Discovery and genomic architecture of *Cercospora zeina* (Crous & U. Braun) effector genes

by

Eugene Ntuntu Kabamba Kabwe

Submitted in partial fulfilment of the requirements for the degree
Master of Science in Bioinformatics

in the

Department of Biochemistry, Genetics and Microbiology
Faculty of Natural and Agricultural Sciences

UNIVERSITY OF PRETORIA
Pretoria

May 2020

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

DECLARATION

I, Eugene Ntuntu Kabamba Kabwe, declare that the dissertation, which I hereby submit for the degree Master of Science in Bioinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

ETHICS STATEMENT

The author, whose name appears on the title page of this dissertation, has obtained, for the research described in this work (where necessary), the applicable research ethics approval.

The author declares that he has observed the ethical standards required in terms of the University of Pretoria's Code of Ethics for Researchers and the Policy guidelines for responsible research.

_____

Signature

_____

Student name

_____

Date

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my supervisor, Prof. Dave K. Berger, for allowing me to pursue an MSc degree in his research group under his supervision. I remember personally asking him for a chance to pursue a project under his supervision, and he accepted me with open arms. He has guided me throughout this process and has constantly challenged my way of thinking. I would also like to thank my co-supervisor, Dr. Tuan A. Duong, for his expertise and for teaching me various skills throughout the process. I convey my gratitude to Prof. Eva H. Stukenbrock and her research group for helping with the SMRT sequencing for this project.

Secondly, I would like to thank the members of the Centre for Bioinformatics and Computational Biology (CBCB). Prof. Fourie Joubert, thank you for hosting me at the centre and for always being willing to help and for creating a very friendly working environment. Prof. Oleg Reva, thank you for teaching me several bioinformatics skills and for organising insightful journal clubs. Mr. Johann Swart, thank you for helping me troubleshoot many of the problems I encountered and for maintaining the technology and equipment within the centre. Dr. Nicky Olivier, thank you for your guidance throughout my journey. I am also grateful to my fellow students within the centre who have made this whole journey enjoyable. I am grateful to Mrs. Glenda Brits for her help with drawing the first two figures of Chapter 1.

Thirdly, I would like to thank Prof. Bernard Slippers for maintaining a great working environment at the Forestry and Agricultural Biotechnology Institute (FABI). FABI is good place to not only grow in your research career, but also to grow as a person. The seminars, discussions and the social interactions have been very valuable. The support from so many colleagues has been valuable as well. I have made many friends at the institute and am thankful for their friendship.

Most importantly, I would like to thank my mother and father for the constant support I received throughout this journey. I am extremely privileged to have such supportive and loving parents. They have provided me with a loving home and so much more throughout my life. I would also like to thank my three brothers for the extra support and encouragement throughout the journey.

This project was supported by grant numbers 118503 and 98977 from the National Research Foundation (NRF) of South Africa.

PREFACE

Discovery and genomic architecture of *Cercospora zeina* (Crous & U. Braun) effector genes

by

Eugene Ntuntu Kabamba Kabwe

Supervisor: Prof. Dave K. Berger

Co-Supervisor: Dr. Tuan A. Duong

Department: Department of Biochemistry, Genetics and Microbiology

University: University of Pretoria

Degree: Master of Science (MSc) in Bioinformatics

Keywords: Dothideomycetes, effectors, *Cercospora zeina*, grey leaf spot, hemibiotroph, genome architecture, "two-speed" genome, effector evolution, repeat-induced point mutation, transposable elements

Diseases caused by plant pathogenic fungi are a threat to forestry and sustainable crop production worldwide. Amongst these fungi are the Dothideomycetes that consists of several economically important pathogens. Upon infection of the host, pathogenic Dothideomycetes often secrete effectors to ensure successful colonisation. Effectors are small secreted proteins that promote infection in the host by supressing host defences or preventing recognition by the host. Understanding the process of pathogenicity and how effector genes evolve will help in creating measures that help control the spread of plant diseases.

The sequencing of the complete genomes of plant pathogenic fungi has allowed researchers to learn about the effector armoury that these pathogens possess as well as how these effector genes evolve. A feature that has been found to be relatively common in many plant pathogenic fungi is the presence of a so-called "two-speed" genome. The "two-speed" genome consists of a

compartmentalized genome divided into a repeat-rich gene-poor region and a repeat-poor gene-rich region.

In the first section of **Chapter 1**, I discuss examples of how a selection of effectors of Dothideomycete biotrophs, hemibiotrophs and necrotrophs perform their function. In the second section, I discuss the evolutionary dynamics of effector genes within the genome. I highlight how the unpredictable nature of repeat-rich regions are a perfect hotspot for effector genes to evolve. In the conclusion section, I highlight areas that still need to be studied to further our understanding of the pathogenicity of Dothideomycetes.

In **Chapter 2**, I present my research results. The research focuses on the maize pathogen *Cercospora zeina* (Crous & U. Braun) that causes the grey leaf spot disease in maize fields in Africa. *C. zeina* is a Dothideomycete that has a hemibiotrophic lifestyle. First, we present the newly sequenced genome of *C. zeina*. It is a nearly gapless assembly that has been sequenced with single-molecule real-time technology. I then aim to identify the effector protein catalogue of *C. zeina* and to discover whether the effector genes reside in the repeat-rich compartments of the genome using bioinformatics analyses.

LIST OF ABBREVIATIONS

BUSCO	Benchmarking universal single-copy orthologues
CNV	Copy number variation
CSEPs	Candidate secreted effector proteins
GFP	Green fluorescence protein
LINEs	Long interspersed nuclear elements
LTR	Long terminal repeat
RIP	Repeat-induced point mutation
RXLR	Arginine, any amino acid, Leucine, Arginine
SINEs	Short interspersed nuclear elements
SMRT	Single-molecule real-time
SSPs	Small secreted proteins
TE	Transposable element

SUMMARY

In Africa, the grey leaf spot disease in maize is caused by the hemibiotrophic Dothideomycete *Cercospora zeina*. Dothideomycetes consist of many economically important phytopathogenic fungi. During infection of the host, phytopathogenic fungi secrete effectors that promote infection. In this dissertation, I review how various effectors perform their function. I then review how the "two-speed" genome of many Dothideomycete phytopathogens is important for the gain of virulence and loss of avirulence effector genes in the constant molecular arms race with the plant host. Furthermore, to improve our understanding of the pathogenicity of *C. zeina*, we aim to identify the effector gene catalogue and analyse its genome architecture. We present a contiguous genome assembly of *C. zeina* generated by PacBio SMRT sequencing technology. The assembly consists of 17 nuclear genome contigs that make up the 41 Mbp genome and contain three possible full chromosomes. The annotation of the genome has revealed a secretome that contains many proteins predicted to have oxidoreductive and peroxidase activities. Effector prediction revealed a total of 274 effectors which included potential homologues of the ECP2, ECP6 and AVR4 effectors from *Cladosporium fulvum*. These results imply that *C. zeina* potentially secretes proteins that prevent recognition by the host and protect against oxidative stress. Additionally, genome architecture analysis of *C. zeina* has revealed a bipartite structure consisting of 33.2% AT-rich compartments and 66.8% GC-rich compartments. However, effector genes are not concentrated in the AT-rich compartments. This study has paved way for the functional characterization of candidate *C. zeina* effectors which could ultimately lead to effector-based breeding of maize.

RESEARCH OUTPUTS

Journal articles

In preparation

Presentations

Oral presentations

FABI Seminars – Pretoria, South Africa – 13 September 2018

Title: "Discovery of pathogenicity effectors from the *Cercospora zeina* genome"

FABI Seminars – Pretoria, South Africa – 7 March 2019

Title: "Effector mining in the *Cercospora zeina* genome"

CBCB Seminars – Pretoria, South Africa – 12 June 2019

Title: "Discovery of pathogenicity effectors from the *Cercospora zeina* genome"

FABI Seminars – Pretoria, South Africa – 1 August 2019

Title: "Analysing the genome architecture of the maize pathogen *Cercospora zeina*"

BGM Seminars – Pretoria, South Africa – 9 September 2019

Title: "Analysing the effector genes and genome architecture of the maize pathogen *Cercospora zeina*"

CBCB Seminars – Pretoria, South Africa – 2 October 2019

Title: "Analysing the effector genes and genome architecture of the maize pathogen *Cercospora zeina*"

Poster presentations

University of Pretoria BGM Research Symposium – Pretoria, South Africa – 6 December 2018

Title: "Discovery of pathogenicity effectors from the *Cercospora zeina* genome"

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 THE EVOLUTIONARY DYNAMICS OF DOTHIDEOMYCETE EFFECTORS

1.1 INTRODUCTION

Food security continues to be one of the pressing global concerns as we reach the halfway point of the century. With the global population estimated to reach approximately 9.7 billion people by the year 2050 (Gerland et al. 2014), there is a great need for nations around the globe to ensure that their populations have enough food. Crops are one of the main sources of foods around the world as they not only serve as staple foods in developing countries but are also required to feed livestock that in turn provide meat and dairy consumption. As a result, it is of utmost importance to maintain the stable supply of crops as a source of food. However, among other factors, disease is one of the main causes of yield losses in crops (Savary et al. 2019). Many organisms that cause diseases are plant pathogenic fungi, which includes those that belong to the Dothideomycetes class.

Dothideomycetes is a class of the Ascomycota division and is a very ecologically diverse group of fungi. Dothideomycetes can be found in almost all areas of earth including Antarctica and in extreme environments (Choe et al. 2018). Some are lichenized (Muggia et al. 2008) whilst others can survive in fresh-water and salt-water aquatic environments (Gerea et al. 2012). Additionally, a few of them have an endophytic lifestyle (Wijesekera et al. 2017). Dothideomycetes is one of the largest classes of fungi with 12 orders containing more than 19,000 species (Kirk et al. 2008). The plant pathogenic fungi that belong to the Pleosporales and Capnodiales are the ones that are of economic importance (Ohm et al. 2012). Their host organisms range from crops such as maize, wheat and rice, to dicots such as soybean, canola and tomato and trees such as pine and poplar.

Dothideomycetes, like many other plant pathogenic fungi, have four main infection strategies (Fig. 1.1), namely biotrophic, necrotrophic, hemibiotrophic and saprotrophic strategies (Lo Presti et al. 2015). Biotrophs, as the name suggests, are pathogens that grow and obtain nutrition from living plant tissue. These pathogens often have to evade or suppress the host's defence responses in order to maintain their life cycle. Necrotrophs on the other hand, are pathogens that rapidly kill the living cells of the host and obtain nutrition from the dead matter. Unlike biotrophs, necrotrophs are usually opportunistic and have a wider host range. Hemibiotrophs initially have a biotrophic lifestyle and then eventually switch to a necrotrophic lifestyle. Lastly, saprotrophs are organisms that obtain nutrition from decaying matter.

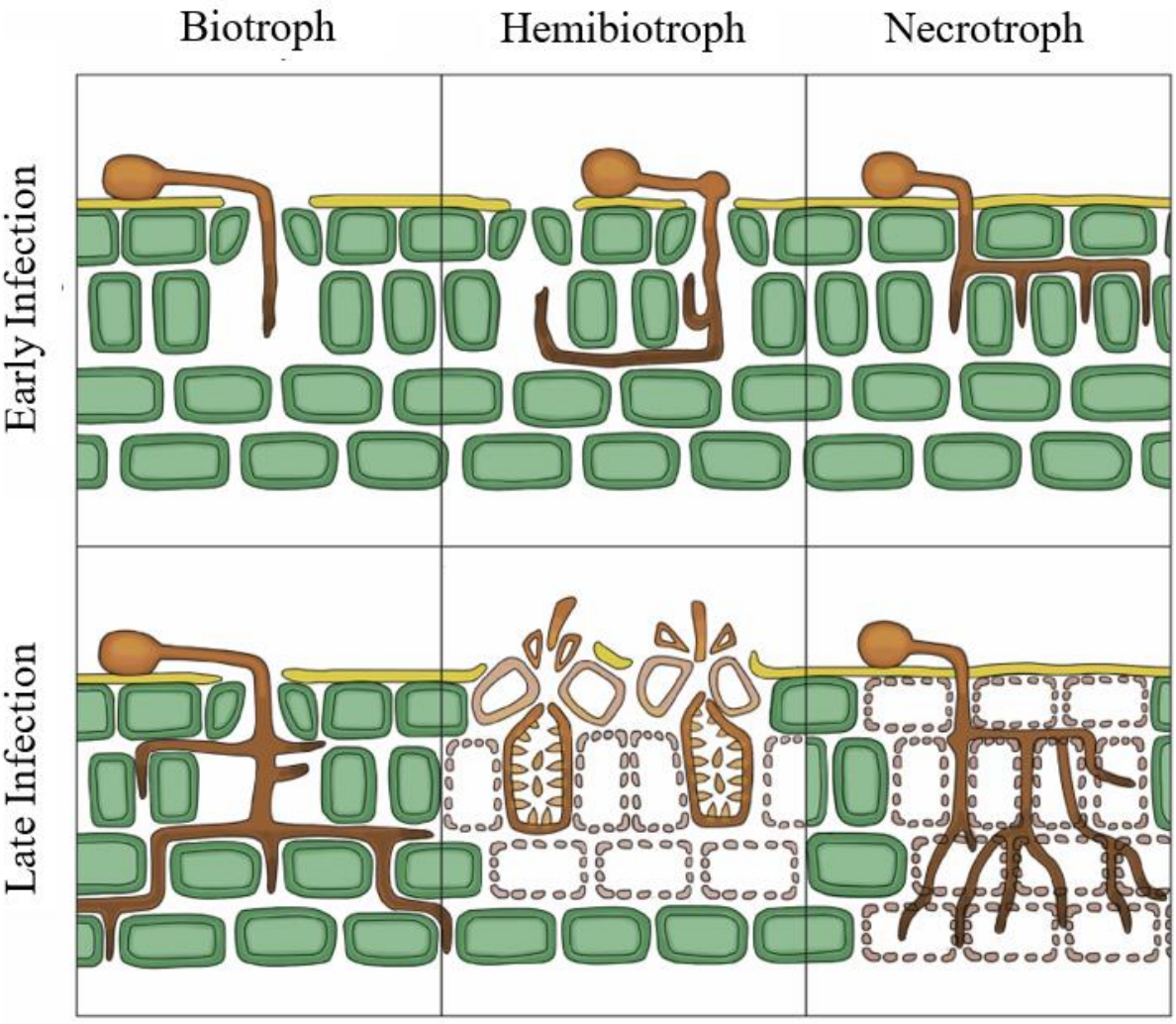


Fig. 1.1 The infection strategies of phytopathogens. The first column illustrates the biotrophic lifestyle of *Cladosporium fulvum* (de Wit et al. 2012). The second column illustrates the hemibiotrophic lifestyle of *Zymoseptoria tritici* (Steinberg 2015). The third column illustrates the

concept necrotrophic lifestyle of *Parastagonospora nodorum* (Oliver et al. 2012). The brown structures are the fungal structures. The green rectangular shapes are the plant cells. The light brown rectangular shapes are the dead plant cells. The yellow top layer is the cuticle. Diagram adapted from (Lo Presti et al. 2015).

Upon infection of the host, pathogens often secrete macromolecules that aid in the infection process. These macromolecules can be toxic secondary metabolites in the case of necrotrophs (Stergiopoulos et al. 2013; Wang et al. 2014), small RNAs (Wang et al. 2015, 2017) and proteins known as effectors (Lo Presti et al. 2015). Effectors are small secreted proteins (SSPs) that promote infection in the host by supressing host defences or preventing recognition by the host. Their functions include manipulating the host cell biology, suppressing host responses and shielding the pathogen to support colonization (Hogenhout et al. 2009). Due to the rapid evolution of the effectors (discussed later), most of these proteins lack amino acid sequence similarity and common domains of known function when compared to other proteins. The obvious expected characteristic of effectors is the presence of a signal peptide because they are secreted. They are often small in size (≤ 300 amino acids), have high cysteine content and lack transmembrane domains (Lo Presti et al. 2015). These are the characteristics that mostly distinguish effectors from other secreted proteins. The definition for pathogenicity effectors can sometimes broadly include all the macromolecules involved in the infection process including small RNAs and secondary metabolites, but for this review, the definition will be restricted to effectors as SSPs. These SSPs usually act as virulence factors that contribute to successful infection of the host. In the presence of a resistance (*R*) protein in the host, the virulence factor becomes an avirulence factor which prevents successful infection.

As Dothideomycetes continue to infect crops around the globe (Zhou et al. 2019; Perelló et al. 2019; Robles-Yerena et al. 2019), it is important to understand the mechanisms in which these pathogens infect their host, so that breeders can counter-act the processes accordingly and effectively. Fungicides have been used extensively but there have been cases of pathogens gaining resistance to the fungicides as well as concerns regarding the environmental side effects (Aktar et al. 2009). Breeding strategies involving the selection of crops harbouring resistance genes have also been suggested (Vleeshouwers and Oliver 2014), although pathogens can evolve to regain virulence via mutations (Daverdin et al. 2012). It is therefore important to understand the molecular mechanisms that are at play in various plant-fungal pathosystems to devise the best strategies to

limit disease and large outbreaks. In this review, I firstly discuss the modes of action of a selection of effector proteins to illustrate the variety of functions (Table 1.1). I then discuss the birth and death of effector genes in Dothideomycete genomes and how their genome localization affects their evolutionary dynamics.

1.2 DOTHIDEOMYCETE EFFECTORS IN ACTION

1.2.1 Biotrophic effectors

Biotrophic plant pathogenic fungi obtain nutrients from living tissue, and as a result have effector catalogues that prevent pattern-triggered immunity caused by recognition of the pathogen. Examples of biotrophic Dothideomycete effectors in this section will mostly include that of the well-studied biotroph *Cladosporium fulvum (syn. Passalora fulva* (Braun et al. 2003)) which causes the leaf mold disease in tomato (*Solanum lycopersicum*). Effectors from *C. fulvum* mostly localize in the apoplastic space as the pathogen does not penetrate into the host cell (Thomma et al. 2005; Mesarich et al. 2018). The first example is the avirulence protein 4 (AVR4) effector (van den Burg et al. 2006). The AVR4 effector prevents hydrolysis of the chitin residues of the fungal cell wall by plant chitinases (Table 1.1, Fig. 1.2). An affinity precipitation assay showed that the AVR4 could only bind to insoluble polysaccharides like chitin and failed to bind to soluble polysaccharides like xylan and cellulose as well as tomato cell walls (van den Burg et al. 2006). Additionally, AVR4 prevented the deleterious effects of plant chitinases on the growth of *Trichoderma viride in vitro* in an AVR4-dependent manner (van den Burg et al. 2006). Further studies showed that the ability of AVR4-silenced transformants of *C. fulvum* to colonize tomato leaves was significantly weakened compared to the AVR4-expressing parent strains (van Esse et al. 2007).

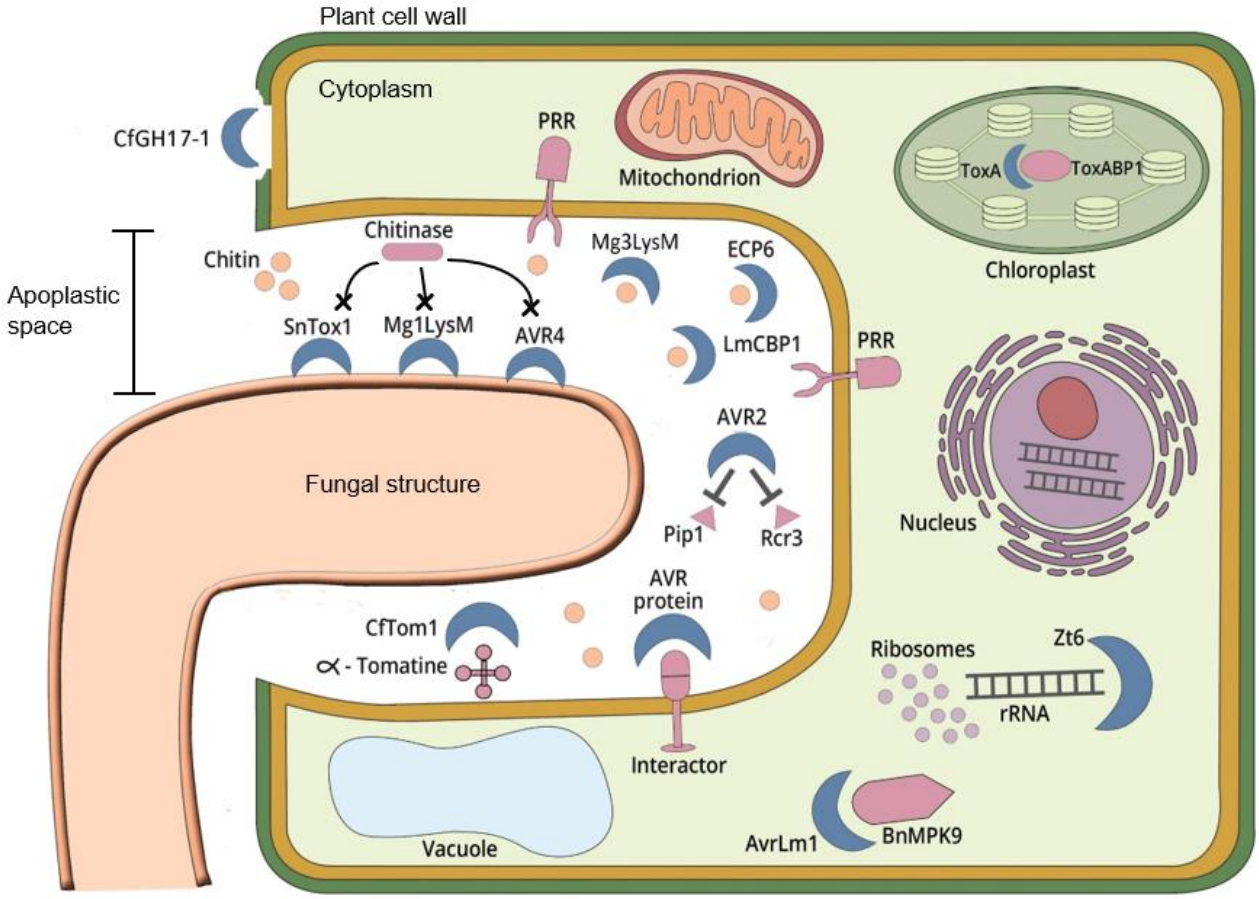


Fig. 1.2 A diagram illustrating the functions of Dothideomycete effectors. The fungal structure on the left is invading the plant cell on the right. The boundary on the fungal structure represents chitin. The green outer layer of the plant cell is the cell wall. The orange inner layer of the plant cell is the cell membrane. Effectors are represented by the blue crescent-moon shapes. The plant molecules that interact with the effectors are all purple. Diagram adapted from (Lo Presti et al. 2015).

In addition to AVR4, another effector that suppresses host immune responses is the LysM domain-containing extracellular protein 6 (ECP6) from *C. fulvum* (Fig. 1.2). The effector sequesters chitin residues that are released from the hyphal cell walls to prevent recognition by the host plant (de Jonge et al. 2010). Similarly to the studies with AVR4, ECP6 was able to associate with chitin but failed to associate with soluble polysaccharides (de Jonge et al. 2010). However, ECP6 was unable to prevent chitinases from inhibiting the growth of *T. viride in vitro* as AVR4 did which showed that ECP6 does not prevent hydrolysis. Treatment of tomato cells in suspension with nanomolar chitin oligosaccharides caused medium alkalinisation. Under the presence of ECP6, the pH changes were inhibited whilst there was no similar reaction with AVR4 which showed that ECP6 prevents chitin recognition by the plant cells.

The third effector that has been characterized in *C. fulvum* is the ECP2 effector (Laugé et al. 1997). The ECP2 effector's function is not entirely known but it is required for full virulence. A study showed that an ECP2-deficient *C. fulvum* strain had an inhibited development in tomato (Laugé et al. 1997). It showed poor colonization of tomato leaves and produced low amounts of other effectors including AVR4. The quantification of sporulation further revealed that the ECP2-deficient strain produced extremely low amounts of conidia compared to the wild-type strain (Laugé et al. 1997). The accumulation of pathogenesis-related (PR) proteins from the susceptible tomato plants inoculated with the *C. fulvum* strains was also quantified. The accumulation was found to be stronger on leaves inoculated with the ECP2-deficient strain compared to the leaves inoculated with the wild-type strain. This showed that the lack of ECP2 production resulted in accumulated PR proteins for the host (Laugé et al. 1997). Ultimately, these results revealed that without ECP2, the virulence of *C. fulvum* on susceptible tomato plants is hampered.

The fourth effector that has been characterized in *C. fulvum* is the AVR2 effector (van Esse et al. 2008). The AVR2 effector inhibits tomato cysteine proteases required for plant defence and acts as a virulence factor (Fig. 1.2). Several AVR2-silenced *C. fulvum* mutants were inoculated on tomato plants and had shown compromised colonization. The mutant strain grew slower and sporulated later when compared to the colonization of the wild-type strain (van Esse et al. 2008). Additionally, the heterologous expression of AVR2 in tomato increased its susceptibility. This was shown by significant necrosis when inoculated with *Botrytis cinerea* which had significantly longer hyphae and also enhanced disease development after inoculation with *Verticillium dahliae* (van Esse et al. 2008). To identify the AVR2 targets, its ability to block the binding of the label DCG-04 to tomato cysteine proteases was analysed via a labelling and purification experiment and resulted in AVR2 having a high affinity for the proteases Rcr3, Pip1 and glycinain (van Esse et al. 2008). Additionally, AVR2 was found to bind to Rcr3 and Pip1 using biotin-labelled AVR2 and streptavidin beads assay.

The fifth characterized *C. fulvum* effector is CfTom1 (Ökmen et al. 2013). The CfTom1 effector is required for the detoxification of the antifungal glycoalkaloid α-tomatine that is secreted by tomato as a defence response (Fig. 1.2). The CfTom1 effector is a carbohydrate active enzyme (CAZyme) that has glycosyl hydrolase 10 (GH10) activity and this was shown by the accumulation of tomatidine (product of α-tomatine degradation) in the apoplastic fluid of tomato plants

constitutively expressing CfTom1 (Ökmen et al. 2013). The *C. fulvum* CfTom1-deletion mutants had 70% reduction in the germination of conidia in the presence of 100 µM α-tomatine *in vitro* compared to the wild-type (Ökmen et al. 2013). In addition, no tomatidine accumulation was found in the deletion mutants in culture filtrates in the presence of α-tomatine whilst there was accumulation in the wild-type culture filtrates. To confirm that CfTom1 was needed for virulence in susceptible tomato plants, the fungal biomass of the deletion mutants in inoculated plants was measured in comparison to the wild-type and there was a reduction by *c.* 63% from 10 days post infection and onwards (Ökmen et al. 2013).

Although *C. fulvum* is traditionally considered a biotroph, recent studies have shown that *C. fulvum* could be a hemibiotroph instead. For example, the CAZyme gene content of *C. fulvum* has been found to be similar to the hemibiotrophs *Zymoseptoria tritici and Leptosphaeria maculans* rather than that of the ascomycete obligate biotroph *Blumeria graminis* and basidiomycete obligate biotrophs *Puccinia* spp. and *Ustilago* spp. (Hane et al. 2020). This in line with the necrosis that is usually seen during the later infection stages of tomato leaf mold disease (Thomma et al. 2005). Indeed, the recently studied CfGH17-1 effector has been shown to be up-regulated during the later necrotrophic stages of infection (Ökmen et al. 2019). The heterologous expression of CfGH17-1 using the PVX expression system via the Agrobacterium-mediated transient transformation assay in Cf-0, Cf-1 and Cf-5 tomato cultivars caused cell death in the leaves 3 weeks post-infiltration (Ökmen et al. 2019). Thin layer chromatography analysis showed that CfGH17-1 was able to hydrolyse laminarin which is a linear 1,3-β-glucan in plant cell walls (Ökmen et al. 2019) (Fig 1.2). This showed that the effector is important for necrosis. Ultimately, most characterized biotrophic Dothideomycete effectors function to prevent recognition.

Table 1.1 List of some of the characterized effectors in Dothideomycetes

Effector	Virulence function or avirulence protein interactor	Species	Order	Lifestyle	References
AVR4	Prevents hydrolysis of chitin residues	*Cladosporium fulvum*	Capnodiales	Biotroph	(van den Burg et al. 2006)
AVR2	Inhibits host cysteine proteases	*Cladosporium fulvum*	Capnodiales	Biotroph	(van Esse et al. 2008)
AVR4E	Interacts with *Cf4-E*	*Cladosporium fulvum*	Capnodiales	Biotroph	(Westerink et al. 2004)
AVR9	Interacts with *Cf9*	*Cladosporium fulvum*	Capnodiales	Biotroph	(Van den Ackerveken et al. 1992)
CfTom1	Detoxifies α-tomatine	*Cladosporium fulvum*	Capnodiales	Biotroph	(Ökmen et al. 2013)
ECP6	Sequesters chitin residues	*Cladosporium fulvum*	Capnodiales	Biotroph	(de Jonge et al. 2010)

Effector	Virulence function or avirulence protein interactor	Species	Order	Lifestyle	References
ECP1	Required for full virulence	*Cladosporium fulvum*	Capnodiales	Biotroph	(Laugé et al. 1997)
ECP2	Required for full virulence	*Cladosporium fulvum*	Capnodiales	Biotroph	(Laugé et al. 1997)
CfGH17-1	Hydrolyses laminarin	*Cladosporium fulvum*	Capnodiales	Biotroph	(Ökmen et al. 2019)
Mg1LysM	Prevents hydrolysis of chitin residues	*Zymoseptoria tritici*	Capnodiales	Hemibiotroph	(Marshall et al. 2011)
Mg3LysM	Sequesters chitin residues	*Zymoseptoria tritici*	Capnodiales	Hemibiotroph	(Marshall et al. 2011)
Zt6	Cleaves rRNA and has antimicrobial activity	*Zymoseptoria tritici*	Capnodiales	Hemibiotroph	(Kettles et al. 2018)
AvrStb6	Interacts with *Stb6*	*Zymoseptoria tritici*	Capnodiales	Hemibiotroph	(Saintenac et al. 2018)

Effector	Virulence function or avirulence protein interactor	Species	Order	Lifestyle	References
Avr3D1	May be recognized by *Stb7* and *Stb12*	*Zymoseptoria tritici*	Capnodiales	Hemibiotroph	(Meile et al. 2018)
DsECP2-1	May be recognized by an immune receptor	*Dothistroma septosporum*	Capnodiales	Hemibiotroph	(Guo et al. 2020)
PfAVR4	Prevents hydrolysis of chitin residues	*Pseudocercospora fijiensis*	Capnodiales	Hemibiotroph	(Arango Isaza et al. 2016)
CfAVR4	Prevents hydrolysis by chitinases and possibly affects cercosporin production	*Cercospora* cf. *flagellaris*	Capnodiales	Hemibiotroph	(Santos Rezende et al. 2020)
AvrLm1	Causes cell death by interacting with BnMPK9	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Ma et al. 2018)
AvrLm2	Interacts with *Rlm2*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Ghanbarnia et al. 2015)

Effector	Virulence function or avirulence protein interactor	Species	Order	Lifestyle	References
AvrLm3	Interacts with *Rlm3*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Balesdent et al. 2002)
AvrLm4-7	Interacts with *Rlm4* and *Rlm7* and masks recognition of AvrLm5-9	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Parlange et al. 2009)
AvrLm5	Interacts with *Rlm5*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Van de Wouw et al. 2014)
AvrLm6	Interacts with *Rlm6*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Fudal et al. 2007)
AvrLm10	Interacts with *Rlm10*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Petit-Houdenot et al. 2019)
AvrLm11	Interacts with *Rlm11*	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Balesdent et al. 2013)
LmCBP1	Binds to chitin	*Leptosphaeria maculans*	Pleosporales	Hemibiotroph	(Liu et al. 2020)

Effector	Virulence function or avirulence protein interactor	Species	Order	Lifestyle	References
ToxA	Binds to ToxABP1 and prevents thylakoid formation	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Manning and Ciuffetti 2005)
ToxB	Required for full virulence	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Singh et al. 2010)
SnTox1	Prevents hydrolysis by chitinases and induces cell death	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Liu et al. 2016)
SnTox2	Interacts with *Snn2*	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Friesen et al. 2007)
SnTox3	Interacts with *Snn3*	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Liu et al. 2009)
SnTox6	Interacts with *Snn6*	*Parastagonospora nodorum*	Pleosporales	Necrotroph	(Gao et al. 2015)

1.2.2 Hemibiotrophic effectors

Hemibiotrophic plant pathogenic fungi first secrete biotrophic effectors to avoid recognition by the host and then secrete necrotrophic effectors at a later stage. Similar to *C. fulvum*, several of the Dothideomycete hemibiotrophs also secrete apoplastic effectors that prevent hydrolysis by plant chitinases and prevent recognition. For example, in a transcriptomic study of the infection cycle of the hemibiotroph *Dothistroma septosporum* which causes Dothistroma needle blight in pine (*Pinus radiata*), it was revealed that DsECP2-1 and DsECP6 (homologues of *C. fulvum* ECP2 and ECP6 respectively) were highly upregulated 31- and 104-fold respectively in the mid stages of infection in comparison to the early stages (Bradshaw et al. 2016). A later study showed that DsECP2-1 was able to trigger cell death in the leaves of the non-host plant *N. tabacum* (Guo et al. 2020). However, in contrast to the *C. fulvum* ECP2, the DsECP2-1 deletion *D. septosporum* mutants had increased colonization of the host *P. radiata* (Guo et al. 2020). This indicated that the *P. radiata* cultivar used in this study possibly had an immune receptor that recognized DsECP2-1.

Another *C. fulvum* effector that has been shown to be conserved in Dothideomycete hemibiotrophs is the AVR4 effector. Homology analyses revealed that this effector was present in *Pseudocercospora fijiensis* and other *Cercospora* species (Stergiopoulos et al. 2010; Mesarich et al. 2016). Using the same methods to determine the function of the *C. fulvum* AVR4 protein, the *P. fijiensis* AVR4 homologue was also shown to protect *T. viride* from chitinases (Stergiopoulos et al. 2010). The *P. fijiensis* induces a hypersensitive response in both tomato harbouring the Cf-4 gene (Stergiopoulos et al. 2010) and the resistant banana variety Calcutta 4 (Arango Isaza et al. 2016). In *Cercospora* cf. *flagellaris*, the AVR4 homologue was shown to not only inhibit chitinase-induced growth suppression *in vitro* but was also shown to be important for production of the secondary metabolite cercosporin which plays a role in plant cell death (Santos Rezende et al. 2020). On solid media, the *C.* cf. *flagellaris* AVR4 disruption mutant only produced around 10% of cercosporin when compared to the wild-type and also had significantly less expression of CTB1 and CTB8 genes (involved in cercosporin production) *in vitro* when compared to the wild-type (Santos Rezende et al. 2020). This is one of the few studies that links effector and secondary metabolite production.

Other apoplastic effectors that bind to carbohydrate molecules are Mg1LysM and Mg3LysM from the hemibiotroph *Z. tritici*. Both effectors were shown to protect the fungal hyphae from plant hydrolytic enzymes like the *C. fulvum* AVR4 effector (Marshall et al. 2011). Although only Mg1LysM was able to prevent chitin-induced plant defense responses, the Mg1LysM deletion mutants did not have a hampered virulence phenotype on susceptible leaves. In contract, Mg3LysM deletion mutants had reduced virulence (Marshall et al. 2011). In the hemibiotroph *L. maculans*, the LmCBP1 effector is a secreted chitin binding protein that lacks chitinase and cellulase activity. A chitin binding assay showed that LmCBP1 was able to bind to four polysaccharides excluding chitosan (Liu et al. 2020). Chitinase and cellulase activity were determined by the production of N-acetyl-D-glucosamine (GlcNAc) in one hour from colloidal chitin and the production of glucose in one hour from cellulose respectively. In both cases, there was no evidence of both activities when compared to the positive controls (Liu et al. 2020). Cotyledons of *B. napus* cultivar 'Westar' were inoculated with LmCBP1 *L. maculans* deletion mutants and the wild-type *L. maculans* JN3 strain for a pathogenicity assay. The lesions caused by the deletion mutants were shown to be significantly smaller than that of the wild-type. After 7 and 11 days post inoculation, trypan blue staining showed that there was a smaller area of cell death in the cotyledons inoculated with the deletion mutant compared to that of the wild-type (Liu et al. 2020). Lastly, the inoculation of the deletion mutant and the wild-type on different *B. napus* cultivars containing different *R* genes showed that those inoculated with the deletion mutant had smaller lesions (Liu et al. 2020).

When hemibiotrophs switch from the biotrophic phase to the necrotrophic phase, they secrete secondary metabolites that act as toxins (Stergiopoulos et al. 2013; Muria-Gonzalez et al. 2015). Additionally, effectors that cause cell death are also secreted. These effectors are usually localized in the cytoplasm of the host cell rather than the apoplastic localization of the biotrophic effectors. Very few of such effectors have been characterized in Dothideomycetes, but recent studies have identified a few of these effectors.

The Zt6 effector in *Z. tritici* has a dual functionality in that in causes cell death in the host plant during the necrotrophic stage and it also has a function in antimicrobial competition (Kettles et al. 2018). This effector is a small ribonuclease that cleaves rRNA of both plants and animals but not rRNA from *Z. tritici* itself (Fig. 1.2). Transcriptome analysis of *Z. tritici* infecting wheat revealed that one of the two peak expression levels of Zt6 was 14 days post infection which coincides with the necrotrophic phase. The co-expression of a GFP transgene with Zt6 in wheat epidermal cells

resulted in the complete abolishment of the GFP fluorescence which indicated that it has toxic effects in the wheat cell cytoplasm (Kettles et al. 2018). A follow-up experiment in the wheat biolistic system showed that the mature Zt6 protein was able to inhibit GFP fluorescence when directed to both the apoplast or cytoplasm. However, when the Zt6 mutant lacking the N-terminal loop (Δ19–40 mutant) was expressed in the cytoplasm, GFP fluorescence was still abolished but the mutant protein was unable to re-enter the cytoplasm when directed to the apoplast. This indicated that the N-terminal loop was needed to re-enter the wheat cell. A cell-free protein expression system also showed that Zt6 and the Δ19–40 mutant successfully cleaved rRNAs from wheat (Kettles et al. 2018).

In *L. maculans*, the AvrLm1 effector is one of the effectors that are vital for the necrotrophic lifestyle switch (Ma et al. 2018). The AvrLm1 effector is a virulence factor that interacts with the *Brassica napus* mitogen-activated protein (MAP) kinase 9 (BnMPK9) which results in cell death (Ma et al. 2018) (Fig. 1.2). To show that AvrLm1 is a virulence factor, *Arabidopsis thaliana* plants expressing the AvrLm1 protein was inoculated with the bacterial pathogen *P. syringae* pv. *tomato* DC3000 and severe chlorosis compared to the inoculation on the wild-type plant was observed (Ma et al. 2018). The AvrLm1 was shown to interact with BnMPK9 in co-expression analysis *in vivo* resulting in early cell death. The agroinfiltration of *Nicotiana benthamiana* with BnMPK9 and AvrLm1 resulted in early cell death compared to the control which only included the GFP gene (Ma et al. 2018). Additionally, AvrLm1 was shown to increase phosphorylation in BnMPK9. Hemagglutinin (HA)-tagged BnMPK9 was co-expressed with AvrLm1 in *N. benthamiana* leaves in parallel to a BnMPK9 and GFP co-expression negative control. After immunoprecipitation using HA-magnetic beads, the level of phosphorylation was determined via immunoblot using anti-phosphothreonine monoclonal antibodies (Ma et al. 2018). The presence of the transgene increased phosphorylation in BnMPK9 compared to the basal level of phosphorylation in the negative control. Lastly, the over-expression of BnMPK9 in *B. napus* inoculated with *L. maculans* was shown to enhance *L. maculans* growth (Ma et al. 2018).

1.2.3 Necrotrophic effectors

Necrotrophic plant pathogenic fungi kill the living tissue of the cells and obtain nutrients from the dead matter. As a result, these host-specific effectors secreted from these fungi are rarely optimized to prevent recognition like for biotrophic fungi and rather allow the rapid killing of the plant cells. A well-studied example of such an effector is the necrosis toxin A (ToxA) from the

Dothideomycete necrotrophs *Parastaganospora nodorum* and *Pyrenophora tritici-repentis* that both infect wheat (*Triticum aestivum*) causing the leaf blotch and the tan spot diseases respectively (Manning and Ciuffetti 2005; Oliver et al. 2012). ToxA interacts with the ToxA binding protein 1 (ToxABP1) in the chloroplast – a protein predicted to be involved in thylakoid formation – and inhibits photosynthetic reactions leading to cell death (Table 1.1, Fig. 1.2). The visualization of GFP-ToxA fusion protein showed that it localized in the cytoplasm and chloroplast in sensitive wheat mesophyll cells (Manning and Ciuffetti 2005). Another study showed that ToxA interacted with ToxABP1 in a pull-down assay (Manning et al. 2007). Furthermore, the transformation of wheat cells with ToxA led to their deaths after the gene was expressed (Manning and Ciuffetti 2005). Sensitive and insensitive wheat cultivars were co-transformed with ToxA lacking the signal sequence and β-glucuronidase (GUS) expression plasmids using the biolistic bombardment technique. Both sensitive and insensitive wheat cells co-transformed with ToxA and GUS had an approximately 50% decrease in GUS expression compared to control cells co-transformed with GUS and a control plasmid (Manning and Ciuffetti 2005).

Another necrotrophic effector characterized in *P. nodorum* is the SnTox1 effector (Liu et al. 2016). The SnTox1 effector has a dual functionality in that in protects the pathogen from chitinases and it induces plant cell death (Fig. 1.2). A polysaccharide-binding assay showed that SnTox1 was able to bind to shrimp shell chitin and chitin beads like AVR4 (Liu et al. 2016). Consequently, the mixture of pathogens *Pyrenophora teres* f. *teres*, *Cercospora beticola* and non-pathogen *Neurospora crassa* with purified SnTox1 was able to defend the pathogens from wheat chitinases in an assay. To test whether SnTox1 alone is able to cause necrosis in sensitive wheat lines, the purified SnTox1 was directly sprayed on the leaf surface and necrosis was observed. Afterwards, transforming the necrotroph *P. teres* f. *teres* with SnTox1 allowed the pathogen to colonize the non-host Snn1 wheat but not the hemibiotroph *C. beticola* and the saprotroph *N. crassa* (Liu et al. 2016).

Whilst many characterized Dothideomycete effectors have been shown to localize in the apoplast, or are known to interact with certain plant receptor proteins (Gao et al. 2015; Meile et al. 2018; Petit-Houdenot et al. 2019), few of them are known to function within the plant cell cytoplasm. In other phytopathogenic fungi, there are currently more reported examples of host cytoplasm-localized effectors. For example, in the Leotiomycete *Sclerotinia sclerotiorum*, there has been a discovery of an effector that inhibits host immunity through the inhibition of salicylic acid formation in the plant cell by interacting with a calcium-sensing receptor in the chloroplast (Tang et

al. 2020). In the Sordariomycete *Colletotrichum higginsianum*, effectors have been shown to localize in the peroxisomes, microtubules and Golgi of the host plant cell (Robin et al. 2018). Lastly, an effector in the basidiomycete *Melampsora larici-populina* has been shown to bind to the promoter of plant genes involved in immunity which thereby suppresses host immunity (Ahmed et al. 2018). Identifying more Dothideomycete effectors that localize in the plant cell cytoplasm will better our understanding of the pathogenicity of these phytopathogens.

1.3 GENOME ARCHITECTURE OF DOTHIDEOMYCETE FUNGI

Second-generation sequencing technologies such as Illumina has changed the scope of genomics studies as it has eased the process of sequencing and lessened the expenses needed for genome sequencing projects. The technology generates millions of short reads with a small error rate (1 error per 1 kbp) (Metzker 2010). However, it is often difficult to combine these short reads into genomic assemblies because of repetitive sequences such as transposable elements (TEs). The emergence of third-generation sequencing technologies such as single-molecule real-time and nanopore sequencing have brought about the ability to sequence long reads (> 1 kbp) which can span repetitive sequences and allow for more complete assemblies (Gibriel et al. 2016). The third-generation sequencing technologies have recovered repetitive sequences in genome sequences that were previously thought to contain a smaller repetitive component (Dutreux et al. 2018) and have thus opened the possibilities for more accurate genomics studies.

The genome architecture of many plant pathogenic fungi is often characterized as having a bipartite structure which consists of rapidly evolving non-conserved and slowly evolving conserved regions. The conserved regions are associated with gene-rich and repeat-poor sequences, whilst the rapidly evolving regions consist of gene-sparse and repeat-rich sequences such as TEs. This dual genome feature is common in many eukaryotes (Costantini and Musto 2017; Bohlin and Pettersson 2019), but what makes the feature stand out in plant pathogenic fungi is the discovery that pathogenicity-related and effector genes are often located in repeat-rich regions. This has led to the concept of “two-speed” genomes where the rapidly evolving regions provide a way for effector genes to diversify and acquire advantageous mutations that will be beneficial when colonizing the host plant (Dong et al. 2015). The “two-speed” concept refers to the different rates at which mutation and

selection occurs whereby the processes in conserved regions are slow, but rapid in repeat-rich regions.

The rapidly evolving repeat-rich regions can occur in different states, such as AT-rich isochores, sub-telomeric regions as well as accessory chromosomes (chromosomes present in select individuals but not the whole population). Isochores are large regions (> 300 kbp) of the genome that contain similarities in base content (Eyre-Walker and Hurst 2001). Diversification in AT-rich isochores can be a result of repeat induced point mutation (RIP) leakage effects. RIP is a fungal defence mechanism that inactivates TEs via cytosine to thymine mutations during sexual reproduction (Cambareri et al. 1991). The effector genes near RIP affected TEs could be subject to RIP leakage effects that could cause favourable mutations (Rouxel et al. 2011). The TEs could also lead diversification via copy number variations through increased ectopic recombination. The TEs are also present in accessory chromosomes, which harbour few genes compared to the core chromosomes. These genes in accessory chromosomes often have more non-synonymous mutations than synonymous mutations indicating a hotspot for rapid evolution (Stukenbrock et al. 2010). In *Z. tritici*, although the accessory chromosomes do not harbour effector genes, a study showed that isolates that had certain accessory chromosomes showed enhanced virulence in the host plant when compared to those that lacked the accessory chromosomes (Stewart et al. 2018). Sub-telomeric regions are more prone to chromosomal rearrangements than other chromosomal areas which provides a hotspot for rapid evolution (Zhong et al. 2017). Repeat-rich regions occur in different genomic proportions in Dothideomycetes (Ohm et al. 2012), whereby species such as *L. maculans* contains up to 34% of repetitive elements whilst in *Alternaria alternata* the proportion is less than 3% (Rajarammohan et al. 2019). Although some plant pathogenic fungi could have small proportions of repeat-rich regions, it is still likely to find effector genes near these regions (Ohm et al. 2012).

The beneficial and deleterious mutations in the rapidly evolving regions could occur at similar rates within a population. It is therefore important to allow reshuffling within a population that would maintain beneficial mutations in various individuals. The best way to maintain the beneficial mutations would be to have effector genes in recombination hotspots. A study in *Z. tritici* showed that among the genes that were found in recombination hotspots, genes encoding secreted proteins were enriched the most (Croll et al. 2015). This genomic localization could facilitate the fixation of the beneficial mutations within the population.

1.4 THE GAIN AND LOSS OF EFFECTOR GENES

For a plant pathogenic fungus to successfully invade the host plant, the plant's immune response must be suppressed. However, in many cases, the plant succeeds to fend off the pathogen. The plant can either trigger immune responses by recognizing pathogen associated molecular patterns such as chitin residues through plant pattern recognition receptors (PRR) or by recognizing effectors through the presence of nucleotide binding-leucine rich repeat (NB-LRR) *R* proteins. In some pathosystems, the evasion or recognition of an effector could be the difference between susceptibility or resistance respectively (Friesen et al. 2006; de Wit 2016). As a result, in the gene-for-gene pathosystem models, there is a constant molecular arms race between the pathogen and the host to have the upper hand and ultimately survive. To maintain its pathogenicity across time, the pathogen must gain new effectors or have the current effector repertoire evolve. Recent genomics studies have provided insight into the effector gene characteristics in the genomes of plant pathogenic fungi. In this section, the evolutionary dynamics of effector genes in Dothideomycetes will be discussed (Fig. 1.3). Examples will mostly include studies done with the model plant pathogens *Z. tritici* and *L. maculans*.

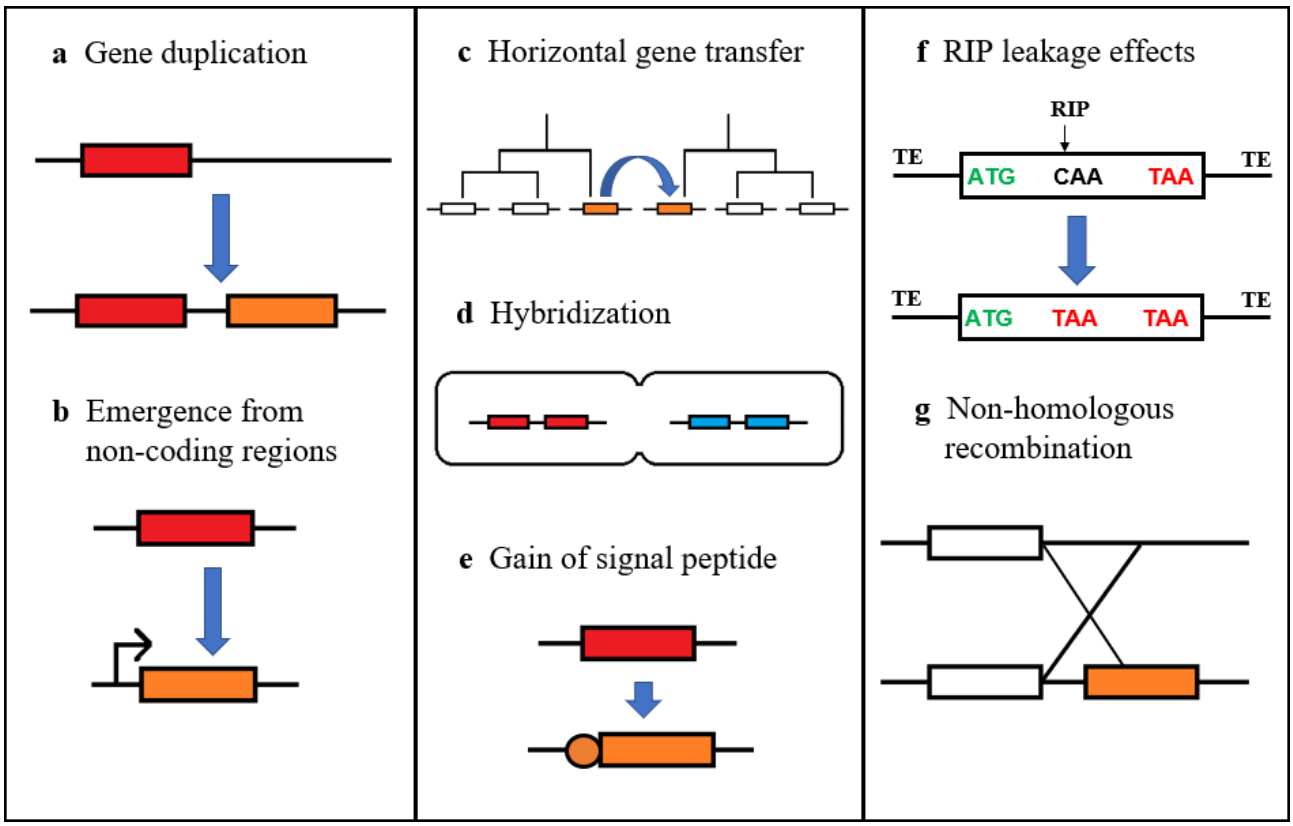


Fig. 1.3 A schematic representation of the evolutionary dynamics of effector genes in Dothideomycetes. The first panel (**a**–**b**) illustrates the *de novo* acquisition of effector genes. The second panel (**c**–**e**) illustrates the emergence and gain of effector genes. The third panel (**f**–**g**) illustrates the loss of avirulence effector genes.

1.4.1 The *de novo* acquisition of effector genes

Effector genes most likely originate like other functional genes within the genome but under a positive selection pressure due to the importance of the gene product in pathogenicity. There are two proposed models that describe how *de novo* genes can be acquired (Tautz and Domazet-Lošo 2011). The first proposes that gene duplication could lead to a paralogue that undergoes diversifying selection to essentially become a new gene with a new function (Fig. 1.3a). This process is known as neofunctionalization. A study in *Z. tritici* showed that a group of sister genes in the A1 and G1 secreted peptidase families experienced accelerated substitution rates after gene duplications (Krishnan et al. 2018). RNA-Seq data revealed that many of the secreted peptidases were differentially expressed during the infection cycle whilst some of them experienced lineage-specific rates of selection when compared to the sister species that infect wild grasses. The study suggests that the selective pressure to have a co-ordinated expression profile with the infection processes will allow the proteins to evolve into virulence proteins that can act as effectors during infection of the host. These effectors would act by suppressing the host's defence responses via proteolysis, a function which can be useful during the biotrophic phase of infection.

The second model proposes that *de novo* genes can emerge from non-coding regions of DNA where mutations lead to the emergence of regulatory elements and start and stop codons (Fig. 1.3b). The creation of *de novo* genes through the second model could occur in multiple proposed ways (McLysaght and Hurst 2016). The promoter of an already existing gene could function bidirectionally to regulate a novel gene in the opposite direction whilst the pre-existing gene can also provide the transcriptional read-through for downstream open reading frames (ORFs) that can act as an extra exon after favourable mutations have occurred (McLysaght and Hurst 2016). Transposable elements (TEs) can transpose a regulatory element to an ORF. Furthermore, mutations to transcription factors (TFs) can allow it to bind to new DNA elements and regulate new genes. There have been a few evidences of the birth of *de novo* genes in *Drosophila*, *Saccharomyces cerevisiae*, *Plasmodium vivax* and *Arabidopsis thaliana* (Begun et al. 2007; Cai et al. 2008; Zhou et al. 2008; Yang and Huang 2011; Donoghue et al. 2011) with very little examples of such cases in

fungi. However, a comparative genomics study of *Z. tritici* between the highly virulent strain 3D7 and the reference genome IPO323 revealed that there were 296 orphan genes that were only found in the 3D7 strain (Plissonneau et al. 2016). These orphan genes had signs of mutational decay and were significantly more likely to be found in clusters. Some of these orphan genes were predicted to encode putative effectors and one of them was highly upregulated during infection (Plissonneau et al. 2016). Although the orphan putative effector genes were transcribed significantly less than the core genes during infection, it is likely that these orphan genes could eventually evolve to have a virulent function because orphan genes are proposed to be ultimately involved in the adaptation of the pathogen to a particular species (Long et al. 2003; Kaessmann 2010). The expression of orphan genes during plant infection would imply a role in plant infection but the level of expression (high or low) could also be a determinant factor. An orphan gene with low expression that could potentially be involved in pathogenesis would benefit more if it was regulated at the exact required amount and time point of infection. Thus, regulatory elements that are bound by the right TFs are arguably as important if not more important than the gene coding a functional protein.

1.4.2 The emergence of effector genes

Effector genes can be acquired through horizontal gene transfer (HGT) (Fig. 1.3c). The phenomenon of HGT was thought to be more prevalent in bacteria but recent studies have shown that such cases do exist in fungi as well. The ToxA effector gene was found to be horizontally transferred from *P. nodorum* to *P. tritici-repentis* (Friesen et al. 2006), two Dothideomycetes that infect wheat plants. Evidence of HGT was shown by results revealing that the ToxA gene was more similar (97%) than the ITS regions (83%) of both species and a transposase gene was found downstream of the gene in both species (Friesen et al. 2006). Additionally, both versions of the ToxA gene produced a functional protein with an identical function despite having two amino acid differences (Friesen et al. 2006). Because (i) the sequence diversity was higher in the ToxA gene from the *P. nodorum* isolates; and (ii) the tan spot disease (caused by *P. tritici-repentis*) reports appeared later than the leaf blotch disease, it was hypothesized that the HGT happened from *P. nodorum* to *P. tritici-repentis* (Friesen et al. 2006). Subsequently, the ToxA gene was also found to be present in the Dothideomycete *Bipolaris sorokiniana* which causes the spot blotch disease in wheat (McDonald et al. 2018). Interestingly, the 12 kbp region containing the gene and few other elements found in *P. nodorum* and *P. tritici-repentis* was also present in *B. sorokiniana* which hinted at another occurrence of HGT. The close association between *P. tritici-repentis* and *B. sorokiniana* in wheat plantations in Central Asia could be the cause of the HGT of the ToxA gene

(McDonald et al. 2018). The mechanism of transfer would likely be through hyphal fusion (Feurtey and Stukenbrock 2018).

Hyphal fusion not only allows the possibility of HGT between fungi, but it can also lead to hybridization which forms a new species (Fig. 1.3d). The Dothideomycete *Zymoseptoria pseudotritici* is proposed to be a hybrid of sister species *Z. tritici* and *Zymoseptoria ardabiliae* (Stukenbrock et al. 2012a). *Z. pseudotritici* infects wild grasses such as *Elymus repens* and *Dactylis glomerata* similarly to its parental species *Z. ardabiliae* which infects *D. glomerata* and *Lolium perenne* (Stukenbrock et al. 2012b). *Z. pseudotritici* was found in grasses that lacked the presence of its parental species which indicated that either the pathogen invaded a new niche or outcompeted both parental species (Stukenbrock 2016). If the hybrid outcompeted both parental species, it could suggest that the hybrid acquired virulence genes that allow it to successfully infect the host as well as a combination of genes that contribute to phenotypic characteristics that provides an advantage in the competition for space. The successful spread of the pathogen shows how hybridization can be beneficial (Stukenbrock et al. 2012a).

The gain of the SnTox1, SnTox3 and SnToxA effectors in *Parastagonospora avenaria* f. sp. *tritici* 1 (Pat1) from *P. nodorum* is another example of the advantage of hybridization. A collection of *Parastagonospora* samples from various continents showed that approximately 20% of the 355 isolates consisted of Pat1 isolates with a prevalence in Canada and Iran (McDonald et al. 2012). Interestingly, 14 isolates showed signs of hybridization. Three *P. nodorum* β-tubulin alleles were found in Pat1 isolates and two Pat1 isolates harboured either a *P. nodorum* MAT1-1 allele or a *P. nodorum* MAT1-2 allele (McDonald et al. 2012). This evidence led to the proposition that hybridization occurred between *P. nodorum* and Pat1 through sexual recombination. A later study showed that Pat1 isolates possessed the *P. nodorum* alleles of the SnTox1, SnTox3 and SnToxA effector genes despite having several differences in conserved housekeeping genes between the two species (McDonald et al. 2013). This study suggests that effector genes can be acquired through hybridization.

For successful proliferation in the host to occur, plant pathogenic fungi must secrete proteins to prevent recognition, immune response or kill the host plant. This secretion process is usually done through the endoplasmic reticulum-Golgi apparatus route and out of the fungal cell whereby the N-terminal secretion signal peptide is required (Lo Presti et al. 2015). Therefore, effector proteins can

emerge via the gain of a secretion peptide (Fig. 1.3e). For host specialization to occur, a species' genes must undergo diversification from its sister species to the point where the genes are only specialized for the specific host. A study showed that a group of genes where positively selected in *Z. tritici* which infects wheat as compared to *Z. pseudotritici* and *Z. ardabiliae* which both infect wild grasses (Poppe et al. 2015). Among these genes, three of them where shown to be involved in virulence. Interestingly, one of the genes (Zt80707) had a sequence encoding a signal peptide whilst the orthologues in *Z. pseudotritici* and *Z. ardabiliae* did not (Poppe et al. 2015). An *in vitro* secretion experiment confirmed that the Zt80707 protein was secreted whilst the *Z. pseudotritici* orthologue failed to be secreted. Additionally, the complementation of the *Z. pseuodotritici* and *Z. ardabiliae* orthologues in the *Z. tritici* Zt80707 mutant failed to rescue the virulence factor in the host plant.

1.4.3 The loss of avirulence effector gene function

Rapidly evolving regions of the genome not only provide a hotspot for effector genes to gain beneficial mutations, but also deleterious mutations. The random and uncertain nature of the hotspot can eliminate effector genes within certain individuals, although depending on the importance of the gene, it is likely that these individuals would die and not pass on their genotype to the next generation. However, these deleterious mutations can also be advantageous in the case of effectors that act as avirulence factors because they would no longer be recognised by the plant's *R* protein. In this section, I discuss a few examples of how deleterious mutations on avirulence effector genes were advantageous to Dothideomycetes.

In the *L. maculans* and oilseed rape pathosystem, there is a gene-for-gene model where the presence of the avirulence effector AvrLm4-7 is recognised by the *R* proteins Rlm4 and Rlm7 in the oilseed rape (Parlange et al. 2009) (Table 1.1). A non-synonymous mutation in the AvrLm4-7 effector allowed the protein to evade recognition by the Rlm4 protein which led to breeding of crops with the Rlm7 gene in some commercial plantations. In these commercial plantations, a study was conducted to analyse the mutations that the AvrLm7 gene would acquire to bypass resistance by Rlm7 containing crops (Daverdin et al. 2012). In the study, they found that multiple mutational signatures including RIP leakage effects were present in various individuals that became virulent on the crops. The RIP mutations created C-to-T mutations in the preferred dinucleotide CA which would lead to premature stop codons (Daverdin et al. 2012). The lack of the avirulence factor prevented the oilseed rape containing the Rlm7 from recognising the pathogen and led to

susceptibility. This shows how RIP mutations could beneficially affect avirulence effector genes (Fig. 1.3f), but it also indicates how it could be deleterious to virulence effector genes.

Similarly, in an effort to escape recognition against wheat plants harbouring the Stb6 resistance gene, the avirulence effector AvrStb6 gene in *Z. tritici* was found to undergo a selection pressure for non-synonymous mutations which would render the gene product non-functional (Zhong et al. 2017). The gene is found in a sub-telomeric region of the chromosome, an area that harbours TEs. In another example, substitution mutations in the effector Avr3D1 led to the evasion of recognition of the virulent *Z. tritici* 3D7 strain from wheat cultivars containing the Stb7 resistance gene (Meile et al. 2018). The Avr3D1 effector contributes to quantitative resistance in the presence of Stb7. The Avr3D1 gene is found in a highly dynamic region of the genome affected by different TE insertion events (Meile et al. 2018). It is likely that this dynamic region led to the mutations. Lastly, the avirulence effector gene Zt_8_609 in *Z. tritici* was found to be located in a region of the chromosome that underwent several chromosomal rearrangements (Hartmann et al. 2017). Isolates that had deletions in the gene as a result of chromosome rearrangements were more virulent on wheat whilst isolates with partial deletions had intermediate virulence (Hartmann et al. 2017). This effector gene was not present in sister species *Z. pseudotritici* and *Z. ardabiliae* which shows that genomic areas with extensive chromosomal rearrangements could also be a region of effector gene emergence. This chromosomal region provides a hotspot for both beneficial and deleterious mutations (Fig. 1.3g). A virulence effector gene found in both sub-telomeric and regions with extensive chromosomal rearrangements could also possibly be subject to deleterious mutations which would hamper its virulence on the host plant.

1.5 THE POPULATION GENETICS OF PHYTOPATHOGENS

For beneficial mutations to become fixed in a population, the magnitude of the effective population size has to be large enough. The effective population size is the number of individuals with a quantity of a specific trait such as alleles that is a representative quantity of the actual population size (Lynch 2007). This means that the rate of loss of heterozygosity would be similar in the effective and actual population size. A large effective population size is essential for beneficial mutations to be fixed or balanced rather than deleterious mutations. Thus, populations that undergo sexual reproduction are more likely to maintain beneficial mutations because of recombination. Smaller effective population sizes would have a higher chance of maintaining deleterious

mutations. There are two evolutionary dynamics in which a beneficial mutation can be maintained, namely the arms race model (as discussed above) and the trench warfare model. In the arms race model, there is a positive directional selection that allows for the fixation of the beneficial mutation at a co-evolving locus (Tellier et al. 2014). In the trench warfare model, there is a positive diversifying selection that allows for multiple alleles to be maintained at a co-evolving locus (Tellier et al. 2014). The arms race model requires a high enough mutational rate for the beneficial mutations to be fixed. This is because the mutation will remain fixed unless a new mutation enters the population. A host that has developed resistance will easily lead to the decline in the size of the pathogen's population if a new mutation does not enter the population. The trench warfare model does not require a high mutational rate because of the balance and presence of multiple alleles at the locus. Due to the diversity of plants in wild environments, the trench warfare model would most likely be present in these wild environments because the presence of multiple alleles in the effector locus would help in combating the diversity. In agricultural plantations however, there is a higher chance of the arms race model being present because agricultural plantations will most likely contain monocultures (McDonald and Stukenbrock 2016). As a result, the high frequency of a certain resistance gene in the crop would pressure the pathogen to ensure a virulence gene that would overcome the resistance gene is fixed in the population.

1.6 CONCLUSIONS AND FUTURE PROSPECTS

In summary, Dothideomycete plant pathogens can infect various economically important crops via different strategies such as the biotrophic, necrotrophic and hemibiotrophic lifestyles. These lifestyles encourage the use of unique effectors that are specialized for the specific lifestyle. Effectors for biotrophs usually allow for recognition evasion, effectors for necrotrophs allow the pathogen to kill the plant cells and hemibiotrophs first use the biotrophic followed by necrotrophic effectors. Because many pathosystems have a gene-for-gene model where the presence of the virulence or resistance protein could be the difference between infection or not, there is a constant molecular arms race between both organisms to continuously evolve and have the upper hand. As a result, many of the effector genes are located in compartments of the genome that allow for rapid evolution. These compartments usually consist of TEs, are in chromosomal rearrangement hotpots or in accessory chromosomes. These rapidly evolving compartments not only create beneficial mutations but also deleterious mutations. Although studies have been conducted to analyse the birth and death of effector genes, there are still questions that remain unanswered.

Studies regarding the regulatory elements of effector genes in Dothideomycetes have been minimal but there have been some recent discoveries regarding their regulation by transcription factors (Tan and Oliver 2017). For example, it was found that the horizontally transferred ToxA effector gene is regulated by a conserved Zn2Cys6 binuclear cluster transcription factor in both *P. nodorum* and *P. tritici-repentis* indicating that the transcription factor is required for necrosis (Rybak et al. 2017). An interesting study would be to determine if the transcription factor only regulates the necrosis related effectors or other genes in the organism involved in infection as well. Similarly, would this transcription factor be responsible for the necrosis switch in hemibiotrophic fungi? If so, then eliminating the transcription factor would be a good target for fungicides. This shows that for an effector gene to be successfully acquired from another fungus, it must be acquired with a promoter which can be recognized by internal transcription factors that regulate the gene during the infection process. Also, because the promoters are near TEs, could RIP leakage effects affect the binding specificity of the transcription factors ensuring that effector genes are regulated at a different time point?

The rapidly evolving genomic compartments are mostly comprised of TEs, and consequently there has been cases of genome expansions mediated by the expansion of TEs which has led to the hypothesis that “bigger is better” (Raffaele and Kamoun 2012). This hypothesis stems from the idea by George C. Williams that lineages that create new species faster and are consequently less likely to become extinct, will dominate biota more than lineages that create species slower. As a result, the hypothesis is that pathogens with bigger flexible and plastic genomes will most likely have a better chance at adapting during co-evolution. However, some plant pathogenic Dothideomycetes have smaller compact genomes with an extremely small proportion of TEs. Would that mean that these pathogens have lower chances of rapidly adapting during co-evolution? If not, then to what extent is the proportion of TEs sufficient or are other random mechanisms at play? Lastly, what TE characteristics are unique to pathogenicity and is there a link between the size of the rapidly evolving compartment and whether the pathogen is specialized to a host or has a broad range of hosts? Answering these questions in future will provide a better understanding of the evolutionary dynamics of plant pathogenic Dothideomycete genomes and how we can overcome this to maintain sustainable crop production.

1.7 REFERENCES

Ahmed MB, Santos KCG dos, Sanchez IB, et al (2018) A rust fungal effector binds plant DNA and modulates transcription. Sci Rep 8:e14718. https://doi.org/10.1038/s41598-018-32825-0

Aktar W, Sengupta D, Chowdhury A (2009) Impact of pesticides use in agriculture: their benefits and hazards. Interdiscip Toxicol 2:1–12. https://doi.org/10.2478/v10102-009-0001-7

Arango Isaza RE, Diaz-Trujillo C, Dhillon B, et al (2016) Combating a Global Threat to a Clonal Crop: Banana Black Sigatoka Pathogen *Pseudocercospora fijiensis* (Synonym *Mycosphaerella fijiensis*) Genomes Reveal Clues for Disease Control. PLOS Genet 12:e1005876. https://doi.org/10.1371/journal.pgen.1005876

Balesdent M-H, Fudal I, Ollivier B, et al (2013) The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene conferring avirulence towards *Brassica rapa*. New Phytol 198:887–898. https://doi.org/10.1111/nph.12178

Balesdent MH, Attard A, Kühn ML, Rouxel T (2002) New Avirulence Genes in the Phytopathogenic Fungus *Leptosphaeria maculans*. Phytopathology 92:1122–1133. https://doi.org/10.1094/PHYTO.2002.92.10.1122

Begun DJ, Lindfors HA, Kern AD, Jones CD (2007) Evidence for *de novo* Evolution of Testis-Expressed Genes in the *Drosophila yakuba* / *Drosophila erecta* Clade. Genetics 176:1131–1137. https://doi.org/10.1534/genetics.106.069245

Bohlin J, Pettersson JH-O (2019) Evolution of Genomic Base Composition: From Single Cell Microbes to Multicellular Animals. Comput Struct Biotechnol J 17:362–370. https://doi.org/10.1016/j.csbj.2019.03.001

Bradshaw RE, Guo Y, Sim AD, et al (2016) Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporum* throughout its infection cycle of the gymnosperm host *Pinus radiata*. Mol Plant Pathol 17:210–224. https://doi.org/10.1111/mpp.12273

Braun U, Crous PW, Dugan F, et al (2003) Phylogeny and taxonomy of *Cladosporium*-like hyphomycetes, including *Davidiella* gen. nov., the teleomorph of *Cladosporium* s. str. Mycol Prog 2:3–18. https://doi.org/10.1007/s11557-006-0039-2

Cai J, Zhao R, Jiang H, Wang W (2008) *De Novo* Origination of a New Protein-Coding Gene in *Saccharomyces cerevisiae*. Genetics 179:487–496. https://doi.org/10.1534/genetics.107.084491

Cambareri EB, Singer MJ, Selker EU (1991) Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. Genetics 127:699–710

Choe Y-H, Kim M, Woo J, et al (2018) Comparing Rock-inhabiting Microbial Communities in Different Rock Types from a High Arctic Polar Desert. FEMS Microbiol Ecol 94:fiy070. https://doi.org/10.1093/femsec/fiy070

Costantini M, Musto H (2017) The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview. J Mol Evol 84:93–103. https://doi.org/10.1007/s00239-017-9785-9

Croll D, Lendenmann MH, Stewart E, McDonald BA (2015) The Impact of Recombination Hotspots on Genome Evolution of a Fungal Plant Pathogen. Genetics 201:1213–1228. https://doi.org/10.1534/genetics.115.180968

Daverdin G, Rouxel T, Gout L, et al (2012) Genome Structure and Reproductive Behaviour Influence the Evolutionary Potential of a Fungal Phytopathogen. PLoS Pathog 8:e1003020. https://doi.org/10.1371/journal.ppat.1003020

de Jonge R, Peter van Esse H, Kombrink A, et al (2010) Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science 329:953–955. https://doi.org/10.1126/science.1190859

de Wit PJGM (2016) *Cladosporium fulvum* Effectors: Weapons in the Arms Race with Tomato. Annu Rev Phytopathol 54:1–23. https://doi.org/10.1146/annurev-phyto-011516-040249

de Wit PJGM, van der Burgt A, Ökmen B, et al (2012) The Genomes of the Fungal Plant Pathogens *Cladosporium fulvum* and *Dothistroma septosporum* Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. PLoS Genet 8:e1003088. https://doi.org/10.1371/journal.pgen.1003088

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

Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C (2011) Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. BMC Evol Biol 11:47. https://doi.org/10.1186/1471-2148-11-47

Dutreux F, Da Silva C, D'Agata L, et al (2018) *De novo* assembly and annotation of three *Leptosphaeria* genomes using Oxford Nanopore MinION sequencing. Sci Data 5:180235. https://doi.org/10.1038/sdata.2018.235

Eyre-Walker A, Hurst LD (2001) The evolution of isochores. Nat Rev Genet 2:549–555. https://doi.org/10.1038/35080577

Feurtey A, Stukenbrock EH (2018) Interspecific Gene Exchange as a Driver of Adaptive Evolution in Fungi. Annu Rev Microbiol 72:377–398. https://doi.org/10.1146/annurev-micro-090817-062753

Friesen TL, Meinhardt SW, Faris JD (2007) The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. Plant J 51:681–692. https://doi.org/10.1111/j.1365-313X.2007.03166.x

Friesen TL, Stukenbrock EH, Liu Z, et al (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet 38:953–956. https://doi.org/10.1038/ng1839

Fudal I, Ross S, Gout L, et al (2007) Heterochromatin-Like Regions as Ecological Niches for Avirulence Genes in the *Leptosphaeria maculans* Genome: Map-Based Cloning of AvrLm6. Mol Plant-Microbe Interact 20:459–470. https://doi.org/10.1094/MPMI-20-4-0459

Gao Y, Faris JD, Liu Z, et al (2015) Identification and Characterization of the SnTox6- Snn6 Interaction in the *Parastagonospora nodorum* –Wheat Pathosystem. Mol Plant-Microbe Interact 28:615–625. https://doi.org/10.1094/MPMI-12-14-0396-R

Gerea AL, Branscum KM, King JB, et al (2012) Secondary metabolites produced by fungi derived from a microbial mat encountered in an iron-rich natural spring. Tetrahedron Lett 53:4202–4205. https://doi.org/10.1016/j.tetlet.2012.05.156

Gerland P, Raftery AE, Ševčíková H, et al (2014) World population stabilization unlikely this century. Science 346:234–237. https://doi.org/10.1126/science.1257469

Ghanbarnia K, Fudal I, Larkan NJ, et al (2015) Rapid identification of the *Leptosphaeria maculans* avirulence gene AvrLm2 using an intraspecific comparative genomics approach. Mol Plant Pathol 16:699–709. https://doi.org/10.1111/mpp.12228

Gibriel HAY, Thomma BPHJ, Seidl MF (2016) The Age of Effectors: Genome-Based Discovery and Applications. Phytopathology 106:1206–1212. https://doi.org/10.1094/PHYTO-02-16-0110-FI

Guo Y, Hunziker L, Mesarich CH, et al (2020) DsEcp2-1 is a polymorphic effector that restricts growth of *Dothistroma septosporum* in pine. Fungal Genet Biol 135:e103300. https://doi.org/10.1016/j.fgb.2019.103300

Hane JK, Paxman J, Jones DAB, et al (2020) "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens – How Many Different Types of Filamentous Plant Pathogens Are There? Front Microbiol 10:e3088. https://doi.org/10.3389/fmicb.2019.03088

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D (2017) A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J 11:1189–1204. https://doi.org/10.1038/ismej.2016.196

Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S (2009) Emerging Concepts in Effector Biology of Plant-Associated Organisms. Mol Plant-Microbe Interact 22:115–122. https://doi.org/10.1094/MPMI-22-2-0115

Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. Genome Res 20:1313–1326. https://doi.org/10.1101/gr.101386.109

Kettles GJ, Bayon C, Sparks CA, et al (2018) Characterization of an antimicrobial and phytotoxic ribonuclease secreted by the fungal wheat pathogen *Zymoseptoria tritici*. New Phytol 217:320–331. https://doi.org/10.1111/nph.14786

Kirk PM, Ainsworth GC, Cannon PF, Minter DW (2008) Ainsworth and Bisby's Dictionary of the Fungi. CABI, Wallingford

Krishnan P, Ma X, McDonald BA, Brunner PC (2018) Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. BMC Evol Biol 18:e7. https://doi.org/10.1186/s12862-018-1123-3

Laugé R, Joosten MHAJ, Van den Ackerveken GFJM, et al (1997) The *In Planta*-Produced Extracellular Proteins ECP1 and ECP2 of *Cladosporium fulvum* Are Virulence Factors. Mol Plant-Microbe Interact 10:725–734. https://doi.org/10.1094/MPMI.1997.10.6.725

Liu F, Selin C, Zou Z, Dilantha Fernando WG (2020) LmCBP1, a secreted chitin-binding protein, is required for the pathogenicity of *Leptosphaeria maculans* on *Brassica napus*. Fungal Genet Biol 136:e103320. https://doi.org/10.1016/j.fgb.2019.103320

Liu Z, Faris JD, Oliver RP, et al (2009) SnTox3 Acts in Effector Triggered Susceptibility to Induce Disease on Wheat Carrying the Snn3 Gene. PLoS Pathog 5:e1000581. https://doi.org/10.1371/journal.ppat.1000581

Liu Z, Gao Y, Kim YM, et al (2016) SnTox1, a *Parastagonospora nodorum* necrotrophic effector, is a dual-function protein that facilitates infection while protecting from wheat-produced chitinases. New Phytol 211:1052–1064. https://doi.org/10.1111/nph.13959

Lo Presti L, Lanver D, Schweizer G, et al (2015) Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol 66:513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Long M, Betrán E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4:865–875. https://doi.org/10.1038/nrg1204

Lynch M (2007) The evolution of genetic networks by non-adaptive processes. Nat Rev Genet 8:803–813. https://doi.org/10.1038/nrg2192

Ma L, Djavaheri M, Wang H, et al (2018) *Leptosphaeria maculans* Effector Protein AvrLm1 Modulates Plant Immunity by Enhancing MAP Kinase 9 Phosphorylation. iScience 3:177–191. https://doi.org/10.1016/j.isci.2018.04.015

Manning VA, Ciuffetti LM (2005) Localization of Ptr ToxA Produced by *Pyrenophora tritici-repentis* Reveals Protein Import into Wheat Mesophyll Cells. Plant Cell 17:3203–3212. https://doi.org/10.1105/tpc.105.035063

Manning VA, Hardison LK, Ciuffetti LM (2007) Ptr ToxA Interacts with a Chloroplast-Localized Protein. Mol Plant-Microbe Interact 20:168–177. https://doi.org/10.1094/MPMI-20-2-0168

Marshall R, Kombrink A, Motteram J, et al (2011) Analysis of Two *in Planta* Expressed LysM Effector Homologs from the Fungus *Mycosphaerella graminicola* Reveals Novel Functional Properties and Varying Contributions to Virulence on Wheat. Plant Physiol 156:756–769. https://doi.org/10.1104/pp.111.176347

McDonald BA, Stukenbrock EH (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. Philos Trans R Soc B Biol Sci 371:e20160026. https://doi.org/10.1098/rstb.2016.0026

McDonald MC, Ahren D, Simpfendorfer S, et al (2018) The discovery of the virulence gene ToxA in the wheat and barley pathogen *Bipolaris sorokiniana*. Mol Plant Pathol 19:432–439. https://doi.org/10.1111/mpp.12535

McDonald MC, Oliver RP, Friesen TL, et al (2013) Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. New Phytol 199:241–251. https://doi.org/10.1111/nph.12257

McDonald MC, Razavi M, Friesen TL, et al (2012) Phylogenetic and population genetic analyses of *Phaeosphaeria nodorum* and its close relatives indicate cryptic species and an origin in the Fertile Crescent. Fungal Genet Biol 49:882–895. https://doi.org/10.1016/j.fgb.2012.08.001

McLysaght A, Hurst LD (2016) Open questions in the study of *de novo* genes: what, how and why. Nat Rev Genet 17:567–578. https://doi.org/10.1038/nrg.2016.78

Meile L, Croll D, Brunner PC, et al (2018) A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. New Phytol 219:1048–1061. https://doi.org/10.1111/nph.15180

Mesarich CH, Stergiopoulos I, Beenen HG, et al (2016) A conserved proline residue in Dothideomycete Avr4 effector proteins is required to trigger a Cf-4-dependent hypersensitive

response. Mol Plant Pathol 17:84–95. https://doi.org/10.1111/mpp.12265

Mesarich CH, Ökmen B, Rovenich H, et al (2018) Specific Hypersensitive Response–Associated Recognition of New Apoplastic Effectors from *Cladosporium fulvum* in Wild Tomato. Mol Plant-Microbe Interact 31:145–162. https://doi.org/10.1094/MPMI-05-17-0114-FI

Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626

Muggia L, Hafellner J, Wirtz N, et al (2008) The sterile microfilamentous lichenized fungi *Cystocoleus ebeneus* and *Racodium rupestre* are relatives of plant pathogens and clinically important dothidealean fungi. Mycol Res 112:50–56. https://doi.org/10.1016/j.mycres.2007.08.025

Muria-Gonzalez MJ, Chooi Y-H, Breen S, Solomon PS (2015) The past, present and future of secondary metabolite research in the Dothideomycetes. Mol Plant Pathol 16:92–107. https://doi.org/10.1111/mpp.12162

Ohm RA, Feau N, Henrissat B, et al (2012) Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLoS Pathog 8:e1003037. https://doi.org/10.1371/journal.ppat.1003037

Ökmen B, Bachmann D, Wit PJGM (2019) A conserved GH17 glycosyl hydrolase from plant pathogenic Dothideomycetes releases a DAMP causing cell death in tomato. Mol Plant Pathol 20:1710–1721. https://doi.org/10.1111/mpp.12872

Ökmen B, Etalo DW, Joosten MHAJ, et al (2013) Detoxification of α-tomatine by *Cladosporium fulvum* is required for full virulence on tomato. New Phytol 198:1203–1214. https://doi.org/10.1111/nph.12208

Oliver RP, Friesen TL, Faris JD, Solomon PS (2012) *Stagonospora nodorum*: From Pathology to Genomics and Host Resistance. Annu Rev Phytopathol 50:23–43. https://doi.org/10.1146/annurev-phyto-081211-173019

Parlange F, Daverdin G, Fudal I, et al (2009) *Leptosphaeria maculans* avirulence gene AvrLm4-7 confers a dual recognition specificity by the Rlm4 and Rlm7 resistance genes of oilseed rape, and circumvents Rlm4 -mediated recognition through a single amino acid change. Mol Microbiol 71:851–863. https://doi.org/10.1111/j.1365-2958.2008.06547.x

Perelló AE, Couretot L, Curti A, et al (2019) First report of spot lesion of wheat caused by *Pyrenophora teres* f. sp *maculata* observed in Argentina. Crop Prot 122:19–22. https://doi.org/10.1016/j.cropro.2019.03.023

Petit-Houdenot Y, Degrave A, Meyer M, et al (2019) A two genes – for – one gene interaction between *Leptosphaeria maculans* and *Brassica napus*. New Phytol 223:397–411. https://doi.org/10.1111/nph.15762

Plissonneau C, Stürchler A, Croll D (2016) The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. MBio 7:e01231-16. https://doi.org/10.1128/mBio.01231-16

Poppe S, Dorsheimer L, Happel P, Stukenbrock EH (2015) Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). PLOS Pathog 11:e1005055. https://doi.org/10.1371/journal.ppat.1005055

Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol 10:417–430. https://doi.org/10.1038/nrmicro2790

Rajarammohan S, Paritosh K, Pental D, Kaur J (2019) Comparative genomics of *Alternaria* species provides insights into the pathogenic lifestyle of *Alternaria brassicae* – a pathogen of the Brassicaceae family. BMC Genomics 20:e1036. https://doi.org/10.1186/s12864-019-6414-6

Robin GP, Kleemann J, Neumann U, et al (2018) Subcellular Localization Screening of *Colletotrichum higginsianum* Effector Candidates Identifies Fungal Proteins Targeted to Plant Peroxisomes, Golgi Bodies, and Microtubules. Front Plant Sci 9:e562. https://doi.org/10.3389/fpls.2018.00562

Robles-Yerena L, Ayala-Escobar V, Leyva-Mir SG, et al (2019) First report of *Cladosporium cladosporioides* causing leaf spot on tomato in Mexico. J Plant Pathol 101:759–759. https://doi.org/10.1007/s42161-018-00218-x

Rouxel T, Grandaubert J, Hane JK, et al (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Commun 2:202. https://doi.org/10.1038/ncomms1189

Rybak K, See PT, Phan HTT, et al (2017) A functionally conserved Zn2Cys6 binuclear cluster transcription factor class regulates necrotrophic effector gene expression and host-specific virulence of two major Pleosporales fungal pathogens of wheat. Mol Plant Pathol 18:420–434. https://doi.org/10.1111/mpp.12511

Saintenac C, Lee W-S, Cambon F, et al (2018) Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. Nat Genet 50:368–374. https://doi.org/10.1038/s41588-018-0051-x

Savary S, Willocquet L, Pethybridge SJ, et al (2019) The global burden of pathogens and pests on major food crops. Nat Ecol Evol 3:430–439. https://doi.org/10.1038/s41559-018-0793-y

Singh PK, Singh RP, Duveiller E, et al (2010) Genetics of wheat–*Pyrenophora tritici-repentis* interactions. Euphytica 171:1–13. https://doi.org/10.1007/s10681-009-0074-6

Steinberg G (2015) Cell biology of *Zymoseptoria tritici*: Pathogen cell organization and wheat infection. Fungal Genet Biol 79:17–23. https://doi.org/10.1016/j.fgb.2015.04.002

Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJGM (2013) Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. FEMS Microbiol Rev 37:67–93. https://doi.org/10.1111/j.1574-6976.2012.00349.x

Stergiopoulos I, van den Burg HA, Okmen B, et al (2010) Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. Proc Natl Acad Sci 107:7610–7615. https://doi.org/10.1073/pnas.1002910107

Stewart E l., Croll D, Lendenmann MH, et al (2018) Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Mol Plant Pathol 19:201–216. https://doi.org/10.1111/mpp.12515

Stukenbrock EH (2016) The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. Phytopathology 106:104–112. https://doi.org/10.1094/PHYTO-08-15-0184-RVW

Stukenbrock EH, Christiansen FB, Hansen TT, et al (2012a) Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc Natl Acad Sci 109:10954–10959. https://doi.org/10.1073/pnas.1201403109

Stukenbrock EH, Jørgensen FG, Zala M, et al (2010) Whole-Genome and Chromosome Evolution Associated with Host Adaptation and Speciation of the Wheat Pathogen *Mycosphaerella graminicola*. PLoS Genet 6:e1001189. https://doi.org/10.1371/journal.pgen.1001189

Stukenbrock EH, Quaedvlieg W, Javan-Nikhah M, et al (2012b) *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the septoria tritici leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). Mycologia 104:1397–1407. https://doi.org/10.3852/11-374

Tan K-C, Oliver RP (2017) Regulation of proteinaceous effector expression in phytopathogenic fungi. PLOS Pathog 13:e1006241. https://doi.org/10.1371/journal.ppat.1006241

Tang L, Yang G, Ma M, et al (2020) An effector of a necrotrophic fungal pathogen targets the calcium-sensing receptor in chloroplasts to inhibit host resistance. Mol Plant Pathol 21:686–701. https://doi.org/10.1111/mpp.12922

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702. https://doi.org/10.1038/nrg3053

Tellier A, Moreno-Gámez S, Stephan W (2014) Speed of adaptation and genomic foorprints of host-parasite coevolution under arms race and trench warfare dynamics. Evolution 68:2211–2224. https://doi.org/10.1111/evo.12427

Thomma BPHJ, van Esse HP, Crous PW, de Wit PJGM (2005) *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. Mol Plant Pathol 6:379–393. https://doi.org/10.1111/j.1364-3703.2005.00292.x

Van de Wouw AP, Lowe RGT, Elliott CE, et al (2014) An avirulence gene, AvrLmJ1, from the blackleg fungus, *Leptosphaeria maculans*, confers avirulence to *Brassica juncea* cultivars. Mol Plant Pathol 15:523–530. https://doi.org/10.1111/mpp.12105

Van den Ackerveken GFJM, Van Kan JAL, de Wit PJ (1992) Molecular analysis of the avirulence gene avr9 of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant J 2:359–366. https://doi.org/10.1111/j.1365-313X.1992.00359.x

van den Burg HA, Harrison SJ, Joosten MHAJ, et al (2006) *Cladosporium fulvum* Avr4 Protects Fungal Cell Walls Against Hydrolysis by Plant Chitinases Accumulating During Infection. Mol Plant-Microbe Interact 19:1420–1430. https://doi.org/10.1094/MPMI-19-1420

van Esse HP, Bolton MD, Stergiopoulos I, et al (2007) The Chitin-Binding *Cladosporium fulvum* Effector Protein Avr4 Is a Virulence Factor. Mol Plant-Microbe Interact 20:1092–1101. https://doi.org/10.1094/MPMI-20-9-1092

van Esse HP, van't Klooster JW, Bolton MD, et al (2008) The *Cladosporium fulvum* Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense. Plant Cell 20:1948–1963. https://doi.org/10.1105/tpc.108.059394

Vleeshouwers VGAA, Oliver RP (2014) Effectors as Tools in Disease Resistance Breeding Against Biotrophic, Hemibiotrophic, and Necrotrophic Plant Pathogens. Mol Plant-Microbe Interact 27:196–206. https://doi.org/10.1094/MPMI-10-13-0313-IA

Wang M, Weiberg A, Dellota E, et al (2017) *Botrytis* small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. RNA Biol 14:421–428. https://doi.org/10.1080/15476286.2017.1291112

Wang M, Weiberg A, Jin H (2015) Pathogen small RNAs: a new class of effectors for pathogen attacks. Mol Plant Pathol 16:219–223. https://doi.org/10.1111/mpp.12233

Wang X, Jiang N, Liu J, et al (2014) The role of effectors and host immunity in plant–necrotrophic fungal interactions. Virulence 5:722–732. https://doi.org/10.4161/viru.29798

Westerink N, Brandwagt BF, De Wit PJGM, Joosten MHAJ (2004) *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the Cf-4 locus (Hcr9-4E) by secretion of a stable avr4E isoform. Mol Microbiol 54:533–545. https://doi.org/10.1111/j.1365-2958.2004.04288.x

Wijesekera K, Mahidol C, Ruchirawat S, Kittakoop P (2017) Metabolite diversification by cultivation of the endophytic fungus Dothideomycete sp. in halogen containing media: Cultivation of terrestrial fungus in seawater. Bioorg Med Chem 25:2868–2877. https://doi.org/10.1016/j.bmc.2017.03.040

Yang Z, Huang J (2011) *De novo* origin of new genes with introns in *Plasmodium vivax*. FEBS Lett 585:641–644. https://doi.org/10.1016/j.febslet.2011.01.017

Zhong Z, Marcel TC, Hartmann FE, et al (2017) A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytol 214:619–631. https://doi.org/10.1111/nph.14434

Zhou K, Yang YH, Wu MD, Li GQ (2019) First Report of *Leptosphaeria biglobosa* Causing Blackleg of Ornamental Kale (*Brassica oleracea* var. *acephala*) in China. Plant Dis 103:770–770. https://doi.org/10.1094/PDIS-06-18-1015-PDN

Zhou Q, Zhang G, Zhang Y, et al (2008) On the origin of new genes in *Drosophila*. Genome Res 18:1446–1455. https://doi.org/10.1101/gr.076588.108

CHAPTER 2 THE GENOME SEQUENCE OF *CERCOSPORA ZEINA* REVEALS A BIPARTITE GC-CONTENT AND PUTATIVE EFFECTOR GENES

2.1 ABSTRACT

The grey leaf spot (GLS) disease is an important foliar disease of maize. In Africa, the disease is predominantly caused by the hemibiotrophic Dothideomycete *Cercospora zeina*. Very little is known about the pathogenesis and effector repertoire of this pathogen. A complete genome assembly can reveal the range of effector and pathogenesis-related genes available for the pathogen. Here, we present a contiguous genome assembly of *C. zeina* generated by PacBio single-molecule real-time sequencing technology that was polished with Illumina read data. The assembly has a total of 22 contigs of which 17 are nuclear genome sequences that make up the 41 Mbp genome. There are at least three possible full chromosomes within the 17 nuclear genome contigs. The annotation of the genome has revealed 12,436 predicted genes of which 1,025 are predicted to have a signal peptide. The secretome contains many proteins predicted to have oxidoreductive and peroxidase activities. Effector prediction revealed a total of 274 effectors which also included potential homologues of the ECP2, ECP6 and AVR4 effectors from *Cladosporium fulvum*. These homologues had high expression counts *in planta*. These results imply that *C. zeina* potentially secretes proteins that prevent recognition by the host plant and protect against oxidative stress. Additionally, genome architecture analysis of *C. zeina* has revealed a bipartite structure consisting of 33.2% AT-rich compartments and 66.8% GC-rich compartments. However, effector genes are not concentrated in the AT-rich compartments as is the case for other phytopathogenic fungi. This study has paved way for the characterization of candidate effectors which could further our understanding of the pathogenicity of *C. zeina*. This study could also ultimately lead to the possibility of effector-based breeding of maize.

2.2 INTRODUCTION

The grey leaf spot (GLS) disease is a foliar yield-limiting disease of maize (*Zea mays*). The disease is characterized by rectangular grey lesions that run parallel to the leaf veins. The lesions prevent the plant from photosynthesizing which ultimately leads to yield loss. There have been reports of up to 79% yield loss in Cameroon and 44% in Iowa, USA in the past (Munkvold et al. 2001; Ngoko et al. 2002), making it a disease of global importance (Ward et al. 1999). In southern Africa, the disease is caused by the plant pathogenic fungus, *Cercospora zeina* (Meisel et al. 2009). The pathogen belongs to the class Dothideomycetes which consists of many plant pathogenic fungi that infect a wide range of plants, including monocots such as wheat and barley and dicots such as pine

and poplar trees. The two largest Dothideomycete orders are the Pleosporales and Capnodiales. The Pleosporales includes necrotrophs such as *Parastagonospora nodorum* and *Pyrenophora tritici-repentis* that both infect wheat. The Capnodiales includes the biotroph *Cladosporium fulvum* (syn. *Passalora fulva*) that infects tomato and the hemibiotrophs *Zymoseptoria tritici* and *Pseudocercospora fijiensis* that infect wheat and banana respectively. *C. zeina* belongs to the Capnodiales (Ohm et al. 2012).

GLS disease symptoms usually appear about 14 to 28 days after infection (Meyer et al. 2017). This latent period has led to the proposition that *C. zeina* is a hemibiotroph rather than a necrotroph as it is widely known. Many Dothideomycete hemibiotrophs and necrotrophs produce more secondary metabolites (SMs) synthesized by enzymes like polyketide synthases (PKS) that aid in the toxification of the host plant (Ohm et al. 2012). The biotroph *C. fulvum* has eight PKS genes but are lowly expressed *in planta* (Griffiths et al. 2015, 2016). This led to the proposition that the lack of SM expression is needed for the stealthy biotrophic stage of colonization (Collemare et al. 2014). Many *Cercospora* species produce the SM toxin cercosporin during infection (Daub and Ehrenshaft 2000; Weiland et al. 2010). However, *C. zeina* lacks the CTB7 gene that is thought to be required for cercosporin production (Swart et al. 2017). This led to the hypothesis that *C. zeina* has another SM toxin that is being produced. Alternatively, other macromolecules like effectors could be important in the infection process.

The sequencing of plant pathogenic fungal genomes has increased in importance as it can provide information about the pathogenicity of an organism. The genome comparisons of the pathogens *C. fulvum* and *Dothistroma septosporum* revealed how each pathogen adapted to their respective lifestyles (de Wit et al. 2012). Additionally, the analysis of the genome sequence of *P. fijiensis* paved way for the identification of a novel PKS gene cluster (Noar et al. 2019). Nowadays, bioinformatics tools have made it possible to identify the full catalogue of effector genes from the genome (Sonah et al. 2016; Dalio et al. 2018; Jones et al. 2018). Effectors are small secreted proteins that promote infection in the host by supressing host defences or preventing recognition by the host. Effectors are secreted upon infection of the host and can work in various cellular compartments, including the apoplastic space, the cytoplasm and organelles (Lo Presti et al. 2015). They can either act as virulence factors whereby their presence causes disease in the plant or prevent the recognition by the host plant.

The genome sequences of plant pathogenic fungi not only provide information about the pathogenicity and lifestyle of the pathogens but can also give information about the evolutionary dynamics of the pathogenicity genes. Many plant pathogenic fungi have a "two-speed" genome that consists of a slowly evolving gene-rich, repeat-poor compartment and a rapidly evolving gene-poor, repeat-rich compartment (Dong et al. 2015). The rapidly evolving compartment can be an AT-isochore, a dispensable chromosome, transposable element (TE)-rich region or a chromosomal breakpoint. The highly dynamic nature of this compartment allows for the rapid evolution of the effector or other pathogenicity-related genes through mutations or rearrangements. Consequently, it has become important to have a complete genome that encompasses the majority of these repeat-rich regions that can harbour effector genes. Second-generation sequencing technologies generate fragmented assemblies that can lead to the under-representation of repetitive elements (Gibriel et al. 2016). This problem has been solved by the availability of third-generation sequencing technologies such as single-molecule real-time (SMRT) sequencing that sequence long reads and generate gapless assemblies. The resequencing of the genome of the pathogen *Verticillium dahliae* with SMRT technology revealed a 3-fold increase in repetitive content when compared to the older genome assembled with Illumina data (Faino et al. 2015).

The genome of *C. zeina* was recently sequenced with the Illumina sequencing technology (Wingfield et al. 2017). To identify candidate effector genes in rapidly evolving genomic compartments, a less fragmented genome assembly was required. This provided the rationale for this study in which effector gene identification and genome architecture analysis was performed on a genome assembly of *C. zeina* that had been sequenced with third-generation sequencing technology. We hypothesized that *C. zeina* harbours a few hundred effector genes like other Dothideomycetes and that the genome has a bipartite structure whereby effector genes localize in AT-rich compartments.

In this study, we sequenced and assembled the *C. zeina* genome with Pacific Biosciences (PacBio) SMRT sequencing technology and used this newly assembled genome for effector identification and characterization. The analysis revealed 274 candidate effector genes and a bipartite genome architecture that has a higher effector gene density in GC-rich compartments.

2.3 MATERIALS AND METHODS

2.3.1 Fungal isolate and culture conditions

The *Cercospora zeina* isolate CMW 25467 originates from the Mkushi region in Zambia and was isolated in March 2007 (Meisel et al. 2009). Glycerol stock of this isolate was plated out and maintained on V8 media. Once a week conidia were patted onto new plates by cutting up the agar blogs containing the fungal spores and patting the pieces upside down on new medium. Plates were kept in the dark at 25°C to initiate growth and sporulation (Beckman and Payne 1983). This isolate was used to infect greenhouse-grown maize plants. The fungus was re-isolated from GLS lesions by separating single spores from the protruding conidiophores observed under a dissecting microscope (90X magnification) and placing them on a new V8 plate (Heystek 2014). A pat culture was maintained to bulk up the fungus until a sufficient amount of spores could be scraped off and stored as a 50% glycerol stock at -80°C in 2012. This culture was then used for DNA extraction and sequencing.

2.3.2 DNA extraction and sequencing

The gDNA was extracted using the CTAB extraction protocol (Allen et al. 2006). The resulting gDNA was quantified on both a 1% agarose gel and Qubit 4 fluorometer. Samples were pooled to get 3.2 µg of DNA in total. PacBio sequencing was performed by the Max Planck Genome Centre (Cologne, Germany) using one SMRT cell.

2.3.3 Genome assembly and filtering

The assembly was performed at the Max Planck Institute. The PacBio raw reads were assembled using HGAP 4 included in SMRTLink version 6 with default parameters. The draft assembly was polished using Arrow as implemented in SMRTLink version 6. The depth of coverage obtained by realignment of reads on the draft assembly was used to filter the assembly: we discarded the contigs with a depth deviating by more than 1.5X from the average coverage across all contigs weighted by the contig length, as in (Plissonneau et al. 2016). The filtered assembly was then aligned with the TruSeq Illumina reads from the paired-end library sequenced in (Wingfield et al. 2017) using the Burrows-Wheeler Alignment (BWA) tool (Li and Durbin 2009) followed by one round of polishing with Pilon (Walker et al. 2014) using the default settings. We used Bowtie (Langmead et al. 2009)

to look for the telomeric repeat CCCTAA in the scaffolds of the polished assembly to identify fully assembled chromosomes (Langmead and Salzberg 2012). This repeat has been identified in diverse organism including fungi and in particular plant pathogens (Fulnečková et al. 2013; King et al. 2015). The entire genome was searched for loci that consisted of 10 or more times the repeat length. Assembly statistics were determined with QUAST v4.4 (Gurevich et al. 2013).

2.3.4 Genome annotation

Before annotating the genome, RNA-Seq reads were mapped onto the genome using HISAT2 v2.1.0 (Kim et al. 2015) with the following parameters: --pen-noncansplice 18, --mp 6,0, --no-softclip, --max-intronlen 10000, -t –reorder. The RNA-Seq data was pooled from 5 different sources. The first was from RNA extracted from *C. zeina* infected CML444 and SC Malawi maize leaves 103 days after planting (DAP) in Baynesfield Estate in KwaZulu-Natal province in South Africa in March 2009. The second was from *C. zeina* infected maize recombinant inbred line (RIL165) derived from a CML444 and SC Malawi cross 103 DAP also in Baynesfield Estate in March 2009 (Meyer et al. 2017). The third was from *C. zeina* infected SC Malawi maize 96 DAP in Baynesfield Estate in March 2012. The fourth was from *C. zeina* inoculated B73 maize leaves 77 DAP in Hildesheim farm, Greytown in KwaZulu-Natal province in South Africa in March 2013 (Swart et al. 2017). The fifth was from *in vitro* cultures of *C. zeina* from seven different conditions (Swart et al. 2017). All RNA-Seq reads were paired-end reads. The BAM file of the combined concordant pair alignments was then used in the annotation using the software BRAKER v2.0.6 (Hoff et al. 2016) with the --fungus parameter activated. The quality of the annotation in terms of recovery of gene space was assessed using BUSCO v3.0.2 (Simão et al. 2015) with the Ascomycota dataset. The functional annotation of the predicted proteins from the PacBio assembly was performed using Blast2GO (Conesa and Götz 2008). BLASTN analysis of the entire contigs 16F, 18F and 20F was performed using the default settings against the non-redundant Nucleotide collection (nr/nt) of the Standard database on the NCBI website (https://www.ncbi.nlm.nih.gov/). Gene prediction of these contigs was performed using AUGUSTUS v3.3 (Stanke et al. 2006) with *ab initio* default settings.

2.3.5 Secretome, CSEP and CAZyme prediction

Proteins with signal peptides were predicted using SignalP v4.1 (Petersen et al. 2011). Transmembrane domains were predicted using TMHMM v2.0 (Krogh et al. 2001). Effector proteins

were predicted with the machine learning program EffectorP v2.0 using the default dataset (Sperschneider et al. 2016). The online tools LOCALIZER v1.0.4 (Sperschneider et al. 2017) and ApoplastP v1.0 (Sperschneider et al. 2018) were used to predict the localization of effector proteins using the default settings. CAZymes were predicted using dbCAN along with HMMER v3.2.1 (e-value < 1e-15, coverage > 0.35) (Finn et al. 2011; Zhang et al. 2018). To identify the potential proteins involved in pathogenicity, BLASTP search was performed against Pathogen-Host Interaction database (PHI-base) (database 4.8) (Winnenburg 2006; Urban et al. 2019) with an e-value of ≤ 1e-5. Tandem repeat domains were predicted using XSTREAM (Newman and Cooper 2007).

2.3.6 Leaf material for RNA-Seq analysis

Susceptible B73 maize leaves were exposed in March 2013 to natural inoculum of *C. zeina* at a hotspot for GLS disease (Hildesheim farm, Greytown in KwaZulu-Natal province in South Africa). GLS lesions initially appeared in the lower leaves then progressed to the upper leaves. Leaves were collected at 77 DAP. At this stage, the digital image analysis determined that the lower leaves had necrotic lesions that covered an average of 8% of the leaf surface area and the upper leaves had chlorotic lesions that covered only an average of 0.2% of the leaf surface area. Two lower leaves (second and third leaf internode below ear) and two upper leaves (send and third internode above ear) were collected for RNA extraction. The two lower leaves and two lower leaves from one plant were pooled respectively for RNA extraction. This collection was done for three randomly selected inoculated maize plants. The leaf material, RNA samples and RNA-Seq raw data was exactly the same as that described in (Swart et al. 2017).

2.3.7 Differential gene expression analysis

RNA-Seq analysis of the B73-*C. zeina* dataset has already been reported in (Swart et al. 2017) (NCBI GEO accession GSE94442), however the reads had been mapped to maize B73 RefGen_v2 (v5b.60) and the Illumina assembly of *C. zeina* CMW 25467 (Wingfield et al. 2017). In this study, improved annotations and assemblies of the genomes were available (maize B73 RefGen_v4 based on PacBio SMRT sequencing (Jiao et al. 2017); and the *C. zeina* assembly based on PacBio SMRT sequencing (this study)). Therefore, the RNA-Seq analysis was redone. RNA-Seq read quality was analysed using the FastQC software (Andrews 2010). Whenever trimming of RNA-Seq reads was necessary, Trimmomatic (Bolger et al. 2014) was used for the trimming of low quality reads. The

reads were mapped to both the B73 RefGen_v4 and the *C. zeina* CMW 25467 PacBio genome assemblies at the same time using HISAT2 v2.1.0 with the same parameters as described above. Read counts were determined using featureCounts (Liao et al. 2014) and the differential expression analysis was performed using DESeq2 (Love et al. 2014). After the DESeq2 analysis was completed, the *C. zeina* transcript data was then extracted, and the maize data was discarded for the purposes of this study.

2.3.8 Repetitive element identification, RIP mutation analysis and genome architecture analysis

Repetitive elements were identified using the PiRATE virtual machine (Berthelier et al. 2018). The REPET pipeline (Flutre et al. 2011) in PiRATE was used for repeat element identification and annotation. This involves identification based on repetitiveness using TEdenovo and RepeatScout (Price et al. 2005). This is followed by clustering with CD-HIT (Li and Godzik 2006; Fu et al. 2012), classification with PASTEC (Hoede et al. 2014) and annotation with TEannot (Quesneville et al. 2005). RIP mutations were analysed using the GC content and alignment-based approach in the RIPCAL software (Hane and Oliver 2008) after it was aligned using MAFFT v7 (Katoh and Standley 2013). The proportion of RIP within the genome was analysed using The RIPper with default settings (van Wyk et al. 2019). The OculterCut v1.1 software (Testa et al. 2016) was used to identify whether the *C. zeina* genome was compartmentalised. The gene density plot was generated using the R script from (Saunders et al. 2014). The circus plot was generated using an R script from (Gu et al. 2014).

2.4 RESULTS

2.4.1 Genome assembly and annotation

Genome assembly using HGAP 4 with the PacBio data generated from a single SMRT cell resulted in a 41 Mbp assembly for *C. zeina* consisting of 22 contigs of which 17 contigs had a mean coverage of 260X (Table 2.1, Appendix Fig. 1). The five contigs not included in the set of 17 had either a low mean coverage less than 260X (contigs 16F, 18F, and 20F) or a very high mean coverage above 1000X (contigs 19F and 21F) (Appendix Table 1, Appendix Fig. 1). A BLASTN search of contigs 16F and 18F against the nr/nt database (NCBI) both resulted in a similarity to chromosome sequences from *Zymoseptoria tritici* and transposable element sequences of

Cercospora zeae-maydis (Appendix Table 1). The BLASTN search of contig 20F did not have any significant match other than two *C. zeina* microsatellite sequences (CzSSR12 and CzSSR11) previously used as population genetics markers (Muller et al. 2016). The BLASTN search of contig 19F resulted in a similarity to ribosomal RNA sequences including the 18S subunit ribosomal RNA gene of various Capnodiales including *C. sojina*, *Z. tritici*, and *C. fulvum*. This suggested that contig 19F contains the ribosomal RNA cistron of *C. zeina*. The BLASTN search of contig 21F resulted in a similarity to mitochondrial genome sequences of fungi such as *Acephala applanata*, *Phialocephala subalpina* and *Aspergillus fumigatus*. This could imply that contig 21F contains the sequences of the *C. zeina* mitochondrial genome. Therefore, for this study, subsequent analyses reported in this dissertation were only performed on the 17 contigs predicted to contain nuclear genome protein-coding sequences.

The contig lengths of the 17 selected nuclear genome contigs ranged from 29,391 to 5,172,692 bp (Table 2.1). The telomeric repeat CCCTAA found in plant pathogens (Fulnečková et al. 2013; King et al. 2015) was searched within the different contigs. These repeat loci were found at both ends of three contigs, and at one end of 12 contigs. This left two contigs in which no telomeric repeat pattern were detected (Table 2.2).

Table 2.1 Statistics of the *C. zeina* PacBio assembly

	Draft assembly[a]	Filtered assembly[b]	Polished filtered assembly[c,d]
Number of contigs	22	17	17
Number of contigs (≥10,000 bp)	21	17	17
Total length	41,824,694	41,717,156	41,717,156
Total length (≥10,000 bp)	41,819,162	41,717,156	41,717,156
Largest contig (bp)	5,193,669	5,193,669	5,193,669
Total length	41,824,694	41,717,156	41,717,156
GC content (%)	47.93	47.96	47.96
N50[e] (bp)	3,977,222	3,977,222	3,977,222
N75 (bp)	2,708,737	2,708,737	2,708,737
L50[f] (# of contigs)	5	5	5

L75 (# of contigs)	8	8	8

[a] Assembly from HGAP and polished with Arrow.

[b] Filtered through realignments and discarding of contigs with a depth deviating by more than 1.5X from the average coverage across all contigs weighted by the contig length.

[c] Assembly polished with Illumina data using Pilon.

[d] Polishing with Pilon made 662 single nucleotide changes but no insertions or deletions.

[e] The N50 statistic is the length of the shortest contig need to make up 50% of the genome length. N75 for 75%.

[f] The L50 statistic is the minimum number of contigs needed to make up 50% of the genome length. L75 for 75%.

The BUSCO genome completeness assessment using BUSCO gene prediction algorithms with the Ascomycota dataset (1,315 genes) showed a genome assembly completeness of 97.2% with 0.1% duplicated, 1.7% fragmented and 1.1% missing (Appendix Table 2). The PacBio sequencing seemed to have recovered a few BUSCO genes as the amount missing decreased from 20 to 15 when compared to the Illumina assembly (Wingfield et al. 2017).

The next step was to annotate the *C. zeina* genome by predicting protein-coding genes using the BRAKER pipeline which uses a two-step method that begins with GeneMark-ET generating initial gene structures followed by AUGUSTUS that uses the initial gene structures for training and integrates RNA-Seq genome alignments to the final gene predictions (Hoff et al. 2016). A total of 12,436 protein-coding genes (including alternative transcripts) were predicted for the 17 nuclear genome contigs of *C. zeina*. This is 326 more than that predicted for the Illumina assembly which had 12,110 predicted protein-coding genes (including alternative transcripts) (Wingfield et al. 2017), indicating that the long-read sequencing with PacBio has revealed additional gene models that would not have been identified in a more fragmented assembly. The number of protein-coding genes predicted in the newly assembled *C. zeina* genome is similar to that of other Dothideomycetes which range from 9,739 in *Mycosphaerella populicola* to 14,186 in *Pyrenochaeta lycopersici* (Ohm et al. 2012; Dal Molin et al. 2018). None of the *in planta* and *in vitro* RNA-Seq reads mapped to contigs 16F, 18F or 20F (data not shown). Therefore, *ab initio* gene prediction was performed on these contigs using AUGUSTUS and no genes were predicted (data not shown). This further shows that these contigs are made up of repetitive sequences and do not appear to contain protein-coding genes such as effectors.

Table 2.2 Features of the 17 selected nuclear genome contigs of the *C. zeina* PacBio assembly

Contig ID[a]	Telomeric repeats	Contig length	Number of protein-coding genes	Number of CSEPs
00F	Downstream only	5,193,802	1,397	33
01F	Downstream only	5,181,497	1,444	40
02F	Both ends	4,949,701	1,359	37
03F	Upstream only	4,383,157	1,226	17
04F	Downstream only	3,977,367	971	29
05F	Downstream only	3,832,675	1,324	24
06F	Both ends	3,460,717	1,008	24
07F	Downstream only	2,708,767	858	15
08F	Both ends	2,565,113	718	25
09F	Upstream only	2,354,587	733	19
10F	Downstream only	1,532,051	361	8
11F	Downstream only	671,136	64	0
12F	Upstream only	490,194	58	2
13F	Upstream only	234,854	28	0
14F	-	97,051	21	1
15F	Upstream only	55,435	0	0
17F	-	29,610	0	0

[a] Contigs 16F, 18F, 19F, 20F and 21F were considered non-protein coding or mitochondrial genome contigs and were therefore not included for annotation analyses.

The BUSCO completeness assessment was performed for the predicted proteins using the Ascomycota lineage because it was the most specific lineage. This assessment assesses the completeness of the predicted protein set. The assessment showed a completeness of 96.8% with 2.7% duplicated, 2.7% fragmented and 0.5% missing (Appendix Table 3). The mean gene length of the protein-coding genes was 1,393 bp with a maximum and minimum length of 26,634 bp and 201 bp respectively (Table 2.3). Functional annotation was performed via Blast2GO and revealed that 8,925 genes matched a homologous gene from the NCBI non-redundant database, 2,676 genes were hypothetical and 835 had no match.

Table 2.3 Nuclear genome features of *C. zeina*

Parameter	Count[c]
Total protein-coding genes (including variants[a])	12,346
Total protein-coding genes (excluding variants)	11,570
Mean transcript length (bp)	1,390
Max gene length[b] (bp)	26,634
Min gene length (bp)	201
Mean intron length (bp)	85
Mean number of introns per gene	2.43
Mean exon length (bp)	499
Percentage of multi-exonic genes	73.54
Mean gene density (genes/Mbp)	277

[a] Variants – predicted alternative transcripts

[b] Including introns

[c] Count per parameter in the 17 nuclear genome contigs.

2.4.2 Carbohydrate-active enzymes (CAZymes)

Plant pathogenic fungi need require carbohydrates from the host plant during colonisation and thus secrete enzymes capable of metabolizing complex carbohydrate structures from the plant such as the cuticle and cell wall components. The CAZyme catalogue in the fungus can provide insight into the various biological interactions that it has with the host. Predicted CAZymes in *C. zeina* revealed a total of 448 proteins, of which 149 were predicted to be secreted. There were 213 glycoside hydrolases (GH), 67 auxiliary activities (AA), 58 carbohydrate esterases (CE), 6 polysaccharide lyases (PL), 96 glycosyltransferases (GT) and 8 carbohydrate-binding modules (CBM). Four of the 8 CBMs were also predicted to contain a GH domain and one was predicted to contain an AA1 domain. The number of CAZymes in *C. zeina* was on the smaller range when compared to other Dothideomycetes where they range from 457 in *Z. tritici* to 744 in *P. lycopersici* (Zhao et al. 2014; McGrann et al. 2016; Dal Molin et al. 2018). The GH group of CAZymes are of particular interest to researchers because of their ability to hydrolyse glycosidic bonds between carbohydrates, a feature which is common for cell wall degrading enzymes (CWDEs). The most abundant GH families in *C. zeina* are GH16 with 16 members, GH3 with 16 and GH13 with 15. The GH16 family consists of enzymes like endoglucanases that degrade hemicellulose (Scheller and Ulvskov 2010),

GH3 consists of β-glucosidases, α-L-arabinofuranosidases and β-D-xylopyranosidases (Harvey et al. 2000; Sørensen et al. 2013) and GH13 consists α-amylases and pullulanases (Stam et al. 2006; Kumar 2010). Plant cell walls contain cellulose, and thus secreted cellulases can break down the cell walls during infection for pathogens. *C. zeina* has genes encoding enzymes that are part of the well-known GH7, GH12 and AA9 cellulase families. When comparing these cellulose degrading CAZymes with other Dothideomycetes, there is a pattern that shows that Pleosporales have generally higher amounts than the Capnodiales (Fig. 2.1). The most notable difference is the AA9 family which appears to increase almost 5-fold in the Pleosporales. The AA9 family (previously GH61) are not necessarily cellulases but consist of lytic polysaccharide monooxygenases that aid cellulases in the cellulose hydrolysis (Agger et al. 2014).

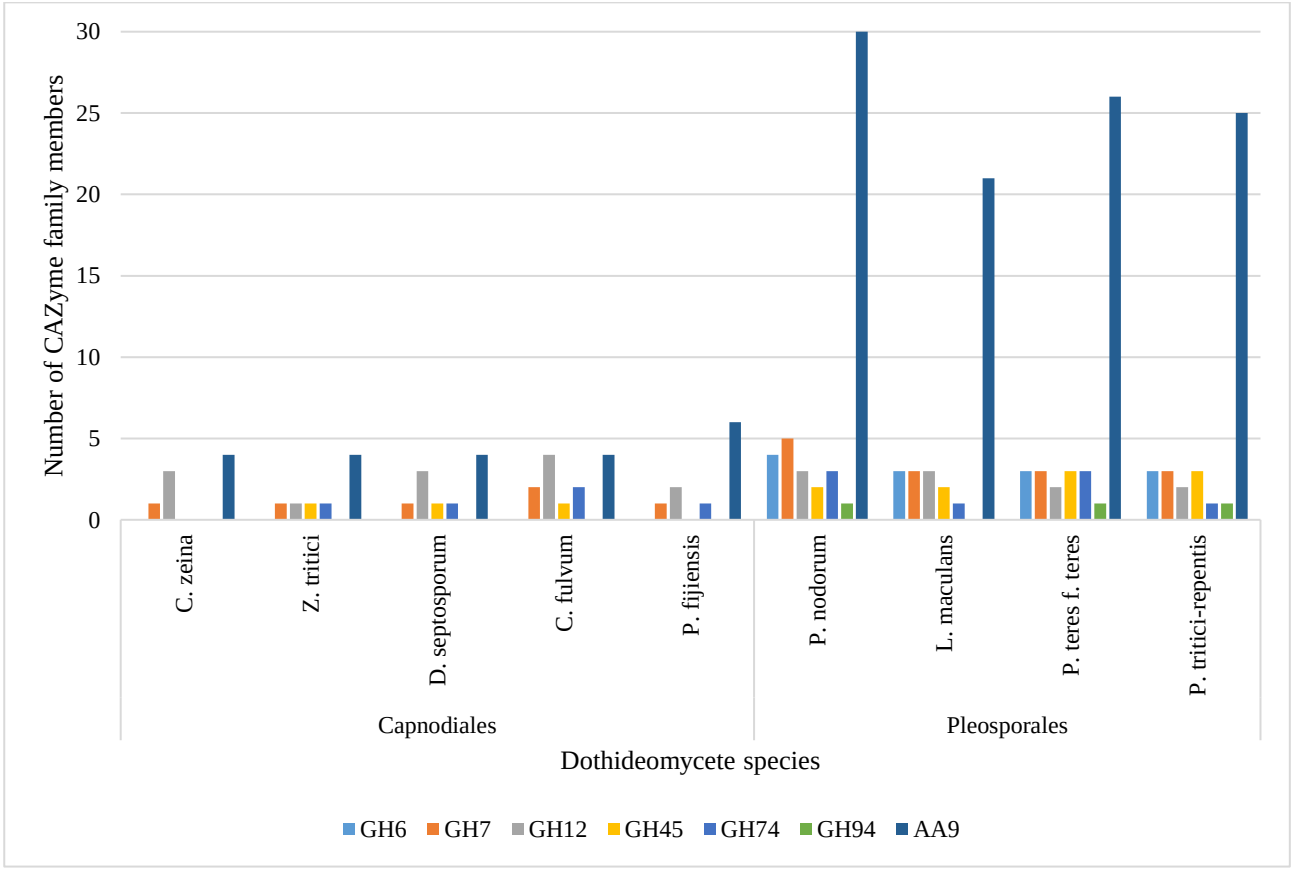


Fig. 2.1 Complement of *C. zeina* cellulose degrading enzymes compared with other Capnodiales and Pleosporales. Capnodiales: *Cercospora zeina*, *Zymoseptoria tritici*, *Dothistroma septosporum*, *Cladosporium fulvum* and *Pseudocercospora fijiensis*. Pleosporales: *Parastagonospora nodorum*, *Leptosphaeria maculans*, *Pyrenophora teres* f. *teres* and *Pyrenophora tritici-repentis*. Data collected from (McGrann et al. 2016; Dal Molin et al. 2018).

2.4.3 Candidate Secreted Effector Proteins (CSEP)

A total of 1,025 genes were predicted to encode a protein containing a signal peptide. Among these proteins, the Gene Ontology (GO) terms were dominated by hydrolase activity, oxidoreductase activity and metal ion binding (Table 2.4). Predicted CAZymes made up 184 proteins in the secretome. A total of 210 signal peptide-containing proteins where predicted to have a transmembrane domain which indicates that they could be membrane-bound proteins. A combination of two approaches were performed to identify the Candidate Secreted Effector Proteins (CSEPs) of *C. zeina*. The first involved the identification of small secreted proteins (SSPs) using a pipeline that has been suggested for fungal genomes (Dalio et al. 2018). The predicted secreted proteins that lacked the transmembrane domain consisted of 815 proteins, of which 358 were less than 300 amino acids in size and 214 of them had ≥ 4 cysteine residues which led to a total of 214 predicted SSPs. The second approach involved the prediction of effector proteins using the EffectorP program that uses a machine learning approach to identify effectors based on already annotated effector proteins from other fungi. A total of 182 predicted effectors were identified using the second approach. The combination of both approaches yielded a total of 274 CSEPs for *C. zeina* (Table 2.5). This number is in the same range as that found in other Dothideomycetes which range from 67 in *Baudoinia compniacensis* to 330 in *Z. tritici* strain 3D7 (Ohm et al. 2012; Arango Isaza et al. 2016; Zaccaron and Bluhm 2017; Wyatt et al. 2018; Richards et al. 2018; Luo et al. 2018; Plissonneau et al. 2018). Among these, 151 had a Blast2GO annotation and 123 were hypothetical. The two predicted *C. zeina* CSEPs g20 and g3999 were found to be homologous to the effector proteins ECP2 and ECP6 respectively from *C. fulvum* (de Wit 2016). Another predicted CSEP g1947 was homologous to the AVR4 effector also from *C. fulvum* (van den Burg et al. 2006). Of the 274 CSEPs, 19 of them were predicted to be CAZymes. Most of them were from the GH and CE families with 8 and 7 respectively. The functions of potential pathogenicity proteins of the secretome were analysed via a BLAST search against the PHI-base. Only 312 of the 1,025 predicted secreted proteins had a significant match in the database with only 40 of them being CSEPs (Table 2.6). This BLAST analysis confirmed the similarity of the *C. zeina* CSEPs g20, g1947 and g3999 to ECP2, AVR4 and ECP6 respectively (Table 2.6).

Table 2.4 The most abundant GO term annotations in the predicted *C. zeina* secretome

Gene Ontology (GO)		
GO ID	GO Term	No. of proteins with GO Term

Molecular function		
GO:0016787	Hydrolase activity	75
GO:0016491	Oxidoreductase activity	47
GO:0004553	Hydrolase activity, hydrolysing O-glucosyl compounds	45
GO:0046872	Metal ion binding	37
GO:0008270	Zinc ion binding	27
GO:0016740	Transferase activity	20
GO:0005524	ATP binding	17
GO:0003677	DNA binding	17
GO:0004601	Peroxidase activity	15
GO:0005509	Calcium ion binding	14
GO:0008061	Chitin binding	14
GO:0004190	Aspartic-type endopeptidase activity	13
GO:0008236	Serine-type peptidase activity	13
GO:0030246	Carbohydrate binding	12
GO:0003824	Catalytic activity	12
GO:0050660	Flavin adenine dinucleotide binding	11
GO:0016614	Oxidoreductase activity, acting on CH-OH group donors	10
GO:0004252	Serine-type endopeptidase activity	10
GO:0008236	Serine-type peptidase activity	10
Biological processes		
GO:0006508	Proteolysis	81
GO:0055114	Oxidation-reduction process	75
GO:0005975	Carbohydrate metabolic process	55
GO:0055085	Transmembrane transport	23
GO:0098869	Cellular oxidant detoxification	18
GO:0008152	Metabolic process	11
GO:0005975	Carbohydrate catabolic process	10
GO:0045491	Xylan catabolic process	10

Effector proteins that localize in the apoplast are usually characterized by a depletion of acidic amino acids, glutamic acid, charged amino acids and contain mostly small amino acids (Sperschneider et al. 2018). By taking into these characteristics into account, the ApoplastP v1.0 program uses a machine learning approach similar to EffectorP to identify effector proteins that localize to the apoplast. A total of 177 *C. zeina* CSEPs were predicted to localize in the apoplast (Table 2.5). Effectors can also localize to the organelles of the host cell by harbouring transit peptides or nuclear localization signals (NLSs) that resemble that of the host cell after the signal peptide (Friesen et al. 2006; Sperschneider et al. 2017). The LOCALIZER v1.0.4 software predicts the transit peptides and NLSs in effector proteins. A total of 27 *C. zeina* CSEPs were predicted to have an NLS, 19 had a chloroplast transit peptide and 9 had a mitochondrion transit peptide (Table 2.5). The *C. zeina* CSEPs homologues of ECP2, ECP6 and AVR4 were predicted to localize in the apoplast.

Table 2.5 Full list of the predicted candidate effector proteins in *C. zeina*

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g20.t1	167	---NA---	PKR94769.1	A	extracellular protein (ECP2)
g43.t1	123	0.917	PKR96327.1	---NA---	hypothetical protein BST61_czeina96g000080
g76.t1	77	0.808	---NA---	A	---NA---
g137.t1	142	0.722	---NA---	N	FK506-binding protein 2
g252.t1	253	---NA---	PKS01155.1	A	CFEM domain-containing protein
g268.t1	152	0.799	PKS01162.1	A	hypothetical protein BST61_czeina19g000180
g291.t1	168	0.638	PKS01180.1	A	antigenic thaumatin-like protein
g342.t1	273	0.673	PKS01218.1	A	glycoside hydrolase family 24 protein
g664.t1	282	---NA---	PKS04211.1	A	Cell wall protein PIR5
g853.t1	179	0.832	PKR98470.1	A; N	hypothetical protein BST61_czeina49g000370
g913.t1	217	---NA---	PKS01340.1	A	hypothetical protein CBER1_01301
g959.t1	133	---NA---	---NA---	A	Guanine-nucleotide dissociation stimulator CDC25
g1066.t1	201	---NA---	PKR96747.1	A	hypothetical protein BST61_czeina84g000180

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g1071.t1	185	0.591	---NA---	C; M; N	---NA---
g1106.t1	340	0.586	PKR98293.1	N	putative mannose 6-phosphate receptor-like protein
g1144.t1	78	0.834	PKR98306.1	A; N	hypothetical protein BST61_czeina52g000520
g1180.t1	191	0.643	PKR99089.1	---NA---	extracellular protein 58-1
g1232.t1	219	---NA---	PKR99061.1	A	glycoside hydrolase family 16 protein
g1358.t1	82	0.773	---NA---	A	---NA---
g1392.t1	190	---NA---	PKS00102.1	A	hypothetical protein BST61_czeina27g000650
g1496.t1	123	0.734	PKR98698.1	---NA---	hypothetical protein BST61_czeina45g000350
g1513.t1	112	0.873	---NA---	A	hypothetical protein CB0940_06777
g1533.t1	131	---NA---	PKR95857.1	A	class II hydrophobin
g1650.t2	444	0.573	PKR96505.1	A	hypothetical protein BST61_czeina90g000060
g1665.t1	131	0.707	PKR96511.1	---NA---	hypothetical protein BST61_czeina90g000120
g1750.t1	222	0.638	PKR94820.1	A	TPA: chitin binding protein, putative (AFU_orthologue; AFUA_1G14430)
g1785.t1	283	---NA---	PKR97028.1	N	hypothetical protein BST61_czeina77g000180
g1790.t1	125	0.731	PKR97050.1	A	predicted protein

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g1802.t1	119	---NA---	PKR97042.1	A	hypothetical protein BST61_czeina77g000320
g1858.t1	100	---NA---	PKS01811.1	A	hypothetical protein BST61_czeina14g000160
g1881.t1	237	0.607	PKS01821.1	A	Endo-1,4-beta-xylanase
g1933.t1	191	0.772	PKS01846.1	A	hypothetical protein BST61_czeina14g000510
g1947.t1	125	0.864	PKS01891.1	A	AVR4 protein
g1988.t1	198	0.783	PKS01942.1	A	hypothetical protein BST61_czeina14g001490
g1994.t1	257	---NA---	---NA---	---NA---	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit
g2004.t1	219	---NA---	PKS01928.1	A	hypothetical protein BST61_czeina14g001340
g2145.t1	228	0.625	PKS02075.1	A	carbohydrate esterase family 5 protein
g2146.t1	225	0.742	PKS02060.1	A	putative endo-1,4-beta-xylanase A
g2157.t1	91	0.557	PKS02064.1	---NA---	hypothetical protein BST61_czeina14g002720
g2173.t1	177	---NA---	PKS02087.1	A	Zip-domain-containing protein
g2197.t1	284	---NA---	PKS02118.1	A	trypsin-domain-containing protein
g2221.t1	331	---NA---	PKS02105.1	A	Endo-1,4-beta-xylanase F3
g2247.t1	229	---NA---	PKR97411.1	A	TRP-domain-containing protein

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g2283.t1	247	0.799	PKS02155.1	A	hypothetical protein BST61_czeina13g000220
g2322.t1	230	---NA---	PKS02187.1	A; C; M	hypothetical protein BST61_czeina13g000540
g2328.t1	154	0.572	PKS02189.1	---NA---	hypothetical protein BST61_czeina13g000560
g2340.t1 + .t2	90	0.836	---NA---	A	---NA---
g2440.t1	272	---NA---	PKR95211.1	---NA---	hypothetical protein BST61_czeina149g000050
g2456.t1	82	0.638	PKR94849.1	---NA---	Surfeit locus protein 1
g2507.t1	243	---NA---	---NA---	C; M	P-loop containing nucleoside triphosphate hydrolase protein
g2516.t1	209	---NA---	---NA---	A	hypothetical protein CBER1_08503
g2532.t1	99	0.882	PKS00517.1	A	hypothetical protein BST61_czeina24g000800
g2614.t3	716	0.586	PKS00470.1	A	carbohydrate-binding module family 18
g2615.t1	243	0.674	PKS00456.1	A	pectate lyase
g2626.t1	172	0.670	PKS00463.1	M	hypothetical protein BST61_czeina24g000240
g2633.t1	169	---NA---	---NA---	---NA---	---NA---
g2634.t1	105	0.927	---NA---	A	hypothetical protein COCMIDRAFT_33555
g2637.t1	203	0.826	PKS00443.1	N	hypothetical protein BST61_czeina24g000040

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g2729.t1	143	0.795	PKR94918.1	---NA---	hypothetical protein BST61_czeina187g000100
g2842.t1	160	0.659	PKR99412.1	---NA---	endoribonuclease l-psp
g2850.t1	195	0.660	PKR99396.1	A	PR-1-like protein
g2856.t1	207	---NA---	PKR94780.1	A	IgE-binding protein
g2860.t1	147	---NA---	---NA---	C; M	---NA---
g2862.t1	103	0.664	PKR94930.1	A	hypothetical protein BST61_czeina184g000030
g2887.t1	267	---NA---	PKS03868.1	A	hypothetical protein BST61_czeina4g000110
g2899.t1	196	---NA---	PKS03909.1	A; C	clock-controlled protein 6
g2965.t1	300	---NA---	PKS03936.1	A	hypothetical protein BST61_czeina4g000800
g3082.t1	191	---NA---	PKS04026.1	A	related to cell wall protein PhiA
g3083.t1	278	---NA---	PKS04054.1	A	Acetyl-CoA synthetase-like protein
g3128.t1	175	---NA---	PKS00741.1	A	hypothetical protein BST61_czeina22g000200
g3134.t1	184	---NA---	---NA---	---NA---	---NA---
g3156.t1	241	0.608	PKS00773.1	A	RmlC-like cupin
g3200.t1	247	0.712	PKS00820.1	A	carbohydrate esterase family 12 protein

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g3276.t1	263	---NA---	PKR96922.1	---NA---	glycoside hydrolase family 128 protein
g3306.t1	82	0.838	---NA---	A	---NA---
g3342.t1	256	0.626	PKR95812.1	---NA---	predicted protein
g3428.t1	172	0.873	PKR96717.1	---NA---	hypothetical protein BST61_czeina85g000250
g3494.t1 + .t2	303	0.603	PKR99163.1	---NA---	Peptidyl-prolyl cis-trans isomerase B
g3553.t1	288	---NA---	---NA---	A	---NA---
g3560.t1	122	0.961	---NA---	---NA---	DNA-directed RNA polymerase I subunit RPA12
g3616.t1	294	---NA---	---NA---	A; C	hypothetical protein DOTSEDRAFT_75987
g3628.t1	170	---NA---	PKS03106.1	---NA---	phosphoglycerate mutase family protein
g3636.t1	225	0.887	PKS03103.1	---NA---	PR-1-like protein
g3640.t1	197	0.873	PKS03101.1	N	hypothetical protein BST61_czeina8g000480
g3680.t1	176	---NA---	---NA---	A	hypothetical protein BST61_czeina89g000060
g3686.t1	201	0.621	PKS03079.1	---NA---	hypothetical protein BST61_czeina8g000260
g3800.t1	66	0.692	PKR98580.1	A	---NA---
g3837.t1	186	0.874	PKR98565.1	---NA---	hypothetical protein BST61_czeina47g000260

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g3843.t1	79	0.777	---NA---	A	---NA---
g3917.t1	213	0.619	PKS01725.1	A	oviduct-specific glyco protein
g3933.t1	197	---NA---	PKS01737.1	A	putative glycerol-3-phosphate dehydrogenase nad-dependent protein
g3938.t1	139	0.663	---NA---	A	lytic polysaccharide monooxygenase
g3999.t1	292	---NA---	PKS95033.1	---NA---	carbohydrate-binding module family 50 protein (ECP6)
g4011.t1	175	---NA---	PKR96099.1	A	GPI anchored CFEM domain protein C
g4018.t1	163	0.808	PKR96084.1	A	cobalt/magnesium transport protein cora
g4070.t1	189	---NA---	PKR97337.1	A	fungal hydrophobin
g4080.t1	146	0.757	PKR97339.1	---NA---	hypothetical protein BST61_czeina69g000210
g4094.t1 + .t2	110	0.758	---NA---	A	related to dehydrogenase/reductase
g4106.t1	233	---NA---	PKR97347.1	---NA---	Lactobacillus up-regulated protein
g4143.t1	147	0.943	---NA---	---NA---	hypothetical protein CB0940_10933
g4149.t1	170	0.893	PKR97066.1	A; C	cobalt/magnesium transport protein cora
g4170.t1	147	0.925	---NA---	---NA---	hypothetical protein CBER1_08721
g4178.t1	186	---NA---	PKR97210.1	A	hypothetical protein BST61_czeina73g000360

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g4226.t1	87	0.832	PKR97318.1	A	hypothetical protein BST61_czeina70g000340
g4247.t1	168	0.749	PKR98359.1	A	hypothetical protein BST61_czeina51g000490
g4278.t1	149	0.553	PKR98337.1	A	Secreted protein NIS1
g4282.t1	226	---NA---	PKR98371.1	A	extracellular protein 20-2
g4360.t1	148	0.718	---NA---	A	extracellular protein 20-2
g4466.t1	119	0.879	PKR97377.1	A	hypothetical protein BST61_czeina68g000160
g4497.t1	154	0.860	PKR97391.1	N	hypothetical protein BST61_czeina68g000300
g4499.t1	299	---NA---	PKR97920.1	---NA---	hypothetical protein BST61_czeina58g000790
g4500.t1	171	0.860	PKR97895.1	A	hypothetical protein BST61_czeina58g000540
g4568.t1	112	0.922	---NA---	A	---NA---
g4569.t1	160	0.880	PKR97896.1	A; C	hypothetical protein BST61_czeina58g000550
g4599.t1	252	---NA---	---NA---	---NA---	2OG-Fe(II) oxygenase
g4652.t1	292	0.601	PKR97827.1	A	Dihydrolipoyl dehydrogenase
g4678.t1	106	0.618	PKR99485.1	A	fungal hydrophobin
g4840.t1	219	---NA---	PKS02590.1	A	hypothetical protein BST61_czeina10g000140

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g4914.t1	320	0.630	PKS02679.1	A; N	carbohydrate esterase family 1 protein
g4972.t1	245	0.660	PKS02698.1	A; C	extracellular protein 60-1
g4976.t1	255	---NA---	PKS02699.1	N	hypothetical protein CBER1_02601
g5083.t2	93	0.555	---NA---	---NA---	---NA---
g5094.t1	214	0.672	PKS03268.1	---NA---	Endoplasmic reticulum vesicle protein 25
g5106.t1	209	0.749	PKS03289.1	---NA---	carbohydrate esterase family 4 protein
g5217.t1	158	0.641	PKS03405.1	A	peptidoglycan-binding domain 1 protein
g5395.t1	187	0.787	PKS01566.1	A; C	Micronemal protein 4
g5579.t1	80	0.639	PKR96388.1	A	---NA---
g5616.t1	134	0.818	---NA---	A	hypothetical protein BST61_czeina94g000140
g5628.t1	283	0.627	PKR98959.1	C	deoxyribonuclease NucA/NucB-domain-containing protein
g5701.t1	159	0.863	PKR95445.1	A	hypothetical protein BST61_czeina134g000100
g5712.t1	99	0.870	PKR95522.1	---NA---	putative zinc finger and scan domain-containing protein 18
g5745.t1	98	0.878	---NA---	A	---NA---
g5829.t1	212	0.709	PKS02497.1	---NA---	hypothetical protein BST61_czeina11g000740

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g5841.t1	278	---NA---	PKS02476.1	---NA---	putative endonuclease exonuclease phosphatase family protein
g5844.t1	225	0.602	PKS02474.1	N	putative O-acetyltransferase CAS1
g5879.t1	251	---NA---	PKS02461.1	A	guanine deaminase
g5885.t1	227	0.712	---NA---	N	MF(ALPHA)1
g5918.t1	175	0.874	---NA---	---NA---	hypothetical protein CB0940_04299
g5953.t1	276	0.686	PKS02785.1	---NA---	carbonic anhydrase
g6053.t1	72	0.783	PKS02865.1	A	Transcription elongation factor S-II
g6073.t1	175	0.704	PKS02872.1	---NA---	Phosphatidylglycerol/phosphatidylinositol transfer protein
g6083.t1	257	0.577	---NA---	N	Lactobacillus up-regulated protein
g6097.t1	156	0.627	---NA---	A	Guanyl-specific ribonuclease F1
g6145.t1	269	---NA---	PKS02929.1	A	hypothetical protein BST61_czeina9g001910
g6160.t2	437	---NA---	PKS02969.1	N	Serine/threonine-protein kinase ppk13
g6172.t1	223	---NA---	PKS02954.1	A; C	hypothetical protein BST61_czeina9g002160
g6200.t1	212	---NA---	PKS02987.1	A	siderophore biosynthesis enzyme
g6230.t1	225	0.791	PKS03031.1	A	cutinase 1

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g6282.t1	164	0.643	PKR95395.1	A	hypothetical protein BST61_czeina136g000060
g6283.t1	216	0.807	---NA---	N	hypothetical protein CBER1_09316
g6284.t1	277	---NA---	---NA---	A; C	lipase 3-like protein
g6365.t1	224	---NA---	PKR98394.1	A	putative cdp-alcohol phosphatidyltransferase protein
g6457.t1	190	---NA---	PKR98891.1	A	hypothetical protein BST61_czeina42g000370
g6465.t1	220	0.735	PKR98858.1	---NA---	Protein ERP1
g6477.t1	215	---NA---	PKR94710.1	A	extracellular protein 51
g6478.t1	133	0.624	PKR96363.1	---NA---	hypothetical protein BST61_czeina95g000140
g6532.t1	138	0.933	PKS02302.1	A; N	hypothetical protein BST61_czeina12g000240
g6576.t1	283	---NA---	PKS02339.1	A	expansin-related protein
g6601.t1	179	---NA---	PKS02372.1	A	GPI anchored serine-threonine rich protein
g6746.t1	299	---NA---	PKS01481.1	M; N	hypothetical protein BST61_czeina17g000760
g6779.t1	187	---NA---	PKS01449.1	A	hypothetical protein BST61_czeina17g000440
g6793.t1	177	---NA---	PKS01459.1	A	hypothetical protein BST61_czeina17g000540
g6819.t1	94	0.608	---NA---	A	Extracellular membrane protein, CFEM domain protein

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g6858.t1	167	---NA---	PKR95125.1	C; M; N	putative 54S ribosomal protein L32, mitochondrial
g6863.t1	77	0.761	---NA---	---NA---	---NA---
g6987.t1	119	---NA---	PKR99799.1	A	hypothetical protein BST61_czeina30g000210
g7017.t1	276	---NA---	PKR99841.1	A; C; N	hypothetical protein BST61_czeina30g000640
g7058.t1	176	0.742	PKR99857.1	A	starch-binding domain-like protein
g7142.t1	279	---NA---	PKR98041.1	---NA---	Extracellular metalloprotease
g7231.t1	185	0.666	PKR97792.1	---NA---	predicted protein
g7302.t1	196	0.747	PKR97720.1	A	hydrophobin-like protein
g7315.t1	81	0.696	---NA---	---NA---	---NA---
g7336.t1	72	0.575	---NA---	A	---NA---
g7351.t1	148	0.742	PKR97578.1	---NA---	UPF0357 protein
g7398.t1	89	0.953	---NA---	A; N	---NA---
g7399.t1	75	0.787	PKR95101.1	A	hypothetical protein BST61_czeina159g000010
g7432.t1	298	0.574	PKR98735.1	A	carbohydrate esterase family 16 protein
g7440.t1	117	0.853	PKR98737.1	---NA---	Long chronological lifespan protein 2

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g7524.t1	73	0.672	---NA---	A	hypothetical protein
g7526.t1	95	0.661	---NA---	A	---NA---
g7527.t1	195	---NA---	PKR97699.1	A	hypothetical protein BST61_czeina61g000480
g7577.t1	147	0.671	PKS00002.1	A	small secreted protein
g7710.t1	213	0.835	PKS00979.1	---NA---	putative membrane protein C17A5.08
g7743.t1	297	---NA---	PKS00918.1	A	glycoside hydrolase family 16 protein
g7791.t1	209	0.802	PKS00884.1	A	hypothetical protein BST61_czeina21g000480
g7857.t1	155	0.670	PKR96298.1	N	hypothetical protein BST61_czeina97g000160
g7864.t1	111	0.596	PKR96301.1	---NA---	hypothetical protein BST61_czeina97g000190
g7868.t1	237	0.767	PKR96304.1	A	D-alanyl-D-alanine carboxypeptidase
g7916.t1 + .t2	158	0.666	PKR98018.1	---NA---	protein disulfide-isomerase domain-containing protein
g7956.t1	183	0.624	PKR95942.1	---NA---	hypothetical protein BST61_czeina112g000020
g8010.t1	211	---NA---	PKR95247.1	---NA---	hypothetical protein BST61_czeina147g000010
g8029.t1	141	---NA---	PKR95479.1	A	extracellular serine-threonine rich protein
g8091.t1	104	0.752	PKR95383.1	A	hypothetical protein BST61_czeina137g000100

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g8133.t1	134	0.862	PKR94712.1	A	hypothetical protein BST61_czeina269g000010
g8176.t1	90	0.776	---NA---	A	---NA---
g8234.t1	75	0.919	---NA---	---NA---	---NA---
g8235.t1	210	---NA---	---NA---	---NA---	DUF3328 domain containing protein
g8252.t1	173	---NA---	PKR97125.1	---NA---	hypothetical protein BST61_czeina75g000370
g8258.t1	85	0.805	---NA---	A	hypothetical protein BST61_czeina70g000340
g8290.t1	220	0.639	PKR97092.1	---NA---	hypothetical protein BST61_czeina75g000040
g8291.t1	321	0.649	PKR97091.1	---NA---	DUF1183-domain-containing protein
g8304.t1	297	---NA---	---NA---	N	Heat shock protein 70
g8310.t1	105	0.771	PKR96858.1	A	hypothetical protein BST61_czeina81g000180
g8311.t1	270	---NA---	---NA---	M; N	---NA---
g8359.t1	82	0.932	---NA---	A	---NA---
g8370.t1	139	0.760	PKR97633.1	A	Guanyl-specific ribonuclease F1
g8382.t1	247	---NA---	PKR97637.1	A; C	hypothetical protein BST61_czeina62g000340
g8412.t1	80	0.897	---NA---	A	---NA---

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g8456.t1	226	---NA---	PKR99013.1	A	Cell differentiation protein rcd1
g8469.t1	166	0.765	PKR99008.1	A	lipase 3-like protein
g8485.t1	197	0.712	PKR94693.1	C	hypothetical protein BST61_czeina311g000010
g8504.t1	234	0.662	PKR96179.1	---NA---	hypothetical protein BST61_czeina101g000120
g8538.t1	129	0.945	---NA---	---NA---	---NA---
g8621.t1	166	0.657	PKR97282.1	A	extracellular protein 52
g8673.t1	221	---NA---	PKR96604.1	A	related to Probable cutinase 1
g8766.t1	219	---NA---	PKS04643.1	A	hypothetical protein BST61_czeina1g000690
g8839.t1	148	0.711	PKS04723.1	A	small secreted protein
g8846.t1	252	---NA---	PKS04744.1	A	hypothetical protein BST61_czeina1g001720
g8864.t1	161	0.555	PKS04736.1	---NA---	major allergen alt a1
g8995.t2	457	0.872	---NA---	---NA---	WSC domain containing protein
g9002.t1	225	0.633	PKR99640.1	A	lipocalin-like domain-containing protein
g9115.t1	81	0.632	---NA---	A	---NA---
g9245.t1	113	0.754	---NA---	---NA---	hypothetical protein CB0940_08017

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g9259.t1	203	0.630	PKR99257.1	A	hypothetical protein BST61_czeina37g000470
g9260.t1	228	---NA---	---NA---	A	hypothetical protein CB0940_08031
g9310.t1	113	0.862	---NA---	A	hypothetical protein CB0940_08044
g9407.t1	229	---NA---	PKS03549.1	A	putative cutinase 1
g9529.t1	297	---NA---	PKS03671.1	---NA---	related to endo alpha-1,4 polygalactosaminidase precusor
g9637.t1	94	0.664	PKS00342.1	A	hypothetical protein BST61_czeina25g001000
g9737.t1	77	0.835	---NA---	A	---NA---
g9797.t1	116	0.954	---NA---	A	---NA---
g9800.t1	98	0.948	PKS03731.1	A	hypothetical protein BST61_czeina5g000610
g9813.t1	125	0.792	---NA---	A	hypothetical protein CB0940_02034
g9883.t1	193	0.647	PKS03806.1	A	Mitochondrial carrier protein RIM2
g9941.t1	151	0.669	PKS03837.1	A	hydrolytic enzyme protein
g10088.t1	95	0.802	---NA---	A	uncharacterized protein RCC_07522
g10111.t1	74	0.743	---NA---	A	extracellular protein 36-1
g10124.t1	252	0.608	PKS00712.1	A	Superoxide dismutase [Mn]

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g10132.t1	178	0.627	PKS00716.1	A	ricin B lectin
g10152.t1	220	0.862	PKR96561.1	---NA---	Peptidase inhibitor 16
g10161.t1	181	---NA---	PKR96541.1	A	hypothetical protein BST61_czeina89g000060
g10264.t1	169	0.616	PKR95914.1	C; M; N	hypothetical protein BST61_czeina114g000230
g10271.t1	189	0.839	PKR95900.1	A	extracellular protein 32-2
g10285.t1	112	0.915	PKR95906.1	A	hypothetical protein BST61_czeina114g000150
g10327.t1	151	0.640	---NA---	C	---NA---
g10386.t1	149	0.925	PKS04545.1	---NA---	hypothetical protein BST61_czeina2g002430
g10398.t1	245	---NA---	---NA---	A	pgap1-like protein
g10403.t1	82	0.704	---NA---	---NA---	---NA---
g10483.t1	185	0.732	---NA---	A	---NA---
g10547.t1	91	0.852	---NA---	A	---NA---
g10572.t1	288	0.631	PKS04407.1	N	deoxyribonuclease nucA/NucB domain-containing protein
g10576.t1	159	---NA---	PKS04384.1	A	extracellular protein 20-2
g10610.t1	195	0.593	PKS04370.1	---NA---	hypothetical protein BST61_czeina2g000670

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g10625.t1	229	0.599	PKS04341.1	A	carbohydrate esterase family 5 protein
g10654.t1	179	0.790	PKS04330.1	A	hypothetical protein BST61_czeina2g000270
g10823.t1	222	0.923	PKR97948.1	A	PR-1-like protein
g10862.t1	80	0.942	---NA---	A	---NA---
g10955.t1	218	0.864	PKR98151.1	---NA---	hypothetical protein BST61_czeina53g000160
g11003.t1	195	---NA---	PKR95358.1	A	carbohydrate-binding module family 63 protein
g11035.t1	289	---NA---	PKS01029.1	---NA---	Aprataxin-like protein
g11042.t1	266	---NA---	PKS01019.1	A	related to malate dehydrogenase
g11101.t1	182	0.873	PKS01112.1	A	hypothetical protein CB0940_03271
g11151.t1	134	0.948	PKS01135.1	A	hypothetical protein BST61_czeina20g001320
g11222.t1	176	---NA---	---NA---	A	hypothetical protein CBER1_00513
g11248.t1	214	---NA---	PKR96481.1	A	hypothetical protein BST61_czeina91g000080
g11258.t1	214	---NA---	---NA---	A; C	spore coat protein SP65-like
g11360.t1	275	---NA---	PKR99922.1	A	Extracellular metalloprotease
g11377.t1	263	---NA---	---NA---	N	hypothetical protein CB0940_03796

Protein IDs[a]	Protein length (AA)	EffectorP score	Corresponding Genbank Protein IDs[b]	Localization[c]	PacBio Blast2GO annotation[d]
g11455.t1	88	0.952	---NA---	A	hypothetical protein SETTUDRAFT_39539
g11466.t1	120	---NA---	PKS03170.1	A	hypothetical protein BST61_czeina8g001170
g11484.t1	136	0.712	---NA---	---NA---	riboflavin aldehyde-forming enzyme
g11514.t1	85	0.777	---NA---	A	---NA---
g11524.t1	284	---NA---	PKR95206.1	---NA---	A-agglutinin anchorage subunit-like
g11558.t1	219	---NA---	---NA---	A	hypothetical protein DOTSEDRAFT_33504

[a] The suffix ".tX" refers to the predicted gene variant.

[b] Protein IDs based on the annotation from the *C. zeina* Illumina assembly (Wingfield et al. 2017).

[c] Abbreviations: A – Apoplast. C – Chloroplast. M – Mitochondrion. N – Nucleus.

[d] Annotations with format BST61_czeinaXXgXXXXXX refer to the gene model from the *C. zeina* Illumina assembly (Wingfield et al. 2017).

Recently, a new hypothesis developed that states that at least one type of tandem repeat domain in non-orthologous effectors has emerged from convergent evolution for plant pathogenicity (Chang et al. 2018). Using the pipeline developed to identify these domains, only three CSEPs g664, g4070 and g10161 were predicted to each harbour different types of these domains. The repeats occurred 2.17, 2.86 and 3.25 times in g664, g4070 and g10161 respectively (Appendix Table 4). A BLASTP search of the tandem repeat for each of the three proteins was performed. For g664, the tandem repeat was identified with indels in predicted cell wall proteins of related *Cercospora* species as well as *Ramularia collo-cygni*. The only virulent protein it had a similarity to was HSP150P from *Saccharomyces cerevisiae* which has been shown to increase virulence. However, the tandem repeat also had indels. For g4070, the tandem repeat was found in hypothetical proteins of the Dothideomycete *Venturia inaequalis* as well as *Fusarium nygamai.* The tandem repeat had no match in any functionally annotated effectors. The tandem repeat of g10161 had no significant matches.

Table 2.6 Annotated PHI-base proteins similar to *C. zeina* CSEPs

Gene IDs	PHI Accession	Target gene name	Species	Phenotype	E-value
g20	PHI:71, PHI:5486, PHI:5544, PHI:5559	ECP2	*Passalora fulva*	Reduced virulence	5.16e-53
g137	PHI:4584	FkpA	*Cronobacter universalis*	Reduced virulence	8.18e-19
g252	PHI:4015	CfmA	*Aspergillus fumigatus*	Unaffected pathogenicity	3.2e-09
g664	PHI:5043	Hsp150p	*Saccharomyces cerevisiae*	Increased virulence	6.27e-34
g1232	PHI:6265	Eng1	*Histoplasma capsulatum*	Reduced virulence	2.50e-34

Gene IDs	PHI Accession	Target gene name	Species	Phenotype	E-value
g1533	PHI:291	CPPH1	*Claviceps purpurea*	Unaffected pathogenicity	1.25e-10
g1881	PHI:2214	Endo-1_4-beta-xylanase I (GH10 family)	*Magnaporthe oryzae*	Reduced virulence	1.22e-84
g1947	PHI:5475	CbAvr4	*Cercospora beticola*	Plant avirulence determinant	9.67e-45
g2145	PHI:69	CUTA	*Botrytis cinerea*	Unaffected pathogenicity	1.74e-07
g2146	PHI:572	XYL2	*Bipolaris zeicola*	Unaffected pathogenicity	1.34e-105
g2197	PHI:652	GIP1	*Phytophthora sojae*	Plant avirulence determinant	3.17e-16
g2221	PHI:2043	XYL-6	*Magnaporthe oryzae*	Unaffected pathogenicity	1.25e-140
g2247	PHI:7693	FlcB	*Aspergillus fumigatus*	Loss of pathogenicity	4.01e-11
g2614	PHI:352	GLO1	*Ustilago maydis*	Loss of pathogenicity	8.57e-60
g2615	PHI:4845	PL1332	*Alternaria brassicicola*	Reduced virulence	4.92e-85
g2626	PHI:810	MGG 02436	*Magnaporthe oryzae*	Reduced virulence	7.32e-24
g2729	PHI:2144	MoAAT	*Magnaporthe oryzae*	Reduced virulence	6.94e-07
g2850	PHI:7144	FvSCP1	*Fusarium verticillioides*	Reduced virulence	1.58e-35

Gene IDs	PHI Accession	Target gene name	Species	Phenotype	E-value
g2856	PHI:2901	BEC1040	*Blumeria graminis*	Reduced virulence	6.33e-11
g3494	PHI:7181	cypB	*Beauveria bassiana*	Reduced virulence	1.94e-109
g3560	PHI:1627	GzTF2S002	*Fusarium graminearum*	Unaffected pathogenicity	2.99e-07
g3636	PHI:7144	FvSCP1	*Fusarium verticillioides*	Reduced virulence	1.36e-19
g3999	PHI:5495, PHI:5543, PHI:5576	Ecp6	*Passalora fulva*	Plant avirulence determinant	3.29e-78
g4011	PHI:4017	CfmC	*Aspergillus fumigatus*	Unaffected pathogenicity	1.66-13
g4070	PHI:458	MHP1	*Magnaporthe oryzae*	Reduced virulence	1.16e-12
g4678	PHI:458	MHP1	*Magnaporthe oryzae*	Reduced virulence	4.91e-14
g4914	PHI:2452	FAED1	*Fusarium graminearum*	Unaffected pathogenicity	3.80e-12
g5106	PHI:8589	PdaA1 (CD630 14300)	*Clostridioides difficile*	Reduced virulence	1.28e-12
g5841	PHI:6473	Rv0888	*Mycolicibacterium smegmatis*	Reduced virulence	2.48e-40
g5879	PHI:7751	Rv3114	*Mycobacterium tuberculosis*	Reduced virulence	5.42e-11
g6160	PHI:1248	FGSG 06832	*Fusarium graminearum*	Unaffected pathogenicity	7.14e-172
g6230	PHI:2383	MfCUT1	*Monilinia fructicola*	Increased virulence	5.14e-06
g6819	PHI:1093	FGSG 02077	*Fusarium*	Reduced	7.31e-11

Gene IDs	PHI Accession	Target gene name	Species	Phenotype	E-value
			graminearum	virulence	
g7142	PHI:479	MEP1	*Coccidioides posadasii*	Reduced virulence	2.44e-71
g8673	PHI:2849	cutA	*Fusarium solani*	Unaffected pathogenicity	7.21e-50
g9407	PHI:2849	cutA	*Fusarium solani*	Unaffected pathogenicity	7.61e-60
g10124	PHI:420	SOD2	*Cryptococcus gattii vgiii*	Loss of pathogenicity	1.46e-47
g10152	PHI:7144	FvSCP1	*Fusarium verticillioides*	Reduced virulence	2.96e-32
g10271	PHI:2901	BEC1040	*Blumeria graminis*	Reduced virulence	5.24e-08
g10823	PHI:7144	FvSCP1	*Fusarium verticillioides*	Reduced virulence	2.12e-35
g11360	PHI:479	MEP1	*Coccidioides posadasii*	Reduced virulence	2.54e-78
g11524	PHI:7664	LysM1	*Penicillium expansum*	Plant avirulence determinant	7.11e-42

2.4.4 Differential gene expression in upper vs lower leaves

The GLS disease lesions usually appear in the lower leaves of the maize plant first before favourable conditions allows the spores to spread to the upper leaves and cause more lesions. We tested whether some CSEPs were differentially expressed between the lower and upper leaves from *C. zeina* inoculated B73 maize leaves 77 DAP in Hildesheim farm, Greytown in KwaZulu-Natal province in South Africa in March 2013. This could indicate which CSEPs are expressed during early infection and which are expressed during the necrotrophic stages. The RNA-Seq reads were mapped on to both the B73 RefGen_v4 maize and *C. zeina* genomes at the same time. Genes that

had zero counts across all samples were removed from further analyses. Differential gene expression analysis was performed using DESeq2. The MA plot shows the proportion of maize and *C. zeina* genes that were differentially expressed (Fig. 2.2).

A total of 10,227 *C. zeina* genes, which included 203 of the 274 CSEP genes, had significantly higher transcript levels in the lower leaves compared to the upper leaves (adjusted p-value < 0.05). A limitation of this analysis was that it is unable to give information on whether a *C. zeina* gene is induced *in planta*, since there was a different *C. zeina* cell count in upper and lower leaves. This was quantified in the Methods S1 file from (Christie et al. 2017). The quantification via a qPCR assay showed that the lower leaves had approximately 6 ng of *C. zeina* gDNA per mg maize gDNA which was significantly higher than the approximate 0.1 ng gDNA per mg maize gDNA in the upper leaves (Christie et al. 2017). Therefore, differences in transcript levels between upper and lower leaves may just reflect differences in *C. zeina* content. Consistent with this, the amount of *C. zeina* reads in the upper and lower leaves was 54,677 and 3,302,537, respectively. This very large difference indicates that transcript numbers for many *C. zeina* genes in upper leaves was potentially too low to give an accurate read count. The difference in fungal content may also explain why so many (10,227) of the 11,570 predicted protein-coding genes of *C. zeina*, had significantly higher transcript levels in lower leaves. In addition, the DESeq2 analysis method did not discard genes that had low or undetected read counts in the upper leaf sample but countable reads in the lower leaf sample, thus contributing to the large number of genes reported.

However, the data was useful to identify *C. zeina* genes that were expressed *in planta*, and most of the counts came from the lower leaves where clear GLS lesions were visible as seen in Figure 1 of the Methods S1 file from (Christie et al. 2017). A total of 5,641 *C. zeina* genes were expressed above a threshold of 18.27 mean value of normalized counts (for *C. zeina* genes only) of both upper and lower leaves, and there were 112 CSEP genes in this category (Data not shown). Additionally, a total of 1,129 *C. zeina* genes had a normalized counts value above the 90th percentile with 47 CSEP genes among them (Appendix Table 5).

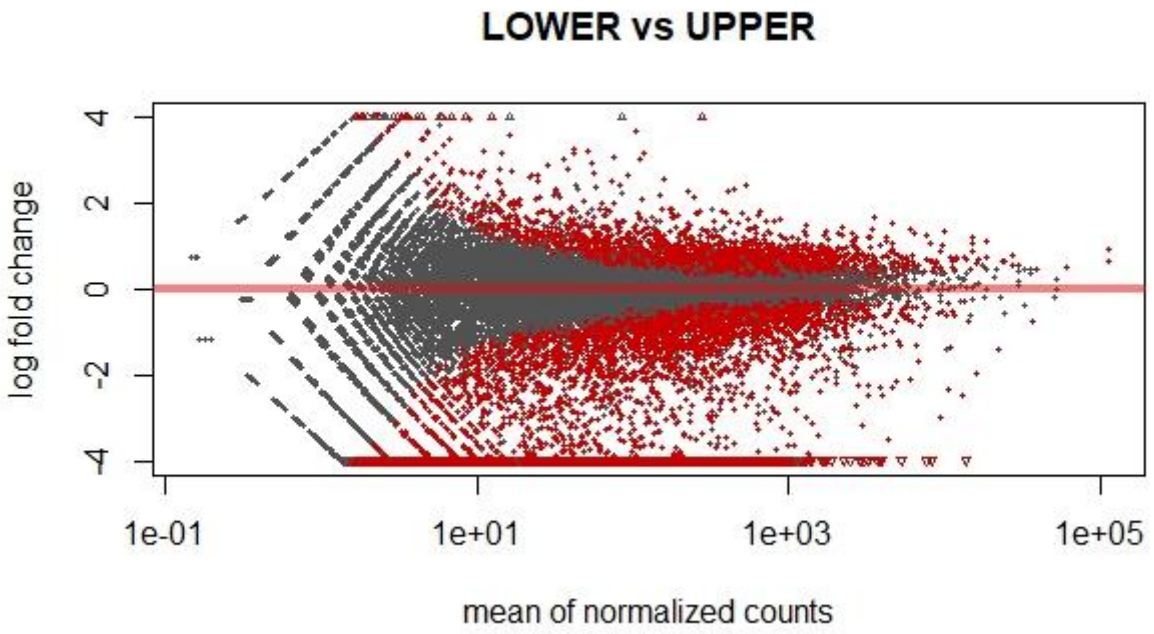


Fig. 2.2 MA plot representing the differentially expressed maize and fungal genes between the upper and lower leaves of B73 maize plants infected with *C. zeina*. Red dots represent genes with significantly different transcript levels between the leaf samples (FDR, 0.05). A total of 62% of the genes with significantly different transcript levels were *C. zeina* genes and 38% were maize genes.

2.4.5 Identification and annotation of transposable elements (TEs)

Along with protein-coding genes, TEs are one of the elements that provide information about an organism's genome. It is therefore important to identify and annotate TEs. TEs are divided into two classes, namely retrotransposons (Class I) and DNA transposons (Class II). Retrotransposons are generally transposed via an RNA intermediate whilst DNA transposons are transposed via a DNA intermediate (Mat Razali et al. 2019). Retrotransposons are further divided into long terminal repeats (LTRs) and non-LTRs which include long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). A proportion of 39% of the *C. zeina* genome consisted of repetitive elements. This is more than that of other Dothideomycetes like *Leptosphaeria maculans* (34%), *Z. tritici* (18%) and other fungi like *Neurospora crassa* (16%). However, it is not as much as the expanded genomes of *C. fulvum* (48%) and *P. fijiensis* (53%) (Fig. 2.3). The majority of the TEs in *C. zeina* consisted of LTR elements, then non-LTR elements, followed by DNA transposons and then unclassified elements. When compared to other fungi, we see a similar pattern of LTR elements being the most abundant (Fig. 2.3). Additionally, Capnodiales have more non-LTR

retrotransposons. The class "Other" includes potential genes, SSRs and ribosomal DNA and other unclassified TEs.

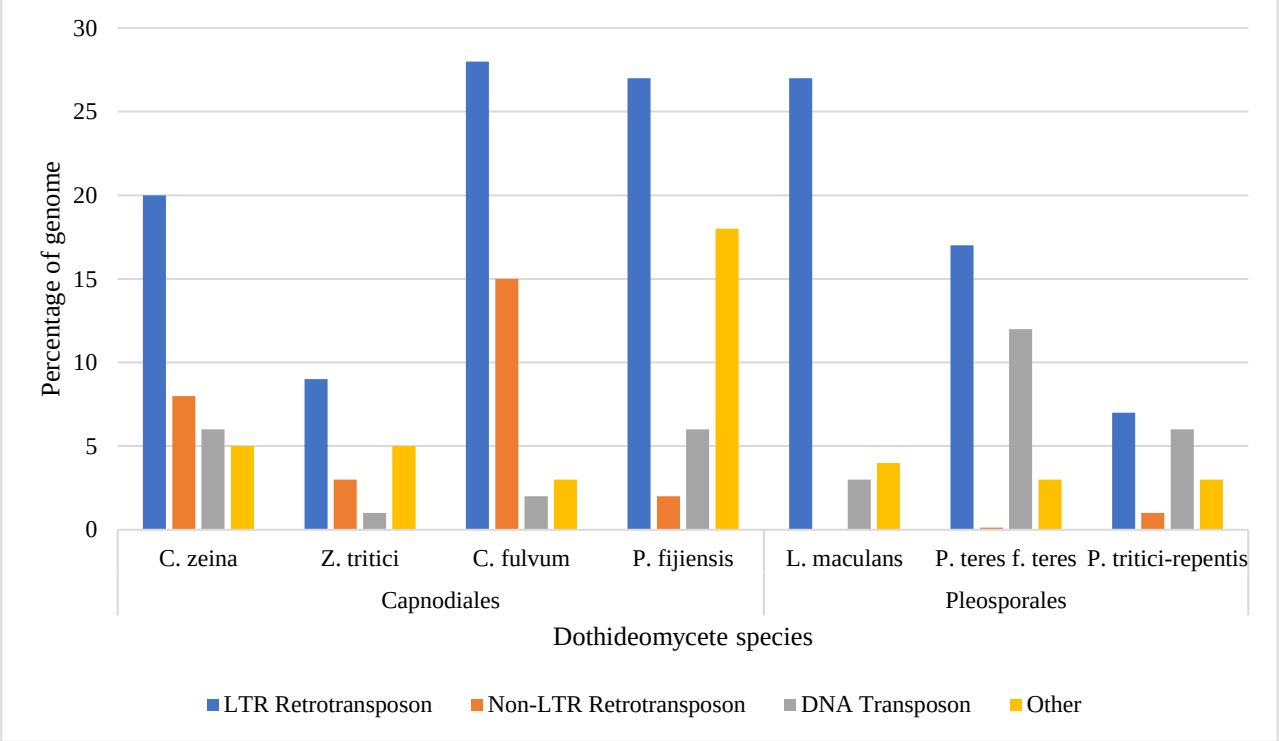


Fig. 2.3 Comparison of repeat classes among Dothideomycete genomes. Capnodiales: *Cercospora zeina*, *Zymoseptora tritici*, *Cladosporium fulvum* and *Pseudocercospora fijiensis*. Pleosporales: *Leptosphaeria maculans*, *Pyrenophora teres* f. *teres* and *Pyrenophora tritici-repentis*. Data collected from (Arango Isaza et al. 2016; Dal Molin et al. 2018).

Many fungi prevent the expansion of TEs in the genome by silencing them with a process known as repeat-induced point mutations (RIP) (Galagan and Selker 2004). These mutations usually involve the C-to-T transition mutations after the cytosine residues were methylated during sexual reproduction (Galagan and Selker 2004). Profiling the genome with The RIPper revealed that 33.31% of the genome was affected by RIP based on the RIP product index value TpA/ApT: $0.8 < x \geq 1.1$ (Appendix Fig. 3). This is almost double the proportion of RIP affected areas in the *N. crassa* genome (van Wyk et al. 2019), an organism where RIP is extensively studied. A total of 301 large RIP-affected regions (LRARs) were identified with an average size of 44,337 bp (Table 2.7). The default size of LRARs is 4,000 bp (van Wyk et al. 2019). The average GC content of these LRARs was only 36.03%. RIP analysis of different TE families was performed using RIPCAL to identify the abundance of dinucleotides associated with RIP in *C. zeina*. The repeat elements were firstly grouped according to their families and were then aligned. RIPCAL then analysed the base

changes from the sequence that had the highest GC content for each family. Analysis of three *Gypsy* LTR families, three *Copia* LTR families and three *Tad1* LINE non-LTR families revealed that there was an abundance of CT-to-TT mutations (Fig. 2.4) rather than the CA-to-TA mutations that are common in other fungi like *N. crassa* and *P. nodorum* (Galagan and Selker 2004; Rouxel et al. 2011) (Appendix Fig. 2). However, this CT-to-TT bias has also been reported in other Dothideomycetes including *Ascochyta rabiei* (Verma et al. 2016) and *P. fijiensis* (Arango Isaza et al. 2016). To confirm that the RIP pathway is functioning in *C. zeina*, we searched for proteins that could be involved in the RIP process. Three *C. zeina* proteins were identified. The proteins g2999, g9350 and g774 had 41%, 37% and 35% similarity to the proteins DIM2, RIPD and a Histone-lysine N-methyltransferase from *N. crassa* respectively. These *N. crassa* proteins are known to be involved in RIP. The Blast2GO annotations of the *C. zeina* proteins were SAL-methionine-dependent methyltransferase for both g2999 and g9350 and SET domain-containing protein for g774.

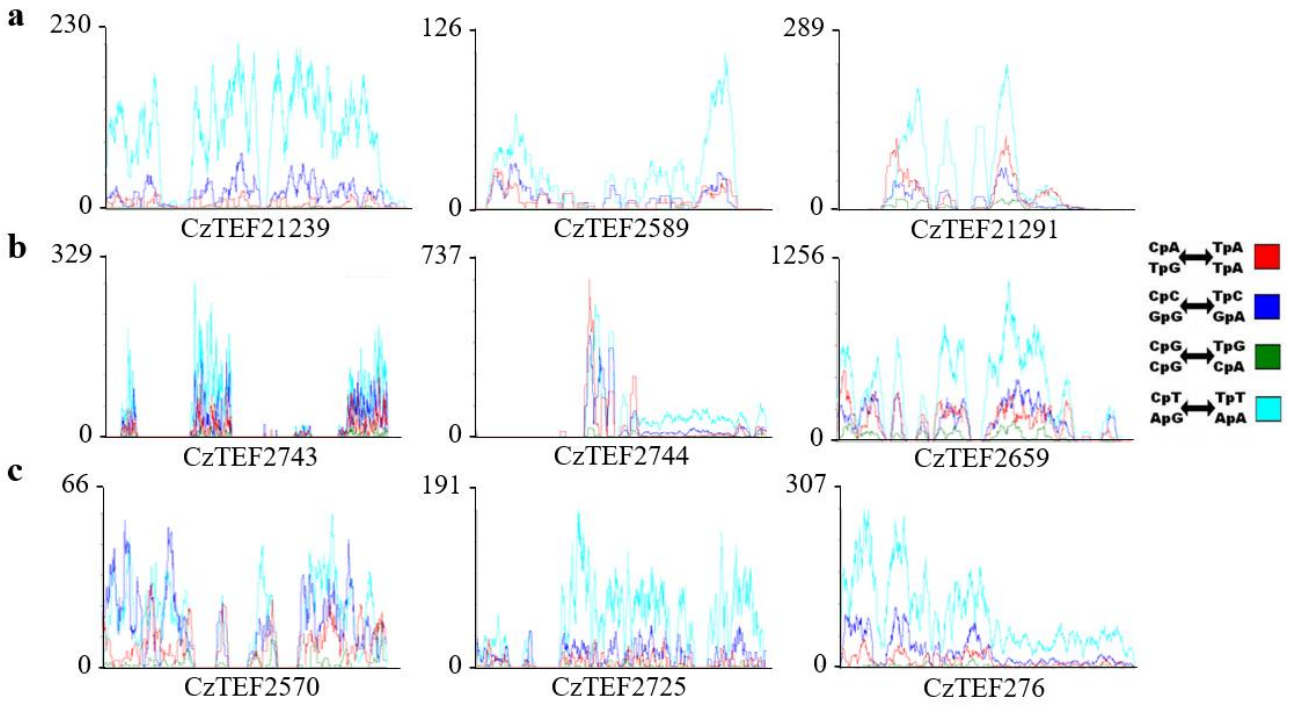


Fig. 2.4 Repeat-induced point mutation (RIP) dinucleotide bias in selected *Cercospora zeina* transposable element families. **a** Three randomly selected *Copia* LTR families with a total of 21, 10 and 32 sequences (from left to right). **b** Three randomly selected *Gypsy* LTR families with a total of 95, 228 and 158 sequences (from left to right). **c** Three randomly selected *Tad1* LINE families with a total 16, 37 and 18 sequences (from left to right). The y-axis illustrates the RIP mutation

frequency along the length of the TE alignment (x-axis). The CzTEFXXX refers to the family names.

Table 2.7 The RIP profile of the *C. zeina* genome obtained from The RIPper

Parameter	
Genome Size	41,717,714
Count of Genomic Windows Investigated	83,435
GC Content of Genome Assembly (%)	47.96
Number of RIP Affected Windows	27,794
RIP Affected Genomic Proportion	33.31
Count of LRARs	301
Average Size of LRARs (bp)	44,336.54
Average GC Content of LRARs (%)	36.03
Genomic Proportion of LRARs (bp)	13,345,300
Product Value for LRARs	1.97
Substrate Value for LRARs	0.11
Composite Value for LRARs	1.86

2.4.6 The genome architecture of *C. zeina*

Many plant pathogenic fungi harbour a “two-speed” genome that consists of a rapidly evolving repeat-rich region and slowly evolving gene-rich region (Dong et al. 2015). Some effector genes have been reported to reside in the rapidly evolving regions of the genome. The OcculterCut program was used to determine if the *C. zeina* genome had a bipartite structure. The genome segmentation process of OcculterCut involves the movement of a border along the genome sequence after which the Jenson-Shannon divergence is calculated at each point (Testa et al. 2016). The segmentation is confirmed at the position where the Jenson-Shannon divergence is maximized and whether the differences between the GC contents of the two adjacent segments are statistically significant (Testa et al. 2016). The minimum segment length is 1 kbp.

OcculterCut analysis in *C. zeina* identified that the genome could be divided up into compartments with different GC content. The first had high GC content (41-100%) and the second with low GC content (0-41%). The low GC content consisted of 33.2% of the genome, which is consistent with

the 33.31% of the genome that was RIP affected (Fig. 2.5). However, only 19 genes (1.08 genes per Mbp) were identified in these AT-rich compartments. This is lower than that of other Dothideomycetes like *Pyrenophora teres* f. *teres* (72.6 genes per Mbp) (Wyatt et al. 2018) and different *Leptosphaeria* species which had a range of 75 to 233 genes in AT-rich regions (Dutreux et al. 2018). There were only two CSEP genes within the set of 19 genes. The two genes are g1513 and g5616 which encode hypothetical proteins that localize in the apoplast and have EffectorP values of 0.873 and 0.818 respectively.

A heatmap showing the density plot of intergenic distances of *C. zeina* genes showed that many effector (CSEP) genes grouped with the majority of genes (Fig. 2.6a) in contrast to the evident separation of RXLR effectors and core genes in *Phytophthora infestans* (Fig. 2.6b). CSEPs were located closer on average to TEs than 274 randomly chosen BUSCO genes (30,637 bp vs 36,895 bp) (Fig. 2.7 and 2.8). However, the Wilcoxon signed rank test showed that this difference was not significant with a p-value of 0.06134.

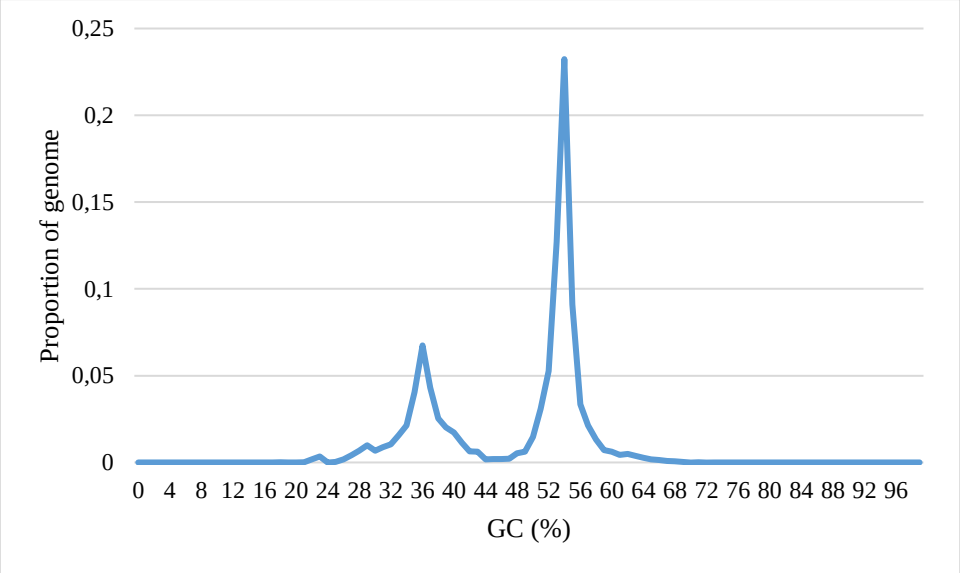


Fig. 2.5 Distribution of GC content in the *C. zeina* genome based on OcculterCut v1.1 analysis. The percentage of GC content is illustrated on the x-axis and the proportion of the genome is shown on the y-axis. A bimodal distribution is observed in which genome compartments fall into the high GC% category (41–100%) or the low GC% category (0–41%). The OcculterCut v1.1 cut-off for distinguishing between the two categories is 41%.

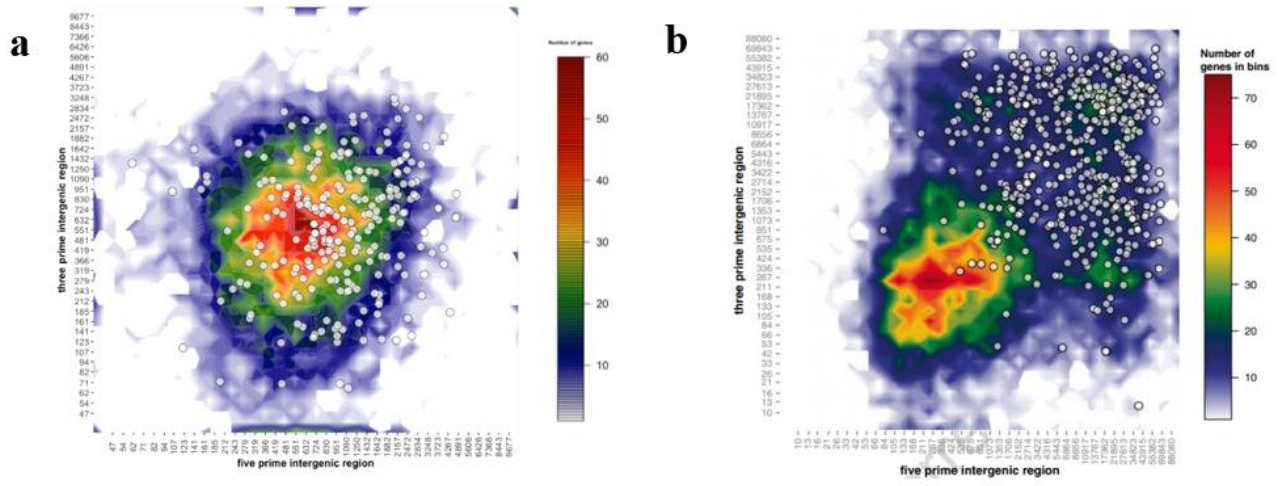


Fig. 2.6 Density plot of the intergenic distances on the 5'-end (x-axis) against the 3'-end (y-axis) for each predicted gene on the genome. **a** Density plot for the *C. zeina* genome. White circles represent CSEPs. **b** Density plot for the *Phytophthora infestans* genome. White circles represent RXLR effectors. *P. infestans* plot obtained from (Saunders et al. 2014).

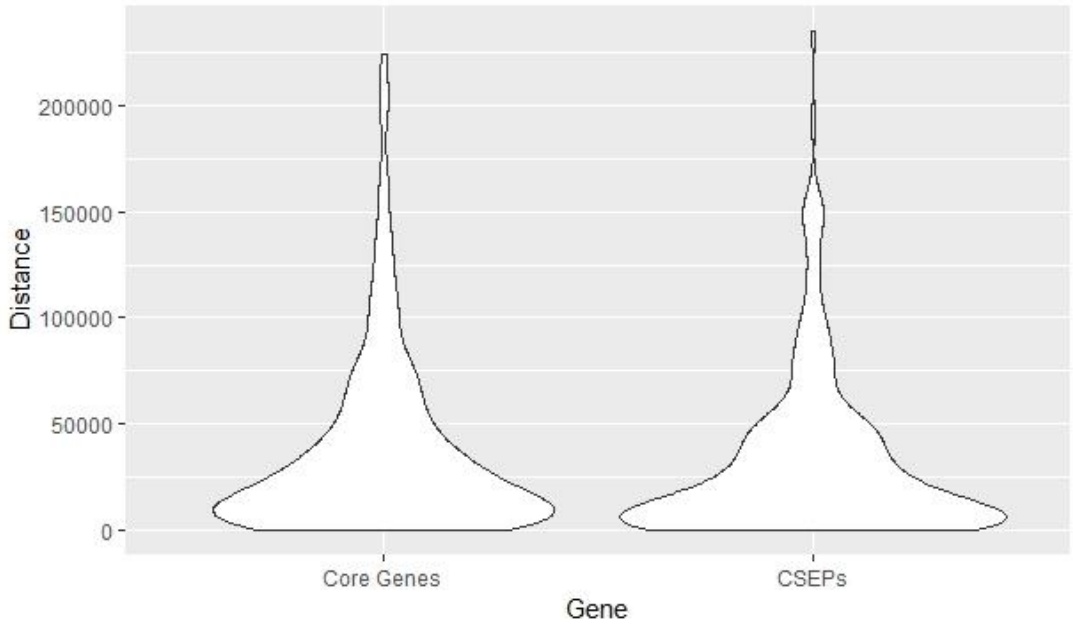


Fig. 2.7 Violin plots of the distances to the nearest transposable elements of the CSEP genes and 274 randomly selected BUSCO core genes of *C. zeina*. Wilcoxon signed rank test showed that the difference in means was not significant with a p-value of 0.06134.

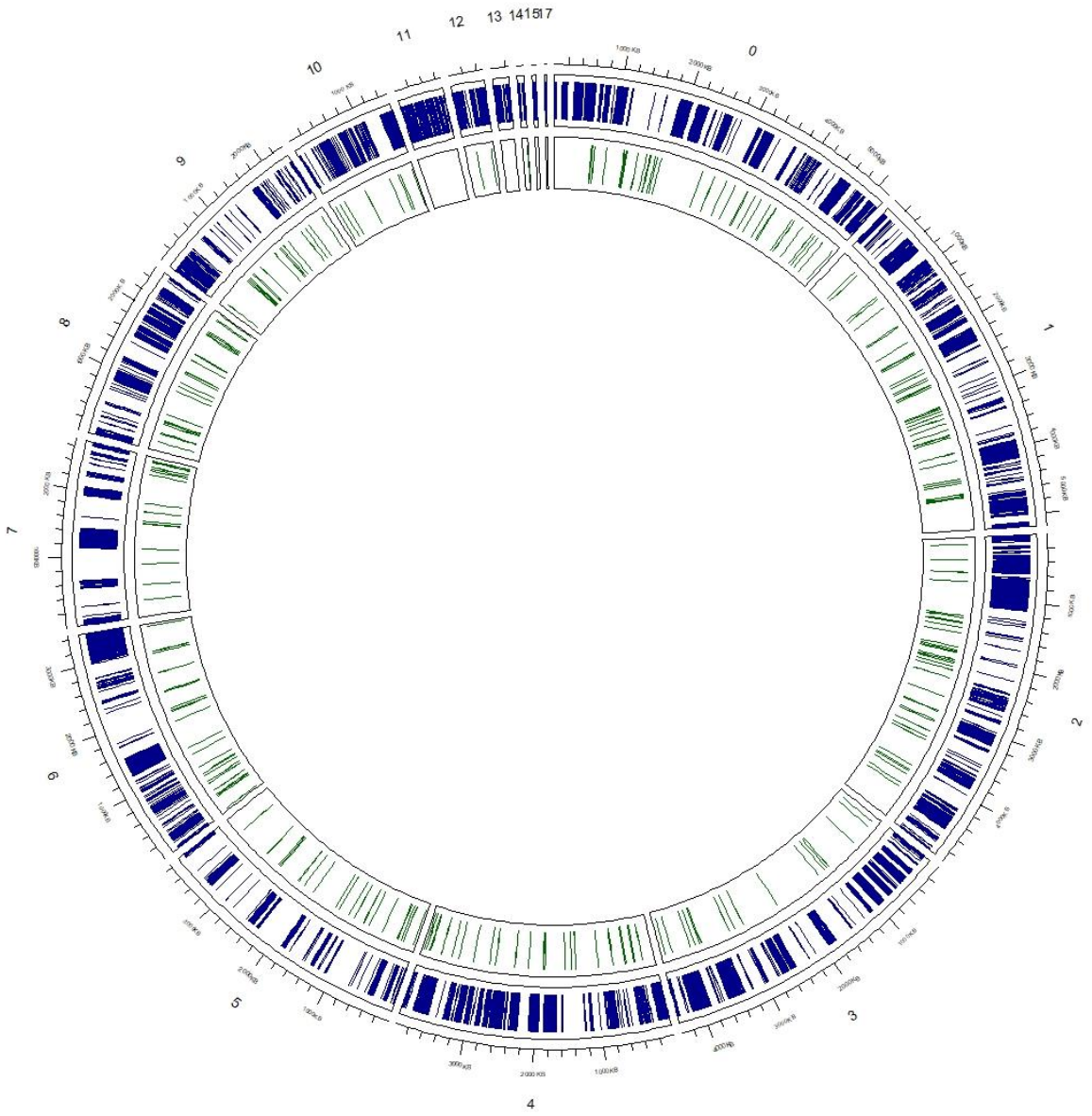


Fig. 2.8 Circos plot depicting the features of the *C. zeina* assembly. Contigs are numbered on the outer perimeter of the plot. Blue lines represent the position of the TEs on the contigs. Green lines represent the positions of the CSEPs on the contigs.

2.5 DISCUSSION

The main outcome of this study was the development of a less-fragmented assembly of the *C. zeina* genome. This assembly has allowed for the discovery of effector genes and to confirm whether *C. zeina* has a "two-speed" genome. The main findings of the study were that (i) *C. zeina* has a similar number of effector genes to other Dothideomycetes; (ii) these genes potentially encode proteins that

aid in the prevention of recognition by the host plant; and (iii) the genome of *C. zeina* is divided into a transposon AT-rich compartment and a gene-dense GC-rich compartment, but effector genes are not populated in the AT-rich compartment.

Third-generation sequencing technologies have allowed researchers to obtain more insight into an organism's genome than was possible before. The ability to have complete and contiguous assemblies not only provides clues to the number of chromosomes present in an organism but also allows the discovery of genes that were previously lost because of the nature of fragmented assemblies (Gibriel et al. 2016). In this study, the genome of *C. zeina* was re-sequenced using the third-generation PacBio SMRT sequencing technology after it was previously sequenced with the second-generation Illumina sequencing technology (Wingfield et al. 2017). The PacBio-sequenced genome was only approximately 1 Mbp larger than the Illumina-sequenced genome and revealed 3 potential full chromosomes.

The completeness assessment with BUSCO, which searches for orthologues of 1,315 genes conserved in Ascomycota, revealed that 15 genes were missing in the SMRT-sequenced genome compared to the 20 of the Illumina-sequenced genome. Some of these "missing" core Ascomycota genes may be absent in a more specific lineage like Capnodiales, so a BUSCO assessment of a more specific gene set would have been more interesting. This result also indicates that the gapless PacBio assembly allowed detection and annotation of five genes from the Ascomycota BUSCO set not found in the Illumina assembly. Applying this 0.4% improvement in gene detection to the C. Zeina Illumina genome assembly with 12,110 predicted genes make a prediction of an additional 46 genes in the PacBio assembly. Annotation of the SMRT-sequenced assembly with BRAKER, which takes into account RNA-Seq data not used in the BUSCO analysis, showed that the SMRT-sequenced genome had a predcted 12,436 genes whilst the unmasked Illumina-sequenced genome had 12,110. The newer annotation recovered 3 missing BUSCO genes (7 vs 10) from the older Illumina annotation and had less fragmented BUSCO genes (35 vs 37) but had more duplicated genes (35 vs 27) (Appendix Table 3). This showed that although the less-fragmented assembly could recover some genes, the gene prediction process for certain genes can still differ.

The secretome of a plant pathogenic fungus is vital for maintaining growth, evading or breaking down host defence responses, penetrating the host and killing the host cell if required. It is therefore important to have information about the secretome in order to understand the pathogen's biology. A

total of 1,025 proteins were predicted to be secreted in *C. zeina*. Many of these proteins had a GO term of oxidoreductase activity. In an infection study of the closely related *D. septosporum*, an upregulation of oxidoreductases was found to occur in the necrotrophic stages with speculation that it may be a response to oxidative stress (Bradshaw et al. 2016). Alternatively, the oxidoreductases may be compensating for the lack of CTB7 production in *C. zeina* (Swart et al. 2017). The CTB7 gene is an oxidoreductase gene part of the cercosporin biosynthetic gene cluster involved in pathogenesis of the *Cercospora* species (Chen et al. 2007; Newman and Townsend 2016). Additionally, 17 proteins were predicted to have the GO term peroxidase activity. A transcriptome study of the infection study of *Z. tritici* revealed that peroxidases were upregulated during the post-penetration symptomless phase of leaf colonization (Rudd et al. 2015). They hypothesized that it was necessary for the protection against oxidative stress during infection. Another abundant GO term was that of peptidases. This could indicate that *C. zeina* might have to fend off against host defences by proteolysis of the host secreted proteins. In another study of *Z. tritici*, the authors found evidence of a high selection of secreted peptidases belonging to the A1 and G1 MEROPS families and they hypothesized that it could lead to neo-functionalization in the form of effectors (Krishnan et al. 2018). They showed that secreted peptidases had significantly higher dN/dS ratios compared to non-secreted peptidases and that several secreted peptidases had dN/dS > 1 compared to the non-secreted peptidases which all had dN/dS < 1 (Krishnan et al. 2018).

Within the cohort of secreted proteins during infection are the small secreted proteins known as effectors. These proteins affect the host organism in a way that promotes the pathogen's infection. A total of 274 CSEPs were predicted in *C. zeina* via a combined approach of manual selection based on a criteria and by using a machine learning tool. This amount is similar to that of other Dothideomycetes (Ohm et al. 2012). CAZymes comprised of only 19 of the CSEPs. Five of these were from the CE5 family. This family consists of cutinases which hydrolyse cutin residues from the cuticle of the host plant and ultimately allows for the penetration by the pathogen (Sweigard et al. 1992). Another two where from the GH16 family which catalyse hemicellulose degradation. This indicates that *C. zeina* could be secreting effectors that aid in the facilitation of penetration into the maize cell. Analysing if these CSEPs are secreted in the early stages of infection might corroborate this. Most of the CSEPs were predicted to localize in the apoplast. This feature further indicates that *C. zeina* might be a hemibiotroph rather than a necrotroph as previously proposed because this feature is common to hemibiotrophs (Lo Presti et al. 2015). Some non-orthologous effector proteins across the fungal kingdom are hypothesized to contain identical tandem repeats

that are required for plant pathogenicity (Chang et al. 2018). Three CSEPs were identified with tandem repeats. However, a search for these tandem repeats yielded no similarity in other functionally characterized effectors except for one in *S. cerevisae*. One these *C. zeina* CSEPs was hypothetical, the other had a Blast2GO annotation of a cell wall protein and the third was annotated as a fungal hydrophobin. In the maize pathogen *Ustilago maydis*, the Rep1 secreted protein containing a tandem repeat domain was revealed to possibly contribute to mycelial hydrophobicity (Teertstra et al. 2006).

The *C. zeina* CSEPs that are of particular interest are the homologues of effectors that have been functionally characterized in closely related pathogens. The orthologue of the *C. fulvum* AVR4 effector protein g1947 in *C. zeina* fits this category and is a good candidate for functional analysis. The *C. fulvum* AVR4 effector prevents hydrolysis of the chitin residues of the fungal cell wall by plant chitinases (van den Burg et al. 2006; van Esse et al. 2007). The AVR4 effector seems to be vital for several Dothideomycetes especially in the Capnodiales. The infiltration of the purified *P. fijiensis* AVR4 homologue in a resistant banana variety resulted in a hypersensitive-like response suggesting that the resistant plant could harbour an *R* protein that recognises the effector (Arango Isaza et al. 2016). Additionally, the AVR4 deletion mutants of *Cercospora* cf. *flagellaris* resulted in reduced cercosporin production and ultimately reduced virulence on soybean (Santos Rezende et al. 2020). This showed that AVR4 was required for cercosporin production. Analysing whether the AVR4 homologue in *C. zeina* can influence the cercosporin gene cluster transcription would be interesting.

Another candidate for functional analysis would be the ECP6 homologue g3999 in *C. zeina*. The ECP6 effector sequesters chitin residues that are released from the hyphal cell walls to prevent recognition by the host plant (de Jonge et al. 2010). The ECP6 deletion mutants of *Z. tritici* infecting specific pattern recognition receptor deletion mutant wheat plants resulted in the successful colonisation by the pathogen (Lee et al. 2014). This showed that the ECP6 homologue was required to prevent PTI. Interestingly, the AVR4 and ECP6 homologues in *C. zeina* were not predicted to contain the CBM14 and CBM50 CAZyme domains respectively that are usually associated with these carbohydrate binding proteins. However, the CSEP g11003 had a CBM63 domain and g2614 had a CBM18 domain, indicating that these proteins could have carbohydrate binding functions.

We performed a differential expression analysis of CSEP genes between the lower leaves and upper leaves of susceptible B73 maize plants challenged with *C. zeina* to test which genes could be involved during late infection and early infection respectively. However, due to the extremely low *C. zeina* read counts in the upper leaves, this analysis would not give a definite answer. We therefore checked for the normalized counts across all samples to identify CSEP genes that had a relatively high expression *in planta*. A total of 47 CSEP genes had normalized counts above the 90th percentile with 22 of these genes predicted to encode small extracellular proteins that are unique to *C. zeina* and had EffectorP values that ranged from 0.671 to 0.927. The high *in planta* expression relative to other genes suggests that these are good candidates for further analyses. Another five of the 47 CSEP genes were predicted to encode cell wall and expansin proteins, suggesting that they play an important role in cell wall growth *in planta*. The *C. zeina* homologues of the ECP2 and ECP6 were among the CSEP genes with high *in planta* expression (Appendix Table 5). This suggests that these effectors are important for the evasion of recognition during the infection process. Surprisingly, the AVR4 homologue did not have normalized counts above the 90th percentile. In a transcriptomic study of the *D. septosporum* and pine pathosystem, they showed that the ECP2 and ECP6 homologues increased 31- and 104-fold in expression respectively during the mid stages of infection when compared to the early stages (Bradshaw et al. 2016). They also showed that the AVR4 homologue had negligible expression throughout the infection stages (Bradshaw et al. 2016). This expression pattern for the AVR4 homologue might be similar for *C. zeina* as well. The last CSEP gene that had normalized counts above the 90th percentile was predicted to encode a hydrophobin protein. Interestingly, in the same transcriptomic study of *D. septosporum*, they showed that a hydrophobin gene was one of the highest expressed genes in the mid and late stages of infection (Bradshaw et al. 2016). However, the deletion of this gene did not hamper the virulence of *D. septosporum* (Bradshaw et al. 2016). A study of a secreted *C. fulvum* hydrophobin showed that the protein had a role in the adhesion of germinating conidia (Lacroix et al. 2008). The high expression counts for the hydrophobin gene in *C. zeina* suggests an important role in cell adhesion to the maize leaf surfaces during infection. A functional study of this gene in *C. zeina* would be needed to confirm this.

Many of the predicted 274 CSEP genes in *C. zeina* were hypothetical and unique to *C. zeina*. The ones with no expression read counts in either *in vitro* or *in planta* datasets will most probably have to be disregarded as they might not be actual genes. However, the ones with relatively low expression levels could be orphan genes. Determining whether these genes are expressed at specific

time points of infection could give an indication that they could potentially be involved in pathogenicity. Orphan genes are thought to ultimately be involved in species-specific adaptation (Long et al. 2003; Kaessmann 2010). Despite having hundreds of CSEPs, it is often the case that the deletion/disruption of a single effector gene will significantly affect the pathogenicity of the plant pathogenic fungus. For example, AVR4-silenced *C. fulvum* mutant strains had a weaker ability to colonise tomato leaves when compared to the wild-type strain (van Esse et al. 2007). Similarly, the CfTom1-deletion *C. fulvum* mutant had 63% reduction in biomass after inoculated onto tomato leaves when compared to the wild-type strain (Ökmen et al. 2013). This leads to the questions why are there many candidate effector genes and are all of them functional? Several reasons could explain this characteristic. Firstly, whilst some effectors prevent recognition or cause host cell death, some might be needed for the growth of structures like haustoria and others might be needed for nutrient absorption in the apoplastic space. Secondly, some effectors might function as complexes and therefore the absence of one protein might make the complex incomplete and ultimately non-functional. Thirdly, the expression of one effector gene might be essential for the expression of the other effector genes. Lastly, some effectors might only be host-specific making the expression of others unnecessary and only functional if the pathogen is colonising another host.

The “two-speed” genome is a term used to describe the bipartite architecture of plant pathogenic fungal genomes (Dong et al. 2015). This includes the rapidly evolving repeat-rich compartment and the slowly evolving repeat-poor compartment. In *L. maculans*, an AT-isochore generated by extensive RIP mutations was revealed to contain several effector genes (Rouxel et al. 2011). This close proximity proposed the idea that RIP leakage effects on the effector genes could lead to their diversification. When comparing the distance of the candidate effector genes to TEs in *C. zeina* with that of core genes, the effector genes were on average slightly closer to the TEs (Fig. 2.7). However, this difference was not significant. Additionally, the density plot of intergenic distances revealed that the effector genes are not clearly distant from the rest of the genes when comparing to the genome architecture of the oomycete *P. infestans* which has a clear “two-speed” genome. A similarly compact genome architecture was seen in the Dothideomycete *R. collo-cygni* which causes ramularia leaf spot in barley (Stam et al. 2018). In this case the genome was referred to as a “one-speed” genome because the TEs only comprised of 6% of the genome. In *C. zeina* however, the TEs comprise of nearly 39% of the genome.

The analysis of the genome architecture of *C. zeina* with OcculterCut has revealed that the genome has two major compartments, namely AT-rich compartments that comprise of 33.2% of the genome and GC-rich compartments that take up 66.8% of the genome. This is usually an indication of a "two-speed" genome, however, seeing that only 19 genes are present "within" the AT-rich compartments, the use of the term "two-speed" should not necessarily be used. The extreme lack of gene presence within these AT-rich compartments is more of an indication that the TEs expand more rapidly and separate from the rest of the genome. If this is the case, would it indicate that *C. zeina* has a "one-speed" genome and the evolution of virulence genes are not as rapid as in other pathogens with "two-speed" genomes? The most used example of a "one-speed" or one-compartment genome is the pathogen causing the barley powdery mildew disease, *Blumeria graminis* f.sp. *hordei*. In this fungus, large-scale compartmentalisation is not present, and the effector genes do not localize close to AT-rich regions (Frantzeskakis et al. 2018). In order to bypass the lack of a rapidly evolving compartment, the evolution of effector genes is aided by the copy number variation (CNV) and heterozygosity of these genes (Frantzeskakis et al. 2018). The CNV feature was not only present for effector genes but also for other core genes. This indicates that non-synonymous mutations in an effector gene paralogue could render a new function in the gene product and lead to diversification and therefore the evolution of the pathogen. Therefore, the lack of a "two-speed" genome does not rule out the evolution of effector genes in the pathogen and this could be the case for *C. zeina*. Another type of the "two-speed" genome is the presence of an accessory chromosome that can harbour effector genes (Ma et al. 2010). This shows that physical compartmentalisation within the genome is not the only evolutionary trajectory that is caused by selection pressures. In *C. zeina*, whether and/or how the large and independent AT-rich compartment affects the evolution of the effector genes will still need to be determined.

The expansion of TEs is common in Capnodiales but rare in the Pleosporales. For example, the repetitive elements make up 39%, 48% and 53% of the *C. zeina* (this study), *C. fulvum* and *P. fijiensis* genomes, respectively. Additionally, the Capnodiales have more non-LTR elements than Pleosplorales (Fig. 2.3). This indicates that the expansion of TEs in the *C. zeina* genome is not entirely unique. However, when considering that some closely related *Cercospora* species have smaller genomes – 37 Mbp in *Cercospora beticola* (Vaghefi et al. 2017), 34 Mbp in *Cercospora canescens* (Chand et al. 2015), 33 Mbp in *Cercospora* cf. *sigesbeckiae* (Albu et al. 2017), 33 Mbp in *Cercospora kikuchii* strain ARG_18_001 (Sautua et al. 2019) and 33 Mbp in *Cercospora arachidicola* (Orner et al. 2015) it could imply a slight recent rapid expansion of TEs in *C. zeina*.

The repetitive elements have only been reported in *C. kikuchii* where they found a genomic proportion of only 0.54% (Sautua et al. 2019). Although, the repetitive elements have not been reported in other *Cercospora* species, the GC content of the *C. zeina* genome (47.96%) is still smaller than that of the others (51.20% in *C. canescens*, 51.5% in *C.* cf. *sigesbeckiae* and 53% in *C. kikuchii*) indicating that *C. zeina* could have more AT-rich compartments that are associated with repetitive elements. Additionally, *C. sojina* has a similar genome size (41 Mbp) to *C. zeina* but has smaller genomic proportion of repetitive elements (25.56%) and higher GC content (53.12%) (Luo et al. 2018). Therefore, these comparisons could suggest a recent rapid expansion of TEs in *C. zeina*.

There are four possible hypotheses that could explain the expansion: firstly *C. zeina* could be immune to RIP, secondly *C. zeina* might have undergone limited sexual life cycles, thirdly there could have been horizontal transfer of TEs and lastly, in cases such as that of *C. kikuchii*, the repetitive elements could be under-represented because of the use of second-generation sequencing technologies. In the first scenario, considering that the CT-to-TT mutation is more common than the CA-to-TA mutation in *C. zeina*, there is an implication that RIP is active but not effective. A CA-to-TA mutation would most likely lead to a premature stop codon and is consequently said to be the preferred RIP mutation (Galagan and Selker 2004). Therefore, the CT-to-TT mutations might not be able to prevent transposase genes from being transcribed which could lead to expansion. The CT-to-TT dominant RIP mutation has also been found in the Dothideomycete *A. rabiei*, however the repetitive elements only comprised 10% of the genome and the prevention of TE expansion by RNA silencing was hypothesized (Verma et al. 2016). In the second scenario, it is possible that *C. zeina* has undergone limited and cryptic sexual life cycles. The RIP mutation process occurs during sexual reproduction (Galagan and Selker 2004), therefore if TE expansion occurs in an organism that possibly has the RIP machinery, then limited RIP may be a result of limited sexual life cycles. It could be that *C. zeina* and perhaps the sister species *Cercospora zeae-maydis* might have undergone fewer sexual life cycles after speciation from other *Cercospora* species. The sexual phase of *C. zeina* has not been observed, and therefore it is thought that this fungus proliferates mostly through asexual reproduction in the maize host. However, population genetics studies of *C. zeina* on maize in South Africa have shown strong evidence for cryptic sex (Muller et al. 2016; Nsibo et al. 2019). In the third scenario, *C. zeina* might have acquired TEs via horizontal transfer from either the host or other maize foliar pathogens. Horizontal transfer has been reported between Dothideomycetes before (McDonald et al. 2019), however reports of evidence of host to pathogen

horizontal transfer have been lacking. Additionally, it would be more likely to happen in biotrophic pathogens because of the longer period spent in symbiosis.

In summary, we have generated a less fragmented genome assembly of *C. zeina*. This assembly has recovered a few genes that were not predicted in the Illumina assembly. More importantly, the assembly has allowed for the discovery of CSEPs that might play a role in pathogenicity and has revealed AT-rich compartments that possibly indicate a rapid expansion of TEs. Contrary to our hypothesis, effector genes do not localize in AT-rich compartments but instead in GC-rich compartments. Future prospects include the functional analyses of the best candidate effector proteins, genomic comparisons with related species to provide more insight on the pathogenicity mechanisms as well as the population genomics of *C. zeina* isolates to reveal effector genes that are under selection. This study has paved the way for the possibility effector-based breeding of maize whereby maize plants containing *R* proteins are selected for (Vleeshouwers and Oliver 2014).

2.6 REFERENCES

Agger JW, Isaksen T, Varnai A, et al (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc Natl Acad Sci 111:6287–6292. https://doi.org/10.1073/pnas.1323629111

Albu S, Sharma S, Bluhm BH, et al (2017) Draft Genome Sequence of *Cercospora* cf. *sigesbeckiae*, a Causal Agent of Cercospora Leaf Blight on Soybean. Genome Announc 5:e00708-17. https://doi.org/10.1128/genomeA.00708-17

Allen GC, Flores-Vergara MA, Krasynanski S, et al (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc 1:2320–2325. https://doi.org/10.1038/nprot.2006.384

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arango Isaza RE, Diaz-Trujillo C, Dhillon B, et al (2016) Combating a Global Threat to a Clonal Crop: Banana Black Sigatoka Pathogen *Pseudocercospora fijiensis* (Synonym *Mycosphaerella fijiensis*) Genomes Reveal Clues for Disease Control. PLOS Genet 12:e1005876. https://doi.org/10.1371/journal.pgen.1005876

Beckman PM, Payne GA (1983) Cultural Techniques and Conditions Influencing Growth and Sporulation of *Cercospora zeae-maydis* and Lesion Development in Corn. Phytopathology

73:286–289. https://doi.org/10.1094/Phyto-73-286

Berthelier J, Casse N, Daccord N, et al (2018) A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. BMC Genomics 19:e378. https://doi.org/10.1186/s12864-018-4763-1

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

Bradshaw RE, Guo Y, Sim AD, et al (2016) Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporum* throughout its infection cycle of the gymnosperm host *Pinus radiata*. Mol Plant Pathol 17:210–224. https://doi.org/10.1111/mpp.12273

Chand R, Pal C, Singh V, et al (2015) Draft Genome Sequence of *Cercospora canescens*: A Leaf Spot Causing Pathogen. Curr Sci 109:2103–2110. https://doi.org/10.18520/cs/v109/i11/2103-2110

Chang H-X, Noel ZA, Sang H, Chilvers MI (2018) Annotation resource of tandem repeat-containing secretory proteins in sixty fungi. Fungal Genet Biol 119:7–19. https://doi.org/10.1016/j.fgb.2018.07.004

Chen H, Lee M-H, Daub ME, Chung K-R (2007) Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. Mol Microbiol 64:755–770. https://doi.org/10.1111/j.1365-2958.2007.05689.x

Christie N, Myburg AA, Joubert F, et al (2017) Systems genetics reveals a transcriptional network associated with susceptibility in the maize-grey leaf spot pathosystem. Plant J 89:746–763. https://doi.org/10.1111/tpj.13419

Collemare J, Griffiths S, Iida Y, et al (2014) Secondary Metabolism and Biotrophic Lifestyle in the Tomato Pathogen *Cladosporium fulvum*. PLoS One 9:e85877. https://doi.org/10.1371/journal.pone.0085877

Conesa A, Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008:1–12. https://doi.org/10.1155/2008/619832

Dal Molin A, Minio A, Griggio F, et al (2018) The genome assembly of the fungal pathogen *Pyrenochaeta lycopersici* from Single-Molecule Real-Time sequencing sheds new light on its biological complexity. PLoS One 13:e0200217. https://doi.org/10.1371/journal.pone.0200217

Dalio RJD, Herlihy J, Oliveira TS, et al (2018) Effector Biology in Focus: A Primer for Computational Prediction and Functional Characterization. Mol Plant-Microbe Interact 31:22–33. https://doi.org/10.1094/MPMI-07-17-0174-FI

Daub ME, Ehrenshaft M (2000) The Photoactivated *Cercospora* Toxin Cercosporin: Contributions

to Plant Disease and Fundamental Biology. Annu Rev Phytopathol 38:461–490. https://doi.org/10.1146/annurev.phyto.38.1.461

de Jonge R, Peter van Esse H, Kombrink A, et al (2010) Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science 329:953–955. https://doi.org/10.1126/science.1190859

de Wit PJGM (2016) *Cladosporium fulvum* Effectors: Weapons in the Arms Race with Tomato. Annu Rev Phytopathol 54:1–23. https://doi.org/10.1146/annurev-phyto-011516-040249

de Wit PJGM, van der Burgt A, Ökmen B, et al (2012) The Genomes of the Fungal Plant Pathogens *Cladosporium fulvum* and *Dothistroma septosporum* Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. PLoS Genet 8:e1003088. https://doi.org/10.1371/journal.pgen.1003088

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

Dutreux F, Da Silva C, D'Agata L, et al (2018) *De novo* assembly and annotation of three *Leptosphaeria* genomes using Oxford Nanopore MinION sequencing. Sci Data 5:180235. https://doi.org/10.1038/sdata.2018.235

Faino L, Seidl MF, Datema E, et al (2015) Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. MBio 6:e00936-15. https://doi.org/10.1128/mBio.00936-15

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–W37. https://doi.org/10.1093/nar/gkr367

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

Frantzeskakis L, Kracher B, Kusch S, et al (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 19:e381. https://doi.org/10.1186/s12864-018-4750-6

Friesen TL, Stukenbrock EH, Liu Z, et al (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet 38:953–956. https://doi.org/10.1038/ng1839

Fu L, Niu B, Zhu Z, et al (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Fulnečková J, Ševčíková T, Fajkus J, et al (2013) A Broad Phylogenetic Survey Unveils the Diversity and Evolution of Telomeres in Eukaryotes. Genome Biol Evol 5:468–483.

https://doi.org/10.1093/gbe/evt019

Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. Trends Genet 20:417–423. https://doi.org/10.1016/j.tig.2004.07.007

Gibriel HAY, Thomma BPHJ, Seidl MF (2016) The Age of Effectors: Genome-Based Discovery and Applications. Phytopathology 106:1206–1212. https://doi.org/10.1094/PHYTO-02-16-0110-FI

Griffiths S, Mesarich CH, Saccomanno B, et al (2016) Elucidation of cladofulvin biosynthesis reveals a cytochrome P450 monooxygenase required for anthraquinone dimerization. Proc Natl Acad Sci 113:6851–6856. https://doi.org/10.1073/pnas.1603528113

Griffiths S, Saccomanno B, de Wit PJGM, Collemare J (2015) Regulation of secondary metabolite production in the fungal tomato pathogen *Cladosporium fulvum*. Fungal Genet Biol 84:52–61. https://doi.org/10.1016/j.fgb.2015.09.009

Gu Z, Gu L, Eils R, et al (2014) circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hane JK, Oliver RP (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9:e478. https://doi.org/10.1186/1471-2105-9-478

Harvey AJ, Hrmova M, De Gori R, et al (2000) Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. Proteins Struct Funct Genet 41:257–269. https://doi.org/10.1002/1097-0134(20001101)41:2<257::AID-PROT100>3.0.CO;2-C

Heystek M (2014) Promoter analysis of three defense gene family members from maize in response to *Cercospora zeina*. University of Pretoria

Hoede C, Arnoux S, Moisset M, et al (2014) PASTEC: An Automatic Transposable Element Classification Tool. PLoS One 9:e91929. https://doi.org/10.1371/journal.pone.0091929

Hoff KJ, Lange S, Lomsadze A, et al (2016) BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661

Jiao Y, Peluso P, Shi J, et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527. https://doi.org/10.1038/nature22971

Jones DA, Bertazzoni S, Turo CJ, et al (2018) Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. Curr Opin Microbiol 46:43–49.

https://doi.org/10.1016/j.mib.2018.01.017

Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. Genome Res 20:1313–1326. https://doi.org/10.1101/gr.101386.109

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317

King R, Urban M, Hammond-Kosack MCU, et al (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics 16:e544. https://doi.org/10.1186/s12864-015-1756-1

Krishnan P, Ma X, McDonald BA, Brunner PC (2018) Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. BMC Evol Biol 18:e7. https://doi.org/10.1186/s12862-018-1123-3

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315

Kumar V (2010) Analysis of the key active subsites of glycoside hydrolase 13 family members. Carbohydr Res 345:893–898. https://doi.org/10.1016/j.carres.2010.02.007

Lacroix H, Whiteford JR, Spanu PD (2008) Localization of *Cladosporium fulvum* hydrophobins reveals a role for HCf-6 in adhesion. FEMS Microbiol Lett 286:136–144. https://doi.org/10.1111/j.1574-6968.2008.01227.x

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lee W-S, Rudd JJ, Hammond-Kosack KE, Kanyuka K (2014) *Mycosphaerella graminicola* LysM Effector-Mediated Stealth Pathogenesis Subverts Recognition Through Both CERK1 and CEBiP Homologues in Wheat. Mol Plant-Microbe Interact 27:236–243. https://doi.org/10.1094/MPMI-07-13-0201-R

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

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

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656

Lo Presti L, Lanver D, Schweizer G, et al (2015) Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol 66:513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Long M, Betrán E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4:865–875. https://doi.org/10.1038/nrg1204

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:e550. https://doi.org/10.1186/s13059-014-0550-8

Luo X, Cao J, Huang J, et al (2018) Genome sequencing and comparative genomics reveal the potential pathogenic mechanism of *Cercospora sojina* Hara on soybean. DNA Res 25:25–37. https://doi.org/10.1093/dnares/dsx035

Ma L-J, van der Does HC, Borkovich KA, et al (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373. https://doi.org/10.1038/nature08850

Mat Razali N, Cheah BH, Nadarajah K (2019) Transposable Elements Adaptive Role in Genome Plasticity, Pathogenicity and Evolution in Fungal Phytopathogens. Int J Mol Sci 20:e3597. https://doi.org/10.3390/ijms20143597

McDonald MC, Taranto AP, Hill E, et al (2019) Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein ToxA between Three Fungal Wheat Pathogens. MBio 10:. https://doi.org/10.1128/mBio.01515-19

McGrann GRD, Andongabo A, Sjökvist E, et al (2016) The genome of the emerging barley pathogen *Ramularia collo-cygni*. BMC Genomics 17:e584. https://doi.org/10.1186/s12864-016-2928-3

Meisel B, Korsman J, Kloppers FJ, Berger DK (2009) *Cercospora zeina* is the causal agent of grey leaf spot disease of maize in southern Africa. Eur J Plant Pathol 124:577–583. https://doi.org/10.1007/s10658-009-9443-1

Meyer J, Berger DK, Christensen SA, Murray SL (2017) RNA-Seq analysis of resistant and susceptible sub-tropical maize lines reveals a role for kauralexins in resistance to grey leaf spot disease, caused by *Cercospora zeina*. BMC Plant Biol 17:197. https://doi.org/10.1186/s12870-

017-1137-9

Muller MF, Barnes I, Kunene NT, et al (2016) *Cercospora zeina* from Maize in South Africa Exhibits High Genetic Diversity and Lack of Regional Population Differentiation. Phytopathology 106:1194–1205. https://doi.org/10.1094/PHYTO-02-16-0084-FI

Munkvold GP, Martinson CA, Shriver JM, Dixon PM (2001) Probabilities for Profitable Fungicide Use Against Gray Leaf Spot in Hybrid Maize. Phytopathology 91:477–484. https://doi.org/10.1094/PHYTO.2001.91.5.477

Newman AG, Townsend CA (2016) Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen *Cercospora nicotianae*. J Am Chem Soc 138:4219–4228. https://doi.org/10.1021/jacs.6b00633

Newman AM, Cooper JB (2007) XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. BMC Bioinformatics 8:e382. https://doi.org/10.1186/1471-2105-8-382

Ngoko Z, Cardwell KF, Marasas WFO, et al (2002) Biological and Physical Constraints on Maize Production in the Humid Forest and Western Highlands of Cameroon. Eur J Plant Pathol 108:893–902. https://doi.org/10.1023/A:1021206028492

Noar RD, Thomas E, Daub ME (2019) A novel polyketide synthase gene cluster in the plant pathogenic fungus *Pseudocercospora fijiensis*. PLoS One 14:e0212229. https://doi.org/10.1371/journal.pone.0212229

Nsibo DL, Barnes I, Kunene NT, Berger DK (2019) Influence of farming practices on the population genetics of the maize pathogen *Cercospora zeina* in South Africa. Fungal Genet Biol 125:36–44. https://doi.org/10.1016/j.fgb.2019.01.005

Ohm RA, Feau N, Henrissat B, et al (2012) Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLoS Pathog 8:e1003037. https://doi.org/10.1371/journal.ppat.1003037

Ökmen B, Etalo DW, Joosten MHAJ, et al (2013) Detoxification of α-tomatine by *Cladosporium fulvum* is required for full virulence on tomato. New Phytol 198:1203–1214. https://doi.org/10.1111/nph.12208

Orner VA, Cantonwine EG, Wang XM, et al (2015) Draft Genome Sequence of *Cercospora arachidicola*, Causal Agent of Early Leaf Spot in Peanuts. Genome Announc 3:e01281-15. https://doi.org/10.1128/genomeA.01281-15

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701

Plissonneau C, Hartmann FE, Croll D (2018) Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol 16:e5. https://doi.org/10.1186/s12915-017-0457-4

Plissonneau C, Stürchler A, Croll D (2016) The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. MBio 7:e01231-16. https://doi.org/10.1128/mBio.01231-16

Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. Bioinformatics 21:i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Quesneville H, Bergman CM, Andrieu O, et al (2005) Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comput Biol 1:e22. https://doi.org/10.1371/journal.pcbi.0010022

Richards JK, Wyatt NA, Liu Z, et al (2018) Reference Quality Genome Assemblies of Three *Parastagonospora nodorum* Isolates Differing in Virulence on Wheat. G3 Genes|Genomes|Genetics 8:393–399. https://doi.org/10.1534/g3.117.300462

Rouxel T, Grandaubert J, Hane JK, et al (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Commun 2:202. https://doi.org/10.1038/ncomms1189

Rudd JJ, Kanyuka K, Hassani-Pak K, et al (2015) Transcriptome and Metabolite Profiling of the Infection Cycle of *Zymoseptoria tritici* on Wheat Reveals a Biphasic Interaction with Plant Immunity Involving Differential Pathogen Chromosomal Contributions and a Variation on the Hemibiotrophic Lifestyle Def. Plant Physiol 167:1158–1185. https://doi.org/10.1104/pp.114.255927

Santos Rezende J, Zivanovic M, Costa de Novaes MI, Chen Z (2020) The AVR4 effector is involved in cercosporin biosynthesis and likely affects the virulence of *Cercospora* cf. *flagellaris* on soybean. Mol Plant Pathol 21:53–65. https://doi.org/10.1111/mpp.12879

Saunders DGO, Win J, Kamoun S, Raffaele S (2014) Two-Dimensional Data Binning for the Analysis of Genome Architecture in Filamentous Plant Pathogens and Other Eukaryotes. In: Birch PRJ, Jones JT, Bos JIB (eds) Plant-Pathogen Interactions, Second. Human Press, pp 29–51

Sautua FJ, Gonzalez SA, Doyle VP, et al (2019) Draft genome sequence data of *Cercospora kikuchii*, a causal agent of Cercospora leaf blight and purple seed stain of soybeans. Data Br 27:e104693. https://doi.org/10.1016/j.dib.2019.104693

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61:263–289. https://doi.org/10.1146/annurev-arplant-042809-112315

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sonah H, Deshmukh RK, Bélanger RR (2016) Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. Front Plant Sci 7:e126. https://doi.org/10.3389/fpls.2016.00126

Sørensen A, Lübeck M, Lübeck P, Ahring B (2013) Fungal Beta-Glucosidases: A Bottleneck in Industrial Use of Lignocellulosic Materials. Biomolecules 3:612–631. https://doi.org/10.3390/biom3030612

Sperschneider J, Catanzariti A-M, DeBoer K, et al (2017) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep 7:e44598. https://doi.org/10.1038/srep44598

Sperschneider J, Dodds PN, Singh KB, Taylor JM (2018) ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. New Phytol 217:1764–1778. https://doi.org/10.1111/nph.14946

Sperschneider J, Gardiner DM, Dodds PN, et al (2016) EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol 210:743–761. https://doi.org/10.1111/nph.13794

Stam MR, Danchin EGJ, Rancurel C, et al (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of -amylase-related proteins. Protein Eng Des Sel 19:555–562. https://doi.org/10.1093/protein/gzl044

Stam R, Münsterkötter M, Pophaly SD, et al (2018) A New Reference Genome Shows the One-Speed Genome Structure of the Barley Pathogen *Ramularia collo-cygni*. Genome Biol Evol 10:3243–3249. https://doi.org/10.1093/gbe/evy240

Stanke M, Keller O, Gunduz I, et al (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. https://doi.org/10.1093/nar/gkl200

Swart V, Crampton BG, Ridenour JB, et al (2017) Complementation of CTB7 in the Maize Pathogen *Cercospora zeina* Overcomes the Lack of *In Vitro* Cercosporin Production. Mol Plant-Microbe Interact. https://doi.org/10.1094/MPMI-03-17-0054-R

Sweigard JA, Chumley FG, Valent B (1992) Disruption of a *Magnaporthe grisea* cutinase gene. Mol Gen Genet 232:183–190

Teertstra WR, Deelstra HJ, Vranes M, et al (2006) Repellents have functionally replaced hydrophobins in mediating attachment to a hydrophobic surface and in formation of

hydrophobic aerial hyphae in *Ustilago maydis*. Microbiology 152:3607–3612. https://doi.org/10.1099/mic.0.29034-0

Testa AC, Oliver RP, Hane JK (2016) OculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes. Genome Biol Evol 8:2044–2064. https://doi.org/10.1093/gbe/evw121

Urban M, Cuzick A, Seager J, et al (2019) PHI-base: the pathogen–host interactions database. Nucleic Acids Res 48:D613–D620. https://doi.org/10.1093/nar/gkz904

Vaghefi N, Kikkert JR, Bolton MD, et al (2017) *De novo* genome assembly of *Cercospora beticola* for microsatellite marker development and validation. Fungal Ecol 26:125–134. https://doi.org/10.1016/j.funeco.2017.01.006

van den Burg HA, Harrison SJ, Joosten MHAJ, et al (2006) *Cladosporium fulvum* Avr4 Protects Fungal Cell Walls Against Hydrolysis by Plant Chitinases Accumulating During Infection. Mol Plant-Microbe Interact 19:1420–1430. https://doi.org/10.1094/MPMI-19-1420

van Esse HP, Bolton MD, Stergiopoulos I, et al (2007) The Chitin-Binding *Cladosporium fulvum* Effector Protein Avr4 Is a Virulence Factor. Mol Plant-Microbe Interact 20:1092–1101. https://doi.org/10.1094/MPMI-20-9-1092

van Wyk S, Harrison CH, Wingfield BD, et al (2019) The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ 7:e7447. https://doi.org/10.7717/peerj.7447

Verma S, Gazara RK, Nizam S, et al (2016) Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. Sci Rep 6:e24638. https://doi.org/10.1038/srep24638

Vleeshouwers VGAA, Oliver RP (2014) Effectors as Tools in Disease Resistance Breeding Against Biotrophic, Hemibiotrophic, and Necrotrophic Plant Pathogens. Mol Plant-Microbe Interact 27:196–206. https://doi.org/10.1094/MPMI-10-13-0313-IA

Walker BJ, Abeel T, Shea T, et al (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Ward JMJ, Stromberg EL, Nowell DC, Nutter FW (1999) Gray leaf Spot: A Disease of Global Importance in Maize Production. Plant Dis 83:884–895. https://doi.org/10.1094/PDIS.1999.83.10.884

Weiland JJ, Chung K-R, Suttle J (2010) The role of cercosporin in the virulence of *Cercospora* ssp. to plant hosts. In: Lartey RT, Weiland JJ, Panella L, et al. (eds) Cercospora Leaf Spot of Sugar Beet and Related Species. The American Phytopathological Society, St. Paul, Minnesota, pp

109–117

Wingfield BD, Berger DK, Steenkamp ET, et al (2017) Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. IMA Fungus 8:385–396. https://doi.org/10.5598/imafungus.2017.08.02.10

Winnenburg R (2006) PHI-base: a new database for pathogen host interactions. Nucleic Acids Res 34:D459–D464. https://doi.org/10.1093/nar/gkj047

Wyatt NA, Richards JK, Brueggeman RS, Friesen TL (2018) Reference Assembly and Annotation of the *Pyrenophora teres* f. *teres* Isolate 0-1. G3 Genes|Genomes|Genetics 8:1–8. https://doi.org/10.1534/g3.117.300196

Zaccaron AZ, Bluhm BH (2017) The genome sequence of *Bipolaris cookei* reveals mechanisms of pathogenesis underlying target leaf spot of sorghum. Sci Rep 7:e17217. https://doi.org/10.1038/s41598-017-17476-x

Zhang H, Yohe T, Huang L, et al (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418

Zhao Z, Liu H, Wang C, Xu J-R (2014) Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics 15:e6. https://doi.org/10.1186/1471-2164-15-6

2.7 APPENDIX

2.7.1 Tables

Appendix Table 1 The sizes and putative content of the five non protein-coding nuclear or mitochondrial genome contigs of *C. zeina*

Contig ID	Contig length (bp)	Match no.	Putative contig content[a]	Accession	E-value	Query cover	Percent identity
16F	21,062	Match 1	Match to *Cercospora zeae-maydis* Maggy-like degenerate transposable element	AY170476.1	2e-33	1%	72.24
		Match 2	Match to *Zymoseptoria tritici* ST99CH_3D7 genome assembly, chromosome 2	LT853693.1	2e-17	0%	80.85
18F	29,885	Match 1	Match to *Cercospora zeae-maydis* Maggy-like degenerate transposable element	AY170476.1	2e-33	1%	72.24
		Match 2	Match to *Zymoseptoria tritici* ST99CH_3D7 genome assembly, chromosome 2	LT853693.1	2e-17	0%	80.85
19F	23,006	Match 1	Match to *Cercospora zebrina* strain CBS 118789 18S ribosomal RNA gene	GU214656.1	0.0	56%	99.46
		Match 2	Match to *Cercospora sojina* strain CPC 12322 18S ribosomal RNA gene	GU214655.1	0.0	56%	99.48
20F	27,256	Match 1	Match to *C. zeina* microsatellite CzSSR12 from (Muller et al. 2016)	KP015841.1	4e-20	0%	94.67

		Match 2	Match to *C. zeina* microsatellite CzSSR11 from (Muller et al. 2016)	KP015840.1	2e-18	0%	93.33
21F[b]	5,517	Match 1	Match to *Acephala applanata* strain T1_K93_131 NADH-ubiquinone oxidoreductase chain 1 (nad1) gene	JN091563.1	1e-13	2%	75.94
		Match 2	Match to *Phialocephala fortinii* strain 7_6_7v NADH-ubiquinone oxidoreductase chain 1 (nad1) gene	JN091554.1	6e-12	2%	75.19

[a] Result of BLASTN analysis of *C. zeina* contig against the nr/nt database (NCBI) with an e-value cut-off of 1e-05.

[b] The "discontiguous megablast" setting was used for the BLASTN search of contig 21F because default settings yielded no significant match.

Appendix Table 2 The comparison of the genomic BUSCO assessment between the *C. zeina* Illumina assembly and the PacBio assembly using the Ascomycota dataset

	Illumina assembly	PacBio filtered assembly[a]	Polished filtered assembly[b]
Complete BUSCOs	1,275 (97.0%)	1,278 (97.2%)	1,278 (97.2%)
Single Copy	1,274 (96.9%)	1,277 (97.1%)	1,277 (97.1%)
Duplicated	1 (0.1%)	1 (0.1%)	1 (0.1%)
Fragmented	20 (1.5%)	22 (1.7%)	22 (1.7%)
Missing	20 (1.5%)	15 (1.1%)	15 (1.1%)
TOTAL	1,315	1,315	1,315

[a] Filtered through realignments and discarding of contigs with a depth deviating by more than 1.5X from the average coverage across all contigs weighted by the contig length.

[b] PacBio assembly polished with Illumina data using Pilon.

Appendix Table 3 The comparison of the annotation (proteins) BUSCO assessment between the *C. zeina* Illumina assembly and the PacBio assembly using the Ascomycota dataset

	Illumina assembly	PacBio assembly[a]
Complete	1,268 (96.5%)	1,273 (96.8%)
Single-copy complete	1,241 (94.4%)	1,238 (94.1%)
Duplicated complete	27 (2.1%)	35 (2.7%)
Fragmented	37 (2.8%)	35 (2.7%)
Missing	10 (0.7%)	7 (0.5%)
TOTAL	1,315	1,315

[a] PacBio assembly polished with Illumina data using Pilon.

Appendix Table 4 The properties of *C. zeina* effectors that contain tandem repeat domains

Protein ID	Protein length	Repeat positions	Period	Copy number	Consensus error	Repeat sequence

g664.t1	282	204 – 229	12	2.17	0.04	APVSQIADGQVQ
g4070.t1	189	56 – 75	7	2.86	0.00	GNGNSGN
g10161.t1	181	29 – 67	12	3.25	0.05	SAPDSTGEKGSG

Appendix Table 5 The normalized read counts of *C. zeina* CSEP genes *in planta*

Gene IDs	Blast2GO annotation	Normalized counts[a]
g1858	hypothetical protein BST61_czeina14g000160	7769.57
g4678	fungal hydrophobin	3991.96
g7577	small secreted protein	3099.47
g3999	carbohydrate-binding module family 50 protein (ECP6)	1748.97
g3082	related to cell wall protein PhiA	1139.66
g664	Cell wall protein PIR5	1101.62
g6576	expansin-related protein	1066.53
g4360	extracellular protein 20-2	912.15
g20	extracellular protein (ECP2)	833.86
g10576	extracellular protein 20-2	764.17
g6601	GPI anchored serine-threonine rich protein	599.46
g10271	extracellular protein 32-2	583.32
g10547	---NA---	554.66
g6863	---NA---	501.73
g2634	hypothetical protein COCMIDRAFT_33555	499.71
g4070	fungal hydrophobin	409.33
g2532	hypothetical protein BST61_czeina24g000800	399.25
g2221	Endo-1,4-beta-xylanase F3	359.61
g1358	---NA---	354.82
g11514	---NA---	353.77
g6073	Phosphatidylglycerol/phosphatidylinositol transfer protein	353.08
g7524	hypothetical protein	290.78
g10124	Superoxide dismutase [Mn]	273.88
g11003	carbohydrate-binding module family 63 protein	255.05
g5885	MF(ALPHA)1	223.07
g3494	Peptidyl-prolyl cis-trans isomerase B	222.72
g8995	WSC domain containing protein	222.55
g3933	putative glycerol-3-phosphate dehydrogenase nad-dependent protein	218.32
g2145	carbohydrate esterase family 5 protein	215.16
g4278	Secreted protein NIS1	201.97
g2899	clock-controlled protein 6	194.74
g4226	hypothetical protein BST61_czeina70g000340	188.76
g7315	---NA---	184.24
g4011	GPI anchored CFEM domain protein C	180.09
g6200	siderophore biosynthesis enzyme	163.74
g8673	related to Probable cutinase 1	163.32

Gene IDs	Blast2GO annotation	Normalized counts[a]
g4282	extracellular protein 20-2	162.54
g10111	extracellular protein 36-1	155.35
g1988	hypothetical protein BST61_czeina14g001490	155.1
g10955	hypothetical protein BST61_czeina53g000160	152.23
g3680	hypothetical protein BST61_czeina89g000060	152.21
g2965	hypothetical protein BST61_czeina4g000800	142.12
g137	FK506-binding protein 2	138.02
g2887	hypothetical protein BST61_czeina4g000110	117.69
g6365	putative cdp-alcohol phosphatidyltransferase protein	113.66
g11466	hypothetical protein BST61_czeina8g001170	107.06
g8359	---NA---	105.74
g6793	hypothetical protein BST61_czeina17g000540	100.48
g1750	TPA: chitin binding protein, putative (AFU_orthologue; AFUA_1G14430)	100.3
g10483	---NA---	99.27
g5829	hypothetical protein BST61_czeina11g000740	99.01
g10862	---NA---	98.89
g3083	Acetyl-CoA synthetase-like protein	94.89
g11455	hypothetical protein SETTUDRAFT_39539	93.73
g5712	putative zinc finger and scan domain-containing protein 18	89.19
g9800	hypothetical protein BST61_czeina5g000610	83.57
g252	CFEM domain-containing protein	81.73
g7399	hypothetical protein BST61_czeina159g000010	78.78
g1513	hypothetical protein CB0940_06777	75.92
g9797	---NA---	74
g5701	hypothetical protein BST61_czeina134g000100	72.29
g6532	hypothetical protein BST61_czeina12g000240	70.98
g43	hypothetical protein BST61_czeina96g000080	70.83
g7710	putative membrane protein C17A5.08	70.02
g8310	hypothetical protein BST61_czeina81g000180	65.13
g8176	---NA---	64.38
g1947	AVR4 protein	63.06
g4497	hypothetical protein BST61_czeina68g000300	59.97
g8370	Guanyl-specific ribonuclease F1	58.17
g913	hypothetical protein CBER1_01301	56.74
g1106	putative mannose 6-phosphate receptor-like protein	54.77
g8029	extracellular serine-threonine rich protein	54.74
g5745	---NA---	54.5
g5094	Endoplasmic reticulum vesicle protein 25	54.18
g7336	---NA---	51.49
g9737	---NA---	50.54
g1180	extracellular protein 58-1	50.42
g7526	---NA---	49.98
g2146	putative endo-1,4-beta-xylanase A	48.65

Gene IDs	Blast2GO annotation	Normalized counts[a]
g10654	hypothetical protein BST61_czeina2g000270	48.59
g6465	Protein ERP1	47.99
g268	hypothetical protein BST61_czeina19g000180	47.13
g8839	small secreted protein	46.12
g6987	hypothetical protein BST61_czeina30g000210	45.27
g8456	Cell differentiation protein rcd1	40.73
g8290	hypothetical protein BST61_czeina75g000040	34.89
g11248	hypothetical protein BST61_czeina91g000080	34.84
g8382	hypothetical protein BST61_czeina62g000340	32.35
g4178	hypothetical protein BST61_czeina73g000360	30.66
g8412	---NA---	29.28
g6478	hypothetical protein BST61_czeina95g000140	28.01
g2456	Surfeit locus protein 1	27.61
g2637	hypothetical protein BST61_czeina24g000040	27.13
g11360	Extracellular metalloprotease	26.68
g6160	Serine/threonine-protein kinase ppk13	23.98
g3800	---NA---	23.27
g10161	hypothetical protein BST61_czeina89g000060	23.22
g1933	hypothetical protein BST61_czeina14g000510	23.06
g1144	hypothetical protein BST61_czeina52g000520	23
g8864	major allergen alt a1	22.06
g9813	hypothetical protein CB0940_02034	21.5
g6172	hypothetical protein BST61_czeina9g002160	21.2
g8258	hypothetical protein BST61_czeina70g000340	21.11
g8133	hypothetical protein BST61_czeina269g000010	20.84
g4914	carbohydrate esterase family 1 protein	20.68
g5879	guanine deaminase	20.56
g3200	carbohydrate esterase family 12 protein	20.21
g7351	UPF0357 protein	19.91
g11042	related to malate dehydrogenase	19.77
g9245	hypothetical protein CB0940_08017	19.29
g9002	lipocalin-like domain-containing protein	18.98
g2850	PR-1-like protein	18.35
g8621	extracellular protein 52	18.17
g9115	---NA---	16.96
g1066	hypothetical protein BST61_czeina84g000180	16.69
g6457	hypothetical protein BST61_czeina42g000370	16.61
g8291	DUF1183-domain-containing protein	15.97
g2626	hypothetical protein BST61_czeina24g000240	15.67
g3628	phosphoglycerate mutase family protein	15.18
g11035	Aprataxin-like protein	14.55
g6858	putative 54S ribosomal protein L32, mitochondrial	14.51
g10264	hypothetical protein BST61_czeina114g000230	14.43
g6477	extracellular protein 51	13.4

Gene IDs	Blast2GO annotation	Normalized counts[a]
g7440	Long chronological lifespan protein 2	13.27
g9310	hypothetical protein CB0940_08044	12.98
g3640	hypothetical protein BST61_czeina8g000480	12.41
g4840	hypothetical protein BST61_czeina10g000140	11.82
g8766	hypothetical protein BST61_czeina1g000690	10.35
g2004	hypothetical protein BST61_czeina14g001340	10.2
g1881	Endo-1,4-beta-xylanase	10.08
g10398	pgap1-like protein	9.99
g10610	hypothetical protein BST61_czeina2g000670	9.5
g9407	putative cutinase 1	9.2
g1496	hypothetical protein BST61_czeina45g000350	8.56
g4652	Dihydrolipoyl dehydrogenase	8.49
g2283	hypothetical protein BST61_czeina13g000220	8.21
g3276	glycoside hydrolase family 128 protein	7.89
g5628	deoxyribonuclease NucA/NucB-domain-containing protein	7.64
g9637	hypothetical protein BST61_czeina25g001000	7.61
g4972	extracellular protein 60-1	7.25
g7142	Extracellular metalloprotease	7.13
g10327	---NA---	7.09
g3560	DNA-directed RNA polymerase I subunit RPA12	7.04
g5579	hypothetical protein BST61_czeina94g000140	7
g1650	hypothetical protein BST61_czeina90g000060	6.79
g5841	putative endonuclease exonuclease phosphatase family protein	6.75
g8234	---NA---	6.7
g4599	2OG-Fe(II) oxygenase	6.55
g2614	carbohydrate-binding module family 18	6.32
g10403	---NA---	6.1
g4499	hypothetical protein BST61_czeina58g000790	5.99
g10386	hypothetical protein BST61_czeina2g002430	5.93
g10152	Peptidase inhibitor 16	5.92
g5083	---NA---	5.66
g6053	Transcription elongation factor S-II	5.51
g7916	protein disulfide-isomerase domain-containing protein	5.46
g11258	spore coat protein SP65-like	5.42
g1232	glycoside hydrolase family 16 protein	5.28
g2615	pectate lyase	5.05
g10625	carbohydrate esterase family 5 protein	4.94
g11484	riboflavin aldehyde-forming enzyme	4.9
g7432	carbohydrate esterase family 16 protein	4.75
g1790	predicted protein	4.74
g1994	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit	4.55
g853	hypothetical protein BST61_czeina49g000370	4.47
g7302	hydrophobin-like protein	4.32
g2862	hypothetical protein BST61_czeina184g000030	4.27

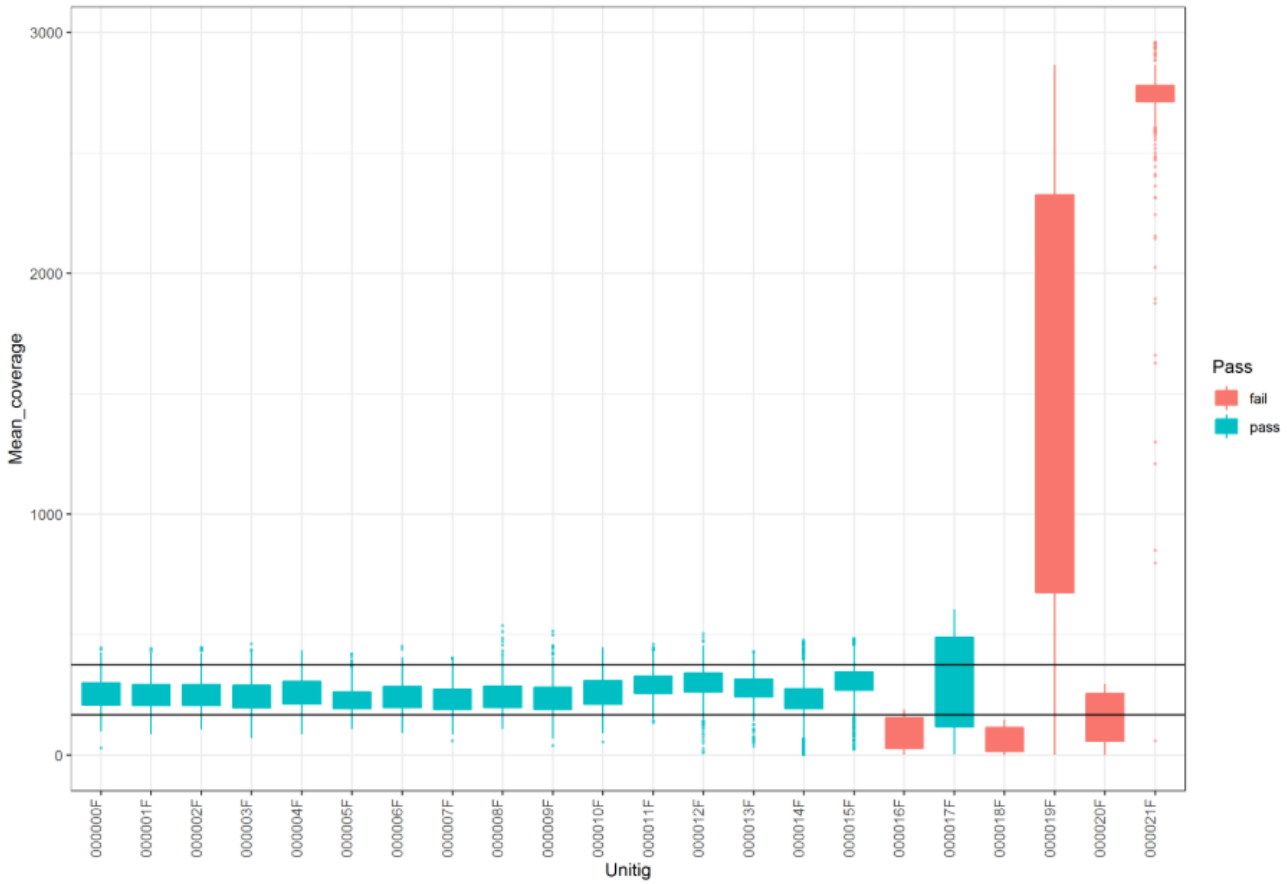
Gene IDs	Blast2GO annotation	Normalized counts[a]
g6284	lipase 3-like protein	4.24
g5918	hypothetical protein CB0940_04299	4.21
g2633	---NA---	4.06
g1802	hypothetical protein BST61_czeina77g000320	3.88
g6283	hypothetical protein CBER1_09316	3.75
g9259	hypothetical protein BST61_czeina37g000470	3.7
g7864	hypothetical protein BST61_czeina97g000190	3.63
g4018	cobalt/magnesium transport protein cora	3.52
g8846	hypothetical protein BST61_czeina1g001720	3.29
g1533	class II hydrophobin	3.29
g4247	hypothetical protein BST61_czeina51g000490	3.23
g9883	Mitochondrial carrier protein RIM2	3.21
g3342	predicted protein	3.08
g5395	Micronemal protein 4	3.01
g11151	hypothetical protein BST61_czeina20g001320	2.93
g2247	TRP-domain-containing protein	2.91
g3128	hypothetical protein BST61_czeina22g000200	2.89
g6282	hypothetical protein BST61_czeina136g000060	2.86
g9941	hydrolytic enzyme protein	2.85
g5217	peptidoglycan-binding domain 1 protein	2.78
g10132	ricin B lectin	2.71
g5106	carbohydrate esterase family 4 protein	2.57
g6083	Lactobacillus up-regulated protein	2.56
g342	glycoside hydrolase family 24 protein	2.55
g10088	uncharacterized protein RCC_07522	2.48
g3156	RmlC-like cupin	2.44
g2173	Zip-domain-containing protein	2.35
g7398	---NA---	2.28
g6145	hypothetical protein BST61_czeina9g001910	2.14
g9529	related to endo alpha-1,4 polygalactosaminidase precusor	2.1
g2197	trypsin-domain-containing protein	2.03
g2322	hypothetical protein BST61_czeina13g000540	1.98
g7743	glycoside hydrolase family 16 protein	1.98
g291	antigenic thaumatin-like protein	1.81
g3553	---NA---	1.8
g3306	---NA---	1.79
g7791	hypothetical protein BST61_czeina21g000480	1.78
g3917	oviduct-specific glyco protein	1.75
g2516	hypothetical protein CBER1_08503	1.71
g4094	related to dehydrogenase/reductase	1.68
g2157	hypothetical protein BST61_czeina14g002720	1.66
g5616	---NA---	1.64
g9260	hypothetical protein CB0940_08031	1.53
g8235	DUF3328 domain containing protein	1.51

Gene IDs	Blast2GO annotation	Normalized counts[a]
g2507	P-loop containing nucleoside triphosphate hydrolase protein	1.46
g10823	PR-1-like protein	1.43
g1665	hypothetical protein BST61_czeina90g000120	1.43
g1785	hypothetical protein BST61_czeina77g000180	1.33
g8504	hypothetical protein BST61_czeina101g000120	1.27
g5844	putative O-acetyltransferase CAS1	1.23
g2340	---NA---	1.22
g6819	Extracellular membrane protein, CFEM domain protein	1.14
g5953	carbonic anhydrase	1.1
g11101	hypothetical protein CB0940_03271	1.07
g4466	hypothetical protein BST61_czeina68g000160	1.04
g2729	hypothetical protein BST61_czeina187g000100	1.04
g4976	hypothetical protein CBER1_02601	1.02
g4569	hypothetical protein BST61_czeina58g000550	0.99
g6230	cutinase 1	0.95
g8538	---NA---	0.94
g8010	hypothetical protein BST61_czeina147g000010	0.94
g4500	hypothetical protein BST61_czeina58g000540	0.93
g8485	hypothetical protein BST61_czeina311g000010	0.88
g3843	---NA---	0.88
g7058	starch-binding domain-like protein	0.87
g4080	hypothetical protein BST61_czeina69g000210	0.76
g3837	hypothetical protein BST61_czeina47g000260	0.75
g11524	A-agglutinin anchorage subunit-like	0.73
g11558	hypothetical protein DOTSEDRAFT_33504	0.73
g8311	---NA---	0.72
g10285	hypothetical protein BST61_czeina114g000150	0.72
g1071	---NA---	0.71
g8304	Heat shock protein 70	0.69
g7956	hypothetical protein BST61_czeina112g000020	0.69
g7857	hypothetical protein BST61_czeina97g000160	0.68
g11377	hypothetical protein CB0940_03796	0.56
g1392	hypothetical protein BST61_czeina27g000650	0.54
g3134	---NA---	0.53
g7868	D-alanyl-D-alanine carboxypeptidase	0.51
g7017	hypothetical protein BST61_czeina30g000640	0.39
g6097	Guanyl-specific ribonuclease F1	0.36
g3938	lytic polysaccharide monooxygenase	0.36
g8252	hypothetical protein BST61_czeina75g000370	0.34
g7527	hypothetical protein BST61_czeina61g000480	0.34
g6779	hypothetical protein BST61_czeina17g000440	0.34
g8469	lipase 3-like protein	0.33
g2328	hypothetical protein BST61_czeina13g000560	0.32
g3428	hypothetical protein BST61_czeina85g000250	0.2

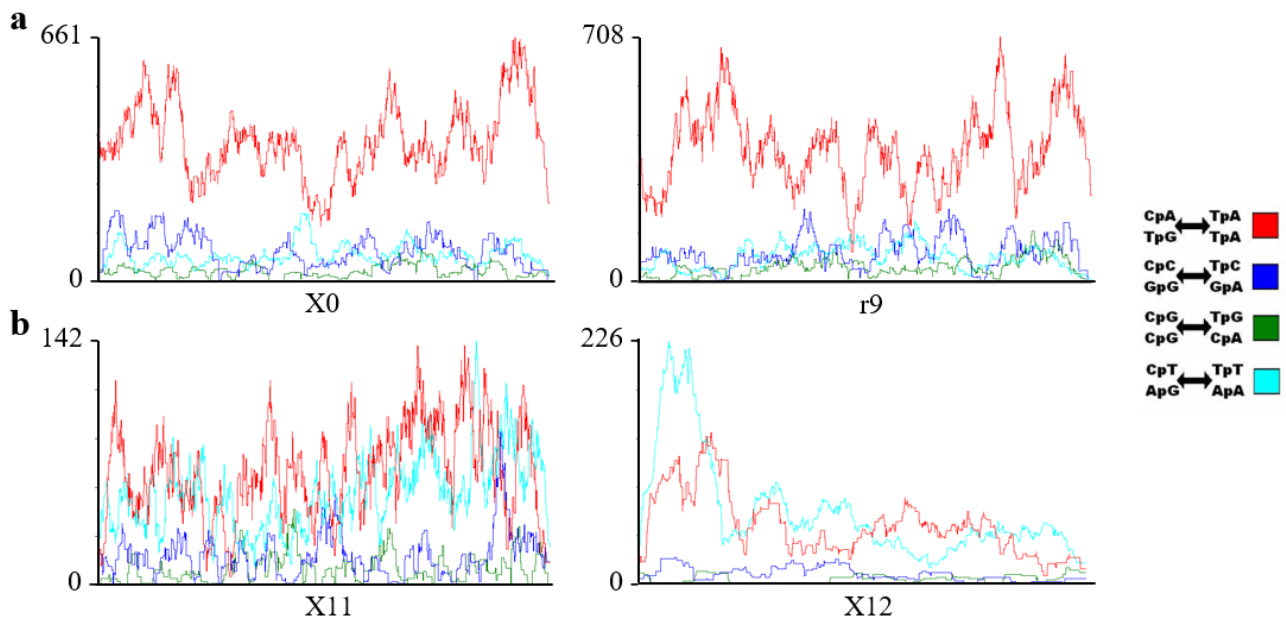
Gene IDs	Blast2GO annotation	Normalized counts[a]
g11222	hypothetical protein CBER1_00513	0.2
g6746	hypothetical protein BST61_czeina17g000760	0.18
g2860	---NA---	0.18
g10572	deoxyribonuclease nucA/NucB domain-containing protein	0.18
g3636	PR-1-like protein	0.17
g959	Guanine-nucleotide dissociation stimulator CDC25	0.14

[a] The 90th percentile of the normalized counts for *C. zeina* genes is 101.11.

2.7.2 Figures

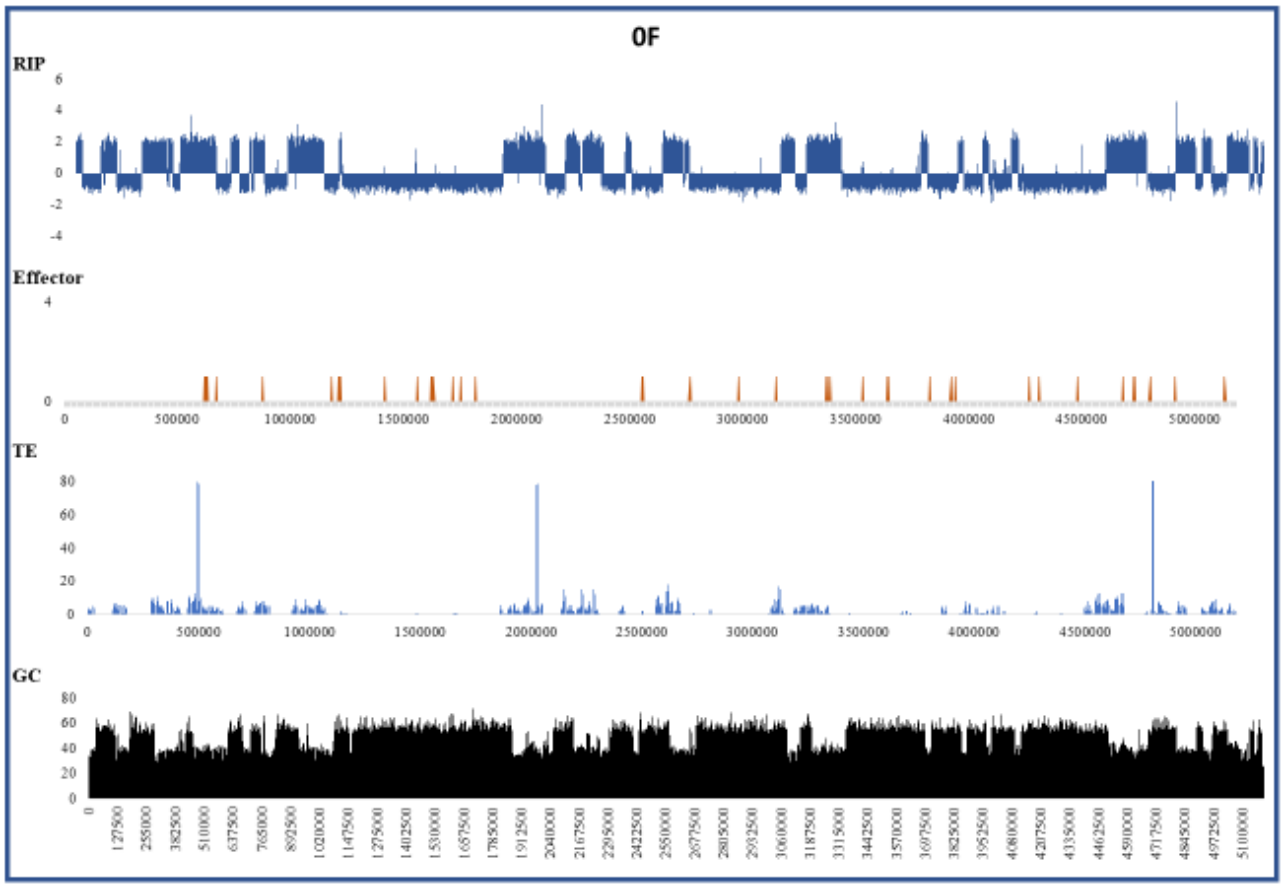


Appendix Fig. 1 Graph illustrating the coverage filter on the *C. zeina* PacBio assembly.

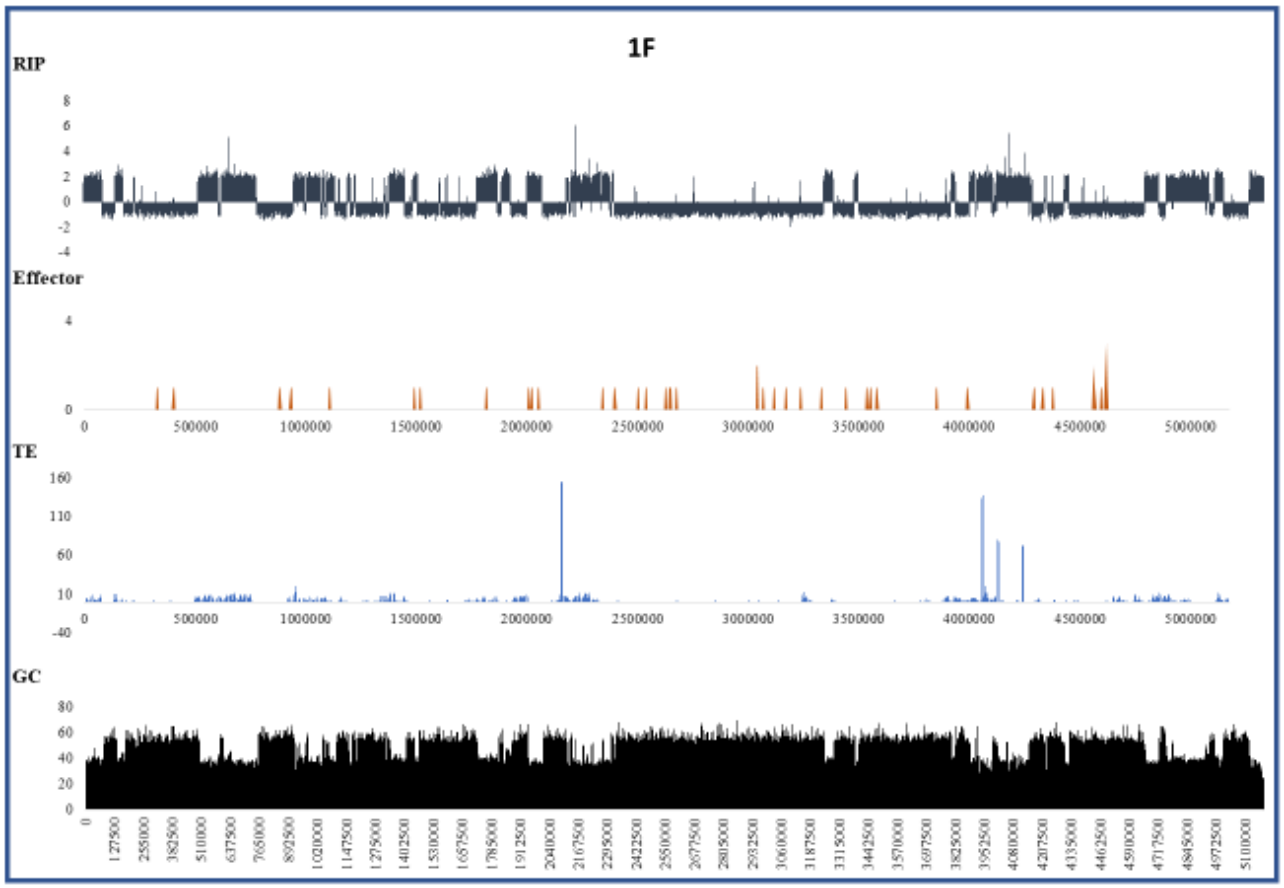


Appendix Fig. 2 Repeat-induced point mutation (RIP) dinucleotide bias in selected *Parastagonospora nodorum* transposable element families. **a** Two non-LTR transposon families with 76 and 72 sequences (from left to right). **b** Two *Gypsy* LTR transposon families with 36 and 29 sequences (from left to right). The y-axis illustrates the RIP mutation frequency along the length of the TE alignment (x-axis). Data collected from (Hane and Oliver 2008).

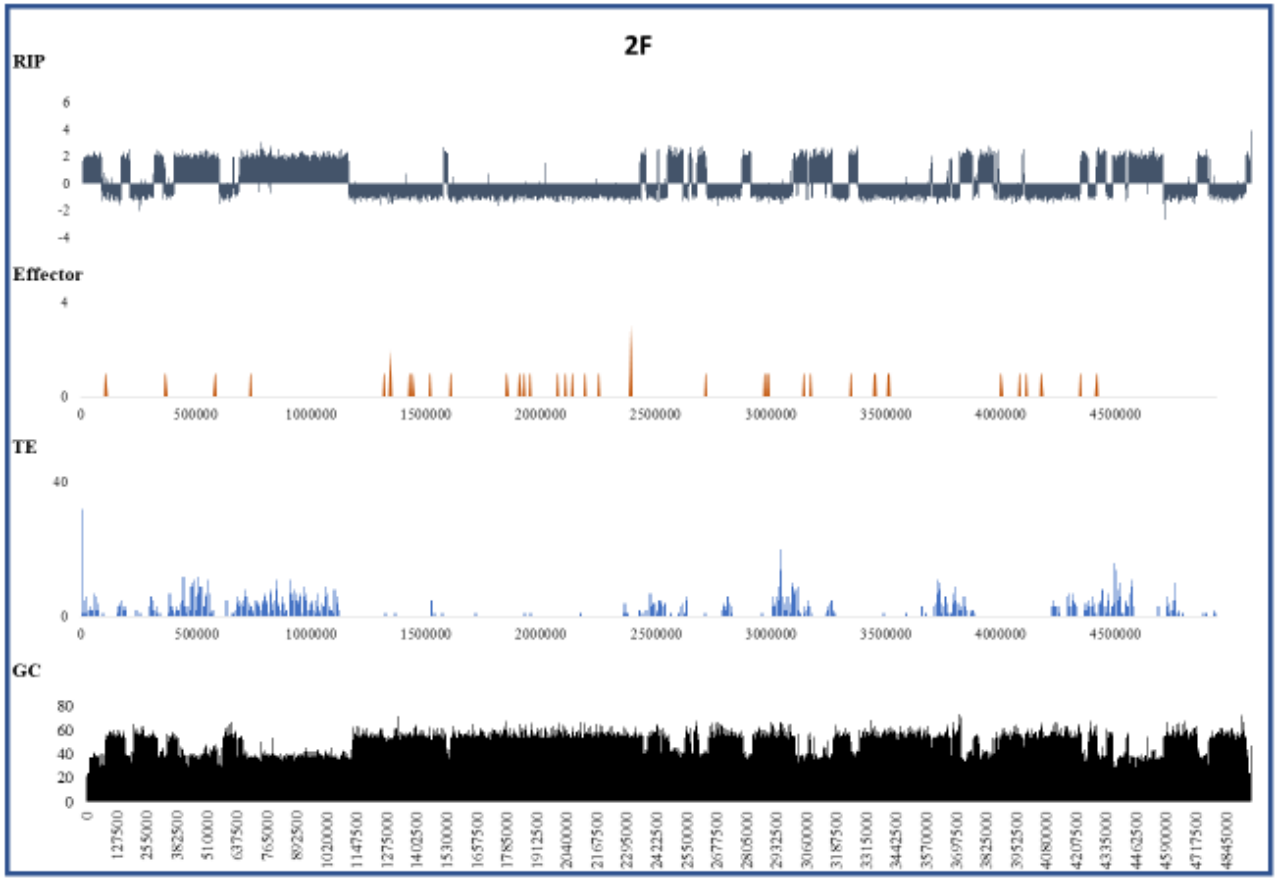
a



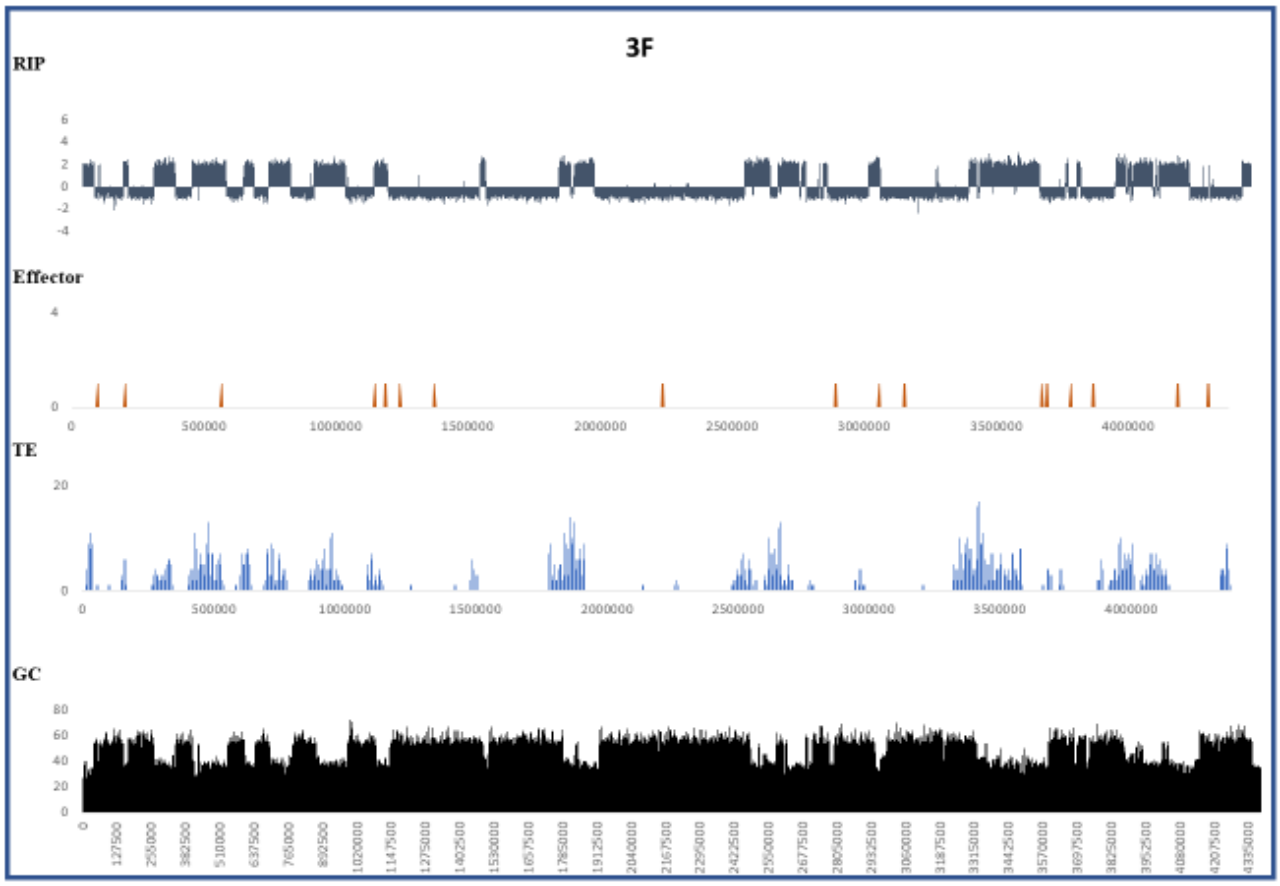
b



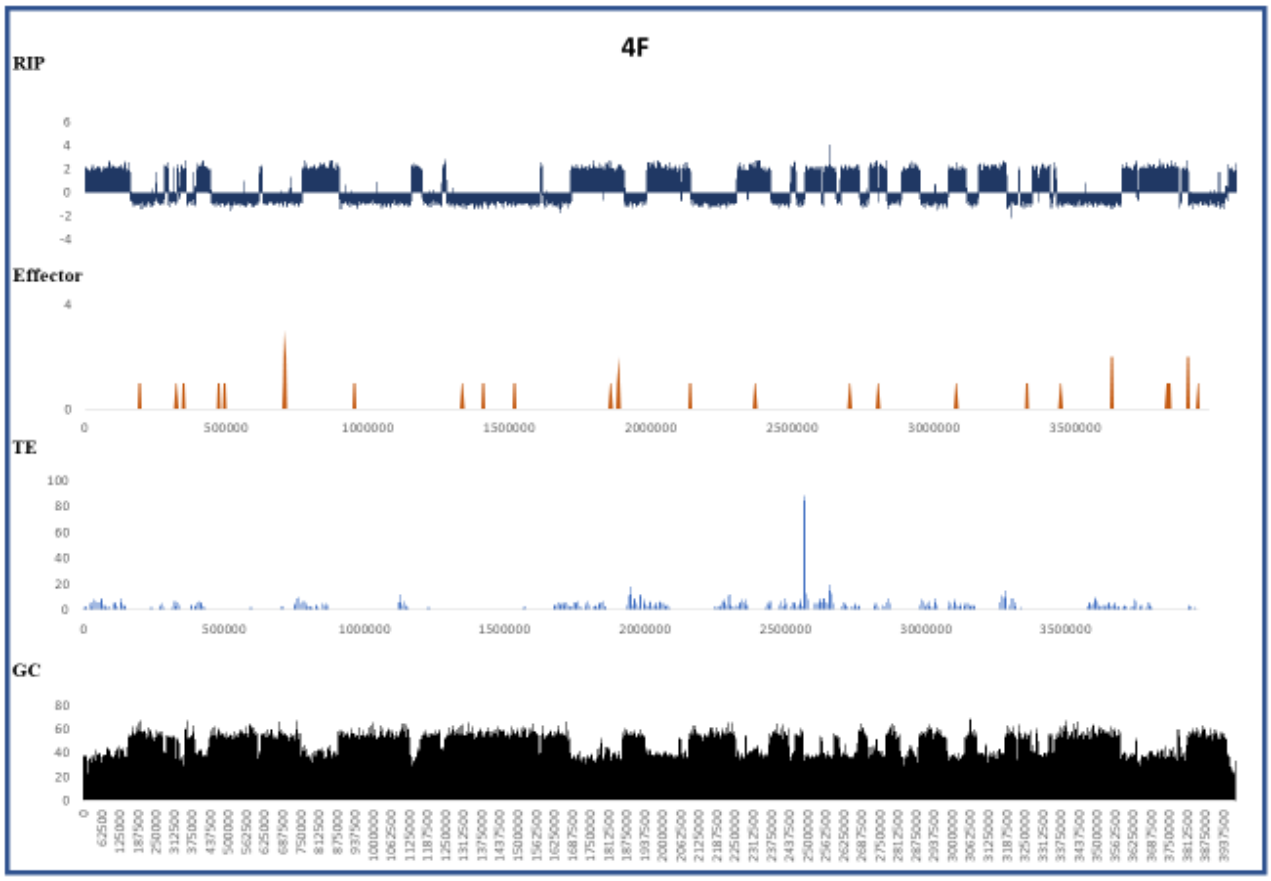
c



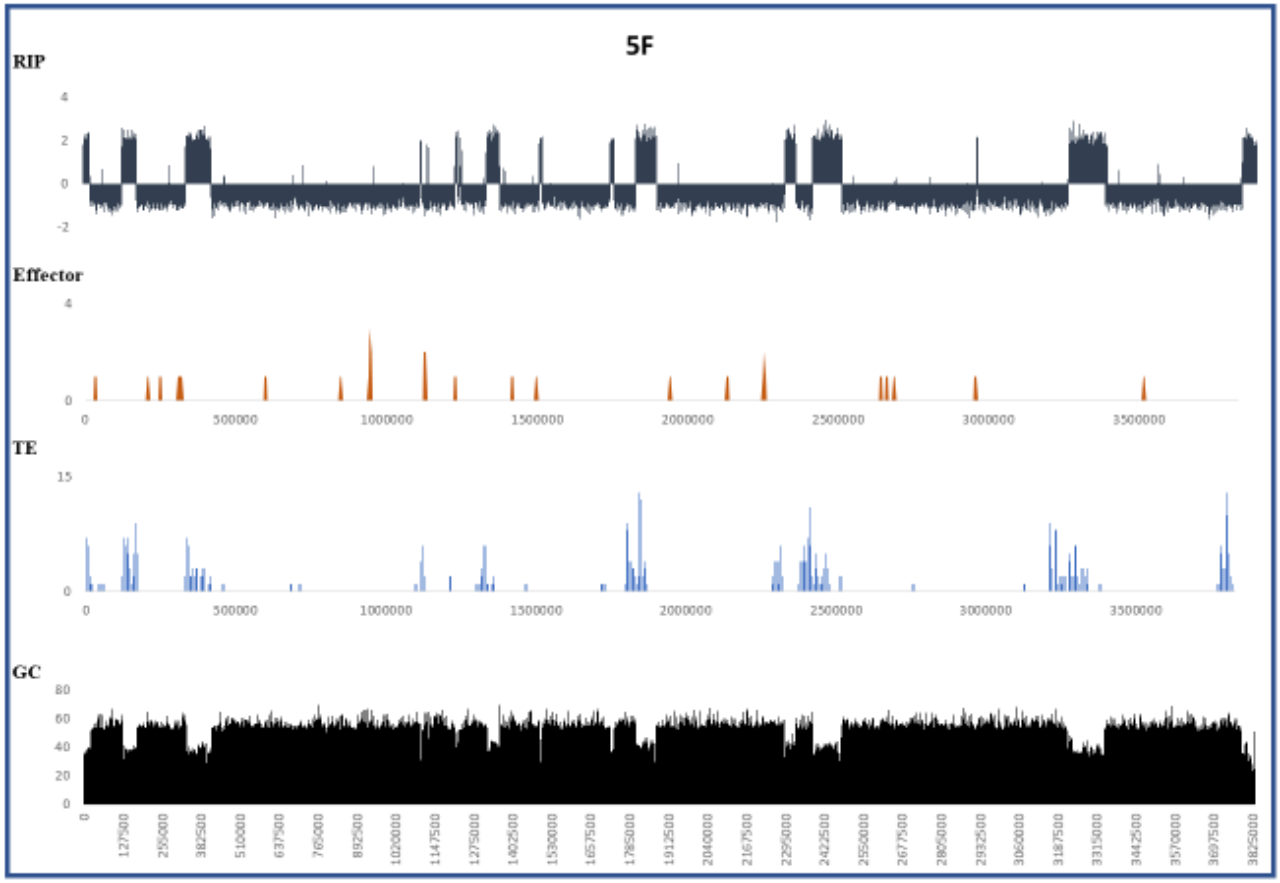
d



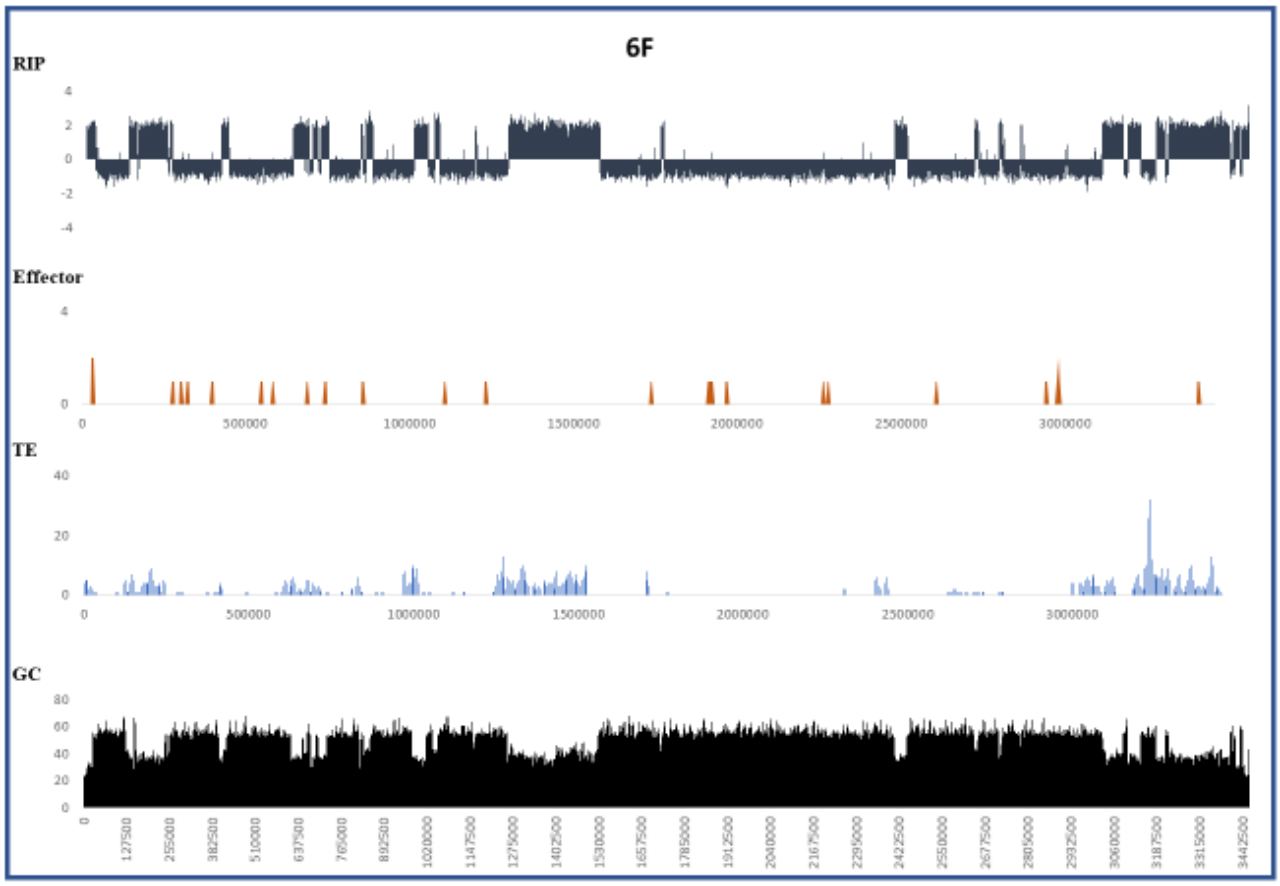
e



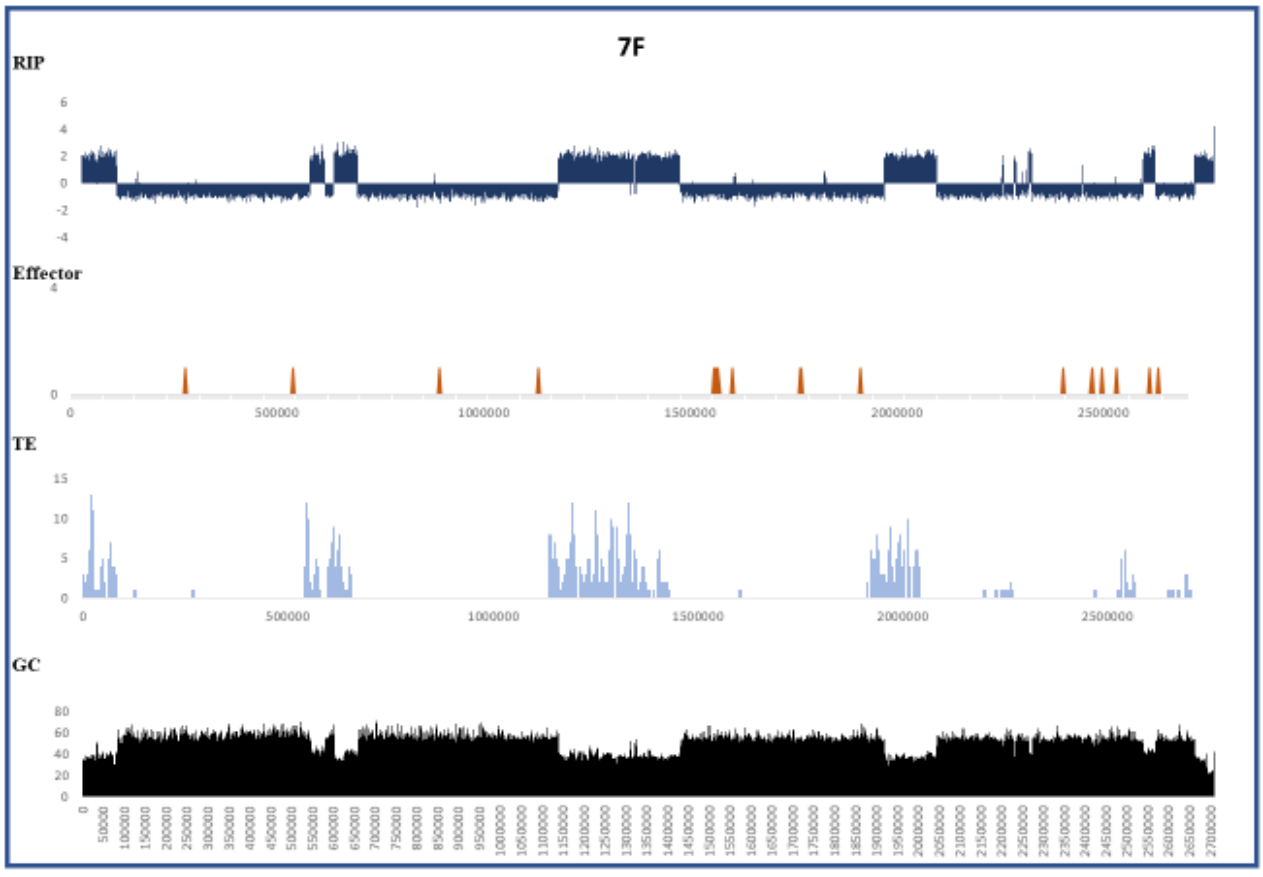
f



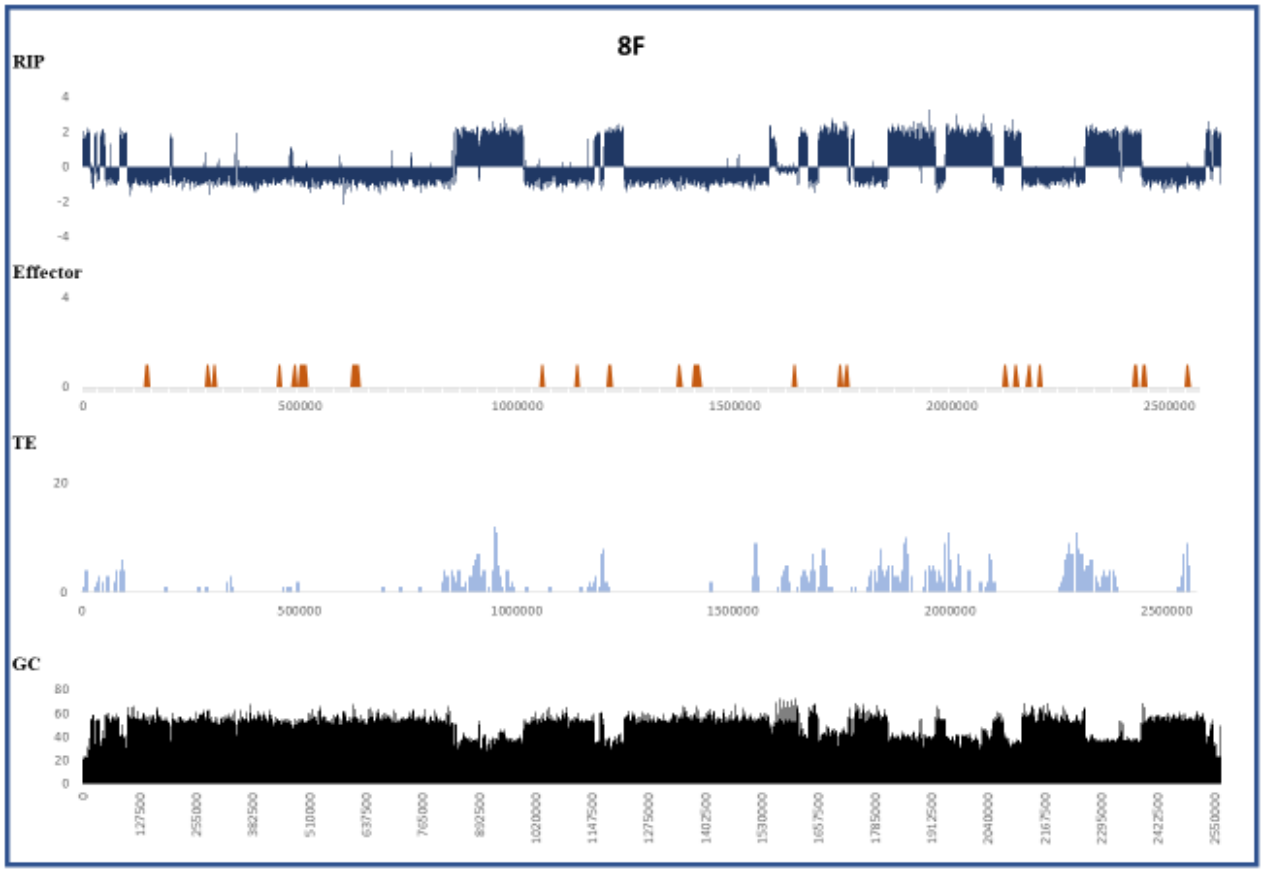
g



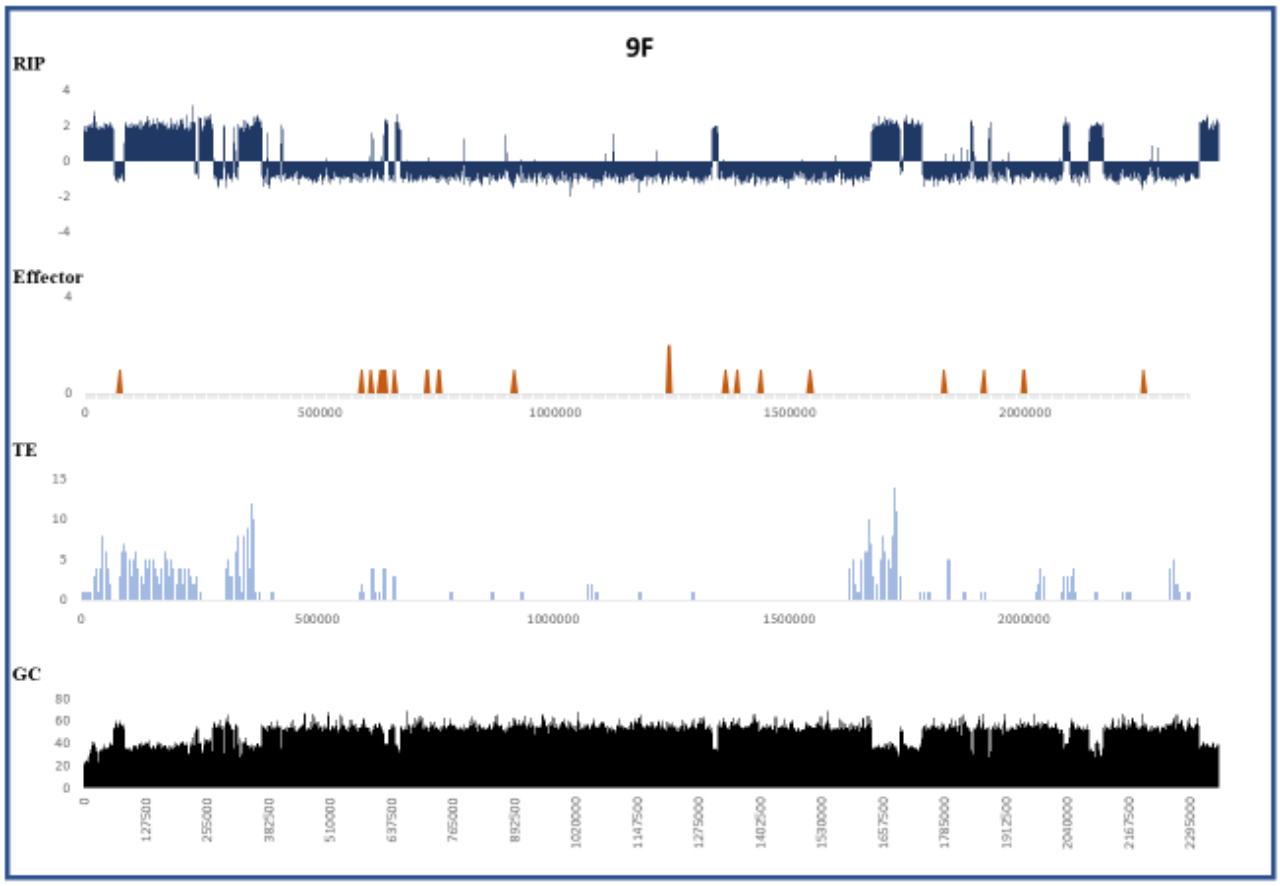
h



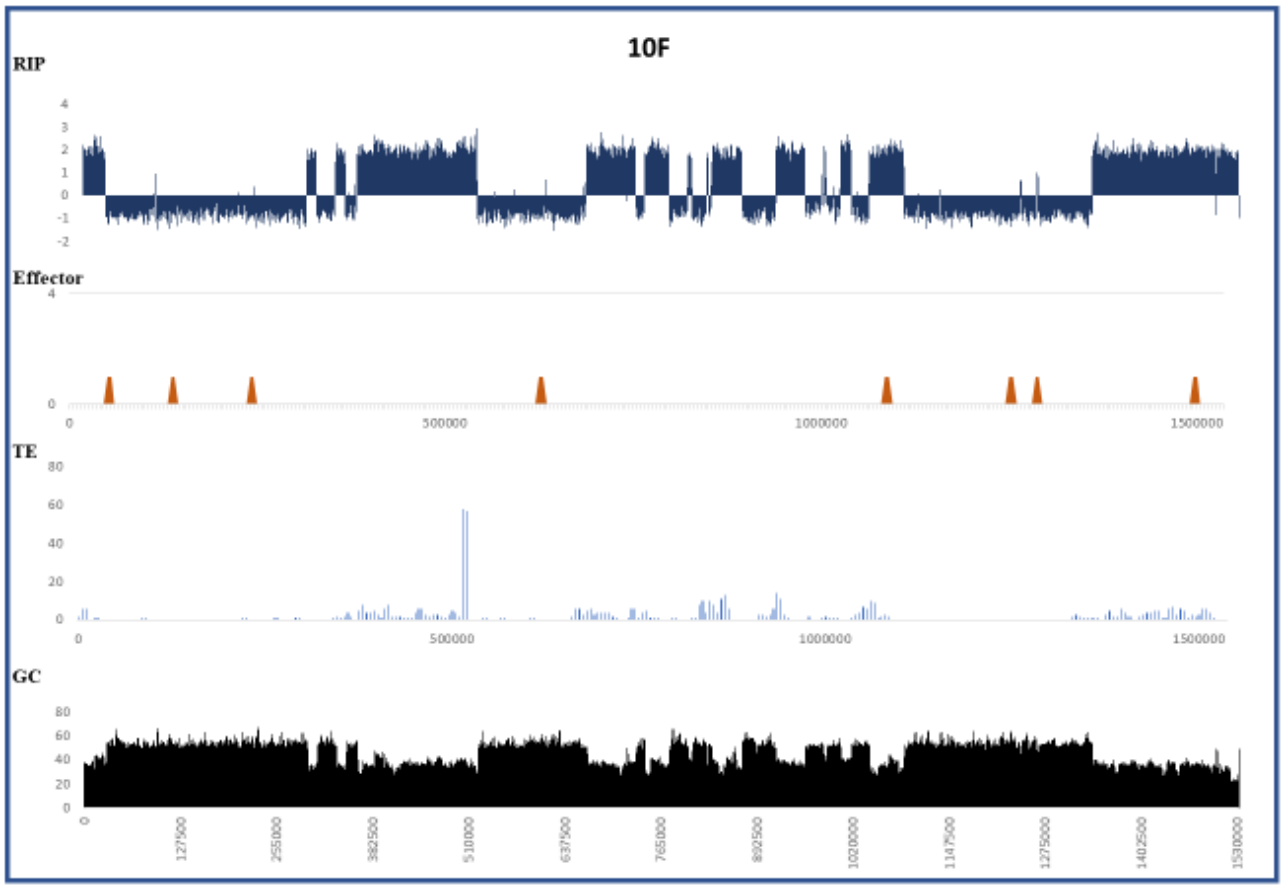
i



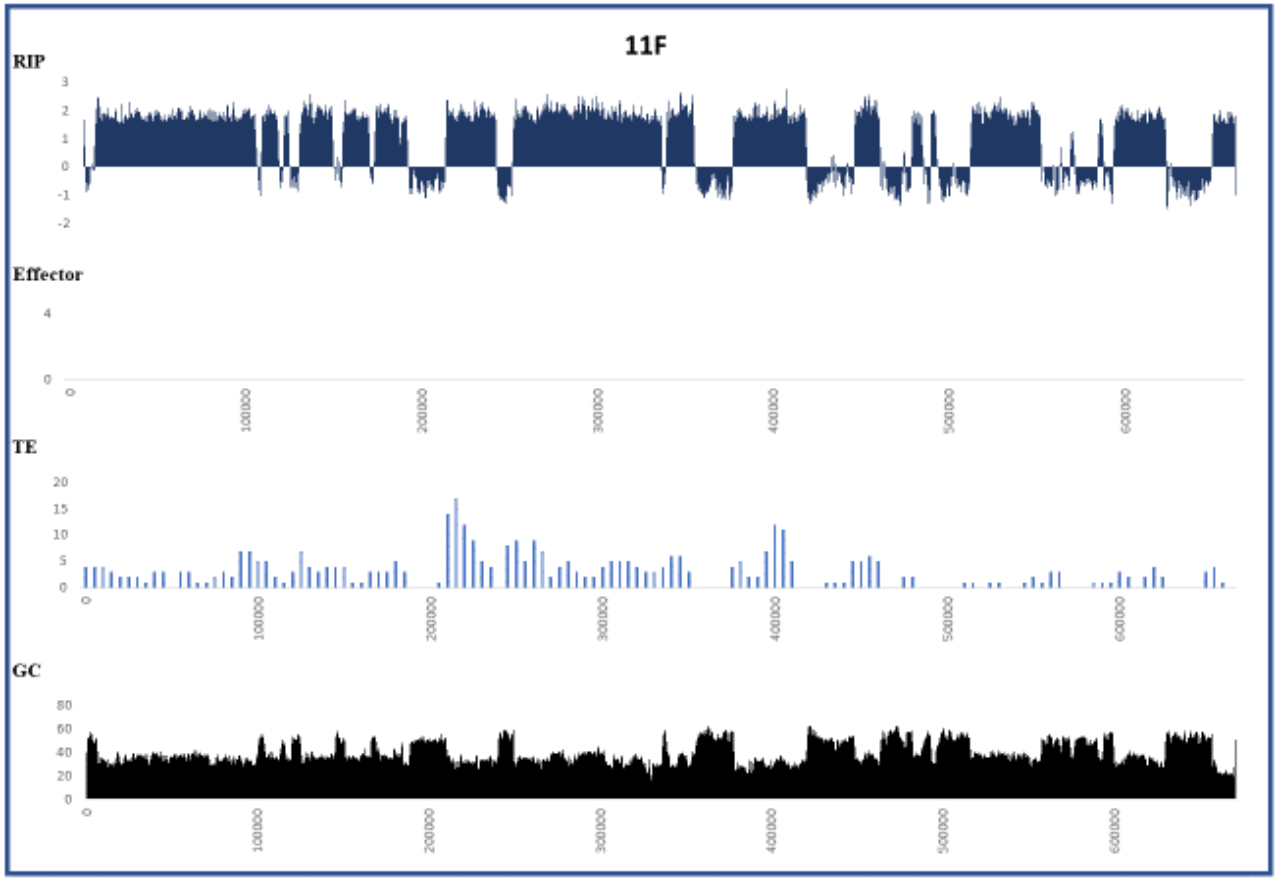
j



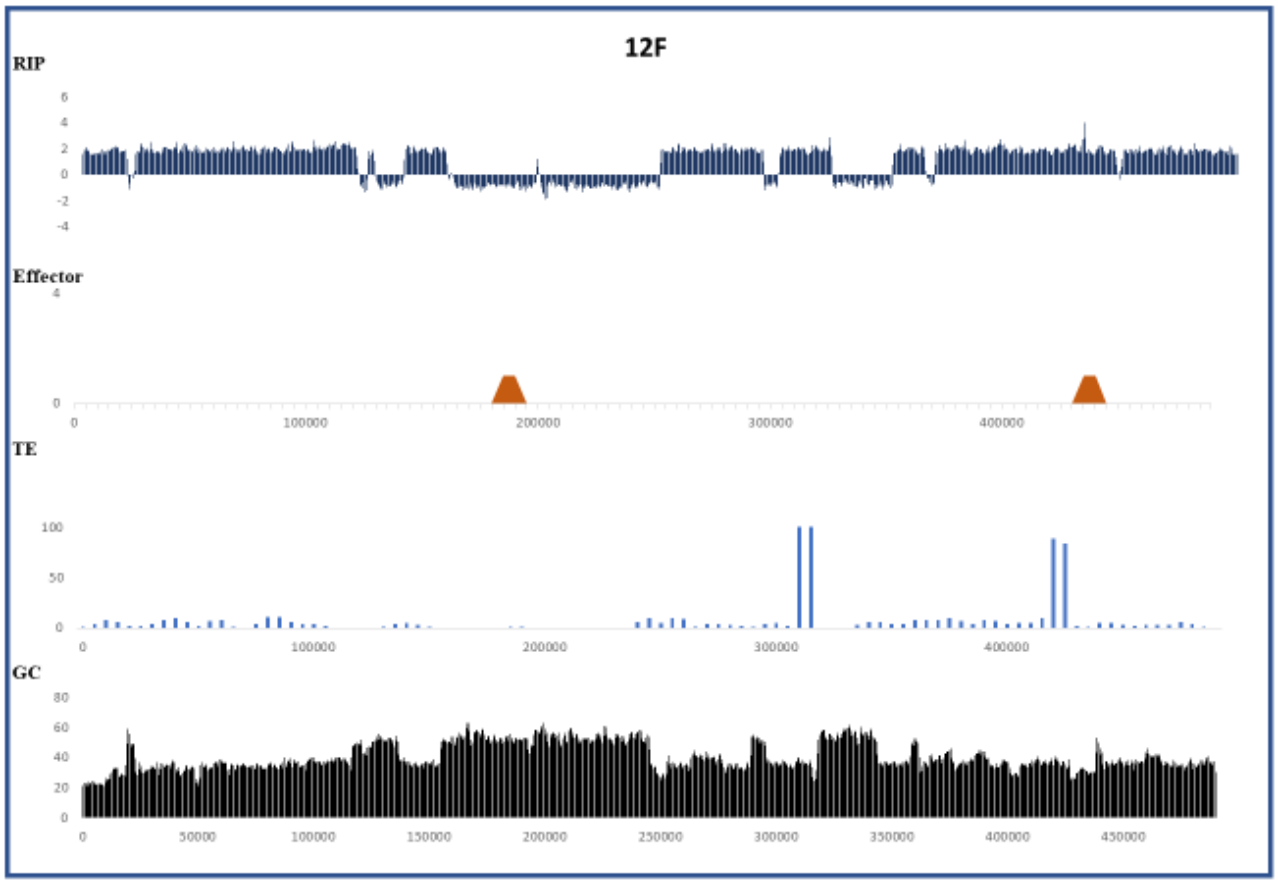
k



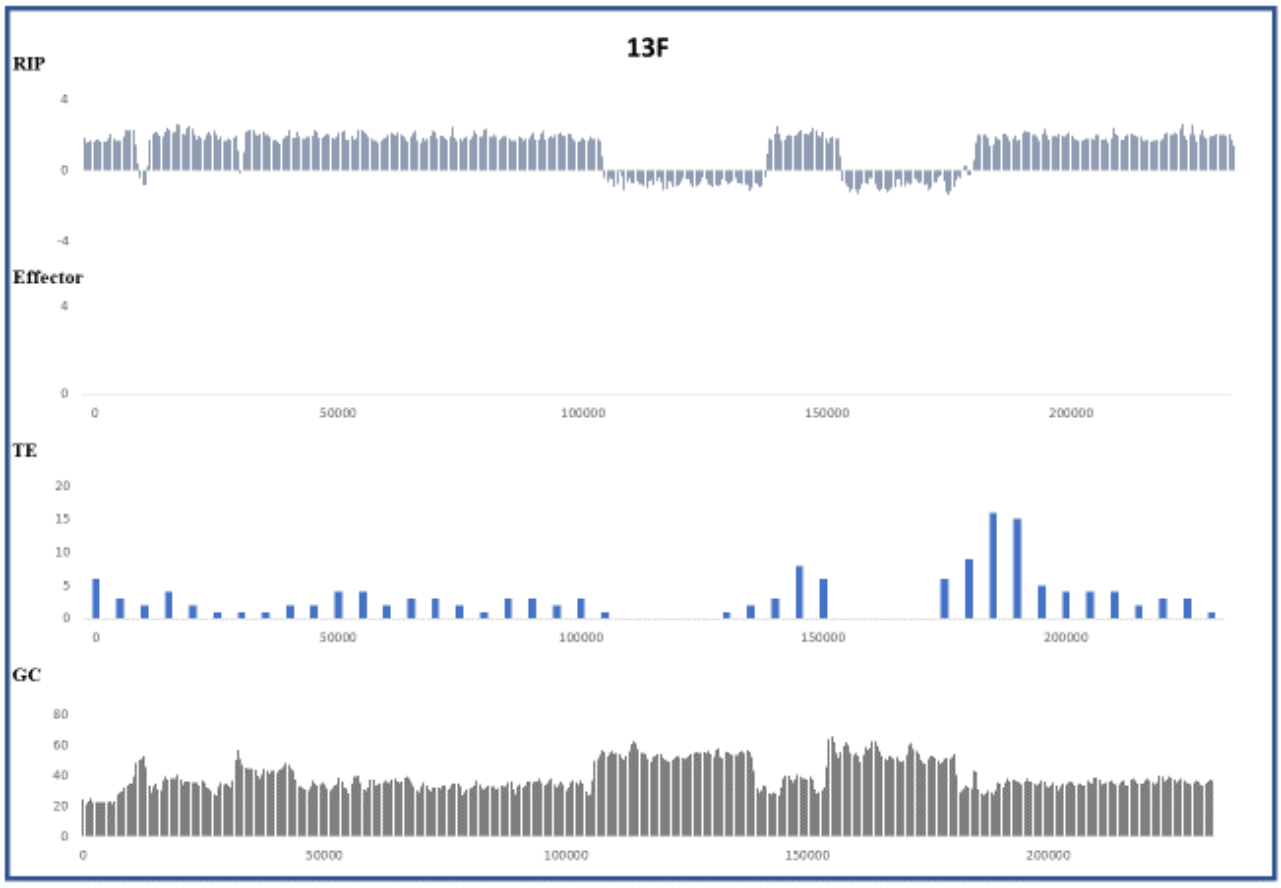
l



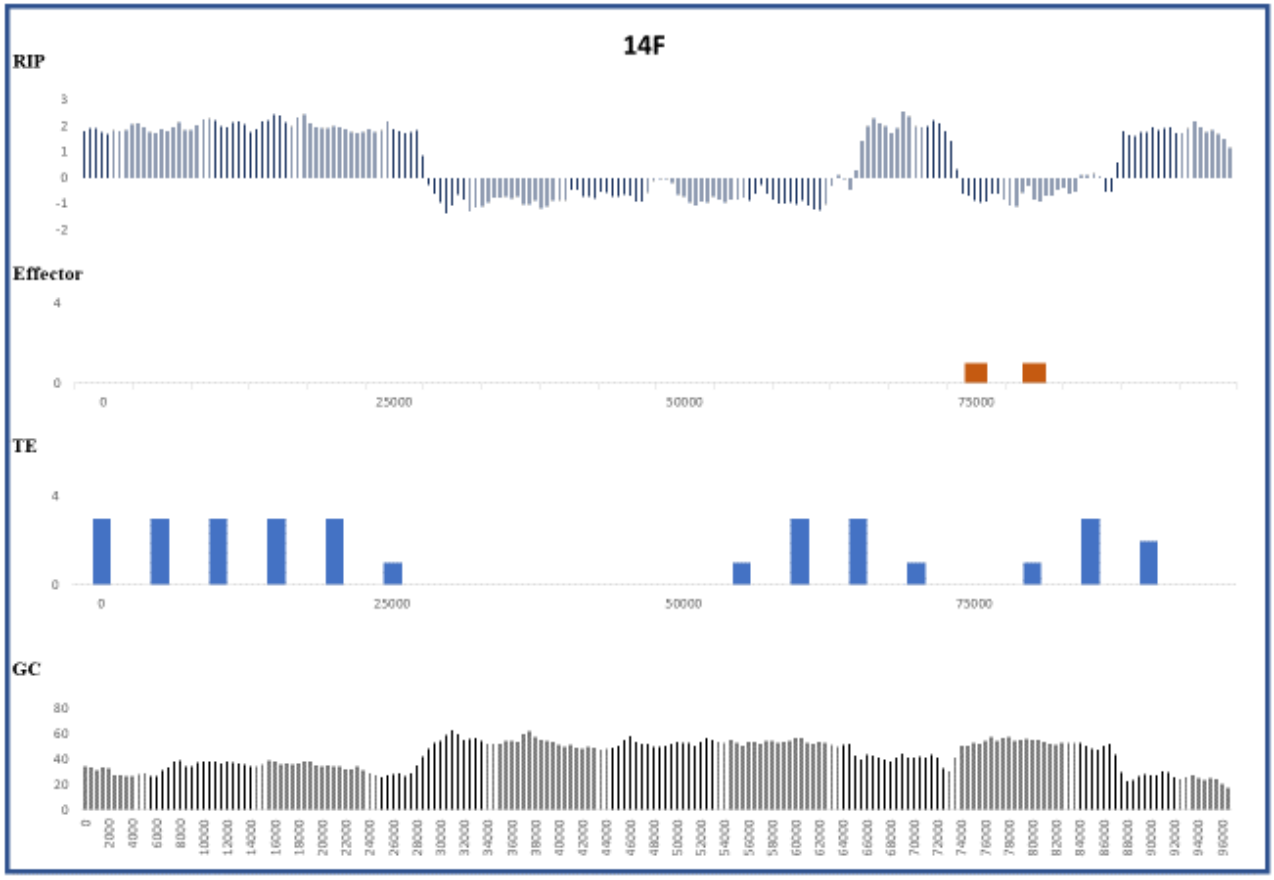
m



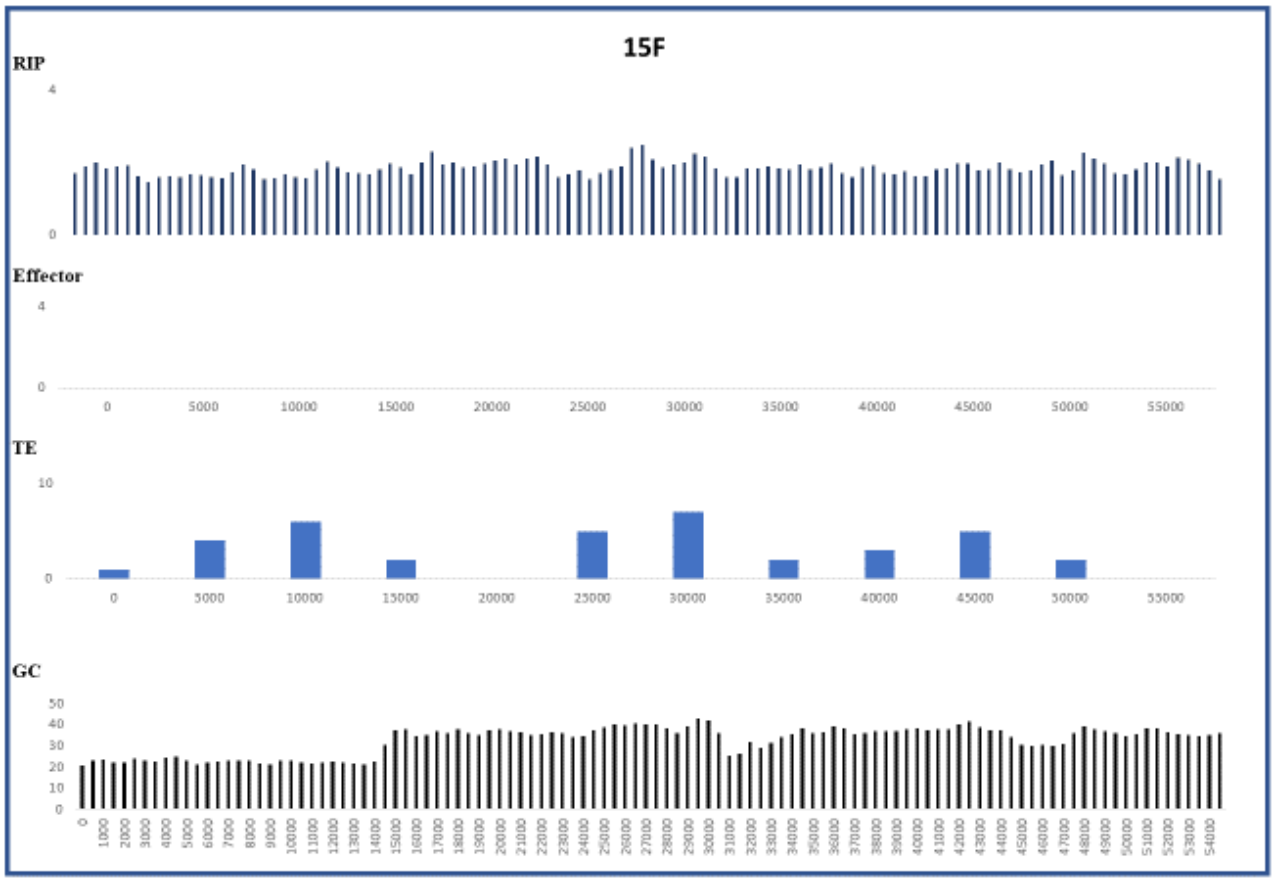
n



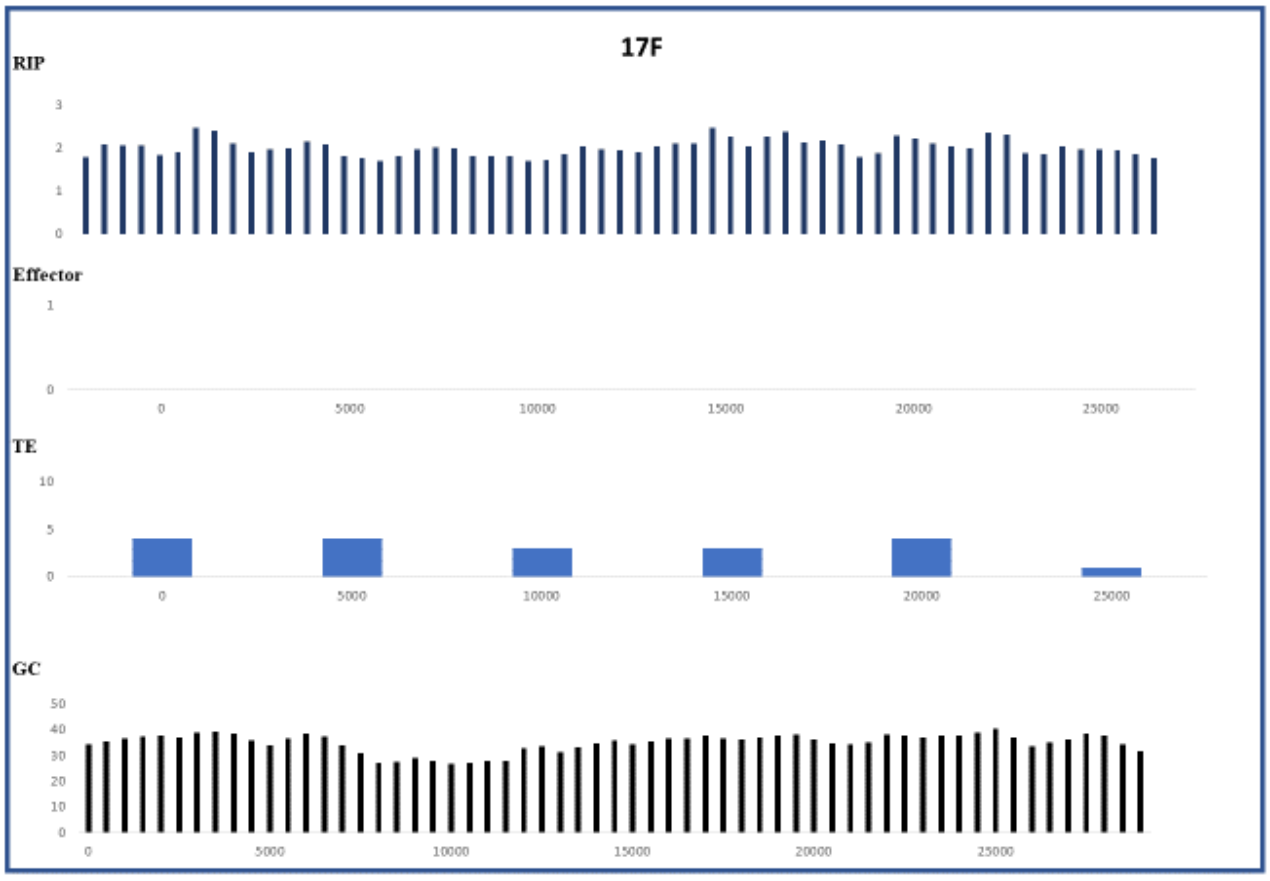
o



p



q



Appendix Fig. 3 Summary of the genetic features from The RIPper of the 17 nuclear genome contigs (**a**–**q**) of the *Cercospora zeina* PacBio assembly. The top panel depicts the changes in RIP composite index values across the length of chromosome using a 1000 bp window and increments of 500 bp. Values above 0 indicate RIP. The second panel depicts the distribution of putative effector genes across the length of the chromosome, 10 kbp and 5 kbp increments. The third panel depicts the distribution of TEs associated repeated sequences (class I and class II) across the length of the chromosome, (10 kbp and 5 kbp increments). The bottom panel illustrates the changes in GC content calculated 1000 bp and 500 bp increments.