant part of transposable elements structure in *H. vastatrix* genome (Talhinhas *et al.*, 2014). Many

studies have been focused on functional analysis of retrotransposons, demonstrating how their location can influence gene expression, plant-pathogen interactions and genome evolution towards adaptative populations by horizontal and vertical transfer (Daboussi, 1997; Daboussi and Capy, 2003; Muszewska *et al.*, 2011; Chadha and Sharma, 2014).

In this context, this work intended to address the putative effect of retrotransposons in genomic variability. Moreover, since retrotransposons can be effectively expressed, it was also intended to better understand how they are expressed during coffee-H. vastatrix interactions, and if differential expression patterns can be related with virulence.

4.2 Intraspecific retrotransposon genetic variability

Genome resources are indispensable for understanding the impact and role of retrotransposons in the diversity and virulence of plant pathogens. Until recently, only two draft genome assemblies were available for *H. vastatrix*, HvCat (Cristancho *et al.*, 2014) and Hv33 (Porto *et al.*, 2019), reporting an extremely high proportion of repetitive regions (ca. 74.4% and 82%, respectively). These studies provided important new knowledge on the content of repetitive elements, describing the predominance of LTR elements (ca. 38% in both HvCat and Hv33), with a larger representation of Gypsy-like elements (more than 60%), and also the presence of non-autonomous long interspersed nuclear elements (LINEs, ca 2%). Nevertheless, the partial annotation provided by these resources were still of limited assistance. In their recent study, Orozco-Arias *et al.* (2022) characterized the transposable elements within these two-coffee leaf rust genome assemblies, identifying LTR-RTs as the vast majority of all the transposable elements detected (reaching 91.7 % in Hv33) and suggesting their significant contribution to the size of the *H. vastatrix* genome. In particular, elements of the Gypsy superfamily corresponded to 93.3% of all retrotransposons identified, contributing to 51.7% (281.2Mbp) of the Hv33 assembly size. Two new *Copia* lineages (Labe and Mapi), and three new *Gypsy* lineages (Soroa, Baco, CO-HUI) were also identified based on phylogenetic analysis, in which the CO-HUI lineage is the most numerous (Orozco-Arias *et al.*, 2022). In addition, analysis of all raw reads revealed that LTR-RTs have a representation higher than the estimation in the assemblies (53.3% vs 38.7% and 37.7% in HvCat and Hv33, respectively), suggesting that the contribution of TEs to genome size must be larger than assumed (Orozco-Arias *et al.*, 2022).

In this study, to assess genomic variability in *H. vastatrix*, three expressed sequence tags (ESTs) predicted as putative retrotransposons were selected to investigate genetic variation and copy number among coffee rust isolates. Annotation of the retrotransposon

candidates (*R190*, *R1057*, *R2407*) using the *H. vastatrix* genome resources available identified them as belonging to the Gypsy superfamily. However, homology searches of our candidate retrotransposons against the transposable element (TE) library (PucciDB) constructed by Orozco-Arias *et al.* (2022), based on de novo re-annotation of 21 Pucciniales genomes, provided a new classification. Contrary to the previous annotation, retrotransposon *R190* was classified as a Mapi type (Copia superfamily), *R1057* as a TRIM type, and the *R2407* as a LARD type, being these last two lineages described as non-autonomous LTR-RT elements. The Copia superfamily is usually described as being half as successful as the Gypsy superfamily in dominating fungal genomes, with the exception of *Saccharomyces cerevisiae* (Muszewska *et al.*, 2011). However, some rust fungi show high proportion of Copia elements, such as *Uromyces viciae-fabae* (86.7%), *Puccinia sorghi* (60%), and *Puccinia novopanici* (56.7%), which suggest that this superfamily is also able to increase its copy number in Pucciniales (Orozco-Arias *et al.*, 2022). The Mapi retrotransposon seems to be a specific lineage of the Pucciniales and was described as being more involved in co-expression with genes through gene-to-gene interaction, dynamically participating in genome evolution (Orozco-Arias *et al.*, 2022). On the other hand, non-autonomous elements are described as lacking most or all of the coding sequences, but are still mobilized by the machinery of functional (autonomous) elements, which provide the required proteins to transpose in trans (Kozyol *et al.*, 2015). The TRIM (terminal repeat retrotransposon in miniature) and LARD (large retrotransposon derivative) elements entirely lack reading frames for retrotransposon proteins, and thus cannot be conventionally classified as Copia or Gypsy like (Sampath *et al.*, 2015). TRIM elements are presumed to be the result of convergent reductive evolution from autonomous elements in different taxonomic groups, and based on characteristic signatures it has been suggested that TRIM could be deletion derivatives of Copia-like elements and LARD of Gypsy-like elements, which could also be the respective candidate autonomous partners (Sampath *et al.*, 2015). These features account for their large distribution across several taxa genomes, in particular plants, where seem to be involved in active restructuring of plant genomes by afecting their promoters and coding regions (Gao *et al*., 2016). In fungi, there is still a great gap of information on the effect and role of these elements, apart from their potential as molecular markers (Santana *et al.*, 2013).

Copy number estimation of *R190*, *R1057* and *R2407* in 21 isolates of *H. vastatrix* revealed a wide content variation: *R190* ranged from 495 copies in Hv92 to 3486 copies in Hv3305; *R1057* varies from 1 single copy in some isolates to 103 copies in Hv1065; *R2407* goes from the minimum of 304 copies in Hv999a to the maximum value of 4457

copies in Hv3305. Given the large amount of copies of the *R190* and *R2407* retrotransposons in our results, that reach up to 3000 copies and more, these retrotransposons could have a contribution to the large genome size, as well as to the genetic variability due to high proliferation levels in the genome.The extremely high capacity of expansion of the *R190* could be expected as this is an autonomous LTR-RT element, but not for *R2407*, since as an non-autonomous element it depends on an autonomous functional partner to transpose. Nevertheless, this LARD LTR-RT element could reach the highest copy number detected in our study, suggesting that it evolved an efficient "parasitism" mechanism to increase genome invasion. Kalendar *et al.* (2004) first described LARD elements in 13 plant species and estimated their copy number by dot blot as 1300, a much lower abundance than that found in *H. vastatrix*. In contrast, the moderate activity of *R1057* is in agreement with the copy number range found in more than 50 plant species (Witte *et al.*, 2001; Sampath *et al.*, 2015; Gao *et al.*, 2016). Orozco-Arias *et al.* (2022) estimated the copy number of transposable elements in the Hv33 assembly using the 80-80-80 classification rule, detecting the highest copy number value in the CO-HUI lineage (12.380), followed by V_Clade (1,803) and less than 200 copies in the Mapi lineage. Our results may provide a more realistic range since the high sensitivity and specificity of qPCR can translate into a significantly more accurate measurement. Although the use of genome assemblies to estimate the proportion of retrotransposons in fungi is well reported, the current data deriving from experimental-based estimations of gene copy number is still scarce. To our knowledge, no study of copy number estimation by qPCR has been reported in fungi, not even for protein coding genes, and the studies dwelling on this matter are focused on the determination of gene copy number variants associated with human disorders (Doleschall *et al.*, 2022).

In this work, no direct correlation could be established between retrotransposon copy number data and the genome size. Rust isolates with bigger genomes do not correspond to the representative data of the biggest copy number determination for each retrotransposon. The copy number estimation is very variable among the 3 retrotransposons. For example, the isolate Hv3305 ($v_{1,2,4,5,6,7,8,9,+}$) described as one of the most virulent pathotypes has the greatest number of copies for 2 out of 3 retrotransposons (*R190* and *R2407*) although its genome size determination (792Mbp) remains 87Mbp below the maximum determination (879Mbp). The data resultant from the qPCR may be representative of the rapid proliferation and diversification of the retrotransposons within *H. vastatrix* genome, with a wide range of both copy number and genetic variation.

The genetic variation of retrotransposon copies was investigated by cloning and sequencing *R190*, *R2407* and *R1057* for 9-10 Hv isolates with contrasting pathotypes. Sequence data analysis and haplotype networks revealed a very high genetic variability for all retrotransposons, either among or within isolates. In fact, sequenced copies were quite divergent and all different, with none shared among rust isolates. In all retrotransposons studied, for each *E. coli* colony sequenced from the same isolate, a different haplotype was retrieved, leading to the assumption that many more retrotransposon copies, than those detected, must exist. This apparent remarkable RT activity with a high mutational rate could represent an important mechanism to generate diversity in an assumed asexual pathogen as *H. vastatrix*. Nowadays, it is well recognized and supported by continuously growing evidence during the last decade that transposable elements are powerful drivers of fungal genetic diversity and genome evolution (Lorraine *et al.*, 2021).

In agreement with the copy number data, *R190* and *R2407* haplotype networks show a very complex pattern of diversified copies. The higher copy number is also associated to a higher mutation level. Indeed, more proliferative retrotransposons tend to have more complex diversification patterns (Muszewska, *et al.*, 2011; Bonchev and Parisod, 2013; Chadha and Sharma, 2014) Nonetheless, it is interesting to note that although *R2407* is more proliferative than *R190*, its copies harbour less substitutions and indels, which may be related with its non-autonomous status as the impact of autonomous or non-autonomous retrotransposons can be dramatically different in the host genome architecture, by inducing structural variation within populations (Lorrain *et al.*, 2021). Specifically, the *R190* median-joining network suggests rapid divergence and expansion, as different copies from the same isolate can differ in more than 30 SNPs and up to 84 bp indels. In contrast, the *R1057* haplotype network shows a more stable structure with a central haplotype, from where all the other haplotypes may have diverged possibly through continuous and steady mutations. In the filamentous fungus *Magnaporthe oryzae*, a positive linear correlation was found between copy number and transposition rate, measured as mutation rate, of two LTR-retrotransposons (*Pyret* and *MGLR3*), which were identified as the most unstable (Chadha and Sharma, 2014). Moreover, Chadha and Sharma (2014) showed that LTR-retrotransposons *Pyret* and *MAGGY* are key components of high genomic instability induced by heat shock and copper stress that mediates genomic rearrangements in *M. oryzae*. According to these authors, these environmental cues induce local bursts of TEs, generating high genetic diversity and subsequently new strains of the fungal pathogen.

On the other hand, effector genes of many plant pathogens are located either within or in proximity to transposable element-rich regions (Lovelace *et al.*, 2023). The strong correlation between effector location and transposable element abundance led to the proposal that these elements contribute to genetic variability by facilitating the swift evolution of effector-rich regions in response to plant immunity, driving pathogen adaptation (Seidl and Thommas, 2017). In the rice blast fungus, diversification by point mutations and genomic instability through deletion or insertion events of the *Avr-Pita* effector enables *Magnaporthe oryzae* isolates to overcome the recognition by the Pita immune receptor in rice, leading to resistance breakdown (Dai *et al.*, 2010). Indeed, insertion of transposable elements in the *Avr-Pita* and *Avr-Pib* regulatory regions result in gain of virulence on rice resistant plants (Kang *et al.*, 2001; Li *et al.*, 2023). Our results suggest that *R190*, *R2407* and *R1057* are highly active in the *H. vastatrix* genome, and that in particular *R190* and *R2407* could be responsible for genomic instability. Although no apparent relation of genetic variability with isolate/pathotype was found, Hv3305, known to having break down resistance of one of the original HDT major sources of resistance (CIFC 832/2) (Silva *et al.*, 2022), shows the highest level of genomic proliferation and copy polymorphism of *R190* and *R2407*. Such pattern could mean that these retrotransposons have an active role in *H. vastatrix* adaptive evolution, given the previously suggested contribution of retrotransposons in *H. vastatrix* lineage differentiation and consequent local host-adaptation/specialization (Rodrigues *et al.*, 2022).

4.3 Expression profiling of the selected retrotransposons in coffee-rust compatible interactions

In fungi, the majority of TE copies are transcriptionally inactive except under stress conditions (Fouché *et al.*, 2020; Lorrain *et al.*, 2021). The mechanism by which some TEs can escape silencing, which is largely governed through epigenetic control, is rather unknown, but distinct stress conditions can induce de-repression of different sets of TEs (Fouché al., 2020).

Host infection constitutes the major stress factor and evolutionary driver in the life cycle of rust fungal pathogens (Rodrigues *et al.*, 2022). In this work, the expression profiles of the three selected retrotransposon were analyzed in two rust isolates with different virulence profiles under a compatible interaction, along the infection process, using qPCR to assess if these could be potentially involved in virulence differentiation. Quantitative real-time PCR (qPCR) has become the most useful research tool to quantify mRNA levels providing a highly accurate and sensitive method to quantify gene expression, even at a single-cell level. To obtain reliable results from the quantification,

a previous normalisation step is required through the quantification of the gene of interest relatively to an internal control (Huggett *et al.*, 2005; Derveaux *et al.*, 2010). Housekeeping genes as reference genes are the most appropriate genes to normalise qPCR data (Dheda *et al.*, 2005). In this study, previously established reference genes for the condition *in planta* [*GADPH* (glyceraldehyde 3-phosphate dehydrogenase), *40S* (40S ribosomal protein) and *CytIII* (cytochrome c oxidase subunit III)] were used (Vieira *et al.*, 2011), and samples from 1 dai were used as control. Retrotransposon expression analyses revealed a completely distinct pattern between the isolates. While for the compatible interaction Hv1427 (race II [v_5]; host *C. arabica* var. Caturra 19/1 [S_H5]) all retrotransposons were upregulated in comparison to 1 dai throughout the infection process, for the compatible interaction Hv70 (race XXIV [$v_{2,4,5}$]; host H152/3 [$S_H2,4,5$]), all retrotransposons were mainly downregulated, particularly in the later stages of infection. This suggests a specific response of these retrotransposons to infection depending on the isolate genotype.

For the isolate Hv70, the retrotransposons were slightly upregulated in only one early time-point of the five studied along the infection process. From 4 dai (days after inoculation; where haustoria differentiation is alrealy largely distributed) onwards, all retrotransposons were downregulated in comparison to 1 dai. Moreover, the 4 dai-stage of infection is where a higher level of repression is detected. Therefore, and although this relative expression values can only be considered as higher or lower than the expression measured at 1 dai, the retrotransposons studied seem to be de-repressed in the early stages of infection, for this isolate. In fact, the early stages are considered the most stressful period for the pathogen since it has to face the host immune responses and impose growth under limited nutrient conditions (Fouché *et al.*, 2020). In contrast, for the Hv1427 under a compatible interaction, all three retrotransposons were highly expressed along the course of infection in comparison to 1 dai, with very similar values of relative expression, except for *R2407* for which expression reaches extremely high fold changes. In this latter case, the retrotransposon activation is particularly strong at 8 dai, where a significant peak of expression is achieved, although at a stage of prolific colonization the pathogen should be submitted to a lesser level of stress. Nevertheless, this isolate present the slowest infection progression. Overall, our results clearly reveal that this set of retrotransposons respond differently to infection led by rusts with different virulence profiles, indicating that genetic background plays a role in retrotransposon activation. It would seem that these retrotransposons could have a more relevant role in virulence for Hv 1427 than for Hv70. We also find descripancies with the expression profiles inferred from the original 454-data, based on EST relative abundance, for these

same retrotransposons in a different Hv isolate under compatible interaction (Hv178a) (race XIV [$v_{2,3,4,5}$]; host H147/1 [S_H2,3,4,5]) (Talhinhas *et al.*, 2014). None of the retrotransposon expression profiles reveal a similar pattern with our relative expression data for both our isolates: Hv00190 profile [Uredospore germination (1 dai) = Appressoria formation (2 dai) > Haustoria formation (3 dai)]; Hv01057 profile [Uredospore germination (1 dai) < Appressoria formation (2 dai) > Haustoria formation (3 dai)]; and Hv02407 profile [Uredospore germination (1 dai) = Appressoria formation (2 dai) < Haustoria formation (3 dai)] do not match our data at any time-point. Although considering that different techniques are being compared, with different levels of resolution for transcript quantification, this provides an addional indication of the possible determination of the retrotransposon response to infection by the pathogen genetic background.

In *Zymoseptoria tritici* (Mycosphaerellaceae; synonym: *Septoria tritici*), a fungal wheat pathogen that causes septoria leaf blotch, Fouché *et al.* (2020) found that retrotransposon expression under host infection differed between strains, suggesting distinct epigenetic environments between different strains or differences in the genomic locations of specific elements, that could affect distance to genes or proximity to different genes. In addition, a joint epigenetic regulation of genes and TEs in close physical proximity, in response to stress, could be argued, as genes close to TEs were upregulated during early infection consistently with the de-repression observed for TEs (Fouché *et al.*, 2020). This pattern of expression agrees with that detected for our retrotransposons in Hv70. However, *H. vastatrix* has a quite distinctive infection strategy and biotrophic pathogens may be subjected to more prolonged periods of stress, which would be more consistent with the retrotransposons profile in Hv 1427. Even considering the different speed of infection progression between the isolates, the significantly differential pattern of retrotransposon expression observed may relate to the relevance of the genes co-expressed for overcoming host defenses in the specific virulences of Hv1427 and Hv70.

Effector genes are recognizedly major players in virulence and are often located in TE-rich regions, becoming expressed upon plant infection. This raises the possibility that the de-repression of TEs interacts with the expression of effectors (Lorrain *et al.*, 2021). Effector gene expression is known to be epigenetically regulated in plant pathogens (Soyer *et al.*, 2014). The transcriptome analysis of four strains of *Z. tritici* (1A5, 1E4, 3D1 and 3D7) exposed to host infection stress, revealed that the expression profiles of TEs nearby the Avr3D1 effector locus matched the effector expression across infection stages. In fact, the presence of strongly silenced TEs, which do not respond to to stress triggers caused by infection, likely contributed to reduced expression of Avr3D1 in the

virulent strain 3D7, leading to reduced virulence on a specific wheat cultivar (Fouché at al., 2020). More recently, Porquier *et al.* (2021) proposed LTR-retrotransposons as pathogenicity factors that manipulate host plant gene expression in *Botrytis cinerea* (Sclerotiniaceae), a necrotrophic fungus that infects various important crops such as tomato. In this pathosystem, *B. cinerea* small RNAs (BcsRNAs) transcribed by LTR-RT are delivered to plant cells to supress host immunity gene expression. The analysis of multiple *B. cinerea* strains revealed that the repression of retrotransposon derived BcsRNAs leads to less aggressiveness on tomato host species. Moreover, transgene-induced production of retrotransposon BcsRNAs in a non-retrotransposon BcsRNAs producing *B. cinerea* strain led to enhanced aggressiveness through the promoted expression of the Gypsy-like *B. cinerea* retrotransposon *BcGypsy3* mRNAs (Porquier *et al.*, 2021).

The data provided by our study on the differential expression of a set of LTR-retrotransposons in different coffee rust pathotypes is very promising on the possible strong contribution of these elements for *H. vastatrix* virulence. However, the association of virulence in coffee leaf rust with the upregulated expression of transposable elements, or their interconnection with effectors still requires a large effort of in-depth analyses in the future.

5 CONCLUSION

In summary, our results reinforce the status of *Hemileia vastatrix* as one of the rust fungi species with the largest genomes, among Pucciniales, showing high variation in genome size among isolates. The possible involvement of retrotransposons on the intra-specific genome size variation could not be demonstrated, as there is no evidence in our data that supports a correlation between retrotransposon proliferation and the increment of the genome size among various isolates. However, this study presents for the first time detailed data on three LTR- retrotransposons (a Copia element and two non-autonomous elements) in different *H. vastatrix* isolates, suggesting a rapid proliferation and diversification of the retrotransposons within *H. vastatrix* genome that could represent an important mechanism to generate genetic diversity and driver of genome evolution in *H. vastatrix*. In addition, clearly distinct expression profiles of these retrotransposons were found between two *H. vastatrix* pathotypes, suggesting an involvement of these elements in their specific virulence response to overcome coffee defenses. This raises the possibility of these elements co-expressing with effector genes or acting directly as pathogenicity factors, which is an exciting topic that should be considered for further investigation.

6 REFERENCES

Aime, M. C., McTaggart, A. R., Mondo, S. J., and Duplessis, S. (2017) 'Phylogenetics and Phylogenomics of Rust Fungi', *Advances in Genetics*, 100, pp. 267–307. doi: 10.1016/bs.adgen.2017.09.011.

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

Alhudaib, K. and Ismail, A.M. (2024) 'First Occurrence of Coffee Leaf Rust Caused by *Hemileia vastatrix* on Coffee in Saudi Arabia', *Microbiology Research*, 15(1), pp.164-173.

Alzohairy, A.M., Gyulai, G., Ramadan, M.F., Edris, S., Sabir, J.S., Jansen, R.K., Eissa, H.F. and Bahieldin, A. (2014) 'Retrotransposon-based molecular markers for assessment of genomic diversity', *Functional Plant Biology*, 41(8), pp. 781–789. doi: 10.1071/FP13351.

Ángel C., C. A., Marín-Ramírez, G. A. and Maldonado, C. E. (2023) 'Genome sequence of *Hemileia vastatrix* Berk. and Br. (Race I), the causal agent of coffee leaf rust, isolate from Risaralda, Colombia', *Microbiology Resource Announcements*. 12(11), pp.e00444-23

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P., Anzueto, F., Hruska, A.J. and Morales, C. (2015) 'The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions', *Food Security*, 7(2), pp. 303–321. doi: 10.1007/s12571-015-0446-9.

Bekele, W. (2010) 'The history of the kingdom of Kaffa: The birthplace of coffee 1390-1935', *ARCCIKCL, Hawassa*.

Bekele, K. B., Senbeta, G. A., Garedew, W., Caixeta, E. T., Ramírez-Camejo, L. A., and Aime, M. C. (2022) 'Genetic diversity and population structure of *Hemileia vastatrix* from Ethiopian Arabica coffee', *Archives of Phytopathology and Plant Protection*, 55(13), 1483–1503. doi: 10.1080/03235408.2021.1983385.

Bandelt, H.-J., Forster, P. and Röhl, A. (1999) 'Median-Joining Networks for Inferring Intraspecific Phylogenies', *Molecular Biology and Evolution*, 16(1), pp. 37–48. doi: 10.1093/oxfordjournals.molbev.a026036.

Batista, D., Guerra-Guimarães, L., Talhinhas, P., Loureiro, A., Silva, D.N., Gonzalez, L., Pereira, A.P., Vieira, A., Azinheira, H.G., Struck, C. and Silva, M.C. (2010, October)

‘Analysis of population genetic diversity and differentiation in *Hemileia vastatrix* by molecular markers’, *In Proceedings of the 23rd International Conference on Coffee Science* (pp. 3-8).

Bettencourt, A. and Rodrigues Jr, C. (1988) ‘Principles and practice of coffee breeding for resistance to rust and other diseases’, *Coffee Agronomy*, 4, pp. 199–234.

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., & Moscou, M. J. (2014) ‘Nonhost resistance to rust pathogens–a continuation of continua’, *Frontiers in Plant Science*, 5, 664. doi: 10.3389/fpls.2014.00664.

Bewick, A.J., Hofmeister, B.T., Powers, R.A., Mondo, S.J., Grigoriev, I.V., James, T.Y., Stajich, J.E. and Schmitz, R.J. (2019) ‘Diversity of cytosine methylation across the fungal tree of life’, *Nature Ecology and Evolution*, 3(3), pp. 479–490. doi: 10.1038/s41559-019-0810-9.

Bock, K. R. (1962) ‘Dispersal of uredospores of *Hemileia vastatrix* under field conditions’, *Transactions of the British Mycological Society*, 45(1), pp. 63–74. doi: 10.1016/s0007-1536(62)80035-7.

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985) ‘Ty elements transpose through an RNA intermediate’, *Cell*, 40(3), 491-500.

Boeke, D. and Corces, V. G. (1989) ‘Trranscription and reverse transcription of retrotransposons’, *Annual Review of Microbiology*, 43, pp. 403–434.

Bonchev, G. and Parisod, C. (2013) ‘Transposable elements and microevolutionary changes in natural populations’, *Molecular Ecology Resources*, 13(5), pp. 765–775. doi: 10.1111/1755-0998.12133.

Borgognone, A., Castanera, R., Morselli, M., Lopez-Varas, L., Rubbi, L., Pisabarro, A.G., Pellegrini, M. and Ramírez, L. (2018) ‘Transposon-associated epigenetic silencing during Pleurotus ostreatus life cycle’, *DNA Research*, 25(5), pp.451-464. doi: 10.1093/dnares/dsy016.

Bourne, E.C., Mina, D., Gonçalves, S.C., Loureiro, J., Freitas, H. and Muller, L.A. (2014) ‘Large and variable genome size unrelated to serpentine adaptation but supportive of cryptic sexuality in *Cenococcum geophilum*’, *Mycorrhiza*, 24, pp. 13–20. doi: 10.1007/s00572-013-0501-3.

Braun, E. J. and Howard, R. J. (1994) ‘Adhesion of fungal spores and germlings to host plant surfaces’, *Protoplasma*, 181, pp. 202–212. doi: 10.1007/BF01666396.

Bremer, B. (1996) ‘Combined and separate analyses of morphological and molecular data in the plant family Rubiaceae’, *Cladistics*, 12(1), pp. 21–40. doi: 10.1006/clad.1996.0003.

Bustin, S. A. *et al.* (2005) ‘Quantitative real-time RT-PCR - A perspective’, *Journal of Molecular Endocrinology*, 34(3), pp. 597–601. doi: 10.1677/jme.1.01755.

Cabral, P.G.C., Maciel-Zambolim, E., Oliveira, S.A.S., Caixeta, E.T. and Zambolim, L. (2016) ‘Genetic diversity and structure of *Hemileia vastatrix* populations on *Coffea* spp.’, *Plant Pathology*, 65, pp. 196–204. doi: 10.1111/ppa.12411.

Carvalho, A. and Ferwerda, F.P. (1969) ‘*Coffea arabica* L. and *Coffea canephora* Pierre ex Froehner In: Outlines of perennial crops breeding in the tropics’, *Miscellaneous Papers*, 4, pp. 189–241.

Carvalho, G.M.A., Carvalho, C.R., Barreto, R.W. and Evans, H.C. (2014) ‘Coffee rust genome measured using flow cytometry: Does size matter?’, *Plant Pathology*, 63(5), pp. 1022–1026. doi: 10.1111/ppa.12175.

Carvalho, R., Silva, E., Loureiro, J., Ramos, A.P. and Talhinhas, P. (2018) ‘Characterisation of *Puccinia hemerocallidis* causing the first outbreak of daylily rust in Europe’, *Revista das Ciências Agrárias*, 41 (Especial): 110-115. doi: 10.19084/rca.17075.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A.G., Grigoriev, I.V. and Stajich, J.E. (2016) ‘Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles’, *PLoS Genetics*, 12(6), 1006108. doi: 10.1371/journal.pgen.1006108.

Cerda, R., Avelino, J., Gary, C., Tixier, P., Lechevallier, E. and Allinne, C. (2017) ‘Primary and secondary yield losses caused by pests and diseases: Assessment and modeling in coffee’, *PLoS ONE*, 12(1), e0169133. doi: 10.1371/journal.pone.0169133.

Chadha, S. and Sharma, M. (2014) ‘Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen *Magnaporthe oryzae*’, *PLoS ONE*, 9(4):e0094415. doi: 10.1371/journal.pone.0094415.

Charrier, A. and Berthaud, J. (1985) ‘Botanical classification of coffee. In: Clifford M.N., Willson K.C. (eds) Coffee.’, Springer, pp. 13–47. doi: 10.1007/978-94-011-2326-6_2.

Charrier, A. and Eskes, A. B. (2004) 'Botany and Genetics of Coffee', in Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers, pp. 25–56. doi: 10.1002/9783527619627.ch2.

Chevalier, A. (1947) 'Les caféiers du globe, fasc. 3: systématique des caféiers et faux-caféiers maladies et insectes nusibles.', Encyclopédie Biologique.

Chinnappa, C. C. and Sreenivasan, M. S. (1968) 'Cytology of *Hemileia vastatrix*', Caryologia, 21(1), pp. 75–82. doi: 10.1080/00087114.1968.10796284.

Conagin, C.H. and Mendes, A.J.T. (1961) 'Pesquisas citológicas e genéticas em três espécies de Coffea: auto-incompatibilidade em *Coffea canephora* Pierre ex Froehner', *Bragantia*, 20, pp.788-804.

Cook, D. E., Mesarich, C. H. and Thomma, B. P. H. J. (2015) 'Understanding Plant Immunity as a Surveillance System to Detect Invasion', *Annual Review of Phytopathology*, 53, pp. 541–563. doi: 10.1146/annurev-phyto-080614-120114.

Cristancho, M.A., Botero-Rozo, D.O., Giraldo, W., Tabima, J., Riaño-Pachón, D.M., Escobar, C., Rozo, Y., Rivera, L.F., Durán, A., Restrepo, S., Eilam, T., Anikster, Y. and Gaitán, A.L. (2014) 'Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales', *Frontiers in Plant Science*, 5, 594. doi: 10.3389/fpls.2014.00594.

d'Oliveira, B. 1954. As ferrugens do cafeeiro. *Revista do Café Português*, 1:5-13.

Daboussi, M. J. and Capy, P. (2003) 'Transposable Elements in Filamentous Fungi', *Annual Review of Microbiology*, 57, pp. 275–299. doi: 10.1146/annurev.micro.57.030502.091029

Daboussi, M. J. (1996) 'Fungal transposable elements: Generators of diversity and genetic tools', *Journal of Genetics*, 75(3), pp. 325–339. doi: 10.1007/BF02966312.

Daboussi, M. J. (1997) 'Fungal transposable elements and genome evolution', *Genetica*, 100(1–3), pp. 253–260. doi: 10.1007/978-94-011-4898-6_25.

Daboussi, M.J., Daviere, J.M., Graziani, S. and Langin, T. (2002) 'Evolution of the *Fot1* transposons in the genus *Fusarium*: Discontinuous distribution and epigenetic inactivation', *Molecular Biology and Evolution*, 19, pp. 510–520. doi: 10.1093/oxfordjournals.molbev.a004106.

Dai, Y., Jia, Y., Correll, J., Wang, X. and Wang, Y. (2010) 'Diversification and evolution of the avirulence gene *AVR-Pita1* in field isolates of *Magnaporthe oryzae'*, *Fungal Genetics and Biology*, 47(12), pp. 973–980. doi: 10.1016/j.fgb.2010.08.003.

Davis, A. P. (2003) 'A new combination in *Psilanthus* (Rubiaceae) for Australasia, and nomenclatural notes on *Paracoffea*', *Novon*, 13(2), pp. 182–184. doi: 10.2307/3393514.

Davis, A.P., Govaerts, R., Bridson, D.M. and Stoffelen, P. (2006) 'An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae)', *Botanical Journal of the Linnean Society*, 152(4), pp. 465–512. doi: 10.1111/j.1095-8339.2006.00584.x.

Davis, A.P., Tosh, J., Ruch, N. and Fay, M.F. (2011) 'Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*', *Botanical Journal of the Linnean Society*, 167(4), pp. 357–377. doi: 10.1111/j.1095-8339.2011.01177.x.

Davis, A.P., Gole, T.W., Baena, S. and Moat, J. (2012) 'The Impact of Climate Change on Indigenous Arabica Coffee (*Coffea arabica*): Predicting Future Trends and Identifying Priorities', *PLoS ONE*, 7(11), e0047981. doi: 10.1371/journal.pone.0047981.

Delplace, F., Huard-Chauveau, C., Berthomé, R. and Roby, D. (2022) 'Network organization of the plant immune system: from pathogen perception to robust defense induction', *Plant Journal*, 109(2), pp. 447–470. doi: 10.1111/tpj.15462.

Derveaux, S., Vandesompele, J. and Hellemans, J. (2010) 'How to do successful gene expression analysis using real-time PCR', *Methods*, 50(4), pp. 227–230. doi: 10.1016/j.ymeth.2009.11.001.

Dheda, K., Huggett, J.F., Chang, J.S., Kim, L.U., Bustin, S.A., Johnson, M.A., Rook, G.A.W. and Zumla, A. (2005) 'The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization', *Analytical Biochemistry*, 344(1), pp. 141–143. doi: 10.1016/j.ab.2005.05.022.

Diniz, I., Talhinhas, P., Azinheira, H.G., Várzea, V., Medeira, C., Maia, I., Petitot, A.S., Nicole, M., Fernandez, D. and do Céu Silva, M. (2012) 'Cellular and molecular analyses of coffee resistance to *Hemileia vastatrix* and nonhost resistance to *Uromyces vignae* in the resistance-donor genotype HDT832/2', *European Journal of Plant Pathology*, 133(1), pp. 141–157. doi: 10.1007/s10658-011-9925-9.

Doleschall, M., Darvasi, O., Herold, Z., Doleschall, Z., Nyirő, G., Somogyi, A., Igaz, P. and Patócs, A. (2022) 'Quantitative PCR from human genomic DNA: The determination

of gene copy numbers for congenital adrenal hyperplasia and RCCX copy number variation', *PLoS ONE* , 17(12), e0277299. doi: 10.1371/journal.pone.0277299.

Doležel, J., Sgorbati, S. and Lucretti, S. (1992) 'Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants', *Physiologia Plantarum*, 85, pp. 625-631. doi: 10.1111/j.1399-3054.1992.tb04764.x.

Doležel, J. and Bartoš, J. (2005) 'Plant DNA flow cytometry and estimation of nuclear genome size', *Annals of Botany*, 95, pp. 99–110. doi: 10.1093/aob/mci005.

Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L. and Chiu, R. (2011) 'Obligate biotrophy features unraveled by the genomic analysis of rust fungi', *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), pp. 9166–9171. doi: 10.1073/pnas.1019315108.

Etienne, H., Anthony, F., Dussert, S., Fernandez, D., Lashermes, P., & Bertrand, B. (2002) 'Biotechnological applications for the improvement of coffee (*Coffea arabica* L.)', *In Vitro Cellular & Developmental Biology-Plant*, 38(2), pp. 129–138. doi: 10.1079/IVP2001273.

Feschotte, C., and Pritham, E. J. (2007) 'DNA transposons and the evolution of eukaryotic genomes', *Annual Review of Genetics*, 41, pp. 331-368. doi: 10.1146/annurev.genet.40.110405.090448

Fernandez, D., Santos, P., Agostini, C., Bon, M.C., Petitot, A.S., SILVA, M.C., Guerra-Guimarães, L., Ribeiro, A., Argout, X. and Nicole, M. (2004) 'Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*)', *Molecular Plant Pathology*, 5(6), pp. 527–536. doi: 10.1111/J.1364-3703.2004.00250.X.

Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A.S., Loureiro, A., Poulain, J., Da Silva, C., Silva, M.C. and Duplessis, S. (2012) '454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals *in planta*-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction', *Molecular Plant Pathology*, 13(1), pp. 17–37. doi: 10.1111/J.1364-3703.2011.00723.X.

Fink, G. R., Boeke, J. D. and Garfinkel, D. J. (1986) 'The mechanism and consequences of retrotransposition', *Trends in Genetics*, 2, pp. 118–123. doi: 10.1016/0168-9525(86)90200-3.

Flor, H. H. (1942) ‘Inheritance of pathogenicity in *Melampsora lini*’, *Phytopathology*, 32, pp. 653–669.

Fouché, S., Badet, T., Oggenfuss, U., Plissonneau, C., Francisco, C.S. and Croll, D. (2020) ‘Stress-Driven Transposable Element De-repression Dynamics and Virulence Evolution in a Fungal Pathogen’, *Molecular Biology and Evolution*, 37(1), pp. 221–239. doi: 10.1093/molbev/msz216.

Freitag, M. (2017) ‘Histone Methylation by SET Domain Proteins in Fungi’, *Annual Review of Microbiology*, 71, pp. 413–439. doi: 10.1146/annurev-micro-102215-095757.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P. and Firoozabady, E. (1983) ‘Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues’, *Science*, 220(4601), pp. 1049–1051. doi: 10.1126/science.220.4601.1049.

Ganesh, D., Petitot, A.S., Silva, M.C., Alary, R., Lecouls, A.C. and Fernandez, D. (2006) ‘Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (*Hemileia vastatrix*) using real-time quantitative RT-PCR’, *Plant Science*, 170(6), pp. 1045–1051. doi: 10.1016/j.plantsci.2005.12.009.

Gao, D., Li, Y., Kim, K. D., Abernathy, B., & Jackson, S. A. (2016) ‘Landscape and evolutionary dynamics of terminal repeat retrotransposons in miniature in plant genomes’, *Genome Biology*, 17, 7. doi: 10.1186/s13059-015-0867-y

Geletu, K. T. (2006) *Genetic diversity of wild Coffea arabica populations in Ethiopia as a contribution to conservation and use planning*, Ecology and Development Series.

Gichimu, B. M. and Omondi, C. O. (2010) ‘Morphological characterization of five newly developed lines of arabica coffee as compared to commercial cultivars in Kenya’, International *Journal of Plant breeding and Genetics*, 4(4), pp. 238–246. doi: 10.3923/ijpbg.2010.238.246.

Gouveia, M. M. C., Ribeiro, A., Várzea, V. M., and Rodrigues Jr, C. J. (2005) ‘Genetic diversity in *Hemileia vastatrix* based on RAPD markers’, *Mycologia*, 97(2), pp. 396–404. doi: 10.3852/mycologia.97.2.396.

Guerra-Guimarães, L., Chaves, I., Pinheiro, C., Leclerq, C., Resende, M. L. V., Jenny, R., Cândido, R.P. and Várzea, V. (2021) ‘The first insight on the *Hemileia vastatrix* urediniospores proteome’, in 28th International Conference on Coffee Science (ASIC). Montpellier, p. 223. doi: 10.1371/journal.pone.0215598.28.

Guzzo, S. D., Harakava, R. and Tsai, S. M. (2009) ‘Identification of Coffee Genes Expressed During Systemic Acquired Resistance and Incompatible Interaction with

Hemileia vastatrix', *Journal of Phytopathology*, 157(10), pp. 625–638. doi: 10.1111/j.1439-0434.2008.01538.x.

Heath, M. C. and Skalamera, D. (1997) 'Cellular Interactions between Plants and Biotrophic Fungal Parasites', *Advances in Botanical Research*, 24, pp. 195–225.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007) 'qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data', *Genome Biology*, 8, R19. doi: 10.1186/gb-2007-8-2-r19.

Hindorf, H., & Omondi, C. O. (2011) 'A review of three major fungal diseases of *Coffea arabica* L. in the rainforests of Ethiopia and progress in breeding for resistance in Kenya', *Journal of Advanced Research*, 2(2), 109-120. doi: 10.1016/j.jare.2010.08.006

Hoffmann, J. (2018). *The World Atlas of Coffee: From beans to brewing-coffees explored, explained and enjoyed*. Hachette UK.

Hua-Van, A., Langin, T. and Daboussi, M. J. (2001) 'Evolutionary history of the impala transposon in *Fusarium Oxysporum'*, *Molecular Biology and Evolution*, 18, pp. 1959–1969. doi: 10.1093/oxfordjournals.molbev.a003736.

Huda, A., Mariño-Ramírez, L. and Jordan, I. K. (2010) 'Epigenetic histone modifications of human transposable elements: Genome defense versus exaptation', *Mobile DNA*, 1, 2. doi: 10.1186/1759-8753-1-2.

Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005) 'Real-time RT-PCR normalisation; strategies and considerations', *Genes and Immunity*, 6(4), pp. 279–284. doi: 10.1038/sj.gene.6364190.

Jones, J. D. G. and Dangl, J. L. (2006) 'The plant immune system', *Nature*, 444(7117), pp. 323–329. doi: 10.1038/nature05286

Kalendar, R., Vicient, C.M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A. and Schulman, A.H. (2004) 'Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes', *Genetics*, 166(3), pp.1437-1450. doi: 10.1534/genetics.166.3.1437

Kang, S., Lebrun, M. H., Farrall, L., and Valent, B. (2001) 'Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene', *Molecular Plant-Microbe Interactions*, 14(5), pp. 671–674. doi: 10.1094/MPMI.2001.14.5.671.

Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Research*, 30(14), pp.3059-3066.

Keith, L.M., Sugiyama, L.S., Brill, E., Adams, B.L., Fukada, M., Hoffman, K.M., Ocenar, J., Kawabata, A., Kong, A.T., McKemy, J.M. and Olmedo-Velarde, A. (2022) 'First Report of Coffee Leaf Rust Caused by *Hemileia vastatrix* on Coffee (*Coffea arabica*) in Hawaii', *Plant Disease*, 106(2), p.761. doi: 10.1094/PDIS-02-16-0237-PDN.

Keith, L.M., Matsumoto, T.K., Sugiyama, L.S., Fukada, M., Nagai, C., Pereira, A.P., Céu Silva, M. and Várzea, V. (2023) 'First report of the physiological race XXIV of *Hemileia vastatrix* (coffee leaf rust) in Hawai'i', *Plant Disease*, 107(8), p.2528. doi: 10.1094/pdis-03-23-0460-pdn

Kelkar, Y. D. and Ochman, H. (2012) 'Causes and consequences of genome expansion in fungi', *Genome Biology and Evolution*, 4(1), pp. 13–23. doi: 10.1093/gbe/evr124.

Kolmer, J. A., Liu, J. Q. and Sies, M. (1995) 'Virulence and molecular polymorphism in *Puccinia recondita* f. sp. *tritici* in Canada', *Phytopathology*, 85(3), pp. 276–285. doi: 10.1094/Phyto-85-276.

Kosaraju, B., Sannasi, S., Mishra, M. K., Subramani, D., and Bychappa, M. (2017) 'Assessment of genetic diversity of coffee leaf rust pathogen *Hemileia vastatrix* using SRAP markers', *Journal of Phytopathology*, 165(7-8), pp. 486-493. doi: 10.1111/jph.12583

Krishnan, P., Meile, L., Plissonneau, C., Ma, X., Hartmann, F.E., Croll, D., McDonald, B.A., and Sánchez-Vallet, A. (2018) 'Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat', *BMC Biology*, 16, 78. doi: 10.1186/s12915-018-0543-2.

Krug, C. A. and Carvalho, A. (1951) 'The genetics of Coffea - Volume IV', *Advances in Genetics*, 4, pp. 128–158.

Kumar, S., Sudisha, J. and Sreenath, H. L. (2008) 'Genetic relation of Coffea and Indian *Psilanthus* species as revealed through RAPD and ISSR markers', *International Journal of Integrative Biology*, 3(3), pp. 150–158.

Kushalappa, A. C. and Eskes, A. B. (1989) 'Advances in coffee rust research', *Annual Review of Phytopathology*, 27, pp. 503–531. doi: 10.1146/annurev.py.27.090189.002443

Koziol, U., Radio, S., Smircich, P., Zarowiecki, M., Fernandez, C., & Brehm, K. (2015), 'A novel terminal-repeat retrotransposon in miniature (TRIM) is massively expressed in *Echinococcus multilocularis* stem cells'. *Genome Biology and Evolution*, 7(8), 2136-2153. doi: 10.1093/gbe/evv126

Larsson, A. (2014) 'AliView: A fast and lightweight alignment viewer and editor for large datasets', *Bioinformatics*, 30(22), pp. 3276–3278. doi: 10.1093/bioinformatics/btu531.

Leigh, J. W. and Bryant, D. (2015) 'POPART: full-feature software for haplotype network construction', *Methods in Ecology and Evolution*, 6, pp. 1110–1116. doi: 10.1111/2041-210X.12410.

Li, P., Lu, Y.J., Chen, H. and Day, B. (2020) 'The Lifecycle of the Plant Immune System', *Critical Reviews in Plant Sciences*, 39(1), pp. 72–100. doi: 10.1080/07352689.2020.1757829.

Li, J., Lu, L., Li, C., Wang, Q. and Shi, Z. (2023) 'Insertion of transposable elements in *AVR-Pib* of *Magnaporthe oryzae* leading to LOSS of the avirulent function', *International Journal of Molecular Sciences*, 24(21), 15542. doi: 10.3390/ijms242115542

Liebig, T., Jassogne, L., Rahn, E., Läderach, P., Poehling, H.M., Kucel, P., Van Asten, P. and Avelino, J. (2016) 'Towards a collaborative research: A case study on linking science to farmers' perceptions and knowledge on Arabica Coffee Pests and Diseases and Its Management', *PLoS ONE*, 11(8), e0159392. doi: 10.1371/journal.pone.0159392.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method', *Methods*, 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.

Lorrain, C., Oggenfuss, U., Croll, D., Duplessis, S. and Stukenbrock, E. (2021) 'Transposable elements in fungi: coevolution with the host genome shapes, genome architecture, plasticity and adaptation', *In Encyclopedia of Mycology* (pp. 142-155). Elsevier.

Loureiro, A., Nicole, M.R., Várzea, V., Moncada, P., Bertrand, B. and Silva, M.C. (2012) 'Coffee resistance to *Colletotrichum kahawae* is associated with lignification, accumulation of phenols and cell death at infection sites', *Physiological and Molecular Plant Pathology*, 77(1), pp. 23-32. doi: 10.1016/j.pmpp.2011.11.002

Loureiro, A., Tavares, S., Azinheira, H.G., Talhinhas, P., Pereira, A.P., Fernandez, D. and Silva, M.C. (2014) 'Identification and characterisation of Hemileia vastatrix effectors',

In *Proceedings of the 25th International Conference on Coffee Science (ASIC)*, *Armenia*, *Colombia* (pp. 8-13).

Luo, S. and Lu, J. (2017) 'Silencing of Transposable Elements by piRNAs in Drosophila: An Evolutionary Perspective', *Genomics, Proteomics and Bioinformatics*, 15(3), pp. 164–176. doi: 10.1016/j.gpb.2017.01.006.

Lyko, F. (2018) 'The DNA methyltransferase family: A versatile toolkit for epigenetic regulation', *Nature Reviews Genetics*, 19(2), pp. 81–92. doi: 10.1038/nrg.2017.80.

Martin, G. B. (1999) 'Functional analysis of plant disease resistance genes and their downstream effectors', *Current Opinion in Plant Biology*, 2, pp. 273–279. doi: 10.1016/S1369-5266(99)80049-1.

Mayne, W. W. (1932) 'Physiological Specialisation of *Hemileia vastatrix* B. and Br', *Nature*, 129, p. 510. doi: 10.1038/129510a0.

McCain, J. W. and Hennen, J. F. (1984) 'Development of the Uredinial Thallus and Sorus in the Orange Coffee Rust Fungus, *Hemileia vastatrix'*, *Phytopathology*, 74(6), pp. 714–721. doi: 10.1094/phyto-74-714.

McCook, S. (2006) 'Global rust belt: *Hemileia vastatrix* and the ecological integration of world coffee production since 1850', *Journal of Global History*, 1(2), pp. 177–195. doi: 10.1017/S174002280600012X.

McCook, S. and Vandermeer, J. (2015) 'The Big Rust and the Red Queen: Long-term perspectives on coffee rust research', *Phytopathology*, 105(9), pp. 1164–1173. doi: 10.1094/PHYTO-04-15-0085-RVW.

McHale, M.T., Roberts, I.N., Noble, S.M., Beaumont, C., Whitehead, M.P., Seth, D. and Oliver, R.P. (1992) 'CfT-I: an LTR-retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato', *Molecular & General Genetics*, 233(3), pp. 337–347. doi: 10.1007/BF00265429.

McTaggart, A.R., Shivas, R.G., van der Nest, M.A., Roux, J., Wingfield, B.D. and Wingfield, M.J. (2016) 'Host jumps shaped the diversity of extant rust fungi (Pucciniales)', *New Phytologist*, 209(3), pp. 1149–1158. doi: 10.1111/nph.13686.

Muñoz-López, M., and García-Pérez, J. L. (2010) 'DNA transposons: nature and applications in genomics', *Current Genomics*, 11(2), 115-128. doi: 10.2174/138920210790886871

Muszewska, A., Hoffman-Sommer, M. and Grynberg, M. (2011) ‘LTR retrotransposons in fungi’, *PLoS ONE*, 6(12), e29425. doi: 10.1371/journal.pone.0029425.

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. (2019) ‘Transposable elements contribute to fungal genes and impact fungal lifestyle’, *Scientific Reports*, 9(1), 4307. doi: 10.1038/s41598-019-40965-0.

Noronha-Wagner, M. and Bettencourt, A. J. (1967) ‘Genetic study of the resistance of *Coffea* spp. to leaf rust’, *Canadian Journal of Botany*, 45(11), pp. 2021–2031. doi: 10.1139/b67-220.

Nunes, C.C., Maffia, L.A., Mizubuti, E.S.G., Brommonschenkel, S.H. and Silva, J.C. (2009) ‘Genetic diversity of populations of *Hemileia vastatrix* from organic and conventional coffee plantations in Brazil’, *Australasian Plant Pathology*, 38(5), pp. 445–452. doi: 10.1071/AP09021.

Orbach, M.J., Farrall, L., Sweigard, J.A., Chumley, F.G. and Valent, B. (2000) ‘A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi*-ta’, *Plant Cell*, 12(11), pp. 2019–2032. doi: 10.1105/tpc.12.11.2019.

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

Orozco-Arias, S., Isaza, G. and Guyot, R. (2019) ‘Retrotransposons in plant genomes: Structure, identification, and classification through bioinformatics and machine learning’, *International Journal of Molecular Sciences*, 20(15), 3837. doi: 10.3390/ijms20153837.

Porquier, A., Tisserant, C., Salinas, F., Glassl, C., Wange, L., Enard, W., Hauser, A., Hahn, M. and Weiberg, A. (2021) ‘Retrotransposons as pathogenicity factors of the plant pathogenic fungus *Botrytis cinerea*’, *Genome Biology*, 22(1), 225. doi: 10.1186/s13059-021-02446-4.

Porto, B.N., Caixeta, E.T., Mathioni, S.M., Vidigal, P.M.P., Zambolim, L., Zambolim, E.M., Donofrio, N., Polson, S.W., Maia, T.A., Chen, C. and Adetunji, M. (2019) ‘Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility’, *PLoS ONE*, 14(4), e0215598. doi: 10.1371/journal.pone.0215598.

Prakash, N.S., Marques, D.V., Varzea, V.M.P., Silva, M.C., Combes, M.C. and Lashermes, P. (2004) ‘Introgression molecular analysis of a leaf rust resistance gene

from *Coffea liberica* into *C. arabica* L.', *Theoretical and Applied Genetics*, 109(6), pp. 1311–1317. doi: 10.1007/s00122-004-1748-z.

Quispe-Apaza, C.S., Mansilla-Samaniego, R.C., López-Bonilla, C.F., Espejo-Joya, R., Villanueva-Caceda, J. and Monzón, C. (2017) 'Genetic diversity of *Hemileia vastatrix* of two coffee producing areas in Peru', *Revista Mexicana de Fitopatología*, 35(3), pp.418-436. doi: 10.18781/r.mex.fit.1612-7.

Quispe-Apaza, C., Mansilla-Samaniego, R., Espejo-Joya, R., Bernacchia, G., Yabar-Larios, M. and López-Bonilla, C. (2021) 'Spatial and temporal genetic diversity and population structure of *Hemileia vastatrix* from Peruvian Coffee Plantations', *Plant Pathology Journal*, 37(3), pp. 280-290. doi: 10.5423/PPJ.OA.10.2020.0192.

Ramírez-Camejo, L.A., Eamvijarn, A., Díaz-Valderrama, J.R., Karlsen-Ayala, E., Koch, R.A., Johnson, E., Pruvot-Woehl, S., Mejía, L.C., Montagnon, C., Maldonado-Fuentes, C. and Aime, M.C. (2022) 'Global analysis of *Hemileia vastatrix* populations shows clonal reproduction for the coffee leaf rust pathogen throughout most of its range', *Phytopathology*, 112(3), pp. 643-652. doi: 10.1094/PHYTO-06-21-0255-R

Ramos, A.P., Tavares, S., Tavares, D., Silva, M.D.C., Loureiro, J. and Talhinhas, P. (2015) 'Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported-2489Mbp', *Molecular Plant Pathology*, 16(9), pp. 1006–1010. doi: 10.1111/mpp.12255.

Ritschel, A. (2005) *Monograph of the genus Hemileia (Uredinales)*. Tübingen, Germany: Gebrüder Borntraeger Verlagsbuchhandlung.

Rodrigues, A.S.B., Silva, D.N., Várzea, V., Paulo, O.S. and Batista, D. (2022) 'Worldwide Population Structure of the Coffee Rust Fungus *Hemileia vastatrix* Is Strongly Shaped by Local Adaptation and Breeding History', *Phytopathology*, 112(9), pp. 1998–2011. doi: 10.1094/PHYTO-09-21-0376-R.

Rodrigues Jr, C., Bettencourt, A. and Rijo, L. (1975) 'Races of the pathogen and resistance to coffee rust', *Annual Review of Phytopathology*, 13, pp. 49–70. doi: 10.1146/annurev.py.13.090175.000405

Rovenich, H., Boshoven, J. C. and Thomma, B. P. H. J. (2014) 'Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes', *Current Opinion in Plant Biology*, 20, pp. 96–103. doi: 10.1016/j.pbi.2014.05.001.

Ruijter, J.M., Lorenz, P., Tuomi, J.M., Hecker, M. and van den Hoff, M.J. (2014) 'Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend

on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency', *Microchimica Acta*, 181, pp. 1689–1696. doi: 10.1007/s00604-013-1155-8.

Rutherford, M. A. and Phiri, N. (2006) 'Pests and Diseases of Coffee in Eastern Africa : A Technical and Advisory Manual', Wallingford, UK: CAB International.

Sampath, P., Lee, J., Cheng, F., Wang, X., Yang, TJ. (2015). Miniature Transposable Elements (mTEs): Impacts and Uses in the Brassica Genome. In: Wang, X., Kole, C. (eds) The Brassica rapa Genome. Compendium of Plant Genomes. *Springer*, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-47901-8_6

Santana, M.F., Batista, A.D., Ribeiro, L.E., de Araújo, E.F. and de Queiroz, M.V. (2013) 'Terminal repeat retrotransposons as DNA markers in fungi', *Journal of Basic Microbiology*, 53(10), pp. 823–827. doi: 10.1002/jobm.201200453.

Santana, M.F., Zambolim, E.M., Caixeta, E.T. and Zambolim, L. (2018) 'Population genetic structure of the coffee pathogen *Hemileia vastatrix* in Minas Gerais, Brazil', *Tropical Plant Pathology*, 43(5), pp. 473–476. doi: 10.1007/s40858-018-0246-9.

Schwessinger, B. and Rathjen, J. P. (2017) 'Extraction of high molecular weight DNA from fungal rust spores for long read sequencing', *in Methods in Molecular Biology*, pp. 49–55. doi: 10.1007/978-1-4939-7249-4_5.

Seidl, M. F. and Thomma, B. P. (2017) 'Transposable elements direct the coevolution between plants and microbes', *Trends in Genetics*, 33(11), pp. 842-851. doi: 10.1016/j.tig.2017.07.003.

Siddiqi, M. A. and Corbett, D. C. M. (1963) 'Coffee bark diseases in Nyasaland', *Transactions of the British Mycological Society*, 46(1), pp. 91–101. doi: 10.1016/s0007-1536(63)80011-x.

Silva, D.N., Várzea, V., Paulo, O.S. and Batista, D. (2018) 'Population genomic footprints of host adaptation, introgression and recombination in coffee leaf rust', Molecular Plant Pathology, 19(7), pp. 1742–1753. doi: 10.1111/mpp.12657.

Silva, M.D.C., Nicole, M., Guerra-Guimarães, L. and Rodrigues Jr, C.J. (2002) 'Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves', *Physiological and Molecular Plant Pathology*, 60(4), pp. 169–183. doi: 10.1006/pmpp.2002.0389.

Silva, M.D.C., Várzea, V., Guerra-Guimarães, L., Azinheira, H.G., Fernandez, D., Petitot, A.S., Bertrand, B., Lashermes, P. and Nicole, M. (2006) 'Coffee resistance to the main

diseases: leaf rust and coffee berry disease', *Brazilian Journal of Plant Physiology*, 18(1), pp. 119–147. doi: 10.1590/s1677-04202006000100010.

Silva, M.C., Guerra-Guimarães, L., Loureiro, A. and Nicole, M.R. (2008) 'Involvement of peroxidases in the coffee resistance to orange rust (*Hemileia vastatrix*)', *Physiological and Molecular Plant Pathology*, 72(1–3), pp. 29–38. doi: 10.1016/j.pmpp.2008.04.004.

Silva, M.D.C., Guerra-Guimarães, L., Diniz, I., Loureiro, A., Azinheira, H., Pereira, A.P., Tavares, S., Batista, D. and Várzea, V. (2022) 'An Overview of the Mechanisms Involved in Coffee-*Hemileia vastatrix* Interactions: Plant and Pathogen Perspectives', *Agronomy*, 12(2), 326. doi: 10.3390/agronomy12020326.

Slotkin, R. K. and Martienssen, R. (2007) 'Transposable elements and the epigenetic regulation of the genome', *Nature Reviews Genetics*, 8(4), pp. 272–285. doi: 10.1038/nrg2072.

Soyer, J.L., El Ghalid, M., Glaser, N., Ollivier, B., Linglin, J., Grandaubert, J., Balesdent, M.H., Connolly, L.R., Freitag, M., Rouxel, T. and Fudal, I. (2014) 'Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus *Leptosphaeria maculans'*, *PLoS Genetics*, 10(3), e1004227. doi: 10.1371/journal.pgen.1004227.

Spanu, P. D. (2012) 'The genomics of obligate (and nonobligate) biotrophs', *Annual Review of Phytopathology*, 50, pp. 91–109. doi: 10.1146/annurev-phyto-081211-173024.

Sperschneider, J., Jacques, S., Xu, B., Upadhyaya, N.M., Mago, R., Jones, A.W., Schwessinger, B., Figueroa, M., Singh, K.B., Stone, E.A. and Wang, M.B. (2018) 'The stem rust fungus *Puccinia graminis* f. sp. *tritici* induces waves of small RNAs with opposing profiles during wheat infection', *bioRxiv*, p.469338.

Steenwyk, J. L. and Rokas, A. (2018) 'Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation', *Frontiers in Microbiology*, 9, 288. doi: 10.3389/fmicb.2018.00288.

Talhinhas, P., Azinheira, H.G., Vieira, B., Loureiro, A., Tavares, S., Batista, D., Morin, E., Petitot, A.S., Paulo, O.S., Poulain, J. and Da Silva, C. (2014) 'Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection', *Frontiers in Plant Science*, 5, 88. doi: 10.3389/fpls.2014.00088.

Talhinhas, P., Ramos, A.P., Tavares, D., Tavares, S. and Loureiro, J. (2015) 'Genome size estimates for six rust (Pucciniales) species', *Revista de Ciências Agrárias*, 38(2), pp. 176–183. doi: 10.19084/rca.16912.

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

Talhinhas, P., Carvalho, R. and Loureiro, J. (2021) 'The use of flow cytometry for fungal nuclear DNA quantification', *Cytometry Part A*, 99(4), pp. 343–347. doi: 10.1002/cyto.a.24335.

Talhinhas, P., Carvalho, R., Tavares, S., Ribeiro, T., Azinheira, H., Ramos, A.P., Silva, M.D.C., Monteiro, M., Loureiro, J. and Morais-Cecílio, L. (2023) 'Diploid Nuclei Occur throughout the Life Cycles of Pucciniales Fungi', *Microbiology Spectrum*, 11(4), e01532-23. doi: 10.1128/spectrum.01532-23.

Tavares, S., Ramos, A.P., Pires, A.S., Azinheira, H.G., Caldeirinha, P., Link, T., Abranches, R., Silva, M.D.C., Voegele, R.T., Loureiro, J. and Talhinhas, P. (2014) 'Genome size analyses of Pucciniales reveal the largest fungal genomes', *Frontiers in Plant Science*, 5, 422. doi: 10.3389/fpls.2014.00422.

Tobias, P.A., Schwessinger, B., Deng, C.H., Wu, C., Dong, C., Sperschneider, J., Jones, A., Luo, Z., Zhang, P., Sandhu, K. and Smith, G.R. (2021) '*Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements', *G3 Genes|Genomes|Genetics* , 11(3), p.jkaa015. doi: 10.1093/g3journal/jkaa015.

Tobias, P.A., Edwards, R.J., Surana, P., Mangelson, H., Inácio, V., do Céu Silva, M., Várzea, V., Park, R.F. and Batista, D. (2022) 'A chromosome-level genome resource for studying virulence mechanisms and evolution of the coffee rust pathogen *Hemileia vastatrix*', bioRxiv, pp.2022-07.

Urbina, H. and Aime, M.C. (2024) 'First US Continental Record of Coffee Leaf Rust (*Hemileia vastatrix*) on Coffee (*Coffea arabica*) in Southwest Florida, USA', *Plant Disease*, 108(1), p.228. doi: 10.1094/PDIS-09-23-1869-PDN.

Vossen, H.V.D. and Walyaro, D.J. (2009) 'Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in arabica coffee (*Coffea arabica* L.)', *Euphytica*, 165(1), pp. 105-111. doi: 10.1007/s10681-008-9769-3

Várzea, V. and Marques, D. V (2005) '*Population variability of Hemileia vastatrix vs Coffee durable resistance*', Universidade Federal de Viçosa Pub. Brazil, pp. 53–74.

Vieira, A. *et al.* (2011) 'Validation of RT-qPCR reference genes for in planta expression studies in *Hemileia vastatrix*, the causal agent of coffee leaf rust', *Fungal Biology*, 115(9), pp. 891–901. doi: 10.1016/j.funbio.2011.07.002.

Vleeshouwers, V. G. A. A. and Oliver, R. P. (2014) 'Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens', *Molecular Plant-Microbe Interactions*, 27(3), pp. 196–206. doi: 10.1094/MPMI-10-13-0313-IA.

Waller, J. M., Bigger, M. and Hillocks, R. J. (2007) '*Coffee pests, diseases and their management'*, CABI. doi: 10.1079/9781845931292.0000.

Witte, C.P., Le, Q.H., Bureau, T. and Kumar, A. (2001) 'Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes', *Proceedings of the National Academy of Sciences*, 98(24), pp.13778-13783. doi: 10.1073/pnas.241341898.

De Wit, P. J. G. M. (2007) 'How plants recognize pathogens and defend themselves', *Cellular and Molecular Life Sciences*, 64(21), pp. 2726–2732. doi: 10.1007/s00018-007-7284-7.

De Wit, P.J., Mehrabi, R., Van den Burg, H.A. and Stergiopoulos, I. (2009) 'Fungal effector proteins: Past, present and future: Review', *Molecular Plant Pathology*, 10(6), pp. 735–747. doi: 10.1111/j.1364-3703.2009.00591.x.

Zadoks, J. C. and Schein, R. D. (1979) *Epidemiology and plant disease management*. XIII. New York: Oxford University Press Inc.

Zemojtel, T., Kielbasa, S. M., Arndt, P. F., Chung, H. R., & Vingron, M. (2009) 'Methylation and deamination of CpGs generate p53-binding sites on a genomic scale', *Trends in Genetics*, 25(2), pp. 63–66. doi: https://doi.org/10.1016/j.tig.2008.11.005.

Zhang, L., Yan, L., Jiang, J., Wang, Y., Jiang, Y., Yan, T. and Cao, Y. (2014) 'The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans'*, *Virulence*, 5(6), pp. 655–664. doi: 10.4161/viru.32180.

https://wfoplantlist.org/plant-list/taxon/wfo-4000008851-2023-06?page=1 (accessed on March 2nd of 2020).

https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:325985-2 (accessed on March 2nd of 2020).

https://varieties.worldcoffeeresearch.org/arabica-2/history-of-arabica (accessed on March 2nd of 2020).

https://www.pacb.com/wp-content/uploads/2015/09/SharedProtocol-Extracting-DNA-usinig-Phenol-Chloroform.pdf (accessed in July 15th of 2020).

SUPPLEMENTARY INFORMATION

DNA extraction buffers:

Buffer A

1) 0.35M sorbitol
2) 0.1M Tris-HCl, pH 9.0
3) 5mM EDTA, pH 8.0

Buffer B

1) 0.2M Tris-HCl, pH 9.0
2) 50mM EDTA, pH 8.0
3) 2M NaCl
4) 2% (w/v) CTAB

Buffer C

5% Sarkosyl (*N*-lauroylsarcosine sodium salt)

DNA Purification protocols

Ammonium acetate protocol

To each eppendorf tube, 0.5 volumes of NH_4CH_3COO (8M) were added and incubated on ice for 30 minutes, followed by a centrifugation at 10.000 x *g* for another 30 minutes. The supernatant was transferred to new eppendorfs, added 2 volumes of absolute ethanol, and kept at -20°C for 1 hour. Then, a 10-minute centrifugation at 5000 x *g* was done to remove the ethanol. Another 500 µl of cold ethanol (70%) were added to the tubes, centrifuged 5 minutes at 10.000 x *g*, removed the ethanol and final centrifugation for 1 minute to remove any ethanol residue. Eppendorf tubes were opened, air dried and DNA was eluted in 25 µl of Tris-HCl (10mM; pH 8.5).

Sodium acetate protocol

To each eppendorf tube, 0.1 volumes of $NaCH_3COO$ (3M; pH 5.5) + 2 volumes of absolute ethanol were added and incubated in -20°C for 1 hour. After that, a 4°C centrifugation at 13500 rpm was done for 30 minutes, followed by a 500 µl cold ethanol (70%) wash in each tube. One cold centrifugation at 13500 rpm for 10 minutes to collect the DNA and remove the ethanol, and another cold centrifugation at 13500 rpm for 1 minute to remove the ethanol residues. Pellets were air dried and final elution in 25 µl of Tris-HCl (10mM; pH 8.5).

Supplementary Table 1 – List of candidate retrotransposons obtained from BLASTn analysis of *Hemileia vastatrix* transcripts (Talhinhas *et al.,* 2014) against both genome NCBI databank for *H. vastatrix* (Cristancho *et al.,* 2014; Porto *et al.,* 2019) and diagnostic SNPs (Silva *et al.,* 2018). For each sequence, the best hit is presented.

Hv transcript	Hv genome seqs	% identity	Alignment length	Mis-matches	gap opens	e-value
Hv00003	gHv_v2_14327	91,11	270	22	2	1e102
Hv00008	gHv_v2_12293	98,06	155	3	0	3e-74
Hv00075	gHv_v2_29145	93,38	272	6	3	3e-112
Hv00190	gHv_v2_11007	86,18	579	75	5	7e-180
Hv01057	gHv_v2_02343	88,37	215	22	3	6e-70
Hv01204	gHv_v2_25486	90,82	207	15	4	4e-76
Hv01207	gHv_v2_21344	95,90	244	10	0	2e-112
Hv01212	gHv_v2_20261	93,52	216	13	1	6e-91
Hv01227	gHv_v2_20946	89,97	329	27	6	3e-120
Hv01263	gHv_v2_00493	90,53	169	10	5	2e-59
Hv01271	gHv_v2_01726	98,62	218	2	1	2e-109
Hv01411	gHv_v2_25486	80,70	746	97	13	3e-172
Hv01796	gHv_v2_18676	87,18	195	25	0	4e-60
Hv02407	gHv_v2_09834	94,84	155	8	0	5e-67
Hv04547	gHv_v2_10773	90,57	159	12	3	1e-55
Hv05720	gHv_v2_14327	90,53	243	17	6	6e-89
Hv05889	gHv_v2_19772	92,56	336	24	1	4e-138
Hv06501	gHv_v2_00244	78,57	504	95	9	2e-89
Hv06694	gHv_v2_20948	97,42	349	5	4	2e-171
Hv07809	gHv_v2_24390	96,92	130	3	1	1e-58
Hv08122	gHv_v2_01583	94,15	171	9	1	1e-71
Hv transcript	**Locus SNPs**	**% identity**	**Alignment length**	**Mis-matches**	**gap opens**	**e-value**
Hv00003	locus167194	95,35	86	4	0	1e-37
Hv00008	locus253736	95,24	84	4	0	2e-36
Hv00075	locus189673	94,19	86	5	0	3e-37
Hv00097	locus261779	98,61	72	1	0	2e-35
Hv00190	locus177093	93,1	87	5	1	3e-34

Hv00333	locus146683	91,86	86	7	0	2e-33
Hv01057	locus81027	92,77	83	5	1	3e-32
Hv01204	locus151971	94,19	86	5	0	1e-36
Hv01207	locus41404	94,25	87	4	1	2e-36
Hv01212	locus19530	90,7	86	6	1	5e-32
Hv01227	locus72423	91,86	86	6	1	2e-32
Hv01251	locus243296	89,41	85	9	0	1e-30
Hv01263	locus193181	98,84	86	1	0	2e-43
Hv01271	locus112918	94,19	86	5	0	5e-36
Hv01411	locus163881	92,68	82	6	0	3e-32
Hv01602	locus273879	94,19	86	4	1	2e-35
Hv01796	locus177495	93,75	80	3	2	7e-33
Hv02407	locus163961	93,75	80	5	0	6e-36
Hv04547	locus81027	93,98	83	4	1	4e-34
Hv05720	locus261779	95,89	73	3	0	8e-33
Hv05889	locus167403	94,12	85	5	0	8e-36
Hv06238	locus261779	95,83	72	3	0	5e-32
Hv06501	locus243296	92,77	83	6	0	6e-34
Hv06694	locus177785	93,1	87	5	1	1e-34
Hv07809	locus243296	94,12	85	4	1	2e-36
Hv08119	locus101333	91,46	82	7	0	4e-32
Hv08122	locus158600	91,57	83	7	0	1e-31
Hv08124	locus243296	95,24	84	4	0	1e-38

Supplementary Table 2 – Description of the selected candidate ESTs (Talhinhas *et al.,* 2014) to study retrotransposons genetic diversity.

Candidate EST	Contig size (bp)	Description (Accession number) *	e-value	Bit Score	Accession length
Hv00075	297	*Hemileia vastatrix* clone HvCat_contig_26833 retrotransposon Gypsy-like, complete sequence (**JX569469.1**)	$9e^{-33}$	135	12308
Hv00190	1722	*Hemileia vastatrix* clone HvCat_contig_74737 retrotransposon Gypsy-like, complete sequence (**JX569700.1**)	0.0	647	10333
Hv01057	1300	*Hemileia vastatrix* clone HvCat_contig_71800 retrotransposon Gypsy-like, complete sequence (**JX569688.1**)	$4e^{-72}$	268	6626
Hv01207	720	*Hemileia vastatrix* clone HvCat_contig_41834 retrotransposon Gypsy-like, complete sequence (**JX569681.1**)	$3e^{-106}$	381	8479
Hv02407	155	*Hemileia vastatrix* clone HvCat_contig_25390 retrotransposon Gypsy-like, complete sequence (**JX569403.1**)	$1e^{-18}$	87.9	11014
Hv04547	795	*Hemileia vastatrix* clone HvCat_contig_71800 retrotransposon Gypsy-like, complete sequence (**JX569688.1**)	$5e^{-69}$	257	6626
Hv08122	302	*Hemileia vastatrix* clone HvCat_contig_20578 retrotransposon Gypsy-like, complete sequence (**JX569411.1**)	$4e^{-46}$	180	5729

* **NCBI megablast (blastn)**

Supplementary Table 3 – Calculations for various selected DNA dilutions equivalent to different number of copies as to set up calibration curves for each retrotransposon. A quantity scale was determined to convert in decimal logarithm (Log10) values so that the curves could be done along with the corresponding cycle threshold values from the real-time quantitative PCR reactions.

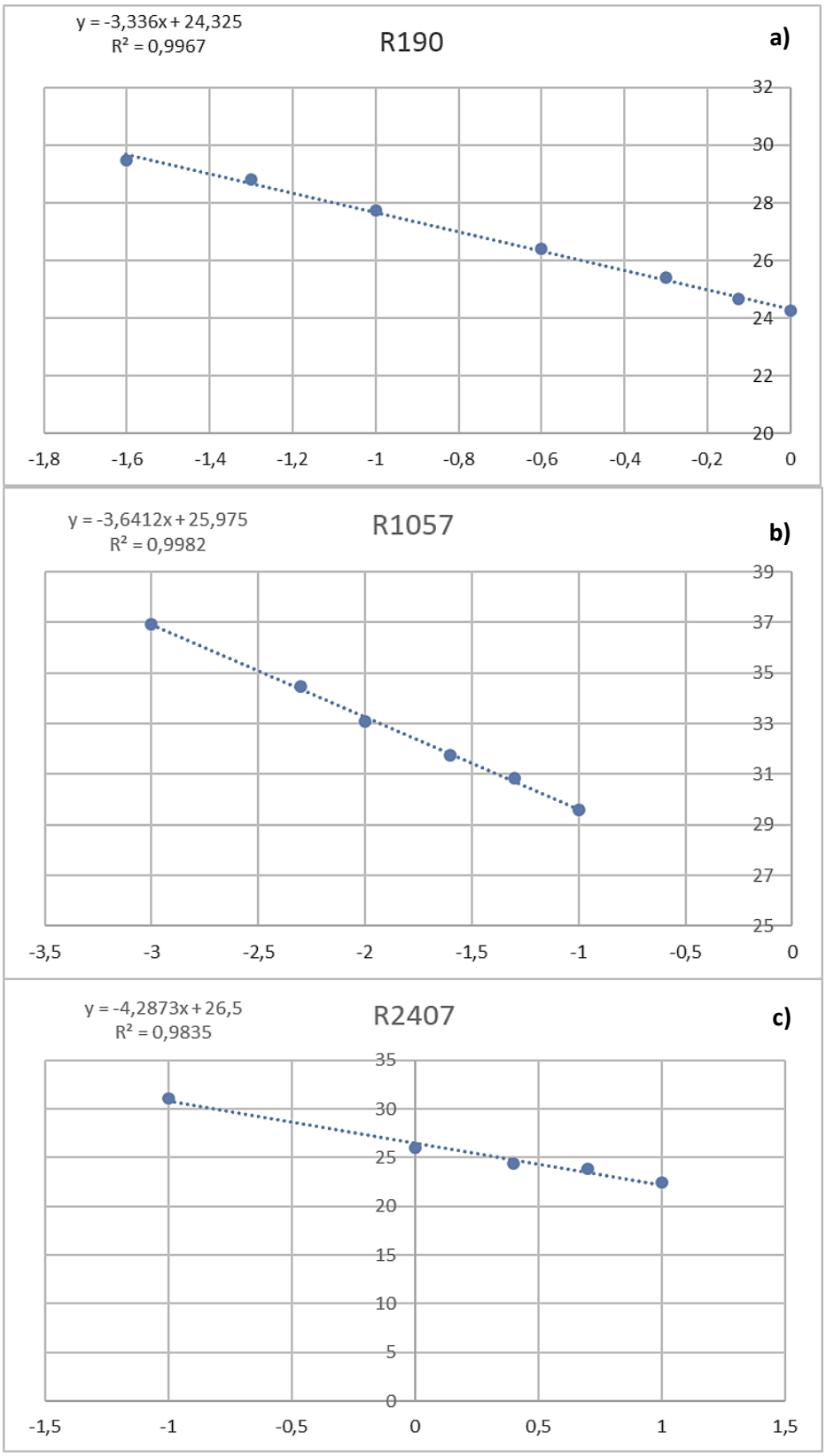
Number of copies	**Quantity**	**Log_{10} value**
1000	1	0,0
750	0,75	-0,1
500	0,50	-0,3
250	0,25	-0,6
100	0,10	-1,0
50	0,050	-1,3
25	0,025	-1,6
10	0,010	-2,0
5	0,005	-2,3
1	0,001	-3,0

Supplementary Table 4 – Comparative p-values from the statistical tests Mann-Whitney and T-Student for the RT-qPCR data for all three retrotransposons (*R190, R1057, and R2407*) under study, comparing two compatible interactions with the rust isolates Hv70 and Hv1427. Red values are indicative of non-statistical significance differences, where the green values are representative of statistically significant differences.

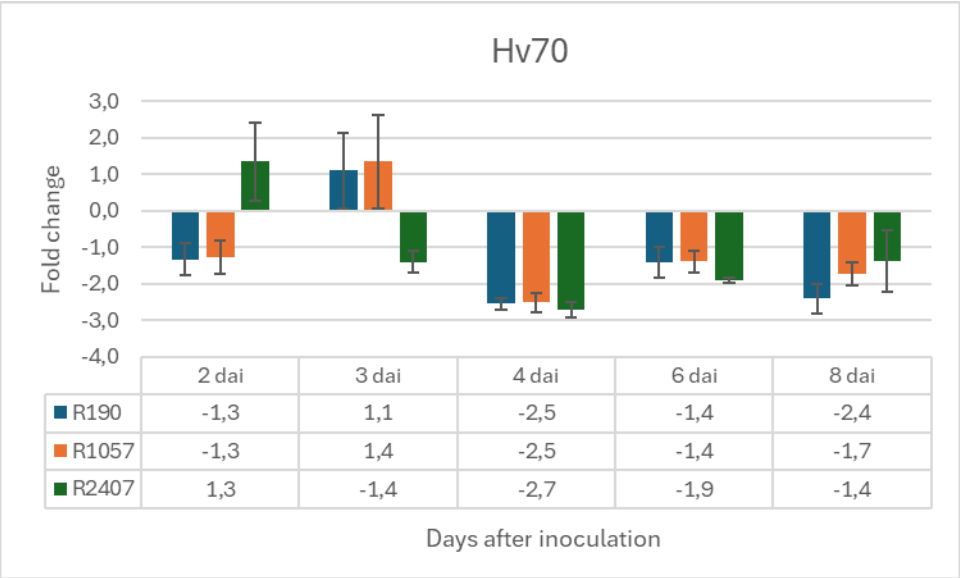
	Mann-whitney U test	**T-student**
R190	p-value	p-value
2 dai	0,148916	0,000904
3 dai	0,148916	0,016127
4 dai	0,148916	0,002002
6 / 8 dai	0,148916	0,002990
8 / 11 dai	0,148916	0,005279

	Mann-whitney U test	**T-student**
R1057	p-value	p-value
2 dai	0,148916	0,005412
3 dai	0,148916	0,023069
4 dai	0,148916	0,000053
6 / 8 dai	0,148916	0,010014
8 / 11 dai	0,148916	0,001267

	Mann-whitney U test	**T-student**
R2407	p-value	p-value
2 dai	0,148916	0,000122
3 dai	0,148916	0,020205
4 dai	0,148916	0,000005
6 / 8 dai	0,148916	0,000003
8 / 11 dai	0,148916	0,009958

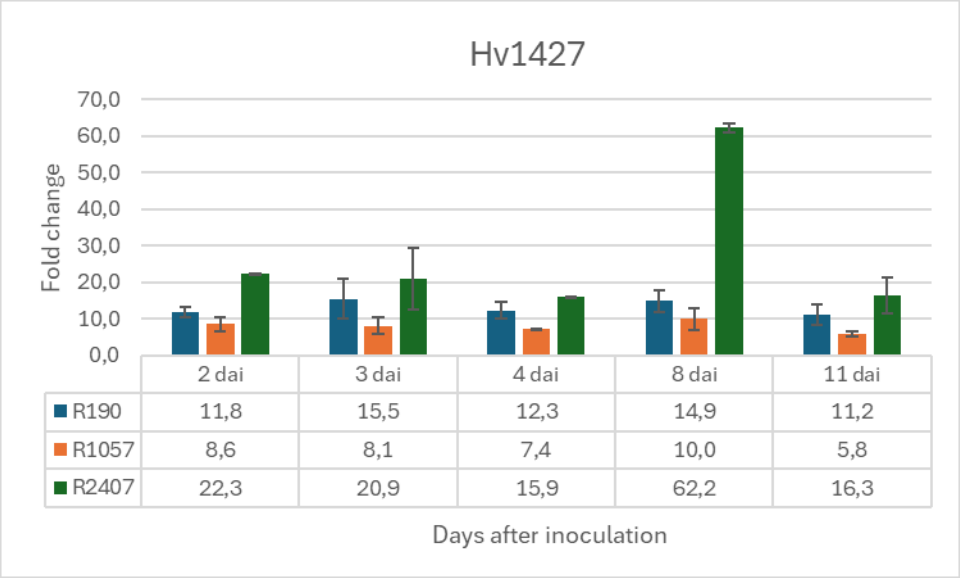


Supplementary Figure 1 – Calibration curves for each retrotransposon under study to estimate the number of copies per genome through qPCR. The x-axis is representative of decimal logarithmic values corresponding to a certain number of copies, as for the y-axis is representative of the cycle threshold (Ct) values of the qPCR reactions; a) *R190* calibration curve; b) *R1057* calibration curve; c) *R2407* calibration curve.



	2 dai	3 dai	4 dai	6 dai	8 dai
R190	-1,3	1,1	-2,5	-1,4	-2,4
R1057	-1,3	1,4	-2,5	-1,4	-1,7
R2407	1,3	-1,4	-2,7	-1,9	-1,4

Supplementary Figure 2 – RT-qPCR expression profiles of all three retrotransposons (*R190, R1057 and R2407*) for the Hv70 compatible interaction, from the 2 dai (days after inoculation) to 8 dai, using the first 24 hours after inoculation as the control sample. Fold change values are representative of relative expression against the control. Positive fold change values over 1 represent upregulated genes and negative fold change values represent downregulated genes.



	2 dai	3 dai	4 dai	8 dai	11 dai
R190	11,8	15,5	12,3	14,9	11,2
R1057	8,6	8,1	7,4	10,0	5,8
R2407	22,3	20,9	15,9	62,2	16,3

Supplementary Figure 3 – RT-qPCR expression profiles of all three retrotransposons (*R190, R1057 and R2407*) for the Hv1427 compatible interaction, from the 2 dai (days after inoculation) to 11 dai, using the first 24 hours after inoculation as the control sample. Fold change values are representative of relative expression against the control. Positive fold change values over 1 represent upregulated genes.