Seventy years of *Fusarium* wilt - coffee interactions: historical genomics reveals pathogen emergence and divergence

Lily Peck

November 2022

Department of Life Sciences
Imperial College of Science, Technology and Medicine
Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy and the Diploma of Imperial College London

Seventy years of *Fusarium* wilt - coffee interactions: historical genomics reveals pathogen emergence and divergence

Lily Peck

Abstract

In the early 20th century, coffee wilt disease devastated coffee farms across Africa, with catastrophic impacts on production and farmers: millions of dollars were lost, the major coffee variety *Coffea excelsa* failed commercially, and widespread farm closures occurred. Improved sanitation measures and plant breeding programmes in the 1950s briefly overcame the effects of this first disease outbreak, but just two decades later a second prolific epidemic attacked robusta coffee (*C. canephora* robusta) in Uganda, Tanzania and the Democratic Republic of Congo. Separately, a third epidemic affected arabica coffee (*C. arabica*) in Ethiopia. Coffee wilt disease is caused by *Fusarium xylarioides*, a soil-borne fungal pathogen that induces vascular wilt on coffee trees. Little is understood about coffee wilt disease and *F. xylarioides* compared with other *Fusarium* plant pathogens although they all have a similar propensity to cause devastating crop and economic losses. Over the past century, coffee wilt disease is the most significant disease to have affected coffee production, and has caused over US$1 billion in losses to national economies in central Africa. This thesis explores the use of living fungal strains preserved in culture collections to elucidate the evolutionary past of a major disease outbreak. The history of coffee wilt disease highlights the risk of disease re-emergence, and the importance of rapid strategic action in response to outbreaks. The knowledge held in culture collections around the world could help prepare us better for managing future outbreaks.

The first experimental chapter of this thesis (chapter two) used genome sequencing

of six historical culture collection strains spanning 52 years to identify the evolutionary processes behind repeated outbreaks of coffee wilt disease. Phylogenomic reconstructions and a screen for putative effector genes showed that the host-specific *F. xylarioides* groups have diverged in gene content and sequence mainly by vertical processes within lineages, and through expansion of certain carbohydrate-active enzyme families. A subset of putative effector genes, however, showed evidence for horizontal acquisition and close homology to genes from *F. oxysporum*.

Subsequently, the next chapter (chapter three) added seven more historic *F. xylarioides* genomes and a long-read sequenced reference genome to test the *F. xylarioides* host-specific groups found in chapter two and the nature of the putative horizontal transfer. Four genetically differentiated populations were present: arabica, robusta and two populations amongst the strains isolated in the 1950s from the initial 1920s-1950s outbreak. In addition, six other *Fusarium* strains were sequenced from amongst species that are also found on coffee wilt diseased trees: five *F. oxysporum* strains and one *F. solani*. Analyses showed evidence for multiple horizontal transfers of over 20 kilobase long regions between the *F. oxysporum* mobile chromosomes and the *F. xylarioides* populations. In addition, *F. xylarioides* is enriched with *miniature impala* transposons that are found in *F. oxysporum* and linked with the horizontal transfer of pathogenicity. These results indicate that horizontal gene transfer occurred repeatedly in the *F. xylarioides* clade and could be one of the drivers behind the repeated emergence of coffee wilt disease during the last century via the transfer of gene- and repeat-rich genomic regions.

The final experimental chapter (chapter four) investigates whether any genes highlighted in the previous chapters were expressed in arabica *F. xylarioides* strains infecting arabica coffee host plants. RNA sequence data reveals major up-regulation of pectin lyase enzymes, suggesting a key role for them in the onset of vascular wilt infection. It also revealed up-regulation of three *F. oxysporum* secreted in xylem (*SIX*) effectors, which are all less than 1 kilobase from *miniature impala* transposons. This, combined with the close phylogenetic relationship with *F. oxysporum* for a subset of several hundred up-

regulated genes, highlights the potential role of horizontal gene transfers in the spread of pathogenicity between *F. oxysporum* and *F. xylarioides* in shared vascular wilt infections.

These results demonstrate how comparative genomics and transcriptomics of historic strains can reveal mechanisms that allow fungal pathogens to keep pace with our efforts to control them. Knowledge of horizontal transfer mechanisms and putative donor taxa could help to design future intercropping strategies that minimize the risk of new and emerging pathogens via effector gene transfers between closely related *Fusarium* taxa.

Acknowledgements

As I sit here and gaze at my fully compiled thesis document, I am awash with many feelings. Tiredness, mostly, but also relief and amazement that I have managed to get it done. And, as always, there are many people who I will subsequently thank, but the first person is my partner James. Always believing in me (even to the extent that I would make my self-imposed submission deadline and organised me a surprise thesis submission party for... two days ago (sorry James)) and always scooping me up after a failure in the lab or a whole day of 'computer says no'.

The next person I would like to thank is my supervisor Tim. I could not have asked for a more supportive supervisor over the years and Tim is the reason I have managed to get to this point. Always enthusiastic, he quietly listens to my ideas before jumping onto google to see if they are possible (and most of the time, they have been). Tim has been kind, respectful, and an inspiring person to work with. He has pushed me when I have needed it and has always had an ear for my problems.

I also thank Pietro and Julie. Neither is my official supervisor, yet both have provided me with endless advice, ideas, and support throughout my PhD. Matt, my co-supervisor, helped to set up this project and has navigated the CABI collection with me. I thank everyone in Barralab from the past four years, and in particular Reuben, for bioinformatics support and helping me get through the aforementioned 'computer says no' days, and Theo, firstly for the inspiration to sequence a reference genome and latterly for the technical help to do so, and Chris for lots of thoughtful advice. Anne and Christiane have provided me with endless help and support through the DTP for which I am extremely

grateful. Following covids disruption and a switch in my aims to look at gene expression data, Aaron Davis set me up with coffee plants from Kew and Martin Selby took care of them at Silwood. I also thank Tom Bell and Martin Bidartondo for sitting on my progress review panel and giving me kind and constructive feedback. And thanks also to the CABI guys, particularly Helen, for showing me how to recognise my 'bananas' and 'bunches of bananas' under the microscope.

Thinking back to my Sheffield days, my first yearnings to do a PhD started with Estrella and Jurriaan many moons ago, and Estrella has been a mentor to me since. Even introducing me to Ana, who has shown me how to work and live in a more mindful and productive way. This thesis would never have been written without my frequent trips to London Fields lido, West reservoir and the various Better saunas in Hackney, all helping me get my thoughts into order.

Merci beaucoup à tout l'équipe GEE (génétique, écologie et évolution). To Tatiana for having me for three months and welcoming me into the team better than I could have dreamed. I learned so much about cheese fungi and the dirtiest of programming in my 3 months in GEE from Jeanne, Ricardo, Bastien the tidyverse sensei, Samuel, and everyone else. And I will never forget the importance of enjoying a French lunch break with Thibault, Bastien, and Elizabeth everyday. Thanks also to the Stevenson fund at Imperial College for funding me to complete this research placement.

Finally, I'd like to thank my mum, for showing little me how far she could reach, and my family for endless emotional (and financial) support, the three of you have helped me get through the last four years. Even if you never stop calling me a 'student'.

As I look upon the next chapter of my life, I am grateful to everyone and a little sad that this journey is nearly over. I hope to keep in touch with everyone and continue to build projects together to try and make the world a better place long into the future.

This thesis was funded by the Grantham Institute at Imperial College's doctoral training partnership 'science and solutions for a changing planet'. Natural Environment Research Council [NE/L002515/1]

In addition, I would like to thank my examiners Mat and Sarah for a thoroughly engaging, informative, challenging and constructive viva.

Statement of originality

This work is original and my own. All experiments were conceived by myself and T. Barraclough, with input from P. Spanu, J. Flood, M. Ryan and T. Giraud. The maps in chapter one have also been published in the British society of plant pathology report (see publications). A paper has been published covering the material in chapter two. In this chapter, R. Nowell annotated genes and transposable elements in the six fungal genomes and assisted with other analyses. In chapter three, scripts to call single nucleotide polymorphisms and population genomic analyses were from J. P. Vernadet and B. Bennetot at Paris Saclay University respectively, the reference genome was sequenced by T. Llewellyn and S. O'Donnell helped quantify the presence of putative horizontal transferred regions and with the final reference assembly. A paper has been published covering the material from section 3.5.1. The original collectors of the fungal strains used in this work have been acknowledged in the appropriate chapters.

Copyright statement

Publications

Figures 1.4 and 1.6 from chapter one were published here:

Peck, L. (n.d.), Coffee wilt disease - the forgotten threat to coffee. In press, Technical report, British Society for Plant Pathology, London.

A modified version of chapter two was published here:

Peck, L. D., Nowell, R. W., Flood, J., Ryan, M. R. & Barraclough, T. G. (2021), 'Historical genomics reveals the evolutionary mechanisms behind multiple outbreaks of the host-specific coffee wilt pathogen *Fusarium xylarioides*', *BMC Genomics* 22(1), 1–24

Section 3.5.1 from chapter three was published here:

Ryan, M. J., Peck, L. D., Smith, D., Flood, J. & Barraclough, T. G. (2022), 'Culture collections as a source of historic strains for genomic studies in plant pathology', *Journal of Plant Pathology* 1, 1–6.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Emerging fungal disease is a problem

Fungal disease has devastated plants and animals for millennia, with the first such records dating from biblical plagues (Fones et al. 2017). History is littered with examples of famines and harvest failures as a result of outbreaks of fungal disease, often with wide-ranging and important human consequences. The Irish potato famine is an example of an oomycete disease outbreak (Donnelly 2002). Such outbreaks are caused by emerging pathogens, namely pathogens that have increased their propensity to cause disease, often by introduction into a new ecological niche, or acquisition of enhanced pathogenicity by acquiring novel genes or a more favourable climate (Anderson et al. 2004). Modern farming practices have inadvertently assisted the processes by which emerging pathogens appear (Fisher et al. 2012). Since the green revolution, farming has favoured the planting of genetically identical monocultures, often protected against disease by single resistance genes and the widespread application of environmentally damaging plant protection products, such as fungicides. Repeated hectares of identical hosts and relatively weak protections have resulted in outbreaks that have decimated crop harvests upon which humanity depends. Recent examples include soybean rust (Langenbach et al. 2016) and wheat stem rust (Leonard & Szabo 2005).

It is well understood that when fungal pathogens encounter host plants, the invader is met by an innate immune system (Jones & Dangl 2006). It is now recognised that the plant innate immune system also comprises a heritable memory with local (i.e. at the site of infection) as well as systemic (i.e. in distant plant parts) inducible immune responses (Reimer-Michalski & Conrath 2016). Consequently, most pathogenic microbe infection attempts fail. However, a small number do succeed thanks to the constant evolutionary pressure on pathogens to overcome host plant inducible defences. In this way, there is a co-evolutionary arms race over millennia between pathogen and host that causes boom and bust cycles of disease, as one gains advantage over the other (Misra & Chaturvedi 2015). Whether the pathogen can take the upper hand depends on its evolutionary potential, namely its genetic adaptability in overcoming host defences that previously kept it at bay (Fones et al. 2017). Such beneficial traits include features which increase variability, such as mutation, recombination, hybridization and horizontal gene transfer (Möller & Stukenbrock 2017, Sanchez-Vallet et al. 2018, Stukenbrock 2013). These processes can result in the evolution of new effectors, proteins which are secreted by pathogens to co-opt host plant physiology during attack, which typically suppress host defences or protect pathogens themselves from an attack in response which may target their growth (de Jonge et al. 2011). Such processes can also result in adaptation to new hosts after host jumps and subsequent speciation: dynamic regions of the genome of *Phytophthora infestans*, the Irish potato famine pathogen, accelerated evolution of effector genes following host jumps (Raffaele et al. 2010). Therefore, whilst the broad mechanisms of how diseases emerge are understood, a major challenge is predicting when and where new types of disease will emerge.

1.2 Fungal vascular wilt disease on plants

Plant pathogens have evolved to target their hosts in a variety of ways, affecting aerial and subground plant parts, as well as the internal vascular system. The plant vascular

system is made up of phloem, whereby living cells transport sugars under high osmotic pressure, and xylem, dead tracheae elements which transport water under low osmotic pressure. Surprisingly, most vascular wilt pathogens invade the nutrient-poor xylem and not the nutrient-rich phloem, probably because it is easier to access dead cells at low pressure (Yadeta & Thomma 2013). The xylem vessels are interconnected by pits, which are small openings (Boer et al. 2003) and which consist of cellulose microfibrils embedded in a hemicellulose and pectin polysaccharide matrix (Tyree & Zimmermann 2013). Water and solutes can move through this mesh-like polysaccharide structure, but the vascular wilt pathogens cannot because they are too large for the pit membrane pores (Choat et al. 2004). Therefore, they have to breach the secondary xylem walls which are highly structured and rigid to colonise and proliferate in the xylem, blocking the flow of water to trigger a characteristic wilt (Figure 1.1a). Vascular wilt diseases are some of the most destructive plant pathogens worldwide and ultimately result in the host plants death (Agrios 2005). Despite the diversity of the fungal kingdom, only four genera are known to colonise this nutrient-poor niche and induce wilts: *Fusarium, Verticillium, Ophiostoma* and *Ceratocystis* (Agrios 2005). Fungal vascular wilt diseases of plants were first described over 50 years ago in extremely accurate detail, considering the absence of molecular techniques at the time (Mace et al. 1981, Talboys & Busch 1970, Talboys 1972). Elucidating the mode of infection was complicated by the fact that the pathogen colonises the interior of the plant, but subsequent studies provided more detail with increasingly powerful molecular and genomic techniques.

Early studies on the precise mechanisms of vascular wilts used pathogenicity-deficient fungal mutants. A diverse range of genes were shown to be required for colonisation of host plant tissue by vascular wilt inducers. These ranged from a mitochondrial carrier protein, where a *fow1* -defective *F. oxysporum* f. sp. *melonis* mutant showed normal growth but significantly reduced pathogenicity against its melon host (Inoue et al. 2002), to an *F. oxysporum* f. sp. *lycopersici* mutant defective in a mitogen-activated protein kinase which showed reduced growth on tomato fruits and severely reduced expression

of a pectate lyase (Di Pietro et al. 2001). Pectate lyases are part of the carbohydrate-active enzyme family (CAZymes), which is thought to be important for vascular wilt fungi in colonising the xylem. Plant cell walls contain cellulose microfibrils embedded in a matrix of hemicelluloses, pectins, and glycoproteins (Carpita & Gibeaut 1993), and these carbohydrates make up the dominant carbon source for invading vascular wilt pathogens (Hervé et al. 2010). The metabolism of cell wall polymers requires a diverse arsenal of enzymes, including those within the CAZyme family (glycoside hydrolases, pectate and polysaccharide lyases and carbohydrate esterases) (Mace et al. 1981). In fact, some CAZymes contain carbohydrate-binding modules (Boraston et al. 2004), which bind to cell wall polymers to increase the catalytic efficiency of the enzymes by improving contact (Hervé et al. 2010, Carrard et al. 2000).

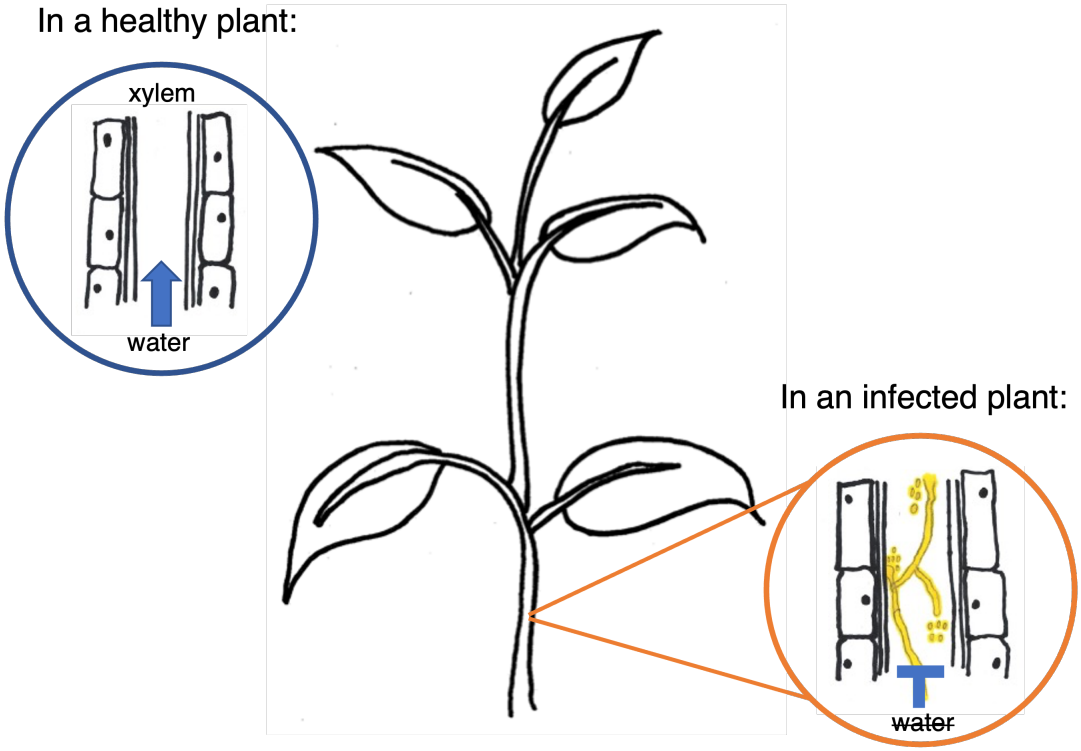
Recently the genomes of three important vascular wilt pathogens were published, *V. dahliae, V. albo-atrum* and *F. oxysporum* (Klosterman et al. 2011, Ma et al. 2010). This enabled the analysis of genome-wide pathogenicity analyses, rather than single gene knock-out studies, and indeed the *Verticillium* and *F. oxysporum* genomes encode more pectin-degrading enzymes and various CAZymes than other fungal pathogens, which suggests a high capacity to break down plant pectins (Klosterman et al. 2011). In *F. oxysporum* secreted in xylem (*SIX*) effectors were first identified by isolation of the proteins from the xylem sap of infected plants (Rep 2005, Houterman et al. 2007). Following the publication of the *F. oxysporum* chromosomal assembly (Ma et al. 2010), all twelve *SIX* effectors which were identified at this time were localised to a single chromosome, with four clusters of *SIX* effectors across its length (Schmidt et al. 2013). This chromosome is now understood to determine the host range of the *F. oxysporum* formae speciales (Van Dam et al. 2017, Li et al. 2020, 2021, Ayukawa et al. 2021). Indeed, recent work using discovery proteomics has uncovered an endophytic role for *F. oxysporum* when interacting with nonhost plants, to complement its well-characterised pathogenicity (Redkar et al. 2022). A new type of effectors has been characterised, termed early root colonisers, with a role in endophytic multihost compatibility in *F. oxysporum*.

The biology of *F. oxysporum* encapsulates the importance of understanding the evolutionary potential of a pathogen. Many of its host-specific *formae speciales* have polyphyletic origins (van Dam et al. 2016) and are defined by accessory chromosomes which carry the *SIX* effectors, pathogenicity genes and transposable elements (Ma et al. 2010, Schmidt et al. 2013). Its genome contains a core set of chromosomes which are shared by *F. oxysporum, F. verticillioides* and the distantly-related *F. graminearum* as well as four lineage-specific (and not shared with other *Fusarium* species) accessory chromosomes. Such genomes are known as 'two-speed' genomes, where genes found in repeat-rich and gene-poor regions such as on the accessory chromosomes have higher rates of evolutionary change, whilst slowly evolving housekeeping genes are found in repeat-poor and gene-rich regions (Dong et al. 2015, Frantzeskakis et al. 2018, Raffaele & Kamoun 2012). More recently, use of the term 'two-speed' is being reviewed, possibly with a preference to 'compartment' over 'speed' (e.g. (Frantzeskakis et al. 2019, Torres et al. 2020). Nonetheless, to use the original terminology, these 'fast-evolving' regions feature higher rates of mutation, recombination and horizontal gene transfers from other species (Stukenbrock 2013). In *F. oxysporum*, the accessory chromosomes are mobile and one carries all known *SIX* effector genes, which transfers pathogenicity between *formae speciales* (Ma et al. 2010, Schmidt et al. 2013). The ability to transfer pathogenicity suggests horizontal gene transfer, a trait common to bacteria and prokaryotes, but considered unusual in eukaryotes (Van Etten & Bhattacharya 2020). Recent findings suggest that it is much more widespread in eukaryotes than previously thought, and horizontal transfers are likely to be important contributors to eukaryotic genomes (McDonald et al. 2019, Rossoni et al. 2019, Ropars et al. 2015).

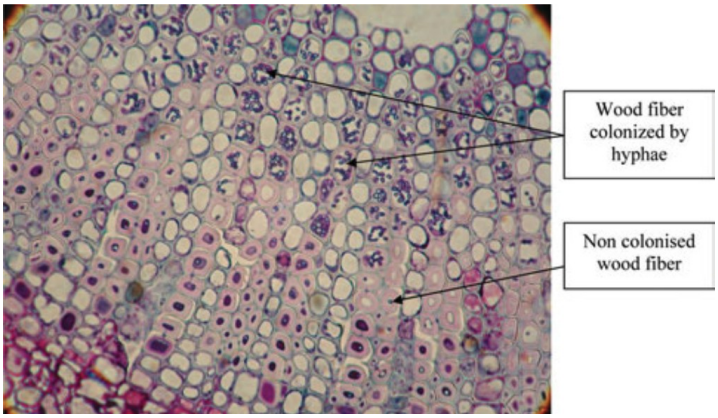
1.3 The genus *Fusarium*

Fusarium is a genus of ascomycetous fungal plant pathogens, which is widely distributed with devastating impacts on crop yields and the quality of staple grains. It can infect

all vegetative and reproductive plant parts, and over 80% of all crops grown globally are associated with diverse *Fusarium*-related disease (Nelson et al. 1983, Leslie & Summerell 2006, Arif et al. 2011). There are thought to be 20 monophyletic species complexes within



(a)



(b)

Figure 1.1: The mechanisms of a vascular wilt infection. A Schematic drawing comparing a healthy and infected plant. In a healthy plant, water passes unimpeded through the xylem vascular tissue. In a plant infected with a vascular wilt pathogen (shown in yellow), the fungus proliferates in the xylem, blocking the passage of water as it spreads and colonises. **B** The interaction between *Fusarium xylarioides* and an infected *Coffea canephora* seedling at the cellular level. Photo from D. Bieysse via CABI.

the *Fusarium* genus with diverse infection strategies across them. For example, as well as *F. oxysporum* and vascular wilt, others carry accessory chromosomes with evidence of horizontal transfer, such as *F. solani* and root rot (Coleman et al. 2009), and even those lacking accessory chromosomes have regions that are 'fast-evolving' and associated with pathogenicity, such as *F. graminearum* and head blight (Sperschneider et al. 2015). Some *Fusarium* produce mycotoxins, which are virulent to sensitive hosts and can cause safety risks with food and feed, affecting both humans and animals (Ward et al. 2008). One of such species complexes is the polytypic species complex known as *Fusarium fujikuroi* Wollenweber (O'Donnell, Cigelnik & Nirenberg 1998). The species within the *F. fujikuroi* complex cause a range of destructive plant diseases to economically important crops, including maize from *F. verticillioides* (Gerlach & Nirenberg 1982), sorghum from *F. thapsinum* (Klittich et al. 1997) and pigeon pea from *F. udum* (Booth 1971). Species within the *F. fujikuroi* complex are separated into African, American or Asian sub-clades based on phylogenetic species recognition and the putative biogeographic origin of their hosts (O'Donnell, Cigelnik & Nirenberg 1998). *Fusarium udum* and *F. xylarioides* are the only species within the *F. fujikuroi* complex to induce true vascular wilts in their hosts, a trait usually associated with the *F. oxysporum* complex. The ability to cause vascular wilts is inferred to be a plesiomorphic trait in the *F. oxysporum* and *F. fujikuroi* complexes (O'Donnell, Cigelnik & Nirenberg 1998, Geiser et al. 2005), or, given the various infection strategies within *Fusarium*, the trait could have been horizontally transferred.

Fusarium oxysporum, through its well-studied *forma speciales*, is able to cause vascular wilt disease on over 120 plant species (Michielse & Rep 2009). Another and less well-studied *Fusarium* wilt is coffee wilt disease, caused by the soil-borne fungal pathogen *F. xylarioides* (Rutherford 2006). Coffee wilt disease typically results in *F. xylarioides* proliferating in and subsequently blocking the plant xylem, restricting the transport of water around the coffee plant (Figure 1.1b). This manifests in wilting and desiccation of leaves, with defoliation and die-back (Figure 1.3a). Characteristic blue-black staining underneath the bark develops, as well as cracks to the bark with perithecia (the fungal

fruiting bodies) noticeable and a swelling trunk (Figure 1.2 and 1.3B-C). Other *Fusarium* species are often isolated from coffee wilt diseased trees, including *F. oxysporum, F. solani, F, stilboides* and *F. lateritium*, but only *F. xylarioides* results in coffee tree death (Tshilenge-Djim et al. 2016, Serani et al. 2007). Over the past century, coffee wilt disease is the most significant disease to have affected coffee production, and has caused over US$1 billion in losses to national economies in central Africa (Phiri & Baker 2009).

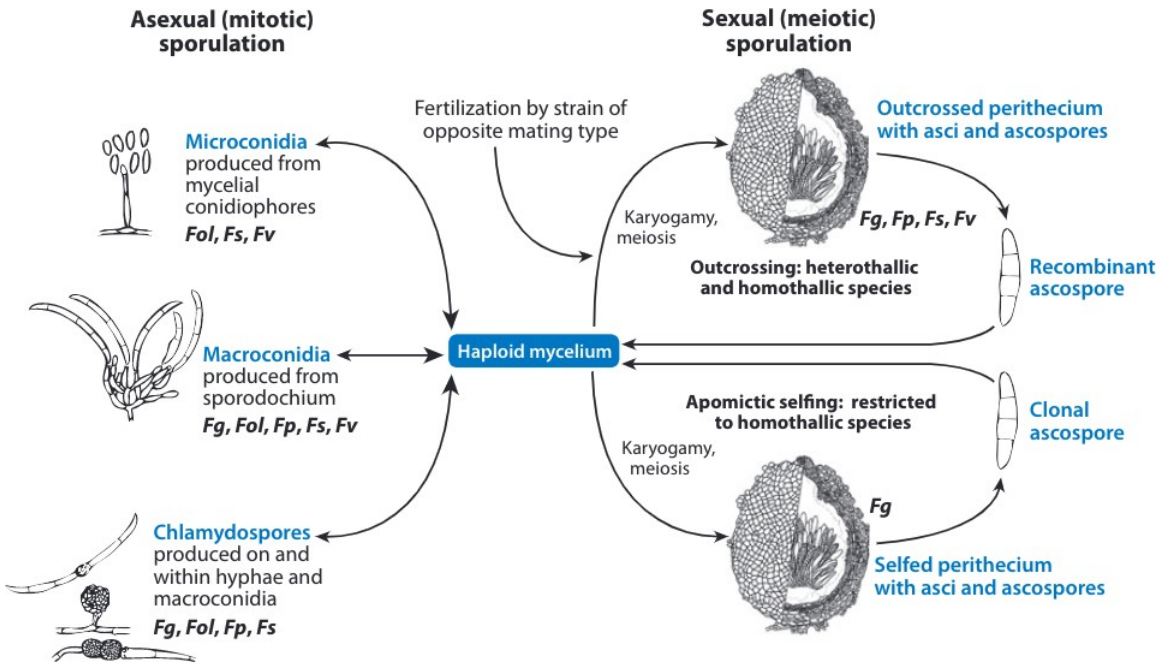


Figure 1.2: Generalised life cycle of *Fusarium*. *Fusarium* mostly features a haploid life cycle, with a brief diploid stage which precedes meiosis and the production of sexual haploid ascospores inside the perithecium. *Fusarium* is both heterothallic (which are self-sterile) and homothallic (which are capable of self-fertilisation). Across the genus, three types of asexual spore are produced: microconidia; macroconidia; and chlamydospores, which are thick-walled long-lasting spores. *Fusarium xylarioides* does not produce chlamydospores, instead perithecia are believed to be the most important source of inoculum in coffee wilt epidemics (Girma et al. 2009) and could constitute the means of survival for the pathogen on dead and decaying plant material in the soil (Rutherford et al. 2009). Abbreviations: *Fg, F. graminearum; Fp;. F. oxysporum* f. sp. *lycopersici; Fp, F. pseudograminearum; Fs, F. solani* f. sp. *pisi; Fv, F. verticillioides*. From Ma et al. (2013).

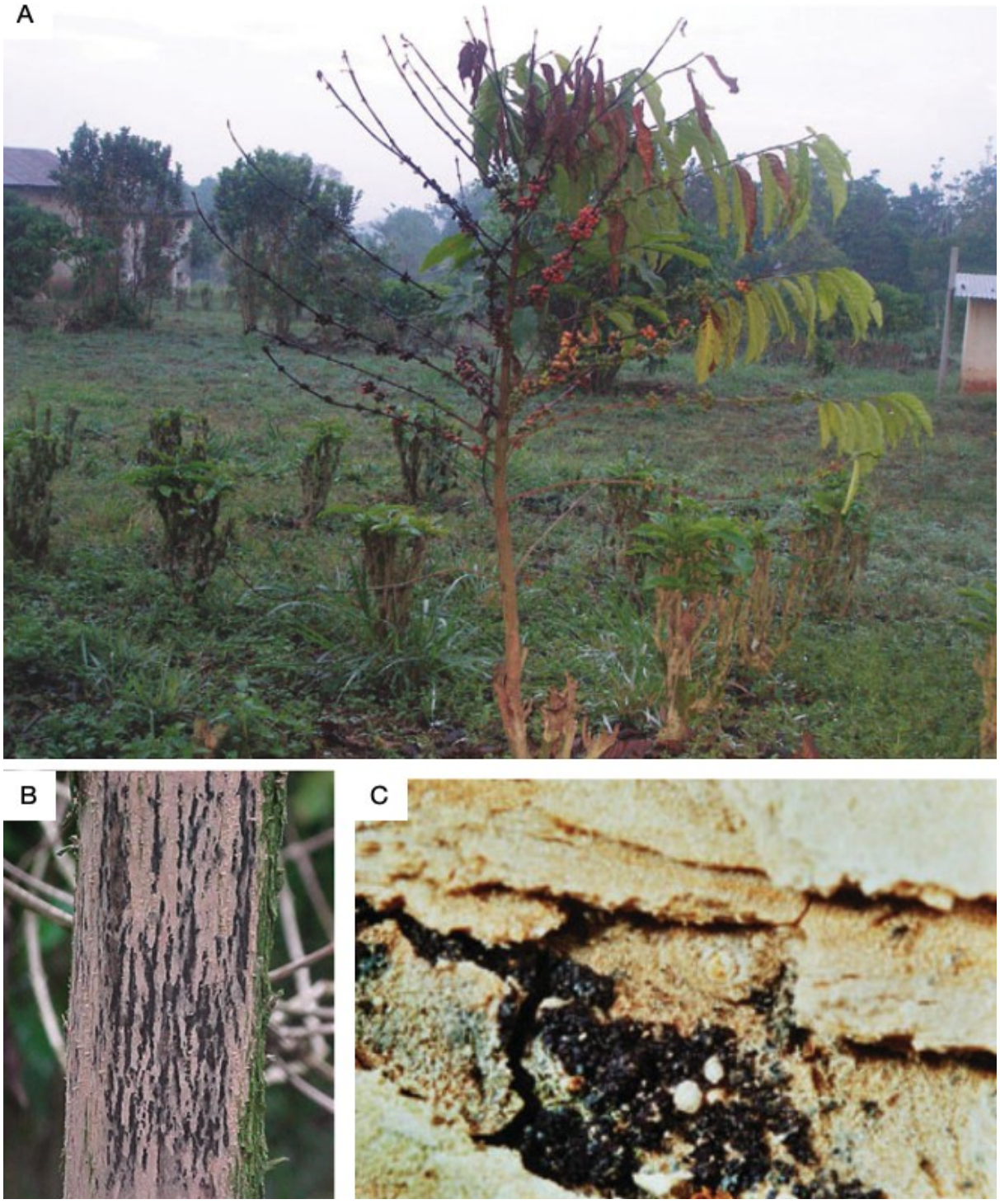


Figure 1.3: The progression of coffee wilt disease. **A** A robusta coffee wilt diseased tree in Uganda. **B** Stromatic fruiting bodies (perithecia and ascospores) of *Gibberella xylarioides* (the sexual stage of *Fusarium xylarioides*) in the wood of a dead arabica coffee tree. **C** Perithecia produced by *Gibberella xylarioides* in cracks of tree bark. Photos from CABI.

1.4 Coffee as a crop

Coffee is a key commodity crop and a multibillion-dollar industry largely dependent upon millions of small-scale farmers. Smallholder farms contribute 80% of all coffee grown in Africa, and its value to livelihoods is enormous (ICO 2022, Vega et al. 2003, Oduor et al. 2003). Indeed, it is the most valuable crop exported from the tropics (Craparo et al. 2015). Globally, its serious diseases are mostly caused by fungal plant pathogens: coffee leaf rust, coffee berry disease and coffee wilt disease. Coffee berry borer is the exception, which is spread via a beetle. Historically, coffee leaf rust and coffee berry disease have received the most attention (Hakiza et al., 2009). However, coffee wilt disease, being soil-borne, is harder to control and treat: either trees die or land must be left fallow, both of which carry a high long-term cost (Girma, 2004).

Coffee originated in east and central Africa and is a key foreign export: in 2013 it was worth nearly US$2 billion (FAO 2022), and a number of countries, including Ethiopia and Uganda rely on it for over half of their foreign exchange earnings (Oduor et al. 2003). Global trade largely depends upon two species: *Coffea arabica*, and *C. canephora* robusta (Mabberley 1997, ICO 2022). *Coffea arabica* (hereafter called 'arabica') grows at high altitudes (1500-2300m above sea level) across east Africa, while *C. canephora* robusta (hereafter called 'robusta') is hardier and found at warmer low-middle altitudes (800-1500m) throughout tropical Africa (Davis et al. 2006, Hakiza et al. 2009, Kalonji-Mbuyi et al. 2009).

1.5 Coffee wilt disease

1.5.1 Coffee wilt disease and its spread across Africa

Coffee wilt disease was first observed in 1927 on *C. excelsa* in Oubangui-Chari (now the Central African Republic, Figueres (1940)). Over the next two decades, diverse coffee crops in west and central Africa were devastated by the then unknown disease as it spread

widely across Africa. The fungal pathogen was isolated and described in 1948 and 1950 respectively (Heim & Saccas 1950, Steyaert 1948) and by the late 1950s, coffee wilt disease was detected in west and east Africa: Guinea and the Ivory Coast by 1958 (Chiarappa 1969), and Ethiopia by 1957 (Figure 1.4) (Stewart 1957, Lejeune 1958, Kranz & Mogk 1973). An international consensus to address escalating disease levels was reached, with three key pillars: the systemic uprooting and burning of all infected trees; seeking natural host resistance in both wild and cultivated varieties; and breeding programmes to utilise this host-plant resistance (Musoli et al. 2009). These measures successfully reduced disease levels across west and central Africa and led to a recovery of robusta coffee production. Coffee wilt disease declined to minor disease levels, but less so in the east where it was already widespread in Ethiopia in arabica coffee by the 1970s (Kranz & Mogk 1973, Van Der Graaff & Pieters 1978). Coffee wilt disease has persisted in Ethiopia on arabica coffee since this time, albeit at lower severity than the robusta outbreaks in east and central Africa (Oduor et al. 2003). Separately, in the mid-1970s, coffee wilt disease re-emerged on robusta coffee in abandoned farms in the Democratic Republic of Congo to reach epidemic levels in the 1980s (Phiri & Baker 2009). By the early 1990s it had spread from the Democratic Republic of Congo into neighbouring western Uganda, reaching eastern Uganda and becoming endemic across the country by the latter half of the decade (Birikunzira & Hakiza 1997). It was reported in northern Tanzania for the first time in 1995 (Kilambo & Kaiza 1997).

A quote by Caleb Dengu in Flood (2009) encapsulates the impact that the re-emergence of coffee wilt disease had on farmers in the 1990s. ’This is the story of an African coffee farmer in Uganda who I interviewed at the start of the project at the end of 1999. This farmer had 1000 coffee trees on his one hectare farm. He was earning US$ 1000 per annum from approximately 1000 kg of green coffee he produced from his field. He had a family of seven that included five children who ranged between 5 and 15 years. He did not hire any outside labour to assist with crop harvesting and processing, and he supplemented his income by intercropping his coffee field with bananas, which are staple food crops in

the area. . . By 1999, the yield from the coffee farm had reduced from 1000 kg to 400 kg, a 60% decline in income for the farmer. In absolute terms, this represents a decline of income from US$ 1000 to US$ 400 per annum for the family of seven. . . In real terms, this represents children losing access to education, healthcare, and sometimes basic food for survival.'

In the next sections, the nature of the initial outbreak and subsequent re-emergence will be considered in more detail.

1.5.2 The first coffee wilt disease outbreak, between the 1920s and 1950s

Today, global coffee production depends on two main commercial crops: arabica and robusta (ICO 2022). However, a hundred years ago, at the time of the first outbreak of coffee wilt disease, various types of coffee were grown throughout western and central Africa. Following the discovery of coffee wilt disease in the 1920s in the Central African Republic, it was widespread in *C. excelsa* and prevalent in other coffee species, namely

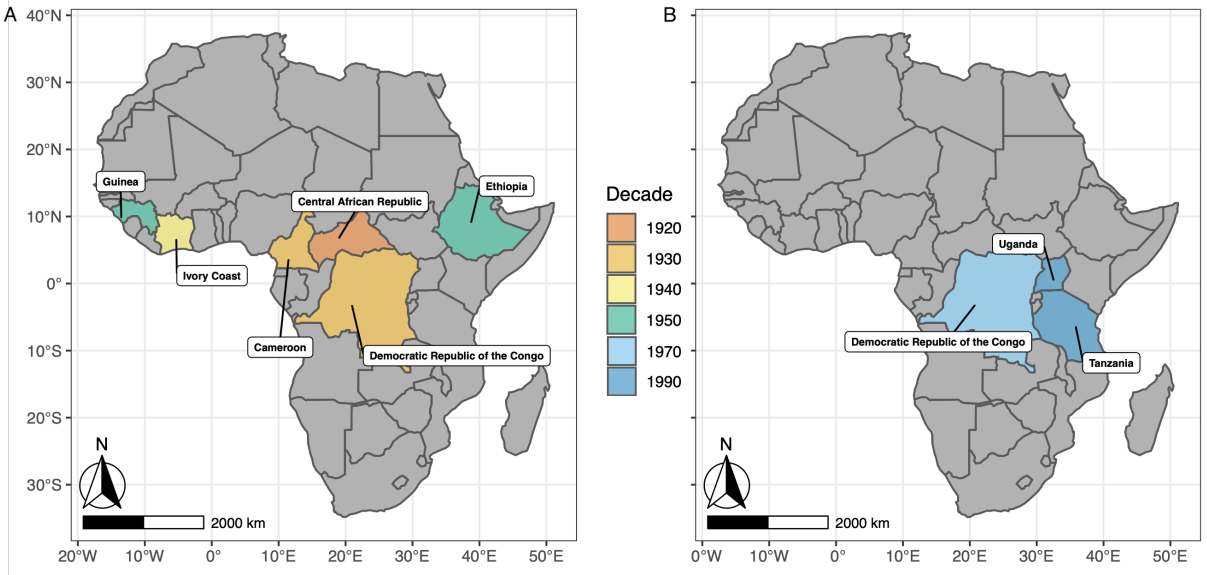


Figure 1.4: The identification of coffee wilt disease symptoms across Africa. A The first outbreak of coffee wilt disease, following its discovery in the Central African Republic in 1927. **B** The identification of coffee wilt disease symptoms across eastern Africa in its second outbreak. Data from ICO, accessed 05/07/2022.

C. canephora and *C. neo-arnoldiana* by the 1940s. In this same period, it was also found across neighbouring countries: in *C. excelsa* plantations in the north-east of Belgian Congo (today's Democratic Republic of Congo) and close to the border with Sudan (today's South Sudan) (Fraselle 1950, Steyaert 1948); in Cameroon in *C. excelsa* plantations; and in the Ivory Coast in *C. canephora* plantations (Flood 2009). Coffee wilt disease was probably present in sub-Saharan Africa at this time, but the disease could have been missed, or reported but not picked up internationally in the interceding countries between those listed above. In fact, there is a report of *F. xylarioides* isolated from rotting tomato fruits in a market in western Nigeria (Onesirosan & Fatunla 1976). So coffee wilt disease was at least present, even if at a low level, across the intervening west African countries.

The outbreak had severe consequences for coffee growers, with *C. excelsa* production in the Central African Republic and Cameroon destroyed. The disease spread to epidemic levels in robusta coffee in the Democratic Republic of Congo and Cote d'Ivoire, destroying more than half of its production and a key robusta coffee cultivar (Flood 2009). By the late 1950s, coffee wilt disease had spread west to Guinea, where it reduced coffee production by 50%(Chiarappa 1969), and east to Ethiopia on arabica coffee (Lejeune 1958). The commercial production of *C. excelsa* did not recover, fundamentally changing African coffee production from a largely *C. excelsa* crop to its replacement by robusta coffee in the middle of the century (Rutherford, 2006).

1.5.3 The second coffee wilt disease outbreak, from the 1970s onwards

After the implementation of systematic management procedures in the 1950s (i.e., the uprooting and burning of infected trees and cultivar breeding programmes), all published reports stated that coffee wilt disease no longer impacted arabica and robusta coffee production (Flood 2009). Indeed, it was not detected again in Guinea, Ivory Coast, Cameroon, and the Central African Republic. However, it appears that it was spreading throughout the 1970s and 1980s in remote areas of the Democratic Republic of Congo,

with reports from towns nearly 700km apart in the north-east state of Orientale of a wilt-like disease on robusta coffee (Phiri & Baker 2009). Following independence from Belgium in 1960, many foreign-owned plantations were abandoned to a local population who, as a result of Belgian governance, lacked the technical capacity to deal with an emerging disease problem (Peck n.d.). Thus *F. xylarioides* was able to proliferate across coffee plantations over two decades before it was detected, reaching Uganda and Tanzania in the mid-1990s (Flood 1996). It rapidly destroyed coffee crops but with a notable absence in Rwanda.

One theory for its spread into Uganda in the mid-1990s was through illegal coffee trafficking, which accelerated with the onset of the civil war (Flood 2009). War broke out in the Democratic Republic of Congo from 1997 to 1998, with illegal border crossings into Uganda appearing. Farmers who lived on this route became aware of coffee wilt disease symptoms, without understanding what was causing them, with reports of farms losing over half of their yields and profits (see section 1.5.1). In 1995, coffee wilt disease was formally identified when Philip Betts, the Managing Director of Esco Zaire, sent diseased plant material for testing to CABI, which was followed by a visit to the affected region in the Democratic Republic of the Congo (Flood 1996).

Fact-finding surveys were subsequently completed in Uganda in 1997, with coffee wilt disease present in 10 major coffee-growing districts (Hakiza et al. 2009). Research grant proposals were initiated and the regional coffee wilt programme was created with funding from the common fund for commodities, the European union and the UK department for international development (Flood 2009). However, management activities to limit its spread were not initiated until the regional coffee wilt programme was launched, some 25 years after its re-emergence in the Democratic Republic of Congo and five years in Uganda (Phiri et al. 2009).

In 2009, it was estimated that a third of coffee farms in the Democratic Republic of the Congo had been abandoned (Kalonji-Mbuyi et al. 2009) and exports had collapsed from an average of 85k tonnes between 1985 and 1995 to below 20k tonnes in the following

decade (ICO 2022). From 1996, the Ugandan coffee trade began to suffer and shortly thereafter the trade collapsed. Almost half of Uganda's robusta crop was destroyed with economic implications >US$100 million and affecting nearly half of Uganda's smallholder farms (Mbalibulha et al. 2015, Phiri & Baker 2009). Ugandan coffee exports continued to decline for nearly a decade before beginning to recover in 2005 but did not exceed the 1997 export volume until 2020 (Peck n.d.). The Democratic Republic of Congo coffee industry has never recovered, and tens of coffee production factories, which used to process up to

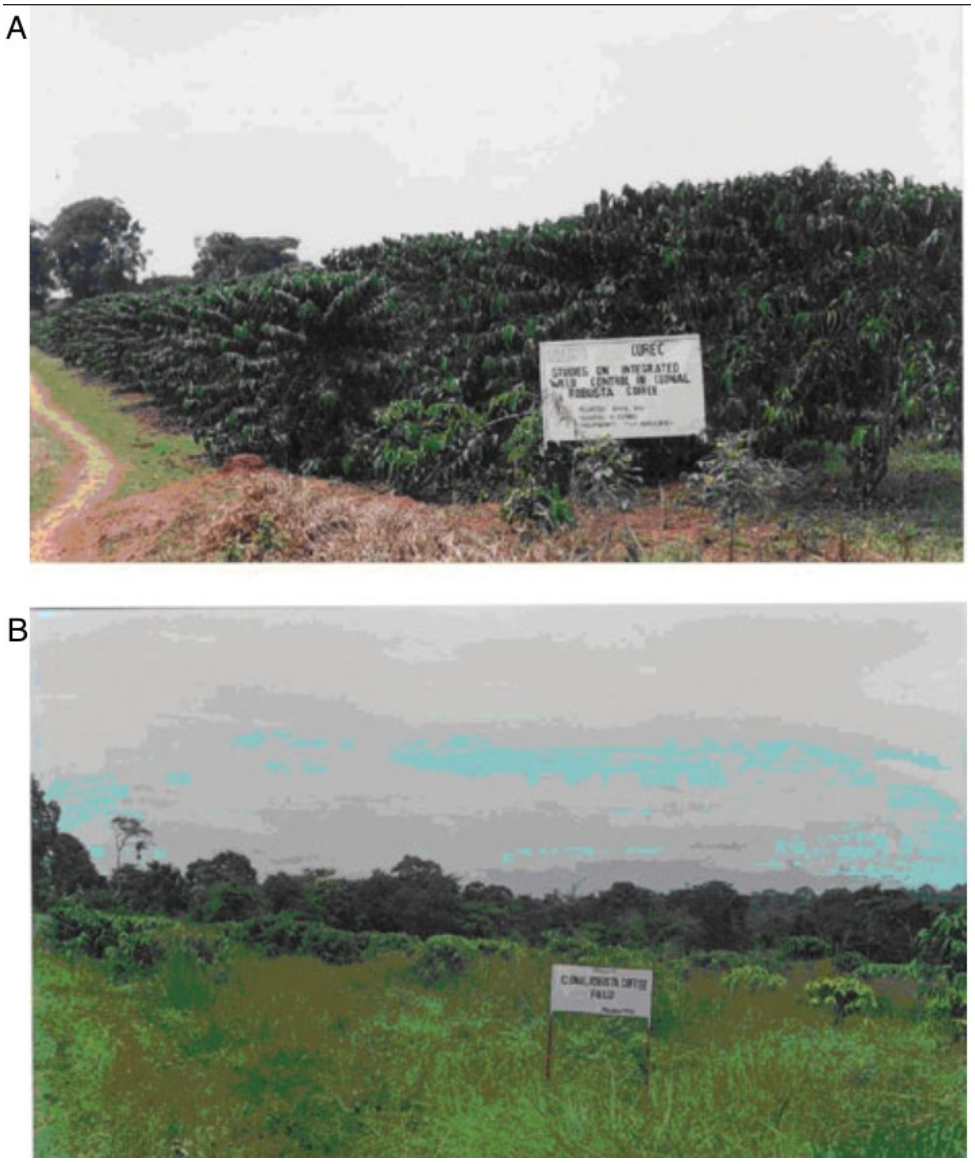


Figure 1.5: A coffee field is destroyed by coffee wilt disease. A A healthy and productive Ugandan coffee field in 1993, and **B** The same field in 1998 has few coffee trees remaining. Photo from G. Hakiza, via CABI.

three thousand tonnes per year, now stand empty in the eastern Democratic Republic of Congo (Betts 2022). On the contrary, the recovery of Ugandan coffee production fared much better: clonal resistant coffee seedlings were developed and widely adopted, some production was replaced with cocoa and bananas, farm extension workers were trained to disseminate information and coffee growing was moved to less traditional areas in the north (Hakiza et al. 2009).

1.5.4 Coffee wilt disease in Ethiopia

Ethiopia has a long tradition of coffee consumption, as the centre of origin of arabica coffee (Girma et al. 2009). The arabica-specific coffee wilt disease was discovered in the 1950s, so after the initial outbreak and before the re-emergence of the robusta-specific disease (Lejeune 1958). By the mid-2000s, incidence and severity remained low, and consequently, Ethiopian coffee production did not suffer large drops in exports (ICO 2022).

However, its prevalence and impacts are reported to have risen over the past two decades with increasing economic losses (Mulatu & Shanko 2019, Oduor et al. 2003). The areas with the lowest incidence are the forest and semi-forest coffee systems, where wild and semi-wild coffee trees are harvested from within tropical rainforest systems (Girma et al. 2009). Lower disease incidence and severity is possibly thanks to local coffee varieties with higher quantitative resistance present. These heterogeneous coffee populations likely possess high levels of genetic diversity and resistance, preventing a rapid emerging epidemic such as could occur on homogeneous farms.

1.6 *Fusarium xylarioides* contains host specialists

A widespread biological and socioeconomic assessment was completed in 2003 in eastern and central Africa (specifically Ethiopia, Uganda, Democratic Republic of Congo, Rwanda, and Tanzania) as part of the regional coffee wilt programme to establish baseline data for coffee wilt disease. Disease incidence in farms across the region (i.e. the

percentage of coffee farms with infection) and severity (i.e. percentage of coffee trees infected per farm) were measured in over five thousand farms (Oduor et al. 2003). In general, incidence and severity were variable, with coffee wilt disease endemic in Ethiopia and Uganda, spreading in Democratic Republic of Congo, restricted in Tanzania to its robusta-growing north and absent in arabica-growing Rwanda (Figure 1.6). Uganda was badly affected, with coffee wilt disease present in 90% of farms surveyed, with 45% of trees infected per farm, while Ethiopia had a medium incidence with very low severity (Figure 1.6b-c). The average incidence was greater than 30% which, for a systemic disease that ends with tree death, is high (Oduor et al. 2003).

Two distinct *F. xylarioides* host-specific strains were revealed, infecting arabica coffee in Ethiopia and robusta coffee in Uganda, Democratic Republic of Congo and northern Tanzania, which were unable to infect non-host coffee trees even when adjacent to infected neighbours. This explains the absence in Rwanda: Rwanda predominantly grows arabica coffee, yet borders the three robusta-infected countries. Consequently, it was proposed that *F. xylarioides*, rather than a single species, could be a species with *formae speciales* (Girma 2004), a species complex containing cryptic species (Geiser et al. 2005) or multiple species including *F. abyssiniae* (the arabica population) and *F. congoensis* (the robusta population) (Lepoint 2006). It is possible that each host-specific *F. xylarioides* strain emerged along with the host coffee partners, with the centre of origin for both plant and pathogen in Ethiopia and tropical lowland Africa for arabica and robusta respectively.

Identification and diagnosis was historically based on morphological species recognition, requiring taxonomic expertise. There has been considerable debate over the taxonomic labelling of *F. xylarioides* since its initial description by Steyaert (1948), which was then formally and incorrectly described by C. Booth in 1971. Booth (1971) described the species with separate ’male’ and ’female’ strainsm which was later questioned by Gerlach & Nirenberg (1982) and Nelson et al. (1983), both of whom, as well as the original description by Steyaert (1948), only described the ’female’ strain morphology. In fact, Booth’s (1971) ’male’ *F. xylarioides* strains was later found to be *F. stilboides* (Gerlach

1978), before Geiser et al. (2005) used phylogenetic species recognition to reveal that the 'female' strains are true *F. xylarioides* and belong to the *F. fujikuroi* species complex, whilst the 'male' strains correspond to the *F. lateritium* clade. However, F. xylarioides is

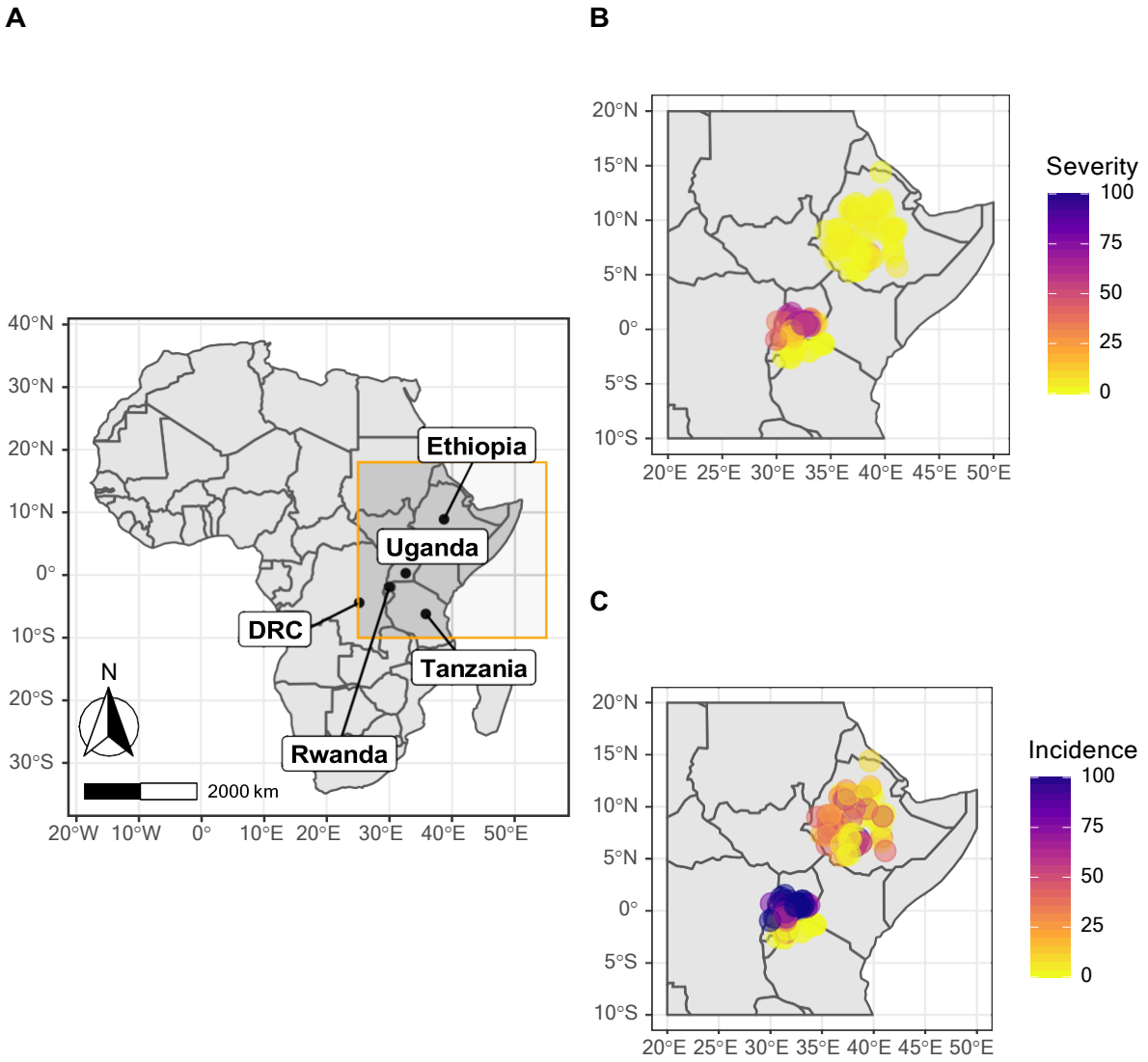


Figure 1.6: Coffee wilt disease incidence and severity in Ethiopia, Uganda and Tanzania. A Affected countries and the area surveyed shown in orange. Only arabica coffee was infected in Ethiopia and only robusta coffee in Uganda and Tanzania. **B** Severity in farms surveyed. Severity indicates the percentage of trees affected on each farm visited. **C** Incidence in farms surveyed. Incidence indicates the percentage of farms with coffee wilt disease presence across the region surveyed. Data is from Oduor et al. (2003), no data for the Democratic Republic of Congo (DRC) and no disease presence recorded in Rwanda.

heterothallic, with morphologically identical separate mating types (Lepoint et al. 2005).

Identification of species in *F. xylarioides* is further complicated because spore morphological characters cannot distinguish between arabica and robusta strains. The only morphological test able to do so is growth on a nutrient-rich potato dextrose agar medium with exposure to UV light – under these conditions the robusta strains develop an orange pigmentation, which is absent from the arabica strains (Figure 1.7) (Lepoint 2006). Whilst *F. xylarioides* does not have heterothallic male and female strains, it is heterothallic in mating type, so biological tests are possible because perithecia are formed when compatible mating types are brought together. Lepoint et al. (2005) initially identified three biological species within *F. xylarioides*: robusta strains and two species each containing two historic strains isolated in the initial outbreak (1920s-1950s). However, molecular methods offer the greatest accuracy in identification. Individual gene studies combining

Figure 1.7: Different pigmentations in the arabica and robusta strains grown under UV light on potato-dextrose agar. The failure to produce orange pigmentation characterises the arabica host-specific *Fusarium xylarioides* strains (white colonies) from the robusta strains (orange colonies). Photo from P. Lepoint and H. Maraite via CABI.

three genes (*Translation elongation factor 1-a, Calmodulin* and *Histone 3*) were able to identify an additional species to make four phylogenetic species within *F. xylarioides*: arabica, robusta, and the two historic species groups (Lepoint et al. 2005). Only in the last two decades, the genomic basis to host-specific disease populations has begun to be investigated.

1.7 The genomic basis to emerging plant pathogens

Recent decades have seen a rise in, on the one hand the number of pathogenic fungi affecting agricultural, natural and human environments, and on the other hand, high-throughput sequencing and phenotyping technologies (Ball et al. 2020). These are allowing the diversity and evolution behind such emerging plant pathogens to be understood, in a way that was not possible even two decades ago at the height of the coffee wilt disease epidemic. Various fungi are emerging as pathogens in diverse ways, by acquiring new types of or higher levels of pathogenicity, or colonising new niches (Fisher et al. 2020). This ability to rapidly increase in pathogenicity or range is underpinned by complex and dynamic genomes that can undergo rapid changes to overwhelm host defences and invade new ecological or host niches (Gladieux et al. 2014). Therefore, it is important to understand the processes by which these new fungal traits can evolve to better understand the risk of the emergence of new types of fungal pathogenicity (Farrer & Fisher 2017). Initiatives such as the Darwin tree of life aim to deliver this, which, among other eukaryote genomes, will sequence plant pathogens on a major scale (Blaxter 2022).

Previously, the analysis of historic genetic material involved the use of highly degraded museum specimens and, in the best case, molecular markers and target probes in a small number of genes (Brewer et al. 2019). However, fungal culture collections represent an ideal opportunity to harness the new technologies and historic assemblages of plant pathogens. They contain tens of thousands of fungal isolates collected over the past century and preserved in a living state. If collections are connected with initiatives

such as the Darwin tree of life, new -omics studies could reveal such processes at a new inter- and intra-species level (Ryan et al. 2022). Fungi undergo divergence and speciation through genomic processes that include gene family expansion, gene gain, introgression, and horizontal gene transfer (Douglas & Langille 2019, Gladieux et al. 2014, Ropars et al. 2015, Borneman et al. 2016). Recently, studies have shown the value of comparative genomics in elucidating the role of these changes in domesticated fungi and plant pathogens, including a host jump in *P. infestans* followed by adaptive specialisation (Raffaele et al. 2010), experimental verification of the transfer of pathogenicity into a previously non-pathogenic strain in *F. oxysporum* (Ma et al. 2010), and the transfer of regions that allow adaptation to cheese environments in distantly related species (Ropars et al. 2015). This thesis explores such processes to investigate the emergence and divergence of coffee wilt disease, using bioinformatic techniques and various *Fusarium* strains from the CABI-IMI culture collection.

1.8 Aim and questions of this thesis

This PhD research project investigated the coevolutionary processes between pathogen and host in a food system, with a focus on coffee wilt disease and its causal agent *F. xylarioides*. Efficient food production methods are needed that guarantee adequate supply and economic returns while limiting environmental impacts. A key limiting factor in global food availability is the control of pathogens and their subsequent disease epidemics, and a challenge is predicting when and where new types of disease will emerge. Despite the importance of pest and disease control for food security, our understanding of what drives this co-evolutionary process, and therefore the spread of disease, is limited. Specific research questions have been developed on the basis of this literature review and are detailed below.

1.8.1 Research approach

This project used genomes of historic fungal pathogens to reconstruct historic genetic changes to establish links between those changes and parameters relating to the disease environment. This was carried out in collaboration with CABI, the project CASE partner, and an agricultural biosciences research organisation. The CABI culture collection of cryopreserved living fungal strains and whole genome sequencing techniques were used to achieve a historic understanding of *F. xylarioides*. This work built upon CABI's regional coffee wilt programme and initial work to verify *F. xylarioides* as the causal agent behind the re-emergence of coffee wilt disease in the 1990s.

1.8.2 Aim and questions

The overarching aim of this thesis was to understand the evolutionary processes behind repeated outbreaks of fungal pathogens. Genomes from the CABI-IMI collection were analysed to understand the pathogen and its divergence into host-specific races. Additionally, it was hoped that this research could inform policy or growers to be better informed and prepared for a future re-emergence.

Specific questions for this thesis were:

1. Is *Fusarium xylarioides* a single species, or a species complex?
2. Did later outbreaks descend from the earlier initial outbreak, or were they separate origins from different wild sources?
3. What genes might underlie pathogenicity in *Fusarium xylarioides*?
4. What genetic changes underpinned the emergence of the host specialists and did horizontal gene transfer bring in new sets of pathogenicity genes?

Chapter 2

Comparative genomics reveals differences between host-specific strains of the *Fusarium* coffee wilt pathogen

2.1 Acknowledgements

I thank G. Senbeta, D. Kilambo, J. Flood, G. Hakiza, J. Kranz and M. Moreau who originally collected the fungal isolates and deposited them in the CABI-IMI collection, without which none of this work would have been possible.

2.2 Abstract

This chapter uses genome sequencing of six historical culture collection strains spanning 52 years to identify the evolutionary processes behind these repeated outbreaks. Phylogenomic reconstructions and a screen for putative effector genes showed that the host-specific *F. xylarioides* groups have diverged in gene content and sequence mainly by

vertical processes within lineages, and through expansion of certain carbohydrate-active enzyme families. A subset of putative effector genes, however, showed evidence for horizontal acquisition and close homology to genes from *F. oxysporum*.

2.3 Introduction

Fungal diseases have devastated major crop yields throughout history and continue to do so (Fones et al. 2017, Strange & Scott 2005). Large-scale planting of crops generates strong selection for new pathogens to emerge, which leads to further rounds of plant breeding to develop new resistant genotypes. This leads to 'boom and bust cycles' that intensify the natural co-evolutionary dynamics of hosts and pathogens. A key goal for sustainable agriculture is therefore to predict disease outbreaks and design robust evolutionary solutions for long-term protection (Burdon & Thrall 2008). A first step towards this goal is to understand the genetic and evolutionary mechanisms by which pathogens overcome resistance and infect new host species. Plants have innate defence responses to detect and overcome pathogen attack (Jones & Dangl 2006). In response, an emerging pathogen can evolve new mechanisms to suppress and overwhelm basal plant defences. These could arise by mutation (including gene duplication or loss), recombination and selection operating within a single population, or from hybridization and/or horizontal gene transfer between species to generate new pathogenicity variants (Möller & Stukenbrock 2017, Sanchez-Vallet et al. 2018). Strong selection to evade plant immunity also leads to host-specificity, whereby pathogens evolve to target particular species or varieties (Sanchez-Vallet et al. 2018).

Comparative genomics is revealing the mechanisms that promote rapid evolution of effector proteins and host specificity in fungal pathogens. Effector genes are often found in highly mutable parts of the genome (de Jonge et al. 2011). For example, in ascomycete fungi effectors often occupy AT-rich compartments of the genome with high mutation rates or cluster with transposable elements (TEs), which increase variation via duplica-

tions, deletions, insertions and inversions (Ma et al. 2010, Klosterman et al. 2011, Rouxel et al. 2011, Chuma et al. 2011, Schmidt et al. 2013). In addition, many ascomycetes have mechanisms to facilitate horizontal transfer of effector genes between taxa either by ”pathogenicity islands” in which pathogenicity genes and TEs cluster in chromsosomal segments depleted in GC (Han et al. 2001) or by whole mobile chromosomes carrying suites of effectors (Ma et al. 2010). For instance, the host-specific virulence protein ToxA was transferred among three wheat pathogens on an 14 kb DNA fragment that is rich in transposons and still actively mobile in one of the species (McDonald et al. 2019), whereas the ability of *F. oxysporum* f. sp. *lycopersici* (hereafter called *F. o. lycopersici*) to infect tomatoes derives from a lineage-specific mobile chromosome, which can be transferred experimentally between strains. Pathogenicity can therefore evolve by mutation, recombination and selection operating within a single lineage, or from horizontal gene transfer between strains to generate new pathogenicity variants.

Although comparative genomics has uncovered mechanisms behind host specialisation in several fungal pathogens, exactly how these processes play out during disease cycles remains less clear. Studies have mostly compared contemporary lineages with different host specialisations, rather than tracking genetic changes over time. For example, understanding the roles of within-lineage evolution versus horizontal transfer in generating new effector gene combinations would benefit from comparing genomes before and after boom-bust cycles, as well as between differentially adapted host-specialists.

Here, six historic strains collected over 52 years are used to investigate successive outbreaks and the origin of host specialisation in *Fusarium xylarioides* Steyaert, a soil-borne fungal pathogen that causes coffee wilt disease. Coffee wilt disease first emerged as a devastating disease of *Coffea excelsa* and *C. canephora* crops in west and central Africa from the 1920s to 1950s (Figueres 1940, Saccas 1956) (Figure 2.1). Improved crop sanitation and breeding programmes in the 1950s successfully reduced its impact but coffee wilt disease later re-emerged in the 1970s, spreading extensively throughout the 1990s and 2000s (Hakiza et al. 2009, Girma et al. 2009, Kalonji-Mbuyi et al. 2009).

At around the same time that coffee wilt disease re-emerged on 'robusta' *C. canephora* robusta coffee, it was also reported in Ethiopia on 'arabica' coffee (*C. arabica*) (Stewart 1957, Lejeune 1958) and *F. xylarioides* was confirmed as the causal agent (Kranz & Mogk 1973, Van Der Graaff & Pieters 1978). By the 1990s, coffee wilt disease was causing widespread destruction of arabica coffee in Ethiopia (Girma et al. 2009), and robusta coffee in the north-east Democratic Republic of Congo, Uganda and northern Tanzania (Hakiza et al. 2009, Kalonji-Mbuyi et al. 2009). It now comprises two host-specific and geographically separated populations, one on robusta coffee in Uganda, Tanzania and

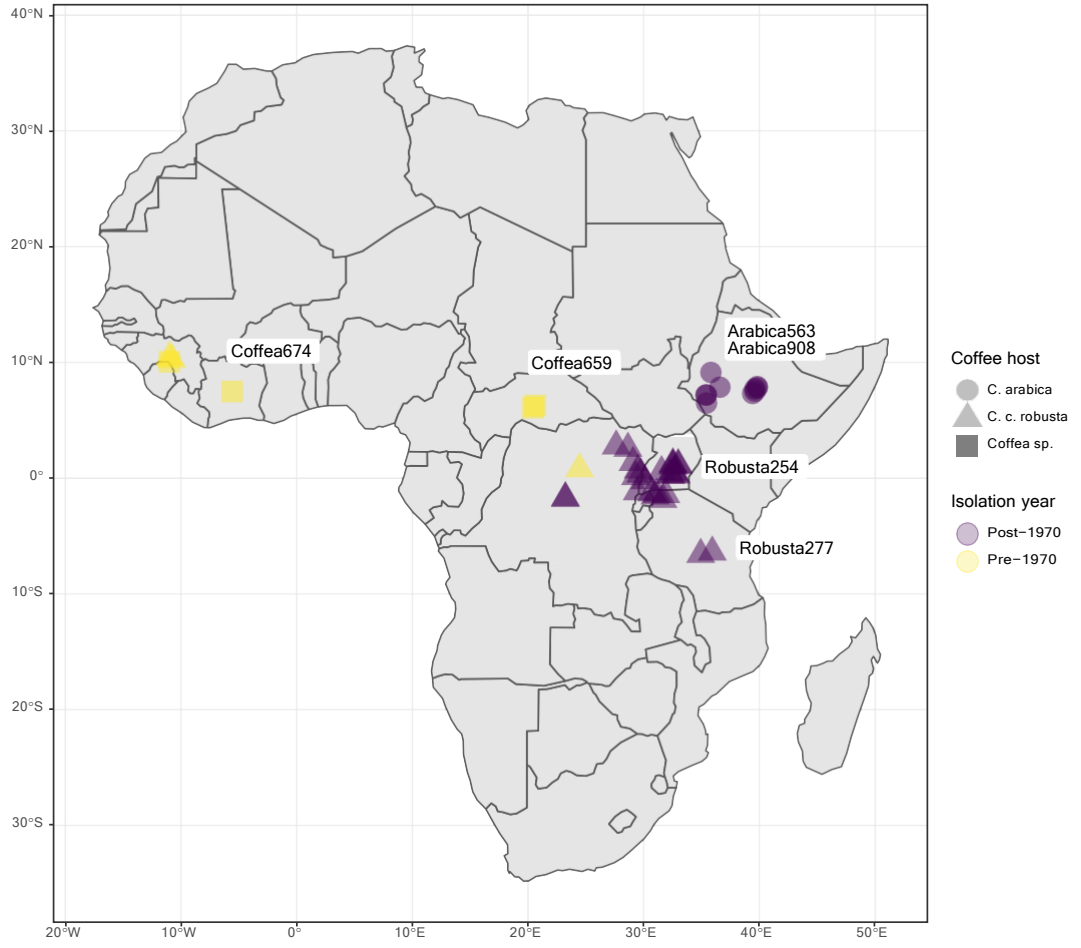


Figure 2.1: The emergence and spread of *Fusarium xylarioides*. A map of Africa detailing the year collected, country of origin and coffee plant host for the 62 *Fusarium xylarioides* strains in the CABI-IMI culture collection. These strains illustrate the spread west of coffee wilt disease from the 1920s-1950s outbreak strains to the 1970s onwards strains, and the emergence of the host-specific arabica and robusta populations. The six strains sequenced in this study are labelled on the map as: coffea674, from Cote d'Ivoire; coffea659 from the Central African Republic; robusta254, from Uganda; robusta277, from Tanzania; arabica563 and arabica908, from Ethiopia. Full strain biogeographic information is in Table 2.1.

Democratic Republic of Congo and the other on arabica coffee in Ethiopia (Figure 2.1). Both populations have caused significant losses to the coffee cash crop, on Africa's two most valuable species (Olal et al. 2018, Mulatu & Shanko 2019). *F. xylarioides* therefore offers a unique study system with repeated epidemics and the apparent emergence of two host-specific populations (Lepoint et al. 2005, Flood 2005, Phiri & Baker 2009). Critically, historical living strains from the initial (1920s-1950s) outbreak as well as the more recent, are optimally cryopreserved in a living state in culture collections.

Previous work described the pathology of the epidemics (Rutherford et al. 2009), clarified molecular taxonomy (Buddie et al. 2015, Flood 2005), showed reproductive isolation between the host-specialists (Lepoint et al. 2005), and reported the first genomes for the robusta population (Olal et al. 2019, Wingfield et al. 2019). The genetic basis for successive outbreaks and host-specialisation remains unexplored, however. Wilting occurs when a pathogen proliferates in and blocks the host xylem, so restricting water transport (Tjamos & Beckman 1989, Mace et al. 1981). In order to colonize the xylem vascular system, effector proteins, including carbohydrate-active enzymes (CAZymes) such as cellulases and pectinases, are required by the fungus to degrade and penetrate the root system (Sharma et al. 2016, Clérivet et al. 2000). In *F. oxysporum*, effector proteins behind wilt induction (termed *SIX* for secreted in xylem) are encoded by a single mobile, pathogenicity chromosome (Ma et al. 2010, Schmidt et al. 2013). As a result, the same host-specific *formae speciales* can have polyphyletic origins, as the ability to infect a particular host is transferred horizontally (Abo et al. 2005, Ellis et al. 2014, 2016, O'Donnell, Kistler, Cigelnik & Ploetz 1998). Whether similar mechanisms apply for *F. xylarioides* and coffee wilt disease, and how pathogenicity is restored between successive disease outbreaks, remains unknown. Intriguingly, coffee is intercropped with banana, and *F. xylarioides* and *F. oxysporum* have been co-isolated from roots of both plants in Uganda, and from coffee roots in Ethiopia (Flood 2005, Serani et al. 2007). Indeed, *F. oxysporum* is able to infect coffee, where it induces a wilt but does not result in the trees' death (Serani et al. 2007). These findings raise the possibility that *F. xylarioides* may

have acquired certain pathogenicity genes from *F. oxysporum*, that has facilitated the recent outbreaks on coffee.

To address these questions, the genomes of six representative historical *F. xylarioides* strains from the CABI-IMI living culture collection were sequenced and compared: two strains derive from the 1950s and therefore the initial outbreak in Central African Republic and Cote d'Ivoire: coffea659 and coffea674 (Table 2.1. I call these coffea strains because of their ability to infect multiple coffea species including robusta, and their original hosts are mostly unknown. Coffea674 is the ex-type and in common with most strains from the initial (1920s-1950s) outbreak infects robusta and other *Coffea* species but not arabica (Lepoint et al. 2005). Coffea659 is one of the few strains able to also infect arabica in trials and therefore is a true host generalist (Girma 2004). The remaining four sequenced strains comprise two arabica strains (arabica563 and arabica908), and two robusta strains (robusta254 and robusta277), all collected five years apart between 1995-2005 (Figure 2.1). Current evidence from molecular markers and crossing experiments supports the distinctiveness of the arabica and robusta populations (Buddie et al. 2015, Lepoint et al. 2005), but varies with respect to relationships between them and with the initial outbreak's coffea strains. Differing studies show the 1990-2000s arabica and robusta populations as sister clades (Lepoint 2006), or the 1990-2000s robusta population grouped with the coffea strains from the pre-1970s outbreak (Flood 2005), suggesting it arose from a subset of older strains from the initial outbreak whereas the arabica population is more divergent. Thus, I first tested the hypothesis that the robusta population derived from the initial (1920s-1950s) outbreak with the arabica population emerging separately. I then compared putative effector genes between the strains. Specifically, I ask (i) whether the 1990-2000s robusta epidemic is genetically different to the earlier outbreak; (ii) whether the host-specialist robusta and arabica populations share similar sets of derived effector genes or whether their similar pathologies evolved independently; (iii) I test whether changes in pathogenicity and host-specialism involved horizontal transfer of effector genes, or was restricted to within-lineage evolution in ancestral sets of effector genes using comparisons

with potentially co-occurring and closely related *Fusarium* species, and (iv) explore the possible role of mobile chromosomes or transposable elements in any putative cases of horizontal transfer.

2.4 Methods

2.4.1 Strain information

Six *F. xylarioides* strains were selected for Illumina MiSeq sequencing (Table 2.1). All strains were originally collected over a 52 year period from coffee wilt diseased trees and stored in the CABI-IMI culture collection (CABI, Egham, Surrey, TW20 9TY). Strains were grown on synthetic low nutrient agar at 25^{o}C, and potato dextrose agar (200g potatoes, 20g agar, 15g glucose, 1 litre distilled water) with UV irradiation in order to confirm morphological identification following (Buddie et al. 2015). Following Cubero et al. (1999), strains were grown in GYM (glucose, yeast and malt) broth and genomic DNA was extracted from <20mg of washed mycelium, which was frozen in liquid nitrogen before extraction following the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) standard protocol.

2.4.2 Genome sequencing and assembly

For each strain, a single library was prepared with the Illumina TruSeq PCR-free kit and sequenced with the Illumina MiSeq v3 600 cycle kit with 2 x 300 bp paired end sequencing and 350 bp insert size at the Department of Biochemistry (University of Cambridge, UK). Low quality bases and adaptors were identified and trimmed (quality score <20) using TrimGalore 0.6.0 by Cutadapt (https://github.com/FelixKrueger/TrimGalore, Martin (2011)). Overlapping paired-end reads were merged using FLASH 1.2.11 (Magoč & Salzberg 2011) before being assembled using MEGAHIT 1.2.8 (Li et al. 2015*a*) with a final k-mer length of 189. Assembly metrics were computed using QUAST 5.0.2 (Mikheenko et al. 2018) (Tables 2.1 and A.1). The quality of the genomes in terms of the presence

of 303 core eukaryotic genes and 290 core fungal genes was assessed using BUSCO v3.0.2 (Simão et al. 2015). For comparison, I selected eleven genomes of related species (Table A.10) selected for being: closely related and wilt-inducers (*F. udum, F. oxysporum* f. sp. *lycopersici* and *F. oxysporum* f. sp. *cubense*); closely related which occupy the same coffee-plant niche (*F. verticillioides , F. solani* and *F. oxysporum*); closely related (*F. fujikuroi, F. mangiferae, F. proliferatum*); and wilt-inducers (*V. dahliae* and *V. albo-atrum*).

Chromosomal and lineage-specific regions in each *F. xylarioides* assembly were identified by whole-genome alignment to the chromosomal level assembly of *F. verticillioides* (GenBank accession GCA_003316975.2, Ma et al. 2010). The coffea659 assembly (chosen for having the larger genome of the two coffea strains) was aligned against *F. verticillioides* using RaGOO v1.1 with the '-C' parameter (Alonge et al. 2019), with the remaining *F. xylarioides* strains aligned in-turn against the assembled RaGOO coffea659. I used BLAST (with 1.00e-50) to identify scaffold specific groups between the closely related *F. xylarioides, F. udum* and *F. verticillioides*: scaffolds which did not match to *F. verticillioides* but were present in each *F. xylarioides* strain and the *F. udum* genome (GCA_002194535.1, Srivastava 2018) were interpreted as *Fusarium xylarioides* and *udum*-specific (FXU). Scaffolds which were present only in *F. xylarioides* were interpreted as *F. xylarioides*-specific (FXS), while scaffolds which did not match to the historic coffea659 genome nor the *F. verticillioides* assembly nor *F. udum* were concluded to be lineage-specific (LS). Here, the LS scaffolds relate to the host-specific populations.

2.4.3 Gene prediction and orthologous clustering

Protein-encoding genes were predicted using the BRAKER v2.1.2 pipeline (Stanke et al. 2006, 2008, Li et al. 2009, Barnett et al. 2011, Hoff et al. 2016, 2019), using RNA-seq data from the closely related *F. verticillioides* to guide gene models. RNA-seq reads (SRR10097610) were mapped to each genome using STAR v2.7.3a (Dobin et al. 2013) and the resultant BAM file was input to BRAKER with default settings. BUSCO

analysis was performed to assess quality of predicted proteins, and the GC content of coding sequences was calculated. This was used to calculate GC-equilibrated regions in 20 kb windows, where GC-rich blocks have a GC content within 1% of that of the coding sequences, as well as AT-rich blocks, where the GC content is <= 0.7 of the coding sequences content (Rouxel et al. 2011). A Kruskal-Wallis chi-squared test confirmed association between genes and the GC-rich blocks, with a Wilcoxon rank test (Bonferroni adjustment) to confirm significance between pairs. Three programs were used to infer functional classification: InterPro v5.35-74.0 (Jones et al. 2014); Superfocus (Silva et al. 2016) which classifies proteins by their SEED categories (Aziz et al. 2012); and NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using megaBLAST searches and the nr database. All proteins lacking significant hits were annotated as hypothetical proteins. OrthoFinder v2.3.8 (Emms & Kelly 2017, 2019) was used to determine orthologous groups amongst the six *F. xylarioides* strains and eleven related species (Table A.10). OrthoFinder also reconstructed a species tree from gene trees of the entire set of orthologs reconstructed by FastML (Figure 4). Concordance of the set of individual gene trees for single ortholog groups was summarised using PhyParts (https://bitbucket.org/blackrim/phyparts/src/master/), while concordance of the ortholog groups was summarised using venn diagrams drawn with the Venn package in RStudio. Correlations between strains were tested for an excess of sharing using the 'SuperExactTest' package in RStudio ((Wang et al. 2015).

2.4.4 Transposable elements and repeat annotation

Repeats and transposable elements (TEs) were identified directly from the assembled nucleotides. RepeatModeler v2.0 was run with parameters '-LTRStruct -pa 32' to first construct a custom repeat library for each genome (Flynn et al. 2020). This library was then used to call TEs and other repeats from the genome scaffolds using RepeatMasker v4.1.0 with the parameters "-pa 32 -lib $LIB -dir . -alignments -gff -no_is" (Smit et al. 2015). The distribution of TEs across the genome was represented as the fraction of

base pairs within 20 kb windows assigned to TEs. In the same way as for genes, the association between TEs and AT-rich areas was tested using a Kruskal-Wallis chi-squared and a Wilcoxon rank test (bonferroni adjustment). The separate distribution of class 1 RNA retrotransposons and class 2 DNA transposons (Daboussi & Capy 2003) was also calculated. For comparison, repeats and TEs were also identified from the raw reads using DNAPipeTE (Goubert et al. 2015) for the *F. xylarioides* strains, as well as *F. o. lycopersici* (SRA Accession SRR3142258).

The presence of repeat-induced point mutations (RIP) were confirmed using The RIP-per bioinformatics tool (Van Wyk et al. 2019), with default settings applied to identify large RIP affected regions, and the RIPCAL (Hane & Oliver 2008) RIP-index scan with default threshold settings applied and a scanning subsequence length of 10 kb for the core chromosome scaffolds and 300 bp for the FXU and LS scaffolds to reflect the average sequence length across the different scaffold groups.

2.4.5 Searching for putative effector genes involved in wilt disease

A four-pronged approach was used to search for putative effector genes involved in the host-specificity between the arabica and robusta populations. Only single gene copy orthologous groups were analysed for ease of comparison. I used BLAST to verify inferred patterns of presence and absence based on the orthogroups, and corrected in the twelve cases (annotated with an ”a” in figure 2.6) where annotation had missed a full length copy of the gene in certain taxa.

Pre-characterised effectors

Putative effector sequences described in closely related *Fusarium* species were downloaded from Genbank (Table A.3). Using BLAST, the sequences were searched for against the *F. xylarioides* predicted genes and scaffolds. Proteins were judged to be present if a blast match of >70% and 1.00E-50 was obtained and the sequence retrieved from the genome

scaffold, checking to include the full protein region if the blast match only returned a partial match.

Small and cysteine-rich putative effectors

Many fungal effector proteins are small (<400 amino acids), cysteine-rich (>4 cysteine residues) and secreted (Van Esse et al. 2008, de Jonge et al. 2011). Putative effectors were therefore searched for within each genome based on size, cysteine content, secretion signal and putative function. First, annotated genes were sorted according to size and number of cysteine residues using SeqKit (Shen et al. 2016), to select those with <400 amino acids and >4 cysteine residues. Next, the presence of a signal peptide and a signal peptide cleavage site on those proteins were predicted using TargetP 1.01 (Emanuelsson et al. 2000, Nielsen et al. 1997) with an RC score cut-off 1-3 to increase specificity. Finally, subcellular localizations for these proteins were predicted using WoLF PSORT (https://wolfpsort.hgc.jp/) to identify secreted extra-cellular proteins. The orthologous gene sets for those genes were identified from the OrthoFinder results.

CAZyme effectors

The CAZymes are carbohydrate-active enzymes thought to be important in the infection pathway for vascular wilt pathogens. Carbohydrates in plant cell walls provide the main source of carbon for fungal pathogens (Hervé et al. 2010) and consist of cellulose microfibrils embedded in a matrix of hemicelluloses, pectin polysaccharides and glycoproteins (Carpita & Gibeaut 1993). The breakdown of cell wall polymers requires a range of enzymes, including glycoside hydrolases, pectate and polysaccharide lyases and carbohydrate esterases (Mace et al. 1981). CAZymes that target cell walls also contain carbohydrate-binding modules (Boraston et al. 2004), which bind to cell wall polymers and increase the enzymes' catalytic efficiency by improving contact (Hervé et al. 2010, Carrard et al. 2000). Carbohydrate-binding modules show specificity for particular polymers (McCartney et al. 2006). Broadly, *Fusarium* is enriched for CAZymes (Table A.5).

Therefore, to identify putative CAZyme-encoding effector genes, and the carbohydrate-binding modules and their associated carbohydrate active modules differentially expressed across the wilt-inducing and non-wilt inducing fungal strains I used the CAZy database (www.cazy.org, Lombard et al. (2013)) to identify CAZyme-encoding orthologous groups. I looked for those present in only the wilt-inducing strains using BLAST (e-vale 1.00e-50), additionally allowing for those which were present in just one non-wilt inducing strain.

Effector genes with TEs in their promoter regions

In particular, class 2 MITES (miniature inverted-repeat transposable elements) have been shown to associate with effector proteins (Schmidt et al. 2013). The abundance of class II transposons recognised by RepeatModeler, miniature and full impala transposons, as well as recently described transposons across the genomes was tested using the accessions detailed in Table A.6 and by Bergemann et al. (2008), with a BLAST (score >70%, 1.00e-50, length >150bp) confirming their presence. The presence of an intact impala ORF was confirmed with the intact FOM24 transposase (Genbank accession AF282722.1) using BLAST, 1e-50.

Analyses for all putative effector classes

For each set of putative effectors, nucleotide sequences were aligned using MAFFT and maximum likelihood trees reconstructed using PhyML 3.0 (Guindon et al. 2009) with a Generalised Time-Reversible model with invariant sites and gamma distributed variation in substitution rates across 4 rate classes of sites, implemented in Geneious v9.1. I also located each putative effector on the genome scaffolds in each species mapped against the chromosomal level *F. verticillioides* assembly. The site test of positive selection in PAML v4.8 (Yang 2007) was used to test for positive selection among codons in the DNA sequences in each phylogenetic tree by estimating the ratio between synonymous and nonsynonymous substitutions (ω). PAML detects positive selection when it is significantly greater than 1 for a subset of codons. I compared a null model of “NearlyNeutral” selection

that includes a class of codons under purifying selection and a class that evolve neutrally, against an alternative model of "PositiveSelection" that additionally includes a class with >1. Log-likehood ratios were calculated and compared with a chi-squared test with one degree of freedom to confirm significance. A Bayes Empirical Bayes approach was used to identify specific amino acid sites under positive selection within gene sequences which were confirmed to be under positive selection (Yang et al. 2005). Horizontal gene transfer was detected by searching for our 64 putative effector genes across the genomes in Table A.10 and *F. oxysporum* ff. spp. in Figure A.7. BLAST hits were aligned using MAFFT and maximum likelihood trees reconstructed using approximate Likelihood Ratio Tests >0.95. The decision tree (Figure 2.7) was then followed to decide if each gene displayed evidence of horizontal gene transfer: is the effector present in other *F. fujikuroi* complex species (Y / N); if Y, does *F. xylarioides* nest with the *F. fujikuroi* gene copies (Y / N); if N to either of previous decisions, does *F. xylarioides* nest with *F. oxysporum* thus disrupting the Fo phylogeny (Y / N); branch support values (BSV) for *F. xylarioides* with *F. oxysporum* (n); if N, does *F. xylarioides* nest with *F. oxysporum* by distance i.e. less distance from *F. xylarioides* to *F. oxysporum* than greatest distance from *F. oxysporum* to *F. oxysporum*; pairwise id % for whole branch with support value for *F. xylarioides* nested with *F. oxysporum*.

2.5 Results

2.5.1 General features of the genomes in comparison with other *Fusarium* species

Genomes were assembled that ranged in size from 58 Mb in the coffea strains to 61 Mb for the robusta strains and 63 Mb for the arabica strains. The robusta genomes are larger than those previously sequenced (55 Mb), however they have a comparable size if scaffolds <500 bp are excluded (comprising 4Mb in total, Table A.1). These short scaffolds were retained for further analyses. These genomes are assumed to be complete,

based on synteny with previously published *F. xylarioides* genomes (Figure A.1), high levels of contiguity (N50 > 40 kb) and 100% predicted proteome completeness based on the presence of BUSCO genes (Table 2.1).

To further evaluate these genomes, they were compared to published genomes from a range of *Fusarium* taxa (Figures 2.2, A.2 and A.3, Table A.10). *Fusarium xylarioides* has a larger genome than its closely related species from the *Fusarium fujikuroi* species complex African clade (also known as the *Gibberella fujikuroi* complex) (Kvas et al. 2009), *F. udum* (56.4 Mb, Srivastava et al. (2018)), which causes wilt on pigeon pea, and that of *F. verticillioides* (41.7 Mb, Ma et al. (2010)), which is a non-wilt plant pathogen of maize. The genomes are similar in size, however, to the more distantly related *F. oxysporum* f. sp. *lycopersici (Fol)* strain 4287 (60 Mb, Ma et al. (2010)). Representative whole-genome alignments revealed the presence of the 11 syntenic core chromosomes shared by *F. verticillioides, F. oxysporum* and more distantly-related *Fusarium* taxa (Ma et al. 2010) in *F. xylarioides* (Figures A.2 and A.3), and the additional genomic material compared with *F. verticillioides* (Figure A.3). To understand precisely which *F. xylarioides* scaffolds

Table 2.1: Genome statistics for sequenced *Fusarium xylarioides* strains.

Name	Coffea674	Coffea659	Robusta277	Robusta254	Arabica563	Arabica908
Strain	IMI392674 CBS258.52	IMI127659i DSMZ62457	IMI392277	IMI392254	IMI389563	IMI375908i
Year	1951	1955	2003	1997	2002	1997
Origin	Cote d'Ivoire	Central African Republic	Tanzania	Uganda	Ethiopia	Ethiopia
Host	coffea	*C. excelsa*	*C. c.* robusta	*C. c.* robusta	*C. arabica*	*C. arabica*
Size Mb	57.2	59.6	61.3	60.3	63.4	62.6
No. scaffolds	3969	4843	5484	5369	7078	7264
N50 bp	62845	55667	54978	56291	42705	41762
BUSCO %	100	100	100	100	100	100
GC %	43.3	43.5	43.1	43.1	42.1	42.2
Coding Genes	14529	14732	14852	14588	14589	14654
Coverage %	36	35	34	34	33	33

matched these syntenic chromosomes, reference-guided scaffolding oriented contigs into chromosomes based on the chromosomal assembly of *F. verticillioides* (Ma et al. 2010). This mapped 85% of contigs for each genome to its syntenic chromosomes and un-aligned scaffolds (labelled 'FV') of *F. verticillioides*, with the remaining 15% comprising shorter un-aligned scaffolds (Figure 2.3). These un-aligned scaffolds were then classified based on their presence and absence across other *F. fujikuroi* complex species: those which are absent from *F. verticillioides* but which are present in *F. udum* and the historic coffea659 strain are labelled 'FXU' (*F. xylarioides* and -*udum* specific); those which are absent from *F. verticillioides* and *F. udum* and are shared with coffea659 are labelled 'FXS' (*F. xylarioides*-specific); and those which are not shared with coffea659 and are unique to each *F. xylarioides* strain are labelled 'LS' (lineage-specific). The FXS scaffolds make up <0.5% of the genomes, whilst LS scaffolds are 3.5% and FXU scaffolds are 12% (Table 2.1). Genome size differences in the arabica and robusta strains are largely explained by the FXU and LS scaffolds (11 Mb and 9 Mb respectively) (Table A.2).

Fusarium xylarioides and *F. udum* are the only known members of the *F. fujikuroi* complex to infect their target hosts with a wilt disease (Rutherford 2006, Geiser et al. 2005). To test whether *F. xylarioides* might have evolved new wilt capabilities by acquiring a pathogenic chromosome from *F. o. lycopersici*, the *F. xylarioides* genomes were mapped to the *F. o. lycopersici* genome assembly. Chromosomes 3, 6, 14 and 15 in the *F. o. lycopersici* genome are supernumerary, or mobile, as well as parts of chromosomes 1 and 2 (Ma et al. 2010). Chromosome 14 is also pathogenic, housing all known *F. o. lycopersici* wilt effector genes. No large-scale matches to any of the fully mobile chromosomes were found in *F. xylarioides* (Figures 2.2 and A.2), ruling out the recent transfer of the whole *F. o. lycopersici* mobile pathogenic chromosome. However, there was a significant excess of smaller matches between the LS and FXU scaffolds and the *F. o. lycopersici* mobile chromosomes (Figure A.4, Pearson's chi-sq test: $p<0.001$). Potentially, this is consistent with smaller pieces having been transferred: 7% of the FXU scaffolds and 20% of the LS scaffolds matched to the four fully mobile *F. o. lycopersici* chromosomes (3, 6, 14 and

15), whereas <2% matched to non-mobile chromosomes. With short-read sequencing it is not possible to infer the chromosomal assembly of the LS and FXU scaffolds nor rule out that *F. xylarioides* has novel mobile chromosomes not closely related to those in *F. oxysporum*. However, the pattern of strong hits to parts of the *F. o. lycopersici* mobile pathogenic chromosome but lack of hits to the majority of it does raise the possibility of transfer of smaller genome regions that are explored further below using alternative evidence.

Annotation of genes and transposable elements (TEs) in the MEGAHIT assemblies reveal the whole genomes, along with *F. udum*, contain on average 14,650 predicted protein-coding genes, covering 34% of the genomes and 5 Mb of repeats, covering 9% (Tables 2.1 and A.1). Within the reference-guided assemblies, the FXU and LS scaffolds

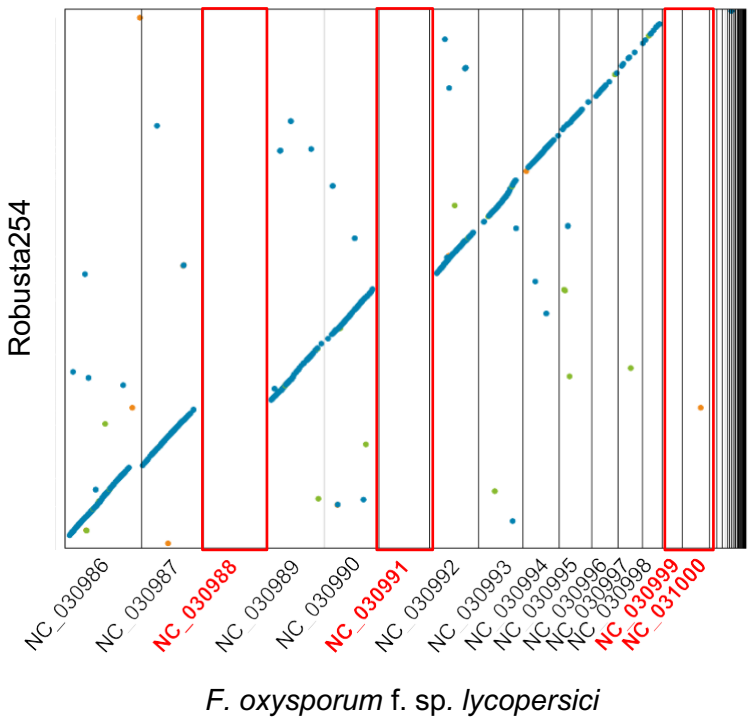


Figure 2.2: Representative whole-genome alignments of *Fusarium xylarioides* strain robusta277 against the 15 *Fusarium oxysporum* f. sp. *lycopersici* chromosomes. The 11 core chromosomes shared with sister *Fusarium* species (Ma et al. 2010) are present and the 4 mobile chromosomes are annotated in red, and . Each dot represents chromosomal correspondence, with absences representing the absent *Fusarium oxysporum* f. sp. *lycopersici* chromosomes. Alignments for each *Fusarium xylarioides* genome are in Figure A.2. Blue indicates forward alignments, green indicates reverse alignments, orange indicates repetitive alignments.

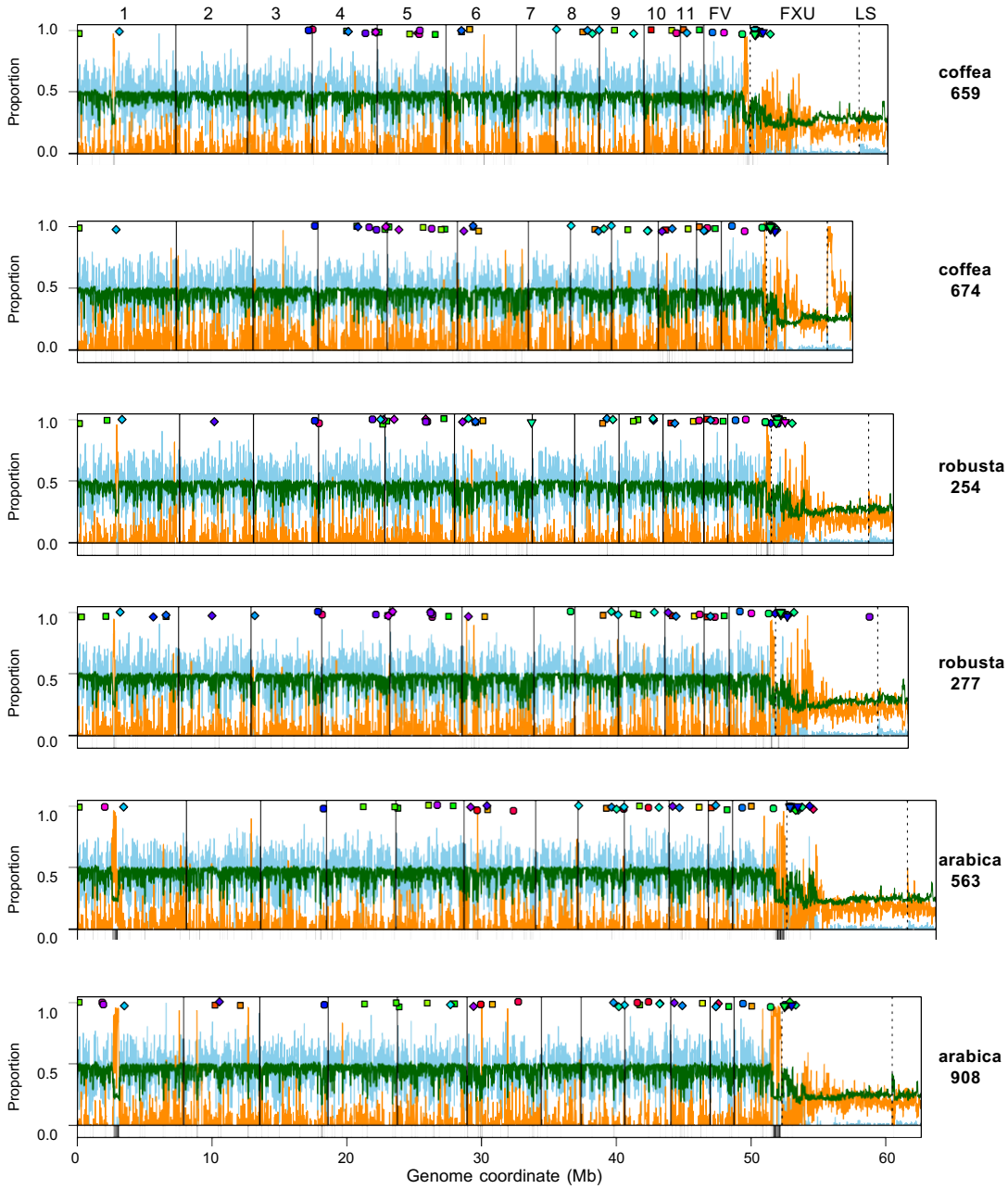


Figure 2.3: Global view of genome metrics plotted across 20 kilobase windows of each *Fusarium xylarioides* genome. Light blue = fraction of nucleotides annotated as a gene. Orange = fraction of nucleotides annotated as transposable element. Dark green = %GC. Vertical solid lines demark the 11 inferred chromosomes matching to the *Fusarium verticillioides* genome. Dashed lines demark: FV, scaffolds that are shared with *Fusarium verticillioides* but not assembled into chromosomes; FXU, scaffolds that are absent from *Fusarium verticillioides* but are shared with coffea659 and *Fusarium udum*; LS, scaffolds that are not shared with *Fusarium verticillioides*, coffea659 nor *Fusarium udum* and are therefore lineage specific. Coloured points indicate the location of putative effector genes: square = pre-defined in published literature, circle = carbohydrate-active enzyme, diamond = small and cysteine-rich proteins. Colours allocated across spectrum arbitrarily but same colour indicates belongs to same ortholog. The lower strip on each plot shows large RIP affected regions in black.

differed considerably in overall genomic content. The 11 core chromosomes along with the FV scaffolds were gene-rich and repeat-poor, whereas the FXU and LS scaffolds were gene-poor and repeat-rich (Figure 2.3). The distribution of genes attributed to particular functional classes was highly concordant amongst the six *F. xylarioides* strains, and within each core chromosome except chromosome 11, which is enriched for putative functions related to stress response and protection from reactive oxygen species (Figure A.2). Interestingly, genes on FXU and LS scaffolds are also enriched for the putative function 'phage major capsid proteins' (Figure A.5b), which could be involved with movements of mobile elements. In terms of TEs, consistent content across both the MEGAHIT and reference-guided genome assemblies was found of 5 Mb of interspersed repeats made up of simple repeats, retroelements and DNA transposons. Slightly less (4 Mb) were found if the raw reads were analysed, showing that my methods find more TEs than an assembly-free method (Table A.1). In contrast, the FXU and LS scaffolds were enriched for transposable elements, which overall comprise 24% of DNA in these scaffold groups, compared to 5% in the 11 core chromosomes. Within this, retrotransposons have contributed to genome expansion in *F. xylarioides*, as is commonly observed in other fungi (Schmidt et al. 2013), making up 16% of DNA in these scaffold groups, compared to 3% in the 11 core chromosomes. The LS scaffolds are further enriched for DNA transposons, with 150% more than is found in the FXU scaffolds and 360% more than in the core chromosomes. The proportion of the whole genome made up of transposable elements also varies among species. Some of these differences could reflect bias due to different assembly methods. For instance, *F. oxysporum* has a high repetitive sequence and TE content: I calculated 6.2 Mb TEs from the *F. o. lycopersici* long-read assembly (Table A.1), which is considerably more than the number calculated from the *F. o. lycopersici* raw reads (1.2 Mb, Table A.1). It is possible that future long-read sequencing would also reveal more TEs in *F. xylarioides*. Despite this potential bias, however, *F. xylarioides*, has nearly 1000% more TEs than recovered from the whole chromosome assembly of *F. verticillioides* (Table A.1).

The *F. xylarioides* genomes display another common structural feature of plant pathogenic ascomycetes, namely the presence of AT-rich genome blocks. Specifically, the Repeat-Induced Point-like (RIP) mutations pathway is a genome defence mechanism against the invasion of mobile transposons that has been linked to enhanced mutagenesis of effector genes in some plant pathogens (de Jonge et al. 2011). It is unique to ascomycete fungi and mutates cytosine bases to thymine in repeated genome sequences (Van Wyk et al. 2019). This leads to AT-rich sequences that are typically low in coding gene content, and can segment the genome into equilibrated compartments (as in *Leptosphaeria maculans*, Rouxel et al. (2011)). Across both MEGAHIT and reference-guided assemblies, significant evidence was found for AT-rich blocks (averaging 21.5% GC), which were spread throughout the core chromosomes and FV scaffolds, and the FXU and LS scaffolds (figure 2.3). On average, these blocks were significantly poor in genes but rich in repeats, in common with patterns in other fungi. The RIP regions made up >33.5% to >40% of each genome, with >7.7% to >8.5% of the genomes in so-called large RIP affected regions (Figure 2.3). Having described the broad features of the genomes, now I will address my questions concerning the multiple outbreaks.

2.5.2 The arabica population arose independently from the robusta and coffea strains

The genome data supports the previous hypothesis that the arabica and robusta populations emerged independently within *F. xylarioides* (Flood 2005). Gene annotations from the *F. xylarioides* and *F. udum* genomes together with ten other published *Fusarium* and *Verticillium* wilt genomes identified 18 569 orthologous gene sets, encompassing 25,0056 genes or 95.6% of all annotated genes. The species tree based on the concordance of 13,782 gene trees of all orthologue groups supports the established order of the *F. fujikuroi* species complex (Kvas et al. 2009), as well as the monophyly of *F. xylarioides*, with over 87% of genes supporting monophyly of the clade (Figure 2.4). No alternate topology was consistently found for the remaining gene trees (the next most common

was supported by just 1.1% of genes), confirming that *F. xylarioides* did not originate by hybridisation or a major influx of genes from other taxa. Within the *F. xylarioides* clade, arabica strains were recovered as a sister clade to robusta and coffea strains with high concordance of gene trees, consistent with the arabica strains constituting genetically isolated taxa. Whilst the two robusta strains are also monophyletic with high concordance, there is little concordance for their branching order with respect to coffea strains: loci vary in whether robusta strains are closer to coffea659 or to coffea674. Whilst analysis of more strains is needed to confirm this conclusion, it fits the hypothesis that the robusta population emerged from within a wider recombining population of the more genetically diverse coffea strains, whereas the arabica population is a more divergent lineage within the *F. xylarioides* clade.

This conclusion is further supported by patterns of presence and absence of genes. Over 96% of single gene copies were shared between the two arabica and between the two robusta strains respectively, while there is less similarity between other *F. xylarioides* comparisons: coffea-robusta, 93%; coffea-coffea, 92%; coffea-arabica, 92%; arabica-robusta, 91%; and *F. xylarioides-F. udum*: 84% (Figure 2.5A). The robusta strains share a significantly higher proportion of orthologous gene sets with coffea659 (SuperExactTest, $p<0.001$), and they generally display more concordance with the coffea strains, while the arabica strains have the most unique orthogroups. The arabica strains share slightly more genes with the ex-type strain coffea674 than with coffea659 despite the latter being the only known strain able to infect both arabica and robusta coffee (Girma et al. 2009, Flood 2005).

2.5.3 A set of putative effector proteins for *Fusarium xylarioides*

The proteome of plant pathogens includes pathogenicity factors and effector proteins, which are secreted to enable pathogen entry, survival and effective colonisation in their host. Thus fungal effector proteins are typically under positive selection and do not

show conserved protein sequences (Jones et al. 2018, de Jonge et al. 2011). Therefore, a multi-dimensional approach was adopted to look for putative effector proteins, namely those which enable the pathogen to become established in its host and trigger the onset of symptoms and a host defence response. Four classes are described: *(i)* those known from other fungal pathogens, *(ii)* small and cysteine-rich proteins, *(iii)* carbohydrate-active enzymes, and *(iv)* genes with class II TEs in their promoter regions, specifically including the presence of *miniature impala (mimp)* transposons which are associated with *F. oxypsorum* effector genes (Schmidt et al. 2013). The following characteristics were also annotated: presence of positive selection indicating genes which are under accelerated

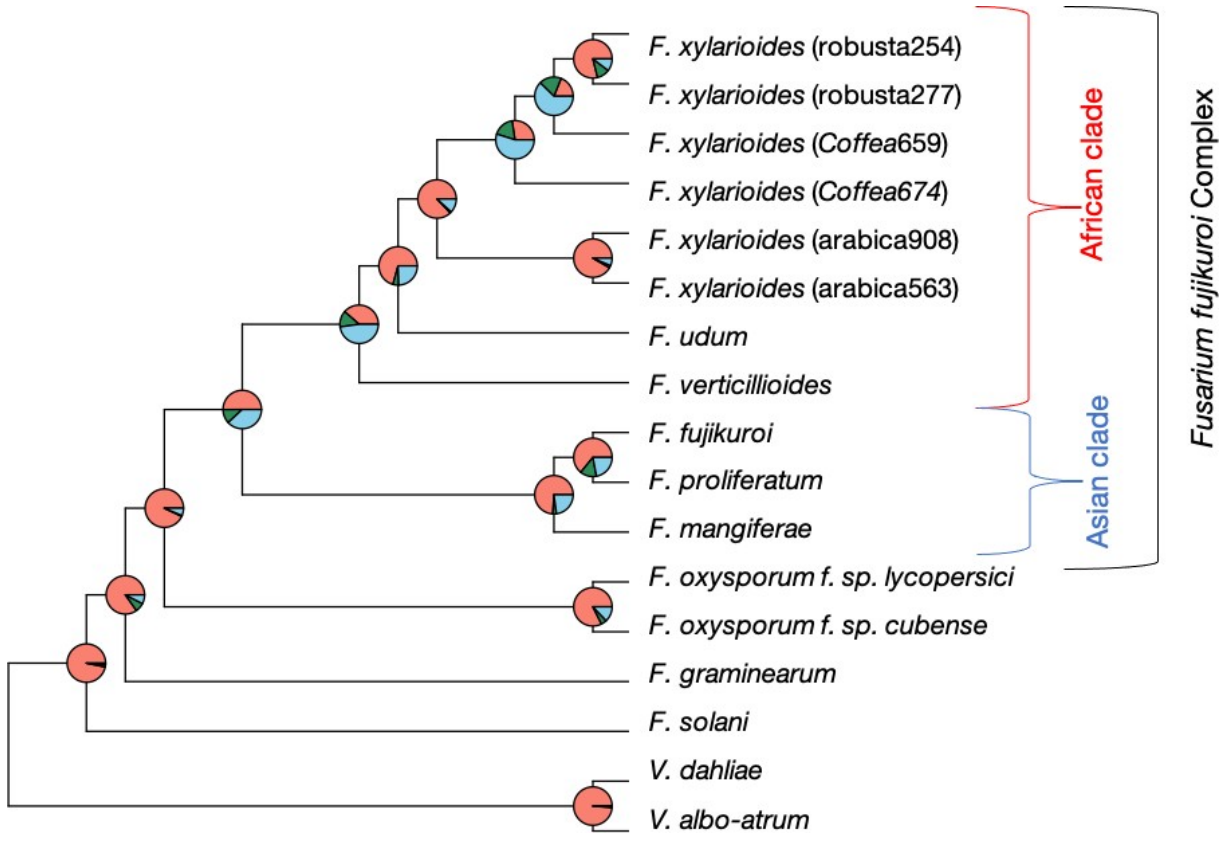


Figure 2.4: Phylogenetic relationships between *Fusarium* species reconstructed from 13,782 orthogroups support monophyly of the *Fusarium xylarioides* clade, with little consistent support for alternate topologies. Strain accession numbers are in Table A.10. The brackets denote the *Fusarium fujikuroi* species complex, with the Asian (blue) and African (red) clades. Pie chart colours: pink = proportion of genes (orthogroups) recovering the depicted node; dark green = the proportion of genes recovering the second most common topology; light blue = the proportion of genes recovering all other topologies.

evolution; presence of a signal peptide for secretion, because it is assumed that effector proteins need to be secreted during host colonisation; genes which are present in an AT-rich genome block as an indicator of potential transposon invasion; the predicted function based on the Pfam domain; and genes which are absent from sister *F. fujikuroi* complex species and share *F. oxysporum* as the closest match with a percent identity above 90%. Altogether, this process identified 64 putative effector genes, which were distributed across scaffolds and the reference-guided scaffold assemblies, and were mostly recovered in the same location in each strain (Figure 2.3). Because effector genes have not been investigated previously in *F. xylarioides*, general findings for each category of effector genes are described first, before focusing on differences between the different host-specific populations.

Putative effector genes known from other species

High BLAST matches in *F. xylarioides* were found to 19 known effector proteins previously characterised in *F. oxysporum* and other wilt-inducing pathogenic fungi, including some that are only recently discovered and unnamed (Table A.4, Figure 2.6). The effector genes *fow1, fmk1, snf1, pelD, chlo-vacu, rho1* (here with two copies *rho1.1* and *rho1.2*), *pep1, pelA, orx1, FOXG 14254* and *catalase-peroxidase* as well as the transcription factor Sge1 were found in all six *F. xylarioides* strains, with *nep1, glucosyltransferase, pda1, six7, six10* and *cytoskeletal* found across the host-specific pairs. The majority of the well-characterised *SIX* effectors from *F. o. lycopersici* chromosome 14 are absent in both *F. udum* and all *F. xylarioides* strains, with the exceptions of the recently described Six10 and Six7 proteins (Schmidt et al. 2013) which are present in arabica and coffea strains, and the transcription factor Sge1, which is required for expression of *SIX* genes (Michielse & Rep 2009), is found in all strains.

Small and cysteine-rich secreted proteins

Many fungal effector proteins are small (<400 amino acids), cysteine-rich (>4 cysteine residues) and secreted (Van Esse et al. 2008, de Jonge et al. 2011). Searching for these properties in the predicted proteins discovered over 2,500 small, cysteine-rich proteins in

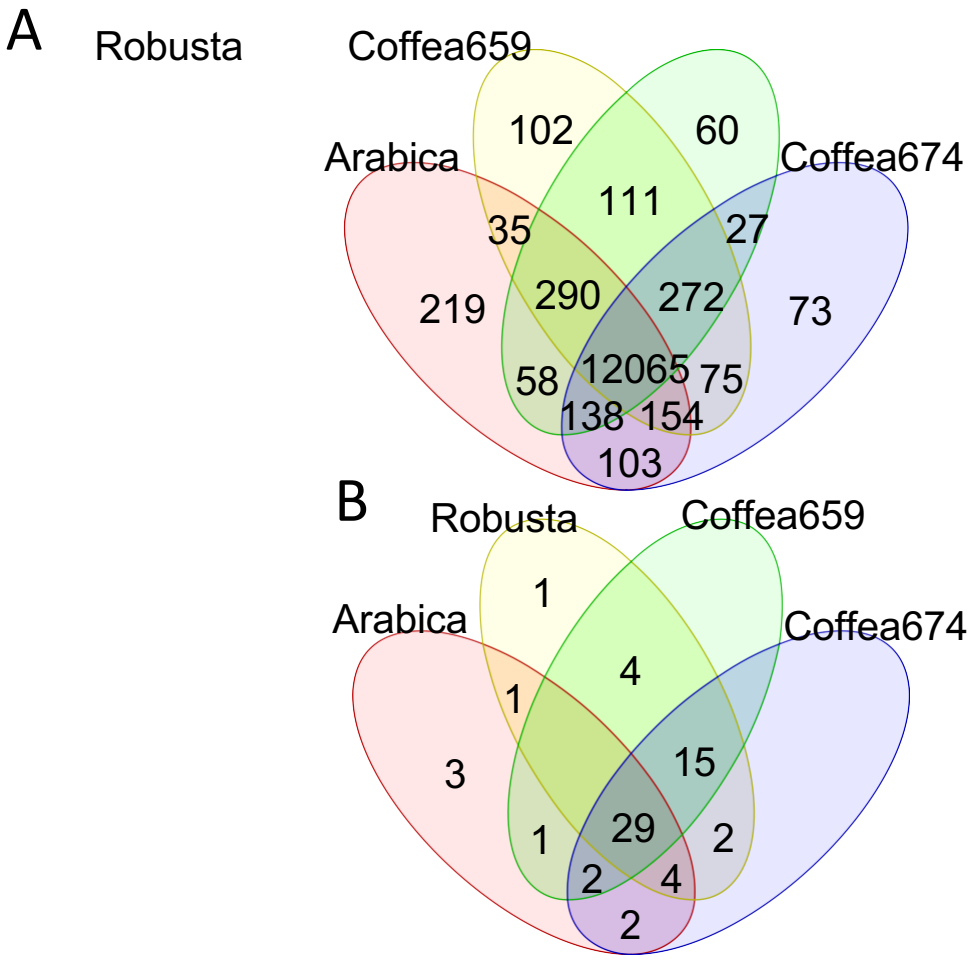


Figure 2.5: Shared features between the *Fusarium xylarioides* strain groups. A Orthogroups shared between *Fusarium xylarioides* arabica, robusta and the two coffea strains. Drawn using 13,782 orthogroups (excluding 449 *Fusarium xylarioides* robusta orthogroups and 175 *Fusarium xylarioides* arabica orthogroups that differed in presence/absence between the the two strains of each population. The two coffea strains are more divergent (with 1123 orthogroups that differ in presence/absence between the two strains) and so are drawn separately. The two arabica strains (563 and 908) share 13 062 genes and the two robusta strains (254 and 277) share 13,104 genes in total. **B** Putative effectors shared between *Fusarium xylarioides* arabica, robusta and the two coffea strains. Drawn using 64 putative effector proteins that differed in presence/ absence across the host-specific populations.

each strain, of which over 500 also have a signal peptide and a signal peptide cleavage site. Focusing on single copy genes for ease of comparison left 132 putative effectors in 21 orthologous groups shared across *Fusarium* and *Verticillium* (Figure 2.6). Patterns of presence and absence were variable among these genes: ten were recovered in all *F. xylarioides* strains. Many of these genes had no recognisable Pfam domain function identified by InterPro (Figure 2.6). This is expected because there are large numbers of unknown small and cysteine-rich proteins secreted in the apoplast (as part of the water-transport pathway) which are typically species- or strain-specific with no known function (de Jonge et al. 2011). However, the orthogroup OG0013477, found in all *F. xylarioides* strains, matches the LysM-containing chitin oligosaccharide protein, which in *Cladosporum fulvum* is required for scavenging of chitin oligosaccharides as a conserved fungal defence to enable undetected invasion (de Jonge et al. 2010). Orthogroup OG0014398 displayed a match, albeit with a relatively low BLAST score, to a number of *six7 F. oxysporum* genes and protein sequences (e.g. MG647014.1) (notably not the *F. o. lycopersici* 4287 sequenced genome used for this comparative analysis), which might indicate that it is a divergent type of Six7 effector compared to previously described copies. Other functional annotations reveal a recurring theme of pectin degradation (expanded in next section): OG0013871 is a hydrolase; OG0015372 is a pectin lyase and OG0014836 a carbohydrate esterase 5 with a carbohydrate binding module (CBM1); while OG0014864 and OG0016323 are major facilitator proteins involved in the transport of solutes, which could influence wilting.

Carbohydrate-active enzymes

Carbohydrates such as cellulose and pectin in plant cell walls provide the main source of carbon for fungal pathogens (Hervé et al. 2010) and so carbohydrate-active enzymes (CAZymes) are thought to be important in the infection pathway. I therefore searched for CAZymes restricted to some or all of the wilt-inducing species of *Fusarium* and *Verticillium* and classified their carbohydrate-binding modules to characterise specificity for

particular polymers. Our search found 56 genes of interest belonging to 16 orthogroups (Figure 2.6). As a whole, *Fusarium* is broadly enriched for lytic enzymes involved in carbohydrate metabolism along with other ascomycete fungi (Reis et al. 2005, Soanes et al. 2008) (Table A.5), while the vascular wilt inducers show enrichment in certain gene families (Figure A.10), also reported by Klosterman et al. (2011)). Fungal growth on pectin requires either the combined activities of pectin esterases and polygalacturonases or the sole activity of pectate lyases (Howell et al. 1976, Mace et al. 1981), and all three enzymes are known from other wilt-inducing fungi (Beckman 1956, Langcake & Drysdale 1975, Talboys & Busch 1970). Interestingly, each share orthologous groups across the wilt-inducing strains analysed in this study (Figure A.10). A polygalacturonase enzyme (glycoside hydrolase (GH) family 28) from *F. oxysporum*, has expanded by one copy number in all *F. xylarioides* strains. Pectate lyase carbohydrate-binding modules (CBM) CBM66 and CBM38, which bind a fructose-hydrolysing GH32, and chitin-cleaving CBM18 and CBM50 are also found across all six *F. xylarioides* strains (as well as *F. udum* and *F. o. lycopersici*). In contrast pectin esterases have expanded differentially among different *F. xylarioides* strains (see next section).

Putative effectors with class II transposable elements in their promoter regions

Effector genes in *F. oxysporum* are associated with class II TEs (Schmidt et al. 2013). There is a certain type of class II TE known as a *mimp* for *miniature impala* (part of the miniature inverted-repeat transposable element (MITE) family) which was first described in *F. oxysporum* (Bergemann et al. 2008). In *F. o. lycopersici*, Schmidt et al. (2013) reported that *mimps* (along with several other newly described types of class II TEs) are mostly found on the accessory chromosomes and within the promoter regions of *SIX* effector genes (Schmidt et al. 2013). Furthermore, *mimps* have recently been found in a few *F. fujikuroi* complex species (van Dam & Rep 2017) and phylogenetic evidence supported repeated transfers between *F. oxysporum* and *F. fujikuroi* complex species. Consequently, a final class of putative effectors was looked for that met four criteria: they

were present in both strains for each host-specific population; they contained *mimps* or the newly described transposon families from Schmidt et al. (2013) in their promoter regions (defined as 1500 bp upstream from the start codon); they were on FXU or LS scaffolds; and they matched regions of *F. o. lycopersici* 's mobile chromosomes. This resulted in eight additional putative effectors to supplement those identified by our three original criteria, with three containing mimps, two containing *hop3*, two containing *yaret2* and one with *yaret1* DNA transposons (Figure 2.6). Additionally, 15 putative effectors from the predefined, small and cysteine-rich and CAZyme classes also shared a scaffold with a transposon, with four genes across the robusta and coffea strains (*catalase peroxidase, OG0014741, OG0014743, OG0014398*) and *six10* across the arabicas with *mimps* in their promoter regions. The potential role of transposition on effector gene evolution is investigated further in section 2.5.6 – for now they are treated simply as another class of putative effector genes.

2.5.4 Host-specific populations differ in their complement of putative effectors

Effector gene presence and absence recapitulated findings from the whole genome analyses: both strains from the same host-specific population nearly always share the same effector gene complement, and the robusta population and coffea strains share greater overlap in effector gene complement (SuperExactTest, p=0.001) than either do with the arabica population (figure 2.5b). Based on this significant association, I looked for changes that might have led to specialisation on robusta coffee and increased virulence, and found just 1 effector unique to the robusta strains, *pda1*. This gene encodes the enzyme pisatin demethylase, which has been shown to detoxify the plant defensive compound (phytoalexin) pisatin by *F. oxysporum* f. sp. *pisi* in peas, and with different alleles present across different *F. oxysporum* host-specific ff. spp. (Covey et al. 2014, Ellis et al. 2016, Coleman et al. 2011). Interestingly, in sugarbeet one *pda1-a* allele (accession AY487143.1) was only found in pathogenic strains (Covey et al. 2014), although the gene copy found in this study is

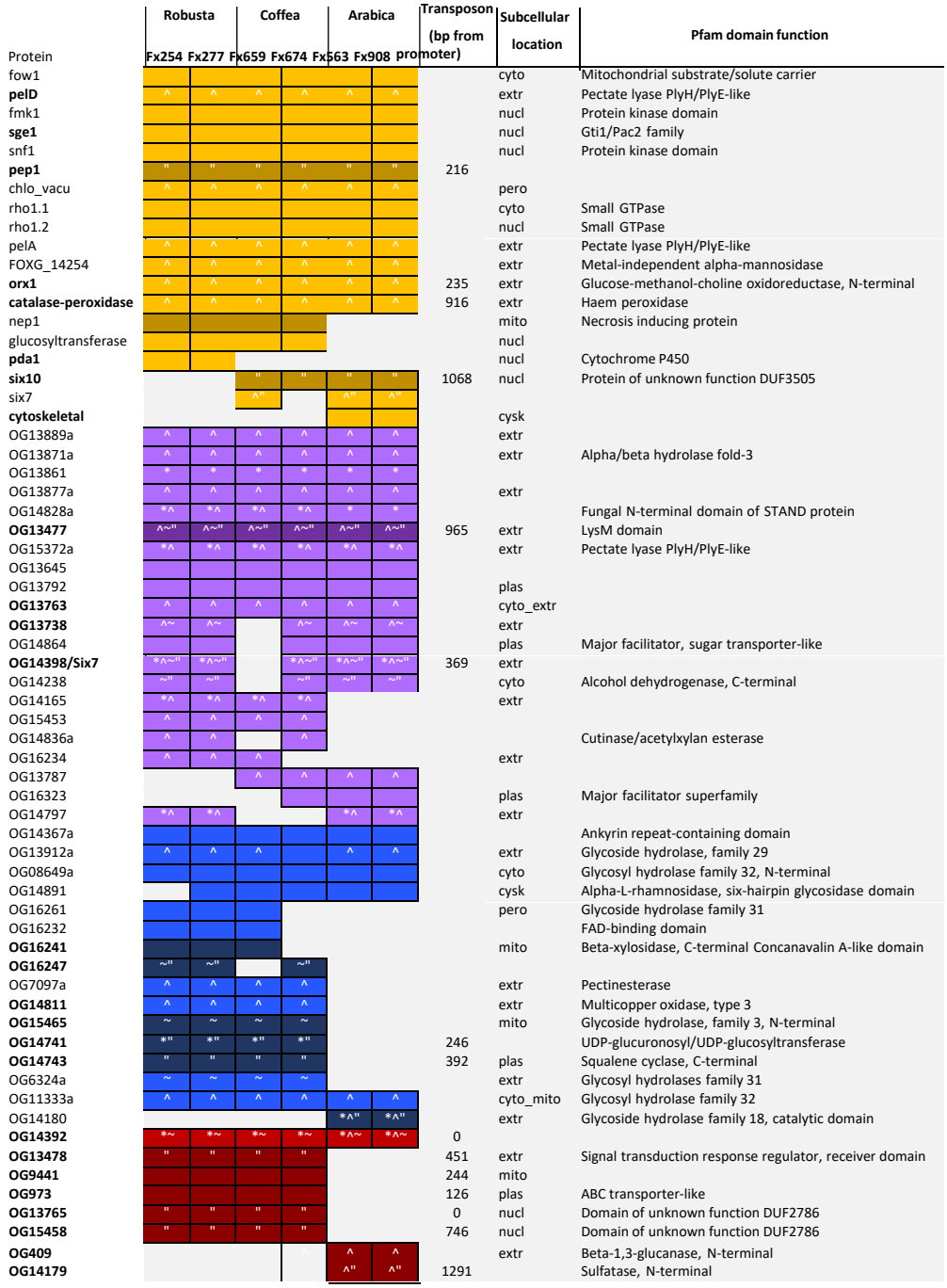


Figure 2.6: Putative effectors' characteristics and presence or absence across *Fusarium xylarioides* strains and *Fusarium udum*. The four effector classes are shown in: yellow for predefined effectors; purple for small and cysteine-rich effectors; blue for carbohydrate-active enzymes; and red for transposon-adjacent effectors. The presence of transposons is represented by names in bold with its distance from the genes promoter described if less than 1500bp (if not, the transposon is over 1500bp away), genes under positive selection by an asterisk, secreted proteins with a signal peptide by a caret, genes in an AT-rich block by a tilde, genes with evidence of horizontal transfer from *Fusarium oxysporum* are a darker shade and function is represented by the Pfam domain, where a hit was returned. Genes which are absent from the FFC with *Fusarium oxysporum* the closest match with a percent identity (%) >=90 are represented by a quotation mark.

97% similar to a *pda1* from *F. oxysporum f. sp. phaseoli* and is only 62% similar to the pathogenic sugarbeet allele. The enzyme could have a direct role against related phytoalexin compounds in coffee or enable growth of robusta strains on alternative hosts. However, given that both coffea674 and coffea659 lack *pda1* it is unlikely this gene is essential for robusta infection. Indeed, with the exception of *pda1*, the robusta population is more unique in the genes that it lacks than in the genes it has gained when compared with the coffea strains. Because plant immune systems are triggered by detection of secreted effector proteins, the loss of genes might also be important for pathogenicity on different hosts, should plants evolve to recognise certain effectors. Four orthogroups that are present in one or both coffea strains are missing from the robusta strains: *six7, six10, OG0013787* and *OG0016323*, so absence of these genes could be further explored as a possible cause of enhanced pathogenicity on robusta.

In contrast, the arabica strains are divergent from both the coffea and the robusta strains in their complement of effector genes. Arabica strains share three unique effectors: *cytoskeletal*, the CAZyme *OG0014180* and *OG0014179* with a TE in its promoter region. The wilt-specific Cytoskeletal protein was recently described in *F. o. lycopersici, V. dahliae* and *V. albo-atrum* and found to be absent from non-wilt inducing *Fusarium* species (Klosterman et al. 2011). The CAZyme *OG0014180* encodes for a GH18 chitinase that has an insertion under positive selection relative to the matching copy in the *F. o. lycopersici* genome, potentially indicating divergent function within the arabica strains. In addition, the arabica strains lack 17 effector orthologs found in one or both coffea strains: *gluco, nep1*, five small cysteine-rich proteins including a hydrolase and a pectate lyase, and ten CAZymes including five glycoside hydrolases, another glucosyltransferase *OG0014741*, and a CE8 pectinesterase *OG007097*. Interestingly, two of these hydrolases, *OG007097* and *OG006324* have apparently inactivated with 20-30% shorter gene copies which were annotated as two separate genes in both arabica strains: in *OG007097* the arabica strains are missing a 500 bp piece of the gene compared with the robusta strains and a stop codon has truncated it into two genes; and in *OG006324* arabica563 has two

copia long terminal repeat (LTR) TE insertions and arabica908 has three unknown TE insertions in the middle of this gene, splitting it and presumably removing functionality. Interestingly, the arabica strains have 20% more copies of these repeats than the robustas, and 35% more than the coffea strains. The two glucosyltransferases (predefined *gluco* and *OG0014741*) are found across the robusta and coffea strains, *F. oxysporum* ff. spp. including *F. oxysporum* f. sp. *cubense* and *lycopersici*, *V. dahliae* and *V. albo-atrum*. The predefined *gluco* glucosyltransferase is required for full pathogenicity in *V. dahliae* (Klosterman et al. 2011), and in *F. xylarioides* it shares a scaffold with the small cysteine-rich putative effector *OG0014165*, which is also present only in robusta and coffea strains (see below). The Nep1 protein induces necrosis and ethylene production in host plants (Pemberton & Salmond 2004) with its family expanded in *F. o. lycopersici* and *V. dahliae* and is allegedly responsible for their wide range of hosts (Yadeta & Thomma 2013). It should be noted that putative CAZyme effectors are not shared exclusively between arabica and coffea strains, whereas robusta strains share seven with the coffea strains.

While these differences confirm separation of the arabica population from the other strains, a few putative effectors displayed contrasting affinities. The four genes highlighted as absent in robusta strains are cases of sharing between coffea and arabica strains. Of these, *six7* is shared with coffea659, the only coffea strain that is also able to infect arabica coffee (Girma et al. 2009, Flood 2005), and therefore of possible interest as pathogenicity factors for growth on arabica coffee. Just one gene, *OG0014797*, is shared by robusta and arabica populations but absent in both coffea strains. Interestingly, this gene gave a significant signal for positive selection and displays considerable amino acid divergence between the two forms (91% Pairwise Identity). However, both copies share *F. anthophilum*, a member of the American *F. fujikuroi* species clade (Kvas et al. 2009), as closest relative (with nearly 80% similarity), while a diverged copy in *F. verticillioides* is closer to *F. nygamai* (84% similarity), another African *F. fujikuroi* species clade member (Kvas et al. 2009). It is possible therefore that this gene diverged rapidly under positive selection for

differential function in each population or that these represent separate acquisitions from alternative sources, despite the same closest BLAST match for *F. xylarioides*. Either way, there is no evidence for substantial recent exchange of effector genes between arabica and the other populations, consistent with genome-wide evidence for their isolation.

I also looked for changes in the number of copies of CAZyme gene families shared by orthologous groups among *F. xylarioides* strains as a possible cause of changes in the ability to degrade plant cell walls and access carbon from their host. Excluding the groups that matched non-wilt-inducing species, differences were found in vascular wilt-inducing plant cell wall-degrading gene families between robusta and arabica populations, with the coffea strains, as before, more similar to robusta. Plant cell walls contain pectins, celluloses, hemicelluloses and other polysaccharides which, along with simple sugars such as mannans, wilt fungi can utilise as a carbon source (Lombard et al. 2013, Mace et al. 1981). When comparing robusta and coffea with arabica, different CAZyme gene families that share functions have expanded across them: the pectinase families CE8, GH78 and GH88 in robusta and coffea, and PL1 in arabica; the xylanase, xylan being a component of hemicellulose, families CE5, GH3, GH43_11, GH43_24 and CBM35 in robusta and coffea, versus CBM42 and GH43_26 in arabica; and the mannanase families CBM13 and GH134 in robusta and coffea, and GT22 in arabica. Whilst not specific to pectin, CBM1-containing carbohydrate esterases that bind cellulose and hydrolyse xylans, mannans and pectins (Kellett et al. 1990, Hogg et al. 2003, McKie et al. 2001) were only present in robusta strains and coffea674. Although it is not possible to infer the functional effects of these changes on the infection process, they suggest differences in carbon usage between the two host-specific populations.

While the main differences among strains are apparent in gene content, cases of divergent selection were also found to act on amino acid sequence of some shared genes: for example, OG0013861 and OG0014828 found in all strains, and OG0014398 found in all strains except coffea659 (Figure 2.6). The latter ortholog is the potentially divergent form of *six7* identified as a small cysteine-rich protein and displays amino acid divergence

(92% Pairwise Identity) in a pattern consistent with strain host-specificity.

2.5.5 Several effector genes have been acquired horizontally by transposable elements

Previous work has shown the importance of horizontal gene transfer in the spread of new fungal plant pathogens, so the panel of putative effectors were screened for evidence of acquisition by horizontal transfer. The hypothesis that effectors associated with wilt formation might have been acquired from *F. oxysporum* was considered. Using BLAST, each effector was tested against a panel of *Fusarium* genomes encapsulating the *F. fujikuroi* complex (the closest relatives of *F. xylarioides*), a wide variety of *F. oxysporum* ff. spp. (as possible sources of horizontally acquired DNA) and distantly related *Fusaria*, and recovered hits with 1.00E>-20 and length >90%. Gene trees were then reconstructed using alignments of matching genes retrieved from each genome. Vertically inherited genes in *F. xylarioides* were expected to group with *F. fujikuroi* complex lineages, as in the core phylogeny (Figure 2.4), whereas horizontally acquired genes were expected to group with *F. oxysporum*.

Across 64 effectors (Figure 2.7), seven display no BLAST hit and their origin is considered to be unknown. All others have positive hits in at least one *F. oxysporum*. Most of these (45) also display at least one hit in an *F. fujikuroi* species complex genome, and indeed in 40 of these effectors the *F. xylarioides* sequences group with *F. fujikuroi* complex sequences, which is consistent with vertical inheritance. Note that nine of these cases had an *F. fujikuroi* complex match only in *F. udum*, the wilt-forming sister lineage to *F. xylarioides* (Table A.8, cases annotated with *). Potentially these cases could represent horizontal acquisition from *F. oxysporum* in the common ancestor of *F. udum* and *F. xylarioides*, but in this instance I focus on putative transfer into the *F. xylarioides* lineage and so these are labelled as vertical. The remaining 5 effectors, out of 45 with positive hits in the *F. fujikuroi* complex, group more closely to *F. oxysporum* lineages with strong support (approximate likelihood ratio test, aLRT support >0.95, Table A.8),

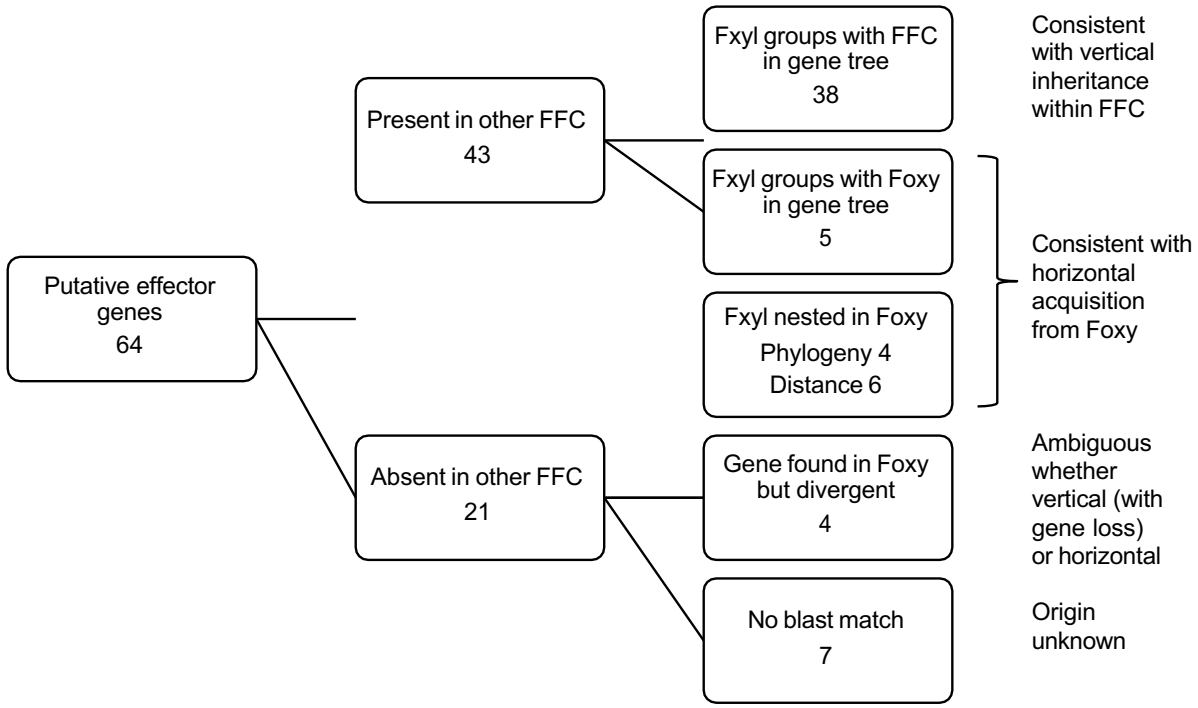


Figure 2.7: Decision tree showing the numbers of putative effector genes in *Fusarium xylarioides (Fxyl)* displaying different patterns in their phylogenetic relationships, focusing on possible horizontal acquisition from *Fusarium oxysporum (Foxy)*. FFC = *Fusarium fujikuroi* complex that *Fusarium xylarioides* belongs to. Interpretations consistent with each pattern are shown on the right.

consistent with horizontal transfer. Among the remaining 14 effectors (of the original 64) with no hits detected in *F. fujikuroi* complex genomes, *F. xylarioides* copies nest within the variation observed among *F. oxysporum lineages* in 10 cases (based on phylogenetic and distance criteria, see methods), which is again consistent with horizontal transfer (e.g. Figure 2.8b). In the final 4 cases, *F. xylarioides* sequences are highly divergent from all *F. oxysporum* sequences and so a hypothesis of recent horizontal acquisition is not strongly supported. Those genes could have been vertically inherited but secondarily lost in all other *F. fujikuroi* complex lineages, or diverged so much that they returned no match in other genomes, or been acquired horizontally from lineages not included in this sample.

Among the 15 putative effectors with evidence for acquisition by transposition, two are shared by all six *F. xylarioides* strains. Of these, one is the putative divergent form of Six7 (OG0014398) that displays evidence of divergent selection between host-

specific strains and the other (OG0013477) matches a LysM domain. Three effectors are in the arabica strains, with one, another SIX protein (Six10) shared by arabica and the coffeas, and two effectors unique to the arabica population and both involved in hydrolysis. The remaining ten effectors are shared by the robusta and at least one of the coffea strains but not arabica. Six of these have recognised functional annotations: Nep1, involved in necrosis; OG0016241 and OG0015465, hydrolysis of sugars; OG0014741,glucosyltransferase; OG0014743, squalene-hopene-cylase; OG0013478, involved in signalling; OG0000973, transporter protein. Thus HGT from *F. oxysporum* appears to have played a greater role to the emergence of the robusta and coffea host-specific strains than in the arabica population. Of the *formae speciales* which consistently match, *F. oxysporum* f. sp. *raphani* shares a percent identity >98% with four of these HGT effectors only found in the robusta and coffea strains (Figure A.7).

Possible causes were further interrogated for 15 putative cases of horizontal transfer from *F. oxysporum* by investigating the flanking regions in the MEGAHIT assemblies. Of particular interest is one scaffold found in the robusta and coffea genomes, which contains four putative effectors matching *F. oxysporum* f. sp. *raphani* with evidence for horizontal transfer (OG0009441, OG0013478, OG0014741 and OG0014743) in each strain (Figure 2.8a, 2.8b, 2.8c, 2.8d). Additionally, this scaffold is absent from *F. verticillioides* and designated as FXU in section 2.5.1, for each strain. This 20 kb scaffold displays high (>97%) identity with *F. oxysporum* (Figure 2.8e), and further features suggestive of horizontal transfer described in section 2.5.6. Among the remaining effectors suggestive of HGT, six were found on scaffolds with high identity to *F. oxysporum* which were mostly absent from *F. verticillioides* too, consistent with transfer of a larger region of DNA, whereas five were assembled into a background of low identity, consistent with transfer of a shorter piece of DNA (Figures A.8 and A.9).

2.5.6 Evidence that class II TEs play a role in horizontal transfer of putative effectors

Previous work implicated *mimp* class II TEs in the horizontal gene transfer of effectors in *F. oxysporum*: *mimp*s are strongly associated with effector genes (Schmidt et al. 2013) and phylogenetic analysis supports multiple transfer events between *F. oxysporum* and other *F. fujikuroi* complex species (van Dam & Rep 2017). Therefore the potential role of *mimps* (Table A.6) in horizontal acquisition of effector genes was investigated in *F. xylarioides*. On average, 46 *mimps* were identified per genome, with 59 in the robusta and coffea strains and 20 in the arabicas, which is at the upper end of estimates in other *F. fujikuroi* complex species (van Dam & Rep 2017) (<20 in each strain except for *F. hostae* which has 40), despite a potential bias to under-estimate numbers of repetitive TEs in assemblies of short-read data (section 2.5.1). Most are from the common *mimp* 1 and 2 families, with 10% in family 4 in the robusta and coffea strains and 10% in the highly divergent and recently described *mimp* family *mn14* in the arabicas (Figure 2.9). One full, and therefore presumably active, copy of the *F. oxysporum* f. sp. *melonis* FOM24 *impala* transposase (BLAST, e=0.00) was found in each strain. To test whether *F. oxysporum* could be the source of these *mimps*, nucleotide sequences were compared. Using BLAST, *mimps* were compared to the nr database as well as the genomes in Table A.10, and only hits to *F. oxysporum* were found: *mimps* in families 1, 4 and the highly divergent family mn4 had >96% percent identity (>97% in 1 and mn4) with *F. oxysporum* with high support values (Figure 2.9) and the *F. xylarioides* and *F. oxysporum mimps* as sister clades.

Of the 15 cases of putative horizontal transfer discovered by phylogenetic analysis of effector gene sequences in the previous section, all are found on scaffolds containing *mimps* and DNA transposons. For example, the 20 kb scaffold from Figure 2.8e, which is present in the robusta and coffea genomes, has close affinity to *F. oxysporum* and is enriched and interspersed with class II DNA transposons and *mimps*. This region matches a chromosomal mini-cluster on *F. o. lycopersici*'s chromosome 14 that contains

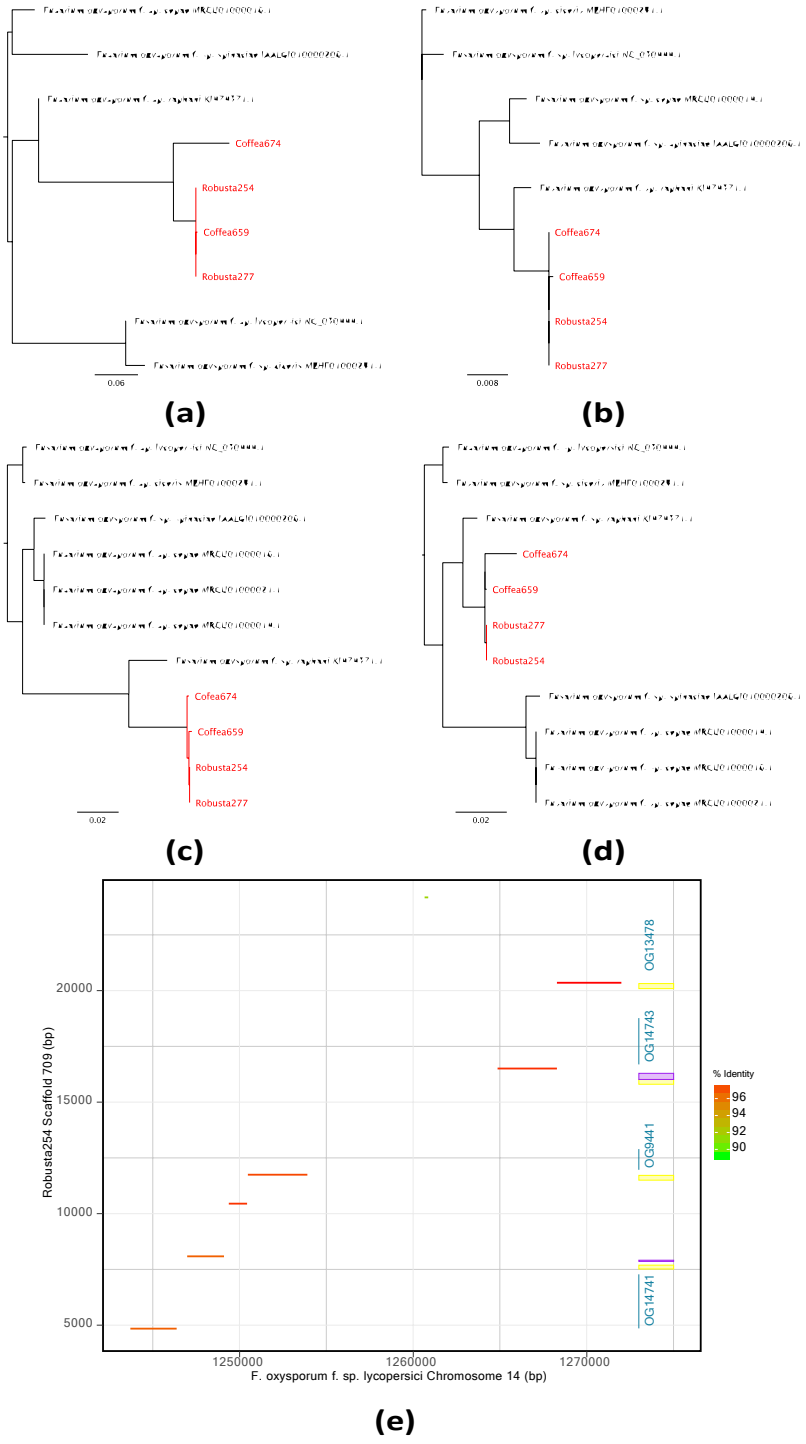
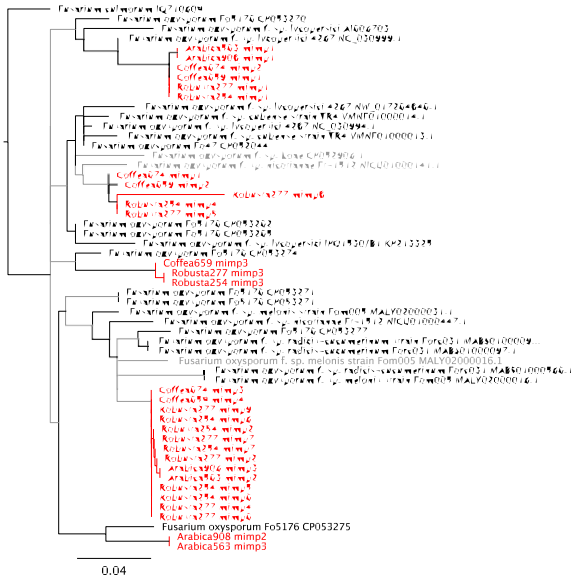


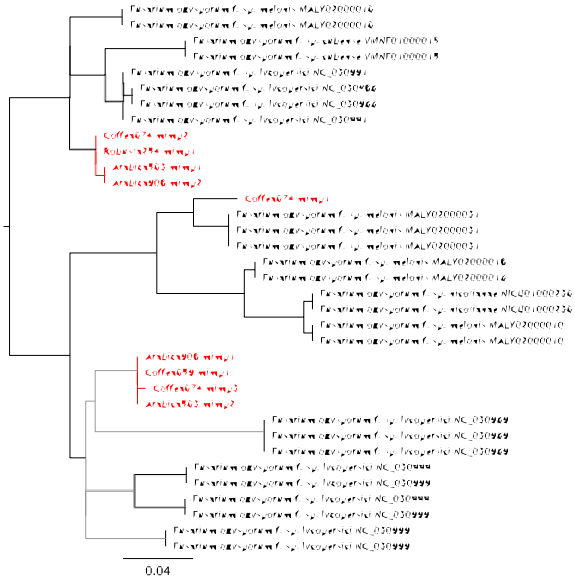
Figure 2.8: Four putative effector genes on an *Fusarium xylarioides* robusta254 scaffold and their phylogenetic trees. A Phylogenetic tree for *OG0014741*, **B** Phylogenetic tree for *OG009441*, **C** Phylogenetic tree for *OG0014743*, **D** Phylogenetic tree for *OG0013478*. For each gene, the *Fusarium fujikuroi* complex is absent, *Fusarium xylarioides* is nested within *Fusarium oxysporum* and *Fusarium oxysporum* f. sp. *raphani* is the closest match (also shown in Figure A.7). All branch support values = 100%, all trees drawn in Geneious 9.1. **E** The four effector genes are surrounded by *mimps* and DNA transposons on a robusta254 scaffold which shares a high (>96%) nucleotide sequence identity with the mobile and pathogenic *Fusarium oxysporum* f. sp. *lycopersici* chromosome 14. Similar scaffolds with the same four putative effector genes are present in robusta277, coffea659 and coffea674. Alignments were made with nucmer (MUMmer3). The blue annotations indicate the effector genes, the yellow annotations indicate *mimps* and the purple annotations indicate DNA transposons.

SIX effectors (being within 1.8 kb to 33 kb of *SIX* genes, Figure A.6, Schmidt et al. (2013)). Schmidt et al. (2013) suggested that effector genes found between interspersed repeat regions on chromosome 14 could coincidentally be transposed with *mimps*, which provides a working hypothesis for how *F. xylarioides* acquired these effectors. In addition, three more putative effectors match to mobile chromosomes that have not been linked with pathogenicity: *six10* and *pep1* match regions of chromosomes 3, 6 and 15, while *OG0014180* matches 3 and 6.

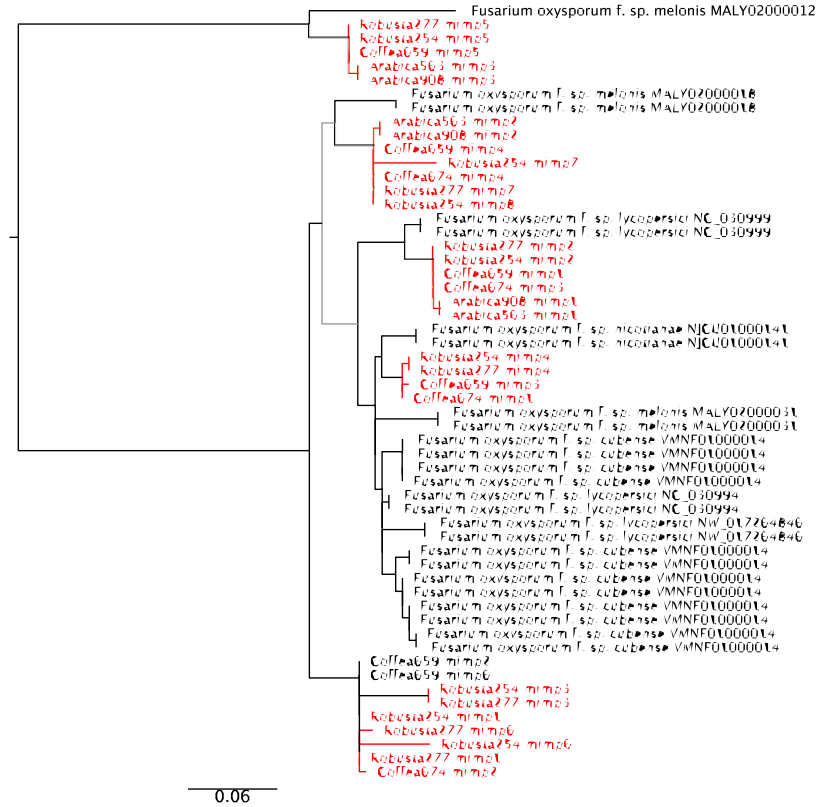
Beyond these examples, the full set of 64 effector loci tend to be close to TEs: the median distance from a TE for the putative effectors ranges from 2.2 kb to 3.9 kb, compared to 4.4 kb to 7.6 kb on average for random subsets of genes chosen from the whole genome (randomisation test, $p<0.05$ in robusta277, coffea674 and arabica563, $p>0.05$ in the other genomes, Table A.7). These numbers might be underestimates of average distances for effector genes found in lineage specific regions that are rich in repeats and TEs that were unable to be assembled. However, not all of the effectors associated with class II transposons display a signal of horizontal gene transfer in our phylogenetic analysis of effector genes. Another possible explanation for the association is that TEs and effector genes cohabit genome islands with high rates of variability, perhaps with TEs directly contributing to increased variability or modified expression of effector genes. In potential support of this idea, effector genes from the reference-guided scaffold assemblies tend to be close to large RIP affected regions (section 2.5.1, median distance from large RIP-affected regions ranges from 6.4 kb to 22.5 kb, compared to 17.5 kb to 28 kb for random subsets of the same number of genes, $p<0.05$ in robusta254, arabica563, and arabica908, $p>0.05$ in the other genomes). It is therefore possible that RIP-mutation and TE transposition within genomes might provide additional mechanisms for enhancing variability of effector genes (Rouxel et al. 2011), which could be explored by further analyses of population variation.



(a) mimp family 1



(b) mimp family 4



(c) mimp family mn4

Figure 2.9: Phylogeny of three *miniature impala* families in *Fusarium oxysporum, Fusarium xylarioides* and *Fusarium culmorum*. A *Mimp* family 1. Tree rooted with *Fusarium culmorum*. **B** *Miniature impala (mimp)* family 4. **C** *Mimp* family mn4. No other species matches were returned compared to the species in Table A.10 and the nr database (BLAST, 1e-50). Nodes which are coloured in black share a branch support value >90%, *Fusarium xylarioides mimps* are annotated in red. Phylogeny was inferred using Chi^2 support values. Drawn in Geneious 9.1.

2.6 Discussion

Using fungal strains collected over a 52-year period and optimally cryopreserved in a living state, the genomic basis of repeated outbreaks of *F. xylarioides* in different commercial coffee species was uncovered. These data support the conclusion that the *F. xylarioides* robusta and arabica groups had divergent origins within the *F. xylarioides* clade, and the robusta population originated within earlier outbreaks of more genetically diverse strains in alternative *Coffea* crops. Furthermore, by cataloguing multiple putative effector and pathogenicity genes, I show how different host-specific groups diverged in gene content and sequence by vertical processes within lineages, as well as by apparently acquiring 15 putative effector genes by horizontal gene transfer, mostly in robusta and coffea strains, from *F. oxysporum*.

The finding that arabica is divergent from the other strains is consistent with earlier work based on crossing experiments and molecular markers, which supports the reclassification of host-specific groups from the 1990s to the 2000s as separate species named *F. abyssiniae* (arabica) and *F. congoensis* (robusta). Evidence for concordant gene trees throughout the genome, as well as substantial differences in gene content, both overall and specifically, for genes believed to be important for the infection process, lends weight to this proposal. Further sampling of additional *F. xylarioides* genomes is needed to further test this. Rather than the presumed recent emergence of these strains, the level of diversity observed is consistent with a divergence far earlier than emergence within the last 50 years of these groups as major disease agents. Moreover, I found little evidence for the transfer of genes from the initial (1920s-1950s) outbreak to the 1990s-2000s epidemic on arabica coffee. I hypothesise that this represents a separate emergence in commercial arabica coffee from *F. xylarioides* strains associated with wild coffee relatives in Ethiopia or previously less noticeable disease symptoms in arabica coffee, but further sampling, including wild relatives, would be needed to test this. On the contrary, the *F. xylarioides* robusta group evolved with relatively minor modifications from the original outbreak in multiple *Coffea* species in central and western Africa, with only two putative effector

genes gained and a handful lost compared to the coffea strains sampled here. The high level of similarity between paired strains within each host-specific population, despite sampling from different locations or countries five years apart, supports the emergence of each population from low initial diversity, that is, each constitutes a single epidemic outbreak replacing earlier variants (Flood 2005, Phiri & Baker 2009).

The observation that 15 effector genes and their flanking regions are class II TE- and *mimp*-enriched, share a high percent identity with *F. oxysporum* and have emerged nested within *F. oxysporum* homologs, supports the role of horizontal transfer in the origin of host-specific differences. The mobile pathogenic chromosome of *F. oxysporum* is widely reported to transfer pathogenicity between different strains (Ma et al. 2010, Schmidt et al. 2013). However, to my knowledge, the transfer of pathogenicity has not been reported across different *Fusarium* species, nor has it been linked with the other mobile chromosomes. *F. oxysporum* is known to undergo conidial anastomosis (Kurian et al. 2018), so such transposition could have taken place in their shared niche, where both *F. xylarioides* and *F. oxysporum* (undescribed *f. sp.*) have been co-isolated from the roots and wood of coffee wilt diseased trees in Ethiopia and central Africa, as well as from banana roots where banana and coffee were intercropped in Uganda (Serani et al. 2007, Flood 2005). *Fusarium oxysporum* f. sp. *cubense* is widespread on banana crops in east and central Africa (Serani et al. 2007). *Fusarium xylarioides* is also known to infect tomato fruits (Onesirosan & Fatunla 1976) and cotton seedlings (Pizzinatto & Menten 1991), offering potential alternate hosts. While Kangire et al. (2002) did not find *F. xylarioides* in 105 different crop and weed species collected surrounding coffee farms in Uganda, the ability of *F. xylarioides* to infect *Solanaceous* crops and the Kayinja banana cultivar were investigated (Phiri & Baker 2009), but with no findings published. Therefore, *F. oxysporum* f. sp. *lycopersici, cubense* and *vasinfectum* (cotton) are of interest, as well as other *formae speciales* which are the closest match to numerous effectors: *F. oxysporum* f. sp. *pisi* and *F. oxysporum* f. sp. *raphani*. If confirmed, this would suggest avoiding the widespread practice of intercropping species that share closely related pathogens, for

example coffee and banana, or weed management for those which share pathogens, for example of *Solanaceous* weeds in crop surroundings. Greater sampling from these taxa and around coffee fields would be needed to further clarify the range of potential donors for horizontally acquired genes.

Contrary to previously described cases in *F. oxysporum* (Ma et al. 2010, Schmidt et al. 2013), I find no evidence that the transfer occurred on a known mobile chromosome, and in fact many of the putative effector genes are spread throughout the genome scaffolds. It remains possible that an unknown mobile chromosome or chromosomes with some homology to the *F. o. lycopersici* mobile chromosomes is involved with some of the variation among strains observed: I was unable to assemble the LS and FXU regions into whole chromosomes, where the matches to the *F. o. lycopersici* mobile chromosomes were concentrated. Future long-read sequencing allied to greater sampling of co-occurring taxa would confirm or refute this possibility.

These results also provide the first insight into the potential genetic basis for pathogenicity in coffee wilt disease. Of the 64 putative effector genes identified, roughly one third were shared by all *F. xylarioides* isolates. These are expected to include genes that contribute to the common pathogenicity traits of these fungi, ie those effectors that migrate to and induce vascular wilting first in the main infected stem and later in the systemic infection of the coffee tree as visualised as premature ripening of coffee berries (Rutherford et al. 2009). The main differences that were detected between host-specific populations were in gene content and the expansion of gene families. For example, while *Fusarium* is broadly enriched for lytic enzymes involved in carbohydrate metabolism (Reis et al. 2005, Soanes et al. 2008), and vascular wilt inducers for certain gene families (as also reported by Klosterman et al. (2011)), I found differences between the arabica and robusta *F. xylarioides* groups, particularly in enzymes involved in the metabolism of plant cell wall components. Specifically, host-specific populations had different enzymes, although they share similar functions in the hydrolysis of chitins, pectins, xylans, mannose, and rhamnose. For example, the arabica strains share seven CAZyme copies, whereas the

robusta strains share 19 different CAZymes with the coffea strains that are involved in the breakdown of the same sugars, as well as cellulose (none are found just in the robusta strains). Although it is not possible to precisely infer the phenotypic effects of these changes, expanded cell wall degrading enzymes imply variation in the capacity or mode of *F. xylarioides* to break down plant cell walls. The ability to use plant carbon suggests that the fungus can live within host coffee plants for a long time before disease symptoms become evident (Klosterman et al. 2011), something which has been reported in the field for *F. xylarioides* (Rutherford et al. 2009). A key benefit of the availability of living samples is that the importance of these differences can now be tested with functional assays.

Chapter 3

Multiple transfer events of large and highly-similar genomic regions between *Fusarium oxysporum* and different populations of the *Fusarium* coffee wilt pathogen

3.1 Acknowledgements

I thank W. Gordon, J. Meyer, M. Rutherford, T. Kibani, G. Armstong, D. Dring, R. Hillocks, E. Khonga and R. Mehrotra, as well as other un-named collectors who originally isolated the fungal pathogens and deposited them in the CABI-IMI collection, without which none of this work would have been possible.

3.2 Abstract

This chapter adds seven more historic *F. xylarioides* genomes and a long-read sequenced reference genome to test: the *F. xylarioides* host-specific groups found in chapter two; and the nature of the putative horizontal transfer. Four genetically differentiated populations were present: arabica, robusta and two populations amongst the strains isolated in the 1950s from the initial outbreak. Furthermore, six other *Fusarium* strains were sequenced from among the species that are also found on coffee wilt diseased trees: five *F. oxysporum* strains and one *F. solani*. The analyses showed evidence for multiple horizontal transfers of regions of more than 20 kilobases long between *F. oxysporum* mobile chromosomes and *F. xylarioides* populations. Furthermore, *F. xylarioides* is enriched with *miniature impala* transposons that are found in *F. oxysporum* and are associated with the horizontal transfer of pathogenicity. These results indicate that horizontal gene transfer occurred repeatedly in the *F. xylarioides* clade and could be one of the drivers behind the repeated emergence of coffee wilt disease during the last century through the transfer of genomic regions rich in genes.

3.3 Introduction

Throughout civilisation, humans have created an array of controlled and homogeneous nutrient-rich environments. This has led to rapid adaptation of microorganisms to exploit these new niches, such as microbial communities in agriculture (Gurr et al. 2011) and domesticated fungi used for fermentation (Dumas et al. 2020). Fungi are a good model system for studying the genomic mechanisms of adaptation in eukaryotes due to the existence of multiple complexes of sibling species with specialisation in different ecological niches and small genomes compared to most eukaryotes (Gladieux et al. 2014). Several mechanisms that are important in microbial adaptation involve gain gain or loss, gene duplication, loss of function due to premature stop codons, and gain of new genes by horizontal gene transfer (HGT) (Douglas & Langille 2019, Gladieux et al. 2014, Ropars

et al. 2015, Borneman et al. 2016). Horizontal gene transfers are known to be essential in prokaryote evolution (Hall et al. 2020), with evidence of frequent gene transfers throughout phyla (Kloesges et al. 2011). In eukaryotes, their larger and more complex genomes have historically made it harder to track HGTs. Gene gains and transfers were assumed to be isolated events in eukaryotes, not significant contributors to eukaryotic genomes (Van Etten & Bhattacharya 2020).

With the rise of long-read genome sequencing, important cases of horizontal gene and chromosome transfer have been described in eukaryotes, particularly between symbionts or parasites that share a common host (McDonald et al. 2019). Recent findings highlight prokaryote-to-eukaryote transfers, such as *Cyanidiophyceae* red algae that transitioned to extreme hot spring environments after horizontal uptake of archaea and bacteria genes (Rossoni et al. 2019). Examples of fungal HGTs in human-altered environments include those between two distantly-related *Pencillium* species which were independently domesticated for cheesemaking (Ropars et al. 2015), and the transfer of a whole mobile chromosome between plant-pathogenic *F. oxysporum* lineages, rendering a previously non-infectious strain pathogenic to a new host (Ma et al. 2010). These examples, as well as cross-kingdom transfers (e.g. (Richards et al. 2011), prove the occurrence of HGTs across many distantly related branches of the eukaryotic tree of life, demonstrating that they are an important force shaping eukaryotic evolution (Van Etten & Bhattacharya 2020).

Fusarium is a diverse genus of ascomycete fungi that is found in both natural and agricultural ecosystems. Plant pathogenic *Fusarium* fungi cause different types of diseases in 80% of all crops, including wilts, cankers and rots, and produce toxic secondary metabolites (mycotoxins) that can be toxic to cattle and humans (Leslie & Summerell 2006). These cosmopolitan traits make *Fusarium* fungi one of the most economically important pathogens (O'Donnell et al. 2013). There are believed to be 20 monophyletic species complexes within the *Fusarium* genus with various infection strategies in them. For example, *F. oxysporum* penetrates the roots before colonising the plant through the vascular system and triggering a wilt, while *F. graminearum* enters the flower and spreads

through its head to new and uninfected parts (Ma et al. 2013). The host-specficity varies across the *Fusarium* genus, with some pathogens infecting one host and others far more. However, the apparent generalism of some pathogens may be due to the existence of cryptic species, and some strains show high host specificity, which are termed *formae speciales (ff. spp.*). Such host-specific lineages have been widely reported for *F. oxysporum, F. graminearum* and *F. solani* (Aoki et al. 2014). For example, *F. oxysporum* has over 120 *ff. spp.* that parasitize important agronomic crops around the world. Some host specificities have repeatedly evolved, with, for example, multiple lineages having independently acquired host specificity on banana and others on tomato (O'Donnell, Kistler, Cigelnik & Ploetz 1998, Ma et al. 2010). The polyphyletic origin of host specificity on tomato is believed to be caused by the horizontal transfer between *F. oxysporum* ff. spp. of supernumerary chromosomes enriched in effector genes involved in pathogenicity (Ma et al. 2010).

The transfer of genes involved in pathogenicity can be mediated by mobile elements, which have been associated with HGT in fungal pathogens (Ma et al. 2010), or by anastomoses, whereby genetic material is transferred via somatic fusions between mycelia (Kurian et al. 2018). Genomes contain mobile transposable elements that can transpose between sites within a genome and replicate themselves. There are two main classes of transposable elements that use RNA (class I) or DNA (class II) in transposition, respectively. Class I elements transpose into RNA and reverse transcribe into cDNA to be inserted into a new locus; hence they are also known as retrotransposons (Muszewska et al. 2011). Class II elements are transposed via a cut-and-paste mechanism using terminal inverted repeats (TIRs) that identify the region to be excised and are known as DNA transposons (Munoz-Lopez & Garcia-Perez 2010). DNA transposons either contain their own transposase (which directs the cut-and-paste mechanism) or are non-autonomous, lacking a transposase, usually due to loss-of-function mutation. However, non-autonomous DNA transposons can still be mobilised by a transposase from a separate full-length copy with a transposase displaying similar TIRs (Feschotte et al. 2002, Bergemann et al. 2008).

Miniature inverted-repeat transposable elements (MITEs) are small, non-autonomous DNA transposons with TIRs, but missing transposase open reading frames (ORF). A particular type of MITE, called *miniature impala (mimp)*, has been identified in *F. oxysporum* which appear to be unable to transpose and are associated with supernumerary chromosomal subregions. These subregions are enriched in DNA transposons and identified effector genes, known as *SIX* that is, secreted in xylem (Schmidt et al. 2013). These effector-rich chromosomal subregions in *F. oxysporum* could exist due to accidental transposition by non-*mimp* DNA transposons (Schmidt et al. 2013). The clustering of pathogenicity genes could be beneficial for a homogeneous genomic environment for transcription (Batada et al. 2007, Schmidt et al. 2013), or result from the transfer of an entire pathogenic region (Shahi et al. 2016). Recently, *mimps* were identified in further pathogenic *Fusarium* species, including *F. hostae* and *F. proliferatum* (van Dam & Rep 2017) and *F. xylarioides* (chapter two).

The *Fusarium fujikuroi* species complex contains approximately 50 species across three clades, corresponding to the biogeographic origins of the plant hosts from which the strains were isolated, that is, African, American and Asian origins (Kvas et al. 2009). The *F. fujikuroi* species complex lineage and the *F. oxysporum* lineages diverged 11 million years ago in the Miocene (O'Donnell et al. 2013). The causal agent of coffee wilt disease, *F. xylarioides*, belongs to the African clade of the *F. fujikuroi* species complex. During the past century, its repeated outbreaks have dramatically changed the landscape of coffee production in sub-Saharan Africa (Hakiza et al. 2009, Girma et al. 2009, Phiri & Baker 2009). The 1920s-1950s coffee wilt outbreak corresponded to a pathogen population able to infect multiple *Coffea* species (this *F. xylarioides* strain group being hereafter called 'coffea'). However, two host-specific *F. xylarioides* groups have recently emerged, the 'arabica' type, which causes disease in *C. arabica* coffee trees in Ethiopia, and the 'robusta' type in *C. canephora* robusta coffee trees in the Democratic Republic of the Congo, Uganda and Tanzania. In chapter two (and Peck et al. (2021)), I show how the *F. xylarioides* arabica and robusta groups are monophyletic but differentiated and have emerged

independently within the *F. xylarioides* clade. Although the robusta strains' clade is monophyletic with high concordance, there is little concordance for their branching order with respect to the coffea strains, from which they appear to have emerged. Furthermore, the genomes contain evidence of multiple different transfers of genetic material from *F. oxysporum* separately into the arabica group on the one hand and into the robusta and coffea groups on the other. Of particular relevance for horizontal gene transfers, various *Fusarium* species have frequently been isolated from coffee wilt-infected trees, including *F. oxysporum*. The role of interactions between these species in the onset of coffee wilt disease is not known, but it is possible that they are the source of genetic material involved in the repeated emergence of coffee wilt disease.

Here, I test the hypothesis that the *F. xylarioides* arabica and robusta groups are genetically isolated lineages that have acquired new host specificity due to horizontal transfer from *F. oxysporum*, mediated by *mimp* mobile elements. Specifically, I addressed the following questions: (1) Are the *F. xylarioides* arabica and robusta groups genetically differentiated populations, one to each other and from the coffea group? (2) Do *F. oxysporum* and the host specialists of *F. xylarioides* show evidence of horizontal gene transfers, namely large and gene-rich regions that are highly similar and lacking in other species? (3) Do the *mimp* copies in *F. xylarioides* originate from *F. oxysporum*, and are they over-represented in the putative transferred regions? To answer these questions, I have sequenced with long-read technology the genome of an arabica strain and used this reference genome to scaffold: the genomes sequenced in chapter two, two published robusta genomes, and five new short-read sequenced arabica, robusta and coffea genomes. These new *F. xylarioides* assemblies have been compared to published *F. fujikuroi* complex genomes and a variety of newly sequenced and published *F. oxysporum* ff. spp. and *F. solani*, from coffee and other hosts, in eastern and central Africa, as well as other parts of the world. Using these methods, I show that the *F. xylarioides* arabica and robusta groups are genetically differentiated populations and differentiated from the coffea group, which is itself made up of two differentiated populations. The *F. xylarioides* arabica and

robusta groups both share large, syntenic, and highly similar gene-rich regions with *F. oxysporum* f. sp. *lycopersici*, as well as highly similar *mimps*, which are all absent from other *Fusarium* species. These likely horizontally transferred regions from *F. oxysporum* to *F. xylarioides* are different in the arabica and robusta groups and therefore could have contributed to the emergence of new disease types.

3.4 Methods

3.4.1 Fungal strain details

All isolates which were used in this chapter are detailed in Tables A.9 and A.10, along with what they were used for.

3.4.2 Pilot study: PCR and Sanger sequencing

To investigate putative horizontal transfers, the presence of a subset of putative effectors from chapter two was determined in 23 additional *Fusarium* strains from the CABI collection (Table 3.1). In total, 19 putative effectors were selected: 15 showed evidence for horizontal transfer from *F. oxysporum (pep1, nep1, six10, OG0013477, OG0014398/ six7, OG0016241, OG0016247, OG0014741, OG0014743, OG0014180, OG0013478, OG0009441, OG0000973, OG0013765, OG0014179*, and three matched to *F. oxysporum*'s mobile pathogenic chromosome (*six7, cytoskeletal* and *OG0000409*). The additional strains were selected for either being isolated from a coffee plant host, isolated from a sub-Saharan African coffee growing region or an *F. oxysporum* f. sp. of interest from chapter two. Primers were designed for the 19 putative effectors and two reference orthologous groups to recapitulate the species tree phylogeny from Figure 2.4 (Table 3.2).

DNA was extracted from each strain following the same protocol as in chapter two. The regions of interest were amplified by PCR using a PTC-200 Peltier Thermal Cycler in a reaction mix containing 3 pmoles of each primer, 1μl of template DNA solution and 10μl of MegaMix Royal (Microzone Ltd, Haywards Heath, UK). MegaMix Royal contains

Table 3.1: Strain bio-geographic details

Species	Short name	Strain	Year	Origin	Host	Sequenced
F. xylarioides	Robusta925	379925	1998	Uganda	*C. c. robusta*	
F. xylarioides	Robusta268	392268	1968	DRC	*C. c. robusta*	Y
F. xylarioides	Robusta272	392272	2002	DRC	*C. c. robusta*	
F. xylarioides	RobustaL0394	L0394	unknown	Uganda	*C. c. robusta*	
F. xylarioides	Coffea676	392676	1964	Guinea?	*Coffea*	Y
F. xylarioides	Coffea035	507035	1979/1963	Guinea	*C. canephora*	Y
F. xylarioides	Coffea113	507113	1950	CAR	*unknown*	Y
F. xylarioides	Arabica038	507038	1971	Ethiopia	*C. arabica*	Y
F. xylarioides	Arabica746	204746d	1977	Ethiopia	*C. arabica*	
F. xylarioides	Arabica567	389567	2004	Ethiopia	*C. arabica*	
F. xylarioides	Robusta089	393089	1998	Uganda	*Musa*	
F. udum		310543	1986	India	*Cajanus cajan*	
F. udum		275452	1983	Malawi	*Cajanus cajan*	
F. verticillioides		375318	1999	Germany	*Zea mays*	
F. oxysporum f. sp. raphani	*F. o. raphani 541*	337541	1989	United States	*Raphanus sativus*	Y
F. oxysporum f. sp. lycopersici	*F. o. lycopersici 54003*	NRRL 54003	unknown	United States		
F. oxysporum	*F. o. coffea 509*	244509	1979	Tanzania	*C. arabica*	Y
F. oxysporum f. sp. pisi	*F. o. pisi 221*	500221	unknown	unknown	*Pisum sativum*	Y
F. oxysporum f. sp. cubense	*F. o. cubense 109*	141109	1969	United States	*Musa*	Y
F. oxysporum f. sp. vasinfectum	*F. o. vasinfectum 248*	292248	1985	Tanzania	*Gossypium*	Y
F. oxysporum	*F. o. coffea 581*	389581	2002	Ethiopia	*C. arabica*	
F. stilboides		392281	2003	Guinea	*C. canephora*	
F. stilboides		389578	2002	Ethiopia	*C. arabica*	
F. solani	*F. solani 280*	392280	1963	Guinea	*C. c. robusta*	Y

Table 3.2: Putative effector primer sequences for polymerase chain reaction

Gene	Forward	Reverse
pep1	CTTATCCACAGGCACCCTCG	GGCTTGCTTTTGACAACCGT
nep1	CAGACATATGCTTGTGCCGC	CCTTGTTGAGGTTGTTGCCG
six7	GCACCTGGAGTCTCGCTATC	AGCCGCAAAAGCAAAAGGAG
six10	CAGGTATTGTGGGCTTCGGT	GGTTGCTGTGTTGCTTCCTG
cytoskeletal	ACCCTGAAGAACGACTTGCC	GCGAGTACCCCTTCAACTGG
OG0013477	CGCCGAGGATGCAAAATGAA	ATGACATACGCGCCATCAGT
OG0014398/Six7	CTTCTGCCTGTTTGCACTGG	CAGCAGCTCCACCACCATT
OG0016241	GTCGAAGTGTCCTTGGTCGT	TCTCTCCATCCGAGTCTGCA
OG0016247	CAGCGTCAACTGCCATCATG	GATAAACGCTCGGCTTGCTG
OG0015465	TTGCTGGGATGGACATGGTC	GTTGATCTGGACGACGGGTT
OG0014741	TTGACGTGGTTGAGGGATGG	CCCGTGATGGAGGCATAGAC
OG0014743	CAGCGTCTCTAGGCTATGGC	GTGAGCTCTGTCGGTTTCCA
OG0014180	CCGTCTTGGCCATTGCGAT	TTTGTAGACCAAGTCCAGCCG
OG0013478	CCAATACTAGCCGCCGATGT	CAGTGATGCCGGCGATATCT
OG0009441	GATGGAGGCAAGGTTGGGAA	TCCTTCTGATGTCGTCCCCT
OG0000973	CAGTTCTATTCGCAACGCCG	AGGGCAGTGACTTTACCAGC
OG0013765	GCACATCCAGTMCATCGTCA	ACAACACAGCTGTAACCCGA
OG0000409	AACCACAGCAGGACTCGAAG	CTGTTGCAGCCCCAAATGTC
OG0014179	CGAGACCGTCTACCGACTTG	GATCTTCTCCGCCACTGGAG

chemically modified Taq polymerase in 2x enhancing buffer (6 mM MgCl2) with 400μM dNTPs. Amplification conditions were: 95°C for 5 min followed by 35 cycles of 30 seconds at 95°C, 45 seconds at 54°C, 120 seconds at 72°C, followed by 10 min at 72°C and held at 10°C. The PCR products were separated by electrophoresis in a 1.5% agarose gel and visualised with a 1kb size ladder (Invitrogen, UK) and compared with a genomic DNA positive control sample.

The PCR products were sequenced by Genewiz/ Azenta (Essex, UK), except for the *F. xylarioides* products in which gene presence in these strains was verified using the expected product size. The Sanger sequence chromatograms were cleaned and aligned with phylogenetic trees drawn in Geneious v9.0 with MAFFT to interpret the presence and phylogeny of the putative effectors, to compare with the species tree phylogeny of the strains using the reference Sanger sequences. This pilot study advised which strains would be used for whole-genome sequencing (Table 3.1).

Table 3.3: Genome statistics for sequenced *Fusarium* strains. Arabica563 is the reference assembly. * indicates a publicly available genome, ˆ indicates a genome sequenced in chapter two.

Short name	Genome size Mb (>1kb)	No. contigs (>1kb)	N50 kb	Complete BUSCO %	GC %	Coding Genes
Arabica563	57.3	21	5,156	100.0	42.2	
Arabica563ˆ						17,457
Arabica038	51.1	4631	24.5	97.9	43.4	15,705
Arabica908ˆ						15,486
Robusta268	50.2	4420	27.1	97.6	44.1	15,575
*Robusta925**						15,478
*RobustaL0394**						15,399
Robusta254ˆ						15,486
Robusta277ˆ						15,578
Coffea035	51.1	3880	30.4	97.9	43.9	15,588
Coffea113	48.7	3795	30.1	98.3	44.3	15,115
Coffea659ˆ						15,464
Coffea674ˆ						15,292
Coffea676	50.0	4362	27.1	97.9	44.4	15,649
F. solani 280	48.4	1867	45.2	96.9	50.9	14,007
Fo coffea 509	49.7	1923	67.9	98.3	47.4	14,475
Fo cubense 109	48.4	1992	60.1	99.0	47.6	14,993
Fo pisi 221	52.9	4299	28.1	95.5	47.5	16,614
Fo raphani 541	53.8	4285	38.1	98.3	47.7	16,421
Fo vasinfectum 248	49.4	1904	72.5	98.6	47.4	14,846

3.4.3 High-molecular weight DNA extraction, long-read genome sequencing and assembly

The *F. xylarioides* strain IMI 389563 (hereafter called 'arabica563') from the arabica host-specific strain group of coffee wilt disease was used for long-read sequencing, and is available from the CABI-IMI culture collection (CABI, Egham, UK).

Arabica563 mycelium was grown in GYM broth for 7 days at 25^{o}C and 175 RPM. GYM broth contains 10g glucose, 1g ammonium di-hydrogen phosphate $NH_4H_2PO_4$, 0.2g potassium chloride, 0.2g magnesium sulphate hydrate $MgSO_4.7H_2O$, 5g yeast extract (oxoid), 1ml 1% zinc sulphate $ZnSO_4.7H_2O$, 1ml 0.5% copper sulphate $CuSO_4.5H_2O$ with 1 litre distilled water. High-quality high molecular weight genomic DNA was extracted from 1-2 g of mycelium ground in liquid nitrogen using a lysis buffer containing sodium metabisulfite, 2M Tris, 500 mM EDTA, 5M NaCl, and 5% Sarcosyl. The gDNA was extracted from the cell lysate using Chloroform, and precipitated using Isopropanol and 70% ethanol, following the protocol of Spanu et al. (2010). The gDNA was eluted in 500μl TE buffer and incubated overnight at 4^{o}C. RNA contaminants were removed using RNAse A (10mg/ ml) and gDNA was purified using Agencourt AMPure XP beads (ThermoFisher). The gDNA quantity was determined using a NanoDrop 2000 (ThermoFisher) and its integrity was determined by electrophoresis with a 0.5% agarose gel. 1 μg of gDNA was diluted with nuclease-free water to a final volume of 47 μl. Half of the DNA was then shared into 20 kb fragments with a COVARIS g-TUBE (Covaris, Inc., Woburn, MA, USA), centrifuging at 4,200 rpm for 1 minute to increase DNA throughput. The shared DNA was converted into Oxford Nanopore libraries using the SQK-LSK110 library kit and the short fragment buffer (SFB) (Oxford Nanopore Technologies, UK). Library prep followed the standard manufacturer's protocol apart from the final elution, which was done for 30 minutes at 37^{o}C. The libraries were sequenced using a SpotON R9.4.1 FLO-MIN106 flowcell for 48 hours.

Nucleotide bases were called from the raw sequence data using the super high accuracy model (SUP) of the GPU version of Guppy v6.0.1 (config file: dna_r9.4.1_450bps_sup.cfg).

Basecalled reads were assembled into contigs by Flye v2.9 with '-nano-hq' parameters (Kolmogorov et al. 2019). The Flye assembly was error-corrected using one round of medaka v1.4.4 (https://github.com/nanoporetech/medaka) and one round of (Hapo-G) polishing (Aury & Istace 2021). Contig completeness was visualised in Tapestry v1.0.0 (Davey et al. 2020), including the presence of the terminal telomere repeat sequence TTAGGG, common to eukaryotes and found in *Metarhizium* telomeres (Inglis et al. 2005) at the ends of each long read. This arabica563 long-read assembly will be referred to as the reference genome henceforth.

3.4.4 DNA extraction for Illumina sequencing and assembly

Based on the results from section 3.4.2, an additional set of 11 strains was selected for Illumina whole-genome sequencing: five *F. xylarioides*; five *F. oxysporum* and one *F. solani* (Table 3.3). All strains will be referred to by their 'short name' (Table 3.3). All strains were from the CABI culture collection (Egham, UK), with IMI 507113 transferred from the Belgian coordinated collection of microorganisms from the Université Catholique de Louvain (Louvain-la-Neuve, Belgium) (under MUCL 47066/ MNHN 709) and IMI 507035 and 507038 transferred from the Agricultural Research Service culture collection (Illinois, USA) (with 507035 under NRRL 25804/ BBA 62721/ CBS 749.49; and 507038 under NRRL 37019/ FRC L96/ BBA 62458). Of the five *F. xylarioides* strains, three were from the coffea strain group (coffea035, coffea113, coffea676, and the remaining two were the earliest isolates available in the collections for each of the arabica and robusta groups (both isolated from around 1970, arabica038 and robusta268).

Strain morphologies were verified and grown in GYM broth (see section 3.4.3 for recipe) following the same protocol as chapter two (Methods). DNA was extracted from 250-500mg of washed mycelium following the NucleoSpin Soil Mini Kit (Macherey Nagel, Germany) standard protocol. For each strain, a single library was prepared with the Illumina DNA PCR-Free Prep and sequenced with the Illumina Novaseq 6000 with 2x 150bp reads. Low quality bases (Phred score <20) and adapters (stringency 4) were removed us-

ing TrimGalore 0.6.0 by Cutadapt (https://github.com/FelixKrueger/TrimGalore, (Martin 2011). Overlapping paired-end reads were stitched together using FLASH 1.2.11 (Magoč & Salzberg 2011) and *de novo* assembled with MEGAHIT 1.2.9 (Li et al. 2015*b*, 2016). Assembly metrics including BUSCO scores were computed using QUAST 5.2.0 (Table 3.3) (Mikheenko et al. 2018)

Each *F. xylarioides* MEGAHIT assembly, as well as the assemblies from chapter two and the publicly available genomes robusta925 and robustaL0394 (Table 3.1), was mapped to the reference genome assembly using RagTag v2.1.0 homology-based scaffolding (Alonge et al. 2019). The parameter '-u' identified unmapped (and therefore absent from arabica563) scaffolds. The *F. oxysporum* and *F. solani* MEGAHIT assemblies were mapped to their closest and least fragmented publicly available assembly (Table A.9). The *F. o. lycopersici* and *F. o. cubense* genomes were the least fragmented (see N50 values).

3.4.5 RNA sequencing to aid annotation of long- and short-read assemblies

The same arabica563 strain used for long-read sequencing was grown under the same culture conditions (section 3.4.3). Mycelia were harvested and total RNA extracted using the standard protocol of the RNeasy Mini kit (Qiagen, Hilden, Germany). The RNA quality assessment (Qubit and Fragment Analyzer), library preparation and Next Generation Sequencing (60M paired-end reads per sample) were performed by Genewiz/ Azenta (Frankfurt, Germany). Reads were trimmed with Trimmomatic v0.39 (parameters: ILLUMINACLIP:$TRIMMOMATIC DIR/adapters/TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25) (Bolger et al. 2014) and quality-checked with FASTQC v0.11.2 (Andrews S et al. 2015). Transcript sequences were *de novo* assembled using Trinity v2.13.2 with the genome-free method (Grabherr et al. 2011).

3.4.6 Gene prediction and orthologous clustering

Protein-encoding genes were predicted using the Funannotate Eukaryotic Genome Annotation Pipeline v1.8.11 (Palmer & Stajich 2020). Before annotation, repetitive contigs were removed using 'funannotate clean' with default parameters and the repetitive elements within the assemblies were soft masked using RepeatModeler and RepeatMasker with 'funannotate mask' and *F. oxysporum* parameters.

Fusarium xylarioides genomes

All 13 *F. xylarioides* genomes were annotated (Table 3.3), including re-annotations for the six genomes from chapter two. The trimmed RNA-seq reads and the trinity assembly were aligned to the arabica563 MEGAHIT assembly using the parameters 'funannotate train –jaccard_clip' which runs PASA to model the gene structures. The PASA gene models were parsed to run Augustus, snap, GeneMark and GlimmerHMM, with the predictions used to run CodingQuarry and the Funannotate Evidence Modeler. The complete set of arabica563 output files were then used to predict protein-encoding genes on the remaining MEGAHIT genome assemblies using 'funannotate predict -p predict_results/fusarium_xylarioides_389563.parameters.json'. The arabica563 MEGAHIT assembly was selected over the reference genome to avoid issues with gene prediction on assemblies of different quality when the parameters were applied to the remaining *F. xylarioides* MEGAHIT assemblies. The gene model predictions were refined (e.g. correcting intron-exon boundaries) and untranscribed regions (UTRs) annotated by re-aligning the RNA-seq data for each genome using 'funannotate update' on the output files from 'funannotate predict'.

The annotations were then mapped to the long-read RagTag assemblies using Liftoff (Shumate & Salzberg 2021) with default parameters.

Other *Fusarium* genomes

The same steps were followed for the five *F. oxysporum* genomes and the *F. solani* genome, using different sources of evidence for the different species. For the best gene model predictions from Funannotate, high-quality protein models from transcripts or RNA-Seq data should be used. Therefore, sorted BAM files (RNA-Seq reads for the closest *f. sp.* match were downloaded from the sequence read archive and aligned to the corresponding assembly using 'bbmap.sh'), and expressed sequence tag data for the closest *f. sp.* match from JGI Mycocosm were used to provide evidence to 'funannotate predict'. The expressed sequence tag data were produced by the US Department of Energy Joint Genome Institute https://www.jgi.doe.gov/ in collaboration with the user community. Evidence sources are listed in Table A.9. The same RNA-Seq data was then re-aligned through 'funannotate update' as in section 3.4.6.

Functional annotation of the protein-coding genes for all genomes was completed with annotations from PFAM, InterPro, UniProtKB, MEROPS, CAZyme and GO Ontology through 'funannotate annotate'. InterProScan v5.56-89.0 was run locally using the parameters '-goterms -iprlookup' and phobius was run through the web browser (https://phobius.sbc.su.se/). All proteins lacking significant hits were annotated as hypothetical proteins.

3.4.7 Transposable elements and repeat annotations

Repeats and transposable elements (TEs) were identified from the nucleotide assemblies. A custom repeat library was constructed using all *Fusarium* RagTag assemblies in RepeatModeler v2.0.3 with the parameter '-LTRStruct' (Flynn et al. 2020). Added to this library were TEs known to be involved in *F. oxysporum* transfer of pathogenicity between strains: *impalas*, a new set of TEs described by Schmidt et al. (2013), as well as manually curated *mimps*. Sequences for *mimp* families 1-4 were downloaded from NCBI (accession numbers: AF076624.1; AF076625.1; EU833100.1; EU833101.1). Sequences for the more divergent *mimp* families mimp5, mimp6, mn2, mn3, mn4, mn8 and mn14 are from Berge-

mann et al. (2008). Manual curation followed the protocol described by Goubert et al. (2022) and involved: identifying conserved domains with sequence homology to these TEs using BLAST; only retaining the best BLAST hits (>= 50% query length and >= 80% identity) for each TE family; extracting and aligning nucleotide sequences for TE families including an additional 500 bp up- and down-stream; trimming alignments to the TIRs at their start and end and removing insertions using T-COFFEE (Notredame et al. 2000). This generated consensus sequences for each *mimp* family that were added to the RepeatModeler custom library and used to call TEs and repeats from genome scaffolds using RepeatMasker v4.1.2 with the parameters '-lib $LIB -gff -no_is' (Smit et al. 2015). The presence of the *impala* transposase ORF from *F. oxysporum* f sp *melonis* was found by a BLAST search 'tblastn -evalue 1.00e-5' using accession AAB33090.2.

3.4.8 Orthologous gene groups and genetic structure

Whole-genome similarity between the protein-encoding genes was assessed using NSimScan, part of the QSimScan package (Novichkov et al. 2016). OrthoFinder was used to determine *Fusarium* orthologous groups between the genomes sequenced in this chapter and a set of publicly available genomes (listed in full in Table A.10). The additional genomes were: eight *Fusarium fujikuroi* complex strains, 10 *F. oxysporum* ff. spp., two distantly-related *F. graminearum* strains and an outgroup *F. solani* strain. The OrthoFinder species tree is drawn using the orthogroups in which all species are present. A custom R-script was used to identify the single-copy orthogroup gene trees closest to the species tree, with 117 of these merged to identify consensus and differences among the single-copy trees using 'ggdensitree' ((Yu et al. 2017). Single nucleotide polymorphisms (SNPs) were called against the arabica563 reference genome using GATK v4.1.2.0 'HaplotypeCaller -ERC GVCF', which provides one Genomic Variant Call Format (gVCF) output file per strain. Using Snakemake v5.3.0, GVCFs were combined using GATK 'CombineGVCFs', genotypes were combined with 'GATKGenotypeGVCFs', 'SelectVariants -select-type SNP' selected the SNPs which were then filtered with 'VariantFiltra-

tion QUAL<30, DP<10, QD<2.0, FS>60.0, MQ<40.0, SOR>3.0, QRankSum<-12.5, ReadPosRankSum<-8.0'. The SNP dataset was used to infer population structure. Principal component analysis of SNP data was completed using the R packages SNPArray and gdsfmt with a linkage disequilibrium threshold of 0.05 (Zheng et al. 2017, 2012). The nucleotide diversity π (Hudson & Wu 1992, Nei & Li 1979), Watterson's ϑ (Watterson 1975), fixation index F_{ST} (Hudson & Wu 1992) and the absolute nucleotide divergence d_{xy} (Nei & Li 1979) were calculated using the R package Popgenome (Pfeifer et al. 2014).

3.4.9 Identifying putative horizontal transfers

Putative horizontal transfers were identified by mapping *F. xylarioides* assemblies and *Fusarium fujikuroi* complex genomes (full list in Tables A.9 and A.10) to an *F. oxysporum* or *F. solani* putative source for horizontal transfers. Specifically, I looked for regions which were present in one or all of the *F. xylarioides* populations and the *F. oxysporum* reference genome, and which were absent from the *Fusarium fujikuroi* species. The analysis was run seven times, each time using a different strain as the reference genome (*F. o. cubense* 109, *F. o. lycopersici* 4287, *F. o. pisi* 221, *F. o. coffea* 509, *F. o. raphani* 541, *F. o. vasinfectum* 248, *F. solani* 280. First the assemblies were mapped to the reference *F. oxysporum* genome using Minimap2 with the parameters '-cx asm10 –secondary=no –cs' (Li 2018). The alignment coverage was calculated in 50 bp sliding windows using 'bedtools makewindows' and the median coverage per window calculated using 'bedtools map -c 4 -o median'. Regions with no or low (<10%) coverage were removed and shared regions across the prescribed groups (arabica, robusta, coffea, *Fusarium fujikuroi* species) were identified by merging windows of coverage <200 bp apart ('bedtools merge -c 4 -o median -i - -d 200') and grouping by population (arabica, robusta, coffea, *Fusarium fujikuroi* complex). Regions were classified as putative horizontal transfers if they were present in all strains in the prescribed group and absent from all others. Their presence and absence were further verified with both the initial minimap2 aligmment and BLAST. Gene presence on the shared regions was calculated by extracting all genes from the published GFF for the

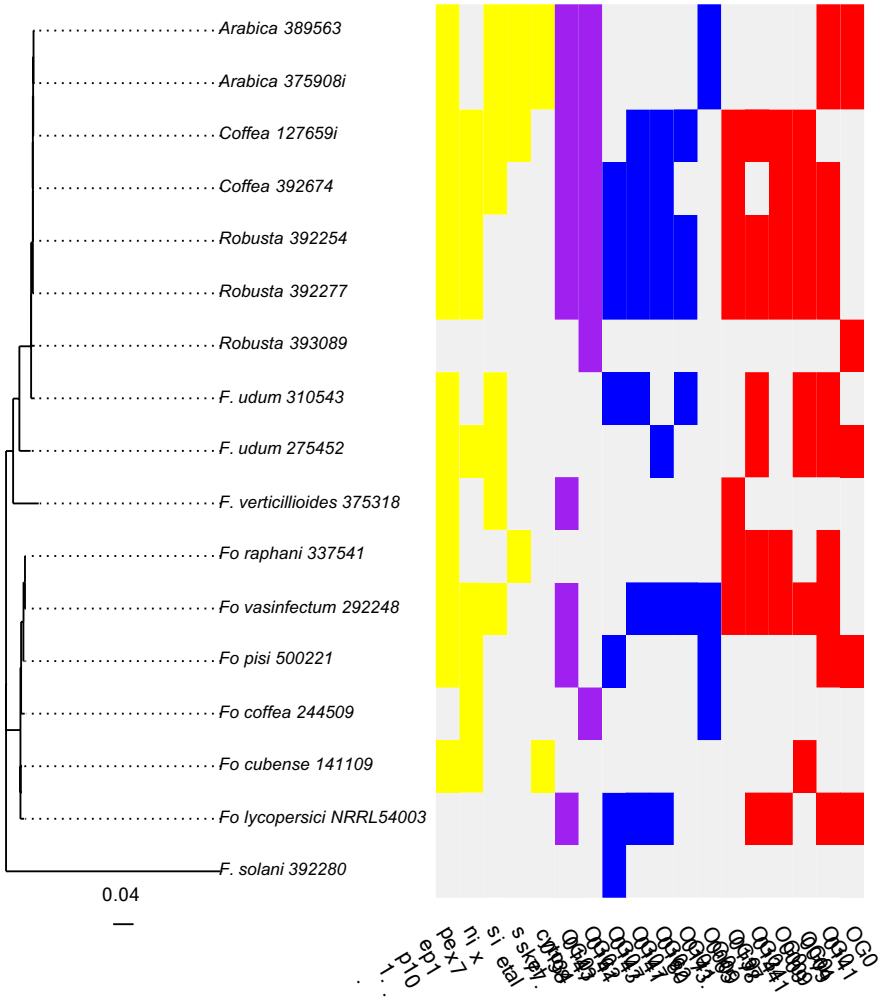


Figure 3.1: Putative effector genes are shared by *Fusarium xylarioides* and *Fusarium oxysporum*, but absent from *Fusarium verticillioides*. Phylogenetic relationships between *Fusarium* strains reconstructed from Sanger sequences of PCR products for two genes which yield congruent topologies with the whole-genome phylogeny in Figure 2.4. The presence of each putative effector gene was tested in each strain, presence is represented by shading of the effector class, as follows: yellow for pre-described plant pathogen effectors; purple for small and cysteine-rich proteins; blue for CAZymes; and red for genes with a transposon in the promoter. Genes which do not show evidence of horizontal transfer are represented with a dot after their name. *F. oxysporum* f. sp. is abbreviated to *Fo* in the tip labels of the tree.

genome used in the RagTag scaffolding (see Table A.9) for the shared coverage region, and looking for highly similar matches using BLAST (length >80%, e-value <1.00e-20, identity > 91%).

3.4.10 Data analysis

All analyses were performed in R using R Statistical Software (v4.2.1; R Core Team 2022) (R Core Team 2021) and in Linux run on the Imperial College Research Computing Service HPC facility (DOI: 10.14469/hpc/2232). The following R packages were used to make figures: tidyverse (Wickham et al. 2019), cowplot (Wilke 2022), ggrepel (Slowikowski 2021), wes anderson (Ram & Wickham 2018).

3.5 Results

3.5.1 A pilot study: *Fusarium xylarioides* shares putative effector genes with *Fusarium oxysporum* and *Fusarium solani*

A preliminary analysis was conducted to choose the genomes to be sequenced. Of the 19 putative effector genes investigated, 16 were shared between *F. oxysporum* and *F. xylarioides*, but absent from *F. verticillioides* (a plant pathogenic member of the *Fusarium fujikuroi* species complex (Ryan et al. 2022). In this pilot study, genes which are absent from *F. verticillioides* but present in *F. xylarioides* and *F. oxysporum* are taken to be suggestive of horizontal transfer. The *F. oxysporum* f. sp. *vasinfectum* 248 genome shares the highest number of putative effector genes with the *F. xylarioides* strains (Figure 3.1). One putative effector was shared between *F. oxysporum* f. sp. *pisi* 221 and *F. solani* with the *F. xylarioides* strains. Hereafter, *F. oxysporum* ff. spp. will be abbreviated to *Fo* followed by their *forma speciales* and strain number.

Consequently, I sequenced the genomes of five *F. oxysporum* ff. spp., one *F. solani*, as well as five additional *F. xylarioides* strains (full details in Table 3.1 and A.9).

3.5.2 A reference assembly for *Fusarium xylarioides*

Assembly of long-read data for *F. xylarioides* IMI 389563 ('arabica563') generated a reference genome with 21 contigs, including 12 contigs over 1 Mb, four more contigs over

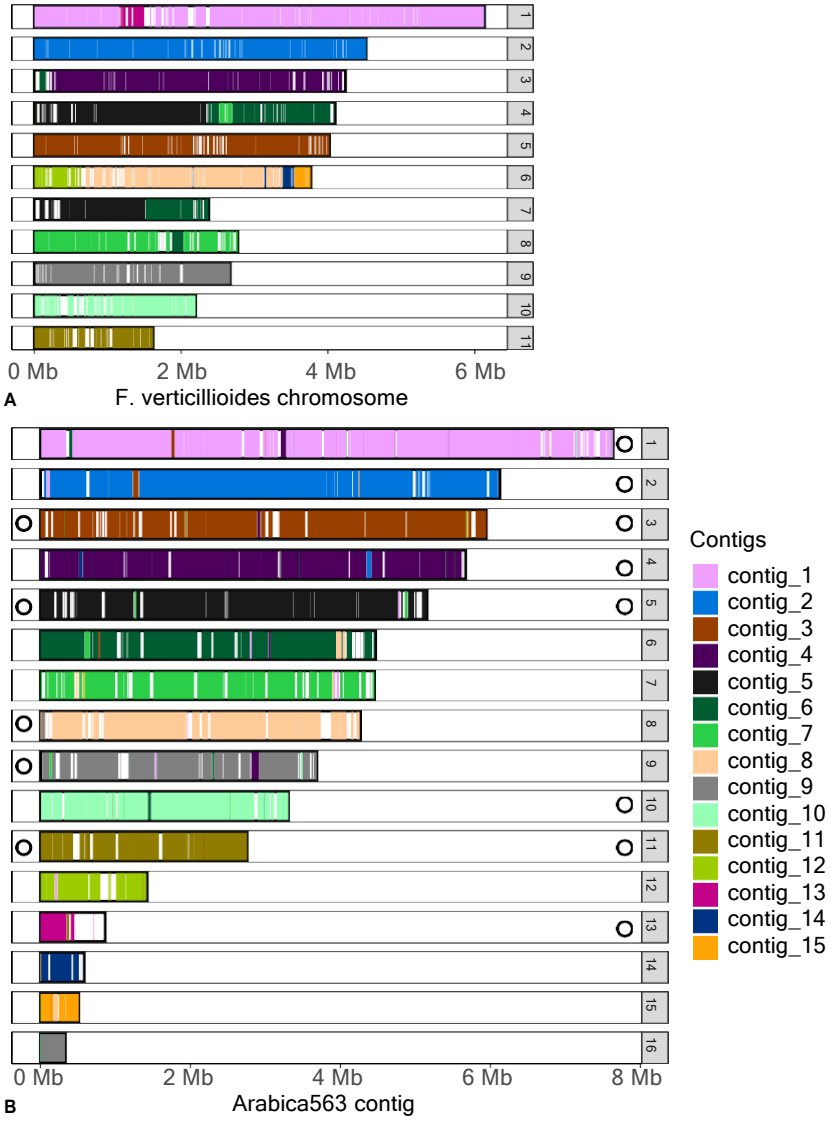


Figure 3.2: Whole-genome alignments (A) *Fusarium xylarioides* arabica563 mapped against *Fusarium verticillioides* chromosomal assembly and (B) *Fusarium xylarioides* robusta254 mapped against the *Fusarium xylarioides* arabica563 contig assembly. A Each bar represents one of the 11 *Fusarium* chromosomes present in *F. verticillioides* that are also present in *Fusarium xylarioides* arabica563, with chromosome size and contig. position on the x axis. Chromosome numbers are annotated in-line with the bars. Shaded regions indicate the presence of syntenic *Fusarium xylarioides* arabica563 contigs, the colour of which identifies the specific mapped contig from *Fusarium xylarioides* arabica563. Unshaded regions indicate contigs found only in *F. verticillioides* i.e. absent from *Fusarium xylarioides* arabica563. *F. verticillioides* chromosomes 1-11 correspond to CM010967.1-CM010977.1 in the genome accession GCA_000149555.1. **B** Each bar represents one of the 16 *Fusarium xylarioides* arabica563 contigs that are also present in *Fusarium xylarioides* robusta254, with contig size and position on the x axis. Shaded regions indicate the presence of syntenic *Fusarium xylarioides* robusta254 contigs, the colour of which identifies the specific mapped contig from *Fusarium xylarioides* robusta254. Unshaded regions indicate contigs found only in *Fusarium xylarioides* arabica563 that are absent in *Fusarium xylarioides* robusta254. Telomeric repeats are annotated as black circles on their corresponding scaffolds. Contigs <1 kb not shown.

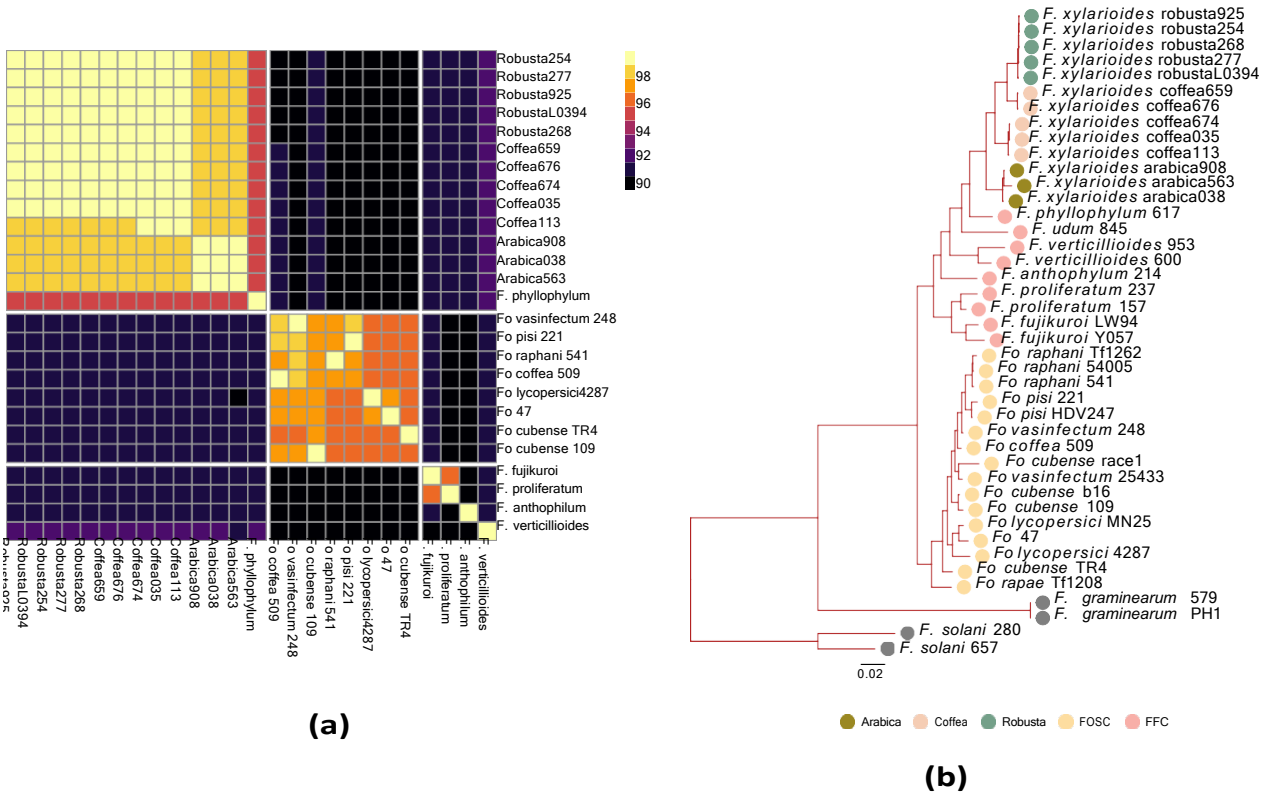


Figure 3.3: Whole-genome similarity and phylogeny between *Fusarium oxysporum* ff. spp., *Fusarium xylarioides* and *Fusarium fujikuroi* complex species. A Whole-genome similarity. Each cell indicates the nucleotide similarity across all predicted coding sequences between the two strains, with the shading scale. Arabica, robusta and coffea genomes all belong to *F. xylarioides*. One genome is shown for each species/ *formae speciales*. **B** Species tree drawn using STAG (Emms & Kelly 2018) with 3544 orthogroups in which every strain has at least one gene copy. Abbreviations: Fo, *F. oxysporum*; the last 3 digits of each strain number is shown. Full strain details in Tables A.9 and A.10.

Figure 3.4: Phylogeny of the genus *Fusarium*. A density tree drawn using a subset of single-copy gene trees. The tree is ordered by the most common topology. Tree tip labels are shaded by phylogenetic group: the arabica, robusta and coffea *F. xylarioides* strains; FFC, *Fusarium fujikuroi* species complex; FOSC, *F. oxysporum* species complex. Outgroups are shown with a grey tip label. Full strain details in Tables A.9 and A.10.

100 kb and five short contigs smaller than 1 kb. The assembly length was 57 Mb (with 55 Mb in the 12 longest contigs) with an N50 over 5 Mb (Table 3.3). The 16 longest contigs mapped to the *F. verticillioides* chromosomal assembly (Figure 3.2a). The *Fusarium* genus contains 11 syntenic chromosomes which are shared by closely and distantly related species, and which make up the 11 *F. verticillioides* chromosomes (Ma et al. 2010). Arabica563's contigs five and six map to *F. verticillioides* chromosomes four and seven but with evidence of major chromosomal translocations between them. The *F. xylarioides* contigs also have additional unique material compared to *F. verticillioides* chromosomes, making up approximately 30% of the total length of each contig (sequence lengths in Figure 3.2a, 3.2b). In addition, I present 11 Illumina genome assemblies, with five new *F. xylarioides* strains and six *Fusarium* species (five *F. oxysporum* and one *F. solani* genomes (Table 3.3). The mapping of each *F. xylarioides* short-read genome to the long-read assembly resulted in contiguous assemblies (Figure 3.2b). These contiguous assemblies allowed identification of five large arabica-specific regions over 700 kb on contigs 6, 8, 9, 12 and with the largest at 2 Mb on contig 13. Short regions of mapping differences between arabica563 and robusta254 contigs (e.g. contig 1 in arabica563 matches contigs 1, 6, 3 and 4 in robusta254) probably represent repetitive regions. Telomeric repeats were identified on both ends of three and one end of seven contigs (Figure 3.2b).

3.5.3 Gene orthology and analysis

Predictions of protein-encoding genes across the *F. xylarioides* genomes sequenced in this chapter as well as two publicly available genomes revealed on average 15,500 genes (Table 3.3). More proteins were predicted than in chapter two (Table 2.1) thanks to the inclusion of RNA-seq data. A similar number of genes were predicted across the *F. oxysporum* and *F. solani* genomes.

To search for transfers between *F. xylarioides* and *F. oxysporum* ff. spp., whole-genome gene predictions were compared with additional published *Fusarium* genomes. The strains within the robusta and coffea *F. xylarioides* types were 99% similar to each

other, except coffea113 which ranged from 98-99% similar to the other robusta and coffea strains. All robusta and coffea *F. xylarioides* strains were 98% similar to *F. xylarioides* arabica strains (Figure 3.3a). The *F. xylarioides* genomes were broadly 90% similar to *F. oxysporum* ff. spp. genomes. *Fusarium xylarioides* is most similar to *F. phyllophylum* (95%), and 91-92% similar to the other species from the *Fusarium fujikuroi* complex. The *F. oxysporum* ff. spp. are 96-99% similar to each other.

Over 26k orthologous groups across all predicted protein-encoding genes were found in the set of analysed species. Over 3500 orthogroups are present in all strains, of which 1685 are single-copy genes. With the inclusion of additional species, and as in chapter two, the species tree based on the concordance of 3544 orthogroups supports the established genetic relationships in the *Fusarium* genus (Figure 3.3b). The inclusion of additional coffea strains returns the coffea population in two separate clades, with the robusta strains monophyletic with one of the coffea clades. The genealogies of the single-copy orthogroups are strongly congruent for the relationships within the *F. fujikuroi* complex species and the outgroups (represented by clear lines, Figure 3.4). There is little concordance between the gene genealogies within the *F. xylarioides* clade, nor within the *F. oxysporum* ff. spp. clade. In addition, the tree topologies reveal shared genealogies across the genus between

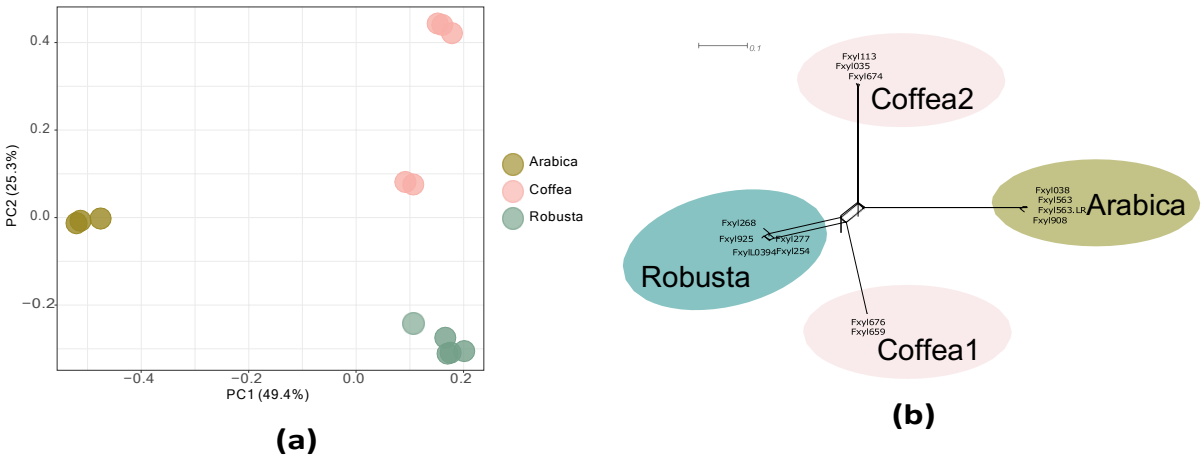


Figure 3.5: Genetic clusters in *Fusarium xylarioides*. A Principal component analysis based on 239,944 single nucleotide polymorphism (SNP) in 13 *Fusarium xylarioides* strains. The points are shaded according to their host population. **B** A neighbour-net (SplitsTree) analysis based on a single nucleotide polymorphism distance matrix. The scale bar represents 0.01 substitutions per site for branchlengths. The points are shaded according to their host population.

the gene trees shared by the *F. oxypsorum* ff. spp. clade and the *Fusarium fujikuroi* species complex, including *F. xylarioides*, clade.

3.5.4 *Fusarium xylarioides* displays genetic subdivision

Within *F. xylarioides*, 239,944 single nucleotide polymorphisms (SNPs) were identified. The maximum-likelihood tree (Figure 3.3b), principal component analysis (Figure 3.5a and neighbour-net (SplitsTree) (Figure 3.5b) analyses based on the SNPs identified the same four genetic clusters. These clusters correspond to an arabica group, a robusta group and two groups among the coffea strains. As in the densitree (Figure 3.4), the coffea group containing coffea659 and coffea676 is the most closely related to the robusta group. The absence of reticulation in the arabica and coffea branches are consistent with a lack of recombination, whilst this analysis reveals more diversity within the robusta clade than previously thought (Figre 3.5b, chapter two). Genetic differentiation between all populations was further confirmed by the high fixation index (F_{ST}) and absolute divergence (d_{xy}) values, with (d_{xy}) lower between robusta and both coffeas (Table 3.4). Therefore, I report the presence of four genetically differentiated populations within *F. xylarioides*: an arabica population and a robusta population, and two populations among the coffea strains from the first (1920s-1950s) outbreak.

Nucleotide diversity (d_{xy})	Arabica	Robusta	Coffea1
Robusta	2.18E-03		
Coffea1	2.30E-03	1.78E-03	
Coffea2	2.29E-03	1.36E-03	1.71E-03

Fixation index (F_{ST})	Arabica	Robusta	Coffea1
Robusta	0.972		
Coffea1	0.963	0.982	
Coffea2	0.956	0.963	0.957

Table 3.4: Population genetics statistics related to genetic differentiation (F_{ST}, d_{xy}) in three *Fusarium xylarioides* populations. Measured in 100 kb windows.

3.5.5 The *Fusarium xylarioides* populations differ in transposable element content

In terms of total TE content, the *F. xylarioides* coffea population has significantly fewer retro- and DNA transposons (5.4 Mb) than arabica (6.7 Mb) and robusta (6.3 Mb) *F. xylarioides* populations (Wilcoxon rank sum exact test, p=0.05 coffea-arabica, p=0.048 coffea-robusta) (Figure 3.6). This difference is primarily driven by *Ty3* (formerly known as *Gypsy*, see Wei et al. (2022) long-terminal repeats (LTRs) in the retrotransposon class.

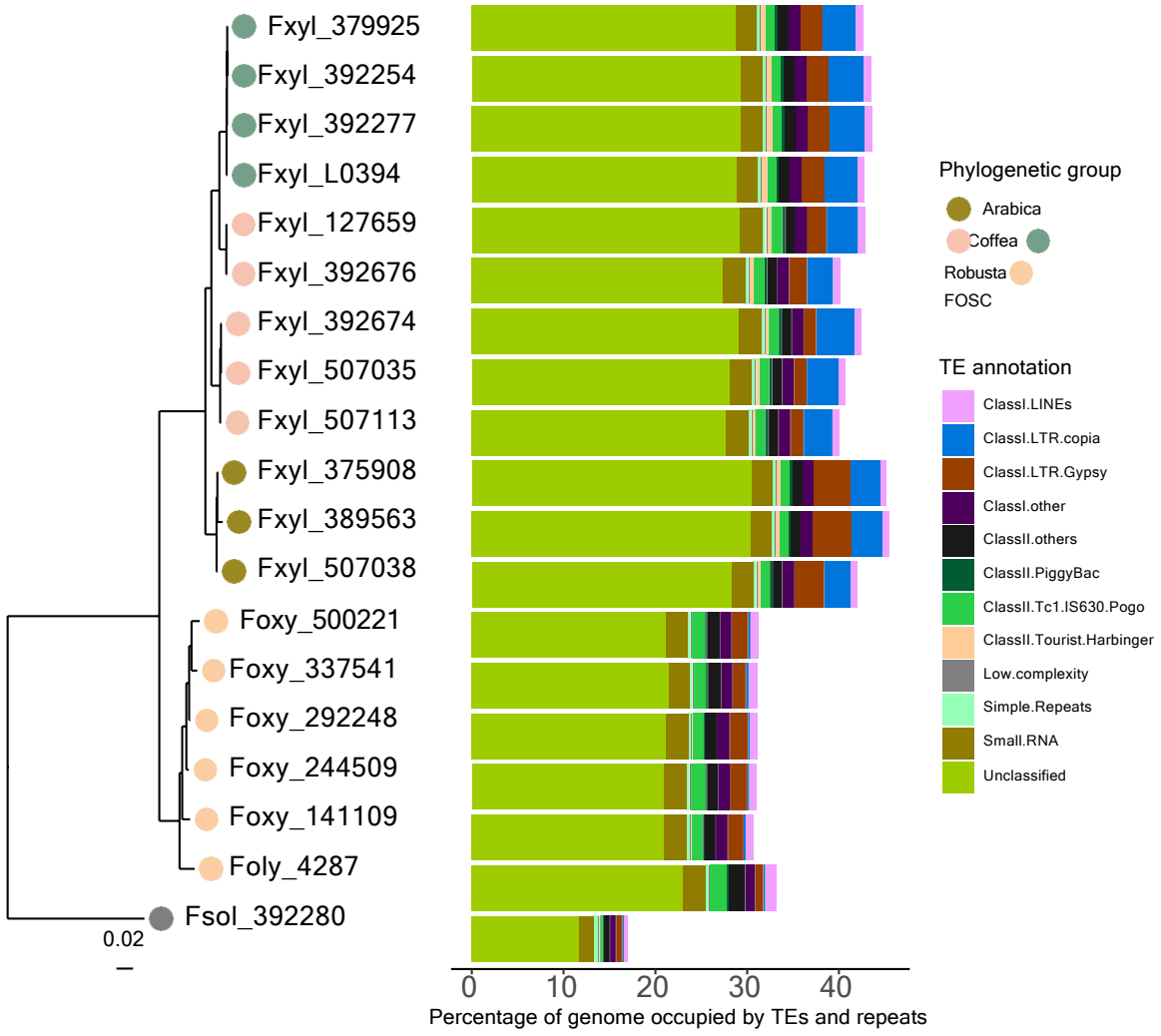


Figure 3.6: Transposable element composition of *Fusarium* genomes. Transposable elements (TE) across the whole-genome are shown as percentage of occupied base pairs across the *Fusarium* phylogeny. Each colour represents a transposable element category. Tree tip labels are shaded by phylogenetic group: the arabica, robusta and coffea *Fusarium xylarioides* strains; FOSC, *Fusarium oxysporum* species complex. Full strain details in Tables A.9 and A.10.

Indeed *Ty3* content is significantly different between the arabica and the robusta genomes (arabica = 2.2 Mb, robusta = 1.4 Mb, Wilcoxon rank sum exact test, $p<0.05$), while they are broadly similar in total TE numbers (6.7 Mb in arabica, 6.3 Mb in robusta). There are also significant differences in DNA transposon content between the arabica and robusta genomes, with arabica enriched in Piggybac (arabica = 180 kb, robusta = 165 kb, Kruskal-Wallis rank sum test, $p<0.05$) and robusta enriched in Tourist/ Harbinger (arabica = 221 kb, robusta = 318 kb, Kruskal-Wallis rank sum test, $p<0.05$).

When comparing the different species, *F. xylarioides* genomes have the most base pairs occupied by TEs, with 11% content, compared with 7% in *F. oxysporum* and 4% in *F. solani*. The *F. oxysporum* genomes have similar total content of class II DNA transposons as the *F. xylarioides* genomes and even more long-interspersed nuclear elements (LINEs) than *F. xylarioides* or *F. solani* (Wilcoxon rank sum exact test, $p<0.01$). The difference between the species is, again, driven by LTRs and specifically the *Ty1/Copia* class (*F. xylarioides* = 1.9 Mb, *F. oxysporum* = 122 kb, Wilcoxon rank sum exact test, $p<0.01$).

3.5.6 Expansion of *miniature impala* elements in *Fusarium oxysporum* and *Fusarium xylarioides*

Across the *F. xylarioides* and *F. oxysporum* genomes sequenced, 450 full-length *mimps* were detected (Figure 3.7). No *mimps* were found in the genomes of the *F. fujikuroi* species complex. The *F. oxysporum* genomes contained the most *mimps*, with *F. o. lycopersici* and *F. o. raphani* containing particularly high copy numbers (Figure 3.7). The *mimp* families 1, 2 and 4 are the most numerous in all genomes. Within *F. xylarioides*, the strains showed similar numbers within their host-specific groupings and over twice as many in the robusta strains than in the arabica strains, a difference primarily driven by *mimp* families 1 and 2.

In addition to the presence of *mimps*, all 13 *F. xylarioides* genomes and most of the *F. oxysporum* genomes contain intact (that is, not interrupted by a stop codon) *impala* transposases (Figure 3.7). Eight publicly available *F. oxysporum* genomes also contain at

least one *impala* transposase, whereas none of the genomes from the *F. fujikuroi* complex species do. Intact *impala* ORFs were also found in the gene annotations from chapter two in the six initial *F. xylarioides* genomes sequenced. Therefore, both the *miniature imapala* elements as well as intact *impala* class II transposon are found across the *F. xylarioides* and *F. oxysporum* genomes and are absent in the remainder of the *Fusarium*

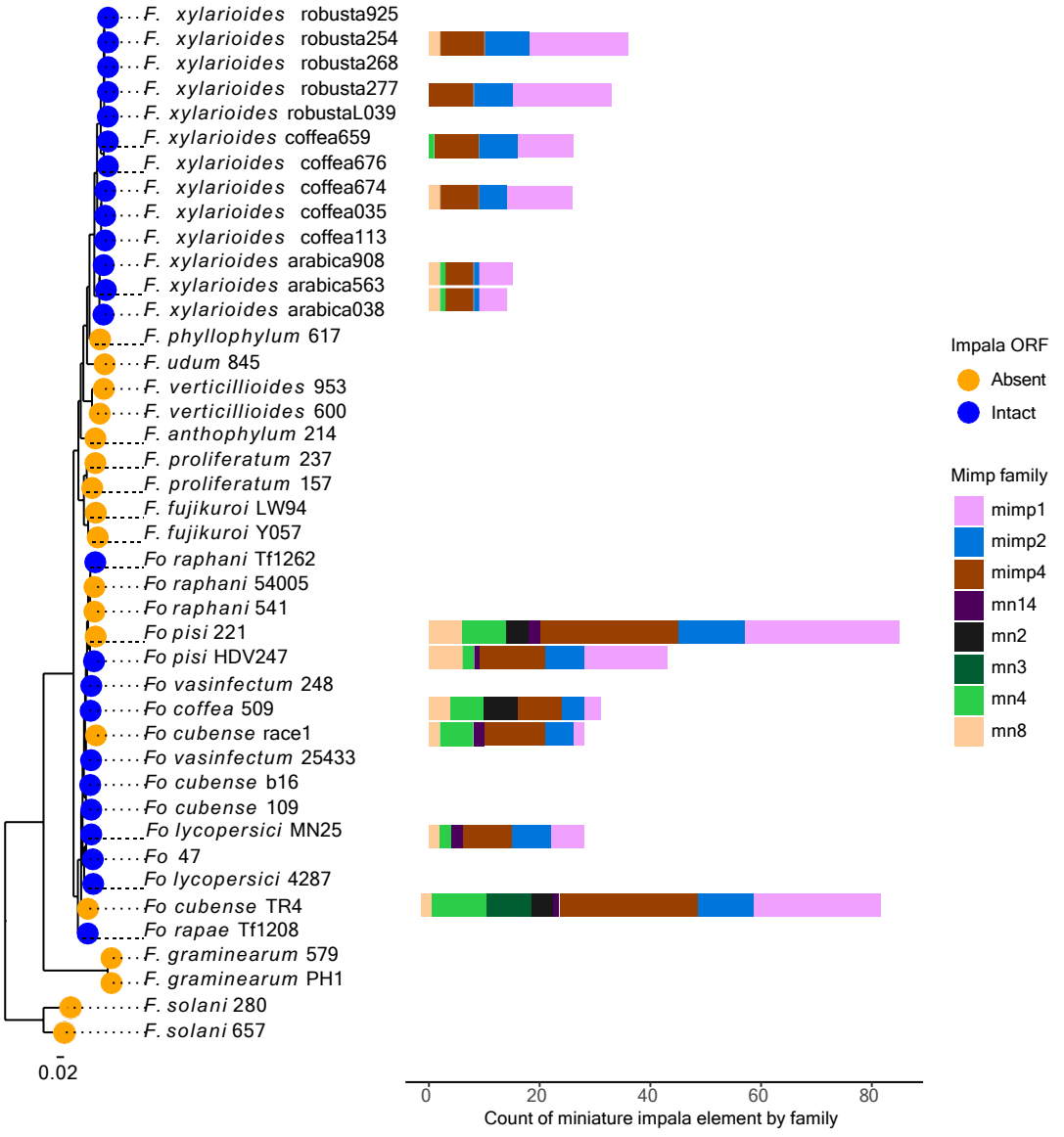


Figure 3.7: ***Miniature impala* composition of *Fusarium* genomes.** *Miniature impala (mimp* elements across the whole-genome are shown as a count per family. They were identified in six representative *Fusarium xylarioides* genomes and each *Fusarium oxysporum* f. sp. genome sequenced in this chapter. Each colour represents the family to which the *mimp* sequence belongs. The species tree is created using the same methods as Figure 3.3b and the tip label corresponds to the presence of an intact *impala* transposase open reading frame.

fujikuroi complex species.

To investigate whether the *mimps* presence in only the two species *F. xylarioides* and *F. oxysporum* could be a result of horizontal transfers, their nucleotide sequences

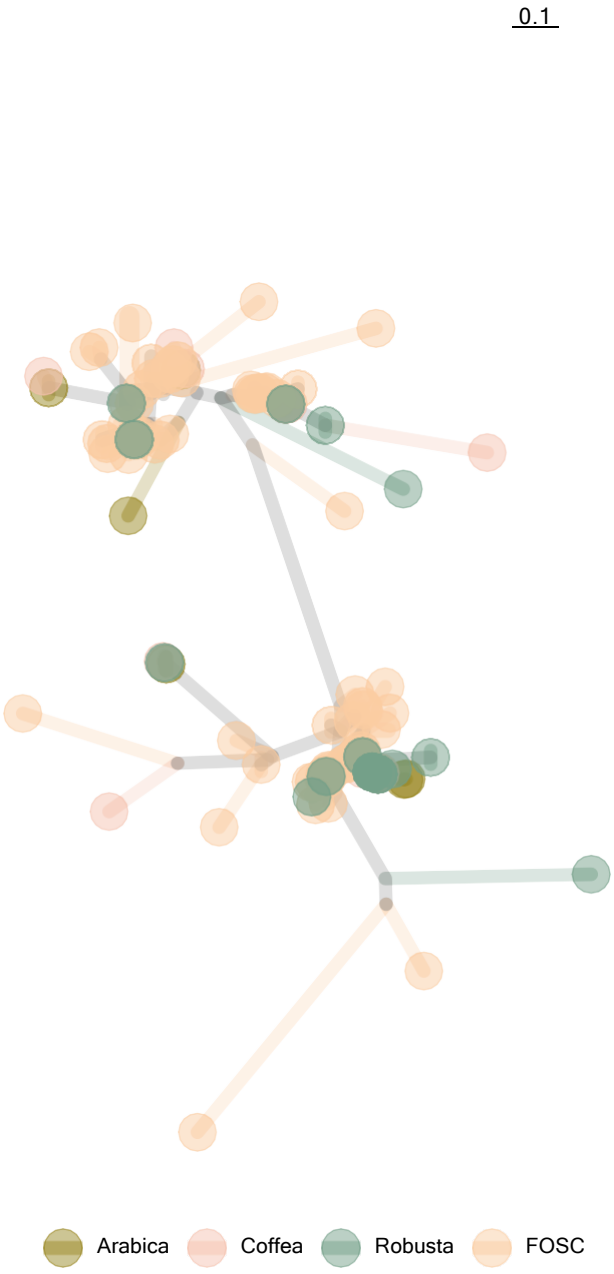


Figure 3.8: ***Fusarium oxysporum* and *Fusarium xylarioides miniature impala (mimp)* family 1 sequence genealogy**. All *mimp* 1 sequences have been extracted and aligned with macse (Ranwez et al. 2018), with the phylogeny represented by a daylight (unrooted) tree layout. Each point corresponds to a *mimp* sequence and the treescale shows their evolutionary distance. Tree tip labels are shaded by phylogenetic group: the arabica, robusta and coffea *F. xylarioides* strains; FOSC, *Fusarium oxysporum* species complex. Full strain details in Table A.9.

were compared with those in other genomes: the *F. o. lycopersici* genome has the most diverse *mimp* content, with various members of the recently described and highly diverse families (starting with 'mn'). These 'mn' families are not monophyletic themselves or with the other *mimp* families, and so sequences from the main *mimp* families were aligned. All *mimp* sequences in families 1, 2 and 4 form distinct clades but *F. xylarioides* and *F. oxysporum mimp* copies scattered throughout all clades, and arabica and robusta *F. xylarioides* copies sharing separate branches with various *F. oxysporum* ff. spp. (Figure 3.8).

3.5.7 Multiple large and gene-rich regions are shared between *Fusarium xylaroides* and different *Fusarium oxysporum* ff. spp.

I identified five regions shared between *F. xylarioides* and the *F. oxysporum* genomes sequenced in this chapter, despite being absent from the other genomes belonging to the *F. fujikuroi* species complex. These five regions are all over 40 kb, have a gene density >0.05 (the average is <0.03 per 100 kb) (Table 3.5) and are differentially present across the *F. xylarioides* populations (arabica, robusta and two coffeas). In the *F. o. lycopersici* and *F. o. cubense* genomes, each region of high coverage in *F. xylarioides* maps to a single chromosome (Table 3.5). The genomes of *F. o. raphani, F. o. pisi* and *F. o. vasinfectum* are too fragmented to analyse the presence of these regions (Tables A.11). I analysed separately the regions shared by *F. o. lycopersici* and *F. o. cubense* with the *F. xylarioides* arabica population on the one hand and those shared by *F. o. lycopersici* and the robusta population on the other hand. Despite high-quality assembly in *F. o. coffea* and the fact that this *F. oxysporum* strain was isolated from coffee and therefore was a strong candidate for horizontal transfers, there were no similar, gene-rich regions shared between this strain and any *F. xylarioides* strain from either population. There were also no shared high similarity regions between *F. xylarioides* and *F. solani* (Table

3.5), eliminating *F. solani* 392280 isolated from a coffee wilt infected tree as a possible source of pathogenicity by horizontal transfer.

Contig 13 in the *F. xylarioides* arabica strains appears to have been partially acquired by horizontal transfer. One half is unique to the arabica population of *F. xylarioides* and this part is highly similar to *F. o. lycopersici* and *F. o. cubense*, sharing a number

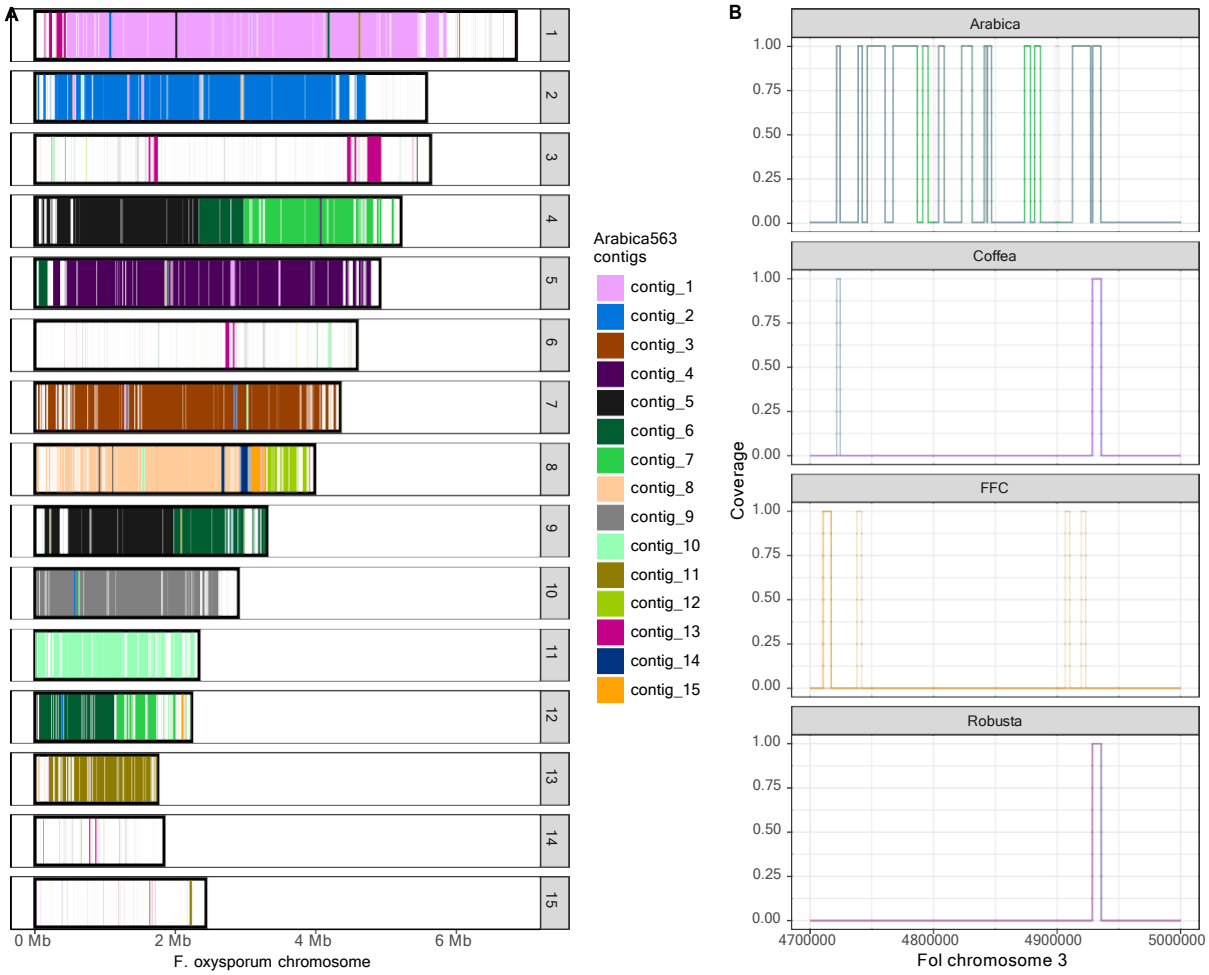


Figure 3.9: A whole-genome alignment between the *Fusarium xylarioides* arabica563 reference assembly and the *Fusarium oxysporum* f. sp. *lycopersici* chromosomal assembly. **A** Each bar represents one of the 15 *F. oxysporum* chromosomes, with chromosome size and contig position on the x axis. Chromosome numbers are annotated in-line with the bars. Shaded regions indicate syntenic *Fusarium xylarioides* arabica563 contigs, the colour of which identifies the specific mapped contig from *Fusarium xylarioides* arabica563. Unshaded regions indicate contigs found only in *Fusarium oxysporum* and which are are absent in *Fusarium xylarioides* arabica563. The *Fusarium oxysporum* chromosomes 1-15 correspond to NC_030986.1-NC_030100.1 in the genome accession GCF_000149955.1. **B** Assembly mapping of the arabica, robusta and coffea *Fusarium xylarioides* populations, and the *Fusarium fujikuroi* complex (FFC) genomes to *Fusarium oxysporum* f. sp. *lycopersici (F. o. lycopersici*. The *Fusarium xylarioides* populations are grouped by their host-specificity with the coffea strains shown together. This figure shows a region between 4.7-5 Mb of chromosome 3 in *F. o. lycopersici*. The x axis represents the genome coordinates and the y axis represents the coverage compared with *F. o. lycopersici*.

Table 3.5: Length and gene number in regions which are shared between the *Fusarium xylarioides* populations and various *Fusarium oxysporum* ff. spp. but absent from the genomes of the *Fusarium fujikuroi* complex species. *F. oxysporum* 47 is a non-pathogenic isolate. The coffea populations are shown together. Cells are shaded according to their value with red representing higher values. Full strain details in Tables A.9 and A.10). Abbreviations: *Fo* = *Fusarium oxysporum* f. sp.

	Length (bp)			
	Arabica-specific	**Arabica & coffea**	**Robusta-specific**	**Robusta & Coffea**
F. o. 47	25,400	7,750	2,450	16,850
F. o. pisi	25,200	14,350	20,300	48,450
F. o. raphani	59,700	11,300	1,050	25,650
F. o. lycopersici	57,200	2,500	2,450	48,750
F. o. cubense	55,700	9,800	4,050	24,750
F. o. coffea	13,800	6,150	4,000	30,250
F. o. vasinfectum	**34,202**	**15,000**	**5,500**	**9,250**
F. solani 280	-	-	-	-
	Genes (n)			
	Arabica-specific	**Arabica & coffea**	**Robusta-specific**	**Robusta & Coffea**
F. o. 47	18	6	2	15
F. o. pisi	18	16	10	44
F. o. raphani	29	10	2	13
F. o. lycopersici	36	2	2	40
F. o. cubense	25	11	4	17
F. o. coffea	13	6	4	33
F. o. vasinfectum	**21**	**18**	**5**	**8**
F. solani 280	-	-	-	-

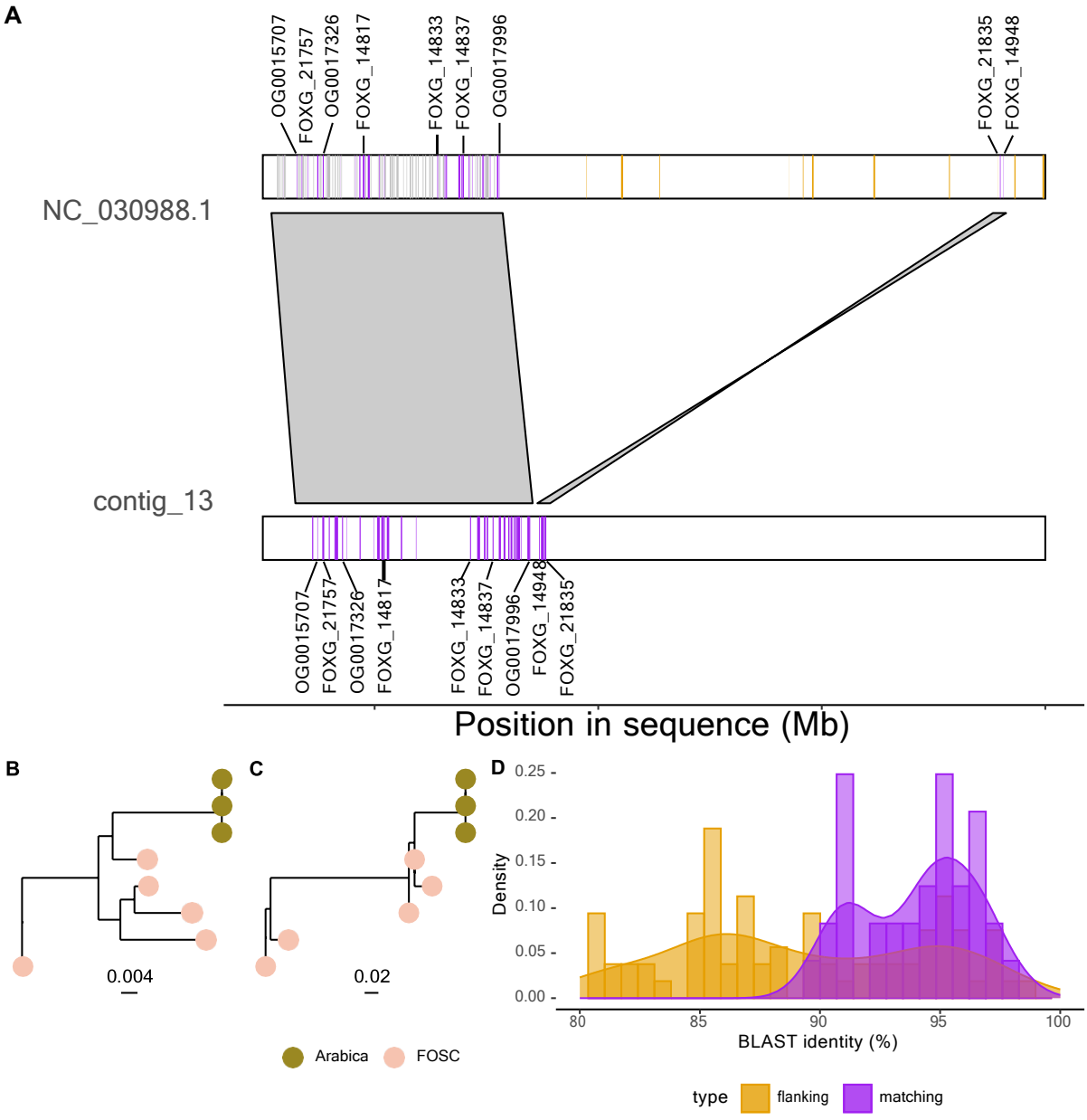


Figure 3.10: A large, gene-rich region is shared between *Fusarium oxysporum* f. sp. *lycopersici* and *Fusarium xylarioides* arabica563. A Two regions are shared, spanning 4.7-4.9 and 5.37-5.38 kb, represented by the shaded grey areas. Highly similar genes shared by the two species are shaded in purple with a subset labelled with the *Fusarium oxysporum* gene locus tag. Non-matching genes found in the shared region are shaded grey. The genes in the flanking regions with a match in *Fusarium xylarioides* (BLAST length >80%) are shaded yellow. Overall this gene class has a lower sequence identity and none of the *Fusarium xylarioides* copies are found on contig 13. **B** A rooted gene tree for the orthologous group OG0017326, found in the shared region. The tree is created in and resolved using the OrthoFinder hybrid species-overlap/duplication-loss coalescent model (Emms & Kelly 2015). **C** Gene tree for the orthologous group OG0017996, found in the shared region. Drawn in the same way as **B**. **D** The genes shared between the two species in the matching region have a higher BLAST sequence identity (mean 94%, dotted line) than those found in the flanking regions either side (mean 88%, dotted line).

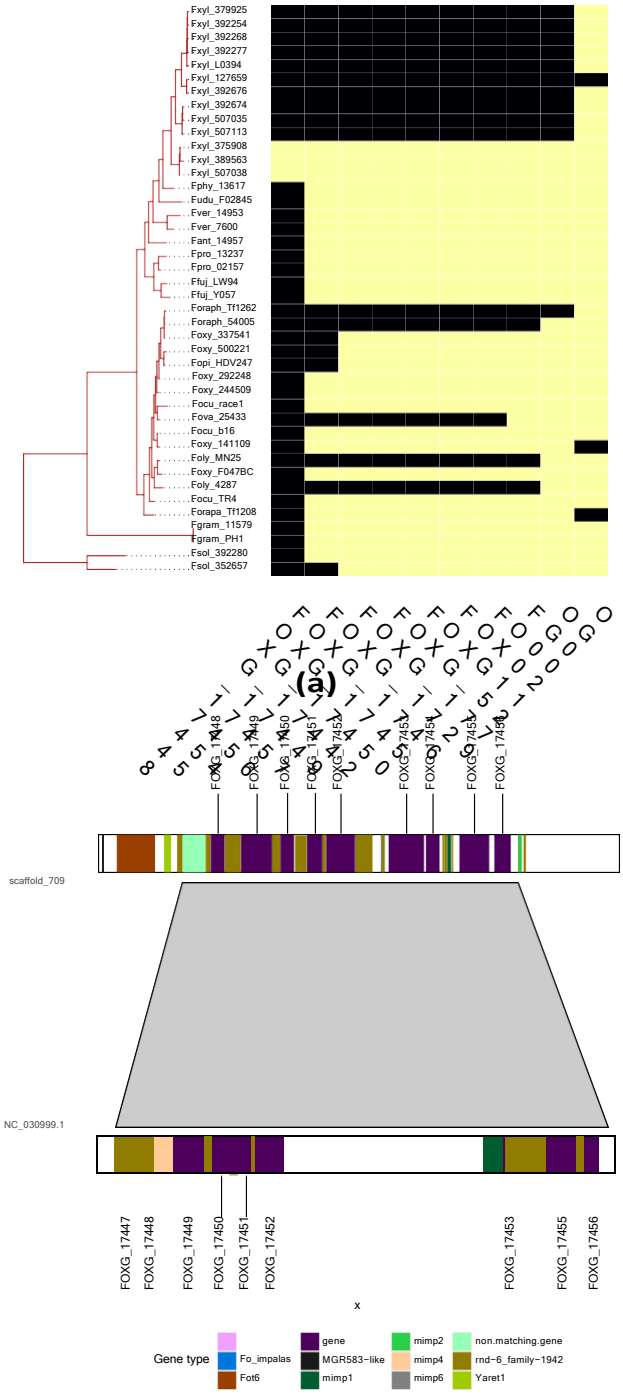


Figure 3.11: A large, gene-rich region is shared between *Fusarium oxysporum* f. sp. *lycopersici* and *Fusarium xylarioides* robusta254. A Heatmap showing presence of orthologous groups found in the shared region, compared with *Fusarium* phylogeny. Present genes are shaded in dark and all share >96% identity. The phylogenetic species tree was created following Figure 3.3b. Full strain details in Tables A.9 and A.10. **B** A highly similar region in *Fusarium oxysporum* f. sp. *lycopersici* chromosome 14 (NC_030999.1, bottom plot) matches to the *F. xylarioides* robusta scaffold 709 (top plot). Genes shared by the two species are shaded according to the legend and labelled with the *Fusarium oxysporum* gene locus tag. Repeats of interest are shaded according to the legend, Fo_impala indicates a fragment of an *impala* transposase. Each shared gene in this region also shares an orthologous group with *Fusarium oxysporum*. Drawn to scale

of highly-similar genes with *F. o. lycopersici*. It matches to different regions in the *F. o. lycopersici* and *F. o. cubense* genomes. This region of contig 13 is absent from the robusta genomes (as also shown in Figure 3.2b) and does not contain any *mimps*. A large and gene-rich region is shared between the mobile chromosome 3 (NC_030988.1) in *F. o. lycopersici* and contig 13 in the arabica *F. xylarioides* population (Figure 3.9a-b). Whilst this region is broadly absent, flanking regions are present in the coffea and robusta *F. xylarioides* genomes and in the *Fusarium fujikuroi* complex genomes (Figure 3.9b). Within this shared region, there are 77 genes on chromosome 3, of which 39 are also present in the corresponding region in contig 13 (Figure 3.10a), with a high sequence identity between the *F. o. lycopersici* and *F. xylarioides* arabica strains (BLAST, >94% identity) compared with a lower identity in both the flanking genes (BLAST, 88% identity, Figure 3.10d) and the whole-genome identity (90%, Figure 3.3a). Two of these genes belong to orthologous groups across the two species (Figure 3.10b-c), and both gene trees reveal the *F. oxysporum* copies are diverse whilst the arabica copies are very similar. A synteny alignment of the two genomes reveals that contig 13 matches to most of *F. o. lycopersici* 's mobile chromosomes: 3, as well as 1, 6, 14 and 15 in smaller pieces (Figure 3.9). This same region of contig 13 also maps to chromosome 10 (SRMI01000010.1) in *F. o. cubense*. Within this region of chromosome 10, there is a sub-region where three genes in *F. o. cubense* match to five genes in the arabica strains (BLAST, 96% identity). One of these matching genes appears to have been duplicated in the arabica strains: TVY62576.1 has three copies on contig 13 in arabica563 (P1J47_002175-T1, P1J47_002174-T1, P1J47_002173-T1). Unlike the match to *F. o. lycopersici*, none of the other genes in this region are present in contig 13. This highly-similar region which is present in the arabica genomes appears to have been acquired by horizontal transfer from a mobile *F. oxysporum* chromosome, although the absence of *mimps* suggests a different mechanism to the *mimp*-assisted transposition suggested in *F. oxysporum*.

The robusta and coffea *F. xylarioides* genomes appear to have acquired a different region from *F. oxysporum* than in the arabica population. This region matches to a different

chromosome and is enriched in *mimps*. This region is absent from the arabica strains and so it is not part of the long-read assembly, but syntenic 25 kb scaffolds are shared by all robusta and coffea *F. xylarioides* strains. This matching region is on chromosome 14, and contributes to 'supercontig 51', an effector-, gene- and transposable element-rich mini-cluster on the mobile pathogenic chromosome in *F. o. lycopersici*. The shared scaffold in the robusta strains contains 9 genes in each genome, all of which match to *F. o. lycopersici* gene copies with a high sequence identity (BLAST, 98%, Figure 3.11a), compared with a 90% whole-genome identity (Figure 3.3a). Interestingly, one of the genes in the shared scaffold (OG0015726) does not match *F. o. lycopersici*, but matches one of the *F. o. raphani* strains. This could indicate *F. o. raphani* as the donor *forma speciales* instead of *F. o. lycopersici*. In addition, the robusta *F. xylarioides* scaffolds contain 2-3 *mimps*, whilst the coffea *F. xylarioides* scaffolds lack these. This scaffold is also enriched for the unnamed TE family rnd-6 family-1942, as well as newly-recognised TEs found on *F. o. lycopersici* 's mobile pathogenic chromosome which flank the shared region(MGR583-like, Fot6, Yaret1, Schmidt et al. (2013)) (Figure 3.11b). In the robusta *F. xylarioides* scaffold, the shared region is flanked by highly-similar copies of the *F. oxysporum* Fot6 element, MGR583-like, Yaret1 and a *mimp 2* mobile elements. The shared and highly-similar *F. oxypsorum* genes (Figure 3.11a) are not *SIX* effectors, but other types of genes, including a glycosyltransferase, Cytochrome P450 monooxygenases, a squalene-hopene-cyclase, a methyltransferase-UbiE protein and a Tri7 homolog. Therefore, the *F. xylarioides* robusta population contains a region which is known to be acquired by horizontal transfer in various *F. oxysporum* ff. spp., possibly via *mimps*. This region contains highly-similar gene copies in both species, as well as various TEs which were identified in *F. oxysporum*.

3.6 Discussion

In order to advance the results of chapter two, a reference genome assembly was sequenced for the *F. xylarioides* arabica563 strain to investigate whether genomic regions that are

shared and highly similar between *F. xylarioides* and *F. oxysporum* have been acquired by horizontal transfer. I show that both the arabica and robusta *F. xylarioides* populations share various highly similar regions with *F. oxysporum* that are lacking in other closely related species, consistent with acquisition by horizontal transfer. The regions shared between *F. oxysporum* and the *F. xylarioides* robusta population are also found in the *F. xylarioides* coffea genomes. Furthermore, these data revealed genetic subdivision in *F. xylarioides*, with multiple differentiated populations occurring in different species of coffee. These results indicated that horizontal gene transfer occurred repeatedly in the *F. xylarioides* clade and could be one of the drivers behind the repeated emergence of coffee wilt disease over the last century through the transfer of genomic regions rich in genes. The inclusion of additional genomes and a reference assembly in a study of historical strains helped to elucidate the evolutionary history of *F. xylarioides*.

The *F. xylarioides* arabica and robusta groups are genetically differentiated populations, and differentiated from the coffea group, which is actually composed itself of two differentiated populations. One of the coffea populations, coffea1, is more closely related to the robusta population. It is interesting that the ranges of the two coffea populations overlapped spatially and temporally (coffea 1 was isolated from Guinea (although this is recorded as Guinea? in the CABI-IMI collection) and Central African Republic, coffea2 was isolated from Guinea, the Central African Republic, and Cote d'Ivoire), yet they remained genetically distinct with the robusta population apparently emerging from within a wider recombining population containing coffea659 and coffea676. These analyses also revealed more genetic divergence within robusta than was observed in whole-genome analyses or reported in chapter two, highlighting the importance of population genetic analyses which include non-coding regions. Possible causes of genetic differentiation between populations could be clonal lineages, which is suggested by low reticulation and lack of recombination between populations. Or reproductive isolation due to pleiotropy between host adaptation and mate choice, in which only strains able to infect the same trees can mate (Giraud et al. 2010) could have driven genetic differentiation between geo-

graphically overlapping populations. Specialisation and reproductive isolation can be the result of pleiotropy if mating occurs within the host, which appears to occur within *F. xylarioides*, as the presence of perithecia on coffee wilt diseased trees indicates (Rutherford et al. 2009). It is an unfortunate aspect of working with historic material that often the host or geographic details were not fully recorded.

Population-specific traits are visible at the genome level, as well as in TE content, where larger arabica genomes show expansion of long-terminal repeat (LTR) TEs. In fact, retrotransposons are often responsible for genome expansion in fungal species, for example Spanu et al. (2010). These genetically differentiated populations support the earlier work of Lepoint (2006) and Girma (2004), which suggested that *F. xylarioides* is a species complex containing distinct arabica and robusta groups, which they called, respectively, *F. xylarioides* f. sp. *abyssiniae* and *F. xylarioides* f. sp. *canephorae*. The robusta population was probably present in the initial (1920s to 1950s) outbreak, with the arabica population presumably also in Ethiopia but at low incidence levels and therefore went undetected. These data refute the hypothesis (discussed in Geiser et al. (2005), Buddie et al. (2015)) that either population emerged after the second (1970s onward) outbreak, because both well-resolved populations contain strains isolated before this, namely: arabica038, isolated in 1971; and robusta268, isolated in 1968. *Fusarium oxysporum* is known as a species complex composed of many different asexual lineages (Ma et al. 2010). Considering that *F. xylarioides*, whilst currently classified as a single species, contains genetically distant populations with similar incongruence between gene trees to those within the *F. oxysporum* species complex, perhaps the *F. xylarioides* formae speciales proposed nearly two decades ago could now be recognised.

Fusarium xylarioides populations contain shared regions from the mobile chromosomes of *F. oxysporum*, indicating multiple horizontal transfers. Several lines of evidence support the acquisition of these regions by horizontal transfer over alternative hypotheses. First, the hypothesis of its presence in the common ancestor and vertical inheritance is contradicted by a high sequence identity greater than 40 kb in *F. xylarioides* arabica and

20 kb in *F. xylarioides* robusta populations. The genome sequence identities between *F. oxysporum* and *F. xylarioides* are otherwise about 90%, and these regions are absent from all other *Fusarium fujikuroi* complex species. Furthermore, the non-homologous locations between the arabica and robusta regions, occupying in each case regions of the genomes which are absent from the other, rule out interbreeding and introgression. However, within each population the locations are homologous between strains, suggesting transfers into a recent common ancestor for each population. The regions appear to have been transferred from *F. oxysporum*, where there is a lower homology between *formae speciales*, to *F. xylarioides*, where each region is identical between strains. However, the incomplete identity (<100%) with *F. oxysporum* indicates that the *F. o. lycopersici* genome is not the specific donor strain. The occurrence of horizontal transfer of pathogenicity via accessory chromosomes was first reported between *F. o. lycopersici* and previously nonpathogenic *F. oxysporum* 47 (Ma et al. 2010). Several more cases have been reported since then, with a putative role for *mimps* in transposition (Schmidt et al. 2013, van Dam et al. 2016, van Dam & Rep 2017). However, this is the first report of transfers of such large genomic regions to a different species. To further verify this finding, additional *F. xylarioides* strains could be sequenced. Furthermore, I show that transfers of large genomic regions from *F. oxysporum* to *F. xylarioides* occurred repeatedly, with different genomic regions acquired by arabica and robusta populations from ancestral regions in the *F. oxysporum* lineages. These transfers could have resulted in the evolution of different pathogenicity found in the host-specific populations and therefore could have resulted in the diseases emergence and subsequent re-emergence.

The *mimp* copies in *F. xylarioides* and *F. oxysporum* are highly similar, with the *F. xylarioides mimps* nested within the *F. oxysporum* copies, and are absent from other strains of the same species complex, suggesting their transfer from *F. oxysporum* to *F. xlyarioides*. No *mimps* were recorded in the *Fusarium fujikuroi* complex species, although van Dam & Rep (2017) found low numbers in some cases. The *mimp* copy number is significantly lower in the *F. xylarioides* arabica population, and indeed there are no

mimps present in the horizontally transferred region it shares with *F. oxysporum*, unlike in the *F. xylarioides* robusta region. This could be because, by chance, no *mimps* were present in the transferred region and therefore fewer *mimps* were able to multiply later. Indeed, it appears that the *F. xylarioides* copies could be active, because all genomes contain an ORF encoding an *impala* transposase (Bergemann et al. 2008). There is even a fragment of this ORF in the putative *F. xylarioides* robusta HGT region. In *F. oxysporum, mimps* always feature in the promoter region of the *SIX* effector genes, as well as other pathogenicity genes expressed throughout infection (Schmidt et al. 2013). Identifying genes in *F. xylarioides* with a *mimp* in the promoter region could reveal further putative effectors. Despite this co-localization, the exact role of *mimps* in the promoters of effector genes is not known (van Dam & Rep 2017, Schmidt et al. 2013) and the differential presence across the *F. xylarioides* populations and putative HGT regions could suggest different mechanisms of horizontal transfer.

The variable *mimp* copy number among the populations in general, and specifically in the horizontally transferred regions, suggests that they could have been acquired by different mechanisms and by chance if the HGT regions contained *mimps*. Horizontal transfers can occur through TE-assisted mechanisms or by transfer of an entire chromosome (Ma et al. 2010). *Fusarium* species are known to form anastomoses - somatic fusions between mycelia, which may facilitate the transfer of genetic material by whole chromosomes (Kurian et al. 2018). Indeed, hyphal fusions can occur between the arabica and robusta *F. xylarioides* populations (Lepoint et al. 2005). The region which robusta and coffea share with *F. o. lycopersici* chromosome 14 is part of supercontig 51, a region of the chromosome which is enriched in TEs, effector genes, and other types of pathogenicity genes (Ma et al. 2010), while the arabica region is part of mobile chromosome 3. Ma et al. (2010) reported the horizontal transfer of the entire mobile pathogenic chromosome to previously non-pathogenic strains, and also the horizontal transfer of a large duplicated region from chromosome 3 to two previously non-pathogenic strains. Therefore, in addition to transferring an entire chromosome, possibly by hyphal fusion, large regions on

other mobile chromosomes can undergo horizontal transfer, enhancing the pathogenicity of the recipient strain. In this instance, it is possible that either whole chromosomes were transferred, with regions later lost, or that specific intra-chromosomal regions underwent horizontal transfer. *Mimps* could have been involved in the transfer of the putative *F. xylarioides* robusta HGT, and by chance could have been absent from the arabica HGT, appearing in the *F. xylarioides* arabica population genomes by another HGT and multiplying, albeit at a lower copy number than in the *F. xylarioides* robusta or coffea genomes.

The presence of *mimps* as well as other *F. oxysporum*-specific mobile elements points towards transposition-mediated transfer or that they were present in the putative HGTS, later multiplying. This interspecies transfer of pathogenicity has not been reported in *Fusarium* before. Transfer could have occurred in a shared niche, since both *F. xylarioides* and *F. oxysporum* have been isolated from the roots and wood of coffee wilt-diseased trees. Furthermore, both species have been co-isolated from banana roots, which farmers frequently intercrop with coffee (Serani et al. 2007) and on which *F. oxysporum* f. sp. *cubense* is widespread (O'Donnell, Kistler, Cigelnik & Ploetz 1998). In addition to banana, *F. xylarioides* has also been isolated from rotting tomato fruits in Nigeria (Onesirosan & Fatunla 1976) and cotton seedlings (Pizzinatto & Menten 1991). Many of the weed species surrounding coffee farms are *Solanaceous* weeds that could potentially be infected by *F. o. lycopersici*, such as tomato-specific *forma speciales*, and many cotton farms are also found in coffee growing regions (Phiri & Baker 2009). In the 1990s-2000s, there were reports of an outbreak of potential alternate hosts for *F. xylarioides* in the many weed species that grow abundantly in coffee-growing areas (Phiri & Baker 2009). Although the presence of *F. xylarioides* was assessed and shown to be absent from 105 different weed and crop species (Kangire et al. 2002), it remains to be questioned whether there are alternative hosts in which the pathogen can exist without causing disease symptoms. Therefore, several plants surround coffee fields on which various *F. oxysporum* ff. spp. could live and interact with *F. xylarioides*.

Chapter 4

Transcriptome profiling of the *Fusarium* coffee wilt pathogen reveals fungal vascular wilt genes, including xylem-associated effectors from *Fusarium oxysporum*

4.1 Abstract

This chapter investigates whether any genes highlighted in the previous chapters were expressed in arabica *Fusarium xylarioides* strains infecting the arabica coffee host. RNA sequence data reveal a major up-regulation of pectin lyase enzymes, suggesting a key role for them in the onset of vascular wilt infection. It also revealed up-regulation of three *F. oxysporum SIX*, that is secreted in xylem, effectors, which are all less than 1 kilobase from *miniature impala* transposons. This, combined with the close phylogenetic relationship with *F. oxysporum* for a subset of several hundred up-regulated genes, highlights the potential role of horizontal gene transfers in the spread of pathogenicity between *F.*

oxysporum and *F. xylarioides* in shared vascular wilt infections.

4.2 Introduction

Over the past 100 years, repeated outbreaks of coffee wilt disease, caused by *Fusarium xylarioides*, have changed the landscape of coffee production in sub-Saharan Africa (Hakiza et al. 2009, Girma et al. 2009). The initial outbreak in the first half of the 20th century corresponded to a disease affecting multiple *Coffea* species; while two host-specific *F. xylarioides* groups later emerged that separately cause disease in arabica and robusta coffee, the mainstays of the African coffee industry (Phiri & Baker 2009). The recent re-emergence of coffee wilt disease that affected robusta coffee in the 1990s-2000s caused over US$1 billion in losses to the Democratic Republic of the Congo and Uganda (Phiri & Baker 2009). Understanding the mechanisms of pathogenicity is essential for the implementation of effective management strategies that can regulate infection and help reduce the potential for future re-emergence and crop losses.

In recent years, RNA sequencing technology has developed to enable the analysis of differences in gene expression. Numerous studies have shown that complex transcriptomic and metabolic changes occur in both pathogen and host upon infection (Ferreira et al. 2007). In plant hosts, one of the first stages of recognition of pathogen attack and the triggering of a defence response is an 'oxidative burst' through the rapid production of reactive oxygen species (Torres et al. 2006), and a 'nitric oxide burst' (Leitner et al. 2009). The host plant recognises carbohydrate structures of an invading pathogen, such as chitin, or even a fragment of its own metabolised plant cell wall through cell surface receptors, using plant lectin proteins as a large part of that initial response (Lannoo & Van Damme 2014). This triggers reactive oxygen species and nitric oxide bursts, as well as a range of downstream signalling. However, some pathogens have evolved effector genes that are secreted into their host to evade detection. Plants, in turn, can evolve to detect these effector genes, which is termed effector-triggered immunity (Jones & Dangl 2006). In

this way, plants and pathogens have co-evolved for millennia. Modern farming practices such as monoculture systems, have inadvertently accelerated this co-evolution between plant and pathogen (Burdon & Thrall 2008). Using deep RNA-seq technology, one can investigate these mechanisms of pathogenicity, in the pursuit of a sustainable solution to repeated disease outbreaks across agricultural systems.

In this chapter, *F. xylarioides* arabica strains (with this group hereafter called 'arabica') isolated in the 1990s-2000s were used to investigate both the mechanism of coffee wilt infection and host-specificity to arabica coffee. Currently, disease levels in areas affected by coffee wilt in eastern and central Africa are much lower and more manageable than they once were. However, any future re-emergence would have catastrophic impacts on African coffee production, farmers and national economies (Phiri & Baker 2009). Furthermore, there is a potential threat to Asian and American coffee production, as the coffee germplasm used in these production areas is susceptible to coffee wilt infection (Phiri et al. 2009). The causal agent *F. xylarioides* is a member of the monophyletic *Fusarium fujikuroi* complex (Geiser et al. 2005). This species complex contains around 50 distinct species, many of which cause disease on agronomically important crops (Kvas et al. 2009). Of these distinct species, only two cause vascular wilt disease in plants: *F. xylarioides* and *F. udum*, which share a recent common ancestor.

Vascular wilts have a unique disease syndrome, characterised by symptoms of wilting, senescence, and eventual death of distal plant parts (Agrios 2005). Four fungal genera (*Fusarium, Verticillicium, Ceratocystis* and *Ophiostoma*) are known vascular wilt pathogens. Fungal wilt pathogens penetrate and enter the vascular system of their host and remain there until the advanced stages of the disease (Mace et al. 1981). The plant xylem system is a nutrient-poor environment, with low concentrations of sugars, amino acids, and oxygen. Vascular wilt pathogens therefore use carbohydrate-active enzymes (CAZymes) (Cantarel et al. 2009) to hydrolyse nutrients from the plant cell wall to survive (Yadeta & Thomma 2013). Carbohydrates such as cellulose and pectin in plant cell walls provide their main source of carbon (Hervé et al. 2010). In fact, *Verticillium* and *Fusar-*

ium wilt fungi have enhanced pectinolytic abilities, with expanded polysaccharide lyase gene families that directly degrade pectin, as well as families that degrade polysaccharide lyase products (Klosterman et al. 2011). Chapter two described a set of 64 putative effector genes and the expansion of the genome of different groups of pectin-degrading enzymes in arabica and robusta strains of *F. xylarioides*. However, without gene expression analysis, it is difficult to verify the role of such enzymes or identify potential effector proteins important in the onset of *Fusarium* wilt.

The onset of vascular wilt infection requires many intercellular signaling pathways that are associated with highly coordinated changes in gene expression (Giri et al. 1998). Transcriptome profiling analyses can reveal such changes in gene expression patterns, such as spore germination in *F. graminearum* (Harris 2017), and in triggering vascular wilt disease in banana by *F. oxysporum* f. sp. *cubense* (Guo et al. 2014). Such analyses could reveal the genes involved in the onset and development of coffee wilt disease. Interesting candidates include pectate lyases and *SIX*, that is secreted in xylem, effector proteins that, together with their homologs, contribute to the pathogenicity of a wide range of *F. oxysporum* ff. spp. in triggering symptoms of vascular wilt (Williams et al. 2016). Furthermore, RNA-seq can reveal whether putative effectors identified in chapter two as present in host-specific arabica are indeed expressed during infection.

Very little information is available on the molecular events and changes in gene expression that occur during the progression of coffee wilt disease. We now know that the arabica and robusta groups are genetically differentiated populations and have been prior to the coffee wilt disease outbreak of the 1970s onwards (chapter three). The goals of this chapter are to uncover the highly expressed genes in the infecting *F. xylarioides* strains as well as the coffee genes that respond to infection. Specifically, I ask: (1) whether the types of putative fungal effectors and carbohydrate-active enzymes described in chapter two and Peck et al. (2021) are up-regulated in infection; (3) whether highly expressed effector-like genes show evidence of horizontal transfer from *F. oxysporum*. In addition, I look for a new set of highly expressed effector-like proteins and report the functional an-

notation for all up-regulated coffee and *Fusarium* genes from infected coffee samples. To answer these questions, mixed transcriptomes were extracted from *C. arabica* infection by two *F. xylarioides* arabica strains and sequenced. Arabica plants and fungal strains were used rather than robusta due to the relative ease in procuring large numbers of arabica coffee plants. Illumina sequencing enabled the analysis of differential gene expression in both arabica strains and their host coffee plants, by comparing transcripts with recently sequenced *F. xylarioides* genomes (chapter three) and the publicly available coffee genome (Zimin et al. 2019). These methods reveal a list of significantly up-regulated pectin lyase enzymes and genes that behave like effectors, including some described in chapter two and three *Secreted In Xylem* effectors previously only described in *F. oxysporum*. This expression analysis provides new insights for further studies directed at understanding the progression of vascular wilt disease. These results will provide a better understanding of the epidemiology and host-specific nature of coffee wilt disease in arabica coffee, as well as identifying the expression of genes related to plant defense and fungal pathogenesis and their functionality.

4.3 Methods

4.3.1 Quantitative PCR pilot study

To find out which parts of infected coffee plants contain fungal RNA, coffee plants were infected and the expression of the housekeeping gene *Translation elongation factor (Tef)* compared to three putative effectors (*pelD, OG0013477, OG0014398*) described in chapter two. Arabica563 inoculum was inserted into a stem wound with a spore concentration of 1.00 x 10^6 spores/ml, following Girma et al. (2009). The plants were left until symptoms developed, after which the stem internode containing the wound site, the internode above and the leaves were all removed and flash frozen in liquid nitrogen. The samples were ground separately in liquid nitrogen with PVPP (polyvinyl polypyrrolidone) following (Rubio-Piña & Zapata-Pérez 2011) in a mortar and pestle. Total RNA was extracted

using the standard protocol of the RNeasy Mini kit (Qiagen, Germany). The quantity of RNA was determined using a NanoDrop 2000 (ThermoFisher) and RNAs with an optical density at 260/280 nm >1.8 were used for further work.

RNA was converted to complementary DNA using SuperScript™ IV Reverse Transcriptase (ThermoFisher) and a 1:1 mixture of random hexamers, following standard protocols. Specific intron-spanning qPCR primers were designed to eliminate the risk of amplifying genomic DNA: *Tef*, TTCGAGAAGGAAGCCGCTG F, GACGGTGACATAGTAGCGAGG R; *pelD*, CTTGGTAGGCTCACACTCGC F, CCTGGGTGACAAGGCCAC R; *OG0013477*, GTTAAGGTCCCTTGCCGTGT F, CCTAGATCGTCGCGTCTCAC R; *OG0014398*, TGTCGGTTTTCTTCTGCCTGT F, TGAAGCACGATCCTTGACGTT R. Thermal curves were analysed to identify their optimal annealing temperatures.

Quantitative PCR was performed in a Roche 480 II LightCycler using PowerUP SYBR Green Master Mix, with 15 μl of cDNA template and 1.5 μl of primer mix in a 30 μl reaction over three 10 μl replicates. Reaction conditions were: 95°C for 2 min, with 40 cycles of 95°C for 5 secs and 59°C for 60 secs. All primer pairs gave a single melt curve peak without shoulders. No amplification was found for negative controls without template DNA. Quantitative expression data were calculated using the Pfaffl method (Pennington et al. 2016).

4.3.2 Fungal and Plant Material

The *F. xylarioides* strains IMI 389563 and IMI 375908 (hereafter called 'arabica563' and 'arabica908') from the arabica host-specific coffee wilt disease population from the CABI-IMI culture collection (CABI, Egham, UK) were used. Strains were grown on synthetic low nutrient agar at 25°C, with the spores verified after five days using lactophenol cotton blue staining under a compound microscope following Geiser et al. (2005). The cultures were harvested with 1 drop of Tween20 and in 10ml sterile water. Two plates were combined to make each fungal inoculum before conidial concentrations were measured with a haemocytometer and adjusted to 1.00 x 10^6 spores/ ml. A third control inoculum

was made in the same way using sterile distilled water added to fresh synthetic low nutrient agar that had not been inoculated. Each inoculum type had four technical replicates (i.e. arabica563 1-4, arabica908 1-4, control inocula 1-4).

Coffea arabica plants were purchased online from Gardeners Dream (https://www.gardenersdream.co and left for 12 weeks to acclimate and grow in a controlled growth room at 25°C with a 12:12 hour light:dark cycle with 120 µmol/m2/s wavelength, and 50%:65% day:night humidity.

4.3.3 Infection assays and controls

The aim was to compare the fungal gene expression of the same arabica strains infecting coffee plants (*in planta*) and grown *in axenic* culture, as well as between the inoculated and control-inoculated coffee plants (Figure 4.1). Arabica coffee plants were infected via a stem wound following Girma (2004), using a sterile needle at the first green (i.e. non-woody) internode. Ten microlitres of fungal inoculum were inserted into each wound site and left to absorb/ air dry. Each inoculum type (arabica563, arabica908, control inoculum) was used to inoculate four plants, making up 16 replicates in total per inoculum type (i.e. four plants in arabica563 1, four plants in arabica563 2, four plants in arabica563 3. four plants in arabica563 4). These were grouped by inoculum type subset (arabica563 1 together, arabica563 2 together etc) and sealed in propagator trays with parafilm to make four trays containing four plants for each inoculum type. Tray location in the growth chamber was randomised and the plants were grown under shade, to mimic field conditions where coffee trees are grown under shade, at 25°C with a 12:12 hour light:dark cycle in a growth room until at least 50% of infected plants showed wilting, yellowing or early leaf senescence as symptoms of infection. The plants were harvested 98 days after inoculation.

Axenic cultures of arabica563 and arabica908 isolates were grown in GYM broth (see section 3.4.3 for recipe) at 25°C with shaking at 175 RPM for seven days. As with the *in planta* inocula, the *in axenic* culture samples included four biological replicates (arabica563 1-4 and arabica908 1-4).

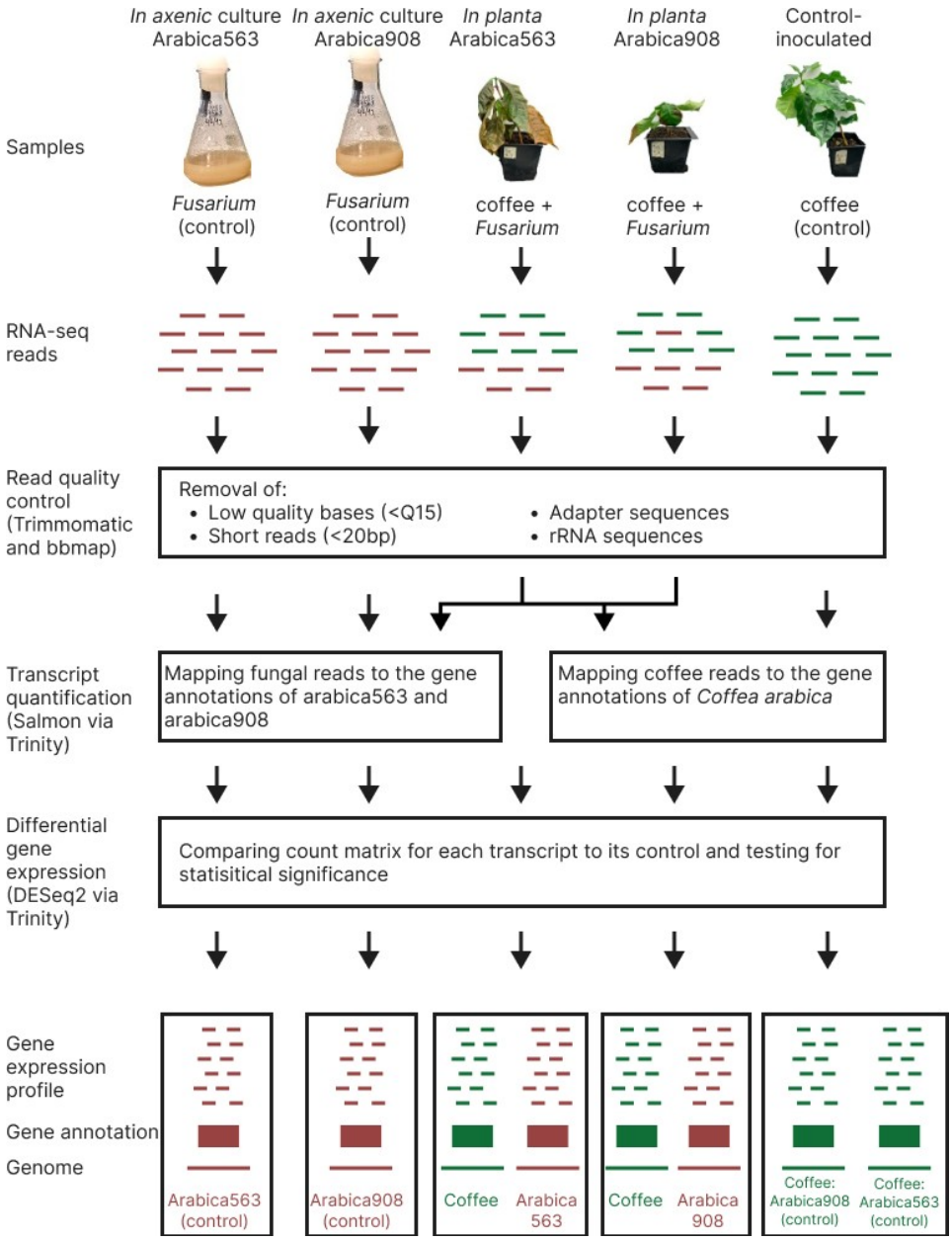


Figure 4.1: Schematic representation of RNA-seq analysis of mixed transcriptomes from *Fusarium xylarioides*-infected arabica coffee plants. Total RNA was extracted from *Coffea arabica* plants 98 days following control-inoculation with water (coffee plants, control) and inoculation with fungal inocula (coffee plants, *Fusarium xylarioides* strains arabica563 or arabica908), as well as from conidial suspensions of the *Fusarium xylarioides* arabica563 and arabica908 strains (fungal strains, control). The RNA-Seq libraries were cleaned to remove low quality bases, adapter sequences and short reads (Trimmomatic), as well as ribosomal RNA (bbmap). Transcript quantification was performed by assembling the libraries to their relevant gene annotations (Salmon), to ensure that only the relevant *Fusarium* and coffee reads were assembled. Finally, the count matrix for each transcript was tested for statistically significant differential expression (DESeq2)

4.3.4 RNA extraction, sequencing and quality control

To harvest, all aboveground plant material was collected and flash-frozen in liquid nitrogen. The biological replicates were ground separately in liquid nitrogen with PVPP (polyvinyl polypyrrolidone) following (Rubio-Piña & Zapata-Pérez 2011) in a mortar and pestle. Approximately 5g of each biological replicate was pooled in a 1.5 ml flash-frozen polypropylene tube before vortexing and re-freezing.

Total RNA was extracted using the standard protocol of the RNeasy Mini kit (Qiagen, Germany). In total, 20 samples were extracted with four technical replicates for each treatment: *in planta* arabica563-infected coffee plants; *in axenic* culture arabica563 ; *in planta* arabica908-infected coffee plants; *in axenic* culture arabica908 ; control (-inoculated with water) coffee plants. RNA was quantified using a NanoDrop 2000 (Thermofisher), with the RNA quality assessment (Qubit and Fragment Analyzer), library preparation and Next Generation Sequencing (60M paired-end reads per sample) performed by Genewiz/ Azenta (Frankfurt, Germany).

Raw sequenced reads were quality- and adapter-trimmed using Trimmomatic v0.39 (parameters: ILLUMINACLIP:$TRIMMOMATIC_DIR/adapters/TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25) (Bolger et al. 2014). Reads were quality-checked with FASTQC v0.11.2 (Andrews S et al. 2015) and retained where both pairs were trimmed. To remove ribosomal RNA sequences, reads were mapped to the SILVA rRNA database using BBTools 'bbmap' with the parameter 'outu=filtered_R*.fq.gz' to keep only the unmapped reads. Finally, the reads were repaired using BBTools 'bbmap' 'repair.sh' to re-pair reads that became disordered or had their pair eliminated. These QC steps removed 61 GB of data. Before using the data to quantify gene expression, reads were mapped using 'bbmap.sh' to the gene annotations of arabica563 and arabica908 from chapter three and from *C. arabica* (GCA_003713225.1, Zimin et al. (2019)). These mapped read libraries were used for all next steps.

4.3.5 Differential expression analysis

Using the gene annotations from chapter three as the target transcriptomes, transcript quantification was performed three times in a genome-free way using Salmon v1.5.2 (Patro et al. 2017) on the mapped read libraries: *in axenic* culture arabica563 and *in planta* arabica563; *in axenic* culture arabica908 and *in planta* arabica908; infected *C. arabica* plants and control-inoculated *C. arabica* plants. The relationships within and between biological replicates for each dataset were visually examined using the 'PtR' script in the Trinity toolkit (Grabherr et al. 2011, Haas et al. 2013). Each transcript count matrix was tested for differential expression using DESeq2 (Love et al. 2014), which uses negative binomial generalized linear models to test for statistical significance. Differential expression was also calculated using edgeR (Chen et al. 2016), which gave similar results, and limma/voom (Ritchie et al. 2015) which handles biases in RNA-seq data differently (Soneson & Robinson 2018) and was too conservative with the read numbers from the *in planta* samples for this experiment. The Benjamini-Hochberg method (Benjamini & Hochberg 1995) was used to adjust P-values for multiple testing to control the false discovery rate (FDR) and strict significance thresholds were applied, with P-values <0.001 and >4-fold to define a set of differentially expressed genes. The top 20% differentially expressed genes were classified as the 'most differentially expressed': 272 genes in arabica563; and 177 in arabica908. *Report BUSCO scores for quality of the transcriptomes

4.3.6 Functional annotation analysis

Gene ontology and InterProScan classifications

All differentially expressed genes were classified according to their gene ontology (GO) term, creating enriched and depleted gene classes across the strains. Functional enrichment analysis for the fungal strains was run using GO-Seq (Young et al. 2010) with gene ontology (GO) terms from chapter three, through the Trinity toolkit analyze_diff_expr.pl with the parameters '–examine_GO_enrichment –GO_annots –gene_lengths'. This script

returns differentially expressed genes (P-value <0.001 and fold change >4) for enriched or depleted GO categories for the up- or down-regulated genes in each comparison. Orthologous relationships between the two *F. xylarioides* arabica strains were determined using OrthoFinder v2.5.4 with default parameters (Emms & Kelly 2015, 2019). The same OrthoFinder output from chapter three was used in this analysis.

Arabica coffee gene annotations (accession: GCF_003713225.1) were assigned InterProScan and Gene Ontology classifications using 'interproscan.sh -iprlookup -goterms'. InterProScan classifications for all samples were analysed using Tidyverse R packages (Wickham et al. 2019).

EffectorP

A new set of putative effector proteins were identified using EffectorP 3.0 (Sperschneider & Dodds 2022) on secreted proteins which were significantly differentially expressed in both fungal strains (total differentially expressed secreted proteins: 224 in arabica908, 380 in arabica563). Only secreted proteins significantly up-regulated in both strains were described as effector-like. EffectorP uses machine learning models to find specific characteristics which may help the pathogen survive in its host plants apoplast or cytoplasm and identifies putative effectors beyond small, cysteine-rich secreted proteins. Its results were combined with the putative effectors identified in chapter two.

Carbohydrate-active enzymes

Plant cell walls are made up of cellulose, hemicelluloses and pectins in a cross-linked matrix (Hervé et al. 2010). Plant pathogens possess an array of enzymes which break down these cell wall polysaccharides such as glycoside hydrolases, carbohydrate esterases and polysaccharide lyases. These enzymes are grouped by family in the Carbohydrate Active EnZymes (CAZy) database (Cantarel et al. 2009). CAZyme-encoding differentially expressed genes were identified in two ways, by comparing the proportion of differentially expressed genes to the genome and taking those where >50% of the genes were significantly

differentially expressed in both strains, and by identifying enriched gene ontology terms with a cell-wall or pectin-degrading enzymatic function.

4.3.7 Annotation of repeats and transposable elements

The RepeatModeler and RepeatMasker outputs from chapter three were used in this analysis (Flynn et al. 2020, Smit et al. 2015).

4.3.8 Data analysis

All programs were run on the Imperial College Research Computing Service high-performance cluster. The following RStudio packages were used for analysis and data presentation: Tidyverse (Wickham et al. 2019); Cowplot (Wilke 2022); ggrepel (Slowikowski 2021), RIdeogram (Hao et al. 2020).

4.4 Results

4.4.1 Phylogenetic orthology

4.4.2 Quantitative PCR analysis of infected coffee

To verify whether the pathogenicity of fungal culture collection strains survived preservation, in this instance for up to 25 years, quantitative PCRs (qPCRs) were set up to test for fungal presence *in planta* following inoculation. In total, two replicated qPCRs were performed, measuring the expression of *pelD, OG0013477* and *OG0014398*, three putative effectors described in chapter two, relative to the control gene *Translation elongation factor (Tef)*, a housekeeping gene widely used to identify *Fusarium* species (Geiser et al. 2005). The analysis was completed across three samples: fungal *in axenic* culture; the stem internode containing the wound site into which fungal inoculum was inserted; and the leaves above the wound site. The Pfaffl expression values for the control genes in the fungal *in axenic* culture were calculated and plotted in relation to the other plant

samples and putative effectors (Figure 4.2B). For each putative effector, the transcript abundance was highest in the coffee leaves, but also present in the stem and the *in axenic* culture sample. Thus, *F. xylarioides* was detectable throughout the aerial parts of inoc-

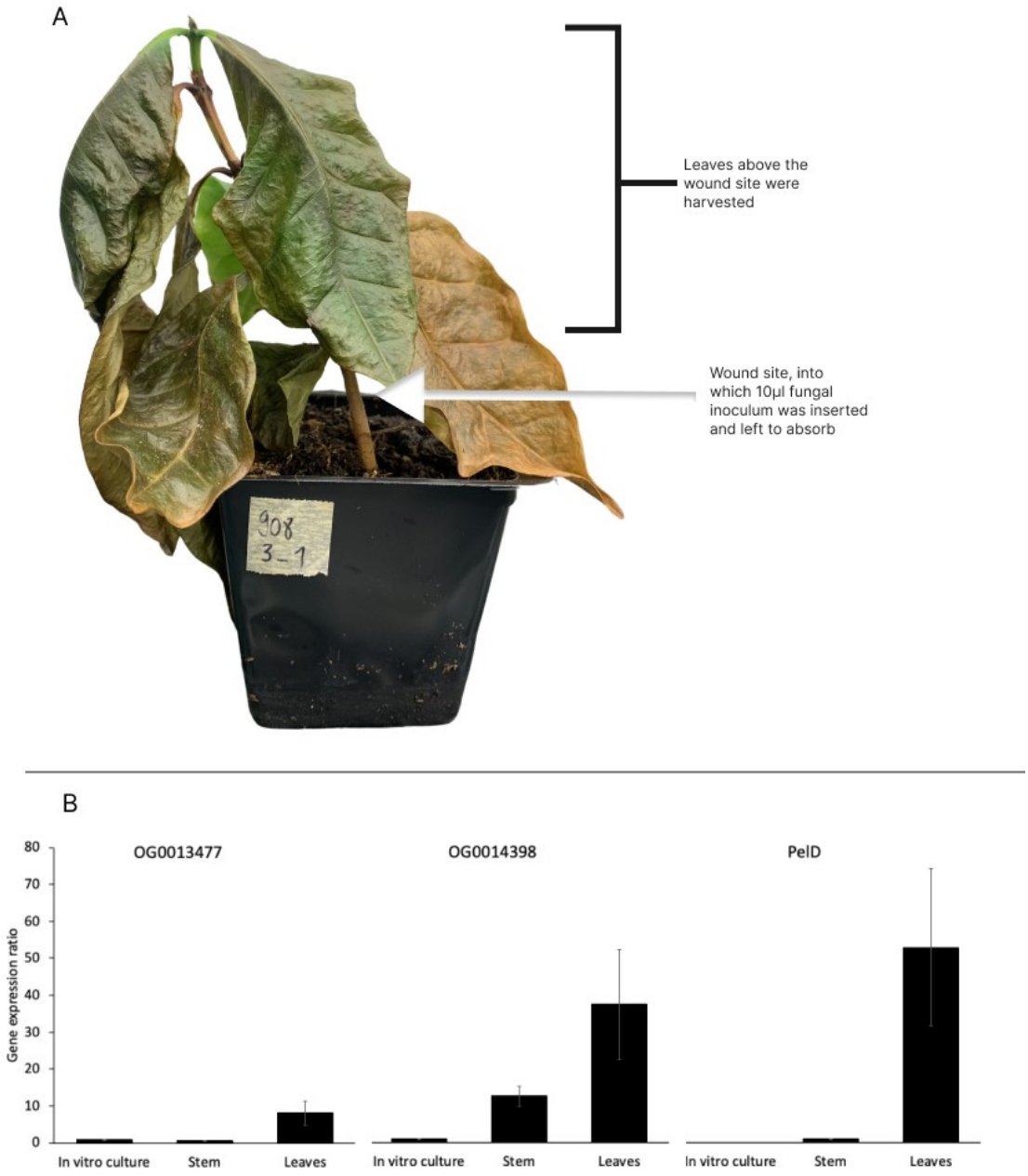


Figure 4.2: Putative effector gene presence across distal coffee plant parts. A An infected arabica coffee plant. Fungal inoculum was inserted into a wound site made on the stem and the plants were monitored until disease symptoms were visible. The stem above the wound site and leaves above the wound site were harvested separately with the expression of the housekeeping gene *Translation elongation factor (Tef)* compared to that of three putative effectors: *pelD, OG0014398* and *OG0013477*. **B** Quantification of RNA/cDNA expression compared to the control gene, *Tef*. The expression levels are Pfaffl expression values relative to *Tef* expression in the fungal *in axenic* culture sample.

ulated coffee plants, so for RNA-Seq analysis of mixed transcriptomes (of *C. arabica* and *F. xylarioides*) all above-ground plant material was harvested to ensure that any proteins expressed in the stem would be included in the samples.

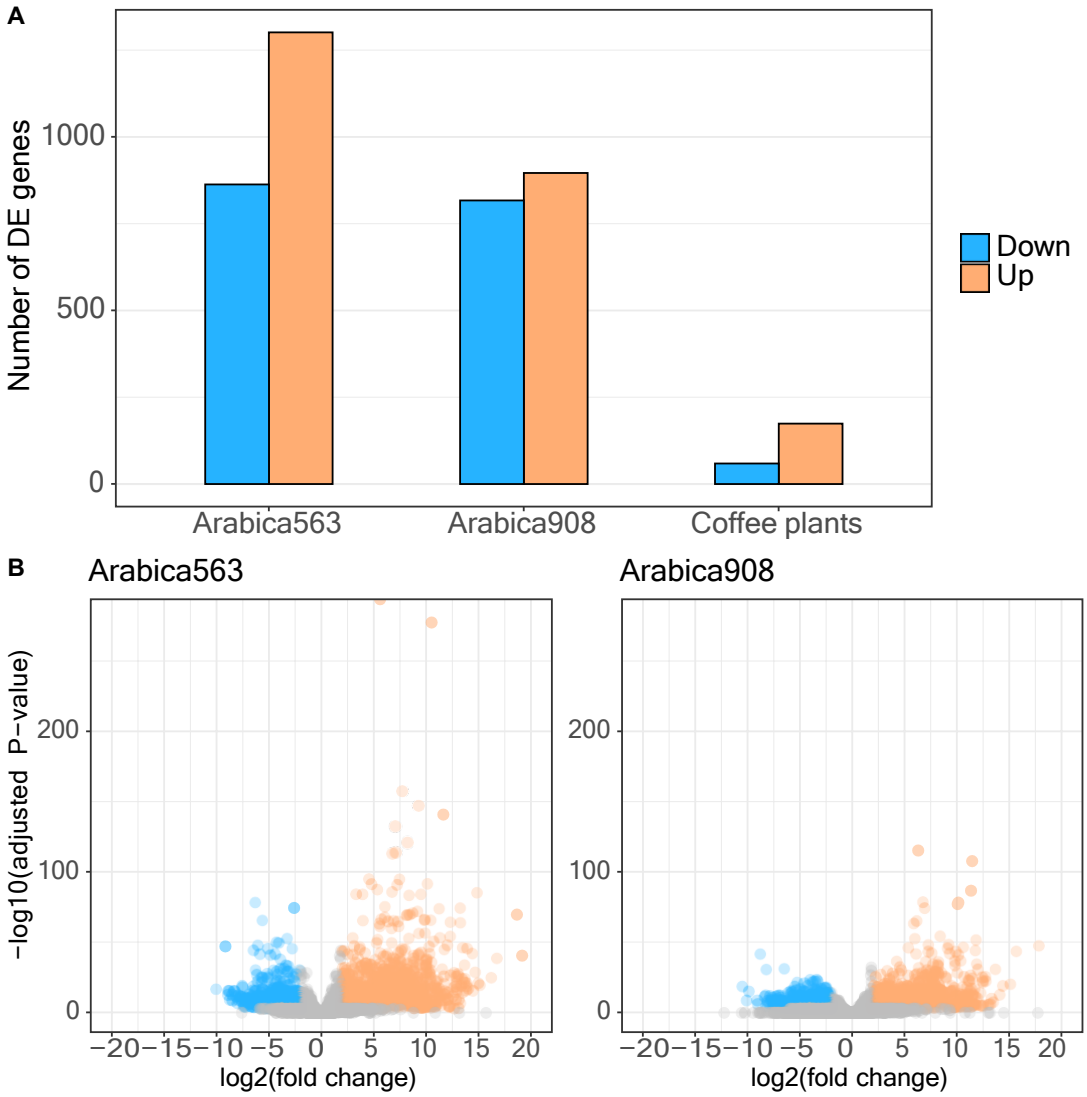


Figure 4.3: The response of *Fusarium xylarioides* pathogens and their plant hosts to experimental inoculation. A Numbers of genes showing significant up-regulation in the infected samples (orange bars) and up-regulation in the control samples (blue bars), for arabica563, arabica908 and coffee plants, relative to their controls (*in axenic* culture samples for the two arabica fungal strains and control-inoculated plants for the infected coffee plants.) Abbreviations: DE = differentially expressed. **B** Volcano plots showing the total differentially expressed genes for arabica908 and arabica563. Each point represents a gene's expression. Orange points are up-regulated *Fusarium* genes in the inoculated coffee samples; blue points are down-regulated genes; grey points are genes which are not significantly differentially expressed. The orange point obscured by the upper limit of the plot in arabica563 has a P-value of exactly 0.00.

4.4.3 Coordinated pathogenic transcription with a corresponding defence response following infection of *Coffea arabica*

The 20 RNA-seq paired libraries yielded 294 Gb of raw sequencing data, reducing to 233 Gb following QC. On average, 0.6% of reads from the *in planta* infected coffee libraries mapped to the *Fusarium* gene annotations, making up 900,000 *Fusarium* reads whilst the *in axenic* culture libraries contained 110 million reads. All *in planta* libraries contained >120 million *C. arabica* reads. Among these expressed genes and relative to their relevant controls, 2164 were differentially expressed across the arabica563 samples; 1713 across the

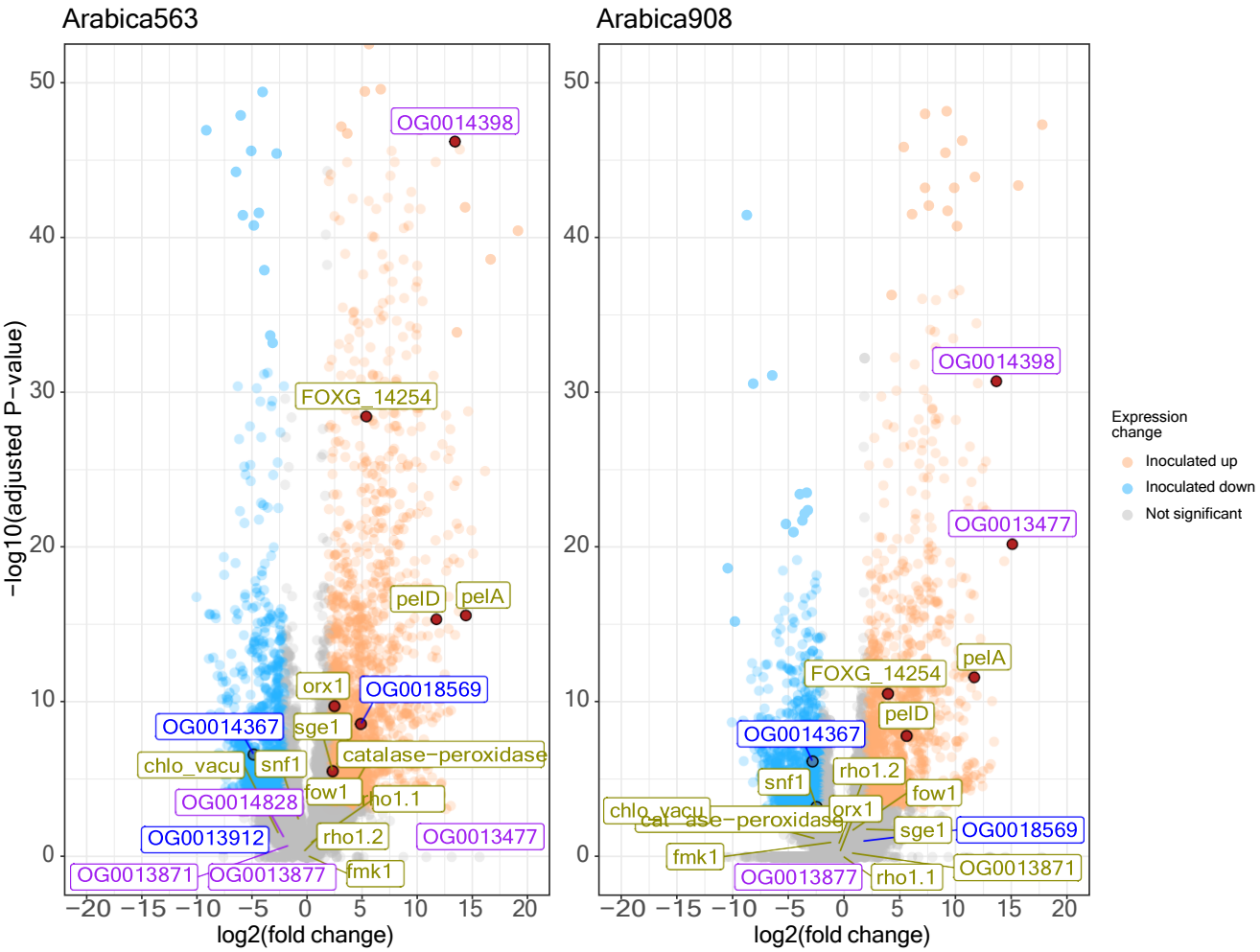


Figure 4.4: Differential expression of genes and putative effector genes in *Fusarium xylarioides* Genes up-regulated in the *in planta* samples are shaded orange and those up-regulated in the *in axenic* culture samples are shaded blue. Putative effectors identified in chapter two are labelled according to their class: yellow, pre-described fungal plant pathogens (Table A.4); purple, small and cysteine-rich proteins; blue, carbohydrate-active enzymes.

arabica908 samples; and ∼233 across the coffee plant samples (Figure 4.3). In the fungal samples, over half of these were upregulated in the *in vivo* infection samples relative to controls, whilst in the coffee samples 80% of the differentially expressed genes (116 in the plants infected with arabica563; 231 with arabica908) were up-regulated in the infected coffee samples. The differentially expressed genes belong to nearly 700 orthologous groups between the two arabica strains.

4.4.4 Significant up-regulation of putative effector genes in the arabica strains during pathogenesis

Of the pathogenicity factors and effector proteins described for *F. xylarioides* in chapter two, four genes were significantly differentially expressed in both strains with two in the 'most differentially expressed' class across both arabica563 and arabica908 (Figure 4.4). Two of these genes, *pelA* and *pelD*, both pectate lyases, are known from other fungal plant pathogns. *OG0014398* is named by its orthologous group and is a hypothetical protein under divergent selection, inferred to be a divergent type of Six7 effector and containing a *miniature impala (mimp)* transposable element linked to pathogenicity in *F. oxysporum* in its promoter region. *FOXG_14254* is an un-named secreted protein identified in *F. oxysporum* as a novel effector candidate and located on its mobile pathogenic chromosome (Schmidt et al. 2013). Just one putative effector, OG0014367, is down-regulated across both strains. Five putative effectors are significantly up-regulated in one of the arabica strains (*orx1, OG0018569, sge1, catalase-peroxidase, OG0013477* (with *OG0013477* expressed but not differentially in the other arabica strain), seven are expressed but not differentially in both strains and six genes expressed but not differentially in one strain (Tables A.12 and A.13).

In addition to the putative effectors described in chapter two, EffectorP 3.0 (Sperschneider & Dodds 2022) identified a new set using the differentially expressed secreted proteins. In both strains, ∼30% of differentially expressed secreted proteins were predicted to be effectors, with twice as many predicted as apoplastic effectors (arabica563,

75; arabica908, 34) compared with cytoplasmic effectors (arabica563, 36; arabica908, 17) and fewer still dual-localised (arabica563, 24; arabica908, 13). A small subset were down-regulated, 28 in arabica563 and 11 in arabica908, of which four belong to orthogroups. However, the majority were up-regulated, 107 in arabica563 and 53 in arabica908, and 41 belonged to orthologous groups in both strains.

Focussing upon the orthologous up-regulated predicted effectors, 22 are apoplastic: six were pectin lyase pectinolytic enzymes which can act as virulence factors (PL1: OG0007961, OG0005885, OG0007346; PL3, OG0000800 (the *pelA* putative effector), OG0009694; PL9, OG0005463) and all in the 'most differentially expressed' class; and nine were similar extracellular degradative enzymes, either those which break down plant cell wall polymers (CE5, OG0013204, OG0002450; CE12, OG0007257; GH5, OG0013750; GH11, OG0011334, OG0011590; GH12, OG0007484) or those which assist the activity of the cellulolytic enzymes (AA9, OG0004566, OG0003742, OG0004766). Six predicted effec-

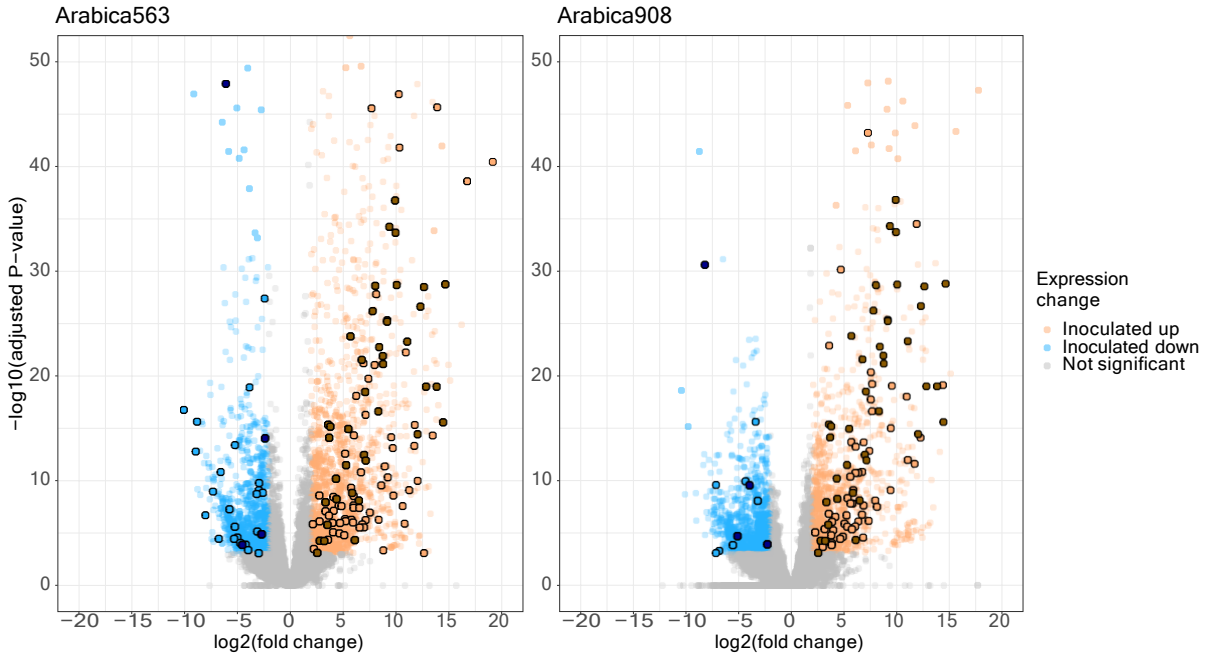


Figure 4.5: Differential expression of predicted effectors and the 'most differentially expressed' orthologous gene groups. Genes up-regulated in the *in planta* samples are shaded orange and those down-regulated shaded blue. Genes which share an orthologous group between the two arabica strains where both sets of genes are in the 'most differentially expressed' class are outlined in black, for both up- and down-regulated genes. Genes which are 'most differentially expressed' and a predicted effector (predicted by EffectorP) are shaded dark orange.

tor orthogroups had no recognised functional domain (OG0009750, OG0013167, OG0016169, OG0009728, OG0003872, OG0000381).

Eight predicted effector orthogroups dual-localize in the apoplast and cytoplasm. Two match fungal effector domains: OG0005427 encodes an Ecp2 effector protein; and OG0009487 contains a 'most differentially expressed' LysM domain, a known fungal effector domain. OG0011892 is involved in tyrosine and phenylanine biosynthesis, and the remainder do not match a functional domain.

There were eleven predicted cytoplasmic effectors. Three with CAZyme functions and 'most differentially expressed' as above (PL3, OG0005899; CE12, OG0005610; GH43, OG0011177) and OG0004761 matching a Necrosis-inducing fungal effector (Nis1). OG0007844 contains a FAS1 domain and OG0009835 contains a PAN/ Apple domain. The remainder had no recognised functional domain.

4.4.5 Significant up-regulation of fungal cell wall- and pectin-degrading enzymes during pathogenesis

There were 37 CAZyme sub-families containing at least one up-regulated *in planta* gene in both arabica strains (214 up-regulated genes in total, Tables A.14 and A.15). All five represented sub-families within the polysaccharide lyase (PL) family show strong up-regulation: PL1, pectate lyase (25 genes, 20 up, 12 'most differentially expressed'); PL3, pectate lyase (11 genes, 10 up, 5 'most differentially expressed'); PL4, rhamnogalacturonan endolyase (6 genes, 3 up, 2 'most differentially expressed'); PL9, pectate lyase (2 genes, both up, both 'most differentially expressed'); PL26, rhamnogalacturonan exolyase (2 genes, both up, one 'most differentially expressed') (Figure 4.6). One gene which is expressed in both genomes belongs to the family Auxiliary Activity 6 (AA6). This is a benzoquinone reductase involved in the protection of fungal cells from reactive quinone compounds. There are 12 Auxiliary Activity 8 genes (AA8, including a cytochrome domain involved in iron reduction) in both strains, with six gene copies significantly up-regulated. There are ten families with a majority of gene copies up-regulated which are

involved in the breakdown of plant cell wall compounds: Carbohydrate Esterase 12, a pectin acetylesterase (8 genes, 6 up); CE 16, which targets various acetyl esters (12 genes, 6 up), Glycoside Hydrolase (GH) 10 (8 genes, 5 up), GH11 (6 genes, 5 up), GH43 (48 genes,

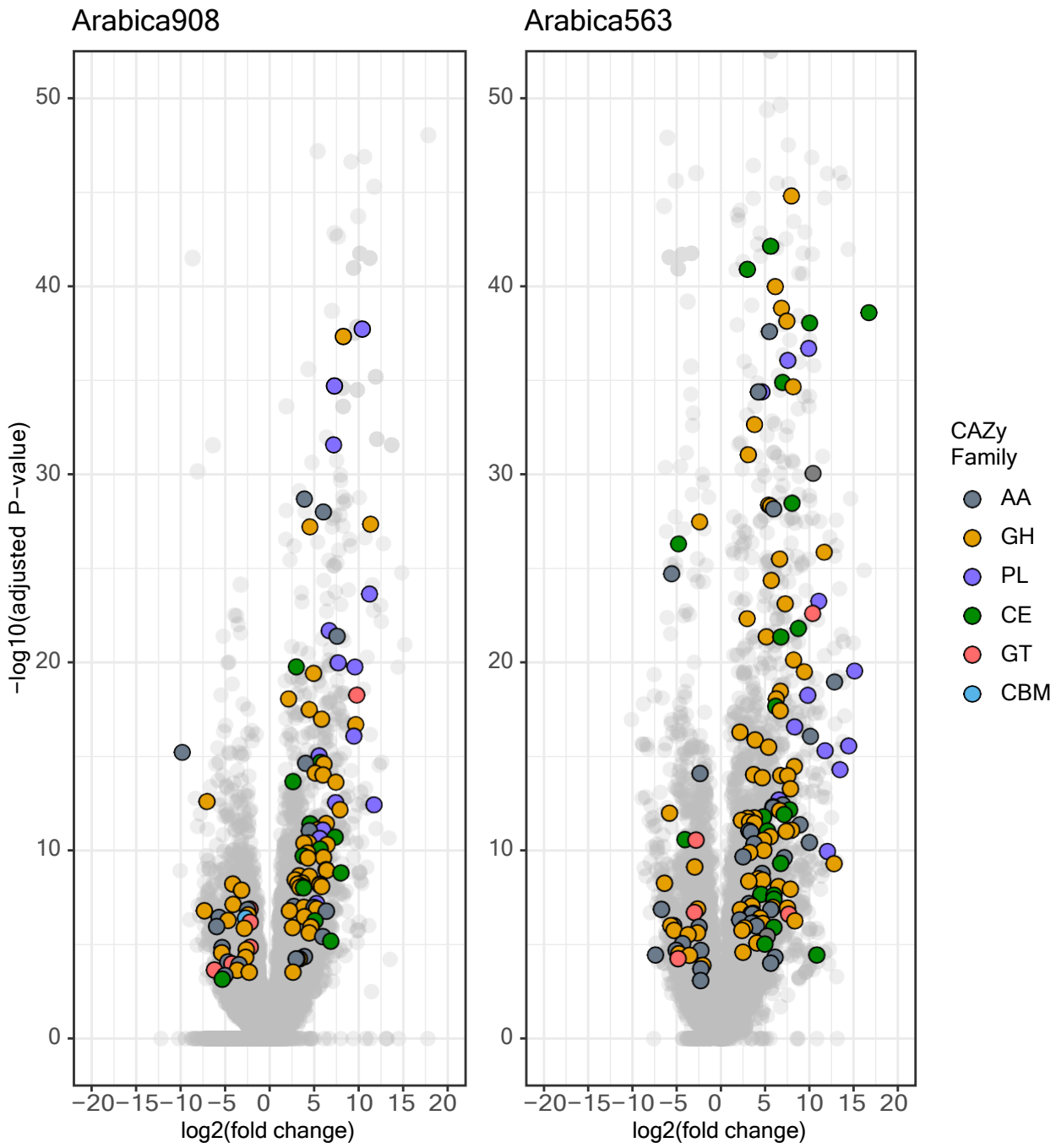


Figure 4.6: Significantly differentially expressed genes which encode a carbohydrate active enzyme (CAZy) family. Each point represents the expression of a gene and is coloured according to the CAZy family. There are no down-regulated polysaccharide lyase genes. Abbreviations: AA, auxiliary activities; CBM, carbohydrate binding module; CE, carbohydrate esterase; GH, glycoside hydrolase, GT, glycosyl transferase; PL, polysaccharide lyase.

27 up) and GH51 (4 genes, 2 up) are all xylanases; GH105, a galacturonidase (8 genes, 7 up); GH37, a trehalase (4 genes, 2 up); GH93, an arabinanase (10 genes, 5 up). There are 11 families which are significantly enriched and secreted by cells and retained within the extracellular space (apoplast) (GO:0005576) (Table A.16): xylan esterases (CE1, CE5), endo- (GH6) and glucanases (GH7, GH5), glucuronidase (GH67), galactosidases (GH2), pectate lyases (PL1, PL3) and copper-dependent lytic polysaccharide monooxygenases (LPMOs) (AA9, formerly GH61), as well as GH10, PL1 and PL3 which are described above. Seven genes in each strain do not match a specific CAZyme family but are 'most differentially expressed' and involved in cellulose binding (GO 0030248), which specifically involves the cellulose-binding superfamily with a catalytic domain. Cellulases and xylanases, important in the microbial degradation of cellulose and xylan, thus include enzymes such as endoglucanases or xylanases along with this cellulose-binding domain. In addition, six genes in each strain are involved in carbohydrate binding (GO 0030246) which includes non-CAZymes as well as pectate lyases (PL4) and galactosidases (GH2, GH31).

4.4.6 Significant up-regulation of other important fungal gene classes during pathogenesis

In addition to the cell wall- and pectin-degrading class, the remaining enriched GO terms were grouped into four main supercategories: transporters; metabolism-related enzymes; ribosomal protein biosynthesis; and iron binding (Table A.16). For transporters, the GO categories 'transmembrane transport', 'transmembrane transporter activity' were enriched with all genes including a transmembrane protein topology prediction. Metabolism-related enzymes include oxidoreductase activity and FAD, FMN and NADP binding. Ribosomal protein biosynthesis includes differentially expressed genes across protein translation and the structural integrity of ribosomes. Finally, there were genes involved in iron ion binding genes which include cytochrome P450 domains across both strains, in addition to the AA8 genes described above.

4.4.7 The genomics underlying a *Fusarium* strain-specific infection

Significant clustering of up-regulated genes

To look for parts of the genome which may be particularly important in the infection of *C. arabica*, regions in which both arabica strains show clusters of differentially expressed genes were examined. Contig 10 has a region which was significantly enriched for differentially expressed genes (Binomial Test, $p<0.01$), whilst contigs 6, 7 and 10 all share regions significantly enriched for up-regulated genes in both arabica strains (Figure 4.7, Binomial Test, $p<0.01$). The predominant gene type in the clusters on contigs 6 and 10 were secreted proteins, with more than 1/3 secreted in both (arabica563: contig 6 = 30/86, contig 10 = 32/73; arabica908: contig 6 = 17/51, contig 10 = 16/43). The contig 6 cluster contained the putative effector *pelD* in both strains, together with two other predicted effectors in arabica908 and an additional 7 putative effectors in arabica563. The contig 10 cluster contained 12 predicted effectors in arabica563 and 7 in arabica908.

There were also differences in the clustering of up-regulated genes between the two arabica strains. In arabica563, as well as contig 10, scaffolds 6, 7, 11 and 13 were all enriched in differentially expressed genes with contigs 11 and 13 enriched in down-regulated genes. In arabica908, contig 3 was enriched in differentially expressed genes, which were up-regulated along with a region of contig 7 and the entirety of scaffold 10.

Up-regulated *Fusarium xylarioides* arabica genes are closely related to *Fusarium oxysporum*, with three highly expressed secreted in xylem effector genes

The phylogeny of the differentially expressed genes with relation to *F. oxysporum*, the *Fusarium fujikuroi* species complex and between the arabica and robusta populations was also explored.

There were nearly 700 orthogroups significantly differentially expressed in both arabica strains. Looking at total phylogenetic distance for individual gene trees, these differen-

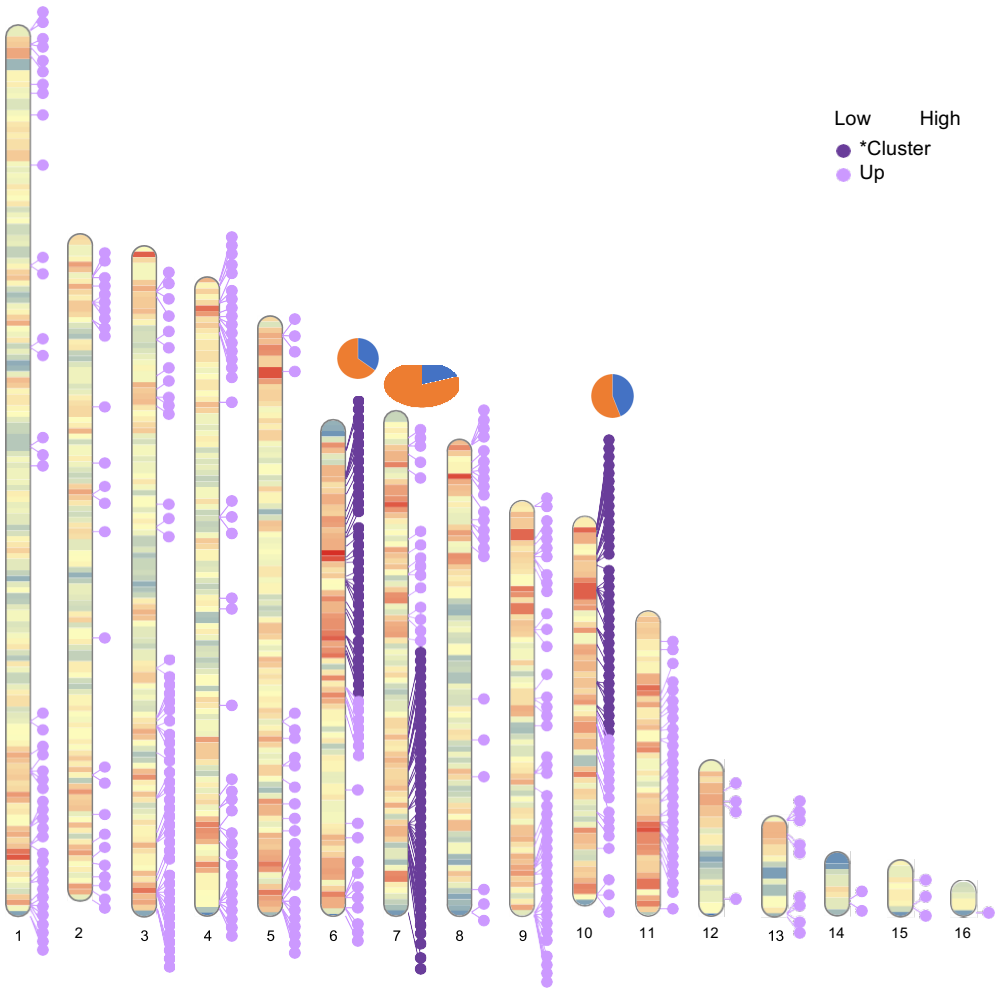


Figure 4.7: Whole-genome view of the loci of significantly up-regulated genes in arabica563. Gene frequency in 100 kb windows is represented by heatmap shading. Up-regulated gene loci which form significantly enriched clusters (Binomial Test, $p<0.001$) are labelled *Cluster with dark purple points (only genes <10kb from their nearest neighbour are shown). The proportion of genes inside these three clusters which are secreted are represented in pie charts (blue = secreted; orange = non-secreted).

tially expressed genes show a closer relationship to *F. oxysporum* than is expected from the whole genome level (Figure 4.8A). The phylogenetic distance between *F. xylarioides* and *F. oxysporum* was 0.1, based on the *Fusarium* species tree (Figure 3.3b). The mean phylogenetic distance between the two species for all 697 differentially expressed genes was 0.08, which reduces further to 0.055 for up-regulated genes. In turn, the differentially expressed genes showed a close phylogenetic relationship between the arabica and robusta genealogies, with a distance of 0.04 on the species tree and a distance of 0.027 for all differentially expressed and 0.022 for up-regulated genes. Therefore, it appears

the up-regulated arabica genes were phylogenetically close to both *F. oxysporum* and the robusta population.

Of the nearly 700 differentially expressed orthogroups, 29 are absent from the *Fusarium fujikuroi* species complex with 20 up-regulated in the *in planta* samples. Of the up-regulated orthogroups, twelve are shared by all the *F. xylarioides* populations (OG0011705, OG0013385, OG0013540, OG0014761, OG0014762, OG0014852, OG0014880, OG0015507, OG0015509, OG0015512, OG0015515, OG0015743), six shared by arabica and various coffea (OG0014579, OG0016169, OG0016940, OG0016944, OG0018160, OG0018949) with just two unique to the arabica population (OG0021712, OG0021744).

The majority of these up-regulated orthogroups do not match annotated functional proteins by BLAST, with the exception of OG0016940, OG0016169, OG0016944 which each match to an *F. oxysporum* effector protein: *Six7*, 91%id; *Six10*, 92%id and *Six12*, 94%id respectively. These three orthogroups are shared by all arabica and all coffea strains and are located within a 4.3 kb region of contig 3. They are also some of the most highly expressed genes across the arabica strains (Figure 4.8B). In addition, *six10* and *six12* have overlapping *mimps*, as well as other TEs, and *six7* is <1 kb from a *mimp4* (Figure 4.9). None of the other genes which overlap with *mimps* are differentially expressed. Of the remaining orthogroups, those which are arabica-specific do not match InterPro or Pfam domains and OG0021744 has a transmembrane domain. Of the remaining up-regulated arabica and coffea-orthogroups, only OG0014579 matches a functional domain with Alpha/Beta hydrolase fold.

Of the 12 up-regulated orthogroups present across all the *F. xylarioides* populations, three showed greater divergence between the arabica and robusta gene copies than the species tree with a phylogenetic branch length distance >0.04: OG0014880, 0.2; OG0015743, 0.157; OG0014761, 0.194. OG0014880 was the highly-expressed putative effector OG0014398 from chapter two and section 4.4.4. OG0014761 contains an FAD-binding domain whilst OG0015743 had no functional domain match. In OG0013385, the arabica and robusta gene copies were highly similar to each other (0.015 distance), as

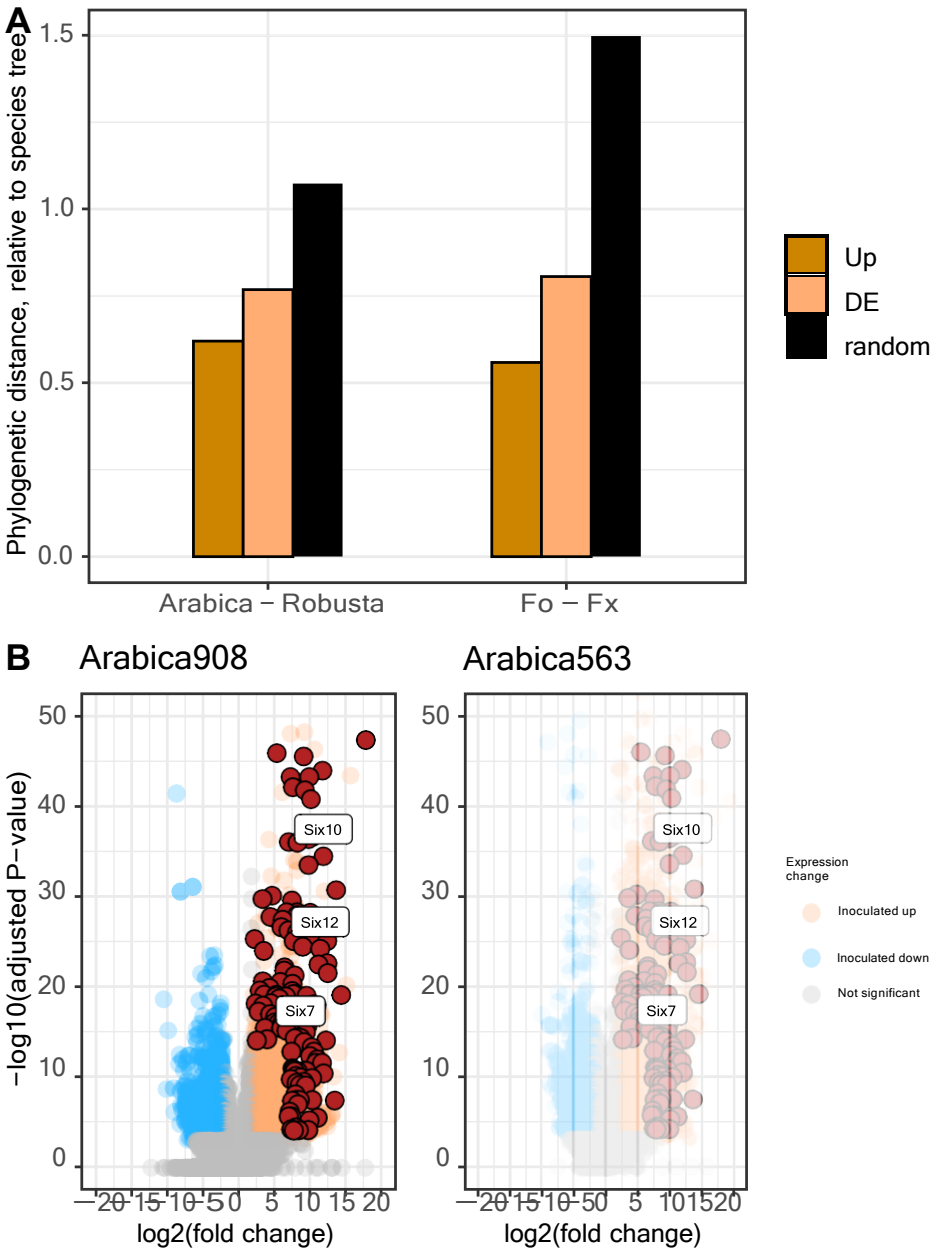


Figure 4.8: Up-regulated genes are closely related to *Fusarium oxysporum* with three highly expressed secreted in xylem *(SIX)* effector genes. A Barplot showing the phylogenetic distance, divided by the comparable distance in the *Fusarium* species tree (Figure 3.3b), between arabica and robusta strains and *Fusarium xylarioides* (Fx) and *Fusarium oxysporum* (Fo) strains for genes which are up-regulated (Up), differentially expressed (DE) and a random set of 697 genes which are present in *Fusarium xylarioides* and *Fusarium oxysporum*. A distance <1 is closer than the species tree, and >1 is more distant than the species tree. **B** Significant up-regulation of three *secreted in xylem* effector genes. Genes up-regulated in the *in planta* samples are in orange and those down-regulated are in blue. Genes which are shaded in dark red are the top 20% most differentially expressed genes which share an orthogroup between the arabica strains. The three orthogroups which encode the *Fusarium oxysporum SIX* effectors are labelled Six7, Six10 and Six12.

well as to *F. oxysporum* (0.075 distance) and contained a Domain of unknown function DUF4334. The remainder showed the expected divergence between arabica and robusta, and with *F. oxysporum*. Three orthogroups formed a cluster on contig 7 and contained domains for: OG0015512, a fungal transcription factor; OG0015509, a winged helix-like DNA-binding domain superfamily; and OG0015507, an MFS transporter superfamily.

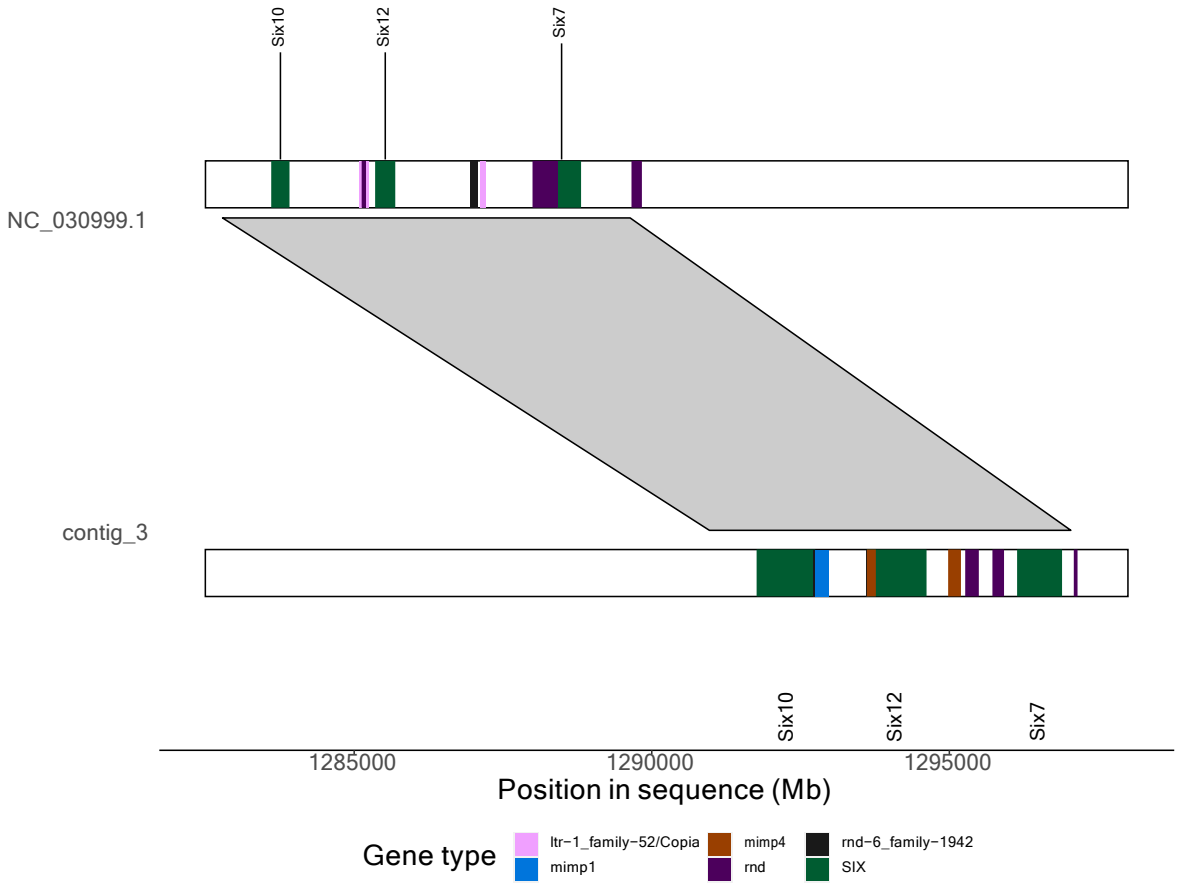


Figure 4.9: A large, gene-rich region is shared between *Fusarium oxysporum* f. sp. *lycopersici* and arabica563. A A region in *Fusarium oxysporum* f. sp. *lycopersici* chromosome 14 (NC_030999.1, top plot) is shared which matches to arabica563 contig 3 (bottom plot, grey rectangle indicates the region which is shared). The shared genes in this 4 kb region are all secreted in xylem *(SIX)* effector genes, which are shaded according to the legend and labelled. Repeats of interest are shaded according to the legend and unclassified repeats which occur once are labelled 'rnd'. The locations for *six12* and *six7* are from Schmidt et al. (2013). Figure drawn to scale.

4.4.8 Up-regulated pathogenesis-related genes in a strongly induced defensive response from *Coffea arabica*

Comparing the inoculated and the control-inoculated coffee plants, there were 116 up-regulated genes in plants inoculated with arabica563, and 231 in the arabica908-inoculated plants (Table A.17). There were also ∼48 down-regulated genes across the two samples. Across all infections, there were 80 significantly up-regulated genes, across 63 functional categories. A subset of these categories include multiple genes, which will be described in turn.

Twelve up-regulated genes encode PR proteins: Thaumatin-like protein synthesis, four genes (XP_027095536.1,XP_027096550.1,XP_027080774.1,XP_027080772.1); binding and oxidoreduction of flavins, two genes (XP_027087183.1,XP_027088006.1); D-mannose binding lectin, two genes (XP_027061914.1 and XP_027077059.1); anthocyanin biosynthesis, two genes (XP_027098144.1; XP_027127315.1); dirigent-like proteins, two genes (XP_027062967.1, XP_027062970.1); and phytoalexins, or lignin biosynthesis, two genes (XP_027062967.1 and XP_027062970.1).

Expansins are proteins which modify plant cell walls which can be expressed by plants in defence, and bacteria and fungi to overwhelm plant defences (Cosgrove 2015). Four expansin proteins are expressed across the infected coffee samples (XP_027083877.1,XP_027082625.1,XP_ and XP_027084129.1).

There are also a range of other defence-related proteins which are significantly expressed. Heat shock proteins (Hsp) account for ten significantly up-regulated genes (Hsp20, XP_027098388.1, XP_027113320.1, XP_027095883.1, XP_027096450.1, XP_027109657.1; Hsp70, XP_027125306.1, XP_027121874.1, XP_027122715.1, XP_027114671.1; Hsp90, XP_027110216.1). Heat shock proteins are so-called because they were identified by high expression at high temperatures, but roles beyond this have now been established and include the protection of newly synthesised proteins during a biotic or abiotic stress event (Ahuja et al. 2010).

4.5 Discussion

Sequencing the mixed transcriptomes of coffee plants inoculated with arabica *F. xylarioides* strains has revealed the molecular basis of a vascular wilt infection, here in arabica coffee. These data supported the critical role of cell wall and pectin-degrading enzymes in successful vascular wilt infection and provided a list of secreted proteins that behave as effectors. Furthermore, both arabica strains appear to have acquired and were highly expressing three *Secreted In Xylem* effectors, apparently by horizontal gene transfer from *F. oxysporum*. Finally, by including coffee plants in the analysis, a list of up-regulated plant genes was provided, many with a putative defensive function, which could be explored in future studies of the molecular basis of infection.

These results confirm the importance of the cell wall- and pectin-degrading enzymes in coffee wilt infection. This finding is consistent with other studies that examined *Fusarium* and *Verticillium*, as well as vascular wilt infections in other crops (Yadeta & Thomma 2013). *Fusarium* and *Verticillium* vascular wilt pathogens enter their hosts by entering and colonising the roots before migrating to the vascular parenchyma cells to enter the xylem vessels (Di Pietro et al. 2003, Klosterman et al. 2009, Schumann et al. 2010). Fungal vascular wilt pathogens are restricted to the xylem, and here *F. xylarioides* is found throughout aerial plant parts and with higher gene expression in the leaves than in the stem. Given the established passage of vascular wilt pathogens into their hosts via root tissue, it would be interesting to expand this experiment to include a root assay in gene expression analysis.

Plant cell walls are primarily composed of cellulose and xylan embedded in a matrix of hemicelluloses, pectic polysaccharides, and glycoproteins (Carpita & Gibeaut 1993). The genomes of wilt pathogens are significantly enriched in cell wall degrading enzymes that are believed to degrade xylem walls to release sugars for the acquisition of nutrients by fungal pathogens (*F. oxysporum* and *Verticillium* (Ma et al. 2010, Klosterman et al. 2011, Fradin & Thomma 2006, Di Pietro et al. 2003), as well as other types of inducers of vascular wilt (such as bacteria, Sun et al. (2005), Jha et al. (2007)). For

example, *F. oxysporum* f. sp. *cubense* races 1 and 4 express an arsenal of secreted CAZymes, mostly the CE family, when infecting their banana hosts (Guo et al. 2014). The high expression of polysaccharide lyase (PL) families that directly metabolise pectins is consistent with these findings and is consistent with reports for *Verticillium* and *F. oxysporum* f. sp. *lycopersici* wilts (Klosterman et al. 2011). A diverse set of polysaccharide lyase families were expressed, including PL1, PL3, PL9 and rhamnogalacturonan lyases in the PL4 and PL26 families. In addition to polysaccharide lyase families that directly metabolise pectins, many other CAZymes that metabolise the products cleaved from pectins by polysaccharide lyase enzymes were also highly expressed, such as GH105, GH51 and GH43 involved in the breakdown of rhamnogalacturonans, or CE12 in homogalacturonan products (Gilkes et al. 1991). Cell wall breakdown releases pectins into xylem vessels, which could unintentionally act as a barrier to further pathogen growth, so pectin-degrading enzymes could break down pectin barriers for the pathogen to spread leading to external symptoms of wilting as infection progresses (Clérivet et al. 2000), further strengthening the case for a strong pectinolytic machinery in fungal vascular wilt pathogens.

In addition to the importance of cell wall-degrading enzymes, these results have provided the first insights into characterising effector-like genes involved in the onset of coffee wilt disease. Pathogens secrete effector proteins into their hosts intracellular (plant cytoplasm) and extracellular (plant apoplast) spaces to modify host immunity and metabolism (Lo Presti et al. 2015). Cytoplasmic effectors may target specific compartments of host cells, while apoplastic effectors are active within the apoplast to target host plant enzymes (De Wit 2016) and can be attached to the fungal cell well (Tanaka & Kahmann 2021). Therefore, proteins secreted in these areas were specifically analysed. Of the putative effectors described in chapter two, four appear to indeed be effectors or at least behave like effectors. Two of these genes, *pelA* and *pelD*, both pectate lyases, are known from other fungal plant pathogens. *OG0014398* is named after its orthologous group and is a hypothetical protein under divergent selection, inferred to be a divergent type of *six7* effector

and containing a *miniature impala (mimp)* transposable element linked to pathogenicity in *F. oxysporum* in its promoter region. *FOXG_14254* is an un-named secreted protein identified in *F. oxysporum* as a novel effector candidate and located on its mobile pathogenic chromosome (Schmidt et al. 2013). I also observed the expansion of some conserved and secreted protein families that are known to have important roles in pathogenesis, such as LysM and Ecp effectors (de Jonge et al. 2010, Lu & Edwards 2016). The LysM domain is believed to be important in the detection of chitin by receptors, and the effector Ecp2 is also known from *Cladosporum fulvum* (Stergiopoulos et al. 2010) and homologous copies (albeit with a low sequence similarity) from *F. graminearum* (Lu & Edwards 2016). This *F. xylarioides* Ecp2 gene also has low sequence similarity, with a 70% BLAST id to the *F. graminearum* copy (FGSG_05341, XM_011325559.1). The effector function of Ecp2 is not fully known, although in *C. fulvum* it induces hypersensitive response cell death in tomato with a cognate resistance gene. As well as the genes that behave like effectors, there are a number of predicted effectors that do not, namely, they are expressed but not differentially. Perhaps if this experiment were repeated at a different time point, they would be. And there are also predicted effectors which are down-regulated in both arabica strains. These could be effectors that coffee plants have evolved to recognise and activate effector-triggered immunity in response to their detection (Jones & Dangl 2006), and therefore are now down-regulated in arabica strains.

The finding that up-regulated genes are phylogenetically closer to *F. oxysporum* than at the whole genome level, and that three *F. oxysporum* secreted in xylem effector genes are highly expressed in both arabica strains, with *mimps* within 1.5 kb of each promoter region, builds on chapter three to further support the role of horizontal transfer in the origin of coffee wilt disease. On the *F. oxysporum* mobile pathogenic chromosome, four clusters are enriched in *SIX* effector genes, genes involved in pathogenicity, class II TEs and *mimps*. Schmidt et al. (2013) suggest these concentrated clusters may originate from the recombination of inverted repeats from similar class II TE families, with adjacent *SIX* genes transposed together with the TE. They report that *mimp* inverted repeats

directly flank *six12* and another *SIX* mini-cluster. Of the *SIX* genes reported here, *six12* and *six10* contain *mimps* and their IRs within the coding-region, while *six7* is <1 kb from a *mimp*, and all three *SIX* genes make up part of a mini-cluster, the so-called supercontig 51, on the *F. oxysporum* f. sp. *lycopersici* pathogenic chromosome. Remarkably, it is a different piece of supercontig 51 from that which is present in the robusta genomes in chapter three. No other *SIX* mini-clusters are present or expressed in *F. xylarioides*. A study looking at transcriptomes of two *F. oxysporum* f. sp. *cubense* races found that both lack mobile chromosomes of *F. oxysporum* f. sp. *lycopersici*, yet still show differing presence and expression of *SIX* genes (Guo et al. 2014). In *F. oxysporum* f. sp. *cubense* race 1, *six1* and *six6* are present, while three orthologs of *six1* as well as *six2, six6* and *six8* are found in race 4. Given the mobile nature of the *SIX* genes, it is perhaps not surprising to find different sets of *SIX genes* which do not match the mini-clusters of *F. oxysporum* f. sp. *lycopersici*. In *F. xylarioides*, the three *SIX* genes are present in each coffea strain as in the arabica strains, but are absent from the robusta genomes and all *Fusarium fujikuroi* complex species. This, combined with the fact that there is no genome-wide pattern of mimp presence and up-regulation (as in *F. oxysporum* f. sp. *lycopersici*, Schmidt et al. (2013)), is suggestive of a transfer event. Furthermore, the general close phylogeny between the up-regulated genes in the arabica strains and *F. oxysporum* could be further evidence of horizontal transfer, or convergence of gene types that are important in vascular wilt infection. Indeed, where robusta contains orthologues for up-regulated genes, the copies of arabica and robusta are phylogenetically close, revealing a close phylogeny between *F. oxysporum*, arabica, and robusta in the gene families of highly-expressed coffee wilt infection genes.

These results provide a detailed insight into the response of the coffee plant to invasion by *F. xylarioides*. The first stage of the inducible defence response in plants is the recognition of PAMPs by plant pathogenicity response proteins and results in pattern-triggered immunity (Jones & Dangl 2006). Plant lectin proteins are part of the receptors involved in the detection of pathogen attack via PAMPs. Plant lectins do so by recognising specific

carbohydrate structures, either from the metabolised plant cell wall or from the pathogen itself (Lannoo & Van Damme 2014). Two mannose-binding plant lectins are up-regulated across the infected coffee samples. Mannose-binding lectins are expressed in the nucleus and cytoplasm and are part of the jacalin-related lectin family, in which proteins have been linked to resistance to disease, wounding, and abiotic stress (Song et al. 2014). One of such proteins is TaJRLL1 in wheat: this jacalin-related lectin binding mannose is expressed after fungal infection by *F. graminearum* and *Blumeria graminis*, and its gene silencing increased susceptibility. It is believed to be involved in salicylic acid and jasmonic acid-dependent defence signalling pathways (Xiang et al. 2011), which are a critical part of the plant defence response (Glazebrook 2001). Another type of inducible immunity in pattern-triggered immunity is the production of antifungal polyphenols. Genes are significantly up- regulated in the coffee samples for two such polyphenols: phytoalexins, or lignan and anthocyanins. Both types of coffee, *C. arabica* and *C. canephora*, produce phytoalexins when challenged with *Hemileia vastatrix*, coffee leaf rust (*C. arabica*; Rodrigues et al. (1975), *C. canephora*; Carlos Florez et al. (2017)). The genes that produce phytoalexins are dirigent-like genes, two of which are expressed in coffee samples. In peas, a phytoalexin is produced in the epidermal layer following infection by a fungal plant pathogen (Borges et al. 2013). Anthocyanins are a type of flavonoid, with proven roles in reducing fungal growth in grapes (Schaefer et al. 2009). These results provide important gene targets that could be targeted in coffee breeding programs that aim to increase resistance to coffee wilt disease.

Chapter 5

General Discussion

5.1 Thesis questions

The overarching aim of this thesis was to understand the evolutionary processes behind repeated outbreaks of fungal pathogens. Genomes from the CABI-IMI collection were analysed to understand the pathogen and its divergence into host-specific races. Additionally, it was hoped that this research could inform policy or growers to be better informed and prepared for a future re-emergence. The questions of this thesis, as stated in the introduction, were:

1. Is *Fusarium xylarioides* a single species, or a species complex?
2. Did later outbreaks descend from the earlier initial outbreak, or were they separate origins from different wild sources?
3. What genes might underlie pathogenicity in *Fusarium xylarioides*?
4. What genetic changes underpinned the emergence of the host specialists and did horizontal gene transfer bring in new sets of pathogenicity genes?

The results of these questions have been reported in the experimental chapters two to four. The first question, whether *F. xylarioides* is a single species or a species complex

was initially addressed in chapter two, with further verification in chapter three. Chapter two showed that the *F. xylarioides* arabica and robusta strain groups independently emerged within the *F. xylarioides* species clade, and that the robusta group appeared to have emerged from within a wider recombining group of strains from the initial outbreak ('coffea'). This interpretation was based on just six strains, and so chapter three sequenced an additional five *F. xylarioides* genomes and included two publicly available genomes to show the presence of four genetically differentiated populations in *F. xylarioides*. These populations included the arabica and robusta groups, and two populations among the coffea strains. Chapters two and three also helped to answer the second question, whereby the arabica population had a separate origin as a sister clade to the robusta and both coffea populations, while the robusta population was more closely related to one of the coffea populations (coffea1), possibly emerging from within this subset. Chapters two and four helped provide gene candidates that might underlie pathogenicity, finding firstly that the *F. xylarioides* genomes have expanded in copy number in certain types of carbohydrate-active enzymes (CAZymes), pectin lyases in particular, which are apparently important in vascular wilt infection, before showing that these same genes were significantly up-regulated across both arabica transcriptomes in chapter four. Additionally, the high expression of *F. oxysporum SIX*, that is, secreted in xylem, effectors in arabica strains is further evidence of the importance of these genes in vascular wilts. Finally, in chapter four, this thesis builds on the suggestion in chapter two of horizontal transfer from *F. oxysporum* by identifying large, syntenic, and highly similar gene-rich regions that the *F. xylarioides* arabica and robusta populations share with *F. oxysporum* f. sp. *lycopersici*, which are all absent from other *Fusarium* species. These likely horizontally transferred regions from *F. oxysporum* into *F. xylarioides* are different in the arabica and robusta groups, and could therefore have contributed to the emergence of new disease types. The answers to these thesis questions are discussed in more detail in subsequent sections, exploring speciation, pathogenicity genes and a putative pathogenicity role of the horizontal transfers.

5.2 Speciation in the *Fusarium xylarioides* clade

One of the original questions at the outset of this thesis was whether *F. xylarioides* is a species, a species with *formae speciales* (Girma 2004), a species complex containing cryptic species (Geiser et al. 2005) or multiple species including *F. abyssiniae* (the arabica population) and *F. congoensis* (the robusta population) (Lepoint 2006).

Initially, to begin answering this question, short-read Illumina *F. xylarioides* genomes from each of the outbreaks were sequenced (chapter two): two strains ('coffea') from the 1920s to 1950s outbreak; two strains from the 1970s onwards robusta outbreak in Uganda, Tanzania and the Democratic Republic of the Congo ('robusta'); and two strains from the arabica outbreak in Ethiopia ('arabica'). Whilst the Illumina genomes revealed differences between these groups of strains, namely that the robusta strains emerged from within a group of wider recombining coffea strains and that the arabica strains are a sister clade to them both, there were also challenges that could not be resolved by a small number of short-read genomes. These included the assembly of apparent lineage-specific regions and recognition of genetic subdivision. Assembly problems are characteristic of short-read genomes, in which high levels of fragmentation are seen in the more complex and difficult-to-assemble regions, which are also often the most interesting. To address these issues and resolve the original question, seven additional genomes were included in the analyses (five sequenced in chapter three and two publicly available) and a long-read genome was sequenced and assembled into a reference quality assembly for an arabica *F. xylarioides* strain (IMI 389563, arabica563). Mapping a larger number of genomes to a reference assembly to analyse single nucleotide polymorphisms enabled the unanswered question to be resolved.

The *F. xylarioides* arabica and robusta groups are genetically diverged populations, and differentiated from the coffea group, which itself is composed of two differentiated populations. Indeed, more diversity is revealed within the robusta population than was apparent from previous analyses (e.g. Buddie et al. (2015) and Lepoint (2006) and in chapter two). The four *F. xylarioides* populations are genetically distant, where high

values of the fixation index between all populations indicated little sharing of genetic material with fixed populations. This supports Lepoint (2006)'s recognition of biological species and incompatible mating between the strain groups. It is interesting that the range of the two coffea populations overlapped spatially and temporally, yet they remained genetically distant. Possible causes of this genetic differentiation could be clonal lineages, as suggested by the low recombination observed between populations, or reproductive isolation due to pleiotropy between host adaptation and mate choice (Giraud et al. 2010), as also suggested by Lepoint (2006)'s biological species. Where mating occurs within the host, specialisation, and reproductive isolation can result if different hosts are targetted. The presence of perithecia on trees affected by coffee wilt indicates that *F. xylarioides* undergoes such processes (Rutherford et al. 2009), so its populations could have become differentiated because only strains able to infect the same coffee tree host could mate. Whilst there were probably more than just two coffea populations, there are unlikely to be many (or even any) additional strains from the initial (1920s-1950s) outbreak, nor any further details surrounding host preference of the existing coffea strains which could be included to further resolve this question. This is a problem caused by a lack of suitable strains collected and correctly stored (see section 5.5).

These populations also reveal that they had differentiated prior to the outbreak of coffee wilt disease in the 1970s: each well-resolved population contains a strain isolated in 1968 (robusta268) and 1971 (arabica038). Therefore, it is probable that all four *F. xylarioides* populations identified here were already present by the 1970s and did not evolve from a common ancestor that later adapted to specific coffee species. Remarkably, this is what some of the early interpretations using simple molecular techniques and field expertise also concluded. This indicates that by the time of the 1970s onwards outbreak in the Democratic Republic of the Congo, the different genotypes were already present. Indeed, it could be that in the vast abandoned coffee plantations across the north-east Orientale province, following independence from Belgium where *F. xylarioides* began to proliferate in the 1970s (see section 1.5.3), a certain strain type became dominant and led

to the later outbreaks in robusta coffee in the 1990s.

The *F. xylarioides* populations show an incongruence between the gene trees similar to that within the *F. oxysporum* species complex. Although *F. xylarioides* is currently classified as a single species, these genetically differentiated populations support the earlier work of Lepoint (2006) and Girma (2004), which suggested that *F. xylarioides* is a species complex containing distinct arabica and robusta groups, which Girma (2004) called respectively *F. xylarioides* f. sp. *abyssiniae* and *F. xylarioides* f. sp. *canephorae*. The fact that the populations are genetically distant and were present prior to the re-emergence of coffee wilt disease following its initial outbreak from the 1920s to 1950s suggests that perhaps the *F. xylarioides* formae speciales proposed nearly two decades ago could now be recognised.

5.3 An enhanced pectinolytic machinery in the *Fusarium* coffee wilt pathogen

The *F. xylarioides* populations show differences between gene content, with certain classes of carbohydrate-active enzymes (CAZymes) expanded between them. The same gene classes were not only present in the genomes sequenced in chapter two, but also in the *F. xylarioides* arabica assemblies mapped to the reference genome in chapter three and were significantly up-regulated in the *F. xylarioides* arabica transcriptomes in chapter four. This consistent finding proves the validity of the methods used throughout this thesis. It is clear from the data that, as is known from other fungal inducers of vascular wilt disease (including *F. oxysporum*), strong pectinolytic machinery is a key pathogenic trait. However, only arabica transcriptomes were analysed, so it is plausible that a different set of CAZymes (possibly robusta-specific families described in chapter two) would be up-regulated in a robusta coffee wilt infection. Despite this, even though only arabica transcriptomes were sequenced, it was possible to compare gene expression with the robusta population by orthologous gene groupings. If a gene with an orthologous copy in

robusta was up-regulated in the arabica strains, it was assumed that it would also be in the robusta strains.

These results confirm the importance of enzymes that degrade the cell wall and pit membranes in the infection of coffee wilt. This finding is consistent with other studies that look at *Fusarium* and *Verticillium* (Klosterman et al. 2011), as well as vascular wilt infections in other crops in which pectin-rich material was deposited around the xylem parenchyma cells and their outgrowths (known as tyloses) and pit membranes (Clérivet et al. 2000). The genomes of wilt pathogens are significantly enriched in cell wall-degrading enzymes that are believed to degrade the walls of the xylem to release sugars for the acquisition of nutrients by fungal pathogens (*F. oxysporum* and *Verticillium* (Ma et al. 2010, Klosterman et al. 2011, Fradin & Thomma 2006, Di Pietro et al. 2003). The high expression of polysaccharide lyase (PL) families that directly metabolise pectins is in good agreement with these findings, including PL1, PL3, PL9 and rhamnogalacturonan lyases in the PL4 and PL26 families. In addition to polysaccharide lyase families that directly metabolise pectins, many other CAZymes that metabolise the products cleaved from pectins by polysaccharide lyase enzymes were also highly expressed. Given that pectin also has a putative defensive role in plants (Clérivet et al. 2000), this thesis further builds the case for a strong pectinolytic machinery in fungal vascular wilt pathogens.

5.4 Horizontal transfers of pathogenicity regions between *Fusarium oxysporum* and *Fusarium xylarioides*

These results demonstrate the multigenic nature of fungal plant pathogens and how they respond to environmental factors. The discovery in chapter two that certain *F. xylarioides* scaffolds and their genes showed homology to *F. oxysporum*, with evidence of their horizontal acquisition, was unexpected. Subsequently, this influenced the hypotheses in

chapter three, in which a variety of *F. oxysporum* and *F. solani* genomes were examined for putative horizontal transfers. It was important to include, and latterly eliminate, *F. solani* as another potential donor *Fusarium* species because it also infects coffee plants and has mobile and pathogenic chromosomes. In looking for a signature of horizontal acquisition, no single gene would confer host specificity, and so a diverse set of pathogenicity genes was expected with the emergence of newly pathogenic or host-specific strains involving shifts in profiles of multiple genes. This could occur by the transfer of a whole pathogenicity-linked region recently, or into a lineage which latterly diverged into independent and significant crop pathogens.

The later work in chapter four was able to build on the initial findings in chapter two, which identified a set of putative effector proteins. Although the effectors from chapter four cannot be described as true effectors, lacking knockout experimental data to verify these, they can certainly be described as effector-like, with high induction across multiple replicates in coffee wilt infection. The methods of Schmidt et al. (2013) to identify new types of effectors in *F. oxysporum* could be further applied to the effector-like list to find candidates with *mimps* in their promoter regions. In addition, the mixed transcriptome experiment in chapter four could be repeated to include the roots of coffee plants. As a soil-borne fungal pathogen, it is more likely to infect through plant roots in the field, and therefore, a different subset of genes could be expressed in root tissue.

The finding that multiple different regions in *F. xylarioides* appear to have been horizontally transferred from *F. oxysporum* is remarkable. Both arabica and robusta populations contain highly similar regions that match different parts of the effector- and transposable element-rich 'supercontig 51' on the *F. oxysporum* f. sp. *lycopersici* mobile pathogenic chromosome. In the robusta genomes, this matches a scaffold that is absent from arabica, and so is absent from the reference assembly, whilst in the arabica genomes this matches to a mimp- and effector-rich part of contig 3 with three *SIX* gene copies significantly up-regulated in both strains in chapter four. It is *six7, six10* and *six12* that were up-regulated, and *six7* and *six10* were additionally identified as putative effectors

in the arabica strains in chapter two. In this *F. xylarioides* example, it is possible that the robusta part of 'supercontig 51' was latterly lost from the arabica genomes, additionally because it does not contain *SIX* effector genes. The whole region could have been acquired by a distant common ancestor to the two populations that has diverged with time between arabica and robusta. Gene loss of regions that are no longer useful to the organism is well reported (Ropars et al. 2020). A reference assembly for the robusta genomes, as well as transcriptome data from a robusta coffee wilt infection, would help to resolve these questions.

There appear to be two possible mechanisms for the acquisition of these regions by horizontal transfer from *F. oxysporum*, both of which are plausible. The first is that of whole-chromosome transfer, which has been experimentally verified in *F. oxysporum*. Hyphal fusion or anastomosis would be required for this, and *F. xylarioides* is known to undergo such processes between its populations. The other mechanism would be transposition-mediated horizontal transfer of specific regions of the *F. oxysporum* lineage-specific and mobile genome, which was recorded together with whole-chromosome transfer by Ma et al. (2010). In terms of an ecological niche, such horizontal transfer could have taken place in the shared coffee tree niche that both species occupy. Numerous studies have co-isolated multiple *Fusarium* species from infected trees. The sequence identity between the *F. xylarioides* and *F. oxysporum* copies of these regions is not complete, i.e., less than 100% similar, which could either indicate an ancient transfer or that the donor sequence used in chapter three is not the donor itself, but possibly a close relative.

What this could mean for both coffee wilt disease and a far broader spectrum of disease is important. If closely-related fungal plant pathogens are able to frequently undergo recombination and the exchange of genetic material, it could mean the infinite adaptability of plant pathogens. Coffee growers in sub-Saharan Africa widely intercrop their coffee with banana crops, and both *F. xylarioides* and *F. oxysporum* have been co-isolated from the roots and wood of coffee wilt infected trees in Ethiopia and central Africa, as well as from banana roots in an intercropped field in Uganda. It could also

have important implications for weed management, if there are additional potential hosts for donor genetic material. Given the difficulties in treating coffee wilt disease once it is established (see section 1.4, management practices that reduce the risk of its re-emergence would be valuable to growers. It is widely viewed by the affected countries that coffee wilt disease has been dealt with, and indeed disease levels today are far lower than they once were. However, as was seen in the 1990s-2000s, coffee wilt can re-emerge and spread rapidly through the environment. Coffee wilt disease is considered to be an African coffee disease, but given that coffee germplasm from other coffee-growing parts of the world (Asia and south America) is also susceptible to *F. xylarioides*, then raising awareness of the threat from coffee wilt disease is needed. Care must be taken in the transportation of coffee germplasm globally to prevent accidental introduction and aid better preparedness in other coffee growing countries.

5.5 The importance of culture collections

None of this work would have been possible without the CABI-IMI and other culture collections around the world. Indeed, the fact that no strains isolated prior to 1950 were preserved has prevented this thesis from looking at the original origins of coffee wilt disease. The CABI collection was originally created as a taxonomic reference resource, yet today it holds tens of thousands of strains isolated from major disease outbreaks, as well as every major biome from around the world. The strains are 'optimally preserved' in a living condition, which means that high-quality genetic material can be obtained and, as found in chapter four, even their pathogenicity is maintained through decades of preservation. This thesis is but one example of how bioinformatic analyses could unlock the total diversity preserved within the CABI culture collection and additionally shows the value in continuing to preserve fungal strains into the future. There is immense value held within culture collections around the world which could have important future applications in managing microorganisms and learning lessons to help with future challenges.

Appendix A

Supplementary Materials

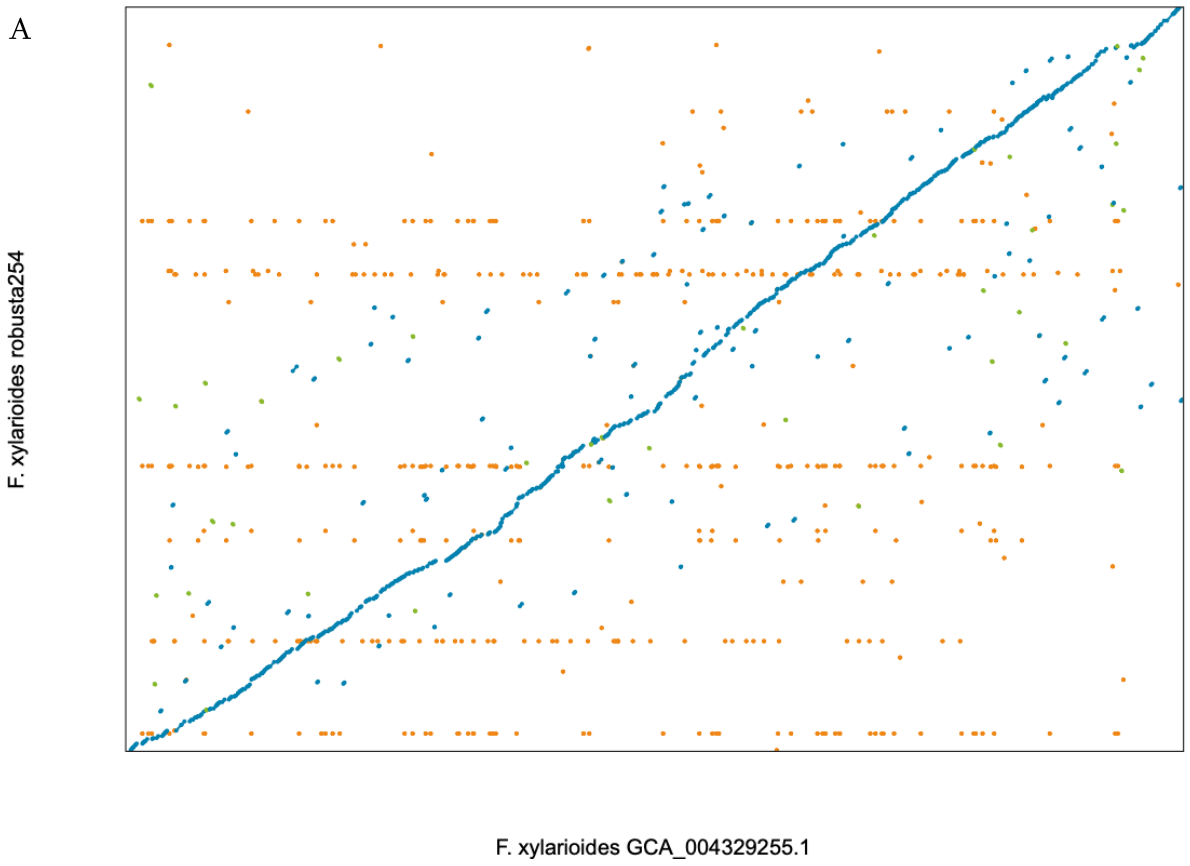


Figure A.1: Representative whole-genome alignments of *Fusarium xylarioides* robusta strains sequenced in this study with publicly available *Fusarium xylarioides* robusta strains. Each dot represents chromosomal correspondence, with absences representing absent chromosomes. Genomes were aligned using nucmer in the MUMmer3 package, with outputs processed using DotPrep.py and visualised using Dot in DNANexus. Blue indicates forward alignments, green indicates reverse alignments, orange indicates repetitive alignments.

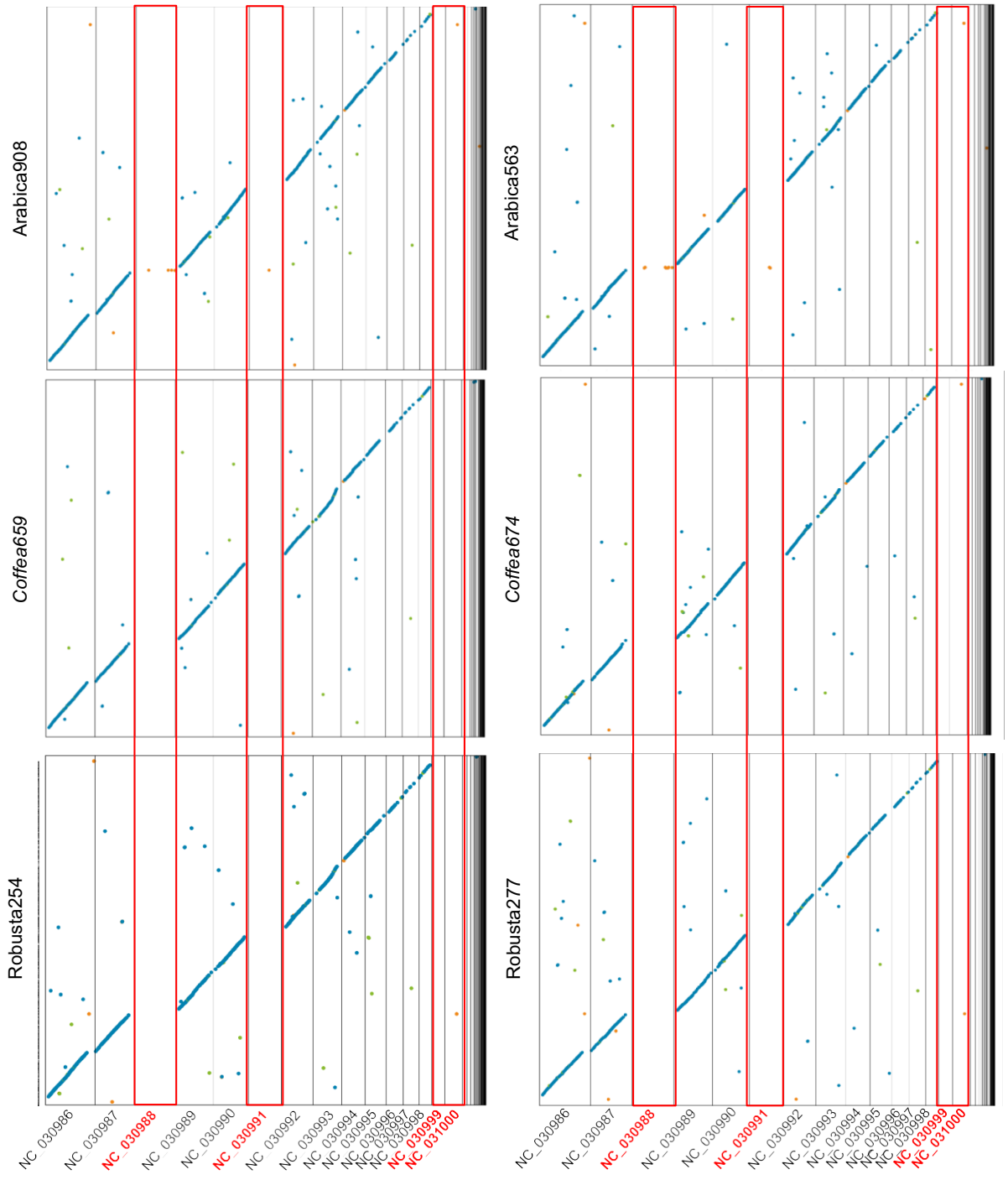


Figure A.2: Representative whole-genome alignments of *Fusarium xylarioides* strains against the 15 *Fusarium oxysporum f. sp. lycopersici* chromosomes. Each *Fusarium oxysporum f. sp. lycopersici* chromosome is labelled, with the four fully mobile chromosomes annotated in red: chromosomes 3, 6, 14 and 15. The remaining 11 chromosomes are the syntenic, core chromosomes shared with sister *Fusarium* species. Each dot represents chromosomal correspondence, with absences representing absent chromosomes. Genomes were aligned using nucmer in the MUMmer3 package, with outputs processed using DotPrep.py and visualised using Dot in DNANexus. Blue indicates forward alignments, green indicates reverse alignments, orange indicates repetitive alignments.

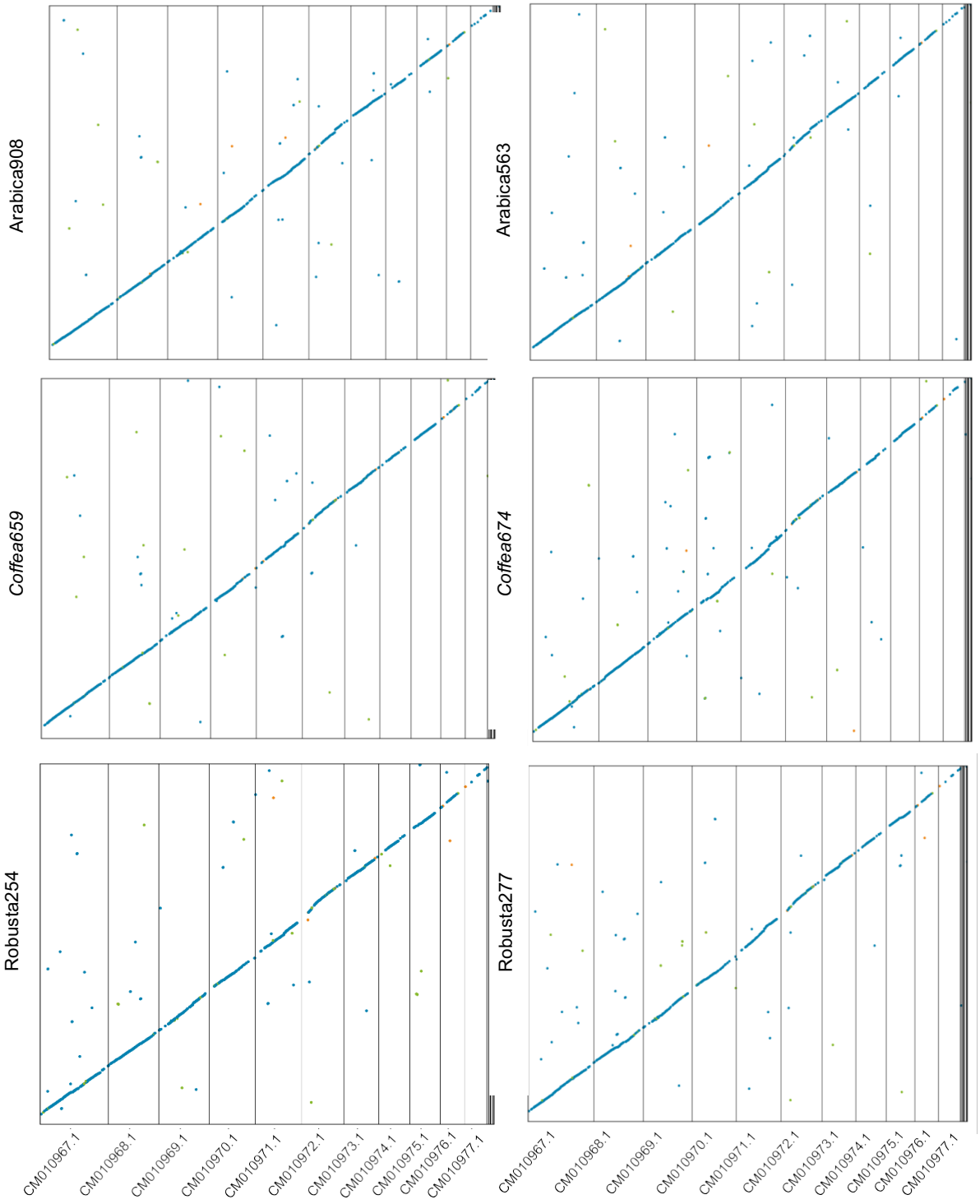


Figure A.3: Representative whole-genome alignments of *Fusarium xylarioides* strains against the 11 *Fusarium verticillioides* syntenic, core chromosomes shared with sister *Fusarium* species. Each dot represents chromosomal correspondence, with absences representing absent chromosomes. Genomes were aligned using nucmer in the MUMmer3 package, with outputs processed using DotPrep.py and visualised using Dot in DNANexus. Blue indicates forward alignments, green indicates reverse alignments, orange indicates repetitive alignments.

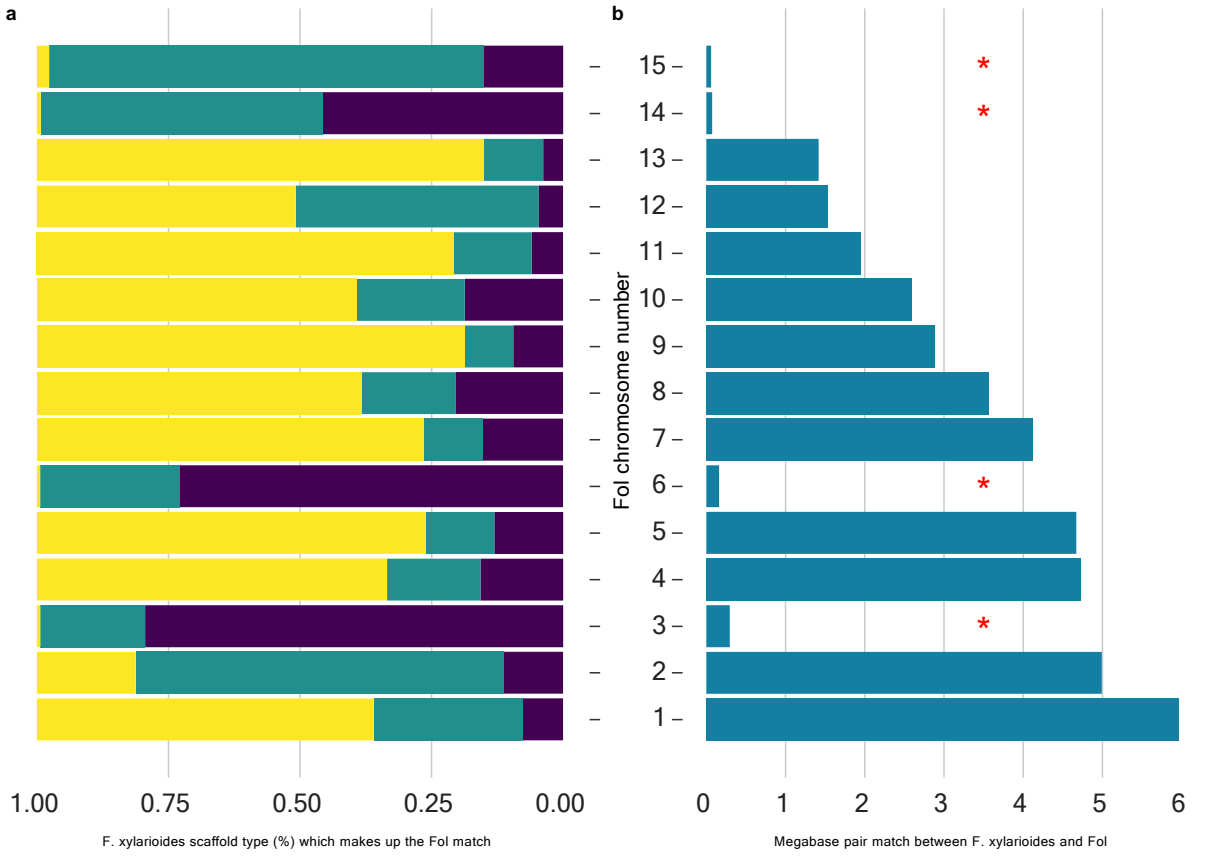


Figure A.4: ***Fusarium xylarioides*** **scaffolds mapped to** ***Fusarium oxysporum*** **f. sp.** ***lycopersici*** **chromosomes using reference-guided scaffolding. A** The type of *Fusarium xylarioides* scaffold group which match each *Fusarium oxysporum* f. sp. *lycopersici* chromosome: yellow, scaffolds which match to core chromosomes in the *Fusarium xylarioides* assembly mapped to *Fusarium verticillioides* (FV); green, scaffolds which match the *Fusarium xylarioides*- and *F. udum* (FXU) specific scaffolds (i.e. shared by these species and absent in *Fusarium verticllioides*); purple, scaffolds which match the Lineage Specific (LS) scaffolds and are not shared across the *Fusarium xylarioides* strains. **B** The total megabase pair match between *Fusarium xylarioides* scaffolds and *Fusarium oxysporum* f. sp. *lycopersici* chromosomes, with mobile chromosomes annotated with a red asterisk.

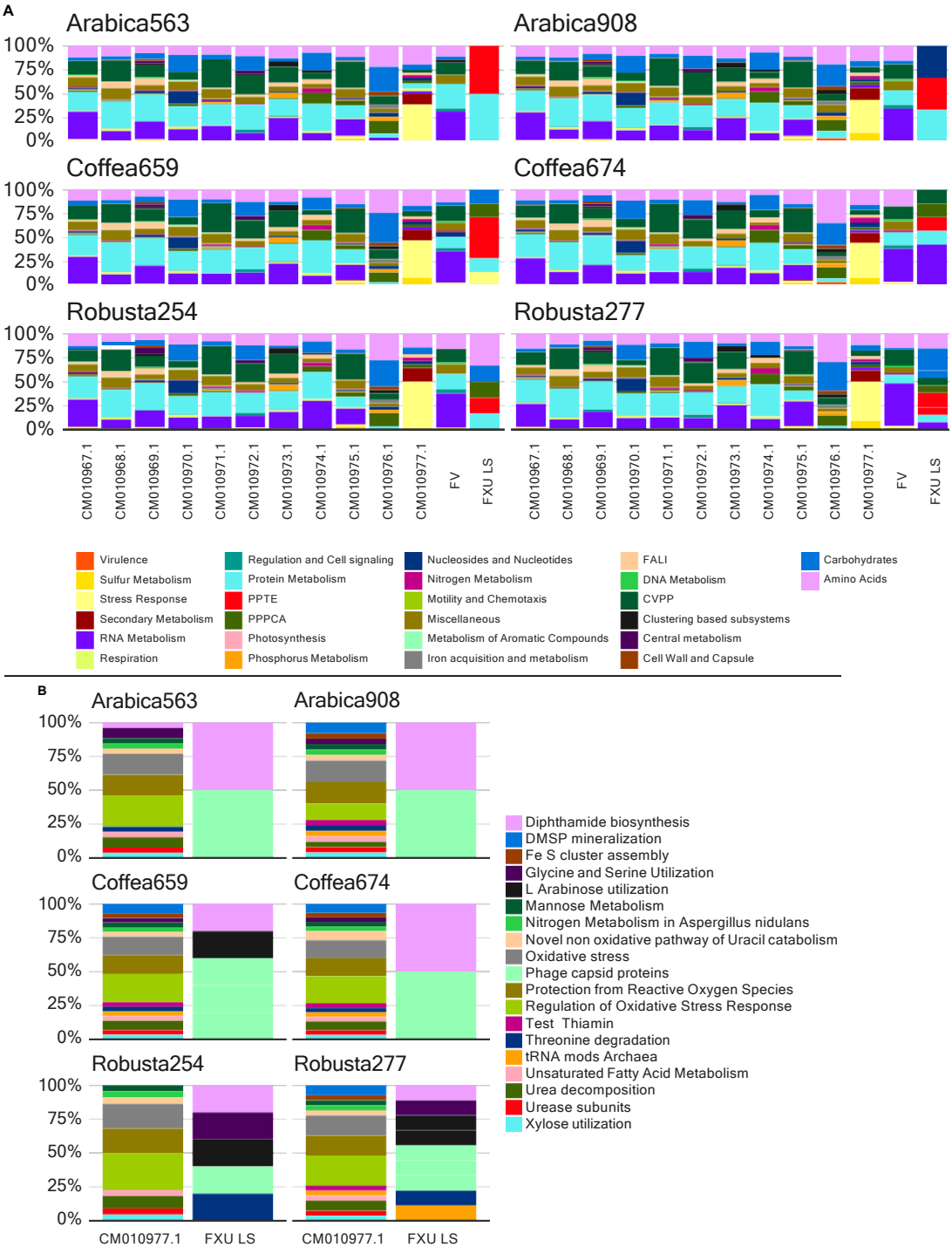


Figure A.5: Functional diversity of *Fusarium xylarioides* across the core chromosomes and by scaffold group: FV, scaffolds which match to un-aligned *Fusarium verticillioides* scaffolds; FXU LS, scaffolds which are absent in *Fusarium verticillioides* but which are shared by *Fusarium xylarioides*- and *F. udum* (FXU) or scaffolds which are unique to each *Fusarium xylarioides* strain and are lineage-specific (LS). A High level functional diversity across the 11 syntenic chromosomes and scaffold groups based on the number of hits to each Level 1 SEED category in SUPER-FOCUS (Silva et al. 2016); *B* Detailed functional diversity across chromosome 11 and the FXU and LS scaffold groups for *Fusarium xylaroides* based on the number of hits to each level 3 SEED category in SUPER-FOCUS. Abbreviations include PPCA, predictions based on plant prokaryote comparative analysis; CVPP, cofactors, vitamins, prosthetic groups & pigments; PPTE, phages, prophages & transposable elements; FALI, fatty acids, lipds and isoprenoids.

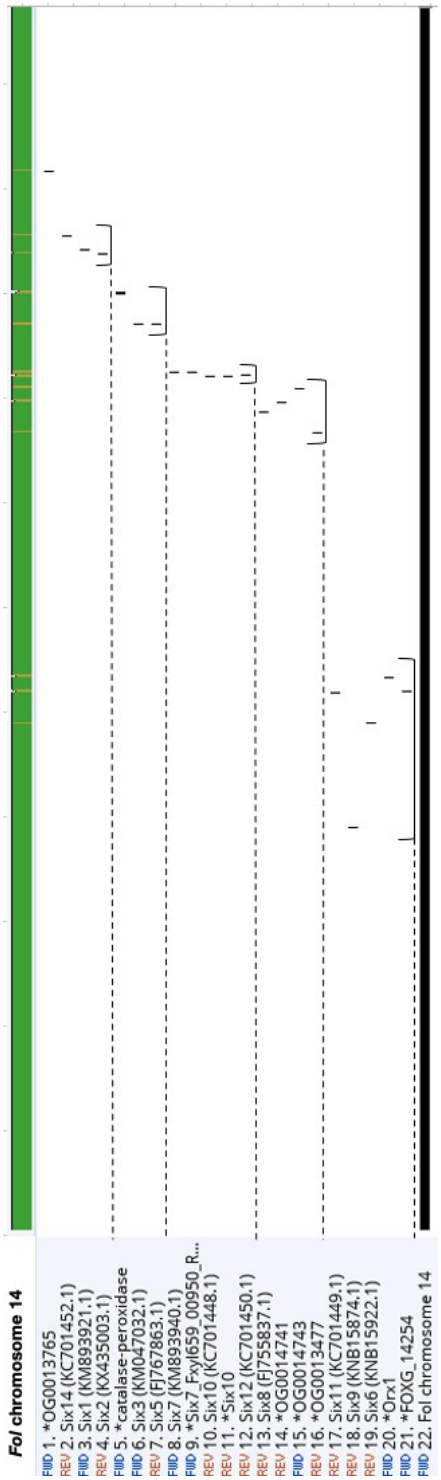
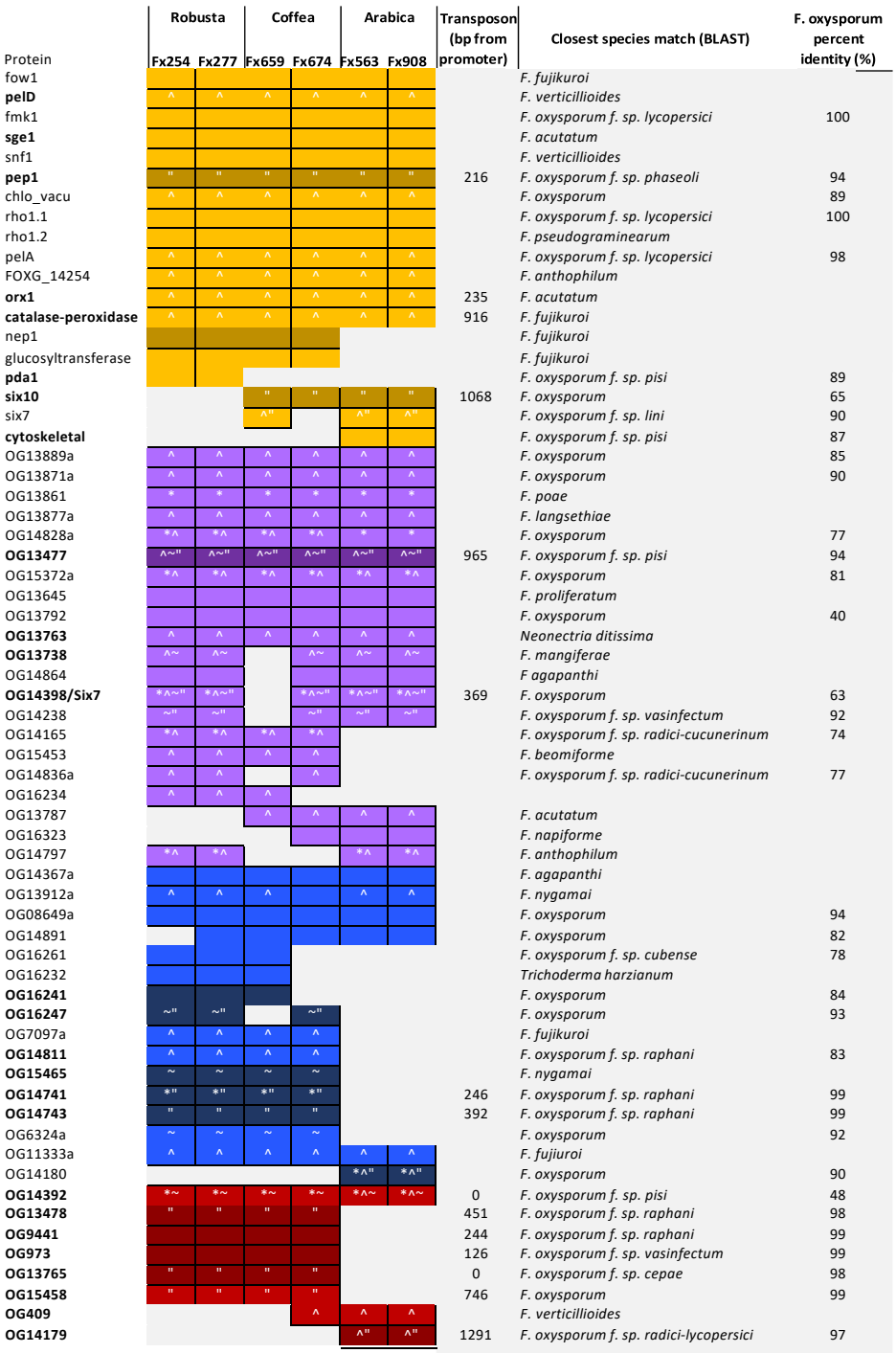


Figure A.6: ***Fusarium oxysporum*** **f. sp.** ***lycopersici*** **chromosome 14 aligned with all known SIX effectors and the nine effectors described in this study which match regions of chromosome 14 (pre-fixed with a *).** Five SIX chromosomal mini-clusters (described by Schmidt et al. (2013)) are marked on the genome plot. Eight effectors all reside close to four of these clusters: *FOXG 14254* and *Orx1* are 1.8 kb and 24 kb from *six11* (which is clustered with *six6* and five genes (including *orx1* in *Fol*) and a transcription factor); OG0013477, OG0014741 and OG0014743 are 30 kb, 12.5 kb and 33 kb respectively from *six8* (which resides in a solo block with class two transposons and a gene encoding an unknown protein); *six7* (from 659 00950) shares a locus with *Fusarium oxysporum* f. sp. *lycopersici six7* and its cluster with *six10* and *six12*. The orthogroups OG0014741 and OG0014743 are both also <40 kb from this *six10, six12, six7* minicluster. Finally, catalase-peroxidase is 45 kb from the *six3, six5* minicluster. OG0013765 is 95 kb from the *six1, six2, six14* minicluster. Sequences aligned with MAFFT and drawn in Geneious 9.1.



Protein	Robusta Fx254	Robusta Fx277	Coffea Fx659	Coffea Fx674	Arabica Fx563	Arabica Fx908	Transposon (bp from promoter)	Closest species match (BLAST)	F. oxysporum percent identity (%)
fow1								*F. fujikuroi*	
pelD	^	^	^	^	^	^		*F. verticillioides*	
fmk1								*F. oxysporum f. sp. lycopersici*	100
sge1								*F. acutatum*	
snf1								*F. verticillioides*	
pep1	"	"	"	"	"	"	216	*F. oxysporum f. sp. phaseoli*	94
chlo_vacu	^	^	^	^	^	^		*F. oxysporum*	89
rho1.1								*F. oxysporum f. sp. lycopersici*	100
rho1.2								*F. pseudograminearum*	
pelA	^	^	^	^	^	^		*F. oxysporum f. sp. lycopersici*	98
FOXG_14254	^	^	^	^	^	^		*F. anthophilum*	
orx1	^	^	^	^	^	^	235	*F. acutatum*	
catalase-peroxidase	^	^	^	^	^	^	916	*F. fujikuroi*	
nep1								*F. fujikuroi*	
glucosyltransferase								*F. fujikuroi*	
pda1								*F. oxysporum f. sp. pisi*	89
six10			"	"	"	"	1068	*F. oxysporum*	65
six7			^"		^"	^"		*F. oxysporum f. sp. lini*	90
cytoskeletal								*F. oxysporum f. sp. pisi*	87
OG13889a	^	^	^	^	^	^		*F. oxysporum*	85
OG13871a	^	^	^	^	^	^		*F. oxysporum*	90
OG13861	*	*	*	*	*	*		*F. poae*	
OG13877a	^	^	^	^	^	^		*F. langsethiae*	
OG14828a	*^	*^	*^	*^	*	*		*F. oxysporum*	77
OG13477	^~"	^~"	^~"	^~"	^~"	^~"	965	*F. oxysporum f. sp. pisi*	94
OG15372a	*^	*^	*^	*^	*^	*^		*F. oxysporum*	81
OG13645								*F. proliferatum*	
OG13792								*F. oxysporum*	40
OG13763	^	^	^	^	^	^		*Neonectria ditissima*	
OG13738	^~	^~		^~	^~	^~		*F. mangiferae*	
OG14864								*F agapanthi*	
OG14398/Six7	*^~"	*^~"		*^~"	*^~"	*^~"	369	*F. oxysporum*	63
OG14238	~"	~"		~"	~"	~"		*F. oxysporum f. sp. vasinfectum*	92
OG14165	*^	*^	*^	*^				*F. oxysporum f. sp. radici-cucunerinum*	74
OG15453	^	^	^	^				*F. beomiforme*	
OG14836a	^	^		^				*F. oxysporum f. sp. radici-cucunerinum*	77
OG16234	^	^	^						
OG13787			^	^	^	^		*F. acutatum*	
OG16323								*F. napiforme*	
OG14797	*^	*^			*^	*^		*F. anthophilum*	
OG14367a								*F. agapanthi*	
OG13912a	^	^	^		^	^		*F. nygamai*	
OG08649a								*F. oxysporum*	94
OG14891								*F. oxysporum*	82
OG16261								*F. oxysporum f. sp. cubense*	78
OG16232								*Trichoderma harzianum*	
OG16241								*F. oxysporum*	84
OG16247	~"	~"		~"				*F. oxysporum*	93
OG7097a	^	^	^	^				*F. fujikuroi*	
OG14811	^	^	^	^				*F. oxysporum f. sp. raphani*	83
OG15465	~	~	~	~				*F. nygamai*	
OG14741	*"	*"	*"	*"			246	*F. oxysporum f. sp. raphani*	99
OG14743	"	"	"	"			392	*F. oxysporum f. sp. raphani*	99
OG6324a	~	~	~	~				*F. oxysporum*	92
OG11333a	^	^	^	^	^	^		*F. fujiuroi*	
OG14180					*^"	*^"		*F. oxysporum*	90
OG14392	*~	*~	*~	*~	*^~	*^~	0	*F. oxysporum f. sp. pisi*	48
OG13478	"	"	"	"			451	*F. oxysporum f. sp. raphani*	98
OG9441							244	*F. oxysporum f. sp. raphani*	99
OG973							126	*F. oxysporum f. sp. vasinfectum*	99
OG13765	"	"	"	"			0	*F. oxysporum f. sp. cepae*	98
OG15458	"	"	"	"			746	*F. oxysporum*	99
OG409				^	^	^		*F. verticillioides*	
OG14179					^"	^"	1291	*F. oxysporum f. sp. radici-lycopersici*	97

Figure A.7: Putative effectors' characteristics and presence or absence across *Fusarium xylarioides* strain and *F. udum*. The four effector classes are shown in: yellow for predefined effectors; purple for small and cysteine-rich secreted effectors; blue for carbohydrate-active enzymes; and red for transposon-adjacent effectors. The symbols highlight: the presence of transposons is represented by names in bold with its distance from the genes promoter described if less than 1500bp (if not, the transposon is over 1500bp away); genes under positive selection by an asterisk; genes in an AT-rich region by a tilde; genes with evidence of horizontal transfer from *Fusarium oxysporum* are a darker shade; genes which are absent from more closely-related *Fusarium* species (namely *F. graminearum*, the Asian clade *Fusarium fujikuroi* complex species and *Fusarium verticillioides* - *F. solani* and *F. udum* were excluded here because *F. solani* also infects coffee ((Rutherford 2005) and thus could be a source of pathogenicity and *F. udum* is also a vascular wilt-inducer) and *Fusarium oxysporum* is the closest match with a percent identity (%) >=90 are represented by a quotation mark; and closest species is shown for each protein with its percent identity (%), where a BLASTp hit was returned.

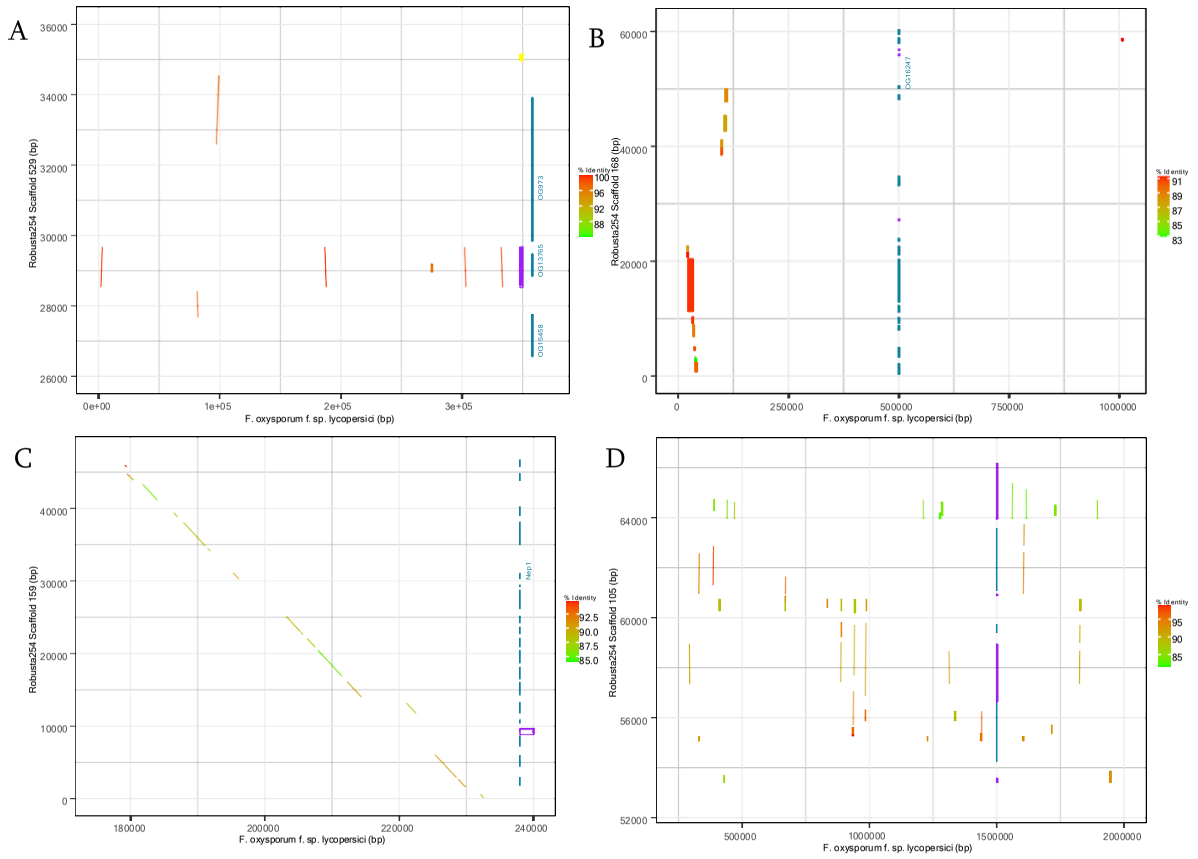


Figure A.8: Representative alignments of *Fusarium xylarioides* robusta254 scaffolds which contain the putative 15 effectors under horizontal gene transfer against the genome of *Fusarium oxysporum* f. sp. *lycopersici* . Line colour indicates the alignment percentage id match, x and y axes have been adjusted to display regions which match. Scaffolds were aligned using nucmer in the MUMmer3 package, with outputs processed in RStudio. An annotated blue line indicates a gene, an annotated and labelled blue line indicates a putative effector gene, yellow boxes indicate *mimps* and purple boxes indicate class 1 and class 2 transposable elements. **A** Scaffold 529 (35kb long) with three putative HGT effectors: OG0015458, OG0013785 and OG000973; a *Hop3* DNA transposon and a *mimp*; **B** Scaffold 168 (79 kb) with 20 genes including one putative HGT effector OG16247, a *hAT* DNA transposon and two *Copia* retrotransposons; **C** Scaffold 159 (80 kb) with 18 genes including the putative HGT effector Nep1 and two *TcMar-Tc1* DNA transposons; **D** Scaffold 105 (96 kb) with three genes, four DNA transposons (*PiggyBac, TcMar-Tc1, MGR583-like, Fot6*.

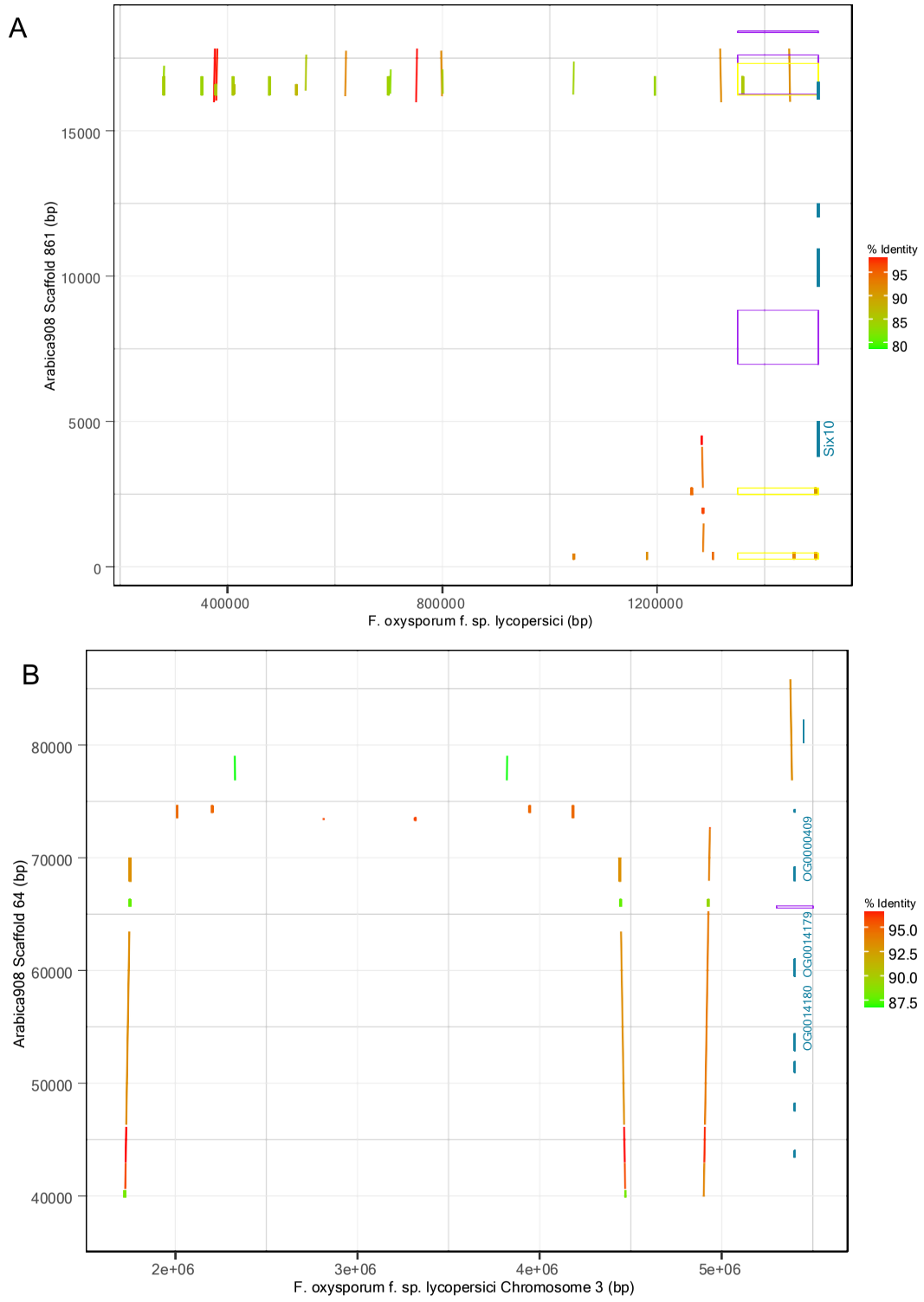


Figure A.9: Representative alignments of *Fusarium xylarioides* arabica908 scaffolds which contain four of the effectors, unique to arabica and one, both or neither of the coffea strains, under horizontal gene transfer, plotted against the genome of *Fusarium oxysporum f. sp. lycopersici* . Line colour indicates the alignment percentage id match, x and y axes have been adjusted to display regions which match. Scaffolds were aligned using nucmer in the MUMmer3 package, with outputs processed in RStudio. An annotated blue line indicates a gene, an annotated and labelled blue line indicates a putative effector gene, yellow boxes indicate *mimps* and purple boxes indicate class 1 and class 2 transposable elements. **A** Scaffold 861 (20kb long) with the Six10 putative HGT effector, two *TcMar-Tc1* class II DNA transposons, one *Copia* retrotransposon and four *mimps*. **B** Scaffold 64 (93 kb long) with three putative HGT effectors: OG14180, OG14179, OG409 and a *Yaret2* class II DNA transposon.

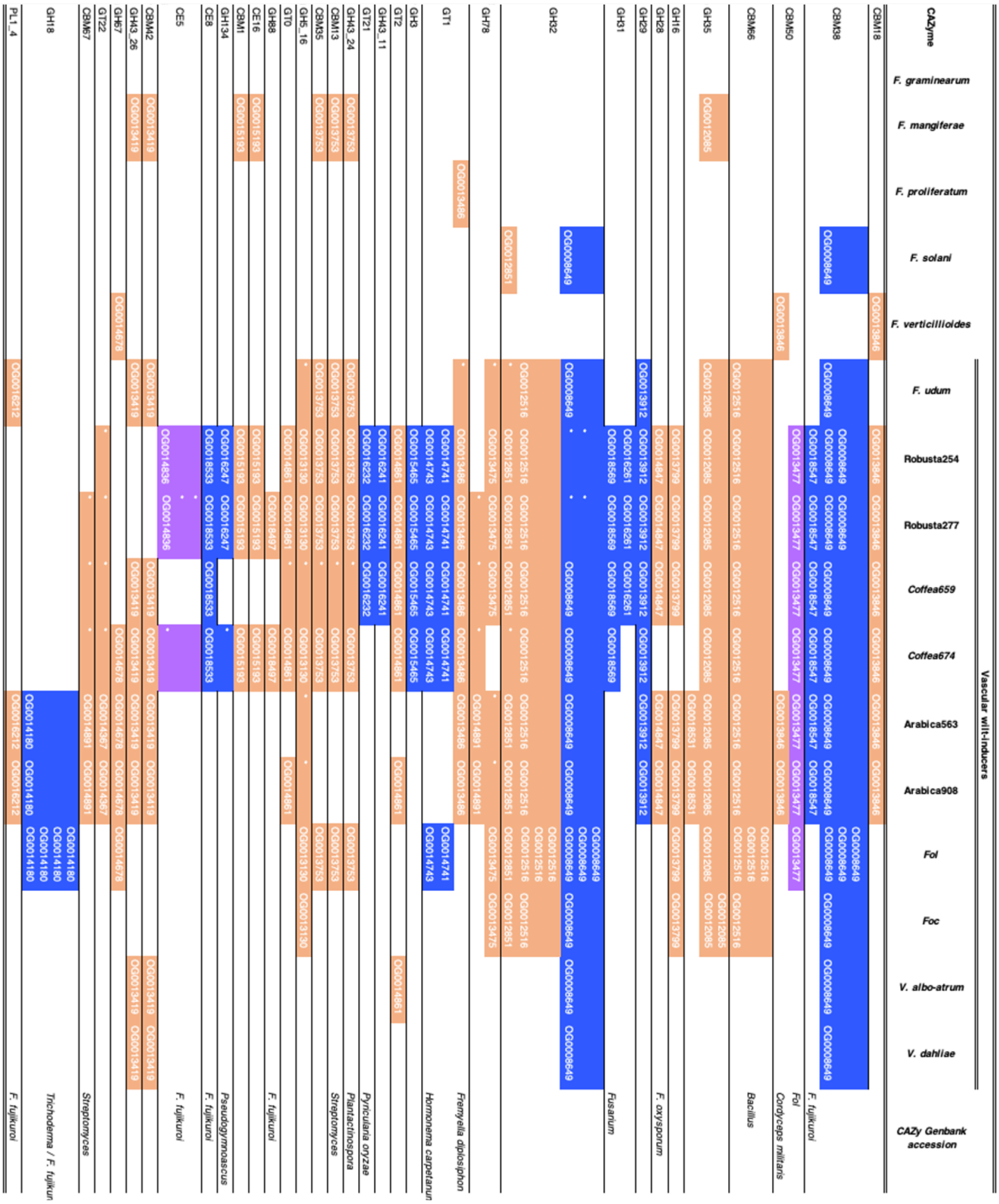


Figure A.10: Gene copy number for carbohydrate-active enzyme-encoding orthologous groups shared across the vascular wilt-inducing *Fusarium* and *Verticillium*. Groups which also included one non-vascular wilt inducer were additionally included, and those which are also a putative effector are shaded the same colour as in figure 5. Where a species has a gene in the orthologous group which is not recognised as a carbohydrate-active enzyme is represented with an asterisk.

Table A.1: Genome statistics for the *Fusarium xylarioides* strains sequenced in this study, compared with sister species. Abbreviations for *Fusarium* sister species: *Fol, Fusarium oxysporum f. sp. lycopersici; Fv, Fusarium verticillioides*. Total repeats includes unclassified repeats. Interspersed repeats include retroelements, DNA transposons, simple and low complexity repeats.

Name	Coffea674	Coffea659	Robusta277	Robusta254	Arabica563	Arabica908	Robusta925	Robusta394	*F. udum*	*Fol*	*Fv*
Strain number	IMI392674	IMI127659i	IMI392277	IMI392254	IMI389563	IMI375908	iIMI379925	FRC L-0394	F02845	4287	FGSC 7600
Reference	This study	This study	This study	This study	This study	This study	Olal et al. 2019	Wingfield et al. 2019	Srivastava et al. 2018	Ma et al. 2010	Ma et al. 2010
Date isolated	1951	1955	2003	1997	2002	1997	1998	2000	2010		
Origin	Cote d'Ivoire	CAR	Tanzania	Uganda	Ethiopia	Ethiopia	Uganda	Uganda	India		
Host	*Coffea*	*C. excelsa*	*C. c.* robusta	*C. c.* robusta	*C. arabica*	*C. arabica*	*C. c.* robusta	*C. canephora*	Pigeonpea	Tomato	Maize
Size (genome assembly) Mb	57.2	59.4	61.3	60.3	63.3	62.6	55.1	55.2	56.4	61.4	42.5
Genome size (Mb, ¿500bp)	54.4	55.1	56.5	56.3	58.0	57.8	55.1	55.2	56.4	61.4	42.5
Total repeats (genome assembly) Mb	13.3	14.9	15.9	15.3	18.6	17.8			12.6	10.8	1.0
Total repeats %	23	25	26	25	30	29			22	18	2
Interspersed repeats Mb	5.8	4.9	5.8	4.3	5.0	5.4			18.9	6.4	0.5
Interspersed repeats %	10	8	9	7	8	9			33	10.5	1
Interspersed repeats (raw reads) Mb	3.9	3.7	3.9	3.8	4.4	3.9				2.4	

Table A.2: Genome statistics for our strains mapped to the chromosomal assembly of *Fusarium verticillioides* using reference-guided scaffolding. Abbreviations: FV, the contigs in each genome mapped to the syntenic chromosomes and un-aligned scaffolds of *Fusarium verticillioides*; FXU, scaffolds which are absent from *Fusarium verticillioides* but which are present in *F. udum* and the historic coffea659 strain (*Fusarium xylarioides* and *-udum* specific); FXS, scaffolds which are absent from *Fusarium verticillioides* and *F. udum* and are shared with coffea659 (*Fusarium xylarioides*-specific); LS, scaffolds which are not shared with coffea659 and are unique to each *Fusarium xylarioides* strain (lineage-specific).

Strain	IMI392674	IMI127659i	IMI392277	IMI392254	IMI389563	IMI375908i
Name	Coffea674	Coffea659	robusta277	robusta254	arabica563	arabica908
Genome Size Mb	57.2	59.6	61.3	60.3	63.4	62.6
Total repeats incl unclassified Mb	13.3	14.9	15.9	15.3	18.6	17.8
Total repeats incl unclassified %	23	25	26	25	30	29
TEs Mb	5.8	4.9	5.8	4.3	5.0	5.4
TEs %	10	8	9	7	8	9
FV scaffolds Mb	51	49	51.6	51.3	52.4	52.3
FXU scaffolds Mb	4.5	7.9	7.5	7.2	8.9	7.7
FXS scaffolds Mb	0.04	-	0.06	0.04	0.04	0.5
LS scaffolds Mb	1.9	2.2	2.3	1.8	2.1	2.1
FV scaffolds %	7	5	5	5	6	6
FXU scaffolds %	8	13	12	12	14	12
FXS scaffolds %	0.1	-	0.1	0.1	0.1	0.8
LS scaffolds %	3	4	4	3	3	3
FV TEs Mb	3.6	2.4	2.8	2.4	2.9	3.2
FXU TEs Mb	1.2	1.9	1.8	1.4	1.6	1.7
LS TEs Mb	1.0	0.6	0.6	0.4	0.5	0.4
FV TEs %	7	5	5	5	6	6
FXU TEs (%)	28	24	23	20.1	18	22
LS TEs (%)	52	26	28	24	23	20

Table A.3: Published genomes used for comparison

Genome	Accession number
F. udum	GCA_002194535.1
F. oxysporum f.sp. lycopersici	GCA_000149955.2
F. oxysporum f.sp. cubense	GCA_005930515.1
F. verticillioides	GCA_003316975.2
F. fujikuroi	GCA_900079805.1
F. mangiferae	GCA_900044065.1
F. solani	GCA_002215905.1
Verticillium dahliae	GCA_000150675.2
V. albo-atrum	GCA_002851705.1
F. graminearum	GCA_000240135.3
F. proliferatum	GCF_900067095.1
F. xylarioides IMI379925	GCA_004329255.1
F. xylarioides FRC L-0394	GCA_013183765.1

Table A.4: Predefined effector protein genes analysed in this study. Genes marked with an * show predicted roles and locations only. Abbreviations for *F.oxysporum* formae speciales sister species: *Fol, Fusarium oxysporum* f. sp. *lycopersici* ; *Foe, Fusarium oxysporum* f. sp. *erythroxyli* ; *For, Fusarium oxysporum* f. sp. *ricini*

Name	Role	Query Accession	Query species	Length	Reference
FOXG_02706.2	Glucosyltransferase*	KNA98333.1	*Fol*	1493	Klosterman et al. (2011)
FOXG_10732.2	Cytoskeletal*	KNB10567.1	*Fol*	449	Klosterman et al. (2011)
FOXG_04660.2	Chloroplast/ vacuole*	KNB01401.1	*Fol*	797	Klosterman et al. (2011)
Nep1	Microbial elicitors of plant necrosis	AF036580.1	*Foe**	2617	Pemberton & Salmond (2004)
Fmk1	MAP kinase	KC257048.1	F. oxysporum	603	Di Pietro et al. (2001)
Fow1	Mitochondrial carrier protein	KC134256.1	F. oxysporum	725	Inoue et al. (2002)
Pda1	Pisatin demethylase	KR855811.1	F. oxysporum	455	Wasmann & VanEtten (1996)
PelA	Pectate lyase	MK918256.1	*Fol*	539	Rogers et al. (2000)
PelD	Pectate lyase	KC294608.1	F. proliferatum	552	Rogers et al. (2000)
Pep1	Pea pathogenicity protein	EU436568.1	Fusarium sp.	216	Han et al. (2001)
Rho1	Rho GTP-ase activating protein	KC017411.1	F. oxysporum	665	Martínez-Rocha et al. (2008)
Sge1	SIX (secreted in xylem) gene expression 1	LC369105.1	*For**	565	Michielse & Rep (2009)
Snf1	Protein kinase sucrose non-fermenting	KU048959.1	F. commune	625	Ospina-Giraldo et al. (2003)
FOXG_14254	Conserved secreted protein	KNB15932.1	*Fol*	1592	Ma et al. (2010)
Orx1	In-planta secreted oxidoreductase enzyme	KNB15937.1	*Fol*	1860	Ma et al. (2010)
Catalase-peroxidase	Secreted enzyme	KNB19974.1	*Fol*	2385	Ma et al. (2010)
SIX10	Secreted in xylem 10	KNB20462.1	*Fol*	736	Ma et al. (2010)

Table A.5: Enriched CAZyme gene familes across *Fusarium* species (*Fusarium xylarioides, F. udum, F. oxysporum* f. sp. *lycopersici, F. verticillioides, F. fujikuroi, F. graminearum*) genomes, compared with three different ascomycete fungi (*Trichoderma reesei, Aspergillus nigris* and *Magnaporthe grisea*). Abbreviations: AA, Auxilliary Activities; CBM, Carbohydrate-Binding Modules; CE, Carbohydrate Esterase; GH, Glycoside Hydrolases; GT, Glycosyltransferases; PL, Pectate Lyases

Species	AA*	CBM*	CE*	GH*	GT*	PL*
Fol	165	330	70	746	396	35
F. xylarioides (coffea674)	111	232	62	488	248	28
F. udum	119	228	63	503	257	30
F. verticillioides	130	265	66	597	336	28
F. fujikuroi	105	220	61	476	274	29
F. graminearum	92	191	52	394	241	27
Trichoderma reesei	57	127	25	304	196	8
Aspergillus nigris	94	151	55	434	321	10
Magnaporthe grisea	118	207	60	378	258	9

Table A.6: Accession numbers and source details for each *impala*, *miniature impala (mimp)* and newly described class II transposable elements

Accession	Transposon
AF076624.1	F. o. repetitive element mimp1
AF076625.1	F. o. repetitive element mimp2
EU833100.1	F. o. f. sp. melonis MITE mimp3
EU833101.1	F. o. f. sp. lycopersici MITE mimp4
AF282722.1	F. o. f. sp. melonis transposon impala transposase gene
AF363407.1	F. o. f. sp. melonis transposon impala M24-impE
AF363412.1	F. o. f. sp. lini transposon impala Ln3-1
AF363413.1	F. o. f. sp. lini transposon impala Ln88-23
AF363414.1	F. o. f. sp. cubense transposon impala Cu-12
AF363416.1	F. o. f. sp. phaseoli transposon impala Ph-5
AF363417.1	F. o. f. sp. phaseoli transposon impala Ph-9
AF363418.1	F. o. f. sp. albedinis transposon impala A-33
AF363419.1	F. o. f. sp. soil transposon impala S47-35
AF363420.1	F. o. f. sp. raphani transposon impala R-8
AF363425.1	F. o. f. sp. melonis transposon impala M24-impD
AF363426.1	F. o. f. sp. melonis transposon impala MK14
AF363427.1	F. o. f. sp. lini transposon impala Ln88-8
AF363430.1	F. o. f. sp. lycopersici transposon impala L15delta5
AF363432.1	F. o. f. sp. lycopersici transposon impala L15-15
AF363433.1	F. o. f. sp. radicis-lycopersici transposon impala RL28-17
AF363434.1	F. o. f. sp. lini transposon impala Ln86-10
AF363435.1	F. o. f. sp. ciceris transposon impala Ci-36
AF363436.1	F. o. f. sp. ciceris transposon impala Ci-16
AF363437.1	F. o. f. sp. melonis transposon impala MK28
AF363438.1	F. o. f. sp. lycopersici transposon impala L15-16
AJ608703.3	F. o. f. sp. lycopersici shh1 gene
AJ608703.3	F. o. f. sp. lycopersici fot5 gene
JX204302.1	F. o. f. sp. fragariae transposon Impala_1
Schmidt et al 2013	FoCrypton
Schmidt et al 2013	FoHelitron
Schmidt et al 2013	Fot6
Schmidt et al 2013	Fot8
Schmidt et al 2013	Hop3
Schmidt et al 2013	Hop6
Schmidt et al 2013	MGR583-like
Schmidt et al 2013	Nht2-like
Schmidt et al 2013	YahAT4
Schmidt et al 2013	YahAT6
Schmidt et al 2013	Yaret1
Schmidt et al 2013	Yaret2

Table A.7: **The median overlap distance to transposable elements and large RIP affected areas for putative effectors and the same number of random genes (randomisation trials, repeated 1000 times, * = p ¡ 0.05) for each *Fusarium xylarioides* strain.**

	Random genes		Putative effectors		P-value	
	Transposons (kb)	LRAR (kb)	Transposons (kb)	LRAR (kb)	Transposons (kb)	LRAR (kb)
Robusta254	6.43	22.39	4.00	15.16	p ¿ 0.05	*
Robusta277	5.79	22.45	2.48	16.69	*	p ¿ 0.05
Coffea659	6.17	24.94	3.00	14.07	p ¿0.05	p ¿ 0.05
Coffea674	7.41	26.99	2.97	22.50	*	p ¿ 0.05
Arabica563	5.26	18.54	2.16	7.51	*	*
Arabica908	4.59	18.14	3.66	6.46	p ¿ 0.05	*

Table A.8: Investigating the support for horizontal acquisition for each putative effector gene in *Fusarium xylarioides (Fx)* with *Fusarium oxysporum (Fo)* as the source of pathogenicity. Following the decision tree in figure (*), this table reports the outcome for each stage of the decision tree. The stages are: is the effector present in other *Fusarium fujikuroi* complex species (Y / N); if Y, does Fx nest with the *Fusarium fujikuroi* complex gene copies (Y / N); if N, does Fx nest with Fo thus disrupting the Fo phylogeny (Y / N); branch support values (BSV) for Fx with Fo (n); if N, does Fx nest with Fo by distance i.e. less distance from Fx to Fo than greatest distance from Fo to Fo; pairwise id % for whole branch with support value for Fx nested with Fo; the HGT class that we assign, classes 2 and 3 display evidence of HGT. FFC = *Fusarium fujikuroi* complex that *Fusarium xylarioides* belongs to.

Effector	Effector present in FFC	Fx nested with FFC	BSV	Fx nested in Fo by phylogeny	BSV	Fx nested in Fo by distance	Pairwise id %	Class
fow1	Y	Y		N				1
pelD	Y	Y		N				1
fmk1	Y	Y		N				1
sge1	Y	Y		N				1
snf1	Y	Y		N				1
pep1	N	N		Y	91		94	3a
chlo_vacu	Y	Y		N				1
rho1.1	Y	Y		N				1
rho1.2	Y	Y		N				1
pelA	Y	Y		N				1
FOXG_14254	Y	Y		N				1
orx1	Y	Y		N				1
catalase-peroxidase	Y	Y		N				1
nep1	Y	N		Y	99		92	2
gluco	Y	Y		N				1
pda1	Y	Y		N				1
six10	N	N		N	100	Y	84	3b
six7	N	N		N			91	4
cytoskeletal	Y*	Y						1
OG13899	Y	Y*	100	N				1
OG13871	Y	Y		N				1
OG13861	Y	Y		N				1
OG13877	Y	Y*		N				1
OG14828	Y	Y		N				1
OG13477	N	N		Y	100		84.5	3a
OG15372	Y	Y		N				1
OG13645	N	no blast match		N				5
OG13792	Y	Y		N				1
OG13763	N	no blast match		N				5
OG13738	Y	Y		N				1
OG14864	Y	Y		N				1
OG14398	N	N					84.4	4
OG14238	N	N		N			88.9	4

OG14165	Y	Y*		N				1
OG15453	N	no blast match		N				5
OG16234	N	no blast match		N				5
OG13787	Y	Y*	80	N				1
OG16323	N	no blast match		N				5
OG14797	Y	Y*	99	N				1
OG14367	N	no blast match		N				5
OG13912	Y	Y						1
OG08649	Y	Y*		Y	100		88.7	1
OG14891	Y	Y		N				1
OG16261	N	Weak Fo blast hit		N				5
OG16232	Y	Y		N				1
OG16241	Y	N		Y	100		89.1	2
OG16247	N	N		N		Y	80.4	3b
OG7097	Y	Y		N				1
OG14811	Y	Y						1
OG15465	Y	N		Y	100		88	2
OG14741	N	N		Y	100	Y	90.1	3a
OG14743	N	N		N		Y	97.9	3b
OG16212	Y	Y						1
OG006324	Y	Y		N				1
OG11333	Y	Y		N				1
OG14180	N	N		N		Y	95.6	3b
OG14392	Y	Y*	46	N				1
OG13478	N	N		N	100	Y	98.8	3b
OG09441	Y	Y/N		N				2
OG00973	Y	N		Y	100		83.6	2
OG13765	N	N		Y	100		86.5	3b
OG15458	N	N		N				4
OG00409	Y	Y						1
OG14179	N	N		Y	99		82	3a

Table A.9: Genomes sequenced in this thesis. The columns describe: species, short species with strain number, type of sequencing completed, RNA-seq data used as evidence in gene prediction, expressed sequence tag (EST) data used as evidence in gene prediction, whether the genome was used in Nsimscan for whole-genome similarity, which publicly available assembly the genome was scaffolded to using RagTag, whether the genome was used as a reference in looking for horizontal transfers.

Species	Short name	Sequencing type	Funannotate: RNA-Seq data	Funannotate: EST data	Nsimscan	RagTag scaffolded to	HGT reference genomes
F. oxysporum f. sp. cubense	Foxy_141109	Illumina	ERR10015931	*F. oxysporum f.sp. cubense II5 v2.0*	Y	GCA_005930515.1	RagTag assembly
F. oxysporum f. sp. coffea	Foxy_244509	Illumina	SRR10123445	*F. oxysporum f.sp. cubense II5 v2.0*	Y	GCF_000149955.1	RagTag assembly
F. oxysporum f. sp. vasinfectum	Foxy_292248	Illumina	SRR12855423	*F. oxysporum f.sp. cubense II5 v2.0*	Y	GCA_000260175.2	RagTag assembly
F. oxysporum f. sp. raphani	Foxy_337541	Illumina	SRR14318405	*F. oxysporum f.sp. cubense II5 v2.0*	Y	GCA_019157275.1	RagTag assembly
F. solani	Fsol_392280	Illumina	SRR19182467	*F. solani FSSC 5 v1.0 from (Mesny et al. 2021)*	Y	GCA_020744495.1	RagTag assembly
F. oxysporum f. sp. pisi	Foxy_500221	Illumina	SRR12855423	*F. oxysporum f. sp. pisi F23 v1.0*	Y	GCA_000260075.2	RagTag assembly
F. xylarioides	Fxyl_127659	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_375908	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_379925	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_389563	Long-read & Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_392254	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_392268	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_392277	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_392674	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_392676	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_507035	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_507038	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_507113	Illumina	Arabica563		Y	Arabica563 reference	
F. xylarioides	Fxyl_L0394	Illumina	Arabica563		Y	Arabica563 reference	

Table A.10: Publicly available genomes analysed in this thesis. The columns describe: species, short species with strain number, accession number, whether the genome was used in OrthoFinder, whether the genome was used in Nsimscan for whole-genome similarity, which publicly available assembly the genome was scaffolded to using RagTag, which genome was used as a reference in looking for horizontal transfers.

Species	Short name	Published genomes	Orthofinder	Nsimscan	HGT reference
F. verticillioides	Fver_7600	GCA_000149555.1	Y	Y	
F. graminearum	Fgram_PH1	GCA_000240135.3	Y	Y	
F. proliferatum	Fpro_01257	GCA_900067095.1	Y	Y	
F. oxysporum f. sp. lycopersici	Foly_MN25	GCA_000259975.2	Y		
F. oxysporum f. sp. pisi	Fopi_HDV247	GCA_000260075.2	Y		
F. oxysporum f. sp. vasinfectum	Fova_25433	GCA_000260175.2	Y		
F. oxysporum f. sp. cubense	Focu_race1	GCA_000350345.1	Y		
F. fujikuroi	Ffuj_LW94	GCA_001023035.1	Y	Y	
F. fujikuroi	Ffuj_Y057	GCA_001023045.1	Y		
F. verticillioides	Fver_14953	GCA_003316975.2	Y		
F. oxysporum f. sp. cubense	Focu_b16	GCA_005930515.1	Y		
F. oxysporum f. sp. cubense	Focu_TR4	GCA_007994515.1	Y	Y	
F. oxysporum 47	Fo_47	GCA_013085055.1	Y	Y	
F. anthophyllum	Fant_25214	GCA_013364935.1	Y	Y	
F. phyllophylum	Fphy_13617	GCA_013396025.1	Y	Y	
F. proliferatum	Fpro_13237	GCA_017309865.1	Y		
F. oxysporum f. sp. raphani	Foraph_Tf1262	GCA_019157275.1	Y		
F. oxysporum f. sp. rapae	Forapa_Tf1208	GCA_019157295.1	Y		
F. solani	Fsol_362657	GCA_020744495.1	Y		
F. oxysporum f. sp. lycopersici	Foly_4287	GCF_000149955.1	Y	Y	GCF_000149955.1

Table A.11: A chromosomal breakdown of the large and gene-rich regions shared between *Fusarium xylarioides* populations and various *Fusarium oxysporum* ff. spp. The coffea populations are shown together.

F. oxysporum f. sp. *lycopersici* chromosomal assembly	Arabica-specific	Arabica & coffea	Robusta-specific	Robusta & coffea
NC_030986.1	-	-	-	6,600
NC_030987.1	3,600	-	-	2,350
NC_030988.1	47,850	-	-	-
NC_030989.1	-	-	-	8,200
NC_030990.1	-	-	-	4,750
NC_030991.1	-	-	-	3,600
NC_030992.1	-	-	-	1,250
NC_030993.1	-	-	2,450	3,850
NC_030994.1	-	-	-	-
NC_030995.1	-	2,500	-	1,550
NC_030996.1	-	-	-	-
NC_030997.1	-	-	-	2,250
NC_030998.1	-	-	-	-
NC_030999.1	-	-	-	14,350
NC_031000.1	850	-	-	-
NW_017264849.1	4,900	-	-	-

F. oxysporum f. sp. *cubense* chromosomal assembly	Arabica-specific	Arabica & coffea	Robusta-specific	Robusta & coffea
SRMI01000001.1	-	-	-	4,400
SRMI01000002.1	1,500	2,050	-	-
SRMI01000003.1	-	1,750	-	15,200
SRMI01000004.1	-	-	-	3,350
SRMI01000005.1	-	-	-	-
SRMI01000006.1	-	1,150	2,250	1,800
SRMI01000007.1	-	2,350	-	-
SRMI01000008.1	-	2,500	-	-
SRMI01000009.1	-	-	1,800	-
SRMI01000010.1	50,950	-	-	-
SRMI01000011.1	-	-	-	-
SRMI01000012.1	3,250	-	-	-

F. oxysporum f. sp. *pisi* chromosomal assembly	Arabica-specific	Arabica & coffea	Robusta-specific	Robusta & coffea
JH650968.1	-	-	-	3,950
JH650969.1	-	-	-	4,350
JH650970.1	-	2,500	-	3,700
JH650971.1	-	1,900	-	-
JH650972.1	-	2,100	-	-
JH650973.1	-	-	2,450	-
JH650974.1	-	-	2,350	-
JH650975.1	2,000	2,500	3,250	2,100
JH650976.1	-	1,650	-	800
JH650977.1	-	-	-	-
JH650978.1	-	-	-	-
JH650980.1	3,150	-	-	1,700
JH650983.1	-	2,100	-	-
JH650984.1	-	-	-	3,100
JH650985.1	-	-	-	-
JH650986.1	-	-	-	-
JH650988.1	-	-	-	2,800
JH651002.1	-	-	-	1,450
JH651009.1	-	-	12,250	-
JH651010.1	-	-	-	1,000
JH651016.1	-	-	-	-
JH651017.1	-	-	-	-
JH651018.1	-	1,600	-	-
JH651021.1	-	-	-	-
JH651023.1	3,250	-	-	-
JH651035.1	-	-	-	4,750
JH651036.1	-	-	-	8,600
JH651058.1	-	-	-	10,150
JH651076.1	2,500	-	-	-
JH651077.1	4,450	-	-	-
JH651086.1	3,250	-	-	-
KI981264.1	6,600	-	-	-

F. oxysporum f. sp. *vasinfectum* chromosomal assembly	Arabica-specific	Arabica & coffea	Robusta-specific	Robusta & coffea
JH657924.1	-	-	-	-
JH657927.1	-	-	-	-
JH657928.1	-	-	-	-
JH657929.1	-	1,300	-	-
JH657931.1	-	2,200	-	-
JH657932.1	-	-	-	-

JH657935.1	-	-	2,250	-
JH657936.1	-	-	-	4,850
JH657938.1	-	-	-	-
JH657940.1	-	1,950	-	-
JH657941.1	-	-	-	-
JH657942.1	-	900	-	-
JH657943.1	-	-	-	1,650
JH657944.1	-	-	-	-
JH657945.1	11,600	1,750	-	-
JH657946.1	-	-	-	-
JH657947.1	-	-	-	-
JH657948.1	-	-	-	-
JH657949.1	-	-	-	-
JH657952.1	-	-	-	-
JH657953.1	-	-	-	-
JH657954.1	-	1,900	-	-
JH657955.1	-	-	-	-
JH657956.1	-	-	-	-
JH657957.1	1,050	-	-	-
JH657959.1	-	-	3,250	-
JH657961.1	-	-	-	-
JH657963.1	-	1,750	-	-
JH657968.1	-	-	-	2,750
JH657969.1	-	-	-	-
JH657975.1	-	1,600	-	-
JH657977.1	-	-	-	-
JH657980.1	-	1,650	-	-
JH657993.1	-	-	-	-
JH657995.1	6,802	-	-	-
JH658011.1	-	-	-	-
JH658025.1	-	-	-	-
JH658063.1	4,000	-	-	-
KK035210.1	-	-	-	-
KK035212.1	-	-	-	-
KK035214.1	-	-	-	-
KK035227.1	2,900	-	-	-
KK035232.1	-	-	-	-
KK035240.1	7,850	-	-	-

F. oxysporum f. sp. *coffea* chromosomal assembly	Arabica-specific	Arabica & coffea	Robusta-specific	Robusta & coffea
NC_030986.1	-	-	-	5,700
NC_030987.1	-	-	-	1,200
NC_030988.1	-	-	-	-
NC_030989.1	-	-	-	2,750

NC_030990.1	3,550	-	-	4,950
NC_030991.1	-	-	-	6,200
NC_030992.1	-	-	-	750
NC_030993.1	2,300	-	-	1,200
NC_030994.1	-	-	1,900	-
NC_030995.1	-	2,500	-	4,000
NC_030996.1	-	-	-	2,800
NC_030997.1	-	3,650	-	-
NC_030998.1	1,850	-	-	-
NC_030999.1	2,850	-	2,100	-
NW_017264847.1	-	-	-	700
NW_017264849.1	3,250	-	-	-

F. oxysporum f. sp. *raphani* chromosomal assembly	Arabica-specific	Arabica & Coffea	Robusta-specific	Robusta & coffea
JAELUR010000001.1	-	-	-	-
JAELUR010000002.1	3,250	-	-	1,850
JAELUR010000003.1	2,900	1,550	-	-
JAELUR010000004.1	-	-	-	-
JAELUR010000006.1	-	-	-	-
JAELUR010000007.1	3,350	-	-	-
JAELUR010000008.1	2,000	-	-	-
JAELUR010000009.1	-	3,250	-	-
JAELUR010000010.1	6,600	1,800	-	-
JAELUR010000011.1	-	-	-	2,800
JAELUR010000012.1	-	2,150	-	3,600
JAELUR010000013.1	-	-	-	-
JAELUR010000014.1	-	-	-	-
JAELUR010000015.1	12,450	-	-	1,850
JAELUR010000016.1	10,650	-	-	7,100
JAELUR010000017.1	-	2,550	-	-
JAELUR010000018.1	2,650	-	-	-
JAELUR010000021.1	-	-	-	8,450
JAELUR010000024.1	-	-	1,050	-
JAELUR010000026.1	-	-	-	-
JAELUR010000028.1	5,850	-	-	-
JAELUR010000029.1	10,000	-	-	-
JAELUR010000038.1	-	-	-	-

Table A.12: Expression data for all putative effectors from chapter 2 across the *Fusarium xylarioides* arabica563 samples. Each gene is described as up-regulated *in planta* ('coffee.up'), *in axenic* ('culture.up') or not differentially expressed ('ns').

Arabica563 gene	log2FoldChange	False Discovery Rate	Predicted EffectorP?	Differentially expressed?	Putative effector
P1J47_011695-T1	13.4154519	7.18E-47	FALSE	coffee.up	OG0014398
P1J47_015569-T1	5.379337	3.84E-29	FALSE	coffee.up	FOXG_14254
P1J47_014967-T1	14.4073093	2.66E-16	TRUE	coffee.up	pelA
P1J47_004459-T1	11.736764	4.71E-16	TRUE	coffee.up	pelD
P1J47_002381-T1	2.5053746	1.88E-10	FALSE	coffee.up	orx1
P1J47_017001-T1	4.9013526	2.62E-09	FALSE	coffee.up	OG0018569
P1J47_007374-T1	-4.8076726	2.45E-07	FALSE	culture.up	OG0014367
P1J47_007948-T1	2.3398743	2.93E-06	FALSE	coffee.up	sge1
P1J47_014428-T1	3.9418829	7.27E-05	FALSE	coffee.up	catalase-peroxidase
P1J47_013555-T1	1.0310702	9.98E-04	FALSE	ns	fow1
P1J47_005049-T1	-0.7238584	4.23E-03	FALSE	ns	snf1
P1J47_012661-T1	-2.5494558	3.20E-02	FALSE	ns	chlo_vacu
P1J47_013807-T1	-2.0638231	5.75E-02	FALSE	ns	OG0014828
P1J47_001679-T1	-2.871467	1.09E-01	FALSE	ns	OG0013912
P1J47_002427-T1	0.3631168	1.23E-01	FALSE	ns	rho1.1
P1J47_003388-T1	-1.6549056	1.82E-01	FALSE	ns	OG0013871
P1J47_013972-T1	-0.2587035	4.35E-01	FALSE	ns	rho1.2
P1J47_016733-T1	-0.3777736	5.01E-01	FALSE	ns	OG0013877
P1J47_016313-T1	0.141086	8.53E-01	FALSE	ns	fmk1
P1J47_017133-T1	15.6417987	1.00E+00	TRUE	ns	OG0013477

Table A.13: Expression data for all putative effectors from chapter 2 across the *Fusarium xylarioides* arabica908 samples. Each gene is described as up-regulated *in planta* ('coffee.up'), *in axenic* ('culture.up') or not differentially expressed ('ns').

Arabica908 gene	log2FoldChange	False Discovery Rate	Predicted EffectorP?	Differentially expressed?	Putative effector
P1J48_007834-T1	13.65516535	2.00E-31	FALSE	coffee.up	OG0014398
P1J48_014203-T1	15.08200554	6.55E-21	FALSE	coffee.up	OG0013477
P1J48_011047-T1	11.68459215	2.55E-12	TRUE	coffee.up	pelA
P1J48_013465-T1	3.95852869	2.95E-11	FALSE	coffee.up	FOXG 14254
P1J48_003935-T1	5.62411861	1.53E-08	TRUE	coffee.up	pelD
P1J48_002710-T1	-2.80463765	6.78E-07	FALSE	culture.up	OG0014367
P1J48_14533-T1	-2.44720929	5.95E-04	FALSE	culture.up	snf1
P1J48_001496-T1	1.96376611	1.62E-02	FALSE	ns	sge1
P1J48_010608-T1	0.74207488	1.98E-02	FALSE	ns	fow1
P1J48_002347-T1	-2.46547337	3.34E-02	FALSE	ns	catalase-peroxidase
P1J48_002576-T1	-2.59379467	6.40E-02	FALSE	ns	chlo_vacu
P1J48_012885-T1	1.72423883	9.40E-02	FALSE	ns	OG0018569
P1J48_012246-T1	-1.08726	1.13E-01	FALSE	ns	fmk1
P1J48_005732-T1	-0.46920318	1.54E-01	FALSE	ns	rho1.2
P1J48_001729-T1	-0.35483043	4.68E-01	FALSE	ns	orx1
P1J48_010428-T1	0.6993708	5.53E-01	FALSE	ns	OG0013871
P1J48_013741-T1	-0.10393013	8.78E-01	FALSE	ns	OG0013877
P1J48_001603-T1	-0.04693784	9.18E-01	FALSE	ns	rho1.1

Table A.14: The number of genes expressed for each carbohydrate-active enzyme sub-family in *Fusarium xylarioides* arabica563. The *in planta* up-regulated genes are shaded, where the warmest colours represent the highest up-regulated gene count.

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
AA1	12	1	4	0
AA11	4	0	1	0
AA12	2	0	0	0
AA13	1	1	0	1
AA14	1	1	0	0
AA16	1	1	0	0
AA2	4	0	0	0
AA3	21	5	1	2
AA3,AA8	6	3	0	0
AA4	6	0	0	0
AA4,AA4	0	0	0	0
AA4,AA7	1	0	0	0
AA5	1	0	0	0
AA5,CBM32	0	0	0	0
AA6	1	1	0	0
AA7	21	6	3	0
AA7,AA4	0	0	0	0
AA8,AA3	0	0	0	0
AA9	14	9	0	2
CBM18	2	1	0	0
CBM18,CE4	0	0	0	0
CBM18,GH18	1	0	1	0
CBM20,GH15	0	0	0	0
CBM21	1	0	0	0
CBM32,AA5	3	0	1	0
CBM38,GH32	0	0	0	0
CBM42,GH54	0	0	0	0
CBM43,GH72	1	0	0	0
CBM50	2	0	1	0
CBM6	1	0	0	0
CBM63	1	0	0	0
CBM67,GH78	5	1	0	1
CE1	4	2	0	0
CE10	54	2	2	0
CE12	4	4	0	2
CE16	6	4	0	1
CE2	1	1	0	0
CE3	4	1	0	0
CE4	9	5	0	3

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
CE5	12	5	0	2
CE7	1	0	0	0
CE8	2	0	0	0
CE9	1	0	0	0
GH1	4	1	0	0
GH10	4	3	0	1
GH105	4	4	0	2
GH106	1	0	0	0
GH11	3	3	0	0
GH114	3	1	1	0
GH115	2	0	0	0
GH12	3	0	1	0
GH125	3	1	0	0
GH127,GH146	0	0	0	0
GH128	3	0	0	0
GH13	8	3	0	0
GH131	1	1	0	1
GH132	2	0	0	0
GH133	1	0	0	0
GH134	1	0	0	0
GH139	1	0	0	0
GH145	2	0	0	0
GH145,PL24	0	0	0	0
GH146	0	0	0	0
GH146,GH127	2	1	0	0
GH146,GH127,GH146,GH127	0	0	0	0
GH15	1	0	0	0
GH15,CBM20	1	0	0	0
GH152	1	0	0	0
GH154	2	0	0	0
GH16	22	4	4	0
GH16,GH64	0	0	0	0
GH162	1	0	0	0
GH17	4	0	0	0
GH18	16	3	2	1
GH2	9	3	0	1
GH20	3	0	1	0
GH24	1	0	0	0
GH28	7	2	0	0
GH28,GH28	0	0	0	0
GH29	3	1	0	0
GH3	20	4	0	1
GH30	1	0	0	0
GH31	8	1	0	0

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
GH32	7	2	0	1
GH32,CBM38	2	1	0	0
GH33	1	0	0	0
GH35	4	3	0	0
GH36	3	1	0	0
GH37	2	1	0	0
GH38	1	0	0	0
GH43	25	14	0	5
GH43,CBM6	0	0	0	0
GH45	1	0	0	0
GH47	10	1	1	1
GH49	1	1	0	0
GH5	21	8	1	4
GH5,GH2	0	0	0	0
GH51	2	1	0	1
GH53	1	1	0	0
GH54,CBM42	1	0	0	0
GH55	2	0	0	0
GH6	1	1	0	0
GH64	3	0	0	0
GH65	1	0	1	0
GH67	2	1	0	0
GH7	3	2	0	0
GH71	2	0	0	0
GH72	2	0	0	0
GH72,CBM43	0	0	0	0
GH75	2	1	0	0
GH76	9	1	1	1
GH78	2	0	0	0
GH78,CBM67	0	0	0	0
GH79	1	0	0	0
GH81	3	1	0	1
GH88	1	1	0	0
GH93	5	3	0	1
GH95	2	0	0	0
GT1	5	1	1	0
GT15	5	0	0	0
GT17	2	0	0	0
GT2_Chitin_synth	8	0	0	0
GT2_Glyco_tranf_2	3	1	0	0
GT2_Glyco_trans_2	3	0	0	0
GT2_Glycos_transf	3	0	0	0
GT2_Glycos_transf,GT2_Glyco_tranf_2	1	1	0	1
GT2_Glycos_transf,GT2_Glycos_transf")	0	0	0	0

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
GT20	2	0	0	0
GT21	1	0	0	0
GT22	4	0	0	0
GT24	1	0	0	0
GT3	1	0	0	0
GT32	5	0	1	0
GT33	1	0	0	0
GT34	3	0	0	0
GT35	1	0	0	0
GT39	3	0	0	0
GT4	5	0	0	0
GT48	0	0	0	0
GT48,GT48”)	0	0	0	0
GT57	2	0	0	0
GT58	1	0	0	0
GT59	1	0	0	0
GT62	3	0	0	0
GT64	2	0	0	0
GT66	1	0	0	0
GT69	2	0	1	0
GT71	2	0	0	0
GT76	1	0	0	0
GT8	6	0	0	0
GT90	5	0	0	0
PL1	12	11	0	8
PL26	1	1	0	0
PL3	6	6	0	4
PL4	3	2	0	1
PL9	1	1	0	1

Table A.15: The number of genes expressed for each carbohydrate-active enzyme sub-family in *Fusarium xylarioides* arabica908. The *in planta* up-regulated genes are shaded, where the warmest colours represent the highest up-regulated gene count.

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
AA1	13	0	3	0
AA11	4	0	3	0
AA12	2	0	0	0
AA13	1	0	0	0
AA14	1	0	0	0
AA16	1	0	0	0
AA2	4	0	1	0
AA3	24	3	1	0
AA3,AA8	0	0	0	0
AA4	5	0	1	0
AA4,AA4	0	0	0	0
AA4,AA7	0	0	0	0
AA5	1	0	0	0
AA5,CBM32	0	0	0	0
AA6	1	1	0	0
AA7	21	3	2	1
AA7,AA4	0	0	0	0
AA8,AA3	6	3	0	0
AA9	14	3	0	1
CBM18	2	0	0	0
CBM18,CE4	1	0	0	0
CBM18,GH18	1	0	0	0
CBM20,GH15	0	0	0	0
CBM21	1	0	0	0
CBM32,AA5	3	0	1	0
CBM38,GH32	0	0	0	0
CBM42,GH54	0	0	0	0
CBM43,GH72	0	0	0	0
CBM50	2	0	0	0
CBM6	0	0	0	0
CBM63	1	0	0	0
CBM67,GH78	0	0	0	0
CE1	4	1	0	0
CE10	55	0	1	0
CE10",CE10")	1	0	0	0
CE12	4	2	0	0
CE16	6	2	0	0
CE2	1	0	0	0
CE3	4	0	0	0
CE4	9	3	0	0

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
CE5	12	3	0	1
CE7	1	0	0	0
CE8	4	1	0	0
CE9	1	0	0	0
GH1	4	0	0	0
GH10	4	2	0	1
GH105	4	3	0	1
GH106	1	0	0	0
GH11	3	2	0	0
GH114	3	0	1	0
GH115	2	0	0	0
GH12	4	0	1	0
GH125	3	1	0	0
GH127,GH146	0	0	0	0
GH128	3	1	0	0
GH13	8	1	0	0
GH131	1	0	0	0
GH132	1	0	0	0
GH133	1	0	0	0
GH134	1	0	0	0
GH139	1	0	0	0
GH145	2	0	0	0
GH145,PL24	0	0	0	0
GH146	1	0	0	0
GH146,GH127	1	1	0	0
GH146,GH127,GH146,GH127	0	0	0	0
GH15	1	0	0	0
GH15,CBM20	1	0	0	0
GH152	1	0	0	0
GH154	2	0	0	0
GH16	21	1	2	0
GH16,GH64	0	0	0	0
GH162	1	0	0	0
GH17	4	0	0	0
GH18	17	2	2	0
GH2	9	3	0	0
GH20	3	0	2	0
GH24	1	0	0	0
GH28	8	2	0	0
GH28,GH28	0	0	0	0
GH29	3	1	0	0
GH3	19	2	1	1
GH30	2	0	0	0
GH31	8	1	0	0

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
GH32	7	1		
GH32,CBM38	2	0	0	0
GH33	1			
GH35	6	1	0	0
GH36	3	1	0	0
GH37	2	1	0	0
GH38	0	0		
GH43	23	13	0	2
GH43,CBM6	1	0	0	0
GH45	1	0	0	0
GH47	10	1	1	0
GH49	1	0	0	0
GH5	18	3		
GH5,GH2	0	0	0	0
GH51	2	1		
GH53	1	0	0	0
GH54,CBM42	1	0	0	0
GH55	3	0	1	0
GH6	1	0	0	0
GH64	3	0	0	0
GH65	1	0	0	0
GH67	2	0	0	0
GH7	2	0	0	0
GH71	2	0	0	0
GH72	2	0	0	0
GH72,CBM43	1	0	0	0
GH75	1	0	0	0
GH76	11	0	1	0
GH78	2			
GH78,CBM67	5	0	0	0
GH79	0	1		
GH81	2	0	0	0
GH88	1	1		
GH93	5	0	0	0
GH95	2	2	0	0
GT1	5	0	0	0
GT15	5	0		
GT17	2	0	0	0
		0	1	0
GT2_Glyco_tranf_2	1	0	1	0
GT2_Glyco_trans_2	3	0	0	0
GT2_Glycos_transf	3	0	0	0
		0		
		0		

CAZyme	Gene annotation (n)	In planta up genes (n)	In axenic up genes (n)	Most differentially expressed gene (n)
GT20	2	0	0	0
GT21	1	0	0	0
GT22	5	0	0	0
GT24	1	0	0	0
GT3	1	0	1	0
GT32	5	0	1	0
GT33	1	0	0	0
GT34	3	0	0	0
GT35	1	0	0	0
GT39	3	0	0	0
GT4	5	0	0	0
GT48	1	0	0	0
GT48,GT48")	0	0	0	0
GT57	2	0	0	0
GT58	1	0	0	0
GT59	1	0	0	0
GT62	3	0	1	0
GT64	2	0	0	0
GT66	1	0	0	0
GT69	1	0	1	0
GT71	2	0	0	0
GT76	1	0	0	0
GT8	6	0	0	0
GT90	4	0	0	0
PL1	13	9	0	4
PL26	1	1	0	1
PL3	5	4	0	1
PL4	3	1	0	1
PL9	1	1	0	1

Table A.16: Gene count for significantly enriched gene ontology terms across the *Fusarium xylarioides* arabica563 and arabica908 *in planta* strains.

Category	Term	Arabica563 (gene n)	Arabica908 (gene n)
GO:0055085	transmembrane transport	194	60
GO:0022857	transmembrane transporter activity	149	45
GO:0003735	structural constituent of ribosome	60	62
GO:0006412	translation	57	60
GO:0005840	ribosome	54	56
GO:0015935	small ribosomal subunit	9	10
GO:0015934	large ribosomal subunit	5	5
GO:0016491	oxidoreductase activity	139	85
GO:0050660	flavin adenine dinucleotide binding	38	17
GO:0071949	FAD binding	34	14
GO:0016705	oxidoreductase activity	38	0
GO:0004497	monooxygenase activity	34	0
GO:0016614	oxidoreductase activity	11	6
GO:0016628	oxidoreductase activity	6	4
GO:0010181	FMN binding	0	8
GO:0050661	NADP binding	0	8
GO:0005506	iron ion binding	42	18
GO:0005975	carbohydrate metabolic process	104	59
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	70	37
GO:0005576	extracellular region	24	9
GO:0030570	pectate lyase activity	16	11
GO:0016829	lyase activity	9	6
GO:0030248	cellulose binding	11	3
GO:0030246	carbohydrate binding	7	6
GO:0016020	membrane	102	0
GO:0016021	integral component of membrane	97	0
GO:0006508	proteolysis	44	0
GO:0020037	heme binding	42	0
GO:0140359	NA	18	4
GO:0008236	serine-type peptidase activity	18	0
GO:0016787	hydrolase activity	0	18
GO:0008237	metallopeptidase activity	10	5
GO:0004252	serine-type endopeptidase activity	13	0
GO:0046873	metal ion transmembrane transporter activity	11	0
GO:0030001	metal ion transport	11	0
GO:0042626	ATPase activity	7	4
GO:0004181	metallocarboxypeptidase activity	7	3
GO:0071704	organic substance metabolic process	9	0
GO:0004601	peroxidase activity	7	0

Category	Term	Arabica563 (gene n)	Arabica908 (gene n)
GO:0016788	hydrolase activity, acting on ester bonds	7	0
GO:0016616	oxidoreductase activity	0	7
GO:0008484	sulfuric ester hydrolase activity	6	0
GO:0043531	ADP binding	6	0
GO:0051537	2 iron, 2 sulfur cluster binding	6	0
GO:0008061	chitin binding	6	0
GO:0031177	phosphopantetheine binding	5	0
GO:0043386	mycotoxin biosynthetic process	4	0
GO:0098869	cellular oxidant detoxification	2	2
GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	2
GO:0006817	phosphate ion transport	0	4
GO:0008199	ferric iron binding	0	4
GO:0006725	cellular aromatic compound metabolic process	0	4
GO:0005315	inorganic phosphate transmembrane transporter activity	0	4
GO:0016702	oxidoreductase activity	0	4
GO:0004499	N,N-dimethylaniline monooxygenase activity	0	4
GO:0050525	cutinase activity	3	0
GO:0004565	beta-galactosidase activity	3	0
GO:0009341	beta-galactosidase complex	3	0
GO:0004089	carbonate dehydratase activity	3	0
GO:0004650	polygalacturonase activity	0	3
GO:0006098	pentose-phosphate shunt	0	3
GO:0022625	cytosolic large ribosomal subunit	0	3
GO:0046855	inositol phosphate dephosphorylation	0	3
GO:0046854	phosphatidylinositol phosphorylation	0	3
GO:0005215	transporter activity	0	3
GO:0006857	oligopeptide transport	2	0
GO:0045493	xylan catabolic process	2	0
GO:0016652	oxidoreductase activity	2	0
GO:0046294	formaldehyde catabolic process	2	0
GO:0004616	phosphogluconate dehydrogenase (decarboxylating)	0	2
GO:0008934	inositol monophosphate 1-phosphatase activity	0	2

Category	Term	Arabica563 (gene n)	Arabica908 (gene n)
GO:0042819	vitamin B6 biosynthetic process	0	2
GO:0042823	pyridoxal phosphate biosynthetic process	0	2
GO:0016805	dipeptidase activity	0	2
GO:0019829	cation-transporting ATPase activity	0	2
GO:0016853	isomerase activity	0	2
GO:0009712	catechol-containing compound metabolic process	0	2
GO:0018576	catechol 1,2-dioxygenase activity	0	2
GO:0004411	homogentisate 1,2-dioxygenase activity	0	2
GO:0006570	tyrosine metabolic process	0	2
GO:0004784	superoxide dismutase activity	0	2
GO:0006559	L-phenylalanine catabolic process	0	2
GO:0016803	ether hydrolase activity	0	2
GO:0046718	viral entry into host cell	0	1
GO:0004834	tryptophan synthase activity	0	1
GO:0003876	AMP deaminase activity	0	1
GO:0032264	IMP salvage	0	1
GO:0043546	molybdopterin cofactor binding	0	1

Table A.17: Gene count and description for significantly enriched InterPro terms in the *Coffea arabica in planta* samples

InterPro category	*Coffea arabica* genes	Gene number	Gene Ontology term	InterPro description	Pfam domain
IPR002068	XP_027098388.1, XP_027113320.1, XP_027095883.1, XP_027096450.1, XP_027109657.1	5	NA	Hsp20/alpha crystallin family	PF00011
IPR001938	XP_027095536.1, XP_027096550.1, XP_027080774.1, XP_027080772.1	4	NA	Thaumatin family	PF00314
IPR007117	XP_027083877.1, XP_027082625.1, XP_027083872.1, XP_027084129.1	4	NA	Expansin C-terminal domain	PF01357
IPR009009	XP_027083877.1, XP_027082625.1, XP_027083872.1, XP_027084129.1	4	NA	Lytic transglycolase	PF03330
IPR013126	XP_027125306.1, XP_027121874.1, XP_027122715.1, XP_027114671.1	4	GO:0005524, GO:0140662	Hsp70 protein	PF00012
IPR001128	XP_027119443.1, XP_027099114.1	2	GO:0004497, GO:0005506, GO:0016705, GO:0020037	Cytochrome P450	PF00067
IPR001155	XP_027087183.1, XP_027088006.1	2	GO:0010181, GO:0016491	NADH:flavin oxidoreductase / NADH oxidase family	PF00724
IPR001480	XP_027061914.1, XP_027077059.1	2	NA	D-mannose binding lectin	PF01453
IPR002213	XP_027098144.1, XP_027127315.1	2	GO:0008194	UDP-glucoronosyl and UDP-glucosyl transferase	PF00201

InterPro category	*Coffea arabica* genes	Gene number	Gene Ontology term	InterPro description	Pfam domain
IPR003959	XP 027077924.1, XP 027113909.1	2	GO:0005524, GO:0016887	ATPase family associated with various cellular activities (AAA)	PF00004
IPR004265	XP 027062967.1, XP 027062970.1	2	NA	Dirigent-like protein	PF03018
IPR008146	XP 027077249.1, XP 027073453.1	2	GO:0004356, GO:0006807	Glutamine synthetase, catalytic domain	PF00120
IPR008147	XP 027077249.1, XP 027073453.1	2	GO:0004356, GO:0006542, GO:0006807	Glutamine synthetase, beta-Grasp domain	PF03951
IPR008914	XP 027068654.1, XP 027069261.1	2	NA	Phosphatidylethanolamine-binding protein	PF01161
IPR013148	XP 027080458.1, XP 027072817.1	2	NA	Glycosyl hydrolases family 32 N-terminal domain	PF00251
IPR013189	XP 027080458.1, XP 027072817.1	2	NA	Glycosyl hydrolases family 32 C terminal	PF08244
IPR000023	XP 027062920.1	1	GO:0003872, GO:0006096	Phosphofructokinase	PF00365
IPR000070	XP 027113015.1	1	GO:0030599, GO:0042545	Pectinesterase	PF01095
IPR000073	XP 027107249.1	1	NA	Alpha/beta hydrolase family	PF12697
IPR000408	XP 027070596.1	1	NA	Regulator of chromosome condensation (RCC1) repeat	PF00415
IPR000490	XP 027067208.1	1	GO:0004553, GO:0005975	Glycosyl hydrolases family 17	PF00332
IPR000642	XP 027077924.1	1	GO:0004176, GO:0004222, GO:0005524, GO:0006508	Peptidase family M41	PF01434
IPR000719	XP 027077059.1	1	GO:0004672, GO:0005524, GO:0006468	Protein kinase domain	PF00069
IPR000726	XP 027075439.1	1	GO:0004568, GO:0006032, GO:0016998	Chitinase class I	PF00182

InterPro category	*Coffea arabica* genes	Gene number	Gene Ontology term	InterPro description	Pfam domain
IPR000743	XP 027119275.1	1	GO:0004650, GO:0005975	Glycosyl hydrolases family 28	PF00295
IPR000863	XP 027095476.1	1	GO:0008146	Sulfotransferase domain	PF00685
IPR000996	XP 027064885.1	1	GO:0005198, GO:0006886, GO:0016192, GO:0030130, GO:0030132	Clathrin light chain	PF01086
IPR001153	XP 027109675.1	1	GO:0042742, GO:0050832	Barwin family	PF00967
IPR001382	XP 027123559.1	1	GO:0004571, GO:0005509, GO:0016020	Glycosyl hydrolase family 47	PF01532
IPR001404	XP 027110216.1	1	GO:0005524, GO:0006457, GO:0016887, GO:0051082, GO:0140662	Hsp90 protein	PF00183
IPR001611	XP 027127483.1	1	GO:0005515	Leucine rich repeat	PF13855
IPR001764	XP 027124556.1	1	GO:0004553, GO:0005975	Glycosyl hydrolase family 3 N terminal domain	PF00933
IPR002241	XP 027100553.1	1	GO:0004553, GO:0005975	Alpha galactosidase A	PF16499
IPR002347	XP 027070209.1	1	NA	short chain dehydrogenase	PF00106
IPR002772	XP 027124556.1	1	GO:0004553, GO:0005975	Glycosyl hydrolase family 3 C-terminal domain	PF01915
IPR002912	XP 027097796.1	1	NA	ACT domain	PF01842
IPR003245	XP 027069566.1	1	GO:0009055	Plastocyanin-like domain	PF02298
IPR003594	XP 027110216.1	1	NA	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	PF02518
IPR003609	XP 027061914.1	1	NA	PAN-like domain	PF08276
IPR004240	XP 027084986.1	1	GO:0016021	Endomembrane protein 70	PF02990
IPR004316	XP 027100714.1	1	GO:0016021	Sugar efflux transporter for intercellular exchange	PF03083
IPR004839	XP 027112899.1	1	GO:0009058, GO:0030170	Aminotransferase class I and II	PF00155

InterPro category	*Coffea arabica* genes	Gene number	Gene Ontology term	InterPro description	Pfam domain
IPR004843	XP_027070343.1	1	GO:0016787	Calcineurin-like phosphoesterase	PF00149
IPR005474	XP_027102624.1	1	NA	Transketolase, thiamine diphosphate binding domain	PF00456
IPR005475	XP_027102624.1	1	NA	Transketolase, pyrimidine binding domain	PF02779
IPR005828	XP_027103991.1	1	GO:0016021, GO:0022857, GO:0055085	Sugar (and other) transporter	PF00083
IPR006094	XP_027088393.1	1	GO:0050660	FAD binding domain	PF01565
IPR006254	XP_027062002.1	1	GO:0004451, GO:0019752	Isocitrate lyase family	PF00463
IPR006501	XP_027113015.1	1	GO:0004857	Plant invertase/pectin methylesterase inhibitor	PF04043
IPR012951	XP_027088393.1	1	GO:0016491, GO:0050660	Berberine and berberine like	PF08031
IPR013149	XP_027120066.1	1	NA	Zinc-binding dehydrogenase	PF00107
IPR015914	XP_027070343.1	1	GO:0003993, GO:0046872	Purple acid Phosphatase, N-terminal domain	PF16656
IPR019378	XP_027070430.1	1	NA	GDP-fucose protein O-fucosyltransferase	PF10250
IPR019791	XP_027102149.1	1	NA	Animal haem peroxidase	PF03098
IPR022003	XP_027086201.1	1	NA	RCD1-SRO-TAF4 (RST) plant domain	PF12174
IPR022702	XP_027127341.1	1	NA	Cytosine specific DNA methyltransferase replication foci domain	PF12047
IPR025733	XP_027070343.1	1	NA	Iron/zinc purple acid phosphatase-like protein C	PF14008
IPR026992	XP_027116487.1	1	NA	non-haem dioxygenase in morphine synthesis N-terminal	PF14226
IPR033248	XP_027102624.1	1	NA	Transketolase, C-terminal domain	PF02780
IPR041233	XP_027100553.1	1	NA	Alpha galactosidase C-terminal beta sandwich domain	PF17801
IPR041569	XP_027077924.1	1	NA	AAA+ lid domain	PF17862
IPR041694	XP_027120066.1	1	NA	N-terminal domain of oxidoreductase	PF16884
IPR044861	XP_027116487.1	1	NA	2OG-Fe(II) oxygenase superfamily	PF03171

Bibliography

Abo, K., Klein, K. K., Edel-Hermann, V., Gautheron, N., Traore, D. & Steinberg, C. (2005), 'High genetic diversity among strains of Fusarium oxysporum f. sp. vasinfectum from cotton in Ivory Coast', *Phytopathology* **95**(12), 1391–1396.

Agrios, G. (2005), *Plant Pathology*, fifth edn, Elsevier.

Ahuja, I., de Vos, R. C., Bones, A. M. & Hall, R. D. (2010), 'Plant molecular stress responses face climate change', *Trends in Plant Science* **15**(12), 664–674.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B. & Schatz, M. C. (2019), 'RaGOO: Fast and accurate reference-guided scaffolding of draft genomes', *Genome Biology* **20**(1), 224.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R. & Daszak, P. (2004), 'Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers', *Trends in Ecology & Evolution* **19**(10), 535–544.

Andrews S, Krueger F, Seconds-Pichon A, Biggins F & Wingett S (2015), 'FastQC. A Quality Control tool for High Throughput Sequence Data'.

Aoki, T., O'Donnell, K. & Geiser, D. M. (2014), 'Systematics of key phytopathogenic Fusarium species: Current status and future challenges'.

Arif, M., Chawla, S., Zaidi, N. W., Rayar, J. K., Variar, M. & Singh, U. S. (2011), 'Development of specific primers for genus Fusarium and F. solani using rDNA sub-unit

and transcription elongation factor (TEF-1α) gene', *African Journal of Biotechnology* **11**(2), 444–447.

Aury, J. M. & Istace, B. (2021), 'Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads', *NAR Genomics and Bioinformatics* **3**(2).

Ayukawa, Y., Asai, S., Gan, P., Tsushima, A., Ichihashi, Y., Shibata, A., Komatsu, K., Houterman, P. M., Rep, M., Shirasu, K. & Arie, T. (2021), 'A pair of effectors encoded on a conditionally dispensable chromosome of Fusarium oxysporum suppress host-specific immunity', *Communications Biology 2021 4:1* **4**(1), 1–12.

Aziz, R. K., Devoid, S., Disz, T., Edwards, R. A., Henry, C. S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Stevens, R. L., Vonstein, V. & Xia, F. (2012), 'SEED Servers: High-Performance Access to the SEED Genomes, Annotations, and Metabolic Models', *PLoS ONE* **7**(10), e48053.

Ball, B., Langille, M. & Geddes-Mcalister, J. (2020), 'Fun(gi)omics: Advanced and Diverse Technologies to Explore Emerging Fungal Pathogens and Define Mechanisms of Antifungal Resistance'.

Barnett, D., Garrison, E., Quinlan, A., Stromberg, M. & Marth, G. (2011), 'BamTools: a C++ API and toolkit for analyzing and managing BAM files', *Bioinformatics* **27**(12), 1691–1692.

Batada, N. N., Urrutia, A. O. & Hurst, L. D. (2007), 'Chromatin remodelling is a major source of coexpression of linked genes in yeast', *Trends in Genetics* **23**(10), 480–484.

Beckman, C. (1956), 'Production of Pectinase, Cellulases, and growth-promoting substance by Ceratostomella Ulmi', *Phytopathology* **46**, 605–609.

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(1), 289–300.

Bergemann, M., Lespinet, O., M'Barek, S. B., Daboussi, M. J. & Dufresne, M. (2008), 'Genome-wide analysis of the Fusarium oxysporum mimp family of MITEs and mobilization of both native and de novo created mimps', *Journal of Molecular Evolution* **67**(6), 631–642.

Betts, P. (2022), 'Personal communication'.

Birikunzira, J. & Hakiza, G. (1997), 'The status of coffee wilt disease (Tracheomycosis) and strategies for its control in Uganda..', *Scientifique International sur le* **17**.

Blaxter, M. L. (2022), 'Sequence locally, think globally: The Darwin Tree of Life Project', *Proceedings of the National Academy of Sciences of the United States of America* **119**(4).

Boer, A. D., Plant, V. V., Environment, C. . & 2003, u. (2003), 'Logistics of water and salt transport through the plant: structure and functioning of the xylem', *Wiley Online Library* **26**(1), 87–101.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-3040.2003.00930.x

Bolger, A. M., Lohse, M. & Usadel, B. (2014), 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics* **30**(15), 2114.

Booth, C. (1971), 'The genus Fusarium.', *The genus Fusarium.* .

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004), 'Carbohydrate-binding modules: Fine-tuning polysaccharide recognition', *Biochemical Journal* **382**(3), 769–781.

Borges, A. F., Ferreira, R. B. & Monteiro, S. (2013), 'Transcriptomic changes following the compatible interaction Vitis vinifera–Erysiphe necator. Paving the way towards an enantioselective role in plant defence modulation', *Plant Physiology and Biochemistry* **68**, 71–80.

Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A. & Schmidt, S. A. (2016), 'Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of Saccharomyces cerevisiae', *G3: Genes, Genomes, Genetics* **6**(4), 957–971.

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs, N., Cowan, R. S., Davies, N. M. J., Dodsworth, S., Edwards, S. L., Eiserhardt, W. L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P. J., Pokorny, L., Leitch, I. J., Forest, F. & Baker, W. J. (2019), 'Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms', *Frontiers in Plant Science* **10**, 1102.

Buddie, A. G., Crozier, J., Rutherford, M. A., Flood, J. & Bridge, P. D. (2015), 'Population development within the coffee wilt pathogen Gibberella xylarioides reflects host-related divergence', *European Journal of Plant Pathology* **142**(2), 291–304.

Burdon, J. J. & Thrall, P. H. (2008), 'Pathogen evolution across the agro-ecological interface: implications for disease management', *Evolutionary Applications* **1**(1), 57–65.

Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009), 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Research* **37**(D233-D238).

Carlos Florez, J., Souto Mofatto, L., do Livramento Freitas-Lopes, R., Siqueira Ferreira, S., Maciel Zambolim, E., Falsarella Carazzolle, M., Zambolim, L. & Teixeira Caixeta, E. (2017), 'High throughput transcriptome analysis of coffee reveals prehaustorial resistance in response to Hemileia vastatrix infection', *Plant Molecular Biology* **95**, 607–623.

Carpita, N. C. & Gibeaut, D. M. (1993), 'Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth', *Plant Journal* **3**(1), 1–30.

Carrard, G., Koivula, A., Söderlund, H. & Beguin, P. (2000), 'Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose', *Proceedings of the National Academy of Sciences* **97**(19), 10342–10347.

Chen, Y., Yin, H., Gao, M., Zhu, H., Zhang, Q. & Wang, Y. (2016), 'Comparative Transcriptomics Atlases Reveals Different Gene Expression Pattern Related to Fusarium Wilt Disease Resistance and Susceptibility in Two Vernicia Species', *Frontiers in Plant Science* **7**(1974).

Chiarappa, L. (1969), 'International assistance in plant pathology in developing countries. With particular reference to FAO programmes.', *FAO Plant Protection Bulletin* **17**, 7–8.

Choat, B., Jansen, S., Zwieniecki, M. A., Smets, E. & Holbrook, N. M. (2004), 'Changes in pit membrane porosity due to deflection and stretching: the role of vestured pits', *Journal of Experimental Botany* **55**(402), 1569–1575.
URL: https://academic.oup.com/jxb/article/55/402/1569/464916

Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., Yoshida, K., Terauchi, R., Fujita, Y., Nakayashiki, H., Valent, B. & Tosa, Y. (2011), 'Multiple translocation of the avr-pita effector gene among chromosomes of the rice blast fungus magnaporthe oryzae and related species', *PLoS Pathogens* **7**(7).

Clérivet, A., Déon, V., Alami, I., Lopez, F., Trees, J. G. & 2000, u. (2000), 'Tyloses and gels associated with cellulose accumulation in vessels are responses of plane tree seedlings (Platanus × acerifolia) to the vascular fungus Ceratocystis fimbriata f', *Springer* **15**(1), 25–31.

Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., Schmutz, J., Taga, M., White, G. J., Zhou, S., Schwartz, D. C., Freitag, M., Ma, L.-j., Danchin, E. G. J., Henrissat, B., Coutinho, P. M., Nelson, D. R., Straney, D., Napoli, C. A., Barker, B. M., Gribskov, M., Rep, M., Kroken, S., Molnár, I., Rensing, C., Kennell, J. C., Zamora, J., Farman, M. L., Selker, E. U., Salamov, A., Shapiro,

H., Pangilinan, J., Lindquist, E., Lamers, C., Grigoriev, I. V., Geiser, D. M., Covert, S. F., Temporini, E. & VanEtten, H. D. (2009), 'The Genome of Nectria haematococca: Contribution of Supernumerary Chromosomes to Gene Expansion', *PLoS Genetics* **5**(8), e1000618.

Coleman, J. J., Wasmann, C. C., Usami, T., White, G. J., Temporini, E. D., McCluskey, K. & Van Etten, H. D. (2011), 'Characterization of the gene encoding pisatin demethylase (FoPDA1) in Fusarium oxysporum', *Molecular Plant-Microbe Interactions* **24**(12), 1482–1491.

Cosgrove, D. J. (2015), 'Plant expansins: diversity and interactions with plant cell walls', *Current Opinion in Plant Biology* **25**, 162–172.

Covey, P. A., Kuwitzky, B., Hanson, M. & Webb, K. M. (2014), 'Multilocus Analysis Using Putative Fungal Effectors to Describe a Population of ¡i¿Fusarium oxysporum¡/i¿ from Sugar Beet', *Phytopathology* **104**(8), 886–896.

Craparo, A., Van Asten, P., Läderach, P., Jassogne, L. & Grab, S. (2015), 'Coffea arabica yields decline in Tanzania due to climate change: Global implications', *Agricultural and Forest Meteorology* **207**, 1–10.

Cubero, O. F., Crespo, A., Fatehi, J. & Bridge, P. D. (1999), 'Plant Systematics and Evolution DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi', *Plant Systematics and Evolution* **216**, 243–249.

Daboussi, M.-J. & Capy, P. (2003), 'Transposable Elements in Filamentous Fungi', *Annual Review of Microbiology* **57**(1), 275–299.

Davey, J. W., Davis, S. J., Mottram, J. C. & Ashton, P. D. (2020), 'Tapestry: validate and edit small eukaryotic genome assemblies with long reads', *bioRxiv* .

Davis, A. P., Govaerts, R., Bridson, D. M. & Stoffelen, P. (2006), 'An annotated taxonomic

conspectus of the genus Coffea (Rubiaceae)', *Botanical Journal of the Linnean Society* **152**, 465–512.

de Jonge, R., Bolton, M. D. & Thomma, B. P. (2011), 'How filamentous pathogens co-opt plants: the ins and outs of fungal effectors', *Current Opinion in Plant Biology* **14**(4), 400–406.

de Jonge, R., Esse, H. P. v., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., Krol, S. v. d., Shibuya, N., Joosten, M. H. A. J. & Thomma, B. P. H. J. (2010), 'Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants', *Science* **329**(5994), 953–955.

De Wit, P. J. (2016), 'Apoplastic fungal effectors in historic perspective; a personal view', *New Phytologist* **212**(4), 805–813.

Di Pietro, A., García-Maceira, F. I., Méglecz, E. & Roncero, M. I. (2001), 'A MAP kinase of the vascular wilt fungus Fusarium oxysporum is essential for root penetration and pathogenesis', *Molecular Microbiology* **39**(5), 1140–1152.

Di Pietro, A., Madrid, M. P., Caracuel, Z., Delgado-Jarana, J. & Roncero, M. I. G. (2003), 'Fusarium oxysporum : exploring the molecular arsenal of a vascular wilt fungus', *Molecular Plant Pathology* **4**(5), 315–325.

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. (2013), 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics* **29**(1), 15–21.

Dong, S., Raffaele, S. & Kamoun, S. (2015), 'The two-speed genomes of filamentous pathogens: waltz with plants', *Current Opinion in Genetics & Development* **35**, 57–65.

Donnelly, J. S. (2002), *The great Irish potato famine*, The History Press.

Douglas, G. M. & Langille, M. G. I. (2019), 'Current and Promising Approaches to Identify

Horizontal Gene Transfer Events in Metagenomes', *Genome Biol. Evol.* **11**(10), 2750–2766.

Dumas, E., Feurtey, A., Rodríguez de la Vega, R. C., Le Prieur, S., Snirc, A., Coton, M., Thierry, A., Coton, E., Le Piver, M., Roueyre, D., Ropars, J., Branca, A. & Giraud, T. (2020), 'Independent domestication events in the blue-cheese fungus Penicillium roqueforti', *Molecular Ecology* **29**(14), 2639–2660.

Ellis, M. L., Cruz Jimenez, D. R., Leandro, L. F. & Munkvold, G. P. (2014), 'Genotypic and Phenotypic Characterization of Fungi in the Fusarium oxysporum Species Complex from Soybean Roots', *Phytopathology* **104**, 1329–1339.

Ellis, M. L., Lanubile, A., Garcia, C. & Munkvold, G. P. (2016), 'Association of Putative Fungal Effectors in Fusarium oxysporum with Wilt Symptoms in Soybean', *Phytopathology* **106**, 762–773.

Emanuelsson, O., Nielsen, H., Brunak, S. & Von Heijne, G. (2000), 'Predicting subcellular localization of proteins based on their N-terminal amino acid sequence', *Journal of Molecular Biology* **300**(4), 1005–1016.

Emms, D. & Kelly, S. (2017), 'STRIDE: species tree root inference from gene duplication events', *Molecular Biology and Evolution* **34**(12), 3267–3278.

Emms, D. M. & Kelly, S. (2015), 'OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy', *Genome Biology* **16**(1), 1–14.

Emms, D. M. & Kelly, S. (2018), 'STAG: Species Tree Inference from All Genes', *BioRxiv* pp. 1–29.

Emms, D. M. & Kelly, S. (2019), 'OrthoFinder: Phylogenetic orthology inference for comparative genomics', *Genome Biology* **20**(238), 1–14.

FAO (2022), 'FAOSTAT data'.

Farrer, R. A. & Fisher, M. C. (2017), 'Describing Genomic and Epigenomic Traits Underpinning Emerging Fungal Pathogens', *Advances in Genetics* **100**, 73–140.

Ferreira, R. B., Monteiro, S., Freitas, R., Santos, C. N., Chen, Z., Batista, L. M., Duarte, J. D., Borges, A. & Teixeira, A. R. (2007), 'The role of plant defence proteins in fungal pathogenesis', *Molecular Plant Pathology* **8**(5), 677–700.

Feschotte, C., Zhang, X. & Wessler, S. R. (2002), 'Minature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons', *Mobile DNA II* pp. 1147–1158.

Figueres, R. (1940), Sur une maladie très grave du caféier en Oubangui, Technical report, Ministère des Colonies, Paris, France.

Fisher, M. C., Gurr, S. J., Cuomo, C. A., Blehert, D. S., Jin, H., Stukenbrock, E. H., Stajich, J. E., Kahmann, R., Boone, C., Denning, D. W., Gow, N. A., Klein, B. S., Kronstad, J. W., Sheppard, D. C., Taylor, J. W., Wright, G. D., Heitman, J., Casadevall, A. & Cowen, L. E. (2020), 'Threats posed by the fungal kingdom to humans, wildlife, and agriculture', *mBio* **11**(3).

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L. & Gurr, S. J. (2012), 'Emerging fungal threats to animal, plant and ecosystem health', *Nature* **484**(7393), 186–194.

Flood, J. (1996), A study of the Tracheomycosis or vascular wilt disease of coffee in Zaire, Technical report, Report presented to Zairian Coffee Organisation (OZACAF).

Flood, J. (2005), Epidemiology and variability of Gibberella xylarioides, the coffee wilt pathogen. R8188 (ZA0505). Final Technical Report, Technical report, CAB International, Wallingford, UK.

Flood, J. (2009), Introduction, *in* J. Flood, ed., 'Coffee Wilt Disease', pp. 1–7.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., Smit, A. F. & Org, A. (2020), 'RepeatModeler2: automated genomic discovery of transposable element families', *Proceedings of the National Academy of Sciences* **117**(17), 9451–9457.

Fones, H. N., Fisher, M. C. & Gurr, S. J. (2017), 'Emerging Fungal Threats to Plants and Animals Challenge Agriculture and Ecosystem Resilience', *Microbiol Spectrum* **5**(2), 787–809.

Fradin, E. F. & Thomma, B. P. (2006), 'Physiology and molecular aspects of Verticillium wilt diseases caused by V. dahliae and V. albo-atrum', *Molecular Plant Pathology* **7**(2), 71–86.

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P. & Panstruga, R. (2018), 'Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen', *BMC Genomics 2018 19:1* **19**(1), 1–23.

Frantzeskakis, L., Kusch, S. & Panstruga, R. (2019), 'The need for speed: compartmentalized genome evolution in filamentous phytopathogens', *Molecular Plant Pathology* **20**(1), 3.

Fraselle, J. (1950), 'Observations préliminaires sur une trachéomycose de Coffea robusta', *Bulletin Agricole du Congo Belge* **XLI**, 361–372.

Geiser, D. M., Lewis Ivey, M. L., Hakiza, G., Juba, J. H. & Miller, S. A. (2005), 'Gibberella xylarioides (anamorph: Fusarium xylarioides), a causative agent of coffee wilt disease in Africa, is a previously unrecognized member of the G. fujikuroi species complex', *Mycologia* **97**(1), 191–201.

Gerlach, W. (1978), Critical remarks on the present situation in Fusarium taxonomy, *in* 'Proceedings of the International Symposium on Taxonomy of Fungi, Part I.', University of Madras, Madras, pp. 115–124.

Gerlach, W. & Nirenberg, H. (1982), 'The genus Fusarium–a pictorial atlas.', *Land-und Forstwirtschaft Berlin-Dahlem* **209**.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., R C Miller, J. & Warren, R. A. (1991), 'Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families', *Microbiological Reviews* **55**(2), 303–315.

Giraud, T., Gladieux, P. & Gavrilets, S. (2010), 'Linking the emergence of fungal plant diseases with ecological speciation', *Trends in Ecology and Evolution* **25**(7), 387–395.

Giri, A. P., Harsulkar, A. M., Patankar, A. G., Gupta, V. S., Sainani, M. N., Deshpande, V. V. & Ranjekar, P. K. (1998), 'Association of induction of protease and chitinase in chickpea roots with resistance to Fusarium oxysporum f.sp. ciceri', *Plant Pathology* **47**(6), 693–699.

Girma, A. (2004), Diversity in Pathogenicity and Genetics of Gibberella xylarioides (Fusarium xylarioides) Populations and Resistance of Coffea spp. in Ethiopia, PhD thesis, Rheinischen Friedrich-Wilhelms-Universitat zu Bonn.

Girma, A., Million, A., Hindorf, H. & Arega, Z. (2009), Coffee wilt disease in Ethiopia, *in* J. Flood, ed., 'Coffee Wilt Disease', CAB International, Wallingford, UK, chapter Four, pp. 50–68.

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., De Vienne, D. M., Rodríguez De La Vega, R. C., Branco, S. & Giraud, T. (2014), 'Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes', *Wiley Online Library* **23**(4), 753–773.

Glazebrook, J. (2001), 'Genes controlling expression of defense responses in Arabidopsis — 2001 status', *Current Opinion in Plant Biology* **4**(4), 301–308.

Goubert, C., Craig, R. J., Bilat, A. F., Peona, V., Vogan, A. A. & Protasio, A. V.

(2022), 'A beginner's guide to manual curation of transposable elements', *Mobile DNA* **13**(1), 1–19.

Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P. & Boulesteix, M. (2015), 'De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti)', *Genome Biology and Evolution* **7**(4), 1192–1205.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011), 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nature Biotechnology 2011 29:7* **29**(7), 644–652.

Guindon, S., Delsuc, F., Dufayard, J. F. & Gascuel, O. (2009), 'Estimating maximum likelihood phylogenies with PhyML', *Methods in Molecular Biology* **537**, 113–137.

Guo, L., Han, L., Yang, L., Zeng, H., Fan, D., Zhu, Y., Feng, Y., Wang, G., Peng, C., Jiang, X., Zhou, D., Ni, P., Liang, C., Liu, L., Wang, J., Mao, C., Fang, X., Peng, M. & Huang, J. (2014), 'Genome and Transcriptome Analysis of the Fungal Pathogen Fusarium oxysporum f. sp. cubense Causing Banana Vascular Wilt Disease', *PLOS ONE* **9**(4), e95543.

Gurr, S., Samalova, M. & Fisher, M. (2011), 'The rise and rise of emerging infectious fungi challenges food security and ecosystem health', *Fungal Biology Reviews* **25**(4), 181–188.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N. & Regev, A. (2013), 'De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature Protocols 2013 8:8* **8**(8), 1494–1512.

Hakiza, G., Kyetere, D., Musoli, P., Wetala, P., Njuki, J., Kucel, P., Aluka, P., Kangire, A. & Ogwang, J. (2009), Coffee wilt disease in Uganda, *in* J. Flood, ed., 'Coffee Wilt Disease', CAB International, Wallingford, UK, pp. 28–49.

Hall, R. J., Whelan, F. J., McInerney, J. O., Ou, Y. & Domingo-Sananes, M. R. (2020), 'Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes', *Frontiers in Microbiology* **11**, 1569.

Han, Y., Liu, X., Benny, U., Corby Kistler, H. & VanEtten, H. D. (2001), 'Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen Nectria haematococca', *Plant Journal* **25**(3), 305–314.

Hane, J. K. & Oliver, R. P. (2008), 'RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences', *BMC Bioinformatics* **9**(478), 1–12.

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G. & Chen, J. (2020), 'RIdeogram: Drawing SVG graphics to visualize and map genome-wide data on the idiograms', *PeerJ Computer Science* **6**, 1–11.

Harris, S. D. (2017), 'Morphogenesis in germinating Fusarium graminearum macroconidia', *Mycologia* **97**(4), 880–887.

Heim, R. & Saccas (1950), 'Gibberella xylarioides R. Heim & Saccas', *Rev. Mycol.* **97**, 97.

Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J. & Knoxa, J. P. (2010), 'Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects', *Proceedings of the National Academy of Sciences of the United States of America* **107**(34), 15293–15298.

Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. (2019), 'Whole-genome annotation with BRAKER', *Methods in Molecular Biology* **1962**, 65–95.

Hoff, K., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. (2016), 'BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS', *Bioinformatics* **32**(5), 767–769.

Hogg, D., Pell, G., Dupree, P., Goubet, F., Martin-Oru; E, S. M., Armand, S. & Gilbert, H. J. (2003), 'The modular architecture of Cellvibrio japonicus mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation', *Biochemical Journal* **371**(3), 1027–1043.

Houterman, P. M., Speijer, D., Dekker, H. L., De Koster, C. G., Cornelissen, B. J. C. & Rep, M. (2007), 'The mixed xylem sap proteome of Fusarium oxysporum-infected tomato plants', *Molecular Plant Pathology* **8**(2), 215–221.

Howell, C., Bell, A. & Stipanovic, R. (1976), 'Effect of aging on flavonoid content and resistance of cotton leaves to Verticillium wilt', *Physiological Plant Pathology* **8**(2), 181–188.

Hudson, R. R. & Wu, C. I. (1992), 'Gene Trees, Species Trees and the Segregation of Ancestral Alleles', *Genetics* **131**(2), 509.

ICO (2022), 'International Coffee Organisation Trade Statistics'.

Inglis, P. W., Rigden, D. J., Mello, L. V., Louis, E. J. & Valadares-Inglis, M. C. (2005), 'Monomorphic subtelomeric DNA in the filamentous fungus, Metarhizium anisopliae, contains a RecQ helicase-like gene', *Molecular Genetics and Genomics* **274**(1), 79–90.

Inoue, I., Namiki, F. & Tsuge, T. (2002), 'Plant Colonization by the Vascular Wilt Fungus Fusarium oxysporum Requires FOW1 , a Gene Encoding a Mitochondrial Protein', *The Plant Cell* **14**, 1869–1883.

Jha, G., Rajeshwari, R. & Sonti, R. V. (2007), 'Bacterial Type Two Secretion System Secreted Proteins: Double-Edged Swords for Plant Pathogens', *https://doi.org/10.1094/MPMI-18-0891* **18**(9), 891–898.

Jones, D. A., Bertazzoni, S., Turo, C. J., Syme, R. A. & Hane, J. K. (2018), 'Bioinformatic prediction of plant–pathogenicity effector proteins of fungi', *Current Opinion in Microbiology* **46**, 43–49.

Jones, J. D. G. & Dangl, J. L. (2006), 'The plant immune system', *Nature* **444**, 323–329.

Jones, P., Binns, D., Chang, H., Fraser, M. & Li, W. (2014), 'InterProScan 5: genome-scale protein function classification', *Bioinformatics* **30**(9), 1236–1240.

Kalonji-Mbuyi, A., Tshilenge-Djim, P. & Saiba, N. (2009), Coffee wilt disease in Democratic Republic of Congo, *in* J Flood, ed., 'Coffee Wilt Disease', CAB International, Wallingford, UK, pp. 7–27.

Kangire, A., Kytere, D., Hakiza, G., Warren, H., Erbaugh, M. & Kabole, C. (2002), Determination of alternate hosts for Fx, the causal agent of Coffee Wilt Disease (CWD) in Uganda, Technical report, CORI, Uganda.

Kellett, L. E., Poole, D. M., Ferreira, L. M. A., Durrant, A. J., Hazlewoodt, G. P. & Gilbert, H. J. (1990), 'Xylanase B and an arabinofuranosidase from Pseudomonas fluorescens subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes', *Biochemical Journal* **272**, 369–376.

Kilambo, D. & Kaiza, D. (1997), Investigation of Fusarium xylarioides on robusta coffee in Misenyi division Bukoba district., Technical report, Lyamungu Misc. Reports.

Klittich, C. J. R., Leslie, J. F., Nelson, P. E. & Marasas, W. F. O. (1997), 'Fusarium thapsinum (Gibberella thapsina): A new species in section Liseola from sorghum', *Mycologia* **89**(4), 643–652.

Kloesges, T., Popa, O., Martin, W. & Dagan, T. (2011), 'Networks of Gene Sharing among 329 Proteobacterial Genomes Reveal Differences in Lateral Gene Transfer Frequency at Different Phylogenetic Depths', *Molecular Biology and Evolution* **28**(2), 1057–1074.

Klosterman, S. J., Atallah, Z. K., Vallad, G. E. & Subbarao, K. V. (2009), 'Diversity, Pathogenicity, and Management of Verticillium Species', *https://doi.org/10.1146/annurev-phyto-080508-081748* **47**, 39–62.

Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B. P., Chen, Z., Henrissat, B., Lee, Y. H., Park, J., Garcia-Pedrajas, M. D., Barbara, D. J., Anchieta, A., de Jonge, R., Santhanam, P., Maruthachalam, K., Atallah, Z., Amyotte, S. G., Paz, Z., Inderbitzin, P., Hayes, R. J., Heiman, D. I., Young, S., Zeng, Q., Engels, R., Galagan, J., Cuomo, C. A., Dobinson, K. F. & Ma, L. J. (2011), 'Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens', *PLoS Pathogens* **7**(7).

Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. (2019), 'Assembly of long, error-prone reads using repeat graphs', *Nature Biotechnology 2019 37:5* **37**(5), 540–546.

Kranz, J. & Mogk, M. (1973), 'Gibberella xylarioides Heim et Saccas on arabica coffee in Ethiopia', *Phytopathologische Zeitschrift* **78**(4), 365–366.

Kurian, S. M., Di Pietro, A. & Read, N. D. (2018), 'Live-cell imaging of conidial anastomosis tube fusion during colony initiation in Fusarium oxysporum', *PLOS ONE* **13**(5), e0195634.

Kvas, M., Marasas, W., Wingfield, B., Wingfield, M. & Steenkamp, E. (2009), 'Diversity and evolution of Fusarium species in the Gibberella fujikuroi complex', *Fungal Diversity* **34**, 1–21.

Langcake, P. & Drysdale, R. (1975), 'The role of pectic enzyme production in the resistance of tomato to Fusarium oxysporum f. lycopersici', *Physiological Plant* **6**(3), 247–258.

Langenbach, C., Campe, R., Beyer, S. F., Mueller, A. N. & Conrath, U. (2016), 'Fighting Asian soybean rust', *Frontiers in Plant Science* **7**(797).

Lannoo, N. & Van Damme, E. J. (2014), 'Lectin domains at the frontiers of plant defense', *Frontiers in Plant Science* **5**(397).

Leitner, M., Vandelle, E., Gaupels, F., Bellin, D. & Delledonne, M. (2009), 'NO signals in the haze: Nitric oxide signalling in plant defence', *Current Opinion in Plant Biology* **12**(4), 451–458.

Lejeune, J. B. H. (1958), Rapport au Gouvernement imperial d'Ethiopie sur la production, Technical report, FAO Rapport PEAT 797, Rome.

Leonard, K. J. & Szabo, L. J. (2005), 'Stem rust of small grains and grasses caused by Puccinia graminis', *Molecular Plant Pathology* **6**(2), 99–111.

Lepoint, P. C. E. (2006), Speciation within the African Coffee Wilt Pathogen, PhD thesis, Universite Catholique de Louvain.

Lepoint, P. C. E., Munaut, F. T. J. & Maraite, H. M. M. (2005), 'Gibberella xylarioides Sensu Lato from Coffea canephora: a New Mating Population in the Gibberella fujikuroi Species Complex', *Applied and Environmental Microbiology* **71**(12), 8466–8471.

Leslie, F. J. & Summerell, B. (2006), *The Fusarium Laboratory Manual*, Blackwell Publishing, Ames, Iowa, USA.

Li, D., Liu, C., Luo, R., Sadakane, K. & Lam, T. (2015*a*), 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics* **31**(10), 1674–1676.

Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. (2015*b*), 'MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics* **31**(10), 1674–1676.

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., Yamashita, H. & Lam, T. W. (2016), 'MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices', *Methods* **102**, 3–11.

Li, H. (2018), 'Minimap2: pairwise alignment for nucleotide sequences', *Bioinformatics* **34**(18), 3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009), 'The sequence alignment/map format and SAMtools', *Bioinformatics* **25**(16), 2078–2079.

Li, J., Fokkens, L. & Rep, M. (2021), 'A single gene in Fusarium oxysporum limits host range', *Molecular Plant Pathology* **22**(1), 108–116.

Li, J., Fokkens, L., van Dam, P. & Rep, M. (2020), 'Related mobile pathogenicity chromosomes in Fusarium oxysporum determine host range on cucurbits', *Molecular Plant Pathology* **21**(6), 761–776.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S. & Kahmann, R. (2015), 'Fungal Effectors and Plant Susceptibility', *Annu. Rev. Plant Biol* **66**, 513–545.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. & Henrissat, B. (2013), 'The carbohydrate-active enzymes database (CAZy) in 2013', *Nucleic Acids Research* **42**(D1), D490–D495.

Love, M. I., Huber, W. & Anders, S. (2014), 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology* **15**(12), 1–21.

Lu, S. & Edwards, M. C. (2016), 'Genetics and Resistance Genome-Wide Analysis of Small Secreted Cysteine-Rich Proteins Identifies Candidate Effector Proteins Potentially Involved in Fusarium graminearum Wheat Interactions', *Phytopathology* **106**(2), 166–176.

Ma, L.-J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'donnell, K., Trail, F., Gardiner, D. M., Manners, J. M. & Kazan, K. (2013), 'Fusarium Pathogenomics', *Annu. Rev. Microbiol.* **67**, 399–416.

Ma, L. J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W. B., Woloshuk, C., Xie, X., Xu, J. R., Antoniw, J., Baker, S. E., Bluhm, B. H., Breakspear, A., Brown, D. W., Butchko, R. A., Chapman, S., Coulson, R., Coutinho, P. M., Danchin, E. G., Diener, A., Gale, L. R., Gardiner, D. M., Goff, S., Hammond-Kosack, K. E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C. D., Koehrsen, M., Kumar, L., Lee, Y. H., Li, L., Manners, J. M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S. Y., Proctor, R. H., Regev, A., Ruiz-Roldan, M. C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D. C., Turgeon, B. G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C. A., Kistler, H. C. & Rep, M. (2010), 'Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium', *Nature* **464**(7287), 367–373.

Mabberley, D. (1997), *The Plant Book*, Cambridge Univeristy Press, Cambridge.

Mace, M., Bell, A. & Beckman, C. (1981), *Fungal Wilt Diseases of Plants*, Academic Press.

Magoč, T. & Salzberg, S. L. (2011), 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics* **27**(21), 2957–2963.

Martin, M. (2011), 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal* **17**(1), 10–12.

Martínez-Rocha, A. L., Roncero, M. I. G., López-Ramirez, A., Mariné, M., Guarro, J., Mart´ınez-Cadena, G. & Di Pietro, A. (2008), 'Rho1 has distinct functions in morphogenesis, cell wall biosynthesis and virulence of Fusarium oxysporum', *Cellular Microbiology* **10**(6), 1339–1351.

Mbalibulha, H., Kibanzanga Christopher, T., Vincent, H. & Sempijja, B. (2015), Annual Report, Technical report, Uganda Coffee Development Authority, Uganda.

McCartney, L., Blake, A., Flint, J., Bolam, D., Boraston, A., Gilbert, H., Knox, J. & 2006, U. (2006), 'Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules', *Proceedings of the National Academy of Sciences* **103**(12), 4765–4770.

McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A. & Solomon, P. S. (2019), 'Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens', *mBio* **10**(5), 01515–19.

McKie, V. A., Vinoken, J. P., Voragen, A. G., Van Den Broek, L. A., Stimson, E. & Gilbert, H. J. (2001), 'A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose', *Biochemical Journal* **355**(1), 167–177.

Mesny, F., Miyauchi, S., Thiergart, T., Pickel, B., Atanasova, L., Karlsson, M., Hüttel, B., Barry, K. W., Haridas, S., Chen, C., Bauer, D., Andreopoulos, W., Pangilinan, J., LaButti, K., Riley, R., Lipzen, A., Clum, A., Drula, E., Henrissat, B., Kohler, A., Grigoriev, I. V., Martin, F. M. & Hacquard, S. (2021), 'Genetic determinants of endophytism in the Arabidopsis root mycobiome', *Nature communications* **12**(1).

Michielse, C. B. & Rep, M. (2009), 'Pathogen profile update: Fusarium oxysporum', *Molecular Plant Pathology* **10**(3), 311–324.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. (2018), 'Versatile genome assembly evaluation with QUAST-LG', *Bioinformatics* **34**, 142–150.

Misra, B. B. & Chaturvedi, R. (2015), 'When plants brace for the emerging pathogens', *Physiological and Molecular Plant Pathology* **92**, 181–185.

Möller, M. & Stukenbrock, E. H. (2017), 'Evolution and genome architecture in fungal plant pathogens', *Nature Reviews Microbiology* **15**(12), 756–771.

Mulatu, A. & Shanko, D. (2019), 'Incidence and Prevalence of Coffee wilt Disease (Gibberella xylarioides) and Its Impact on the Rural Livelihoods in Western Guji Zone , Southern Ethiopia', *American Journal of Bioscience* **7**(1), 7–15.

Munoz-Lopez, M. & Garcia-Perez, J. L. (2010), 'DNA Transposons: Nature and Applications in Genomics', *Current Genomics* **11**(2), 115–128.

Musoli, P., Girma, A., Hakiza, G., Kangire, A., Pinard, F., Agwanda, C. & Bieysse, D. (2009), Breeding for Resistance Against Coffee Wilt Disease, *in* J. Flood, ed., 'Coffee Wilt Disease', CAB International, Wallingford, UK, pp. 155–175.

Muszewska, A., Hoffman-Sommer, M. & Grynberg, M. (2011), 'LTR retrotransposons in fungi', *PLoS ONE* **6**(12).

Nei, M. & Li, W. H. (1979), 'Mathematical model for studying genetic variation in terms of restriction endonucleases.', *Proceedings of the National Academy of Sciences* **76**(10), 5269–5273.

Nelson, P., Toussoun, T. & Marasas, W. (1983), *Fusarium species: an illustrated manual for identification.*, Pennsylvania State University Press, Pennsylvania.

Nielsen, H., Engelbrecht, J., Brunak, S. & Von Heijne, G. (1997), 'Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites', *Protein Engineering* **10**(1), 1–6.

Notredame, C., Higgins, D. G. & Heringa, J. (2000), 'T-Coffee: A novel method for fast and accurate multiple sequence alignment', *Journal of molecular biology* **302**(1), 205–217.

Novichkov, V., Kaznadzey, A., Alexandrova, N. & Kaznadzey, D. (2016), 'NSimScan:

DNA comparison tool with increased speed, sensitivity and accuracy', *Bioinformatics (Oxford, England)* **32**(15), 2380–2381.

O'Donnell, K., Cigelnik, E. & Nirenberg, H. I. (1998), 'Molecular Systematics and Phylogeography of the Gibberella fujikuroi Species Complex', *Mycologia* **90**(3), 465.

O'Donnell, K., Kistler, H. C., Cigelnik, E. & Ploetz, R. C. (1998), 'Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies', *Proceedings of the National Academy of Sciences* **95**(5), 2044–2049.

O'Donnell, K., Rooney, A. P., Proctor, R. H., Brown, D. W., McCormick, S. P., Ward, T. J., Frandsen, R. J., Lysøe, E., Rehner, S. A., Aoki, T., Robert, V. A., Crous, P. W., Groenewald, J. Z., Kang, S. & Geiser, D. M. (2013), 'Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria', *Fungal Genetics and Biology* **52**, 20–31.

Oduor, G., Simons, S., Phiri, N., Njuki, J., Poole, J., Pinard, F., Kyeter, D., Hakiza, G., Musoli, P. & Lukwago, G. (2003), Surveys to assess the extent and impact of coffee wilt disease in East and Central Africa. Final Technical Report., Technical report, CAB International, Nairobi, Kenya.

Olal, S., Bitalo, D. N., Olango, N. D., Mulindwa, J., Ochwo, S., Opiyo, S. O., Arinaitwe, G. & Ogwok, E. (2019), 'De Novo Genome Sequence of a Fusarium xylarioides Race Pathogenic to Robusta Coffee (Coffea canephora) in Uganda', *Microbiol Resource Announcements* **8:e00520-19**.

Olal, S., Olango, N., Kiggundu, A., Ochwo, S., Adriko, E., Nanteza, A., Matovu, E., Gw, L., Kagezi, G., Hakiza, G. J., Wagoire, W., Rutherford, M. & Opiyo, S. (2018), 'Using Translation Elongation Factor Gene to specifically detect and diagnose Fusarium xylaroides, a Causative Agent of Coffee Wilt Disease in Ethiopia, east and central Africa', *Journal of Plant Pathology and Microbiology* **9**(440), 2.

Onesirosan, P. T. & Fatunla, T. (1976), 'Fungal Fruit Rots of Tomatoes in Southern Nigeria', *Journal of Horticultural Science* **51**(4), 473–479.

Ospina-Giraldo, M. D., Mullins, E. & Kang, S. (2003), 'Loss of function of the Fusarium oxysporum SNF1 gene reduces virulence on cabbage and Arabidopsis', *Current Genetics* **44**(1), 49–57.

Palmer, J. M. & Stajich, J. (2020), 'Funannotate v1.8.1: Eukaryotic genome annotation'.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. (2017), 'Salmon provides fast and bias-aware quantification of transcript expression', *Nature Methods 2017 14:4* **14**(4), 417–419.

Peck, L. (n.d.), Coffee wilt disease - the forgotten threat to coffee. In press, Technical report, British Society for Plant Pathology, London.

Peck, L. D., Nowell, R. W., Flood, J., Ryan, M. R. & Barraclough, T. G. (2021), 'Historical genomics reveals the evolutionary mechanisms behind multiple outbreaks of the host-specific coffee wilt pathogen Fusarium xylarioides', *BMC Genomics* **22**(1), 1–24.

Pemberton, C. L. & Salmond, G. P. (2004), 'The Nep1-like proteins - A growing family of microbial elicitors of plant necrosis', *Molecular Plant Pathology* **5**(4), 353–359.

Pennington, H. G., Li, L. & Spanu, P. D. (2016), 'Identification and selection of normalization controls for quantitative transcript analysis in Blumeria graminis', *Molecular Plant Pathology* **17**(4), 625–633.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. (2014), 'PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R', *Molecular Biology and Evolution* **31**(7), 1929–1936.

Phiri, N. & Baker, P. (2009), Coffee Wilt in Africa: Final Technical Report of the Regional Coffee Wilt Programme (2000-2007), Technical report, CABI, Wallingford, UK: CAB International.

Phiri, N., Kimani, M., Negussie, E., Simons, S. & Oduor, G. (2009), Management of Coffee Wilt Disease, *in* J. Flood, ed., 'Coffee Wilt Disease', CAB International, Wallingford, chapter 9, pp. 137–154.

Pizzinatto, M. & Menten, J. (1991), 'Pathogenicity of eight Fusarium species isolated from seeds to cotton seedlings.', *Summa Phytopathologica* **17**(2), 124–134.

R Core Team (2021), 'R: A Language and Environment for Statistical Computing'.

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H. Y., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C. & Kamoun, S. (2010), 'Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage', *Science* **330**(6010), 1540–1543.

Raffaele, S. & Kamoun, S. (2012), 'Genome evolution in filamentous plant pathogens: why bigger can be better', *Nature Reviews Microbiology* **10**(6), 417–430.

Ram, K. & Wickham, H. (2018), 'wesanderson: A Wes Anderson Palette Generator'.

Ranwez, V., Douzery, E. J., Cambon, C., Chantret, N. & Delsuc, F. (2018), 'MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons', *Molecular Biology and Evolution* **35**(10), 2582–2584.

Redkar, A., Sabale, M., Schudoma, C., Zechmann, B., Gupta, Y. K., López-Berges, M. S., Venturini, G., Gimenez-Ibanez, S., Turrà, D., Solano, R. & Di Pietro, A. (2022), 'Conserved secreted effectors contribute to endophytic growth and multihost plant compatibility in a vascular wilt fungus', *The Plant Cell* **34**(9), 3214–3232.

Reimer-Michalski, E. M. & Conrath, U. (2016), 'Innate immune memory in plants', *Seminars in Immunology* **28**(4), 319–327.

Reis, H., Pfiff, S. & Hahn, M. (2005), 'Molecular and functional characterization of a secreted lipase from Botrytis cinerea', *Molecular Plant Pathology* **6**(3), 257–267.

Rep, M. (2005), 'Small proteins of plant-pathogenic fungi secreted during host colonization', *FEMS Microbiology Letters* **253**(1), 19–27.

Richards, T. A., Soanes, D. M., Jones, M. D., Vasieva, O., Leonard, G., Paszkiewicz, K., Foster, P. G., Hall, N. & Talbot, N. J. (2011), 'Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes', *Proceedings of the National Academy of Sciences of the United States of America* **108**(37), 15258–15263.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. (2015), 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research* **43**(7), e47–e47.

Rodrigues, C. J., Medeiros, E. F. & Lewis, B. G. (1975), 'Relationship between a phytoalexin-like response in coffee leaves (Coffea arabica L.) and compatibility with Hemileia vastatrix Berk. & Br', *Physiological Plant Pathology* **6**(1), 35–41.

Rogers, L. M., Kim, Y.-K., Guo, W., Gonzá Lez-Candelas, L., Li, D., Kolattukudy, P. E. & Ryan, C. A. (2000), 'Requirement for either a host-or pectin-induced pectate lyase for infection of Pisum sativum by Nectria hematococca', *National Acad Sciences* **97**(17).

Ropars, J., Didiot, E., Rodríguez de la Vega, R. C., Bennetot, B., Coton, M., Poirier, E., Coton, E., Snirc, A., Le Prieur, S. & Giraud, T. (2020), 'Domestication of the Emblematic White Cheese-Making Fungus Penicillium camemberti and Its Diversification into Two Varieties', *Current Biology* **30**(22), 4441–4453.

Ropars, J., Rodríguez De La Vega, R. C., Ló Pez, M., Dupont, L., Branca, A. & Giraud, T. (2015), 'Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi', *Current Biology* **25**, 2562–2569.

Rossoni, A. W., Price, D. C., Seger, M., Lyska, D., Lammers, P., Bhattacharya, D. & Weber, A. P. (2019), 'The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions', *eLife* **8**.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, R., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., J Kema, G. H., Lapalu, N., Lawrence, C. B., May, K., Meyer, M., Ollivier, D., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Gillian Turgeon, B., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, L., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, N. & Howlett, B. J. (2011), 'Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations', *Nature Communications* **2**(1), 1–10.

Rubio-Piña, J. A. & Zapata-Pérez, O. (2011), 'Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants', *Electronic Journal of Biotechnology* **14**(5).

Rutherford, M. (2005), 'Epidemiology and variability of Gibberella xylarioides, the coffee wilt pathogen. R8188 (ZA0505). Final Technical Report'.

Rutherford, M. (2006), 'Current Knowledge of Coffee Wilt Disease, a Major Constraint to Coffee Production in Africa', *Phytopathology* **96**(6), 663–666.

Rutherford, M., Bieysse, D., Lepoint, P. & Maraite, H. (2009), Biology, taxonomy and epidemiology of the coffee wilt pathogen Gibberella xylarioides sensu lato, *in* J. Flood, ed., 'Coffee Wilt Disease', pp. 99–119.

Ryan, M. J., Peck, L. D., Smith, D., Flood, J. & Barraclough, T. G. (2022), 'Culture collections as a source of historic strains for genomic studies in plant pathology', *Journal of Plant Pathology* .

Saccas, A. (1956), 'Recherches expérimentales sur la trachéomycose des caféiers en Oubangui-Chari', *Agronomie Tropicale* **11**, 7–58.

Sanchez-Vallet, A., Zurich, E., Fouche, S., Hartmann, F., Soyer, J., Tellier, A. & Croll,

D. (2018), 'The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens', *Annual Review of Phytopathology* **56**, 21–40.

Schaefer, H. M., Rentzsch, M. & Breuer, M. (2009), 'Anthocyanins Reduce Fungal Growth in Fruits', *Natural Product Communications* **3**(8), 1267–1272.

Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L. & Rep, M. (2013), 'MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum.', *BMC genomics* **14**(119), 1–21.

Schumann, G. L., D'arcy, C. J., Schumann, G. L. & D'arcy, C. J. (2010), *Essential Plant Pathology*, second edition edn, The American Phytopathological Society, St. Paul, Minnesota, USA.

Serani, S., Taligoola, H. K. & Hakiza, G. J. (2007), 'An investigation into Fusarium spp. associated with coffee and banana plants as potential pathogens of robusta coffee', *African Journal of Ecology* **45**, 91–95.

Shahi, S., Beerens, B., Bosch, M., Linmans, J. & Rep, M. (2016), 'Nuclear dynamics and genetic rearrangement in heterokaryotic colonies of Fusarium oxysporum', *Fungal Genetics and Biology* **91**, 20–31.

Sharma, M., Sengupta, A., Ghosh, R., Agarwal, G., Tarafdar, A., Nagavardhini, A., Pande, S. & Varshney, R. K. (2016), 'Genome wide transcriptome profiling of Fusarium oxysporum f sp. ciceris conidial germination reveals new insights into infection-related genes', *Scientific Reports* **6**(37353), 1–11.

Shen, W., Le, S., Li, Y. & Hu, F. (2016), 'SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation', *PLoS ONE* **11**(10), e0163962.

Shumate, A. & Salzberg, S. L. (2021), 'Liftoff: accurate mapping of gene annotations', *Bioinformatics* **37**(12), 1639–1643.

Silva, G., Green, K., Dutilh, B. & Edwards, R. (2016), 'SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data', *Bioinformatics* **32**(3), 354–361.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. (2015), 'BUSCO: user guide', *Bioinformatics* **31**(19), 3210–3212.

Slowikowski, K. (2021), 'ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'''.

Smit, A. F. A., Hubley, R. & Green, P. (2015), 'RepeatMasker Open-4.0.'.

Soanes, D. M., Alam, I., Cornell, M., Wong, H. M., Hedeler, C., Paton, N. W., Rattray, M., Hubbard, S. J., Oliver, S. G. & Talbot, N. J. (2008), 'Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis', *PLoS ONE* **3**(6), e2300.

Soneson, C. & Robinson, M. D. (2018), 'Bias, robustness and scalability in single-cell differential expression analysis', *Nature Methods 2018 15:4* **15**(4), 255–261.

Song, M., Xu, W., Xiang, Y., Jia, H., Zhang, L. & Ma, Z. (2014), 'Association of jacalin-related lectins with wheat responses to stresses revealed by transcriptional profiling', *Plant Molecular Biology* **84**(1-2), 95–110.

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., Van Themaat, E. V. L., Brown, J. K., Butcher, S. A., Gurr, S. J., Lebrun, M. H., Ridout, C. J., Schulze-Lefert, P., Talbot, N. J., Ahmadinejad, N., Ametz, C., Barton, G. R., Benjdia, M., Bidzinski, P., Bindschedler, L. V., Both, M., Brewer, M. T., Cadle-Davidson, L., Cadle-Davidson, M. M., Collemare, J., Cramer, R., Frenkel, O., Godfrey, D., Harriman, J., Hoede, C., King, B. C., Klages, S., Kleemann, J., Knoll, D., Koti, P. S., Kreplak, J., López-Ruiz, F. J., Lu, X., Maekawa, T., Mahanil, S., Micali, C., Milgroom, M. G., Montana, G., Noir, S., O'Connell, R. J., Oberhaensli, S., Parlange, F., Pedersen, C., Quesneville, H., Reinhardt, R., Rott, M., Sacristán, S., Schmidt,

S. M., Schön, M., Skamnioti, P., Sommer, H., Stephens, A., Takahara, H., Thordal-Christensen, H., Vigouroux, M., Weßling, R., Wicker, T. & Panstruga, R. (2010), 'Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism', *Science* **330**(6010), 1543–1546.

Sperschneider, J. & Dodds, P. N. (2022), 'EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes', *Molecular Plant-Microbe Interactions* pp. 146–156.

Sperschneider, J., Gardiner, D. M., Thatcher, L. F., Lyons, R., Singh, K. B., Manners, J. M. & Taylor, J. M. (2015), 'Genome-Wide Analysis in Three Fusarium Pathogens Identifies Rapidly Evolving Chromosomes and Genes Associated with Pathogenicity', *Genome Biol. Evol.* **7**(6), 1613–1627.

Srivastava, A. K., Kashyap, P. L., Chakdar, H., Kumar, M., Srivastava, A. K., Yadav, J., Jamali, H., Srivastava, R., Sharma, A., Tiwari, P., Singh, A. & Saxena, A. K. (2018), ' First De Novo Draft Genome Sequence of the Pathogenic Fungus Fusarium udum F02845, Associated with Pigeonpea (Cajanus cajan L. Millspaugh) Wilt ', *Microbiology Resource Announcements* **7**(13).

Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. (2008), 'Using native and syntenically mapped cDNA alignments to improve de novo gene finding', *Bioinformatics* **24**(5), 637–644.

Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. (2006), 'Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources', *BMC Bioinformatics* **7**(62), 1–11.

Stergiopoulos, I., Van Den Burg, H. A., Ökmen, B., Beenen, H. G., Van Liere, S., Kema, G. H. & De Wit, P. J. (2010), 'Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots', *Proceedings*

of the National Academy of Sciences of the United States of America **107**(16), 7610–7615.

Stewart, R. B. (1957), 'Some plant diseases occurring in Kaffa province, Ethiopia ', *Imperial Ethiopian College of Agricultural and Mechanical Arts* **3**, 15–16.

Steyaert, R. (1948), 'Contribution à l'étude des parasites des végétaux du Congo', *Bulletin de la Société Royale de Botanique de Belgique* **80**(1-2), 11–58.

Strange, R. N. & Scott, P. R. (2005), 'PLANT DISEASE: A Threat to Global Food Security', *Annu. Rev. Phytopathol* **43**, 83–116.

Stukenbrock, E. H. (2013), 'Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens', *New Phytologist* **199**(4), 895–907.

Sun, Q. H., Hu, J., Huang, G. X., Ge, C., Fang, R. X. & He, C. Z. (2005), 'Type-II secretion pathway structural gene xpsE, xylanase- and cellulase secretion and virulence in Xanthomonas oryzae pv. oryzae', *Plant Pathology* **54**(1), 15–21.

Talboys, P. W. (1972), 'Resistance to Vascular Wilt Fungi', *The Royal Society Publishing* **181**, 319–332.

Talboys, P. W. & Busch, L. V. (1970), 'Pectic enzymes produced by Verticillium species', *Transactions of the British Mycological Society* **55**(3), 367–381.

Tanaka, S. & Kahmann, R. (2021), 'Cell wall-associated effectors of plant-colonizing fungi', *Mycologia* **113**(2).

Tjamos, E. & Beckman, C. (1989), *Vascular Wilt Diseases of Plants: Basic Studies and Control (NATO ASI Series H: Cell Biology)*, Springer-Verlag, Berlin.

Torres, D. E., Oggenfuss, U., Croll, D. & Seidl, M. F. (2020), 'Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model', *Fungal Biology Reviews* **34**(3), 136–143.

Torres, M. A., Jones, J. D. & Dangl, J. L. (2006), 'Reactive Oxygen Species Signaling in Response to Pathogens', *Plant Physiology* **141**(2), 373–378.

Tshilenge-Djim, P., Muengula-Manyi, M., Kingunza-Mawanga, K., Ngombo-Nzokwani, A., Tshilenge-Lukanda, L. & Kalonji-Mbuyi, A. (2016), 'Field Assessment of the Potential Role of Fusarium Species in the Pathogenesis of Coffee Wilt Disease in Democratic Republic of Congo', *Asian Research Journal of Agriculture* **2**(3), 1–7.

Tyree, M. & Zimmermann, M. (2013), *Xylem structure and the ascent of sap*, Springer, Berlin.

URL: https://books.google.com/books?hl=en&lr=&id=wnv1CAAAQBAJ&oi=fnd&pg=PA1&dq=Tyree,+M.+T.,+and+Zimmermann,+M.H.+(2002).Xylem+Structure+and+theAscent+of+Sap.+Berlin:+Springer.&ots=87Migqq_gY&sig=Ms4TsKWITD3Ty48H-e1C1_eSBS4

Van Dam, P., Fokkens, L., Ayukawa, Y., Van Der Gragt, M., Ter Horst, A., Brankovics, B., Houterman, P. M., Arie, T. & Rep, M. (2017), 'A mobile pathogenicity chromosome in Fusarium oxysporum for infection of multiple cucurbit species', *Scientific Reports* **7**(1), 1–15.

van Dam, P., Fokkens, L., Schmidt, S. M., Linmans, J. H., Corby Kistler, H., Ma, L. J. & Rep, M. (2016), 'Effector profiles distinguish formae speciales of fusarium oxysporum', *Environmental Microbiology* **18**(11), 4087–4102.

van Dam, P. & Rep, M. (2017), 'The Distribution of Miniature Impala Elements and SIX Genes in the Fusarium Genus is Suggestive of Horizontal Gene Transfer', *Journal of Molecular Evolution* **85**(1-2), 14–25.

Van Der Graaff, N. A. & Pieters, R. (1978), 'Resistance levels in Coffea arabica to Gibberella xylarioides and distribution pattern of the disease', *Neth. J. Pl. Path* **84**, 117–120.

Van Esse, P. H., Van't Klooster, J. W., Bolton, M. D., Yadeta, K. A., Van Baarlen, P., Boeren, S., Vervoort, J., De Wit, P. J. G. M. & Thomma, B. P. H. J. (2008), 'The Cladosporium fulvum Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense W OA', *Am Soc Plant Biol* **20**, 1948–1963.

Van Etten, J. & Bhattacharya, D. (2020), 'Horizontal Gene Transfer in Eukaryotes: Not if, but How Much?', *Trends in Genetics* **36**(12), 915–925.

Van Wyk, S., Harrison, C. H., Wingfield, B. D., De Vos, L., Van Der Merwe, N. A. & Steenkamp, E. T. (2019), 'The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations', *PeerJ* **7**(e7447).

Vega, F. E., Rosenquist, E. & Collins, W. (2003), 'Global project needed to tackle coffee crisis', *Nature* **425**(6956), 343–343.

Wang, M., Zhao, Y. & Zhang, B. (2015), 'Efficient test and visualization of multi-set intersections', *Scientific Reports* **5**(16923), 1–12.

Ward, T. J., Clear, R. M., Rooney, A. P., O'donnell, K., Gaba, D., Patrick, S., Starkey, D. E., Gilbert, J., Geiser, D. M. & Nowicki, T. W. (2008), 'An adaptive evolutionary shift in Fusarium head blight pathogen populations is driving the rapid spread of more toxigenic Fusarium graminearum in North America', *Fungal Genetics and Biology* **45**, 473–484.

Wasmann, C. & VanEtten, H. (1996), 'Transformation-mediated chromosome loss and disruption of a gene for pisatin demethylase decrease the virulence of Nectria haematococca on pea', *Molecular Plant-Microbe Interactions* **9**(9), 793–803.

Watterson, G. A. (1975), 'On the number of segregating sites in genetical models without recombination', *Theoretical Population Biology* **7**(2), 256–276.

Wei, K. H.-C., Aldaimalani, R., Mai, D., Zinshteyn, D., Satyaki, P., Blumenstiel, J.,

Kelleher, E. S. & Brooks, E. C. (2022), 'Rethinking the "gypsy" retrotransposon: A roadmap for community-driven reconsideration of problematic gene names', *BioRxiv* .

Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019), 'Welcome to the Tidyverse', *Journal of Open Source Software* **4**(43), 1686.

Wilke, C. (2022), 'Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2''.

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., Kidd, B. N., Anderson, J. P., Ghosh, R., Garg, G., Lichtenzveig, J., Kistler, H. C., Shea, T., Young, S., Buck, S. A. G., Kamphuis, L. G., Saxena, R., Pande, S., Ma, L. J., Varshney, R. K. & Singh, K. B. (2016), 'Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting Fusarium oxysporum formae speciales facilitates identification of candidate effectors', *BMC genomics* **17**(1).

Wingfield, B. D., Fourie, A., Simpson, M. C., Bushula-Njah, V. S., Aylward, J., Barnes, I., Coetzee, M. P. A., Dreyer, L. L., Duong, T. A., Geiser, D. M., Roets, F., Steenkamp, E. T., van der Nest, M. A., van Heerden, C. J. & Wingfield, M. J. (2019), 'IMA Genome-F 11', *IMA Fungus* **10**(13).

Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., Ma, Z. & Wang, Y. (2011), 'A jacalin-related lectin-like gene in wheat is a component of the plant defence system', *Journal of Experimental Botany* **62**(15), 5471–5483.

Yadeta, K. A. & Thomma, B. P. H. J. (2013), 'The xylem as battleground for plant hosts and vascular wilt pathogens', *Frontiers in Plant Science* **4**(97).

Yang, Z. (2007), 'PAML 4: phylogenetic analysis by maximum likelihood', *Molecular Biology and Evolution* **24**(8), 1586–1591.

Yang, Z., Wong, W. S. W. & Nielsen, R. (2005), 'Bayes empirical bayes inference of amino acid sites under positive selection.', *Molecular biology and evolution* **22**(4), 1107–18.

Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. (2010), 'Gene ontology analysis for RNA-seq: accounting for selection bias', *Genome Biology* **11**(2).

Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. (2017), 'ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data', *Methods in Ecology and Evolution* **8**(1), 28–36.

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C. & Levine, D. (2017), 'SeqArray—a storage-efficient high-performance data format for WGS variant calls', *Bioinformatics* **33**(15), 2251–2257.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C. & Weir, B. S. (2012), 'A high-performance computing toolset for relatedness and principal component analysis of SNP data', *Bioinformatics* **28**(24), 3326–3328.

Zimin, A., Yepes, M., Maldonado, C. E., Navarro, L., Kovaka, S., Pertea, M., Florez, C., Gongora, C. E., Moncada, P., Yorke, J., Gaitan, A. & Aldwinckle, H. (2019), Coffea arabica variety Caturra and C. eugenioides Genome Assembly and Annotation, *in* 'Plant and Animal Genome Conference XXVII'.