CHAPTER 2

Identification and annotation of *Cercospora zeina Avr4*, *Ecp2* and *Ecp6* effectors

1. ABSTRACT

Most fungal effector proteins identified have been species-specific and shared no sequence similarity with any other known protein sequences. Recently, fungal effectors such as *Cladosporium fulvum Avr4*, *Ecp2*, and *Ecp6* have been shown to have homologs in other fungal species belonging to the Dothideomycete class. Various fungal effectors can show an even wider distribution, as seen with *C. fulvum Ecp2* and *Ecp6* that is present in some pathogenic and even some non-pathogenic fungal species. *Cercospora zeina* also belongs to the Dothideomycete fungal class and therefore recent availability of a draft *C. zeina* (Mkushi) genome sequence provided an opportunity to determine whether *C. zeina* effector homologs for *Avr4*, *Ecp2*, and *Ecp6* exist. Through BLAST searches it was shown that homologs of *Avr4*, *Ecp2*, and *Ecp6* are present in the *C. zeina* (Mkushi) genome. The *C. zeina Avr4*, *Ecp2*, and *Ecp6* gene structures were predicted by the gene prediction programmes Augustus and FGENESH, and the gene structures were validated by gene specific amplification and mapping of *in planta* and *in vitro* RNAseq reads to the predicted gene regions. Homologs of the predicted *C. zeina* effectors were also found to be present in *C. zeae-maydis*, *M. fijiensis, M. graminicola*, and *C. fulvum* through BLAST searches of the respective genome sequences. Conserved domain predictions indicated that the *C. zeina*, *C. zeae-maydis*, and *M. fijiensis* Avr4 proteins all have the conserved chitin-binding Peritrophin-A domain (CBM14), which has been shown to be involved in the binding of chitin present in fungal cell walls and protecting it against hydrolysis by host chitinases. The *C. zeina* Ecp2 contains a putative necrosis-inducing Hce2 conserved domain. The *C. zeina* and *M. fijiensis* Ecp6 proteins contain conserved chitin-binding LysM domains, which function by sequestering chitin oligomers to be able to avoid a host defence response. The presence of the *Avr4*, *Ecp2*, and *Ecp6* effectors in the *C. zeina* (Mkushi) genome together with their putative conserved domains provide insight into the possible roles that these proteins might play during maize infection.

2. INTRODUCTION

Effectors play an important role in plant-pathogen interactions. Effectors are the pathogen's strategic method that enables it to invade its host and to remain undetected within the host without triggering a host immune response. An identified model system for plant-pathogen interactions is the biotrophic fungus *Cladosporium fulvum* that causes leaf mould on tomato (*Solanum lycopersicum*) (THOMMA *et al.* 2005). This model system has been studied extensively and in the research process, effectors have been identified and characterized. The effectors relevant to this study are the *C. fulvum Avr4*, *Ecp2*, and *Ecp6*.

Most fungal effectors identified so far were species-specific, but some effectors such as *Avr4*, *Ecp2* and *Ecp6* were found to be conserved between fungi belonging to the Dothideomycete class or even across fungal species. It has been proposed that three classes of effector proteins exist based on the distribution of effectors between fungal species (STERGIOPOULOS *et al.* 2010). The first class includes effectors that are species-specific and do not share identity with any other known fungal proteins. Most effectors identified so far fall into this class. The second class of fungal effectors includes effectors that have homologs between fungal species belonging to the same class as seen with *C. fulvum Avr4* homologs identified in fungi belonging to the Dothideomycete class (STERGIOPOULOS *et al.* 2010). The third class of fungal effectors has widely distributed fungal homologs present in many fungal species as seen with *C. fulvum Ecp2* (STERGIOPOULOS *et al.* 2012) and *C. fulvum Ecp6* (DE JONGE and THOMMA 2009).

The predicted *C. fulvum* Avr4 effector protein is 135 amino acid residues in length. The mature Avr4 protein consists of 117 amino acids as a result of cleavage of the N-terminal signal peptide (SP) of 18 amino acids and contains eight cysteine residues and a chitin-binding Peritrophin-A (CBM14) functional domain (JOOSTEN *et al.* 1994; VAN DEN BURG *et al.* 2006). The Avr4 protein was found to be localized to the apoplastic space during infection where it binds to chitin present in the fungal cell walls and protects fungal hyphae against the enzyme activity of chitinases (VAN DEN BURG *et al.* 2006). In the presence of the *Cf-4* resistance gene, *C. fulvum* Avr4 effector induces a hypersensitive response (HR) (VAN DEN BURG *et al.* 2006). Homologs of *C. fulvum* Avr4 have been identified in *Cercospora apii* (CaAvr4), *Cercospora beticola* (CbAvr4), *Cercospora nicotianae* (CnAvr4), *Cercospora zeina* (CzAvr4), and *Mycosphaerella fijiensis* (MfAvr4) (STERGIOPOULOS *et al.* 2010). The MfAvr4 has been shown to be a functional ortholog of *C. fulvum* Avr4 and even though it shares low sequence identity with *C. fulvum* Avr4, MfAvr4 could be detected by the *Cf-4* resistance gene.

The predicted *C. fulvum* Ecp2 effector protein is 165 amino acid residues in size. The mature Ecp2 protein consists of 143 amino acids as a result of cleavage of the N-terminal SP of 22 amino acids and contains four cysteine residues and a putative necrosis inducing Hce2 domain (VAN DEN ACKERVEKEN *et al.* 1993). The Ecp2 effector is secreted into the apoplastic space during infection, but the exact function of Ecp2 still needs to be determined. *C. fulvum* Ecp2 has been shown to be an important virulence factor, as disruption of the effector lead to reduced virulence (VAN DEN ACKERVEKEN *et al.* 1993). Ecp2 causes an HR in the presence of the *Cf-Ecp2* resistance gene (VAN DEN ACKERVEKEN *et al.* 1993). Homologs identified for *C. fulvum* Ecp2 are present in *M. fijiensis* (Mf-Ecp2, Mf-Ecp2-2, and MfEcp2-3) and *Mycosphaerella graminicola* (Mg-Ecp2, Mg-Ecp2-2, and MgEcp2-3) (STERGIOPOULOS *et al.* 2010). The Mf-Ecp2 homolog has been shown to be able to trigger an HR in the presence of the *Cf-Ecp2* gene (STERGIOPOULOS *et al.* 2010). Ecp2 homologs (termed Hce2) form part of an ancient superfamily of putative effectors that are widely distributed across the fungal kingdom (STERGIOPOULOS *et al.* 2012).

The *C. fulvum* Ecp6 effector is 222 amino acid residues in length that includes an N-terminal SP of 23 amino acids. After cleavage of the N-terminal SP, the mature Ecp6 protein consists of 199 amino acids that includes eight cysteine residues and three chitin-binding LysM domains (DE JONGE and THOMMA 2009). *C. fulvum* Ecp6 is secreted into the apoplastic space during infection, where it functions by sequestering chitin oligomers to prevent possible detection and subsequent activation of the host defence mechanisms (DE JONGE and THOMMA 2009). The *C. fulvum* resistance gene for Ecp6 remains unknown. Fungal homologs of *C. fulvum* Ecp6 have been detected in a variety of pathogenic and non-pathogenic fungal species (DE JONGE and THOMMA 2009).

This research chapter aimed to identify if the *C. zeina Avr4*, and *C. fulvum Ecp2* and *Ecp6* effectors are present within the *C. zeina* (Mkushi) genome. Due to the conserved nature of the above mentioned effectors in Dothideomycetes and even across the fungal kingdom, the study hypothesis was that homologs of *Avr4*, *Ecp2* and *Ecp6* would be present in the *C. zeina* (Mkushi) genome. It was also hypothesized that if these homologs were present in *C. zeina* (Mkushi), they would contain similar functional domains as generally found in Avr4, Ecp2, and Ecp6 proteins. The search for putative homologs of *Avr4*, *Ecp2*, and *Ecp6* were also expanded to the genome of the sibling species of *C. zeina*, namely *Cercospora zeae-maydis*.

3. MATERIALS & METHODS

All reagents used in this study were obtained from Sigma (Aston Manor, South Africa) unless otherwise stated. Primers were synthesized by IDT (Coralville, Iowa) and sequencing reactions were done by Inqaba Biotec (Pretoria, South Africa). All centrifugation steps were done by using a bench top MiniSpin® centrifuge (Eppendorf, Hamburg, Germany), unless otherwise stated.

3.1 Available genomes and effector sequences

A draft genome sequence of an African isolate (Mkushi) of *C. zeina* (CMW25467) (MEISEL *et al.* 2009) has been sequenced through next generation Illumina sequencing (Prof. D. K. Berger, University of Pretoria 2011) and was available at the time of this study. The draft *C. zeina* (Mkushi) genome sequence was available in a BLAST interface format that was accessed from an internal server (http://sequenceserver.bi.up.ac.za). *C. zeae-maydis* filtered gene models (*C. zeae-maydis* V1.0), the *M. fijiensis* V2.0 genome, the *M. graminicola* V2.0 genome, and the *C. fulvum* V1.0 genome were accessible from the DOE Joint Genome Institute website (http://genome.jgi.doe.gov/). Gene sequences for the effectors *Avr4*, *Ecp2*, and *Ecp6* were obtained from the GenBank database accessed from the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The gene sequences with their accession numbers are summarized in Table 1.

Table 1. GenBank accession numbers for Dothideomycete *Avr4*, *Ecp2*, and *Ecp6* effector genes used in this study. Also included are the sizes of the respective gene sequences.

Effector gene	Accession number	Size (bp)	Citation
Avr4			
C. fulvum Avr4 gene (mRNA)	X78829.1	408	(JOOSTEN *et al.* 1994)
C. zeina Avr4 precursor gene (gDNA) (collection number: CBS118820)	GU574327.1	405	(STERGIOPOULOS *et al.* 2010)
C. apii Avr4 precursor gene (gDNA)	GU574326.1	408	(STERGIOPOULOS *et al.* 2010)
C. nicotianae Avr4 precursor gene (gDNA)	GU574325.1	405	(STERGIOPOULOS *et al.* 2010)
C. beticola Avr4 precursor gene (gDNA)	GU574324.1	408	(STERGIOPOULOS *et al.* 2010)
Ecp2			
C. fulvum Ecp2 gene (gDNA)	Z14024.1	554	(VAN DEN ACKERVEKEN *et al.* 1993)
Ecp6			
C. fulvum Ecp6 gene (mRNA)	EU730588.1	687	(BOLTON *et al.* 2008)

3.2 BLAST searches for the *Avr4*, *Ecp2*, and *Ecp6* effectors

Basic Local Alignment Search Tool (BLAST) searches (ALTSCHUL *et al.* 1990) against the available *C. zeina* (Mkushi) genome were made to determine if the effector genes *Avr4, Ecp2, and Ecp6* were present. BLASTn searches against *C. zeina* (Mkushi) genome were made through an internal BLAST interface (http://sequenceserver.bi.up.ac.za) that was developed by Mr F. van Staden at the University of Pretoria in 2012. BLAST searches were done by using the *C. zeina Avr4* (GU574327), *C. fulvum Ecp2* (Z14024) and *C. fulvum Ecp6* (EU730588.1) gene sequences. Reciprocal BLAST analysis involved BLASTn and tBLASTx searches against the NCBI database using the predicted *C. zeina* (Mkushi) gene sequences (*Avr4*, *Ecp2*, and *Ecp6*) as a query. The annotated *C. zeina* (Mkushi) *Avr4*, *Ecp2* and *Ecp6* gene sequences were used to perform BLAST searches (BLASTn and tBLASTx) to the *C. zeae-maydis* V1.0, *M. fijiensis* V2.0, *M. graminicola* V2.0, and *C. fulvum* V1.0 genome sequences available from the DOE Joint Genome Institute (www.jgi.doe.gov/). All BLAST results were obtained by using an expected (E) cut-off value of 1.0E-5 and BLOSUM as a scoring matrix. BLAST hits were assessed based on the percentage identity (identical bases between the query sequence and the BLAST hit alignment), the bit score (indicates how accurate the alignment is), and the expected value (E-value) (the statistical significance of the alignment). A higher bit score indicates a good alignment, and lower E-values indicate significant hits.

3.3 Gene predictions and annotations

The downloaded contigs from the BLAST results for the *C. zeina, C. zeae-maydis, M. fijiensis*, *M. graminicola,* and *C. fulvum Avr4*, *Ecp2*, and *Ecp6* effector genes were first trimmed (500 bp up and downstream of the aligned region) and then put through two gene prediction programmes (FGENESH and Augustus) to predict intron and exon boundaries by using an annotated genome of the model fungus (*Neurospora crassa*) as a reference. The Augustus prediction programme can be used in the form of a web interface (http://bioinf.uni-greifswald.de/augustus/) where the contig sequence (FASTA format) is uploaded and a reference genome is chosen to base the predictions on. The FGENESH prediction programme was downloaded (http://linux1.softberry.com). The ascomycete mould fungus (*N. crassa)* genome was chosen as a reference to base the predictions on as it is well-annotated, best-characterized among filamentous fungi, and serves as model for fungal genome annotations (DAVIS 2000). For each gene the generated predictions from both programmes were inspected and compared by using GenomeView software (http://genomeview.org). GenomeView allows you to visually inspect gene structure and organization over a stretch of DNA translated into three different reading frames. The

predicted gene regions (start/stop sites and intron/exon boundaries) can therefore also be manually annotated in GenomeView based on the structure of the predicted sequences and their homologs, and exported as a FASTA file. These annotated genes were then subjected to reciprocal BLAST (BLASTn and tBLASTx) analysis against the NCBI database. Conserved domains for the predicted *C. zeina* (Mkushi) Avr4, Ecp2, and Ecp6 proteins were determined through the conserved domain search interface from NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Protein sequences can be queried against a database containing well annotated multiple sequence alignment models for full protein sequences (MARCHLER-BAUER *et al.* 2013). The predicted protein sequences for each *C. zeina* (Mkushi) effector (Avr4, Ecp2, and Ecp6) were queried against the conserved domain database (CDD v.3.11) to determine if there were any conserved functional regions present in these sequences.

3.4 *In vitro* culturing of *C. zeina* and gDNA extractions

C. zeina cultures were initiated from a glycerol stock of a virulent *C. zeina* (Mkushi) (CMW25467) (MEISEL *et al.* 2009) strain (V2A) that was isolated through single conidial isolations from *C. zeina* infected maize leaves from a glasshouse trial in 2011 (Monique Heystek, MSc Dissertation, University of Pretoria, 2014). Culturing was done by plating 200 µl of the fungal stock onto V8 media (200 ml V8 juice, 2 g $CaCO_3$, 15 g bacteriological agar, and 800 ml double distilled water) (BECKMAN and PAYNE 1982). The plates were then stored at room temperature and in the dark for approximately five to seven days. After this time, the fungal conidia produced were transferred to new V8 media to encourage *C. zeina* conidia formation and prevent vegetative growth. The fungal conidia were transferred to the new V8 media plates by a process called pat culturing. Pat culturing involves cutting the media in half and then transferring the conidia to the new V8 plate by patting the cut half (containing the fungal conidia) onto the fresh V8 plate (B. Bluhm personal communication). Continuous pat culturing was done every five to seven days until the desired density of conidia was obtained. The final transfer of conidia was made to V8 media covered with a single layer of cellophane. Glycerol stocks of these *C. zeina* cultures were made by cutting out a block (approximately 0.5 cm x 0.5 cm in size) of conidiating fungal culture and preserving it by storing the fungal conidia in 10% glycerol and freezing it away at -80˚C.

Total genomic DNA from *C. zeina* (Mkushi) cultures was extracted by using a large-scale CTAB DNA extraction protocol (optimized by Prof. B. H. Bluhm, University of Arkansas). Briefly, the cellophane sheet containing the *C. zeina* (Mkushi) fungal growth was peeled from the V8 plate and weighed. Approximately 1.13 g of the sample (fungal material and

cellophane) was ground to a fine powder with liquid nitrogen, where after 10 ml cetyltrimethyl ammonium bromide (CTAB) buffer (1 ml 1 M Tris HCl (pH 8.0), 2.8 ml 5 M NaCl, 0.4 ml 0.5 M ethylenediaminotetraacetic acid (EDTA), 0.2 g CTAB, and ~5.8 ml distilled sterile water) with 0.4 g polyvinylpyrroilidone (PVP) and 50 µl β-Mercaptoethanol (BME) was added. The mixture was vortexed and incubated at 65˚C for at least one hour, mixing once every 30 minutes. 10 ml chloroform was added and mixed by inversion for one minute. The mixture was then centrifuged for 20 minutes at 10 000 rpm (12633 g). The aqueous phase was transferred to a new tube and the volume was estimated. Cold 7.5 M ammonium acetate (0.08 volumes) and 0.54 volumes of ice-cold isopropanol was added to the aqueous phase and mixed by inverting the tube 30 times. The aqueous phase mix was incubated at -20˚C for an hour and then centrifuged for 20 minutes at 10 000 rpm (12633 g). The supernatant was discarded and 1ml sterile water (65˚C) was added to the DNA pellet to dissolve the pellet. The tubes containing the pellet and sterile water were put in a water bath (65˚C) to quicken the process. RNAse A enzyme (5 µl) was added to the dissolved pellet and the solution was mixed by gently flicking the tube. The solution was incubated at 37˚C for 30 minutes. After the incubation step, 1 ml chloroform was added and mixed by inversion for one minute. The mixture was centrifuged for 10 minutes at 13 000 rpm (21350 g). The centrifugation step was repeated for clarity of the phase interface. The DNA was precipitated by adding two volumes of ice-cold 100% ethanol and was thoroughly mixed by inversion followed by an incubation step at -20˚C for 30 minutes. The mixture was centrifuged for five minutes at 13 000 rpm (21350 g) to pellet the DNA. The DNA pellet was washed by adding 750ul ice-cold 70% ethanol and inverting the tube four times. The mixture was then centrifuged again at 13 000 rpm (21350 g) for five minutes and the supernatant was discarded. The DNA pellet was washed once more by adding 750ul ice-cold 95% ethanol and inverting the tube four times. The centrifugation step was repeated again at 13 000 rpm (21350 g) for five minutes and the supernatant was discarded. The DNA pellet was air dried and resuspended in 50 µl of 1X TE buffer (10 mM Tris HCl (pH 8.0), 1mM EDTA (pH 8.0) and distilled sterile water). Extracted gDNA was quantified spectrophotometrically with a NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Waltham, Massachusetts). The gDNA quality was analysed by standard gel electrophoresis and gDNA samples were stored away at -20˚C until needed.

3.5 Primer design and amplification of effector genes

Gene-specific primers were designed for each predicted *C. zeina* effector gene (*Avr4*, *Ecp2* and *Ecp6*) by using CLC Bio Main Workbench software (Aarhus, Denmark). The primer design parameters were: primer length of 18-22 nucleotides, melting temperature between

50-60˚C, and GC content of 40-60%. The designed primer sets were subsequently analysed by using the OligoAnalyser 3.1 tool from IDT (Integrated DNA Technologies) (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/). Primer sequences designed for this study are summarized in Table 2. The primers were synthesized by IDT (Coralville, Iowa) and primer specificity was investigated through conventional PCR analysis. The PCR cycling conditions for all primer sets were as follows: denaturation at 94˚C for 5 min, followed by 35 cycles of (94˚C for 30 sec, 57/59˚C for 45 sec, and 72˚C for 2 min), and a final elongation step at 72˚C for 7 min. For PCR amplification, reaction volumes of 20 µl consisted of 1X NH_4 PCR reaction buffer, 0.2 mM dNTP mix (Bioline, London, UK), 1.5 mM $MgCl_2$, 0.3 µM of each of the primers, 0.025 U of BIOTAQ$_{TM}$ DNA Polymerase (Bioline, London, UK), and made up to 20 µl with sterile distilled water. *C. zeina* (Mkushi) gDNA template (10 ng) was used in each reaction, with PCR negative control reactions containing water instead of template. The PCR products were visualized by gel electrophoresis through agarose in a 1X TAE (Tris-acetate-EDTA) buffer.

Table 2. A summary of the gene-specific primers designed in CLC Bio to confirm the *C. zeina* (Mkushi) contig assemblies. The table includes primers sequences for each gene, primer names and sequences, and melting temperatures.

Target gene	Primer Name	Primer Sequence (5'-3')	Melting Temperature (˚C)
C. zeina Avr4	Avr4 UPS	TCCAAAACAAACAGCAGAACCA	57.54
	Avr4 F	ATGTTCGGCCTCTTCCACCTC	58.80
	Avr4 INT	TGGGTGATTGCGAAGTTC	56.06
	Avr4 R	TGTGAAAACGGGCAAGAATC	57.80
	Avr4 DWN	CTCTATCGTGTGCATATATTCCTC	57.09
C. zeina Ecp2	Ecp2 UPS	ACTCCTTCAACTTCCACC	56.31
	Ecp2 F	CTAGTTCGATGGGTTGTA	57.01
	Ecp2 INT	TTCAGTGTCCGCGTCATC	57.59
	Ecp2 R	ATGCTTTTCAACGTCGCT	58.06
	Ecp2 DWN	GTAGAAGAGAGTTTTGCAGAAG	56.63
C. zeina Ecp6	Ecp6 UPS	CATATGGAGACACGACCGA	55.61
	Ecp6 F	TTAAGAGGAGGTGATGTTGT	54.46
	Ecp6 R	ATGAAGTCTTACTTGTTCGC	55.49
	Ecp6 DWN	TTTATTCTCTGCCCTCGC	54.52
	Ecp6F2A	CGAGAGCAGGTTTGTGACA	59.96
	Ecp6F2B	AGGATTACTCGCCTGGGT	59.74
	Ecp6R2A	CCCATCTAACTACCCCGATCCTAA	59.40
	Ecp6R2B	GCGAAGGTTTGGTGCAATG	59.35

3.6 Gene annotation through mapping of RNAseq reads

RNAseq read mapping to gene regions serves as an effective way to validate and refine gene predictions and annotations. RNAseq mapping was done with the help of Mr N. Olivier from the University of Pretoria. RNAseq reads were mapped to the gene regions by using TopHat splice junction mapper (TRAPNELL *et al.* 2009) and viewed in GenomeView. The RNAseq read colours all represent different mapped reads. Green reads represent reads mapped to the forward strand from a sense fragment. Blue reads represent reads mapped to the reverse strand from a sense fragment. Cyan reads represent reads mapped to the reverse strand from an anti-sense fragment. Red reads represent reads that have a gap or deletion in the read. Purple/Pink reads represent reads that aligned over a splice junction and black reads indicate an insertion in the read. The *in planta* RNAseq reads were obtained from a pool of unmapped RNAseq reads that were generated as part of a maize eQTL project "Genomics of quantitative disease resistance in African maize varieties" from the Department of Plant Science, University of Pretoria. This project generated data from RNA isolated from sets of field infected maize plants, showing various levels of susceptibility to Grey Leaf Spot. The harvested time point was done at symptom development. RNA was sequenced and aligned to the maize genome. The pool of reads that did not align to the maize genome contained *C. zeina* transcripts. *In vitro* RNAseq data was generated by Miss V. Birkenbach from the University of Pretoria. The *in vitro* RNAseq data consisted of RNA extracted from *C. zeina* (Mkushi) grown on eight different *in vitro* growth conditions that was subsequently sequenced.

3.7 Sequencing *C. zeina* effector genes *Avr4*, *Ecp2*, and *Ecp6*

3.7.1 Clean-up of PCR products

PCR reactions were purified by using a Sephadex G-50 clean-up protocol. Sephadex slurry was prepared by adding 45 ml distilled water to 3 g Sephadex G-50 powder and microwaving the solution until dissolved. Sephadex cleaning columns were made by spinning the prepared Sephadex slurry (650 µl per eppendorf tube) in a bench-top centrifuge at 2000 rpm (505 g) for three minutes. The prepared column was then added to a new 1.5 ml eppendorf tube and the PCR product was gently pipetted onto the middle of the column and spun in the centrifuge at 2000 rpm (505 g) for three minutes. The purified PCR product was in the elute. After the PCR products were purified, the concentrations were determined spectrophotometrically with a NanoDrop™ 1000 (Thermo Scientific, Waltham, Massachusetts).

3.7.2 Cloning and sequencing

Cloning of purified PCR products was performed by using the InsTAclone™ PCR Cloning Kit (Thermo Scientific, Waltham, Massachusetts), following the manufacturer's specifications. Chemically competent cells were prepared from *E. coli* JM109 cells by means of a bacterial transformation kit included in the InsTAclone PCR cloning kit. Briefly, C-medium was inoculated with a single, fresh (not older than 10 days) bacterial (*E.coli* JM109) colony and incubated overnight at 37˚C with shaking. The 2 ml culture was enough for 26 transformations. Transformation reactions were carried out according to the manufacturer's specifications. Recombinant clones were identified based on blue/white screening which involved the plating out of transformed bacterial colonies on agar plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and isopropyl-β-D-thiogalactopyranoside (IPTG). Positively transformed colonies were white in colour, while negatively transformed colonies were dark blue in colour. Five positively transformed colonies were further analyzed through colony PCR. The cloning kit makes use of the pTZ57R/T cloning vector (Appendix, Figure S1) and insert DNA can be amplified by using standard M13/pUC primers. The PCR cycling conditions for the M13/pUC primer set were as follows: denaturation at 94˚C for 2 min, followed by 30 cycles of (94˚C for 30 sec, 55˚C for 30 sec, and 72˚C for 1 min), and a final elongation step at 72˚C for 7 min. For PCR amplification, reaction volumes of 20 µl consisted of 1X NH_4 PCR reaction buffer, 0.2 mM dNTP mix (Bioline, London, UK), 1.5 mM $MgCl_2$, 0.3 µM of each of the primers, 0.025 U of $BIOTAQ_{TM}$ DNA Polymerase (Bioline, London, UK), and 15,7 µl sterile distilled water. The PCR products were visualized by standard gel electrophoresis to determine the presence and expected length of each product. Cloned products were sequenced as a further step of validation. Plasmid extractions for sequencing purposes were done by using the GeneJET™ Plasmid Miniprep (Thermo Scientific, Waltham, Massachusetts). Extractions were done according to the manufacturer's specifications. Extracted plasmid samples were analyzed with a Nano-Drop™ 1000 spectrophotometer (Thermo Scientific, Waltham, Massachusetts) to determine the concentrations and also analysed by standard agarose gel electrophoresis to determine the product size and to confirm product presence after the extraction protocol. Plasmid samples with a concentration of at least 100ng/µl were used for sequencing reactions. Sequencing reactions were done by Inqaba Biotec with standard M13F and M13R primers. Forward and reverse sequencing reactions were done for each sample. Sequenced products were analysed and aligned by using CLC Bio Main Workbench software (Aarhus, Denmark).

3.8 Alignments and phylogenetic analysis

Protein alignments for the Avr4, Ecp2, and Ecp6 protein homologs were generated by using Mafft software (http://www.ebi.ac.uk/Tools/msa/mafft/). The *C. fulvum* Avr4, Ecp2, and Ecp6 protein sequence accession numbers were CAA55403.1, CAA78401.1, and ACF19427 respectively, which corresponds to the *C. fulvum* gene accession numbers X78829.1, Z14024.1, and EU730588.1 respectively. The Mafft alignment was saved as a phylip4 file and visualized and edited in CLC Bio Main Workbench software (Aarhus, Denmark). Pairwise comparisons of sequence identity and similarity of the aligned protein sequences were determined with CLC Bio Main Workbench software (Aarhus, Denmark) and EMBOSS Needle software (www.ebi.ac.uk). Sequence identity was based on the number of amino acids that matched exactly in the alignment, whereas the sequence similarity also took chemically similar amino acid groups into account. For the identity and similarity calculations, gaps were not counted and measurements were relational to the shorter of the two proteins compared. Nucleotide alignments were generated by MUSCLE alignment software that is incorporated in MEGA5 software package (TAMURA *et al.* 2011). The best substitution model for the phylogenetic analysis was predicted through MEGA5 software, and the prediction was based on the nucleotide alignment. The best substitution model was incorporated during the calculation of the phylogenetic trees. Neighbourhood-joining (NJ) and Maximum Likelihood (ML) analysis was done for each nucleotide alignment. Statistical significance of the phylogenetic groupings was calculated with 1000 Bootstrap values and indicated as percentages on the nodes of the combined NJ/ML trees. Bootstrap values greater than 75% were considered statistically significant.

4. RESULTS

4.1 Evidence for the presence of the *Avr4* effector homolog in *C. zeina*

4.1.1 The presence of an *Avr4* homolog in the *C. zeina* (Mkushi) genome

The *C. zeina Avr4* query gDNA sequence (GU574327, culture collection number: CBS118820) (Table 1) was used for BLASTn searches against the draft *C. zeina* (Mkushi) genome. The available *C. zeina* databases included Illumina data assembled with Abyss or Velvet and *C. zeina* (Mkushi) 454 sequencing data assembled with Newbler. BLASTn results showed a higher percentage identity as well as better statistical significance when the *C. zeina* (GU574327) sequence was queried against the *C. zeina* Illumina data assembled with Velvet compared to the other *C. zeina* datasets. The BLAST hit (putative *C. zeina* (Mkushi) *Avr4*) obtained shared 76% identity to the *C. zeina Avr4* sequence (GU574327) with a 209 bp aligned region showing six gaps, a score of 220, and a statistical significance of 2e-50. The gaps seen in the *Avr4* alignment did not affect the alignment of downstream regions (results not shown).

4.1.2 *C. zeina* (Mkushi) *Avr4* preliminary predicted gene structure

Although a putative *Avr4* homolog was found to be present in the *C. zeina* (Mkushi) genome, the *Avr4* homolog sequence needed to be annotated in order to determine the *C. zeina* (Mkushi) *Avr4* gene structure in terms of intron and exon boundaries. The expected gene structure of the preliminary *C. zeina* (Mkushi) *Avr4* gene showed similar predictions when analysed by both Augustus and FGENESH gene prediction programmes. The preliminary *C. zeina* (Mkushi) *Avr4* homolog was predicted to be 476 bp long and to contain a 98 bp intron (Figure 1), which is different to the *C. zeina Avr4* (GU574327) gene sequence that is 405 bp long with no introns (STERGIOPOULOS *et al.* 2010).

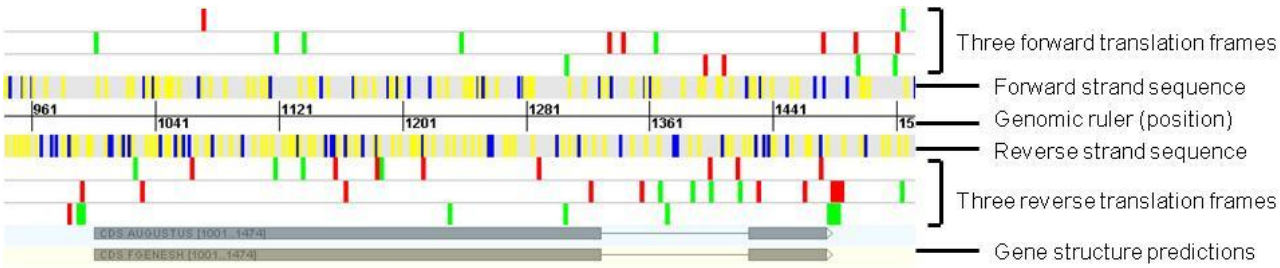


Figure 1. *C. zeina* (Mkushi) *Avr4* preliminary gene prediction visualised in GenomeView comparing the expected gene structures and orientation as predicted by Augustus and FGENESH. The annotations made by both prediction programmes were similar. The grey bar at the bottom of the figure represents the prediction made by FGENESH, whereas the blue bar represents the prediction made by Augustus. Both predictions were present on the forward strand. The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. The blue and yellow vertical lines represent possible acceptor and donor sites for splicing in the forward and reverse sequence.

Reciprocal BLASTn and tBLASTx analysis against the NCBI database indicated that the closest homolog for the preliminary predicted *C. zeina* (Mkushi) *Avr4* gene was still the *C. zeina Avr4* gene (GU574327). The preliminary predicted *C. zeina* (Mkushi) *Avr4* gene shared 76% identity (209 bp alignment region) to the *C. zeina Avr4* gene (GU574327) and the translated *C. zeina* (Mkushi) *Avr4* gene shared 63% identity (120 amino acid alignment region) to the translated *C. zeina Avr4* gene (GU574327) (Table 3). Other significant BLASTn and tBLASTx results including hits to other *Cercospora* species, the *M. fijiensis* carbohydrate-binding module family 14 protein and *C. fulvum* Avr4 are also summarized in Table 3.

Table 3. A summary of the most relevant BLASTn and tBLASTx results obtained from reciprocal BLAST analysis of the preliminary predicted *C. zeina* (Mkushi) *Avr4* gene.

BLASTn Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
C. zeina Avr4	(GU574327)	76	199	2e-47	6	290 bp
C. nicotianae Avr4	(GU574325.1)	75	196	2e-46	6	290 bp
C. apii Avr4	(GU574326.1)	75	190	2e-45	6	290 bp
C. beticola Avr4	(GU574324.1)	75	190	9e-45	6	290 bp
M. fijiensis carbohydrate-binding module family 14 protein	(XM007927856.1)	79	73.4	2e-09	2	94 bp
tBLASTx Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
C. zeina Avr4	(GU574327)	63	361	5e-39	-	120 aa
C. nicotianae Avr4	(GU574325.1)	63	356	3e-38	-	120 aa
C. beticola Avr4	(GU574324.1)	63	331	8e-35	-	112 aa
C. apii Avr4	(GU574326.1)	62	329	1e-34	-	112 aa
M. fijiensis carbohydrate-binding module family 14 protein	(XM007927856.1)	68	328	2e-34	-	76 aa
C. fulvum Avr4	(X78829.1)	54	107	6e-12	-	35 aa

The preliminary *C. zeina* (Mkushi) Avr4 protein was predicted to be 126 aa in size and has a conserved chitin-binding domain named the Chitin Binding Peritrophin-A domain (CBM14) (pfam01607). The CBM14 domain is located at the 45-103 amino acid region of the *C. zeina* Avr4 protein (Figure 2). CBM14 chitin-binding domains are extracellular domains and are found in chitin binding proteins (CASU *et al.* 1997; ELVIN *et al.* 1996; SHEN and JACOBS-LORENA 1998).

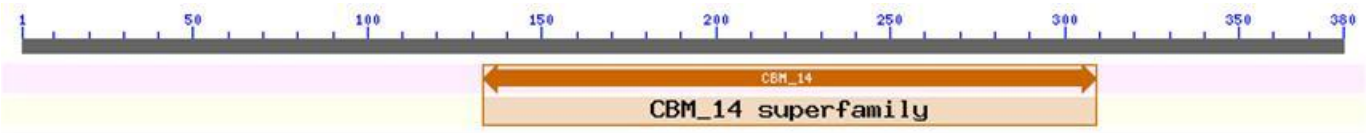


Figure 2. **Conserved chitin-binding domain (CBM14) present in the preliminary *C. zeina* (Mkushi) Avr4 protein.** The preliminary predicted *C. zeina* (Mkushi) *Avr4* has a conserved Chitin Binding Peritrophin-A extracellular domain (CBM14) which is found in chitin-binding proteins. The grey bar represents the predicted *C. zeina* (Mkushi) *Avr4* query sequence and the orange bar represents the region of the conserved domain. The figure is drawn to scale.

4.1.3 Confirmation of *C. zeina* (Mkushi) *Avr4* contig assembly

4.1.3.1 PCR analysis with gene-specific primers

Although the preliminary *Avr4* homolog was identified and annotated in the *C. zeina* (Mkushi) genome sequence, there was still a possibility that the contig might not have been assembled correctly. Polymerase chain reaction (PCR) analysis with gene-specific primers was chosen to verify the *C. zeina* (Mkushi) contig assembly around the *Avr4* effector gene. The gene-specific primers for the preliminary predicted *C. zeina* (Mkushi) *Avr4* effector gene sequence was designed in such a way that different fragments of known length would be amplified from *C. zeina* (Mkushi) gDNA in order to confirm the contig assembly (Figure 3). The *C. zeina* (Mkushi) *Avr4* primer sequences are summarised in Table 2.

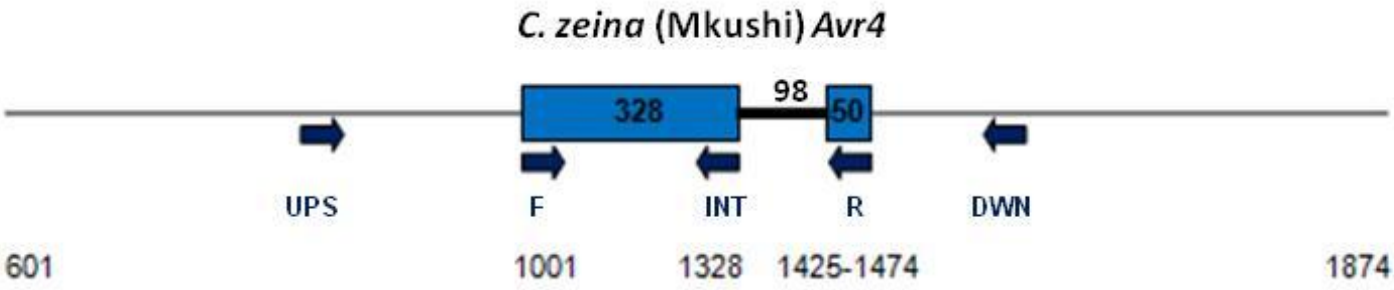


Figure 3. Gene-specific primers designed to the preliminary predicted *C. zeina* (Mkushi) *Avr4* gene. Designed primer regions are indicated by the navy arrows. The blue boxes indicate exons and the bold black line indicates an intron region. The respective sizes of the introns and exons were indicated in the figure. The figure was not drawn to scale.

Specific gene regions of the preliminary predicted *C. zeina* (Mkushi) *Avr4* gene were amplified from *C. zeina* (Mkushi) gDNA through PCR. Cellophane sheets provided an effective way to remove maximum conidia from the V8 media for *C. zeina* gDNA extraction purposes. gDNA extractions from *in vitro* grown cultures produced pure, good quality gDNA (results not shown). The amplified PCR products were analysed on a 1.3% agarose gel and

the product lengths could be determined. For the *Avr4* gene-specific amplifications, six amplicons were amplified (Figure 4). These *Avr4* amplicons were amplified using the following primers combinations: (1) UPS and DWN primers, (2) UPS and R primers, (3) UPS and INT primers (4) F and INT primers, (5) F and R primers, (6) F and DWN primers. The expected amplicon sizes of 953 bp, 863 bp, 727 bp, 328 bp, 464 bp, and 554 bp were obtained for the respective primer combinations (Figure 4). Because these regions could all be amplified from *C. zeina* (Mkushi) gDNA, the *Avr4* contig assembly was assumed to be correct.

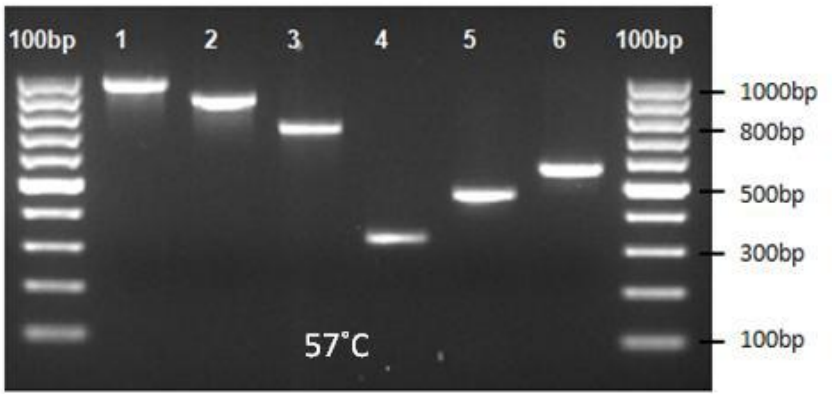


Figure 4. *C. zeina* (Mkushi) *Avr4* contig confirmation with gene-specific primers.
The *C. zeina* (Mkushi) *Avr4* contig was verified through PCR amplification with *Avr4*-specific primers. The subsequent amplification products were analysed on a 1.3% agarose gel. All products were amplified from *C. zeina* (Mkushi) genomic DNA and a 100 bp DNA ladder (Thermo Scientific, Waltham, Massachusetts) was used to measure product sizes. *Avr4* products amplified at 57˚C with UPS and DWN primers (1), UPS and R primers (2), UPS and INT primers (3), F and INT primers (4), F and R primers (5), F and DWN primers (6). Expected amplicon sizes for *Avr4* were 953 bp (1), 863 bp (2), 727 bp (3), 328 bp (4), 464 bp (5), and 554 bp (6).

4.1.3.2 *C. zeina* (Mkushi) *Avr4* annotation through mapping of RNAseq reads

In addition to PCR analysis with gene-specific primers, mapping of RNA sequence (RNAseq) reads to predicted contigs or gene regions is another way to validate and improve gene predictions and annotations based on empirical evidence. *In vitro* and *in planta* RNAseq reads were available from two different studies in our research group as outlined in the materials and methods section. Mapping of the RNAseq reads was done to the full *Avr4* contig obtained from the BLAST results from the *C. zeina* (Mkushi) draft genome. RNAseq reads were mapped to full contigs instead of trimmed contigs, to avoid possible forced reads. Improved annotations based on the empirical evidence were made as the initial annotations were based solely on computational evidence.

The *in planta* RNAseq reads that mapped to the *Avr4* gene region of the contig are shown in figure 5. The annotation of *Avr4* was adjusted by using the mapped RNAseq reads as a reference as well as the *C. zeina Avr4* sequence (GU574327) that is the closest homolog of the preliminary predicted *C. zeina* (Mkushi) *Avr4* gene. The *in vitro* (results not shown) and *in planta* RNAseq reads that mapped to the *Avr4* gene region both supported the same improved annotation. The annotated *Avr4* gene consisted of a single exon of 336 bp, which was smaller than the initial preliminary prediction and does not contain an intron sequence.

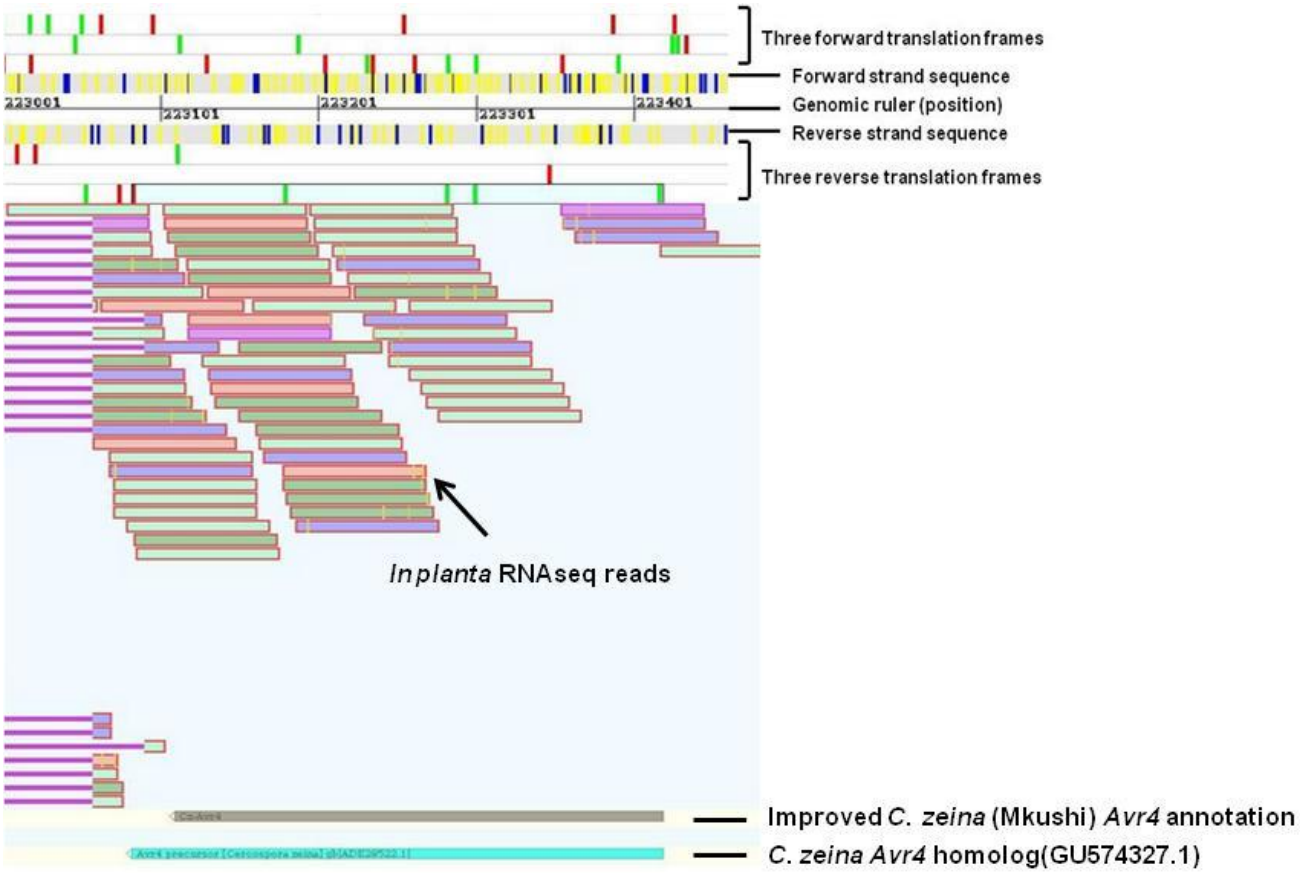


Figure 5. The improved *C. zeina* (Mkushi) *Avr4* annotation.
In planta RNAseq reads were mapped to the contig containing the predicted *C. zeina* (Mkushi) *Avr4* gene. Only the region containing the annotated *Avr4* gene is shown. The coloured bars represent the *in planta* reads mapped to the whole contig. The *C. zeina Avr4* gene homolog (GU574327) was used as a reference for annotation purposes and is represented by the light blue horizontal bar. The annotation of the *Avr4* gene was done manually by taking into account possible translation start and stop sites, the *C. zeina Avr4* homolog, and the mapped RNAseq reads. The improved *Avr4* annotation is represented by the grey horizontal bar and is shorter than the initial preliminary *C. zeina Avr4* gene prediction (not shown). The *in planta* RNAseq reads does not support an intron region. The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. *Avr4* is present on the reverse strand. See materials and methods for RNAseq read colours and their descriptions.

The annotated *C. zeina* (Mkushi) *Avr4* gene was subsequently cloned and sequenced. Because of the change that was made to the *Avr4* gene annotation, the gene-specific UPS and DWN primers (Table 2) still had to be tested before sequencing of the *Avr4* gene as a final validation step. The *Avr4* amplicon was amplified from *C. zeina* (Mkushi) gDNA in a PCR reaction at 57˚C. The PCR product was analysed on a 2% agarose gel (Figure 6) and determined to be approximately 953 bp, which is the expected amplicon size. The amplified *Avr4* product was then purified, cloned into the pTZ57R/T plasmid (Appendix, Figure S1), and sequenced with standard M13 primers. The sequenced *C. zeina* (Mkushi) *Avr4* gene region with up- and downstream regions is shown in figure 7, and supports the *Avr4* gene annotation and genome assembly. The sequenced *C. zeina* (Mkushi) *Avr4* was identical to the *C. zeina* (Mkushi) *Avr4* annotated from the genome sequence.

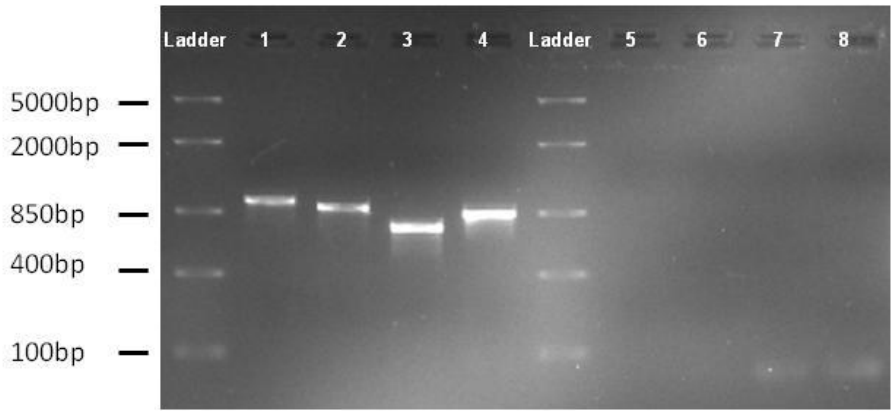


Figure 6. Amplification of full length *C. zeina* (Mkushi) *Avr4*, *Ecp2* and *Ecp6* genes.
Before sequencing, gene-specific primers for the three effector genes (*Avr4*, *Ecp2*, and *Ecp6*) were tested. PCR amplification with gene-specific primers was done at 57˚C for both *Avr4* and *Ecp2*, and at 59˚C for the *Ecp6* fragments. The subsequent amplification products were analysed on a 2% agarose gel. All products were amplified from *C. zeina* (Mkushi) genomic DNA and a middle range ladder (Thermo Scientific, Waltham, Massachusetts) was used to measure product sizes. *Avr4* was amplified with gene-specific UPS and DWN primers, *Ecp2* was amplified with gene-specific UPS and DWN primers, and *Ecp6* was amplified with primer combinations Ecp6F2A & Ecp6R2A and Ecp6F2B & Ecp6R2B. (1) *Avr4* product amplified with gene-specific UPS and DWN primers, (2) *Ecp2* product amplified with gene-specific UPS and DWN primers, (3) *Ecp6* product amplified with Ecp6F2A and Ecp6R2A primers, (4) *Ecp6* product amplified with Ecp6F2B & Ecp6R2B primers, and lanes 5-8 shows the negative controls of the amplified products *Avr4*, *Ecp2*, *Ecp6A* and *Ecp6B* respectively. Expected amplicon sizes were 953 bp for *Avr4*, 896 bp for *Ecp2*, 735 bp for *Ecp6A*, and 848 bp for *Ecp6B*. The faint bands seen in lanes 7 and 8 were primer dimers.

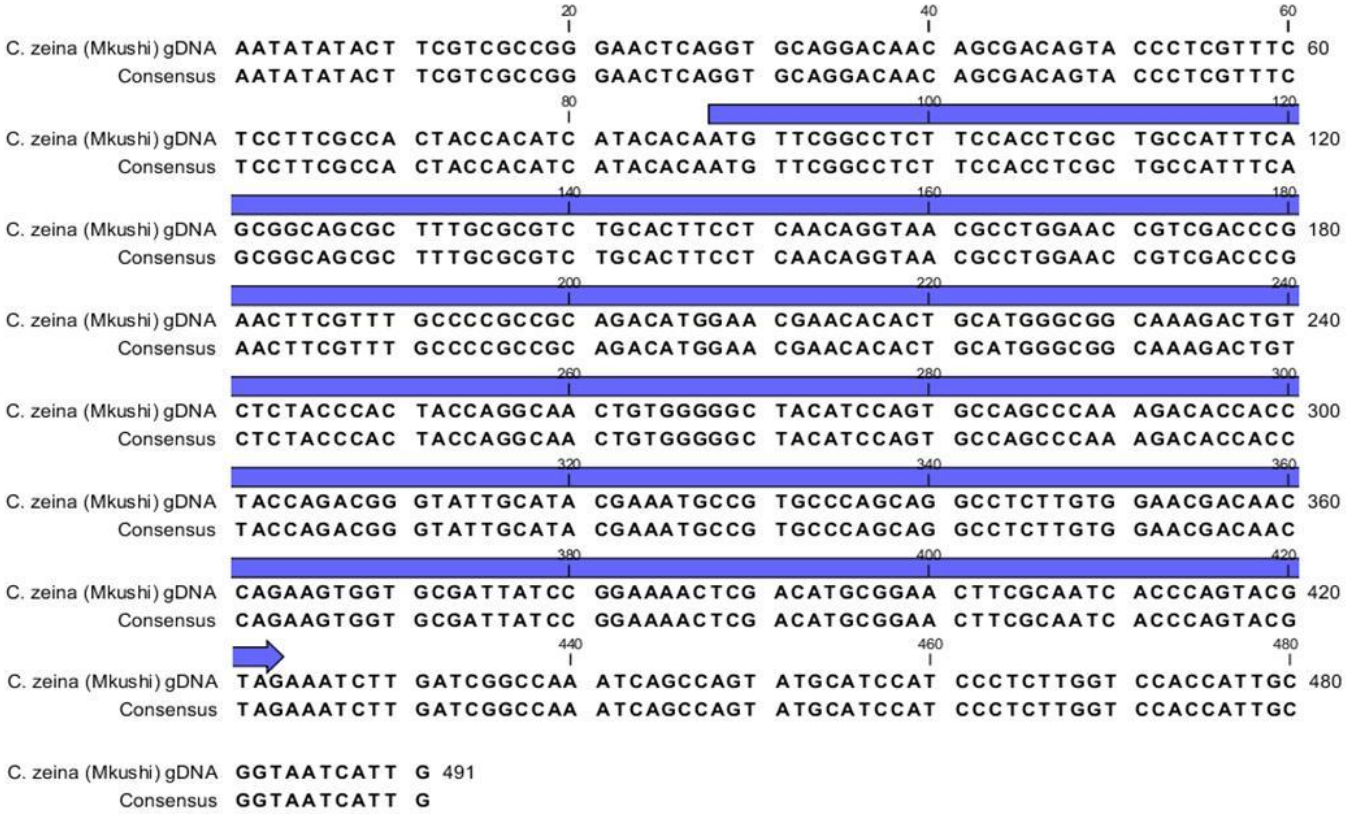


Figure 7. Confirmation of the *C. zeina* (Mkushi) *Avr4* contig assembly.
The *C. zeina* (Mkushi) *Avr4* cloned gene region was sequenced from plasmid DNA using standard M13 primers. The plasmid map is shown in the appendix (Figure S1). Forward and reverse sequencing products were combined as a consensus sequence. The *C. zeina* (Mkushi) genome assembly containing the *Avr4* gene region (blue arrow) with approximately 100 bp up- and downstream regions was used as a reference. The alignment between the two sequences was identical.

4.1.4 *C. zeina* (Mkushi) *Avr4* homologs present in *C. zeae-maydis* and confirmed in *M. fijiensis* and *C. fulvum*.

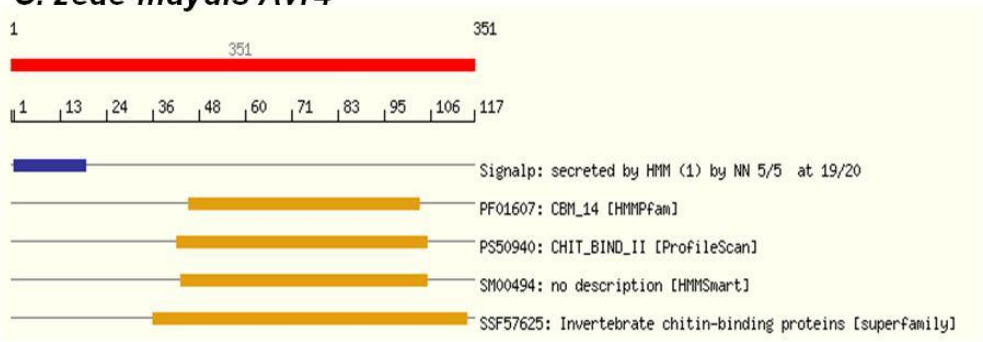
The annotated *C. zeina* (Mkushi) *Avr4* gene sequence was used as a query sequence to BLAST the *C. zeae-maydis*, *M. fijiensis*, *M. graminicola* and *C. fulvum* genomes to determine if *Avr4* homologs were present. Reasons for choosing these genomes were: *C. zeae-maydis* is the sibling species of *C. zeina*; *M. fijiensis* which is closely related to *C. zeina, M. graminicola* is the possible sexual stage of *C. zeina* (GROENEWALD *et al.* 2006), and *C. fulvum* as the model organism. BLASTn and tBLASTx searches were made against the respective genomes and the results are indicated in Table 4. Homologs of the annotated *C. zeina* (Mkushi) *Avr4* were found in *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum*. No homologs were detected in *M. graminicola*.

Table 4. A summary of the most relevant BLASTn and tBLASTx results obtained for the annotated *C. zeina* (Mkushi) *Avr4* effector query sequence to the *C. zeae-maydis*, *M. fijiesis*, *M. graminicola*, and *C. fulvum* genomes.

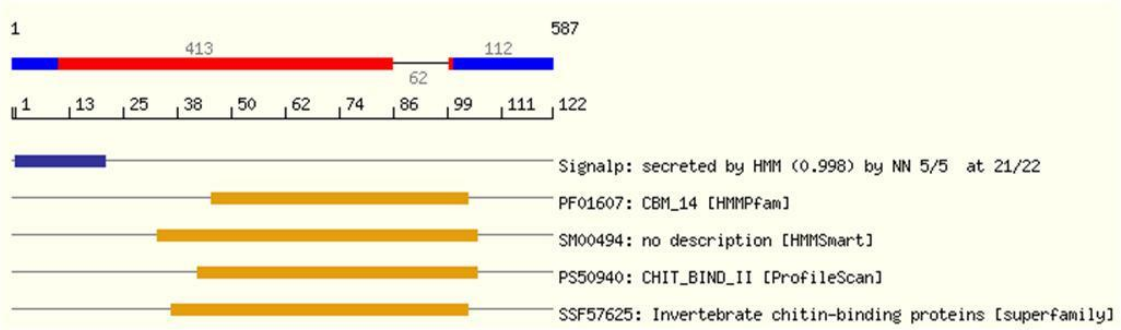
BLASTn query	Genome	BLAST hit information					Protein ID
		Identity	Score	E-value	Gaps	Aligned region	
C. zeina(Mkushi) *Avr4*	*C. zeae-maydis*	86%	112	3e-54	-	128 bp	Cerzm1:85994
C. zeina(Mkushi) *Avr4*	*M. fijiensis*	-	-	-	-	-	-
C. zeina(Mkushi) *Avr4*	*M. graminicola*	-	-	-	-	-	-
C. zeina(Mkushi) *Avr4*	*C. fulvum*	-	-	-	-	-	-
tBLASTx query	**Genome**	**BLAST hit information**					**Protein ID**
		Identity	**Score**	**E-value**	**Gaps**	**Aligned region**	
C. zeina(Mkushi) *Avr4*	*C. zeae-maydis*	62%	140	3e-33	-	107 aa	Cerzm1:85994
C. zeina(Mkushi) *Avr4*	*M. fijiensis*	68%	153	6e-38	-	76 aa	Mycfi2:87167
C. zeina(Mkushi) *Avr4*	*M. graminicola*	-	-	-	-	-	-
C. zeina(Mkushi) *Avr4*	C. fulvum	54%	52	4e-16	-	35 aa	Clafu1:189855

The *C. zeae-maydis Avr4* homolog (Cerzm1:85994) shared 86% identity to *C. zeina* (Mkushi) *Avr4* on a nucleotide level with an alignment region of 128 bp and shared 62% identity with an alignment region of 107 aa on protein level. The predicted *C. zeae-maydis Avr4* gene is 351 bp with no introns (Figure 8). The *C. zeae-maydis Avr4* homolog contains a CBM14 domain (IPR002557) and an N-terminal signal peptide. No *Avr4* homologs have previously been identified in *C. zeae-maydis*. The *M. fijiensis* Avr4 homolog (Mycfi1:87167 or XM007927856) shared 68% identity (alignment region of 76 aa) with *C. zeina* (Mkushi) Avr4 on a protein level. The predicted *M. fijiensis Avr4* consists of 587 bp with a 62 bp intron (Figure 8). The *M. fijiensis* Avr4 homolog also contains a CBM14 domain (IPR002557) and an N-terminal signal peptide. The Avr4 protein homolog found in *C. fulvum* (Clafu1:189855 or CAA55403.1) shares 54% identity (alignment region of 35 aa) on a protein level with *C. zeina* (Mkushi) Avr4. The predicted *C. fulvum Avr4* homolog consists of a single exon of 408 bp (Figure 8). The *C. fulvum Avr4* homolog contains a CBM14 domain (IPR002557) as well as an N-terminal signal peptide. The *M. fijiensis Avr4* homolog (Mycfi1:87167) and the *C. fulvum Avr4* homolog (X78829) have previously been identified and renamed as *MfAvr4* (STERGIOPOULOS et al. 2010) and *Avr4* (JOOSTEN et al. 1994). The confirmation of the homologs in this study therefore provides further evidence that the correct *Avr4* homolog has been identified in *C. zeina* (Mkushi). An *Avr4* homolog in *M. graminicola* could previously also not be identified when using the *C. fulvum Avr4* (X78829) sequence as a query (STERGIOPOULOS *et al.* 2010).

C. zeae-maydis Avr4



M. fijiensis Avr4



C. fulvum Avr4

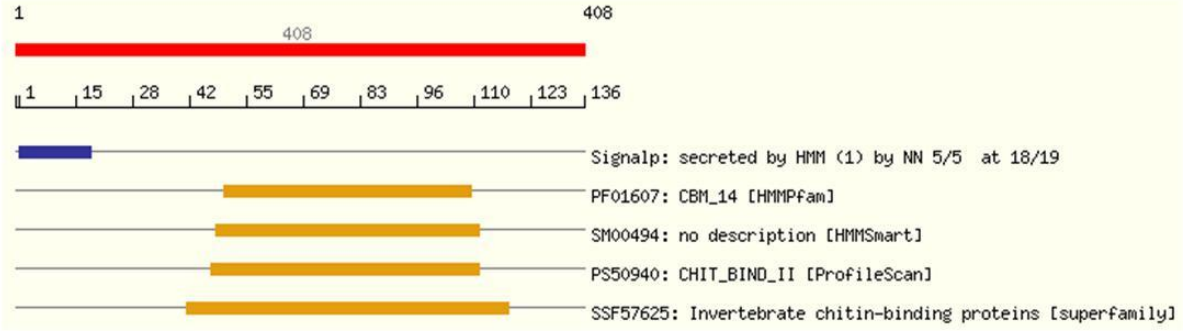


Figure 8. Predicted gene structures of the *C. zeina* (Mkushi) *Avr4* homologs present in *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum*. The coding regions of the predicted gene homologs were indicated by red bars and the untranslated regions (UTRs) were indicated with blue bars if present. Indicated below each gene prediction, are the conserved functional domains (in yellow) and signal peptide (SP) sequences (in blue). *C. zeae-maydis Avr4* was predicted to be 351 bp with no introns, the predicted *M. fijiensis Avr4* gene was 587 bp with a 62 bp intron, and the predicted *C. fulvum Avr4* gene was 408 bp with no introns.

4.1.5 Protein alignments of Avr4

Through BLAST search analysis using the predicted protein sequence of *C. zeina* (Mkushi) Avr4 as query against the protein sequence database available from GenBank, fungal homologs of these predicted proteins were obtained. The reciprocal BLAST result indicated that *C. zeina* Avr4 (GU574327) was still the most homologous to the *C. zeina* (Mkushi) Avr4

protein, sharing 68% identity with the *C. zeina* (Mkushi) Avr4 protein. The best hits as well as the relevant homologous sequences identified in section 4.1.2 and 4.1.4 were aligned (Figure 9). Residue conservation is indicated by the bar graph below the sequences in terms of percentage conservation between sequences in the alignment. Gaps in the alignment are indicated by hyphens. The eight proteins included in the Avr4 protein alignment were the four *Cercospora* Avr4 sequences: *C. apii*, *C. beticola*, *C. nicotianae*, and *C. zeina* (referred to as *C. zeina* (GU574327) in the alignment) that were available (STERGIOPOULOS *et al.* 2010), the predicted *C. zeina* (Mkushi) and *C. zeae-maydis* Avr4 proteins, the *M. fijiensis* Avr4 protein (Mycfi2:87167) (STERGIOPOULOS *et al.* 2010), and the *C. fulvum* Avr4 protein (Clafu1:189855) (JOOSTEN *et al.* 1997).

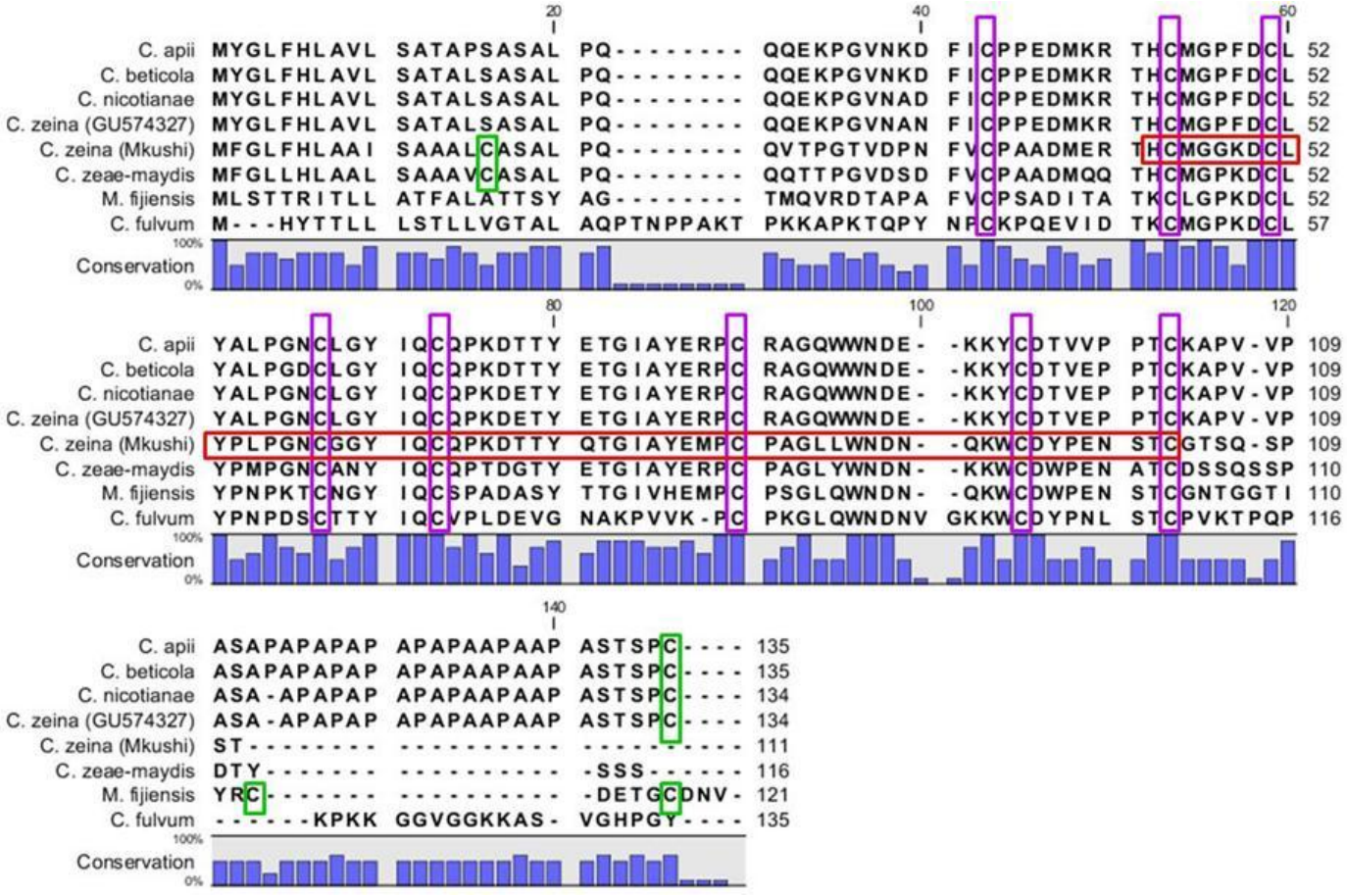


Figure 9. Avr4 protein alignment.
Avr4 protein alignment generated with Mafft software and visualised with CLC Bio (main workbench) software. The alignment includes the most homologous proteins as well as other relevant Avr4 protein sequences identified in this chapter. The percentage conservation between the sequences at each amino acid in the alignment was indicated by bar graphs (Blue). Cysteine residues were indicated with green boxes and conserved cysteine residues were indicated with purple boxes. The red boxes indicate the location of the chitin-binding CBM14 domain position in *C. zeina* (Mkushi) Avr4.

The Avr4 protein alignment showed high sequence conservation, especially in the region where the chitin-binding domain (CBM14) is found (indicated by the red box in figure 9). It is also important to note that seven of the eight conserved cysteine residues in the alignment are found in the conserved chitin-binding region. The eight conserved cysteine residues are indicated with purple boxes in the alignment (Figure 9). The cysteine residues that were not conserved are indicated with green boxes. The *C. apii*, *C. beticola*, *C. nicotianae*, *C. zeina* (GU574327), *C. zeina* (Mkushi), and the *C. zeae-maydis* Avr4 proteins each have nine cysteine residues. The *M. fijiensis* and the *C. fulvum* Avr4 proteins have ten and eight cysteine residues, respectively. Two general gap regions were found at residues 23-30 and 100-101 in the alignment, but are absent in the *C. fulvum* Avr4 protein. The *C. zeina* (Mkushi), *C. zeae-maydis*, and the *M. fijiensis* Avr4 proteins were on average 20 residues shorter than the rest of the Avr4 proteins, causing the gap in the C-terminal region of the alignment.

When comparing the alignment between the two *C. zeina* Avr4 proteins, it was expected that the two proteins would be more or less identical. However, the *C. zeina* (GU574327) and the *C. zeina* (Mkushi) Avr4 proteins differ in length, the amount of cysteine residues, and overall sequence similarity. A pairwise comparison between the *C. zeina* (Mkushi) and *C. zeina* (GU574327) Avr4 proteins showed that the two proteins share 53% sequence identity and 64.40% sequence similarity (Figure 10). Interestingly, the *Cercospora* Avr4 sequences reported in Stergiopoulos *et al.* (2010) all showed an APA amino acid repeat motif at the C-terminal region of the alignment (Figure 9). However, careful examination of the C-terminal region of *C. zeina* (Mkushi) Avr4 did not reveal this motif. The *C. zeina* (Mkushi) and *C. zeae-maydis* Avr4 proteins were more similar in length and shared higher sequence similarity as well as nine cysteine residues. The pairwise comparison indicated that the *C. zeae-maydis* Avr4 protein shares 72.60% sequence identity and 82.10% sequence similarity to *C. zeina* (Mkushi) Avr4 (Figure 10). The *M. fijiensis* Avr4 protein had the highest number of cysteine residues and shares relatively low protein identity and similarity to the rest of the Avr4 proteins in the alignment. The percentage identity of *M. fijiensis* Avr4 to the rest of the proteins in the alignment ranged from 28.90 to 44.10% and the percentage similarity ranged from 38.20 to 55.50% (Figure 10). The *C. fulvum* Avr4 protein is a more distantly related sequence and shared relatively lower sequence identity (ranging from 29.80 to 41.40% identity) and similarity (ranging from 37.70 to 47.90% similarity) with the rest of the Avr4 proteins in the alignment (Figure 10). However, the *C. fulvum* Avr4 protein contains eight cysteine residues, which are all conserved between the Avr4 proteins in the alignment (Figure 9).

1. *C. apii*
2. *C. beticola*
3. *C. nicotianae*
4. *C. zeina* (GU574327)
5. *C. zeina* (Mkushi)
6. *C. zeae-maydis*
7. *M. fijiensis*
8. *C. fulvum*

	1	2	3	4	5	6	7	8
1		98.50	96.30	96.30	59.30	60.41	38.20	38.70
2	97.80		97.80	97.80	60.70	61.90	39.90	38.70
3	96.30	97.00		100.00	60.40	63.00	40.10	37.70
4	95.60	96.30	99.30		60.40	63.00	40.10	37.70
5	51.10	51.90	52.20	53.00		82.10	55.10	41.50
6	52.50	52.50	53.60	52.90	72.60		55.50	39.40
7	28.90	29.10	29.30	29.30	44.10	43.00		47.90
8	30.70	31.30	29.80	29.80	32.40	32.40	41.40	

% Identity

% Similarity

Figure 10. Pairwise comparisons between the aligned Avr4 protein sequences generated with CLC Bio Main Workbench software (Aarhus, Denmark) and EMBOSS Needle. The pairwise comparison was made between the eight Avr4 protein sequences (1) *C. apii*, (2) *C. beticola*, (3) *C. nicotianae*, (4) *C. zeina* (GU574327), (5) *C. zeina* (Mkushi), (6) *C. zeae-maydis*, (7) *M. fijiensis*, and (8) *C. fulvum*. The upper comparison gives the percentage amino acid similarity values in the alignment, whereas the lower comparison gives the percentage identity. The difference between amino acid identity and similarity is outlined in the relevant materials and methods section. This figure compliments the Avr4 protein alignment shown in Figure 9.

4.1.6 Phylogenetic analysis of *Avr4* homologs

Following the Avr4 protein alignment results, phylogenetic analysis was done to determine the evolutionary relationships between the specific aligned sequences. Nucleotide sequence alignments instead of protein alignments were made for the *Avr4* homologs. Most protein sequences available for Avr4 were predicted from the nucleotide sequences and therefore alignments for phylogenetic analysis were made by using nucleotide sequences instead of protein sequences to avoid making further predictions from sequences that have not been validated. Also, phylogenetic analysis from amino acid alignments gave different groupings for neighbour- joining (NJ) and maximum likelihood (ML) analysis (results not shown).

The best predicted substitution model based on the *Avr4* alignment was the Tamura 3-parameter (T92) model with gamma distribution. Figure 11 shows the combined phylogenetic tree of ML and NJ analysis based on the *Avr4* nucleotide alignment. As expected, the four very similar *Avr4* sequences (*C. zeina* (GU574327), *C. nicotianae*, *C. beticola*, and *C. apii*) grouped together in the same clade forming two subgroups. *C. zeina* (GU574327) and *C. nicotianae* formed part of one subgroup and *C. beticola* and *C. apii* formed the other subgroup. Interestingly, *C. zeina* (Mkushi) and *C. zeae-maydis* grouped together, but remained separate from the other four *Cercospora* sequences. The *C. zeina* (GU574327) and *C. zeina* (Mkushi) *Avr4* sequences grouped separately which was also unexpected. The groupings were supported by very high bootstrap percentages. Bootstrap values were obtained for 1000 replications and indicated as percentages on the nodes of the

resulting phylogenetic trees. The *C. fulvum Avr4* sequence was used to root the tree as it was more distantly related to the other sequences. The ML and NJ phylogenetic trees showed identical branching in each case. For each phylogenetic tree, the bootstrap values above 75% are indicated on the nodes as NJ/ML values.

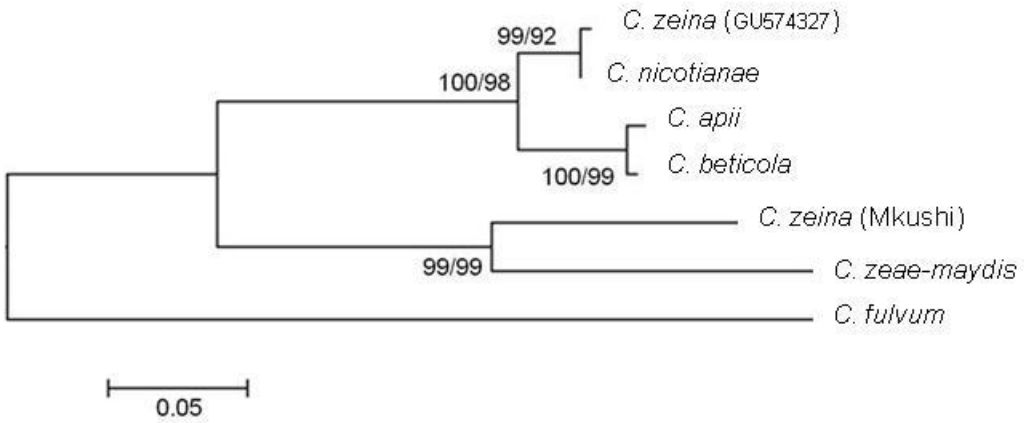


Figure 11. Combined phylogenetic tree for *Avr4* constructed based on the nucleotide alignment of the different *Avr4* gDNA sequences. Bootstrap values are indicated on the nodes as neighbour-joining analysis/maximum likelihood analysis. Bootstrap values were obtained from 1000 replicates and indicated as percentages on the nodes.

4.2 Evidence for the presence of the *Ecp2* effector homolog in *C. zeina*

4.2.1 The presence of an *Ecp2* homolog in the *C. zeina* (Mkushi) genome

The *C. fulvum Ecp2* gDNA sequence (Z14024) (Table 1) was used as a query for BLASTn searches against the draft *C. zeina* (Mkushi) genome. BLASTn results showed a higher percentage identity as well as better statistical significance when the *C. fulvum* (Z14024) sequence was queried against the *C. zeina* Illumina data assembled with Velvet compared to the other *C. zeina* datasets. The putative *C. zeina* (Mkushi) *Ecp2* hit showed 73% identity to the *C. fulvum Ecp2* query sequence (Z14024) with an aligned region of 102 bp with no gaps, a score of 64, and statistical significance of 4e-07.

4.2.2 *C. zeina* (Mkushi) *Ecp2* preliminary predicted gene structure

The preliminary *C. zeina* (Mkushi) *Ecp2* homolog was annotated to be able to determine the gene structure in terms of intron/exon boundaries. Gene prediction programmes (Augustus and FGENESH) both predicted similar gene structures for *C. zeina* (Mkushi) *Ecp2*. The *C. fulvum Ecp2* gene (Z14024) is 554 bp long including an intron of 56 bp (VAN DEN ACKERVEKEN *et al.* 1993). The preliminary predicted size of the *C. zeina* (Mkushi) *Ecp2* was 569 bp with an intron of 65 bp (Figure 12), which is very similar to *C. fulvum Ecp2* (Z14024).

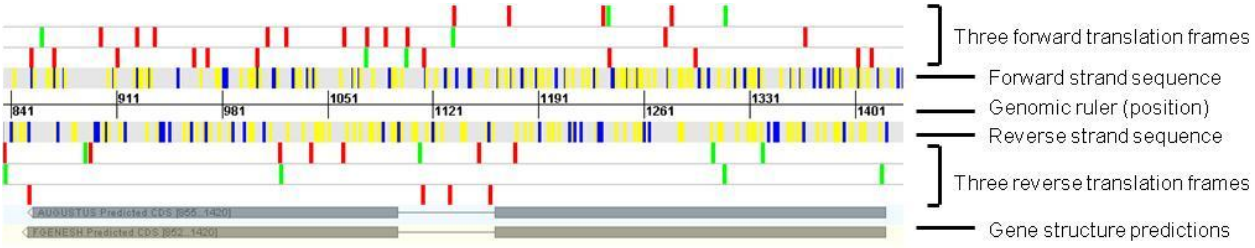


Figure 12. *C. zeina* (Mkushi) *Ecp2* preliminary gene prediction visualised in GenomeView comparing the expected gene structures and orientation as predicted by Augustus and FGENESH. The annotations made by both prediction programmes were similar. The grey bar at the bottom of the figure represents the prediction made by FGENESH, whereas the blue bar represents the prediction made by Augustus. Both predictions were present on the reverse strand. The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. The blue and yellow vertical lines represent possible acceptor and donor sites for splicing in the forward and reverse sequence.

Reciprocal BLASTn and tBLASTx analysis against the NCBI database showed that the *M. fijiensis* hypothetical protein (XM007924386) was the closest homolog to the preliminary predicted *C. zeina* (Mkushi) *Ecp2* gene. The *M. fijiensis* hypothetical protein (XM007924386) contained an Hce2 (**H**omologs of ***C****. fulvum* **E**cp**2**) domain (STERGIOPOULOS *et al.* 2012) and is a putative necrosis inducing factor as described by the authors (STERGIOPOULOS *et al.*

2010). The preliminary predicted *C. zeina* (Mkushi) *Ecp2* gene shared 69% identity (431 bp alignment region) to the *M. fijiensis* hypothetical protein (XM007924386) and the translated *C. zeina* (Mkushi) *Ecp2* gene shared 68% identity (77 aa alignment region) to the translated *M. fijiensis* hypothetical protein (XM007924386) (Table 5). Other significant BLASTn and tBLASTx were also summarized in Table 5 and include hits to *C. fulvum Ecp2*.

Table 5. A summary of the most relevant BLASTn and tBLASTx results obtained from reciprocal BLAST analysis of the preliminary predicted *C. zeina* (Mkushi) *Ecp2* gene.

BLASTn Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
M. fijiensis hypothetical protein	(XM007924386)	69	129	3e-26	21	431 bp
C. fulvum Ecp2	(Z14024.1)	72	54	0.002	0	94 bp
tBLASTx Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
M. fijiensis hypothetical protein	(XM007924386)	68	279	2e-45	0	77 aa
C. fulvum Ecp2	(Z14024.1)	61	186	8e-36	0	61 aa
M. fijiensis hypothetical protein	(XM007929702)	45	104	4e-08	0	42 aa

The preliminary predicted *C. zeina* (Mkushi) Ecp2 protein has a conserved putative necrosis-inducing factor (Hce2) (pfam14856) that corresponds to the Hce2 domain found in the mature *C. fulvum* Ecp2 protein (VAN DEN ACKERVEKEN *et al.* 1993). The preliminary C*. zeina* (Mkushi) Ecp2 protein was predicted to be 168 aa. The Hce2 domain is located at the 104-425 bp region of the *C. zeina* (Mkushi) *Ecp2* gene (Figure 13). An N-terminal signal peptide is present in the preliminary predicted *C. zeina* (Mkushi) Ecp2 (result not shown). Although the exact function of the *C. fulvum* Ecp2 pathogenesis factor has not been determined yet, it has been proposed that Ecp2 is a necrosis inducing factor (LAUGE *et al.* 1998; STERGIOPOULOS *et al.* 2010).

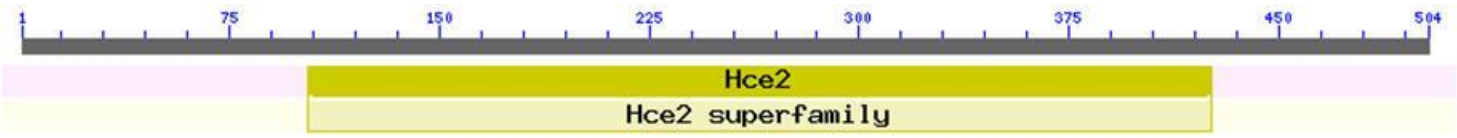


Figure 13. Putative necrosis-inducing factor (Hce2) conserved in *C. zeina* (Mkushi) *Ecp2*
The preliminary predicted *C. zeina* (Mkushi) *Ecp2* has a conserved putative necrosis-inducing factor Hce2 domain which is also present in the *C. fulvum* Ecp2 pathogenicity factor. The grey bar represents the predicted *C. zeina* (Mkushi) *Ecp2* query sequence and the yellow bar represents the region of the conserved domain. The figure was drawn to scale.

4.2.3 Confirmation of *C. zeina* (Mkushi) *Ecp2* contig assembly

4.2.3.1 PCR analysis with gene-specific primers

Gene-specific primers were designed to the *C. zeina* (Mkushi) *Ecp2* gene region to be able to determine if the *C. zeina* (Mkushi) contig containing the preliminary predicted *Ecp2* was assembled correctly. Primer pairs for *C. zeina* (Mkushi) *Ecp2* were designed in such a way that six different overlapping amplicons of known length would be amplified through PCR analysis (Figure 14). The *Ecp2* primer sequences are summarized in Table 2.

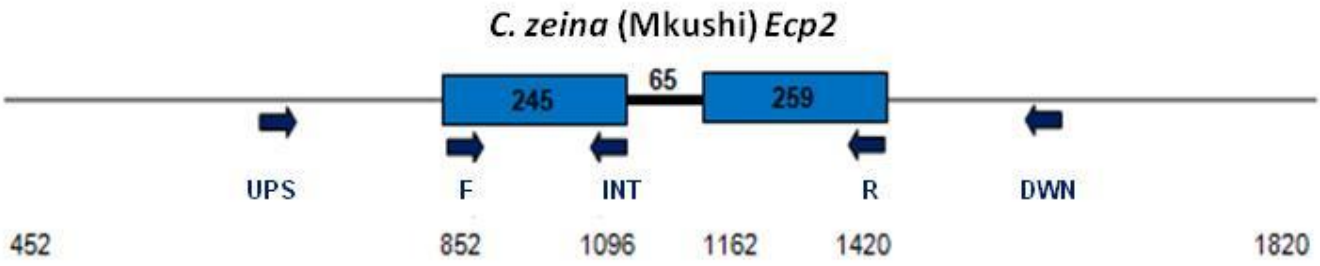


Figure 14. Gene-specific primers designed to the preliminary predicted *C. zeina* (Mkushi) *Ecp2* gene. The navy arrows indicate the designed primer regions. The blue boxes indicate exons and the bold black line indicates an intron region. The respective sizes of the introns and exons were indicated in the figure. The figure was not drawn to scale.

The amplified PCR products were analysed on a 1.3% agarose gel and the product lengths could be determined. The *Ecp2* amplicons were amplified from *C. zeina* (Mkushi) gDNA using primer combinations as follows; (1) UPS and DWN primers, (2) UPS and R primers, (3) UPS and INT primers, (4) F and INT primers, (5) F and R primers, and (6) F and DWN primers. Not all the amplicons could be amplified at the same temperature as the melting temperatures of some primer combinations varied. All of the primer combinations could amplify specific products at 57˚C and 58˚C, except the F and R primers that could only amplify at 58˚C and the UPS and DWN primers that specifically amplified at 57˚C. Therefore, two gel images are seen in figure 15 in order to show amplification of all the products. The expected amplicon sizes were all obtained and were 896 bp, 633 bp, 304 bp, 832 bp, 569 bp, and 240 bp for primer combinations respectively. These amplification products gave more confidence in the assembly of the contig containing *Ecp2*.

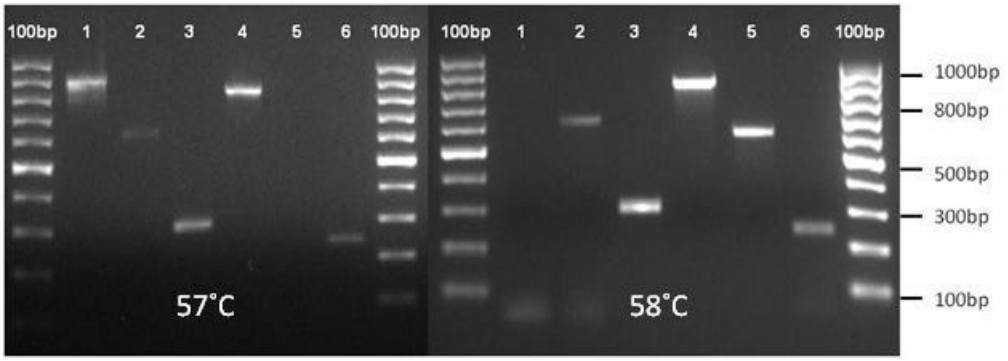


Figure 15. *C. zeina* (Mkushi) *Ecp2* contig confirmation with gene-specific primers.
The *C. zeina* (Mkushi) *Ecp2* contig was verified through PCR amplification with gene-specific primers. The subsequent amplification products were analysed on a 1.3% agarose gel. All products were amplified from *C. zeina* (Mkushi) genomic DNA and a 100 bp ladder (Thermo Scientific, Waltham, Massachusetts) was used to measure product sizes. *Ecp2* products amplified at 57˚C and 58˚C with UPS and DWN primers (1), UPS and R primers (2), UPS and INT primers (3), F and INT primers (4), F and R primers (5), F and DWN primers (6). The expected amplicon sizes for *Ecp2* were 896 bp (1), 633 bp (2), 304 bp (3), 832 bp (4), 569 bp (5), and 240 bp (6).

4.2.3.2 *C. zeina* (Mkushi) *Ecp2* annotation through mapping of RNAseq reads

Mapping of RNAseq reads to the *C. zeina* (Mkushi) *Ecp2* contig was an additional step to validate and improve the preliminary *Ecp2* gene prediction and annotation based on empirical evidence. Mapping of *in vitro* and *in planta* RNAseq reads (outlined in the relevant materials and methods section) was done to the entire *C. zeina* (Mkushi) *Ecp2* contig to avoid possible forced reads.

The *C. zeina* (Mkushi) *Ecp2* gene region with mapped *in planta* RNAseq reads is shown in Figure 16. The *C. zeina* (Mkushi) *Ecp2* gene was annotated based on the mapped RNAseq reads as well as the homolog *C. fulvum Ecp2* sequence (Z14024.1) as a reference. The *in vitro* (results not shown) and *in planta* mapped reads supported the initial predicted annotation that included the intron sequence and therefore no changes were made to the *C. zeina* (Mkushi) *Ecp2* gene. The *Ecp2* gene contains two exon regions and a single intron. The first exon is 245 bp, followed by a small intron of 65 bp, and a second exon that is 259 bp in size.

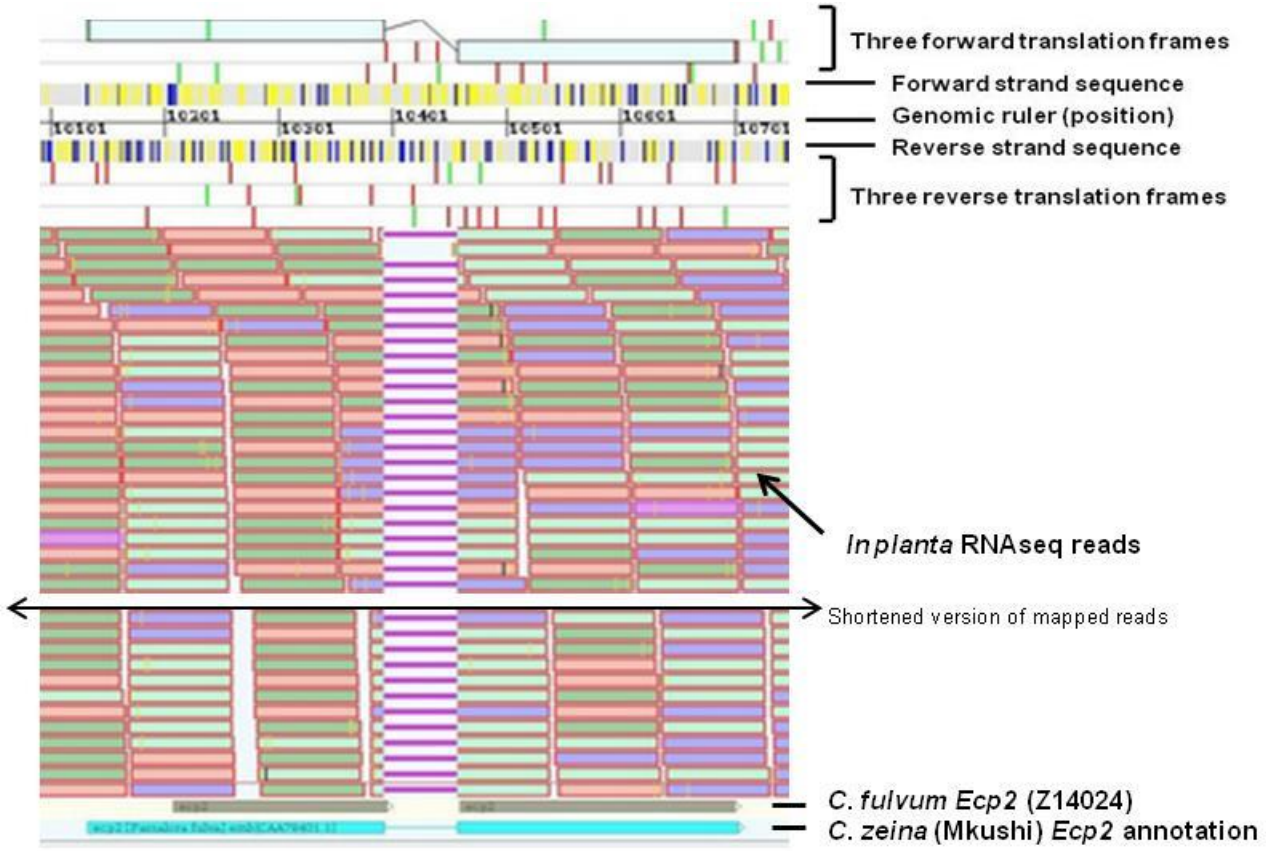


Figure 16. *C. zeina* (Mkushi) *Ecp2* annotation.
In planta RNAseq reads were mapped to the contig containing the *C. zeina* (Mkushi) *Ecp2* gene. Only the region containing the annotated *Ecp2* gene is shown. The coloured bars represent the *in planta* reads mapped to the whole contig. The *C. fulvum Ecp2* gene homolog (Z14024.1) was used as a reference for annotation purposes and is represented by the grey horizontal bar at the bottom of the figure. The annotation of the *Ecp2* gene was done manually by taking into account possible translation start and stop sites, the *C. fulvum Ecp2* homolog, and the mapped RNAseq reads. The manual annotation of *C. zeina Ecp2* supported a similar gene structure as seen with the predicted *C. zeina* (Mkushi) *Ecp2* gene (not shown). The *Ecp2* intron region is well supported by the mapped RNAseq reads. The *C. zeina* (Mkushi) *Ecp2* annotation is represented by the light blue horizontal bar at the bottom of the figure. The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. *Ecp2* is present on the forward strand. See materials and methods for RNAseq read colours and their descriptions.

The annotated *C. zeina* (Mkushi) *Ecp2* gene was subsequently cloned and sequenced. The *C. zeina* (Mkushi) *Ecp2* gene-specific UPS and DWN primers (Table 2) were tested before sequencing was undertaken. The *C. zeina* (Mkushi) *Ecp2* amplicon was amplified from *C. zeina* (Mkushi) gDNA in a PCR reaction with the gene-specific UPS and DWN primers at 57˚C. The PCR product was analysed on a 2% agarose gel and the product length was determined to be approximately 896 bp in size (Figure 6). The amplified *Ecp2* product was cloned into the pTZ57R/T plasmid (Appendix, Figure S1) and sequenced (as outlined in the relevant materials and methods section). The sequencing products of *Ecp2* are shown in figure 17, and support the *Ecp2* gene sequence as determined from the genome assembly.

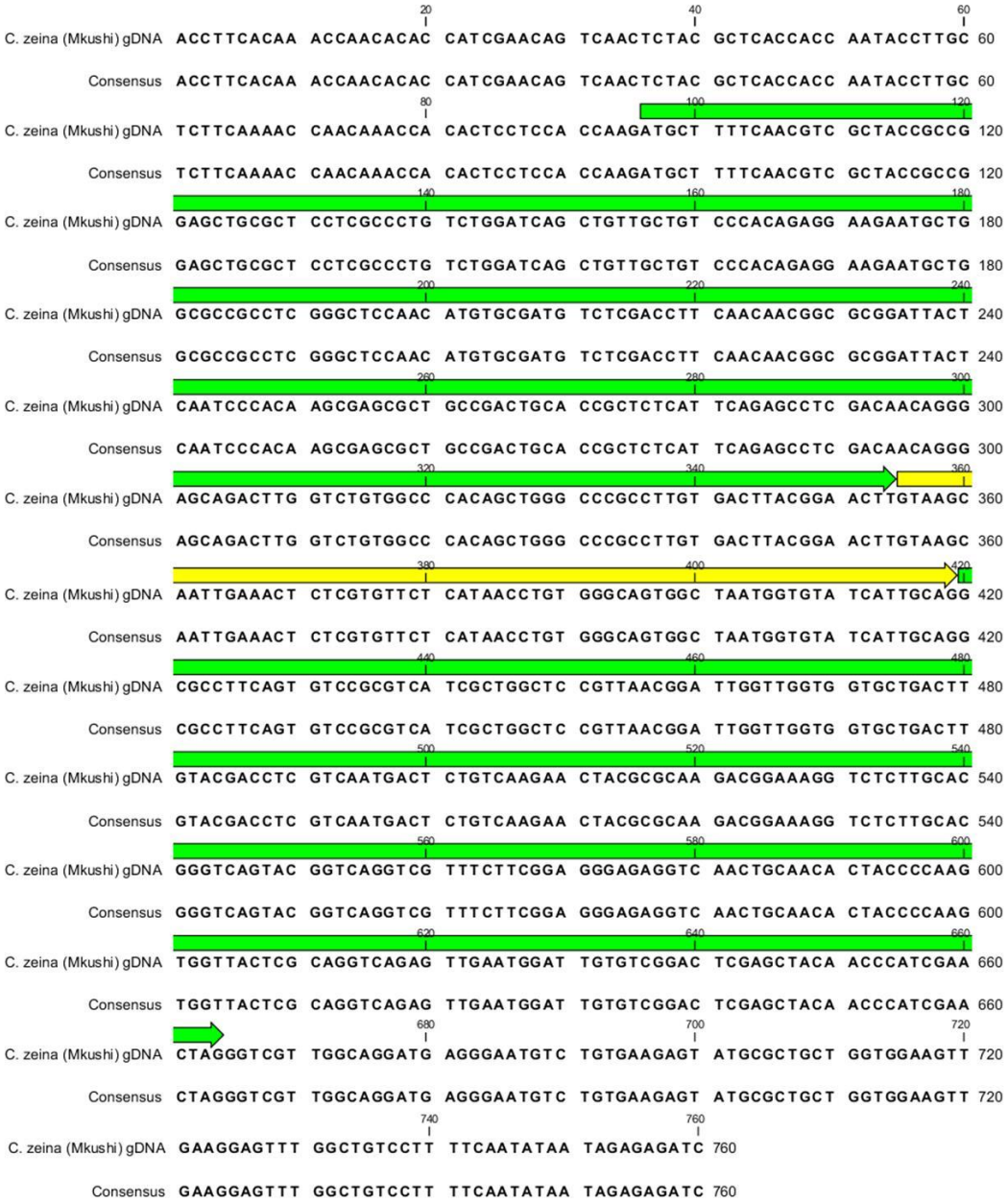


Figure 17. Confirmation of *C. zeina* (Mkushi) *Ecp2* contig assembly.
The *C. zeina* (Mkushi) *Ecp2* cloned gene region was sequenced from plasmid DNA using standard M13 primers. Forward and reverse sequencing products were combined and are represented as a consensus sequence. The *C. zeina* (Mkushi) contig containing the *Ecp2* gene region (green arrows indicate exons and the yellow arrow indicates the intron position) with approximately 100 bp up- and downstream of the predicted *Ecp2* gene was used as a reference.

4.2.4 *C. zeina* (Mkushi) *Ecp2* homologs present in *C. zeae-maydis* and confirmed in *M. fijiensis* and *C. fulvum*.

The annotated *C. zeina* (Mkushi) *Ecp2* gene sequence was used as a query to BLAST the *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum* genomes to determine if *Ecp2* homologs were present. The genomes were chosen based on the reasons mentioned in

section 4.1.4. BLASTn and tBLASTx searches were made against the respective genomes and the results are indicated in Table 6.

Table 6. A summary of the most relevant BLASTn and tBLASTx results obtained for the annotated *C. zeina* (Mkushi) *Ecp2* effector query sequence to the *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum* genomes.

BLASTn query	Genome	BLAST hit information					Protein ID
		Identity	Score	E-value	Gaps	Aligned region	
C. zeina (Mkushi) *Ecp2*	*C. zeae-maydis*	89%	179	2e-66	0	158 bp	Cerzm1:53627
C. zeina (Mkushi) *Ecp2*	*M. fijiensis*	92%	54	2e-07	0	39 bp	Mycfi1:52972
C. zeina (Mkushi) *Ecp2*	*M. graminicola*	-	-	-	-	-	-
C. zeina (Mkushi) *Ecp2*	*C. fulvum*	-	-	-	-	-	-
tBLASTx query	**Genome**	**BLAST hit information**					**Protein ID**
		Identity	**Score**	**E-value**	**Gaps**	**Aligned region**	
C. zeina (Mkushi) *Ecp2*	*C. zeae-maydis*	74%	128	8e-65	0	66 aa	Cerzm1:53627
C. zeina (Mkushi) *Ecp2*	*M. fijiensis*	44%	41	2e-07	0	57 aa	Mycfi1:52972
C. zeina (Mkushi) *Ecp2*	*M. graminicola*	-	-	-	-	-	-
C. zeina (Mkushi) *Ecp2*	*C. fulvum*	61%	88	4e-40	0	61 aa	Clafu1:197200

Homologs of the annotated *C. zeina* (Mkushi) *Ecp2* were found in *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum*. No homologs were detected in *M. graminicola*. The *C. zeae-maydis Ecp2* homolog (Cerzm1:53627) shared 89% identity to the *C. zeina* (Mkushi) *Ecp2* on a nucleotide level with an alignment region of 158bp and shared 74% identity with an alignment region of 66 aa on a protein level (Table 6). The predicted *C. zeae-maydis Ecp2* gene is 566 bp with a 68 bp intron (Figure 18). The *C. zeae-maydis Ecp2* homolog contained a predicted N-terminal signal peptide. No *Ecp2* homolog has previously been identified in *C. zeae-maydis*. The *M. fijiensis Ecp2* homolog (Mycfi2:52972 or XM007924386) shared 92% identity to the *C. zeina* (Mkushi) *Ecp2* on a nucleotide level with an alignment region of 39bp and shared 44% identity with an alignment region of 57 aa on a protein level (Table 6). The predicted *M. fijiensis Ecp2* gene is 736 bp with a 59 bp intron (Figure 18). The *M. fijiensis* Ecp2 homolog contained a predicted N-terminal signal peptide. The *M. fijiensis* Ecp2 homolog (Mycfi2:52972) has previously been identified and renamed as MfEcp2-3 (STERGIOPOULOS *et al.* 2010). A *C. zeina* (Mkushi) Ecp2 homolog was also detected in *C. fulvum* (Clafu1:197200 or CAA78401.1) and shared 61% identity with an alignment region of 61 aa. The predicted *C. fulvum Ecp2* gene is 554 bp with a 56 bp intron (Figure 18). The *C. fulvum* Ecp2 homolog also contained a predicted N-terminal signal peptide. The *C. fulvum* Ecp2 homolog (Clafu1:197200) has also been identified previously (VAN DEN ACKERVEKEN *et al.* 1993). Confirmation of the previously identified Ecp2 effectors in *M. fijiensis* and *C.*

fulvum provides confidence for the predicted *C. zeina* (Mkushi) Ecp2 homolog. An interesting finding is that the *Ecp2* intron seems to be conserved among all the homologs identified in this study.

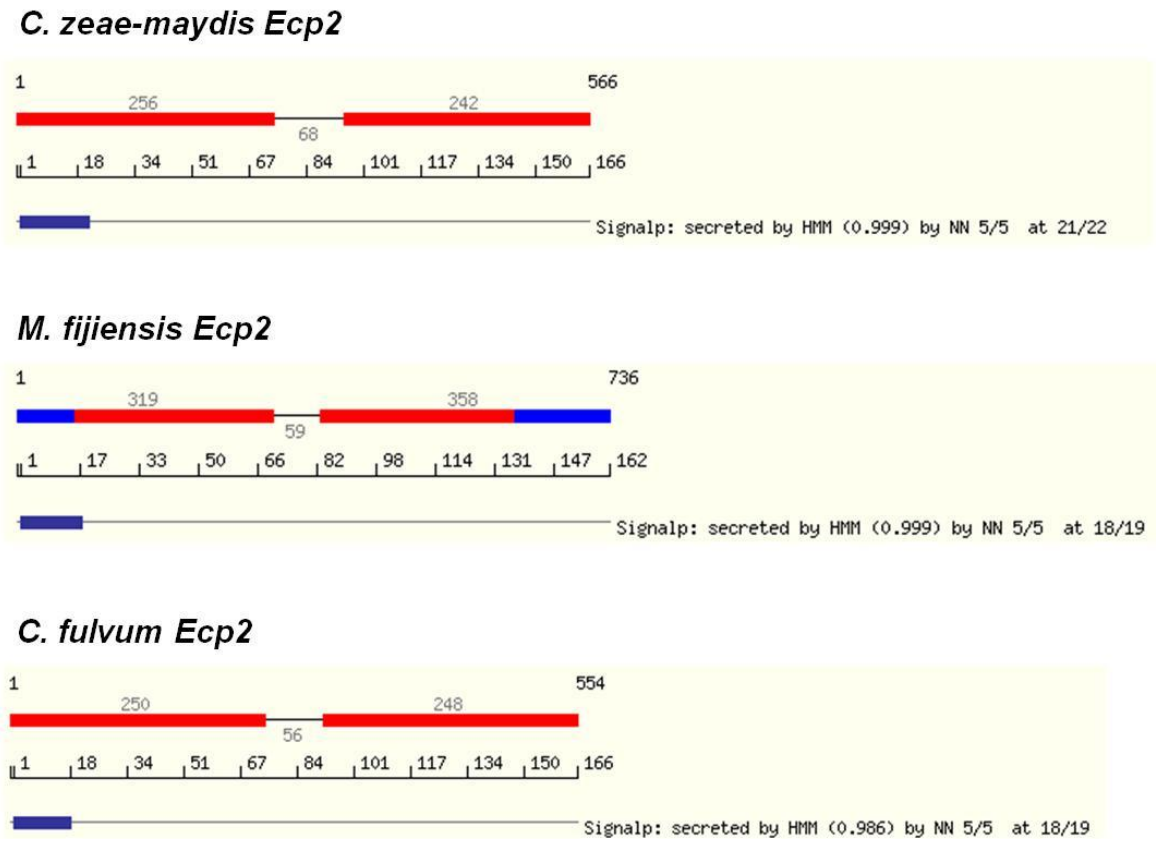


Figure 18. Predicted gene structures of the *C. zeina* (Mkushi) *Ecp2* homologs present in *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum*. The coding regions of the predicted gene homologs were indicated by red bars and the untranslated regions (UTRs) were indicated with blue bars if present. Indicate below each gene prediction, are conserved functional domains (if present) and signal sequences (indicated in blue). *C. zeae-maydis Ecp2* had a predicted size of 566 bp and contained a 68 bp intron. *M. fijiensis Ecp2* was predicted to be 736 bp and has a 59 bp intron. *C. fulvum Ecp2* was predicted to be 554 bp and has a 56 bp intron. The *Ecp2* intron seems to be conserved in the homologs.

4.2.5 Protein alignments of Ecp2

Through BLAST search analysis using the predicted *C. zeina* (Mkushi) Ecp2 protein sequence against the protein sequence database available from GenBank, fungal homologs could be obtained. The closest available homolog of the *C. zeina* Ecp2 protein was *C. fulvum* Ecp2. The *C. fulvum* Ecp2 protein (CAA78401.1) shared 55% identity with the *C. zeina* Ecp2 protein with an alignment region of 169 aa. The best hits as well as the other

relevant homologous protein sequences (identified in section 4.2.2 and 4.2.4) were used in the alignment as seen in Figure 19. Residue conservation was indicated by the bar graph (blue) in terms of percentage conservation between sequences in the alignment. Gaps were indicated as hyphens. Four protein sequences were included in the Ecp2 protein alignment (Figure 19), namely the predicted *C. zeina* (Mkushi) Ecp2 protein, the predicted *C. zeae-maydis* (Cerzm1:53627) and *M. fijiensis* Ecp2 (Mycfi1:52972) proteins, as well as the *C. fulvum* Ecp2 protein (CAA78401.1) (VAN DEN ACKERVEKEN *et al.* 1993).

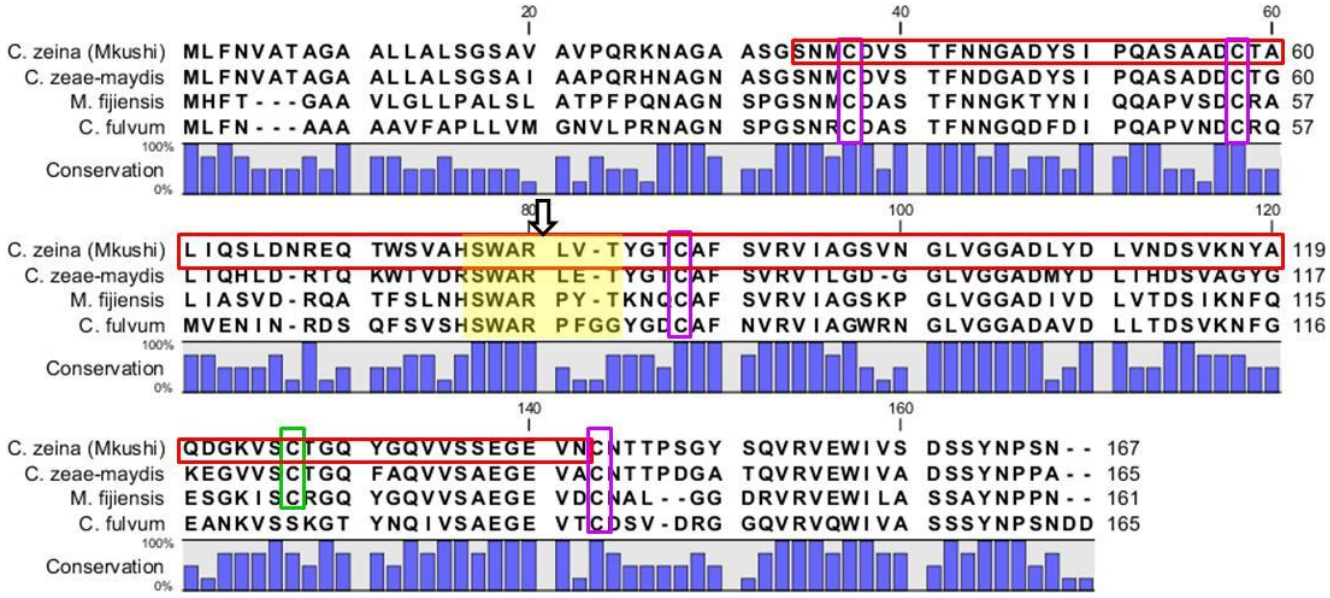


Figure 19. Ecp2 protein alignment generated with Mafft software and visualised with CLC Bio (main workbench) software. The alignment includes the most homologous proteins as well as other relevant protein sequences. The percentage conservation between the sequences at each amino acid in the alignment was indicated by bar graphs (Blue). Cysteine residues were indicated with green boxes and conserved cysteine residues were indicated with purple boxes. The red box indicates the location of the Hce2 domain in *C. zeina* (Mkushi) Ecp2. The yellow highlighted region indicates the residues that flank the splice junction of the intron from the *Ecp2* sequence. The black arrow shows the position where the intron has been spliced out for all four the aligned proteins.

The four Ecp2 proteins shared similar protein sequences, especially at the putative necrosis inducing (Hce2) domain that is indicated by the red box on the *C. zeina* (Mkushi) Ecp2 protein (Figure 19). The Ecp2 alignment contains four conserved cysteine residues, as indicated by the purple boxes (Figure 19). Another cysteine residue present in all the Ecp2 proteins, except *C. fulvum* Ecp2 protein, is found within the Hce2 domain region as indicated by the green box in figure 19. *C. zeina*, *C. zeae-maydis,* and *M. fijiensis* Ecp2 proteins contain five cysteine residues each, whereas the *C. fulvum* Ecp2 protein contains four

cysteine residues. General gap regions are present at the following residues in the alignment: 5-7 (absent in *C. zeina* and *C. zeae-maydis*), 83 and 169-170 (absent in *C. fulvum*). The translated regions that flanked the putatively conserved intron region in the *Ecp2* homologs were indicated by the yellow shading in the Ecp2 alignment (Figure 19). A black arrow was used to indicate the exact region where the intron was spliced out for all four proteins in the alignment. Good identity between all four Ecp2 proteins downstream of the intron splice site indicates that the intron position was annotated correctly. Four residues (77-80 in the alignment) of the upstream flanking region were conserved in all the Ecp2 proteins.

A pairwise comparison was made between the aligned Ecp2 protein sequences to determine the sequence identity and similarity. *C. zeae-maydis* Ecp2 shared the highest sequence identity (76.80%) and similarity (83.90%) to the *C. zeina* (Mkushi) Ecp2 protein (Figure 20). The *C. fulvum* Ecp2 protein shared the lowest identity (54.40%) with the *C. zeina* Ecp2 protein when compared to the other sequences in the alignment. However, the similarity between the *C. fulvum* and *C. zeina* (Mkushi) Ecp2 proteins were 68.40% (Figure 20). Even though the other proteins in the alignment shared relatively low (50.00-57.40%) sequence identity, the sequence similarity ranged from 64.10-71% (Figure 20).

1. *C. zeina* (Mkushi)
2. *C. zeae-maydis*
3. *M. fijiensis*
4. *C. fulvum*

	1	2	3	4
1		83.90	71.00	68.40
2	76.80		64.10	64.70
3	57.40	50.90		75.20
4	54.40	50.00	57.00	

% Similarity

% Identity

Figure 20. Pairwise comparisons between the aligned Ecp2 protein sequences generated with CLC Bio Main Workbench software (Aarhus, Denmark) and EMBOSS Needle. The pairwise comparison was made between the four Ecp2 protein sequences (1) *C. zeina* (Mkushi), (2) *C. zeae-maydis*, (3) *M. fijiensis*, and (4) *C. fulvum*. The upper comparison gives the percentage amino acid similarity values in the alignment, whereas the lower comparison gives the percentage identity. The difference between amino acid identity and similarity is outlined in the relevant materials and methods section. This figure compliments the Ecp2 protein alignment shown in Figure 19.

4.2.6 Phylogenetic analysis of *Ecp2* homologs

Following the Ecp2 protein alignment results, phylogenetic analysis was done to determine the evolutionary relationships between specific aligned Ecp2 sequences. *Ecp2* phylogenetic analysis involved nucleotide alignments for the same reason as mentioned in section 4.1.6. The best predicted substitution model based on the *Ecp2* alignment was the Tamura 3-parameter (T92) model with invariant sites. Figure 21 shows the combined phylogenetic trees of ML and NJ analysis based on the *Ecp2* alignment. The ML and NJ phylogenetic trees showed identical branching in each case. For each phylogenetic tree, the bootstrap values above 75% are indicated on the nodes as NJ/ML values. Bootstrap values indicated two separate clades, with *C. zeina* and *C. zeae-maydis* grouping together, and *M. fijiensis* and *C. fulvum* forming a different clade. Bootstrap percentages of both analyses for both groups were 100%. No rooting sequence was available for the *Ecp2* phylogenetic analysis.

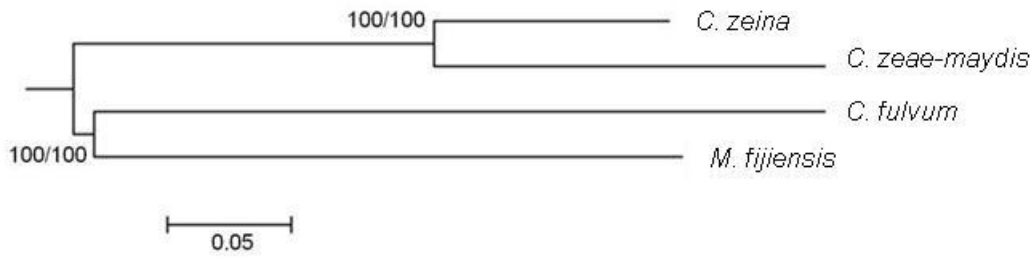


Figure 21. Combined phylogenetic trees for *Ecp2* constructed based on the nucleotide alignment of the different *Ecp2* nucleotide sequences. Bootstrap values are indicated on the nodes as neighbour-joining analysis/maximum likelihood analysis. Bootstrap values were obtained from 1000 replicates and indicated as percentages.

4.3 Evidence for the presence of the *Ecp6* effector homolog in *C. zeina*

4.3.1 The presence of an *Ecp6* homolog in the *C. zeina* (Mkushi) genome

The *C. fulvum Ecp6* query mRNA sequence (EU730588.1) (Table 1) was used for BLASTn searches against the draft *C. zeina* (Mkushi) genome. BLASTn results showed a higher percentage identity as well as better statistical significance when the *C. fulvum* (EU730588.1) sequence was queried against the *C. zeina* Illumina data assembled with Velvet compared to the other *C. zeina* datasets. The putative *C. zeina* (Mkushi) *Ecp6* BLAST hit showed 70% identity to the *C. fulvum Ecp6* query sequence (EU730588.1) with a 358 bp region aligned region with no gaps, a score of 180, and a statistical significance value of 3e-39.

4.3.2 *C. zeina* (Mkushi) *Ecp6* preliminary predicted gene structure

The BLASTn hit indicated a possible *C. zeina* (Mkushi) *Ecp6* homolog; however the sequence needed to be annotated in order to determine the preliminary predicted *C. zeina* (Mkushi) *Ecp6* gene structure in terms of intron/exon boundaries. Both prediction programmes (FGENESH and Augustus) predicted identical *C. zeina* (Mkushi) *Ecp6* gene structures. The preliminary predicted *C. zeina* (Mkushi) *Ecp6* gene was similar in size (765 bp) to the *C. fulvum Ecp6* gene (EU730588.1) (687 bp) (BOLTON *et al.* 2008) and also had no predicted introns (Figure 22).

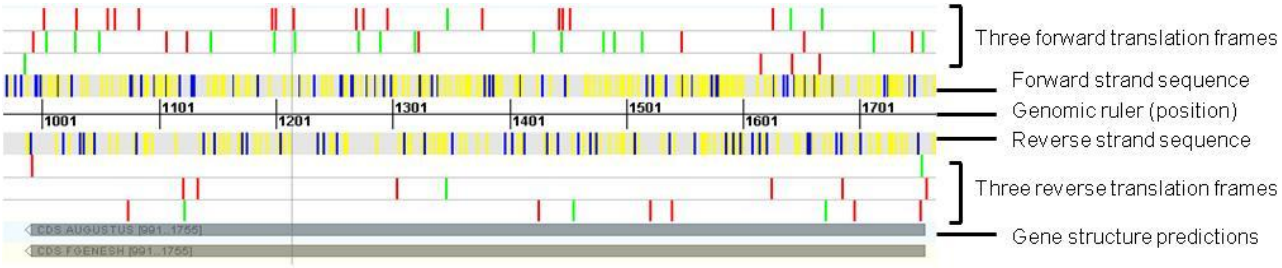


Figure 22. *C. zeina* (Mkushi) *Ecp6* preliminary gene prediction visualised in GenomeView comparing the expected gene structures and orientation as predicted by Augustus and FGENESH. The annotations made by both prediction programmes were similar. The grey bar at the bottom of the figure represents the prediction made by FGENESH, whereas the blue bar represents the prediction made by Augustus. Both predictions were present on the reverse strand. The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. The blue and yellow vertical lines represent possible acceptor and donor sites for splicing in the forward and reverse sequence.

Reciprocal BLASTn and tBLASTx analysis against the GenBank showed that the closest homolog for the preliminary predicted *C. zeina* (Mkushi) *Ecp6* gene was from *M. fijiensis* (XM007931055) that shared 68% identity (514 bp alignment region) on a nucleotide level and 57% identity (133 aa aligned region) on a translated level to the predicted *C. zeina* (Mkushi) *Ecp6* (Table 7). However, *C. fulvum Ecp6* (EU730588.1) was one of the hits obtained and shared 70% identity (358 bp aligned region) on a nucleotide level and 60% identity (121 aa aligned region) on a translated level to the predicted *C. zeina* (Mkushi) *Ecp6* (Table 7). Another significant BLAST hit was from *M. graminicola* (XM003848615) and was summarised in Table 7.

Table 7. A summary of the most relevant BLASTn and tBLASTx results obtained from reciprocal BLAST analysis of the preliminary predicted *C. zeina* (Mkushi) *Ecp6* gene.

BLASTn Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
M. fijiensis carbohydrate-binding module family 50 protein	(XM007931055)	68	206	2e-43	13	514 bp
C. fulvum Ecp6	(EU730588.1)	70	180	3e-36	-	358 bp
M. graminicola hypothetical protein	(XM003848615)	66	140	2e-25	14	576 bp
tBLASTx Hits	**Accession number**	**BLAST hit information**				
		Identity (%)	**Score**	**E-value**	**Gaps**	**Aligned region**
M. fijiensis carbohydrate-binding module family 50 protein	(XM007931055)	57	403	1e-62	-	133 aa
C. fulvum Ecp6	(EU730588.1)	60	412	3e-61	-	121 aa
M. graminicola hypothetical protein	(XM003848615)	57	369	4e-55	-	126 aa

The preliminary *C. zeina* (Mkushi) Ecp6 protein was predicted to be 256 aa in size with three conserved lysin motifs (LysM) (cd00118). The three conserved LysM motifs are located at 67-112, 143-187, and 200-243 aa regions of the predicted *C. zeina* (Mkushi) Ecp6 protein (Figure 23). LysM motifs are generally small, globular domains that are widespread across bacteria and eukaryotes. LysM domains commonly bind to peptidoglycans in bacteria and to chitin in eukaryotes (BATEMAN and BYCROFT 2000; BIRKELAND 1994; BUIST *et al.* 2008; JORIS *et al.* 1992; PONTING *et al.* 1999; SPAINK 2004; TURNER *et al.* 2004).

Figure 23. Conserved LysM domains in the preliminary predicted *C. zeina* (Mkushi) *Ecp6.*
The preliminary predicted *C. zeina* (Mkushi) *Ecp6* has three conserved lysin motifs (LysM). LysM domains are widespread and considered to be small globular domains. LysM domains have been shown to bind chitin monomers in eukaryotes. The grey bar represents the preliminary predicted *C. zeina* (Mkushi) *Ecp6* query nucleotide sequence and the orange bars represent the regions of the conserved LysM domains. The figure was drawn to scale.

4.3.3 Confirmation of *C. zeina* (Mkushi) *Ecp6* contig assembly

4.3.3.1 PCR analysis with gene-specific primers

PCR analysis with *Ecp6* specific primers was undertaken to be able to determine if the *C. zeina* (Mkushi) *Ecp6* contig was assembled correctly. The *Ecp6* primers were designed in such a way that different fragments of known length would be amplified from *C. zeina* (Mkushi) gDNA (Figure 24). The primers designed to the preliminary *C. zeina* (Mkushi) *Ecp6* gene are summarized in Table 8.

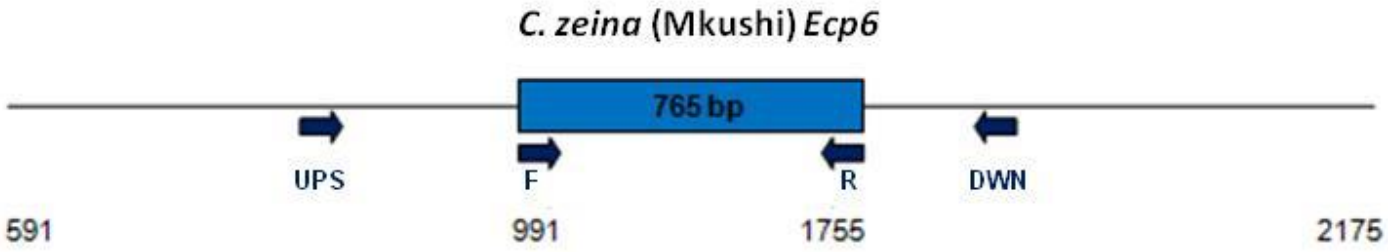


Figure 24. Gene-specific primers designed to the preliminary predicted *C. zeina* (Mkushi) *Ecp6*.
The navy arrows indicate the designed primer regions. The blue box indicates the *Ecp6* exon. The predicted exon size of the preliminary *C. zeina* (Mkushi) *Ecp6* is 765 bp. The figure was not drawn to scale.

The preliminary *C. zeina Ecp6* gene had only one predicted exon with no introns, and therefore primers were designed to amplify four overlapping amplicons. The primer combinations were (1) UPS and DWN primers, (2) UPS and R primers, (3) F and DWN primers, and (4) F and R primers. Primer pairs did not amplify at 54-57˚C (results not shown). However, only two of the primer combinations, combinations (2) and (4), amplified the desired products at 58˚C (Figure 23). The expected product sizes were 1008 bp, 817 bp, 956 bp, and 765 bp respectively. A possible reason why primer combinations (1) and (3) did

not work, could be that the designed DWN primer, which is present in both combinations (1) and (3), was designed to a region which does not form part of the current contig assembly. As these primers were designed to preliminary predicted *C. zeina* (Mkushi) *Ecp6* gene regions, the downstream (3' untranslated region) prediction might not have been accurate. This indicates that the *Ecp6* contig assembly cannot be verified. However, it is important to note that the preliminary predicted *Ecp6* gene could be amplified by using the F and R primers designed to the start and stop regions of the gene. Further validation by means of RNAseq read mapping was needed to give more insight into the *Ecp6* gene structure.

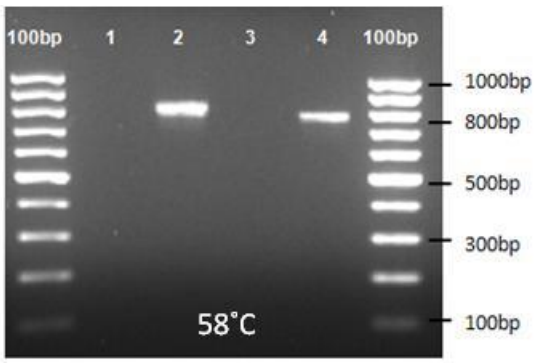


Figure 25. *C. zeina* (Mkushi) *Ecp6* contig confirmation with gene-specific primers.
The *C. zeina* (Mkushi) *Ecp6* contig assembly was verified through PCR amplification with gene-specific primers. The subsequent amplification products were analysed on a 1.3% agarose gel. All products were amplified from *C. zeina* (Mkushi) genomic DNA and a 100 bp ladder (Thermo Scientific, Waltham, Massachusetts) was used to measure product sizes. *Ecp6* products amplified at 58˚C with UPS and DWN primers (1), UPS and R primers (2), F and DWN primers (3), and F and R primers (4). The expected amplicon sizes for *Ecp6* were 1008 bp (1), 817 bp (2), 956 bp (3), and 765 bp (4). No products were amplified from primer combinations (1) and (3).

4.3.3.2 *C. zeina* (Mkushi) *Ecp6* annotation through mapping of RNAseq reads

Mapping of both *in planta* and *in vitro* RNAseq reads as well as the *M. fijiensis Ecp6* homolog (XM007931055) was used as a reference to improve the *Ecp6* annotation. The *in planta* RNAseq reads that mapped to the *C. zeina* (Mkushi) *Ecp6* gene region (Figure 26A) showed very low coverage and therefore could not support the annotation. However, the *C. zeina* (Mkushi) *Ecp6* gene was annotated manually based on the *M. fijiensis Ecp6* gene homolog (XM007931055) and according to the possible start and stop sites in the three different translation frames. The *in vitro* RNAseq reads (Figure 26 B) supported the intron region in the newly annotated *Ecp6* gene.

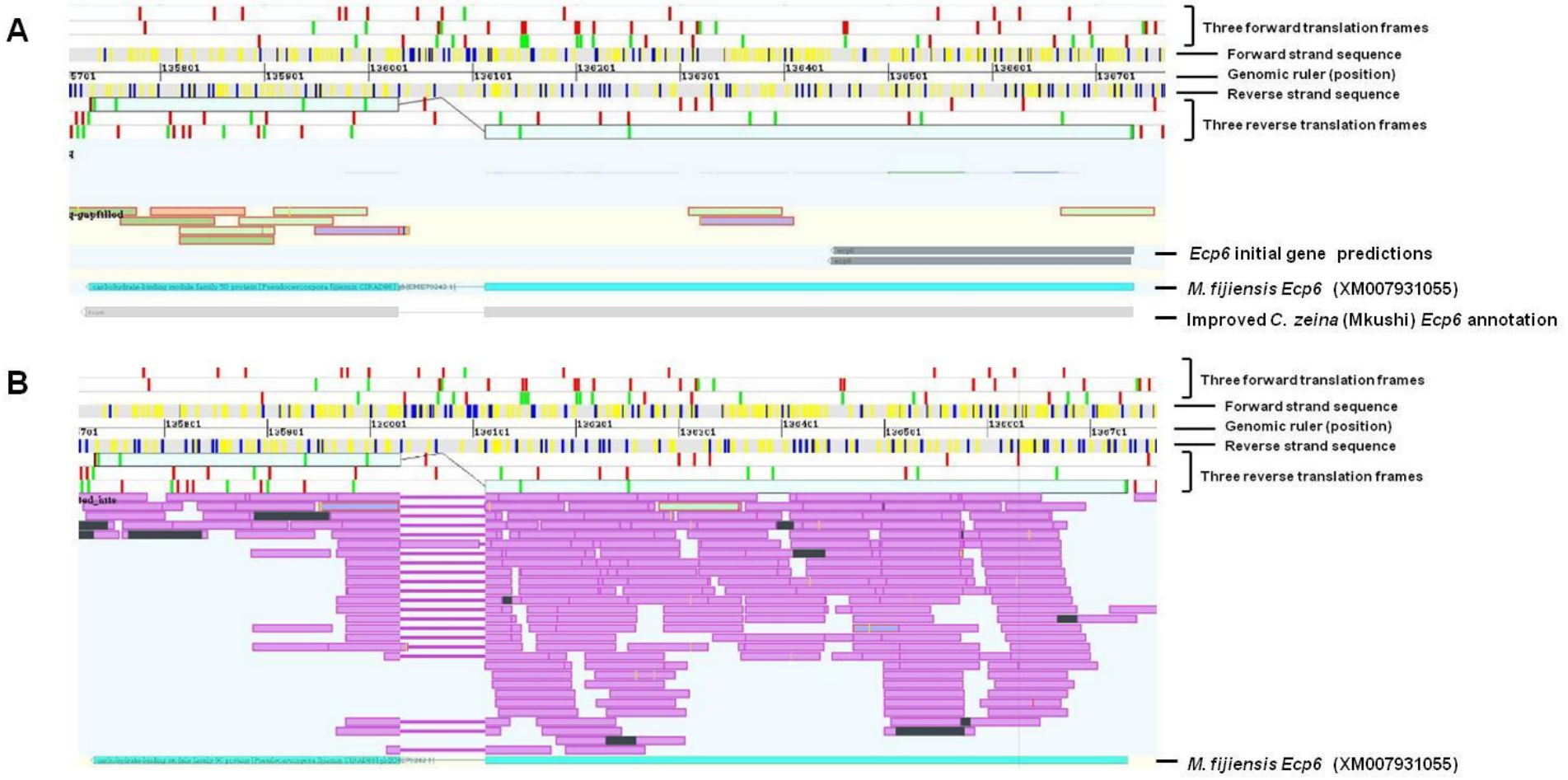


Figure 26. I**mproved *C. zeina* (Mkushi) *Ecp6* annotation.**
In planta RNAseq reads (A) and *in vitro* RNAseq reads (B) were mapped to the contig containing the *C. zeina* (Mkushi) *Ecp6* gene. Only the region containing the annotated *Ecp6* gene was shown. The coloured bars represent the RNAseq reads mapped to the whole contig. Only a few *in planta* RNAseq reads mapped to the *Ecp6* gene region providing very low coverage (A). The *in vitro* RNAseq reads showed a higher coverage and supported the improved annotation (B). The *M. fijiensis Ecp6* gene homolog (XM007931055) was used as a reference for annotation purposes and is represented by the light blue horizontal bar. The annotation of the *Ecp6* gene was done manually by taking into account possible translation start- and stop sites. The improved *Ecp6* annotation is represented by the grey horizontal bar and is longer than the initial *C. zeina Ecp6* gene prediction (dark grey horizontal bars) (A). The red and green vertical lines represent possible translation start and stop sites in the three different reading frames for annotation purposes. *Ecp6* is present on the reverse strand. See materials and methods for RNAseq read colours and their descriptions.

The newly annotated E*cp6* gene is larger than the initial annotation of *Ecp6*. Instead of consisting of a single exon that is 765 bp in size, the improved *Ecp6* annotation consists of two exons and a single intron. The first exon is 624 bp and is followed by an 83 bp intron and a second exon of 297 bp based on empirical evidence. As the annotated *C. zeina* (Mkushi) *Ecp6* gene is larger than the initial preliminary annotation, new gene-specific primers had to be designed for the *Ecp6* gene. The primers were designed to amplify overlapping parts of the *C. zeina* (Mkushi) *Ecp6* gene, as the gene was too big (larger than 1000 bp) to be sequenced in a single reaction. The regions of the new primers designed to the annotated *Ecp6* gene are indicated in figure 27 and information on the primer sequences can be found in Table 2. In order to confirm the new contig assembly according to the *Ecp6* annotation, the Ecp6 fragments were amplified through PCR reactions. The PCR products were analysed on a 2% agarose gel and the product lengths could be determined through a middle range DNA ladder (Thermo Scientific, Waltham, Massachusetts) (Figure 6). For the *Ecp6* amplifications, two amplicons were amplified namely, *Ecp6A* and *Ecp6B*. The amplicons were amplified using primer combinations (1) Ecp6F2A and Ecp6R2A primers, and (2) Ecp6F2B and Ecp6R2B primers. The expected amplicon sizes of 953 bp for *Ecp6A* and 863 bp for *Ecp6B* were obtained at 59˚C as seen in Figure 6. Because both these regions were successfully amplified from *C. zeina* (Mkushi) gDNA, the *Ecp6* contig assembly was assumed to be correct. Both *Ecp6A* and *Ecp6B* amplified products were cloned into the pTZ57R plasmid (Appendix, Figure S1) and sequenced. The sequenced *Ecp6* products are shown in Figure 28, which support the assembly of the contig around the *Ecp6* region.

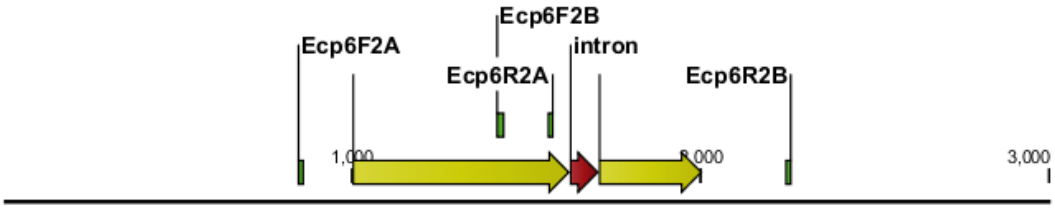


Figure 27. *Ecp6* primer sequences designed to the annotated *C. zeina* (Mkushi) *Ecp6* gene.
The improved *Ecp6* annotation shows a 1004 bp gene region, consisting of two exons that are 624 bp and 297 bp, respectively (indicated by the yellow arrows) and a single intron of 83 bp (indicated by the red arrow). Primer sequences were designed in such a way that the two amplicons would overlap each other. Primers are indicated by the vertical green bars. Ecp6F2A and Ecp6R2A produce an amplicon (*Ecp6A*) of 735 bp, and the primer set Ecp6F2B and Ecp6R2B produce an amplicon (*Ecp6B*) of 848 bp. The overlapping region between the two primer sets is 165 bp in size. The figure was drawn to scale.

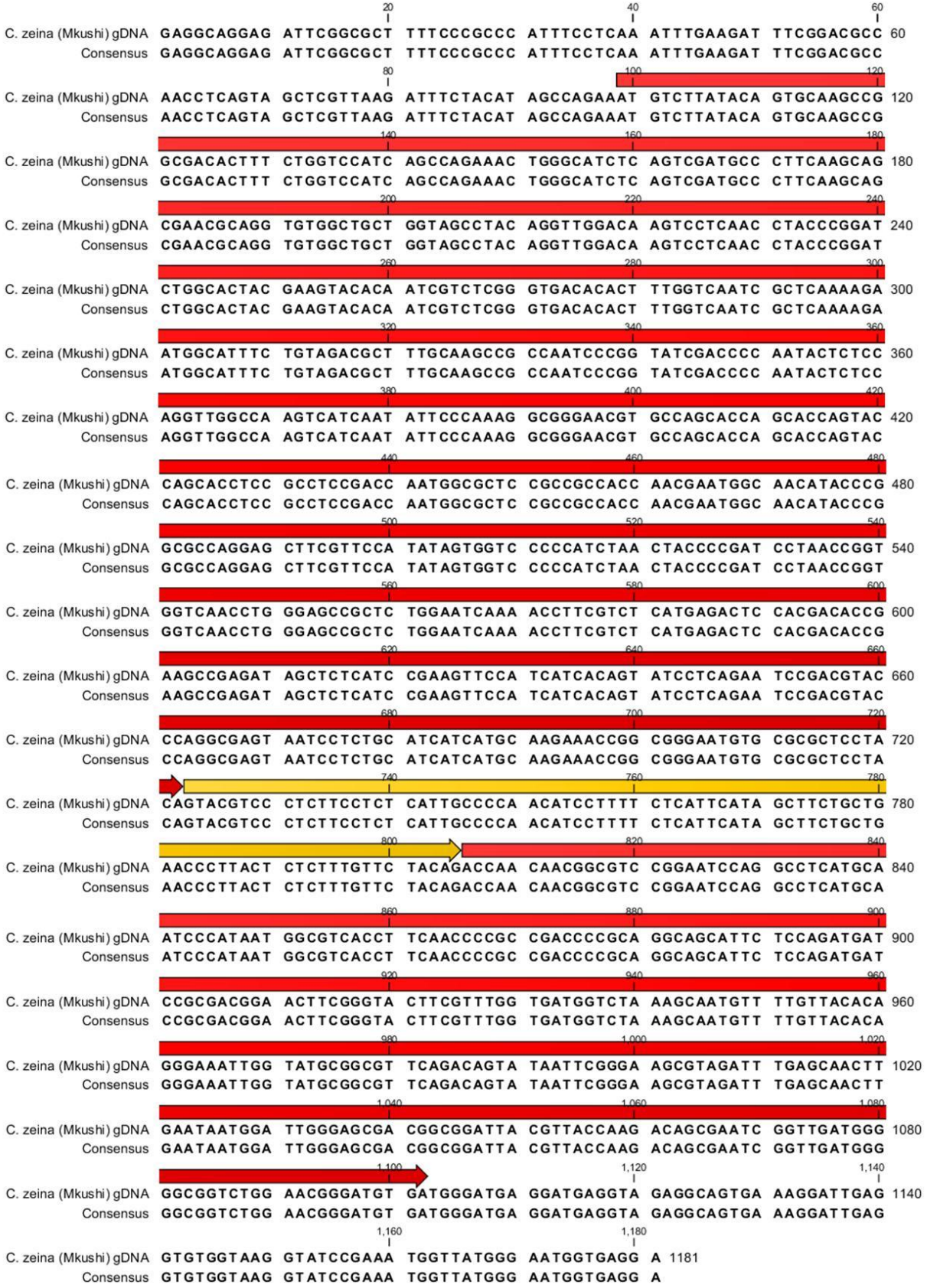


Figure 28. Confirmation of the *C. zeina* (Mkushi) *Ecp6* contig assembly.
The *C. zeina* (Mkushi) *Ecp6* cloned gene regions were sequenced from plasmid DNA using standard M13 primers. Forward and reverse sequencing products were combined and represented as a consensus sequence. *C. zeina* (Mkushi) gDNA containing the *Ecp6* gene region (red arrows indicate exons and the yellow arrow indicates the intron position) with approximately 100 bp up- and downstream regions was used as a reference.

4.3.4 *C. zeina* (Mkushi) *Ecp6* homologs present in *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum*.

The annotated *C. zeina* (Mkushi) *Ecp6* gene was used as a query sequence for BLAST analysis against the *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum* genomes to determine if *Ecp6* homologs were present. The genomes were chosen based on the reasons mentioned in section 4.1.4. BLASTn and tBLASTx searches were made against the respective genomes and the results were summarized in Table 8.

Table 8. A summary of the most relevant BLASTn and tBLASTx results obtained for the annotated *C. zeina* (Mkushi) *Ecp6* effector query sequences to the *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum* genomes.

BLASTn query	Genome	BLAST hit information					Protein ID
		Identity	Score	E-value	Gaps	Aligned region	
C. zeina (Mkushi) *Ecp6*	*C. zeae-maydis*	-	-	-	-	-	-
C. zeina (Mkushi) *Ecp6*	*M. fijiensis*	97%	26	8e-06	-	30 bp	Mycfi1:34332
C. zeina (Mkushi) *Ecp6*	*M. graminicola*	-	-	-	-	-	-
C. zeina (Mkushi) *Ecp6*	*C. fulvum*	-	-	-	-	-	-
tBLASTx query	**Genome**	**BLAST hit information**					**Protein ID**
		Identity	**Score**	**E-value**	**Gaps**	**Aligned region**	
C. zeina (Mkushi) *Ecp6*	*C. zeae-maydis*	35%	45	3e-06	-	60 aa	Cerzm1:88418
C. zeina (Mkushi) *Ecp6*	*M. fijiensis*	63%	144	3e-63	-	96 aa	Mycfi1:34332
C. zeina (Mkushi) *Ecp6*	*M. graminicola*	59%	93	1e-21	-	70 aa	Mycgr1:17609
C. zeina (Mkushi) *Ecp6*	*C. fulvum*	56%	81	1e-19	-	63 aa	Clafu1:194347

C. zeina (Mkushi) *Ecp6* homologs were found in *C. zeae-maydis*, *M. fijiensis*, *M. graminicola, and C. fulvum*. *C. zeae-maydis Ecp6* (Cerzm1:88418) shared 35% identity (60 aa aligned region) with the *C. zeina* (Mkushi) *Ecp6* on a translated level (Table 8). The *C. zeae-maydis Ecp6* gene was predicted to be 923 bp with an intron of 155 bp and contained a conserved Lysozyme-like domain (Figure 29). The *C. zeae-maydis* Ecp6 predicted homolog contained no N-terminal SP. No *Ecp6* homolog has previously been identified in *C. zeae-maydis*. *M. fijiensis Ecp6* (Mycfi1:34332) shared 97% identity (30 bp aligned region) with the *C. zeina* (Mkushi) *Ecp6* on a nucleotide level, but 63% identity (96 aa aligned region) on a translated level (Table 8). The predicted *M. fijiensis Ecp6* gene was 579 bp with no introns and contained two conserved LysM domains (IPR002482) (Figure 29). The *M. fijiensis* Ecp6 homolog also contained an N-terminal SP. The *M. fijiensis* homolog (Mycfi134332) has not been identified previously. The *M. fijiensis Ecp6* homolog identified previously (Mycfi1: 86817) (BOLTON *et al.* 2008) has three LysM domains. *M. graminicola* Ecp6 (Mycgr1:17609) shared 59% identity (70 aa aligned region) to the *C. zeina* (Mkushi) Ecp6 on a translated

level (Table 8). The predicted *M. graminicola Ecp6* gene was 585 bp with no introns (Figure 29). The *M. graminicola* Ecp6 homolog contained no predicted domains and no N-terminal SP. The *M. graminicola* homolog (Mycgr1:17609) has not been identified previously. Three *M. graminicola Ecp6* homologs have been identified previously, namely *Mg1LysM* (294 bp), *MgxLysM* (312 bp), and *Mg3LysM* (699 bp) and contained one, no, and three LysM domains respectively (MARSHALL *et al.* 2011). *C. fulvum Ecp6* (Clafu1:194347) shared 56% identity (63 aa aligned region) with the *C. zeina* (Mkushi) Ecp6 on a translated level (Table 8). The predicted *C. fulvum Ecp6* gene was 1106 bp with an intron of 59 bp and an N-terminal SP, but no conserved domains (Figure 29). The *C. fulvum* homolog (Clafu1:194347) has also not been identified previously. The previously identified *C. fulvum Ecp6* (687 bp) has three predicted LysM domains (BOLTON *et al.* 2008).

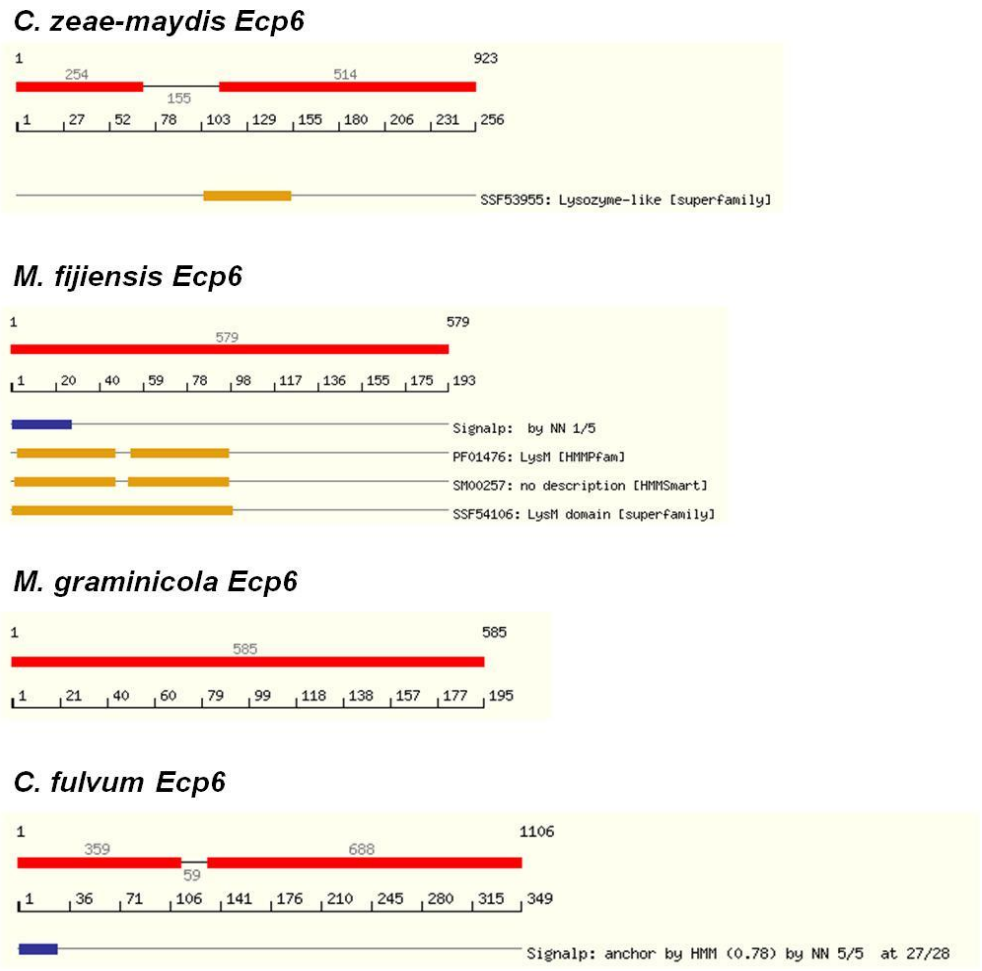


Figure 29. Predicted gene structures of the *C. zeina* (Mkushi) *Ecp6* homologs present in *C. zeae-maydis*, *M. fijiensis*, *M. graminicola,* and *C. fulvum*. The coding regions of the predicted gene homologs were indicated by red bars. Indicated below each gene prediction, are the conserved functional domains (in yellow) and signal sequences (in blue). *C. zeae-maydis Ecp6* had a 935 bp predicted gene region with an intron of 155 bp. The *M. fijiensis Ecp6* gene had a predicted exon of 579 bp with no introns and the *M. graminicola Ecp6* gene was 585 bp, also with no intron sequences. The predicted *C. fulvum Ecp6* gene was 1106 bp with an intron of 59 bp.

4.3.5 Protein alignments of Ecp6

Through BLAST search analysis using the predicted *C. zeina* (Mkushi) Ecp6 protein sequence against the protein sequence database, fungal homologs were obtained. The reciprocal BLASTp result indicated that the closest homolog for *C. zeina* (Mkushi) Ecp6 was *M. fijiensis* carbohydrate-binding module family 50 protein (XP007929625) that shared 59% identity (210 aa aligned region) with the *C. zeina* (Mkushi) Ecp6 on a protein level. The *M. fijiensis* carbohydrate-binding module family 50 protein (XP007929625) is exactly the same protein as the identified homolog *M. fijiensis* (Mycfi1:34332) described in section 4.3.4. The best hits as well as other relevant homologous protein sequences identified in section 4.3.4 were used in the Ecp6 protein alignment (Figure 30). Residue conservation was indicated by the bar graph (blue) below the aligned sequences in terms of percentage conservation between sequences in the alignment. Gaps in the alignment were indicated as hyphens. The Ecp6 protein alignment (Figure 30) contains the following Ecp6 proteins: the annotated *C. zeina* (Mkushi) Ecp6, *M. fijiensis* Ecp6 homolog (Mycfi1:34332), *C. zeae-maydis* Ecp6 homolog (Cerzm1:88418), *M. graminicola* Ecp6 (Mycgr1:17609) proteins as well as the *C. fulvum* Ecp6 protein (EU73088) (BOLTON *et al.* 2008). Many gaps are present in the alignment and most gaps can be attributed to the difference in protein lengths ranging from 192 to 306 residues in size. No conserved cysteine residues could be found in the alignment. The cysteine residues that were present in the alignment were indicated with green boxes. The *C. zeina* (Mkushi) Ecp6 protein has two cysteine residues, whereas the *M. fijiensis* Ecp6 protein (Mycfi1:34332) only has a single cysteine residue. The *C. zeae-maydis* (Cerzm1:88418) and *M. graminicola* Ecp6 (Mycgr1:17609) proteins have five and four cysteine residues, respectively. The *C. fulvum* Ecp6 protein has the most cysteine residues, but only one of the ten cysteine residues are shared with the other proteins.

The Ecp6 proteins shared relatively low sequence similarity, but some sequence identity could be seen in the regions where the three lysin (LysM) domains are found, as indicated by the red, orange, and yellow boxes in the alignment (Figure 30). The region of the Ecp6 protein alignment showing the highest sequence identity was around the first LysM domain (region 116-220 of the alignment). A pairwise comparison (Figure 31) of percentage amino acid identity and similarity between the sequences in the Ecp6 protein alignment was focussed on the 116-220 aligned region. Even though it is expected that the *C. zeina* and *C. zeae-maydis* Ecp6 proteins should be more similar in sequence, the proteins differ in conserved domain regions as well as in protein lengths. The *C. zeina* (Mkushi), *C. zeae-maydis*, and *M. graminicola* predicted Ecp6 proteins shared no sequence identity in the region around the first LysM domain (Figure 31). The *M. fijiensis* Ecp6 protein shared 68.40% sequence identity and 85.70% sequence similarity (Figure 31) to the *C. zeina* Ecp6

protein in the analysed region. The *C. zeina* (Mkushi) and the *C. fulvum* Ecp6 proteins shared low sequence identity (33.60%) and similarity (56.10%) (Figure 31) in the analysed aligned region. Low sequence identity (34.30%) and similarity (54.30) was also observed between the *C. fulvum* and *M. fijiensis* Ecp6 proteins in the aligned region (Figure 31). No comparison could be drawn from the *C. zeae-maydis* and *M. graminicola* Ecp6 sequences in the specified region of the alignment.

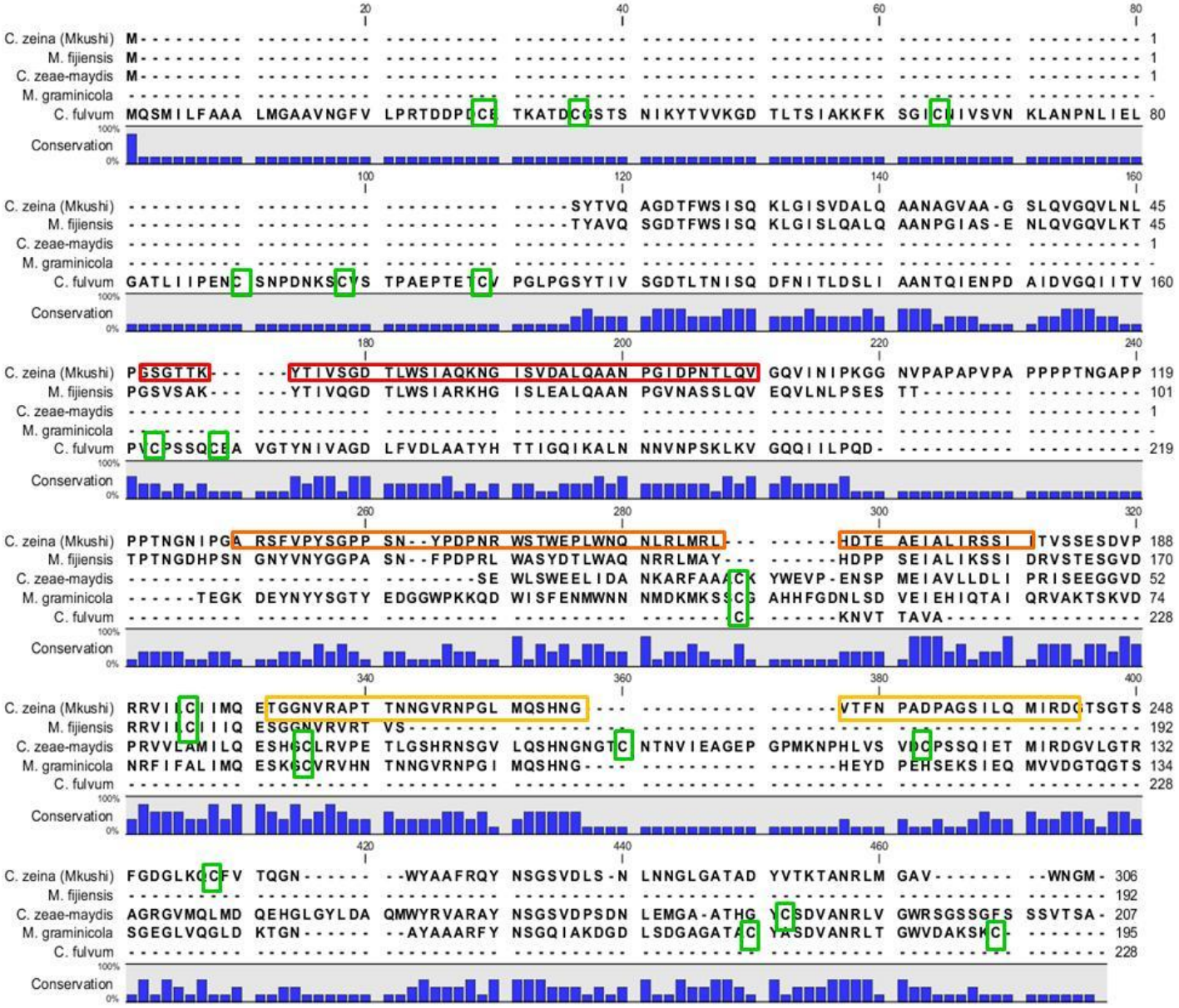


Figure 30. Ecp6 protein alignment generated with Mafft software and visualised with CLC Bio (Main workbench) software. Homologous Ecp6 proteins were included in the alignment. Amino acid residues that differed between proteins were shaded blue and gaps were indicated with hyphens. Cysteine residues were indicated with green boxes. No conserved cysteine residues were present. The red, orange, and yellow boxes indicate the locations of the three putative carbohydrate-binding LysM domains present in the *C. zeina* (Mkushi) Ecp6 protein.

1. *C. zeina* (Mkushi)
2. *M. fijiensis*
3. *C. zeae-maydis*
4. *M. graminicola*
5. *C. fulvum*

	1	2	3	4	5
1		85.70	0.00	0.00	56.10
2	68.40		0.00	0.00	54.30
3	0.00	0.00		N/A	0.00
4	0.00	0.00	N/A		0.00
5	33.60	34.30	0.00	0.00	

% Similarity

% Identity

Figure 31. Pairwise comparisons between the aligned Ecp6 protein sequence region (region 116-220 in figure 30) generated with CLC Bio Main Workbench software (Aarhus, Denmark) and EMBOSS Needle. The pairwise comparison was made between the four Ecp6 protein sequences (1) *C. zeina* (Mkushi), (2) *M. fijiensis*, (3) *C. zeae-maydis*, (4) *M. graminicola*, and (5) *C. fulvum*. The upper comparison gives the percentage amino acid similarity values in the alignment region, whereas the lower comparison gives the percentage identity. The difference between amino acid identity and similarity is outlined in the relevant materials and methods section. This figure compliments the Ecp6 protein alignment region around the first LysM domain 116-220 (Figure 30).

4.3.6 Phylogenetic analysis of *Ecp6* homologs

Following the Ecp6 protein alignment results, phylogenetic analysis was done to determine the evolutionary relationships between specific aligned sequences. Nucleotide sequence alignments instead of protein alignments were made for the putative *Ecp6* effector homologs as described in the *Avr4* phylogenetic analysis section. Alignments were made with the full *Ecp6* nucleotide sequences. The best predicted substitution model based on the *Ecp6* alignment was the Kimura 2-parameter model with uniform rates among sites. Figure 32 shows the combined phylogenetic trees of ML and NJ analysis based on the *Ecp6* alignment and substitution model. The ML and NJ phylogenetic trees showed identical branching in each case. For each phylogenetic tree, the bootstrap values above 75% are indicated on the nodes as NJ/ML values. The combined *Ecp6* tree supports a closer phylogenetic relationship between the *C. zeina* and *M. fijiensis* sequences, as these two sequences grouped together with high statistical significance. The *C. zeae-maydis Ecp6* sequence formed a separate group from the *C. zeina* and *M. fijiensis* sequences. The *C. fulvum* sequence was used to root the tree as it was more distantly related to the other *Ecp6* sequences.

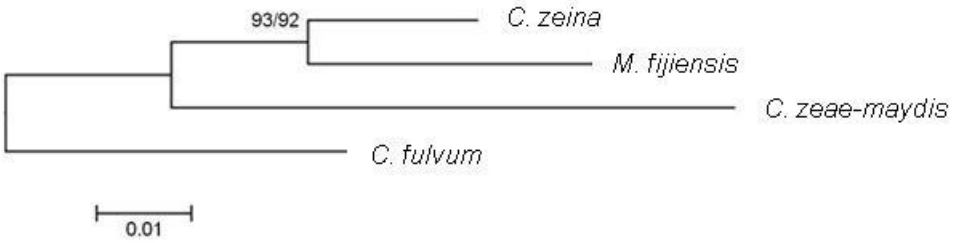


Figure 32. Combined phylogenetic tree for *Ecp6* constructed based on the nucleotide alignment of the different *Ecp6* nucleotide sequences. Bootstrap values are indicated on the nodes as neighbour-joining analysis/maximum likelihood analysis. Bootstrap values were obtained from 1000 replicates and indicated as percentages.

5. DISCUSSION

The results of this study confirmed that the previously identified fungal effectors *Avr4*, *Ecp2*, and *Ecp6* are present in the *C. zeina* (Mkushi) genome. The *C. zeina* (Mkushi) effectors were annotated based on gene-prediction programmes and validated empirically by means of RNAseq reads. Furthermore the amplified gDNA regions of the *C. zeina* (Mkushi) *Avr4*, *Ecp2* and *Ecp6* genes were cloned and sequenced. The protein sequences of the *C. zeina* (Mkushi) Avr4, Ecp2, and Ecp6 were predicted computationally.

The available sequences for the above mentioned effectors included gDNA sequences for *Avr4* from *C. zeina* (GU574327) and *Ecp2* from *C. fulvum* (Z14024.1). An mRNA sequence was available for *C. fulvum Ecp6* (EU730588). The *C. fulvum* Avr4 (CAA55403.1), Ecp2 (CAA78401.1), and Ecp6 (ACF19427.1) protein sequences have been confirmed through mass spectrometry (BOLTON *et al.* 2008; VAN DEN BURG *et al.* 2003).

BLASTn searches were the most sensitive way to search for homologous effector sequences in the draft *C. zeina* (Mkushi) genome as opposed to BLASTp searches due to the degeneracy of the genetic code. Fungal homologs of the annotated *C. zeina* (Mkushi) *Avr4*, *Ecp2*, and *Ecp6* genes in other fungal species of the Dothideomycete class were detected by BLASTn and tBLASTx analysis. BLASTp analysis was not an effective way to detect homologs as the protein sequences of these genes were predicted sequences. Homologs of the annotated *C. zeina* (Mkushi) effectors were shown to be present in the *C. zeina*, *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum* genome sequences. The gene regions and conserved domains of these effectors were predicted in the relevant genome sequences. Some of the homologs identified were already described which provided evidence that the effectors were accurately identified in *C. zeina* (Mkushi). The *Avr4* homolog in *M. fijiensis* (Mycfi2:87167) was MfAvr4 described by Stergiopoulos and colleagues (STERGIOPOULOS *et al.* 2010). The *C. fulvum Avr4* homolog (Clafu1:189855) was CfAvr4 (X78829.1) (JOOSTEN *et al.* 1994). The *M. fijiensis Ecp2* (Mycfi1:52927) identified was MfEcp2-3 (STERGIOPOULOS *et al.* 2010) and the *C. fulvum Ecp2* homolog (Clafu1:197200) was *CfEcp2* (Z14024.1) (VAN DEN ACKERVEKEN *et al.* 1993). Even though none of the *C. zeina* (Mkushi) *Ecp6* homologs determined by this study have been identified previously, the *M. fijiensis* homolog (Mycfi1:34332) had two predicted LysM domains and the *C. zeae-maydis* homolog (Cerzm1:88418) had a predicted lysozyme-like domain.

The *Avr4*, *Ecp2* and *Ecp6* gene regions and structures were confirmed and annotated in the draft genome sequence of *C. zeina* (Mkushi). It was found that a conserved chitin-binding Peritrophin-A (CBM14) domain is present and conserved in the predicted Avr4 protein

sequences of *C. zeina*, *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum*. A conserved putative necrosis inducing (Hce2) domain has been found to be present in the predicted *C. zeina* Ecp2 protein. The predicted Ecp6 protein sequences of *C. zeina* and *M. fijiensis* have three and two conserved Lysin (LysM) carbohydrate-binding domains, respectively.

Avr4 protein alignments of *C. apii*, *C. beticola*, *C. nicotianae*, *C. zeina* (GU574327), *C. zeina* (Mkushi), *C. zeae-maydis*, *M. fijiensis*, and *C. fulvum* shared eight conserved cysteine residues and relatively high sequence similarity within the conserved CBM14 domain region. The protein alignments of *C. zeina*, *M. fijiensis*, *C. zeae-maydis*, and the *C. fulvum* Ecp2 proteins shared four conserved cysteine residues and relatively high sequence identity in the region where the conserved Hce2 domain is found. The protein alignment of the Ecp6 proteins from *C. zeina*, *M. fijiensis*, *C. zeae-maydis*, *M. graminicola*, and *C. fulvum* showed no conserved cysteine residues and shared relatively low sequence similarity. The region of the alignment showing the best conservation could be found at the first LysM domain (116-220 alignment region) and therefore identity and similarity values were calculated for this region only. Phylogenetic analysis based on nucleotide alignments of the Dothideomycete *Avr4*, *Ecp2* and *Ecp6* homologs gave more insight into the evolutionary relationships between the effector sequences. The *C. zeina* (Mkushi) *Avr4* sequence was more closely related to the *C. zeae-maydis Avr4* gene than to the previously identified *C. zeina Avr4* (GU574327) sequence that grouped separately with the other *Cercospora* sequences. The predicted *Ecp2* genes of *C. zeina* and *C. zeae-maydis* were closely related and similarly, *M. fijiensis* and *C. fulvum Ecp2* also shared a close phylogenetic relationship. The *C. zeina* and *C. zeae-maydis Ecp6* sequences were more distantly related as *C. zeina Ecp6* is phylogenetically closer to *M. fijiensis Ecp6*.

Generally, most fungal effectors are believed to be species-specific as not many fungal effectors have been shown to have homologs in other fungal species (STERGIOPOULOS and DE WIT 2009). Results from this study were similar to a recent study done by Stergiopoulos and colleagues that identified homologs of *C. fulvum Avr4* in *C. apii*, *C. beticola*, *C. nicotianae*, *C. zeina*, and *M. fijiensis* (STERGIOPOULOS *et al.* 2010). The current study also identified other *Avr4*, *Ecp2,* and *Ecp6* homologs in other Dothideomycete species, including *C. zeae-maydis*, by using the predicted *C. zeina* (Mkushi) effector sequences. The finding that *Avr4* effector homologs are conserved in different fungal species of the Dothideomycete class (STERGIOPOULOS *et al.* 2010) and that *Ecp2* and *Ecp6* homologs are conserved widely across the fungal kingdom (DE JONGE and THOMMA 2009; STERGIOPOULOS *et al.* 2012), supports the idea that three different classes of fungal effectors exist, namely species-

specific effectors, effectors that occur within a specific fungal class, and effectors that are conserved across fungal classes (STERGIOPOULOS *et al.* 2010).

Conserved functional domains within these effector proteins in different fungal species might indicate the importance of these domains in basic fungal virulence. It was proposed that conserved effectors between different fungal species support common host ancestry (DE WIT *et al.* 2012). The fact that *Ecp2* and *Ecp6* effector homologs are conserved in many species of the fungal kingdom may indicate a common strategy of fungal pathogens to prevent a host immune response (DE JONGE *et al.* 2010; STERGIOPOULOS *et al.* 2012). Conserved domains give an indication of the possible role of a specific effector protein during infection. *C. fulvum* and *M. fijiensis Avr4* homologs both contain the chitin-binding Peritrophin-A domain (CBM14) and has been shown to be important defensive virulence factors that bind to chitin present in fungal cell walls (STERGIOPOULOS *et al.* 2010; VAN DEN BURG *et al.* 2006). The *C. zeina* (Mkushi) and *C. zeae-maydis* Avr4 protein homologs could therefore be possible functional orthologs of *M. fijiensis* and *C. fulvum* Avr4 proteins and also bind to chitin present in fungal cell walls, protecting it against hydrolysis by host chitinases (STERGIOPOULOS *et al.* 2010).

Although the exact function of the *C. fulvum* Ecp2 protein has not been elucidated, it was proposed that this protein might be a necrosis inducing factor as an *M. fijiensis Ecp2* homolog was shown to cause varying degrees of necrosis in the presence and absence of the cognate *Cf-Ecp2* resistance gene (STERGIOPOULOS *et al.* 2010). The putative necrosis inducing (Hce2) domain present in *C. fulvum* Ecp2 is also conserved in the predicted and validated *C. zeina* (Mkushi) Ecp2 protein. Therefore it can be proposed that the *C. zeina* (Mkushi) Ecp2 protein might be a functional ortholog of *C. fulvum* Ecp2 protein causing necrosis that favours the necrotrophic growth stage of the hemibiotrophic fungus. Because no existing functional domains could be detected for the predicted *M. fijiensis* and *C. zeae-maydis* Ecp2 proteins, no functional predictions could be made. The conserved intron-exon boundaries present in all the *Ecp2* homologs identified in this study are similar to what has been found by Stergiopoulos and colleagues (STERGIOPOULOS *et al.* 2010). The conserved *Ecp2* gene structure is another possible indication of inheritance from a common ancestor.

LysM domains have been shown to be involved in carbohydrate-binding and more specifically chitin-binding (BOLTON *et al.* 2008; SANCHEZ-VALLET *et al.* 2013). Therefore the LysM domains present in the Ecp6 homologs identified in *C. zeina* and *M. fijiensis* might indicate that these homologs also function in the sequestering of chitin oligomers to prevent a host defence response in a similar way that has been shown for *C. fulvum* Ecp6 (DE JONGE

and THOMMA 2009; SANCHEZ-VALLET *et al.* 2013). Two of the three LysM domains present in *C. fulvum* Ecp6 are responsible for the formation of a composite binding site or grove that binds chitin oligomers with high affinity (SANCHEZ-VALLET *et al.* 2013). The remaining LysM domain can bind to chitin with low affinity, yet it is still able to prevent detection by the host. Therefore, multiple LysM domains do not necessarily mean a higher chitin-binding affinity. The best characterized LysM protein is the AcmA autolysin from *Lactococcus lactis* which needs three LysM domains for proper enzyme functioning (STEEN *et al.* 2005). More research is needed to determine the exact function of multiple LysM domains present in proteins.

Effector proteins rarely share sequence characteristics and often have no closely related sequences in current protein databases (ELLIS *et al.* 2009). As fungal effector proteins are small cysteine-rich secreted molecules, effector proteins have generally been identified by the presence of an N-terminal secretion protein (SP), depending on the secretion system, and at least two pairs of cysteine residues (HOUTERMAN *et al.* 2007). The presence of a SP generally indicates that the protein is secreted extracellularly by means of the classical secretion system (MANJITHAYA and SUBRAMANI 2011).

Disulphide bonds can be formed between two cysteine residues and are very common in secreted proteins (PEBERDY 1994). Disulphide bonds help to stabilize the structure and activity of the protein after secretion in the high enzymatic environment (apoplast) of the host (BATEMAN and BYCROFT 2000; JOOSTEN and DE WIT 1999; JOOSTEN *et al.* 1997; THOMMA *et al.* 2005). Almost all the predicted Avr4, Ecp2 and Ecp6 effector proteins in this study showed the presence of an N-terminal peptide sequence as well as cysteine residues. The eight conserved cysteine residues in the Avr4 protein alignment are consistent with the finding from Stergiopoulos *et al.* (2010) and indicate the formation of four possible disulphide bonds which might be crucial for Avr4 protein stability. The predicted *M. fijiensis* Avr4 protein has ten cysteine residues, which are shared with MfAvr4 from the Stergiopoulos paper. Interestingly, the *C. zeina* (Mkushi) and the *C. zeae-maydis* Avr4 proteins shared nine cysteine residues which is similar than the amount of cysteine residues present in the *C. zeina* (GU574327) Avr4 protein, but one of the cysteine residues occurs in a different region. An uneven number of cysteine residues might indicate the loss of a cysteine residue. Loss of a cysteine residue, which might be involved in a disulphide bond, might influence the stability of the protein in the intercellular spaces of the host. The fungus might exploit this mechanism to form isoforms of its effector proteins to avoid pathogen recognition. A similar mechanism has been proposed by (JOOSTEN *et al.* 1997). An uneven number of cysteine residues were also found in some of the predicted Ecp2 proteins. All four Ecp2 proteins in the alignment

showed four conserved cysteine residues, which is consistent with previous findings (STERGIOPOULOS *et al.* 2012), that are possibly crucial for protein structure. The *C. zeina*, *C. zeae-maydis*, and *M. fijiensis* Ecp2 proteins in the alignment shared five cysteine residues of which only four were conserved. Even though no conserved cysteine residues were present in the Ecp6 protein alignment, each protein still contained at least two cysteine residues, except for the predicted *M. fijiensis* Ecp6 protein that only contained a single cysteine residue. The *M. fijiensis* Ecp6 protein might therefore be unstable. Ecp6 protein homologs have been identified in a wide range of fungal species (BOLTON *et al.* 2008; DE JONGE and THOMMA 2009) which might explain why the Ecp6 protein sequences share low sequence similarity. This is in contrast with the finding that *Avr* genes have considerably more polymorphisms when compared to *Ecp* genes (STERGIOPOULOS *et al.* 2007).

Phylogenetic analysis of nucleotide sequences of the *Avr4* genes indicated that the predicted *C. zeina* (Mkushi) sequence is more closely related to the *C. zeae-maydis Avr4* gene than the *C. zeina* (GU574327) *Avr4*. This finding also supports the findings from the Avr4 protein alignment, where *C. zeina* (Mkushi) and *C. zeae-maydis* share nine cysteine residues and is similar in protein length, whereas the *C. zeina* (GU574327) Avr4 protein shares its nine cysteine residues as well as protein length with the other *Cercospora* species. A possible reason why the *C. zeina* (GU574327) *Avr4* and *C. zeina* (Mkushi) *Avr4* sequences did not share a closer phylogenetic relationship could be due to the geographic distance found between the two *C. zeina* strains. The *C. zeina* (GU574327) *Avr4* sequence was identified from a strain (CBS118820) originating from KwaZulu-Natal (South Africa), whereas the *C. zeina* (Mkushi) strain (CMW25467) is from Zambia (Africa) (MEISEL *et al.* 2009). These strains might have evolved independently from each other causing the changes in sequence. A similar effect has been observed for the *Avr4* effector when allelic variation between a global population of *M. fijiensis* was compared (STERGIOPOULOS *et al.* 2014).

It was expected that the predicted *Ecp2* genes of *C. zeina* and *C. zeae-maydis* should be more closely related and that the *M. fijiensis* and *C. fulvum Ecp2* genes would group together, as *C. zeina* and *C. zeae-maydis* are sibling species and *M. fijiensis* and *C. fulvum* are more closely related to each other than to the *Cercospora* species. The evolutionary relationships were confirmed by the Ecp2 protein alignment as well as the *Ecp2* phylogenetics. Even though *C. zeina* and *C. zeae-maydis* are sibling species, the *C. zeina* and *C. zeae-maydis Ecp6* sequences did not group together in the Ecp6 phylogenetic tree. The *C. zeina Ecp6* sequence was phylogenetically closer to *M. fijiensis Ecp6*. This might be due to the fact that the *C. zeina* and *M. fijiensis* Ecp6 proteins share two LysM functional

domains, whereas the predicted *C. zeae-maydis* has no predicted LysM domains. The phylogenetic trees for the *Avr4*, *Ecp2* and *Ecp6* genes showed similar groupings of the Dothideomycete species (*C. zeae-maydis*, *M. fijiensis*, *C. fulvum*, and *M. graminicola*) as observed in the phylogenetic tree of Dothideomycetes based on the sequences of three protein-coding genes (Translation elongation factor 1 alpha (*TEF1*), and the first and second largest subunits of DNA-directed RNA polymerase II (*RPB1* and *RPB2*) (OHM *et al.* 2012).

The absence of predicted LysM domains in the *C. zeae-maydis* and *M. graminicola* Ecp6 protein homologs could be due to deletions in important LysM domain regions as seen in the Ecp6 protein alignment. Lack of the Hce2 functional domain in the *C. zeae-maydis* and *M. fijiensis* Ecp2 proteins is not due to deletions in the Hce2 domain region. In fact, both *C. zeae-maydis* and *M. fijiensis* Ecp2 protein sequences shared the four conserved cysteine residues as well as high sequence similarity in the predicted Hce2 domain region. More insight is needed into the function and folding of the Ecp2 protein to be able to understand the importance of the Hce2 functional domain.

As fungal effectors function in evading host immune responses or by disrupting these responses, they are constantly evolving to be able to take part in the evolutionary arms race between the pathogen an its host (DE WIT *et al.* 2009). It has been shown that pathogen effector genes undergo mutations and deletions, which are often characterised as nonsynonymous polymorphisms, to be able to fool its host (MA and GUTTMAN 2008; STERGIOPOULOS *et al.* 2007; STUKENBROCK and MCDONALD 2009). As these effectors undergo constant evolution, it is expected that a specific effector shared between genera and even between different strains of a species, will have sequence differences. It might therefore illustrate the challenge faced in this study when identifying effector homologs in different genera.

One of the limitations of the study was that only the predicted *C. zeina Avr4*, *Ecp2* and *Ecp6* sequences were verified through mapping of *in planta* and *in vitro* RNAseq data and through PCR amplification with gene-specific primers. To be able to verify these effectors gene sequences, the full length cDNA product of each gene would have to be cloned and sequenced. The predicted *C. zeae-maydis*, *M. fijiensis*, and *M. graminicola* effector sequences could not be verified with transcriptomic data.

6. CONCLUSION

The study provided more insight into the classification of fungal effectors. Fungal effectors are not only species-specific, but can be conserved between fungal species of the same class. Homologous fungal effectors often share high sequence similarity and even functional domains. This study showed that homologs of the *C. fulvum Avr4*, *Ecp2* and *Ecp6* are present in other fungal genomes of the Dothideomycete class, including *C. zeina*, *C. zeae-maydis*, *M. fijiensis*, *M. graminicola*, and *C. fulvum*. The chitin-binding Peritrophin-A (CBM14) domain present in *C. fulvum* Avr4, has been shown to be conserved in the *C. zeina (Mkushi)*, *C. zeae-maydis*, and the *M. fijiensis* Avr4 proteins. The Hce2 domain present in putative necrosis inducing *C. fulvum* Ecp2 protein is also present in the *C. zeina* (Mkushi) Ecp2 protein. Three LysM domains are present in the *C. zeina* (Mkushi) Ecp6 protein. The predicted *M. fijiensis* Ecp6 protein contains two LysM domains, while *C. zeae-maydis* Ecp6 contains a lysozyme-like domain. It can therefore be speculated that the *C. zeina*, *C. zeae-maydis*, and *M. fijiensis* Avr4 proteins bind to chitin present in fungal cell walls and protect it against hydrolysis by host chitinases. It can also be speculated that the *C. zeina* Ecp2 protein might have a necrosis inducing function that might favour the necrotrophic growth stage of the fungus. The LysM domains present in the *C. zeina* and *M. fijiensis* Ecp6 proteins might also play a role in the detection of chitin oligomers to avoid a host immune response. More research will need to be done to determine if the predicted Avr4, Ecp2, and Ecp6 proteins from *C. zeina*, *C. zeae-maydis*, *M. fijiensis*, and *M. graminicola* are functional orthologs of the *C. fulvum* Avr4, Ecp2 and Ecp6 proteins by determining if the conserved domains are functional. Confirmation of the effector protein functions would provide more insight into the molecular interactions between *C. zeina* and its host during infection.

7. REFERENCES

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. Journal of Molecular Biology **215:** 403-410.

BATEMAN, A., and M. BYCROFT, 2000 The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). Journal of Molecular Biology **299:** 1113-1119.

BECKMAN, P. M., and G. A. PAYNE, 1982 External growth, penetration and development of *Cercospora zeae-maydis* in corn leaves. Phytopathology **72:** 810-815.

BIRKELAND, N. K., 1994 Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage phi LC3: a dual lysis system of modular design. Canadian Journal of Microbioly **40:** 658-665.

BOLTON, M. D., H. P. VAN ESSE, J. H. VOSSEN, R. DE JONGE, I. STERGIOPOULOS *et al.*, 2008 The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. Molecular Microbiology **69:** 119-136.

BUIST, G., A. STEEN, J. KOK and O. P. KUIPERS, 2008 LysM, a widely distributed protein motif for binding to (peptido)glycans. Molecular Microbiology **68:** 838-847.

CASU, R., C. EISEMANN, R. PEARSON, G. RIDING, I. EAST *et al.*, 1997 Antibody-mediated inhibition of the growth of larvae from an insect causing cutaneous myiasis in a mammalian host. Proceedings of the National Academy of Sciences USA **94:** 8939-8944.

DAVIS, R. H., 2000 *Neurospora: Contributions of a Model Organism*. (Oxford, Great Britain: Oxford University Press).

DE JONGE, R., and B. P. THOMMA, 2009 Fungal LysM effectors: extinguishers of host immunity? Trends in Microbiology **17:** 151-157.

DE JONGE, R., H. P. VAN ESSE, A. KOMBRINK, T. SHINYA, Y. DESAKI *et al.*, 2010 Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science **329:** 953-955.

DE WIT, P. J., R. MEHRABI, H. A. VAN DEN BURG and I. STERGIOPOULOS, 2009 Fungal effector proteins: past, present and future. Molecular Plant Pathology **10:** 735-747.

DE WIT, P. J., A. VAN DER BURGT, B. OKMEN, I. STERGIOPOULOS, K. A. ABD-ELSALAM *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 Genetics **8:** e1003088.

ELLIS, J. G., M. RAFIQI, P. GAN, A. CHAKRABARTI and P. N. DODDS, 2009 Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. Current Opinion in Plant Biology **12:** 399-405.

ELVIN, C. M., T. VUOCOLO, R. D. PEARSON, I. J. EAST, G. A. RIDING *et al.*, 1996 Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*. cDNA and deduced amino acid sequences. Journal of Biology and Chemistry **271:** 8925-8935.

GROENEWALD, M., J. Z. GROENEWALD, T. C. HARRINGTON, E. C. A. ABELN and P. W. CROUS, 2006 Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. Fungal Genetics and Biology **43:** 813-825.

HOUTERMAN, P. M., D. SPEIJER, H. L. DEKKER, D. E. K. CG, B. J. CORNELISSEN *et al.*, 2007 The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. Molecular Plant Pathology **8:** 215-221.

JOOSTEN, M., and P. DE WIT, 1999 The tomato-*Cladosporium fulvum* interaction: A Versatile Experimental System to Study Plant-Pathogen Interactions. Annual Reviews in Phytopathology **37:** 335-367.

JOOSTEN, M. H., T. J. COZIJNSEN and P. J. DE WIT, 1994 Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. Nature **367:** 384-386.

JOOSTEN, M. H., R. VOGELSANG, T. J. COZIJNSEN, M. C. VERBERNE and P. J. DE WIT, 1997 The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4-mediated resistance by producing unstable AVR4 elicitors. Plant Cell **9:** 367-379.

JORIS, B., S. ENGLEBERT, C. P. CHU, R. KARIYAMA, L. DANEO-MOORE *et al.*, 1992 Modular design of the *Enterococcus hirae* muramidase-2 and *Streptococcus faecalis* autolysin. FEMS Microbiology Letters **70:** 257-264.

LAUGE, R., M. H. JOOSTEN, J. P. HAANSTRA, P. H. GOODWIN, P. LINDHOUT *et al.*, 1998 Successful search for a resistance gene in tomato targeted against a virulence factor of a fungal pathogen. Proceedings of the National Academy of Sciences USA **95:** 9014-9018.

MA, W., and D. S. GUTTMAN, 2008 Evolution of prokaryotic and eukaryotic virulence effectors. Current Opinion in Plant Biology **11:** 412-419.

MANJITHAYA, R., and S. SUBRAMANI, 2011 Autophagy: a broad role in unconventional protein secretion? Trends in Cell Biology **21:** 67-73.

MARCHLER-BAUER, A., C. ZHENG, F. CHITSAZ, M. K. DERBYSHIRE, L. Y. GEER *et al.*, 2013 CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Research **41:** 348-352.

MARSHALL, R., A. KOMBRINK, J. MOTTERAM, E. LOZA-REYES, J. LUCAS *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 Physiology **156:** 756-769.

MEISEL, B., J. KORSMAN, F. J. KLOPPERS and D. K. BERGER, 2009 *Cercospora zeina* is the causal agent of grey leaf spot disease of maize in southern Africa. European Journal of Plant Pathology **124:** 577-583.

OHM, R. A., N. FEAU, B. HENRISSAT, C. L. SCHOCH, B. A. HORWITZ *et al.*, 2012 Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathogens **8:** e1003037.

PEBERDY, J. F., 1994 Protein secretion in filamentous fungi--trying to understand a highly productive black box. Trends in Biotechnology **12:** 50-57.

PONTING, C. P., L. ARAVIND, J. SCHULTZ, P. BORK and E. V. KOONIN, 1999 Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. Journal of Molecular Biology **289:** 729-745.

SANCHEZ-VALLET, A., R. SALEEM-BATCHA, A. KOMBRINK, G. HANSEN, D. J. VALKENBURG *et al.*, 2013 Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. Elife **2:** e00790.

SHEN, Z., and M. JACOBS-LORENA, 1998 A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. Cloning, expression, and characterization. Journal of Biology and Chemistry **273:** 17665-17670.

SPAINK, H. P., 2004 Specific recognition of bacteria by plant LysM domain receptor kinases. Trends in Microbiology **12:** 201-204.

STEEN, A., G. BUIST, G. J. HORSBURGH, G. VENEMA, O. P. KUIPERS *et al.*, 2005 AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. FEBS Letters **272:** 2854-2868.

STERGIOPOULOS, I., V. CORDOVEZ, B. OKMEN, H. G. BEENEN, G. H. KEMA *et al.*, 2014 Positive selection and intragenic recombination contribute to high allelic diversity in effector genes of *Mycosphaerella fijiensis*, causal agent of the black leaf streak disease of banana. Molecular Plant Pathology **15:** 447-460.

STERGIOPOULOS, I., M. J. DE KOCK, P. LINDHOUT and P. J. DE WIT, 2007 Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. Molecular Plant Microbe Interactions **20:** 1271-1283.

STERGIOPOULOS, I., and P. J. DE WIT, 2009 Fungal effector proteins. Annu Rev Phytopathol **47:** 233-263.

STERGIOPOULOS, I., Y. A. KOURMPETIS, J. C. SLOT, F. T. BAKKER, P. J. DE WIT *et al.*, 2012 *In silico* characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Molecular Biology and Evolution **29:** 3371-3384.

STERGIOPOULOS, I., H. A. VAN DEN BURG, B. OKMEN, H. G. BEENEN, S. VAN LIERE *et al.*, 2010 Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi

pathogenic on dicots and monocots. Proceedings of the National Academy of Sciences of the United States of America **107:** 7610-7615.

STUKENBROCK, E. H., and B. A. MCDONALD, 2009 Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. Molecular Plant Microbe Interactions **22:** 371-380.

TAMURA, K., D. PETERSON, N. PETERSON, G. STECHER, M. NEI *et al.*, 2011 MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution **28:** 2731-2739.

THOMMA, B. P., V. A. N. E. HP, P. W. CROUS and D. E. W. PJ, 2005 *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic *Mycosphaerellaceae*. Molecular Plant Pathology **6:** 379-393.

TRAPNELL, C., L. PACHTER and S. L. SALZBERG, 2009 TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25:** 1105-1111.

TURNER, M. S., L. M. HAFNER, T. WALSH and P. M. GIFFARD, 2004 Identification and characterization of the novel LysM domain-containing surface protein Sep from *Lactobacillus fermentum* BR11 and its use as a peptide fusion partner in *Lactobacillus* and *Lactococcus*. Applied and Environmental Microbiology **70:** 3673-3680.

VAN DEN ACKERVEKEN, G. F., J. A. VAN KAN, M. H. JOOSTEN, J. M. MUISERS, H. M. VERBAKEL *et al.*, 1993 Characterization of two putative pathogenicity genes of the fungal tomato pathogen *Cladosporium fulvum*. Molecular Plant Microbe Interactions **6:** 210-215.

VAN DEN BURG, H. A., S. J. HARRISON, M. H. JOOSTEN, J. VERVOORT and P. J. DE WIT, 2006 *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Molecular Plant Microbe Interactions **19:** 1420-1430.

VAN DEN BURG, H. A., N. WESTERINK, K. J. FRANCOIJS, R. ROTH, E. WOESTENENK *et al.*, 2003 Natural disulfide bond-disrupted mutants of AVR4 of the tomato pathogen *Cladosporium fulvum* are sensitive to proteolysis, circumvent Cf-4-mediated resistance, but retain their chitin binding ability. Journal of Biology and Chemistry **278:** 27340-27346.