Characterisation of the *Cf-Ecp2* gene encoding for recognition of the conserved fungal effector Ecp2 in *Solanum pimpinellifolium* and *Nicotiana paniculata*

Lucy M^{c}Cann

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy
The John Innes Centre and The Sainsbury Laboratory

August 2016

Abstract

Members of the Capnodiales class of fungi have evolved contrasting lifestyles on plants, ranging from biotrophy, through to endotrophy and necrotrophy. Many of these fungi contain homologs of the effector protein called Ecp2. The 'currant' tomato (*Solanum pimpinellifolium*) resistance (*R*) gene *Cf-Ecp2* recognises Ecp2 produced by the pathogenic fungus, *Cladosporium fulvum*. In this study, *Cf-Ecp2* was fine-mapped to the *Orion* locus, which contains eight *Homologs of Cladosporium resistance gene Cf-9* (*Hcr9s*), four of which share 100% sequence identity within their open reading frame. The *Cf-Ecp2* locus exemplifies the complex nature of many *R* gene loci brought about by successive rounds of tandem duplication. This extensive duplication renders *R* gene loci complex and difficult to resolve. As a consequence, the sequencing of a BAC minimal tiling path across the *Cf-Ecp2* locus required use of both short read and long read sequencing technologies, with MinION providing vital scaffolding reads. A transposon mutagenesis experiment generated two deletion mutants. The mutants had lost the ability to recognise Ecp2 along with the *OR2A* (*2A*) gene from the *Cf-Ecp2* locus. Wild-type tomato (*Solanum lycopersicum)* Cf0 stable transformants, overexpressing *2A* via the cauliflower mosaic virus 35S promoter, partially recapitulated the *S. pimpinellifolium* CfEcp2 phenotype. Lack of penetrance of the phenotype in the transgenic plants was attributed to the functional interference of *2A* overexpression. The non-host, *Nicotiana paniculata* TW99, also recognises Ecp2 from *C. fulvum*. *N. paniculata* CfEcp2 was characterised for its ability to recognise many homologs of Ecp2, including those from *Mycosphaerella fijiensis* (causal agent of Black Sigatoka on banana) and the wilt-causing fungi, *Fusarium oxysporum* and *Verticillium dahliae*. The ability of *Cf-Ecp2* to code for the recognition of an effector from many plant pathogens provides an exciting opportunity to engineer resistance to such pathogens in important crops.

List of contents

List of Tables

Chapter 4

Chapter5

List of Figures

Chapter 4

Chapter 5

List of Abbreviations

Abbreviation	Definition
35S	Cauliflower Mosaic Virus 35S promoter
Ab	Antibiotic selection
AF	Apoplastic fluid
AU	*Aurora*
Avr	Avirulence
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide Basic Local Alignment Search Tool
BMGY	Buffered Glycerol complex medium
BMMY	Buffered Methanol complex medium
bp	Base pair
BWA	Burrow-Wheeler Aligner mapping
CaMV	Cauliflower Mosaic virus
CAPS	Cleaved Amplified Polymorphic Sequences
Carb	Carbenicillin
CC-NB-LRR	Coiled-coiled domain-nucleotide binding-leucine rich repeat
CCS	Circular consensus sequencing
CDPKs	Calcium dependent protein kinase
Cf genes	Genes conferring resistance to *Cladosporium fulvum*
Cf0	*Solanum lycopersicum* Cf0
CfEcp2	*Solanum pinpinellifolium* CfEcp2
CFU	Colony forming unit
cM	CentiMorgan
Col-0	Columbia
COSII	Conserved Orthologous Set II
CRISPR/Cas	Clustered, regularly-interspaced short palindromic repeats/CRISPR-associated
DAMPs	Danger associated molecular patterns
DMSO	Dimethyl sulfoxide
d.p.i.	days post inoculation
d.p.s.	days post sowing

Abbreviation	Definition
Ds	*Dissociation element*
Ecp	Extracellular protein
EDTA	Ethylenediaminetetraacetic acid
EM	Expectation maximum
EMS	Ethyl methanesulfonate
ETI	Effector-triggered immunity
ETS	Effector-triggered susceptibility
EWT	Experiment-wide threshold
F_1	First filial generation
F_2	Second filial generation
F_3	Third filial generation
GM	Genetic modification
GUS	beta-glucuronidase
Hce2	*Homologs of Cladosporium fulvum Ecp2*
Hcr2	*Homologs of Cladosporium resistance gene Cf-2*
Hcr9	*Homologs of Cladosporium resistance gene Cf-9*
HGAP	The Hierarchical Genome Assembly Process
HR	Generic hypersensitive response referenced in wide literature. Includes HR^0 and HR^+
HR^0	Partial hypersensitive response with patchy necrosis
HR^+	Hypersensitive response with confluent cell death
HRec	Homologous recombination
IPTG	Isopropyl-beta-D-thiogalactopyranoside
JIC	John Innes Centre
Kan	Kanamycin
kb	Kilo base pair
LB	Luria and Bertani
Ler	Landsberg erecta
LOD	Logarithm of the odds
LRK	Lectin like receptor kinase
LRR	Leucine rich repeat
MAPKs	Mitogen activated protein kinase

Abbreviation	Definition
Mb	Mega base pair
mRNA	Messenger RNA
MS	Murashige and Skoog
MW	*Milky Way*
NaS	Nanopore synthetic long
NB-LRR	Nucleotide binding-leucine rich repeat
NHEJ	Non-homologous end joining
NL	*Northern Lights*
npt	*Neomycin phosphotransferase*
NR	No response
OR	*Orion*
ORF	Open reading frame
PAMPs	Pathogen associated molecular pattern
PCR	Polymerase chain reaction
PE	Paired end
PFGE	Pulse field gel electrophoresis
PLCP	Papain-like cysteine proteases
PR	Pathogenesis response
PRR	Pattern recognition receptor
PTI	Pattern triggered immunity
PVX	Potato Virus X
QTL	Quantitative trait locus
R gene	*Resistance* gene
RFLP	Restriction fragment length polymorphism
RLK	Receptor-like kinase
RLP	Receptor-like protein
RPM	Rotations per minute
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
S/T PK	Serine/threonine protein kinase
S/TPK–NB-LRR	Serine/threonine-protein kinase–nucleotide binding-Leucine rich repeat

Abbreviation	Definition
sAc	*Stabilised activator-element*
SC	*Southern Cross*
SLP	Seedling lethal phenotype
SMRT	Single molecule real time sequencing
SNP	Single nucleotide polymorphism
SP	Signal peptide
T_1	First transgenic generation
T_2	Second transgenic generation
T_3	Third transgenic generation
T_4	Forth transgenic generation
TALEN	TAL effector nucleases
TIR-NB-LRR	Toll interleukin1 like -nucleotide binding-leucine rich repeat
TGAC	The Genome Analysis Centre
TMV	Tobacco Mosaic Virus
TW99	*Nicotiana paniculata* TW99
TW102	*N. paniculata* TW102
Ws-0	Wassilewskija
WT	Wild type
Xgal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YPD	Yeast Extract Peptone Dextrose medium
ZFN	Zinc finger nucleases
ψ	prefix indicates pseudogene
1178	*S. pimpnellifolium* 1178
1179	*S. pimpinellifolium* 1179
c prefix to *Orion* locus gene	Gene from PacBio assembly C
d prefix to *Orion* locus gene	Gene from PacBio assembly D
2C	*OR2C*
2B	*OR2B*
2A	*OR2A*
2A1	*OR2A1*
c2A1	*cOR2A1*

Abbreviation	Definition
d2A1	*dOR2A1*
2A2	*OR2A2*
cψ2A2	*cψOR2A2*
d2A2	*dOR2A2*
2A3	*OR2A3*
cψ2A3	*cψOR2A3*
d2A3	*dOR2A3*
2A4	*OR2A4*
cψ2A4	*cψOR2A4*
dψ2A4	*dψOR2A4*
c2A5	*cOR2A5*
c2A6	*cOR2A6*
ψ2A5	*ψOR2A5*
ψ2C	*ψOR2C*

Acknowledgements

I would like to thank individuals for their help and support throughout my PhD.

Firstly, I would like to thank my supervisors Brande Wulff, Colwyn Thomas and Mark Banfield for their supervision throughout my PhD. Specifically, I would like to thank Brande Wulff for his consistent help, advice and support throughout my PhD. I would also like to thank Burkhard Steuernagel for general discussions regarding my PhD, specifically bioinformatics and my thesis. Additionally, I thank the rest of the Wulff group and the former 2Blades group, specifically Inma Hernandez-Pinzon, Cintia Kawashima and Mathew Moscou, for advice and discussions.

I would like to thank my father for his help with my PhD thesis and the rest of my family for their constant support throughout my PhD.

I would like to thank individuals for their specific contributions to the PhD project including GopalJee Jha for the generation of the F_2 population from the cross between *Solanum pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0, for the PVX:*Ecp2* screen on this F_2 population leading to the identification and characterisation of the recombinants 4H09 and 2C10, and for the genotyping of 750 F_2 plants and identification of 172 recombinants leading to the identification of 11E01 and 13D08. I would like to thank Kamil Witek for the MinION sequencing of BAC 7B and Michael Giolai, for the PacBIO and MinION sequencing of BAC 7B. I would like to thank Colwyn Thomas and Michael Iiakovidis for the generation and characterisation of the two deletion mutants *S. pimpinellifolium* CfEcp2 1178 and 1179. I would like to thank Stuart Harder for the generation of the F_2 population from the cross between *Nicotiana paniculata* TW99 and *N. paniculata* TW102. Additionally, I would like to thank Stuart Harder and Marie Wulff-Westergaard for their contributions in phenotyping the *Nicotiana* populations. I would also like to thank Marie for her contribution to the first round of Sequenom markers on the F_2 population and Mathew Moscou for his help with the analysis of the Sequenom data and generation of the genetic linkage map in *N. paniculata*.

Where work has been completed in this PhD thesis by another individual they have been cited in the text.

I would like to thank The John Innes Centre (JIC) and The Sainsbury Laboratory (TSL) for their support during my PhD thesis. Finally, I would particularly like to thank the JIC and TSL media kitchens and the horticultural services.

Chapter 1

An introduction to disease resistance in plants

Plants form interactions with a wide range of microorganisms including bacteria, fungi and oomycetes. Such interactions may be beneficial to the plant, such as the rhizobia-forming root nodules where atmospheric nitrogen is fixed into ammonia, a form of nitrogen accessible to the plant, and mycorrhizal associations between roots and fungi that provide the plant with scarce inorganic nutrients. However, some interactions are detrimental and lead to disease. Although pathogenesis is generally a rare outcome of plant-microbe interactions, those pathogens that can infect certain plant species have the potential to cause devastating annual losses to crops. The overall worldwide annual losses have been estimated at approximately 16% (Oerke, 2006). Plants have evolved to protect themselves from such pathogenic attack. Such plant defences include physical, chemical, molecular and genetic barriers.

1.1 Preformed defences

Physical barriers include the production of a waxy cuticle that protects against penetration by non-specialised fungi and enables bacteria to be washed off in wet conditions. Tomato produces saponin α-tomatine, which generates a chemical defence barrier against a broad spectrum of fungi (Ito et al., 2007). This glycoside lycotetraose is composed of an aglycone and tetrasaccharide. It forms complexes with fungal membrane sterols causing pores and inducing programmed cell death in the fungus (Ito et al., 2007).

To infect the plant and cause disease, the pathogen must overcome such barriers. Some pathogens have evolved special growth mechanisms. For example, the rice blast fungus *Magnaporthe grisea* generates a localised osmotic pressure of up to 80 bar in its appressorium, enabling penetration of the rice leaf surface (Howard and Valent, 1996). Furthermore, the cell walls of some microbes are resistant to α-tomatine because the cell walls do not contain sterols (Arneson and Durbin, 1968). This is the case, for example, for some *Fusarium solani* mutants (Defago and Kern, 1983) and the oomycetes *Phythium spp.* and *Phytopthora spp..* In addition, some fungi produce enzymes that degrade and inactivate α-tomatine, such as the tomatinase in *F. oxysporium* f. sp. *lycopersicum* (Lairini

et al., 1996). This enzyme catalyses the hydrolysis of α-tomatine into tomatine and β-lycotetraose.

1.2 Pattern-triggered immunity

Pathogen-associated molecular patterns (PAMPs) are conserved molecules found in a group of pathogens. They are vital to the life of the pathogen but they can trigger defence responses when recognised by complementary pattern-recognition receptors (PRRs) in the host plant (Boller and Felix, 2009; Medzhitov and Janeway, 1997). PAMPs include oligosaccharides, glycoproteins and non-protein molecules such as liposaccharides (Zipfel and Robatzek, 2010). For example, the 22 amino acid peptide flg22 within the bacterial flagellin protein, is recognised specifically by the *Arabidopsis thaliana* PRR FLAGELLIN SENSING 2 (FLS2), which is a receptor-like kinase (RLK) (Chinchilla et al., 2006; Gomez-Gomez and Boller, 2000). Furthermore, in tomato, the leucine-rich repeat (LRR) containing receptor-like protein (RLP) LeEIX1/2 recognises the fungal ethylene inducing xylanase (EIX). The PRRs CHITIN OLIGOSACCHARIDE ELICITOR-BINDING PROTEIN (CEBiP) and CHITIN ELICITOR RECEPTOR KINASE (CERK1) of rice and *Arabidopsis*, respectively, respond to fungal chitin (Kaku et al., 2006; Miya et al., 2007; Ron and Avni, 2004).

Danger-associated molecular patterns (DAMPs) are also detected by the plant. For example, the plant cell wall oligomers released by fungal cell wall-degrading enzymes are detected by the WALL-ASSOCIATED KINASE 1 (WAK1) (Brutus et al., 2010). The recognition of PAMPs/ DAMPs by the plant causes a signalling cascade, involving calcium dependent protein kinases (CDPKs) and Mitogen activated protein kinases (MAPKs) resulting in pattern-triggered immunity (PTI) (Zipfel and Robatzek, 2010). This involves an ion influx, an oxidative burst, callose-deposition and accumulation of the phytohormones ethylene and salicylic acid.

1.3 Effector-triggered susceptibility

During pathogenesis biotrophic pathogens secrete a series of molecules called effectors (Van't Slot and Knogge, 2002). Many effectors have been shown to be important for

virulence. Some effectors suppress PTI or other host defence-responses by targeting PRRs and other molecules in the defence signalling pathways (Ashfield et al., 2014; Gimenez-Ibanez et al., 2009; Mackey et al., 2003; Mackey et al., 2002; Van't Slot and Knogge, 2002). Other effectors do not play a role in defense, but have been shown to make the plant host a better environment for colonisation and reproduction. For example, the transcription activator-like (TAL) effectors of *Xanthomonas oryzae* (that infects rice), activate sugar homeostasis genes (Bogdanove et al., 2010). However, for many effectors a virulence function has yet to be identified. Most often, effectors are specific to the pathogen that secretes them and the host they target. This results in effector-triggered susceptibility (ETS) and corresponding disease in the plant.

1.4 Effector-triggered Immunity

In the so-called 'arms race', plant proteins, encoded by *Resistance* (*R*) genes, evolved by natural selection for those that enable recognition of specific effectors, encoded by Avirulence (*Avr*) genes (Jones and Dangl, 2006). The effector recognition by the protein encoded for by the *R* gene activates effector-triggered immunity (ETI) in a gene-for-gene model (Jones and Dangl, 2006). ETI confers disease resistance via a reactive oxygen species (ROS) burst, callose deposition, pathogenesis response (PR) gene induction and triggering of the hypersensitive response (HR) (Jones and Dangl, 2006; Tsuda and Katagiri, 2010). HR is a localised cell death that is thought to inhibit growth of the biotrophic pathogen by limiting its source of nutrients (Greenberg and Yao, 2004). Inheritance of resistance encoded by *R* genes is generally in a 3:1 manner of resistance: no resistance.

1.5 Non-hosts

When all isolates of a pathogen are unable to infect all accessions of a plant species, the plant species is considered a non-host of the pathogen (Jones and Dangl, 2006; Schulze-Lefert and Panstruga, 2011). This may be due to lack of effectors in the pathogen that are able to overcome PTI, resulting in failure of pathogen growth. Another possibility is that all the plants within a species contain an *R* gene that is able to recognise one or more effectors carried by all strains of the pathogen. ETI is then triggered, resulting in non-host resistance (Schulze-Lefert and Panstruga, 2011).

1.6 Structural groups of Resistance genes

R genes fall into distinct groups. One group of intracellular proteins have a nucleotide-binding (NB)-LRRs group. This group, commonly referred to as NB-LRRs, contains two classes, those containing an N terminal Toll/interleukin-1 like receptor domain (TIR-NB-LRRs) and those containing a coiled-coiled domain (CC-NB-LRRs). Examples for TIR-NB-LRRs are flax L and M (Anderson et al., 1997; Lawrence et al., 1995); One example for a CC-NB-LRR is I2 from *Solanum lycopersicon* I2 (Simons et al., 1998). The intracellular Pto protein of tomato, conferring resistance to *Pseudomonas syringae* pv. *tomato*, is encoded by the only *R* gene, so far cloned, that simply contains a serine/threonine protein kinase (S/T PK) domain (Martin et al., 1993; Pitblado and McNeill, 1983). Some *R* genes encode for RLPs, e.g. the *Cf* genes of tomato, which confer resistance to *Cladosporium fulvum* (Dixon et al., 1998; Dixon et al., 1996; Jones et al., 1994; Thomas et al., 1997). These glycoproteins contain an extracellular LRR domain, a transmembrane domain and a short intracellular N terminal domain. The RLKs encoding *R* genes, e.g. rice *Xa21* conferring resistance to *Xanthomonas oryzae* pv. *oryzae*, are of similar domain structure to the *Cf* genes, yet they contain an N-terminal S/T PK (Song et al., 1995).

1.7 *R* gene clusters and the cloning of *R* genes

1.7.1 *R* gene clusters

R gene variants are generated in the plant by tandem duplication, followed by sequence diversification (Michelmore and Meyers, 1998; Parniske et al., 1997). This plasticity of *R* gene loci generates novel variation, which allows plant hosts to recognise diverse effectors, either directly or indirectly by binding to effector targets.

This method of evolution has resulted in the presence of *R* gene clusters within the genome (Michelmore and Meyers, 1998). A cluster of *R* genes may contain different *R* genes with the ability to recognise more than one effector, and more than one pathogen (Jones et al., 1994; Michelmore and Meyers, 1998; Narusaka et al., 2009b; Panter et al., 2002; Parniske et al., 1997; Takken et al., 1998; Takken et al., 1999; Thomas et al., 1997). In addition, those *R* genes operating in clusters may work together to recognise an effector (Narusaka et al., 2009a; Narusaka et al., 2009b). *R* genes within clusters can be interspersed with other structural classes of genes, which may or may not be required for

R gene function (Andolfo et al., 2013; Martin et al., 1993; Meyers et al., 2003; Michelmore and Meyers, 1998; Salmeron et al., 1996). In addition, there are groups of *R* gene clusters present in some areas of plant genomes (Parniske et al., 1999; Wulff et al., 2009a). For example, on the short arm of chromosome 1 in the tomato genome five *R* gene clusters are found, each containing one to five *R* genes or *R* gene analogues encoding RLPs (Parniske et al., 1999; Wulff et al., 2009a).

1.7.2 Genetic maps enable positioning of *R* genes within the plant genome

To aid in the identification of the *R* gene encoding the resistance to a particular pathogen, genetic maps are utilised to position the *R* gene within the genome (Browns, 2002). Some problems with such maps arise due to the non-random distribution of heterochromatin, which can cause marker clustering, since there is a lower frequency of recombination in heterochromatin than in euchromatin (Browns, 2002). Nonetheless, genetic maps have been very useful in enabling the isolation of *R* genes from many plant species including tomato (Haanstra et al., 1999). Typically, a susceptible parent is crossed to a resistant parent. Mapping depends on the parents containing sequence polymorphisms in the genome. These sequence polymorphisms are then exploited and converted into markers. A mapping population is generated (the F_2 population from the cross between the parents). These F_2 individuals are phenotyped for recognition of an effector/ pathogen and genotyped using the markers. Associations are identified using statistical analysis between the markers and the resistance/ susceptible phenotype. The aim of genetic mapping is to find flanking markers that are on either side of the target locus. This locus can then be further analysed using DNA sequencing to identify the *R* gene.

1.7.3 Genome complexity reduction using Bacterial Artificial Chromosome libraries

The complexity of a genome can be reduced by splitting it into smaller parts. The genome can be split either by mechanical shearing or by partial digestion with a restriction enzyme. The fragments are then transferred into a vector and stored, for example as bacterial artificial chromosomes (BACs). In the case of BACs, a colony of bacterial clones exists for each fragment. Each bacterial clone in that colony carries an additional small chromosome consisting of the genomic fragment and a specific vector that is designed for maintenance of the fragment within the colony. In the case of a map-based *R* gene

cloning project, such a BAC library would be generated from the genome of the resistant parent. Markers that are closely linked to the fine-mapped *R* gene can be used to probe the BAC library with the intention to identify a set of overlapping BAC clones that span the entire target locus from flanking marker to flanking marker. A set of overlapping BAC clones that completely cover a locus is also called the minimum tiling path. Once a minimal tiling path has been generated across the *R* gene locus the isolated BACs can be sequenced (Kubat, 2007). Generally, a complexity reduction method, such as BAC-based sequencing, improves the chance to reconstruct the correct sequence from short reads in comparison to whole genome shotgun sequencing. Still, there can be difficulties in determining the sequence of the loci due to the repetitive nature of *R* gene loci. Short read sequencing technologies tend to be unable to distinguish between the repeats, resulting in the collapse of many genes into one gene sequence (Ashton et al., 2015; Jain et al., 2015; Snyder et al., 2010). However, new sequencing technologies that can sequence longer reads, such as PacBio and MinION, enable the resolution of complex repetitive sequences (Ashton et al., 2015; Jain et al., 2015; Snyder et al., 2010; Van Buren et al., 2015). Genes identified on the selected clones become candidates for the *R* gene.

1.7.4 Transposon-tagging to aid in the identification of *R* genes

Some *R* genes have been isolated by transposon-tagging (Collins et al., 1999; Jones et al., 1994; Takken et al., 1998; Takken et al., 1999). Here, a transposable element, for example a *Dissociation element* (*Ds*), is transformed into the host genome and an event, which is genetically linked to the locus carrying the *R* gene is identified (Rommens et al., 1992). Within the same genome, an unlinked *stabilised activator* (s*Ac*) element enables the transposase function of the *Ds* element. When such a plant is crossed to another, which carries the corresponding effector, the F_1 progeny carrying the active *R* gene and effector are seedling-lethal and die (Jones et al., 1994). Surviving plants will carry the *R* gene, which has been inactivated by *Ds* element insertion or deletions caused by transposon jumping. This method enabled *Cf-9* to be mapped to a 3 kilo base pair (kb) region of the tomato genome and subsequent cloning of the gene (Jones et al., 1994). Similar protocols were also used to clone the tomato *Cf-4E,* maize *Rp1D* gene and tobacco *N* gene (Collins et al., 1999; Takken et al., 1998; Takken et al., 1999; Whitham et al., 1994).

1.7.5 Screening plant genomes for *R* gene analogues

In addition to cloning individual *R* genes that encode for a known resistance, whole plant genomes have been interrogated for their *R* gene analogue complement (Andolfo et al., 2013; Jupe et al., 2012; McHale et al., 2009; Meyers et al., 2003; Monosi et al., 2004). The number of *NB-LRRs* identified in each genome ranged from 111 in tomato to 473 in Rice (Andolfo et al., 2013; Jupe et al., 2012; McHale et al., 2009; Meyers et al., 2003; Monosi et al., 2004). 176 RLPs and 261 RLKs were also identified within tomato (Andolfo et al., 2013).

1.8 The effector - *R* protein interaction

It has been proposed that the LRR domain of R proteins takes part in protein-protein interactions, either directly with the effector or with another plant-derived protein, which is targeted by the effector (Jones and Dangl, 2006; Kobe and Deisenhofer, 1995). These two modes of interaction lend themselves to two separate models for effector recognition by R proteins. In a direct interaction, the R protein directly binds to the effector and triggers a HR (Dangl and McDowell, 2006; Dodds et al., 2006; Jones and Dangl, 2006). This has been shown for the flax L and M, R proteins and their corresponding effectors from *Melampsora lini*, AvrL567 and AvrM, respectively, which interact in a yeast two-hybrid assay (Dodds et al., 2006). Other R protein products indirectly recognise their corresponding effectors. One model, the guard hypothesis, suggests that the R protein guards another protein, which is the direct target of the effector (Dangl and McDowell, 2006; Jones and Dangl, 2006). For example, the *P. syringae* effector AvrPphB targets and cleaves the protein kinase PBS1 (Shao et al., 2003). This cleavage of PBS1 is required to trigger activation of the *R* gene product RPS5 and consequent resistance signalling (Shao et al., 2003).

1.9 *R* gene companions

In order to recognise the effector and to trigger the downstream signalling cascade, some *R* genes require other plant components.

1.9.1 Effector targets

The presence of an avirulence determinant is not beneficial, in evolutionary terms, for the pathogen. As a consequence, the scientific search for the virulence function of effectors began (White et al., 2000). Regardless of whether an effector directly or indirectly interacts with a given R protein, it is believed that some effectors target plant processes. If the R protein, in accordance with the guard hypothesis, monitors the effector target then this second component would be required for ETI (Ade et al., 2007; Rooney et al., 2005).

The role of Avr2 as a virulence factor is well documented. The fungal pathogens *Botrytis cinerea* and *Verticillium dahliae* show greater virulence on *A. thaliana* and tomato when expressing *Avr2* (Van Esse et al., 2008). Avr2 inhibits plant-secreted, papain-like cysteine proteases (PLCP) including RCR3 (Required for *C. fulvum* resistance 3) and PIP1 (*Phytophthora* Inhibited Protease 1) (Rooney et al., 2005; Shabab et al., 2008; Van Esse et al., 2008). *S. pimpinellifolium* RCR3 is the virulence target of the Avr2 effector and triggers HR in the presence of Cf-2 (the guard). Although *Phytophthora infestans* effectors EPIC1 and EPIC2 also inhibit Rcr3 and PIP1. This inactivation does not trigger *Cf-2*–dependent HR (Song et al., 2009). A functional RCR3 must also be present in tobacco as expression of *S. pimpinellifolium Cf-2* and *Avr2* in tobacco triggers HR (M.S.D., unpublished work cited in (Dixon et al., 2000)).

1.9.2 More than one *R* gene may be required for function

Within a locus conferring recognition of an effector, more than one *R* gene may be required to mediate the recognition (Cesari et al., 2014). For example, the *NB-LRRs RPS4* and *RRS1* sit head to head in the Arabidopsis genome (Narusaka et al., 2009a; Narusaka et al., 2009b). Both NB-LRRs are required to mediate the recognition of the *P. syringae* effector AvrRps4, the *Ralstonia solanacearum* effector PopP2 and a *Colletotrichum higginsianum* effector, which is, as of yet, unidentified. In addition, the requirement of two genetically linked *R* genes for recognition of an effector has been found for many other NB-LRRs (Cesari et al., 2014).

1.9.3 Signalling by receptor-like resistance proteins requires other helper proteins

RLPs are believed to require another component to signal recognition of an effector because the short cytoplasmic domain of RLPs does not appear to contain a signalling domain (Dixon et al., 1996; Jones et al., 1994). Conservation of sequence is observed in the last 9.5 C-terminal LRRs of Cf-9 and Cf-2, which may interact with a common protein, most likely an RLK that transduces the disease-resistance signal (Dixon et al., 1996). In fact, Cf-4 has been shown to interact with the RLK Suppressor Of BIR1-1/Evershed (SOBIR1/EVR) in the presence and absence of Avr4 (Liebrand et al., 2013; Postma et al., 2015). In addition, Cf-4 interacts with the RLK BRI1-associated receptor kinase 1 (BAK1), only in the presence of Avr4 (Postma et al., 2015). BAK1 is also required for the PRR FLS2 mediated recognition of flg22 (Chinchilla et al., 2007). It is likely that the interaction between Cf-4, SOBIR1 and Avr4 enables the interaction between Cf-4 and BAK1 and thus the triggering of the downstream signalling cascade that leads to resistance.

1.10 Movement of *R* genes into different plant species

1.10.1 Downstream signalling of effector-recognition is conserved

The capacity for *R* genes to be transferred from their plant of origin to other species has been investigated extensively, but an open debate continues as to the capacity of *R* gene transfer (Wulff et al., 2011). It is believed that, following effector recognition by the R protein and its companions (i.e. the R protein signalling complex), the downstream signalling cascade is conserved across plant species (Holton et al., 2015; Thomas et al., 2000). It is possible therefore to transfer resistance to Powdery Mildew pathogens from *A. thaliana* to *Nicotiana benthamiana* and *N. tobacum* by the transgenic expression of the *A. thaliana NB-LRR* gene *RPW8* in both of these species (Xiao et al., 2003). This conserved resistance response signalling is also present between dicots and monocots. When fused to the LRR domain of *A. thaliana* EFR, the rice Xa21 kinase domain is able to trigger a response to binary protein recognition of EF-TU in *A. thaliana* (Holton et al., 2015). Interestingly, the transfer of the capacity of plants to recognise an effector, from one species into another, can be performed by different methods (Wulff et al., 2011).

1.10.2 Crossing *R* genes into a new plant species

R genes can be crossed into domesticated plant varieties from wild relatives (Hajjar and Hodgkin, 2007; Kerr and Bailey, 1964; Stevens and Rick, 1988). For example, *Cf* genes identified in wild species of tomato, such as the *Cf-4* gene in *S. hirsutum* and the *Cf-9* gene in *S. pimpinellifolium*, have been bred into domesticated commercial tomato lines (Kerr and Bailey, 1964; Stevens and Rick, 1988). However, this causes linkage drag with unwanted characteristics along with the *R* gene being incorporated into the new crop line. For example, the *Cf-4* and *Cf-9*–mediated recognitions were introgressed along with large regions of wild tomato DNA (Parniske et al., 1997; Parniske et al., 1999). Linkage drag is not always directly correlated to the number of backcross generations (Young and Tanksley, 1989). In fact, although some breeding programmes have low linkage drag after a few generations, others are ineffective at limiting linkage drag, especially when many genes are being introgressed (Young and Tanksley, 1989). Another disadvantage of *R* gene species-transfer by crossing is that it is restricted to sexual compatibility.

1.10.3 Genome-editing to generate resistance to pathogens

As an alternative approach to transformation, genome-editing can be used to modify pre-existing genes within a plant species to become a known *R* gene (Li et al., 2012; Wang et al., 2014). For example, the targeted deletion of the three copies of *MLO* from the bread wheat genome, using TAL effector nucleases (TALENs), caused increased resistance to powdery mildew (Wang et al., 2014). Other methods of genome-editing can involve the use of Zinc finger nucleases (ZFN) or clustered, regularly-interspaced short palindromic repeats/CRISPR-associated (CRISPR/Cas) systems to alter specific regions in the plant genome (Belhaj et al., 2013; Gaj et al., 2013; Weinthal et al., 2010). ZFN, TALENs and CRISPR/Cas systems cause double-stranded breaks at target sites. Each of the systems can be paired. For example, two CRISPRS/Cas can be used to generate cut sites at two regions of the genome e.g. at two sites within a gene or gene cluster (Gaj et al., 2013). When a double-strand break has formed, the native plant DNA repair system takes over (Gaj et al., 2013).

There are two forms of DNA repair in plants; non-homologous end joining (NHEJ) and homologous recombination (HRec) (Gaj et al., 2013). If a foreign donor DNA molecule is

present, the repair processes can insert the donor DNA in between the cut sites (Gaj et al., 2013). In the absence of foreign donor DNA, the repair processes mend the break in the DNA (with NHEJ, causing more errors than HRec) thus causing targeted knock out of a gene (Gaj et al., 2013).

Genome editing keeps the amount of foreign DNA introduced into the plant to a minimum (Kuzma, 2016). This method, however, depends on the level of homology between the pre-existing gene and the gene it is targeted to replace. If the genes are very diverse, a higher level of editing is required, i.e. more foreign DNA would be introduced into the host plant. However, this method could, for instance, merely alter the LRRs of a pre-exiting NB-LRR so that it would be able to recognise the target effector. In this case, the edited NB-LRR would highjack the preformed defence system of the plant to trigger resistance following the recognition of the target effector.

1.10.4 Genetic modification

Genetic modification (GM) includes the transfer of alien DNA into an organism. Using transformation, this method can be applied to directly introduce *R* genes into crops without additional linkage drag (Jones et al., 2014; Wulff et al., 2011). Genetic modification breaks down plant species barriers, enabling *R* genes from distantly-related plants to be integrated into other crop species (Mendes et al., 2010). Any other *R* gene companions required for *R* gene-encoded recognition of said effector, which are not already present in the host, must also be identified and cloned to enable transfer of *R* genes widely among crops (Ashfield et al., 2014; Mackey et al., 2002; Rooney et al., 2005; Salmeron et al., 1996).

1.10.5 The durability of *R* genes

The introduction of single *R* genes mediating resistance to a pathogen is readily overcome by selection for variation in the pathogen population that overcomes resistance (McDonald and Linde, 2002a, b). Mutations within effectors that enable evasion of *R* gene-encoded recognition are beneficial to the pathogen and proliferate throughout the population of the pathogen by natural selection. However, mutations in effectors may cause a virulence penalty. In these cases, variants will only be beneficial to the pathogen in a plant population where the *R* gene is maintained.

The virulence function of Avr4 is to bind to chitin in the fungal cell wall and protect it from chitinases. Virulent races of *C. fulvum* on tomato carrying *Cf-4* have been shown to contain a single base change in Avr4, converting the triplet TGT (encoding for cysteine) to TAT (encoding for tyrosine) (Joosten et al., 1994). The mutant proteins are no longer recognised by Cf-4, yet are still able to bind chitin and, when bound to chitin, are less sensitive to proteases (Joosten et al., 1994; Van den Burg et al., 2003). Thus, they retain their virulence function but lose their avirulence function.

In comparison, *C. fulvum* races, which overcome recognition mediated by Cf-9, have lost the *Avr9* gene from the genome and do not produce Avr9 (Vankan et al., 1991). Furthermore, frame-shift mutations induced by single nucleotide insertions or deletions in *Avr2*, at several independent sites, or point mutations inducing an early stop codon, are present in different strains of *C. fulvum* that are virulent on *S. lycopersicum* Cf-2 (Luderer et al., 2002b). Both forms of mutation generate a truncated protein, lacking at least three different cysteine residues and are predicted to be unstable or non-functional (Luderer et al., 2002b). The conservation versus loss of virulence function, caused by mutation and allelic-propagation throughout the population, may be determined by the importance of the virulence function of the effectors, their genetic redundancy or whether the *R* gene/ Avr gene pair are under balancing selection.

Similarly, the *Phytopthora infestans* effector Avr3a exists in two forms; $Avr3a^{KI}$ and $Avr3a^{EM}$ (Armstrong et al., 2005; Cardenas et al., 2011). $Avr3a^{KI}$ triggers a resistance response in the presence of the *R3a R* gene product, whilst $Avr3a^{EM}$ evades this recognition (Armstrong et al., 2005; Cardenas et al., 2011). This non-durable resistance

and infection is described by the zig-zag-zig evolutionary 'arms race' of pathogen-host interactions (Jones and Dangl, 2006).

Pyramiding is a technique of introducing multiple *R* genes, conferring resistance to a single pathogen (Joshi and Nayak, 2010). This would mean that the pathogen has to overcome multiple *R* gene encoded resistances in order to infect the plant, generating more durable resistance (McDonald and Linde, 2002a). This resistance is at higher risk of breaking down if during breeding, the *R* genes are again separated. They must therefore be introduced at the same location in the target genome or into a non-recombinogenic part of the target genome to reduce the chance separation of the *R* genes during downstream breeding and deployment.

1.11 *Cladosporium fulvum* interaction with tomato

C. fulvum causes leaf spot disease on its only host tomato, genus *Solanum* (previously *Lycopersicum*)(Bond, 1938). During both compatible and incompatible infection, conidia land on the leaf surface and germinate, sending off runner-like hyphae along the leaf surface (Thomma et al., 2005). In a compatible interaction, conidiophores emerge from stomata ~10 days post inoculation (dpi) (Thomma et al., 2005). The hyphal growth is straight and intercellular, filling the apoplast of the lower mesophyll in the leaf and the nearby vascular tissue (Thomma et al., 2005). Haustoria or other feeding structures are not formed but the hyphae are in close contact with the host cells (Thomma et al., 2005). During incompatible interactions, the hyphae are not straight and bulbous hyphae may form (Thomma et al., 2005). Growth may be inhibited during stomatal penetration or soon after within the sub-stomatal cavity (Thomma et al., 2005). As *C. fulvum* is an extracellular pathogen, communication between the pathogen and host plant is extracellular (Thomma et al., 2005). *C. fulvum* secretes effectors into the apoplast and host *R* gene products project into this intercellular space to detect such molecules and induce HR (de Wit et al., 1994). So far, all *R* genes cloned from tomato that confer recognition of *C. fulvum* effectors are RLPs.

Tomato is recognised as a good model organism to study disease. The first *R* gene demonstrating a gene-for-gene type resistance, *Pto,* was isolated from tomato (Martin et al., 1993). Tomato has a short generation time from seed to seed, generates multiple

progeny and has simple growth requirements (Kimura and Sinha, 2008). Furthermore, the genome sequence of *S. lycopersicon* Heinz 1706 has been published (Sato et al., 2012). The tomato genome, composed of 900 Mega base pair (Mb) and 12 chromosomes, contains 34,727 genes (Sato et al., 2012). In addition to the sequence of the tomato genome, several high-density genetic maps have been generated (www.solgenomics.net).

1.11.1 The *Cf* genes

There appears to be two forms of RLP-encoding genes in tomato with the potential to confer resistance to *C. fulvum*. A cluster containing *homologs of Cladosporium resistance gene Cf-2* (*Hcr2*) is present on the short arm of chromosome 6 and contains the *R* genes *Cf-2* and *Cf-5* (Figure 1.1) (Dixon et al., 1998; Dixon et al., 1996). *Homologs of Cladosporium resistance gene Cf-9* (*Hcr9s*) are present in clusters on the short arm of chromosome 1 (Figure 1.1) (Parniske et al., 1999; Thomas et al., 1997; Wulff et al., 2009a). There are five such clusters, *Northern Lights* (*NL*), *Milky Way* (*MW*), *Aurora* (*AU*), *Orion* (*OR*) and *Southern Cross* (*SC*) (Figure 1.1). At each locus, more than one *R* gene can code for the recognition of the same effector, recognition of more than one effector can be encoded for by one cluster of *R* genes and the function of the *R* genes can be dependent on the age of the plant. The *MW* locus encodes for *Cf-9*, *Cf-9DC* (or *Hcr9-9D*), *Hcr9-9B*, *Cf-4* and *Cf-4E,* and other *Hcr9s* (Figure 1.1) (Jones et al., 1994; Kruijt et al., 2004; Panter et al., 2002; Thomas et al., 1997; Van der Hoorn et al., 2001a). *9DC* recognises Avr9, and is a natural chimeric variant of *Cf-9* (Kruijt et al., 2004; Van der Hoorn et al., 2001a). On the other hand, *Hcr9-9B*, with 89% identity to *Cf-9*, recognises Avr9B (yet to be cloned) (Panter et al., 2002). Not all of the *Hcr9* and *Hcr2* genes identified have been shown to carry resistance function and some pseudogenes have been identified within the clusters (Figure 1.1) (Wulff et al., 2009a).

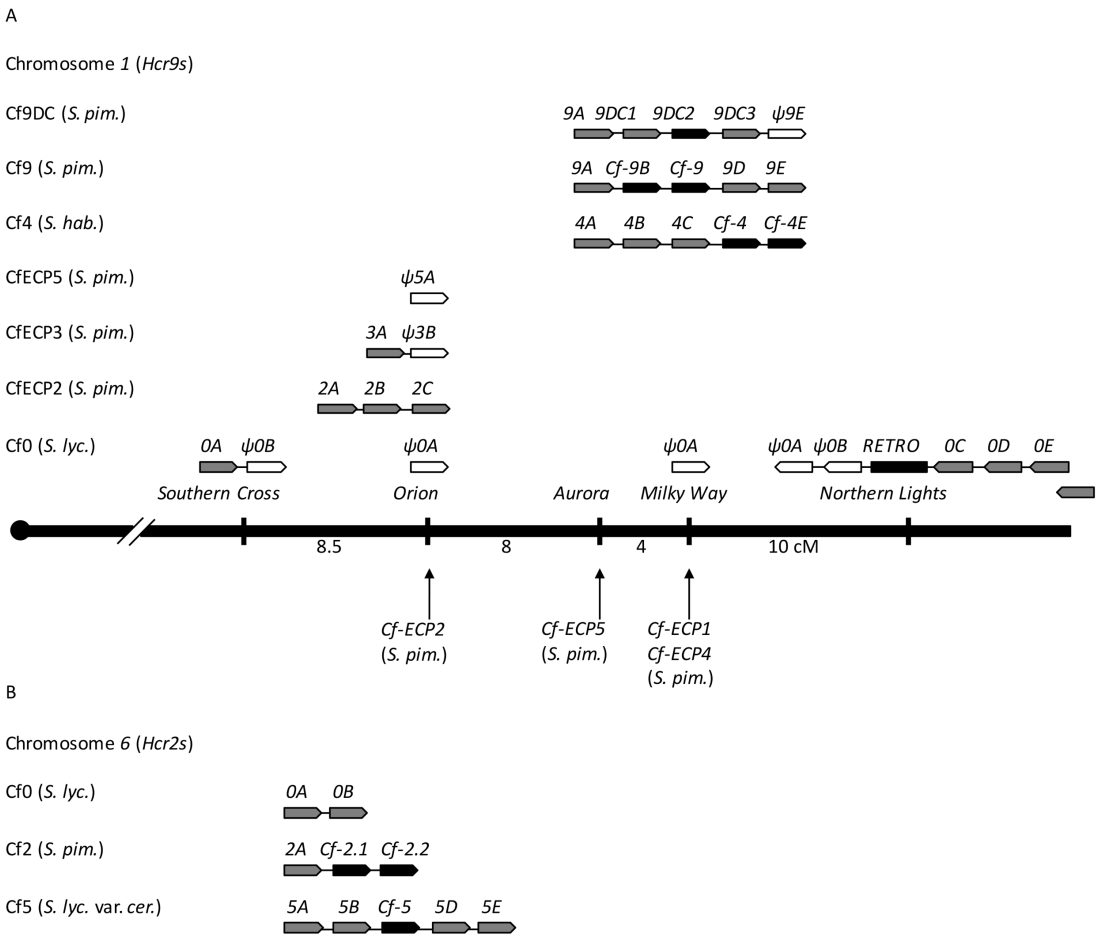


Figure 1.1. Location of identified *Hcr9s* (A) and *Hcr2* (B) genes located on chromosomes 1 and 6, respectively.

In (A) the position of the genes in clusters across chromosome 1 (bold black line) from the centromere (round end of line) are shown with the distance of the clusters from each other in centiMorgan (cM) below the line and the name of each *Resistance* (*R*) gene cluster above the line. Name of *R* genotype and species such a genotype was identified in is in brackets, followed by arrows representing genes in certain chromosomal locations. Names above arrows are those given to the genes. Black arrows are cloned and known to have activity against a certain effector. Grey arrows are cloned yet unknown activity and white arrows are pseudogenes (ψ). *Hcr9s = Homologs of Cladosporium resistance gene Cf-9, Hcr2 = Homologs of Cladosporium resistance gene Cf-2*, *S. pimp = S. pimpinellifolium* (formerly *Lycopersicum pimpinellifolium*), *S. hab = S. habrochaites* (formerly *L. hirsutum*), *S. lyc = S. lycopersicum* (formerly *L. esculentum*) and *S. lyc.* var. *cer. = S*. *lycopersicum* variety *cerasiforme* (formerly *L. esculentum* variety *cerasiforme*) (Peralta et al., 2005). RETRO = retrotransposon. Modified from (Wulff et al., 2009a).

The *Hcr9* and *Hcr2* homologs share a conserved structure consisting of seven domains (Figure 1.2). The signal peptide (SP) ranges from 22 – 34 amino acids (Dixon et al., 1998; Dixon et al., 1996; Jones et al., 1994; Thomas et al., 1997). Domains B-D show homology to the polygalacturonidase inhibitor protein (Dixon et al., 1998; Dixon et al., 1996; Takken et al., 1999). The LRRs form a conserved β-strand β-turn with repeating type A and B

LRRs, both types containing the XXLXLXX consensus sequence (X represents any amino acid) (Dixon et al., 1996). The conserved aliphatic side chains point into the protein core while the residues flanking these point out forming the solvent-exposed surface (Kobe and Deisenhofer, 1993). The LRR length, on average 24 amino acids, is conserved to a greater extent in *Hcr2* proteins compared to *Hcr9s* (Dixon et al., 1998; Dixon et al., 1996). Residues within the LRR domain have been found to be required for recognition of *C. fulvum* effectors (Van der Hoorn et al., 2001b; Wulff et al., 2009b).

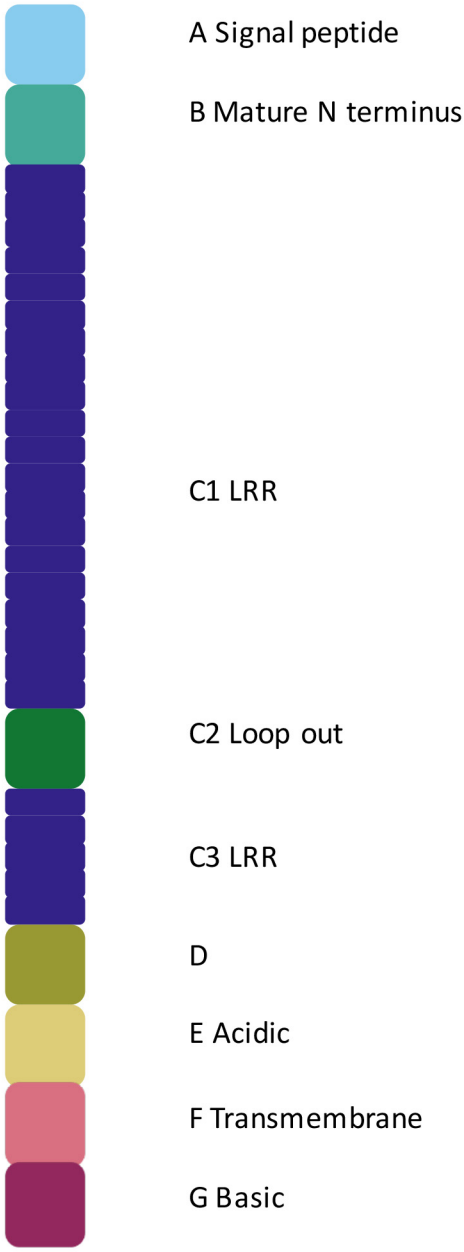


Figure 1.2. The domain structure of Hcr9s and Hcr2s (not to scale).

Cleavage of the N-terminal signal peptide (SP) (A) exposes a new N-terminus of the mature protein (B). This is followed in sequence by a leucine rich repeat (LRR) domain (C1), a loop out (C2) and second LRR domain (C3). Domain D has no apparent function but precedes the membrane-anchor domains E, F, and G containing an acidic (E) and basic group (G) of residues flanking the transmembrane domain (F), anchoring it into the membrane. Hcr9 = Homolog of Cladosporium resistance gene Cf-9, Hcr2 = Homolog of Cladosporium resistance gene Cf-2.

The transmembrane domain contains uncharged amino acids e.g. 37 amino acids in Cf-9, and 24 amino acids in Cf-2 and Cf-5 (Dixon et al., 1998; Dixon et al., 1996; Jones et al.,

1994). This domain is flanked by an extracellular acidic domain (e.g. in Cf-9 consisting of 10 negatively-charged amino acids and no positively-charged amino acids) and an intracellular basic domain (e.g. 21 amino acids in Cf-9, eight of which are positively charged) (Jones et al., 1994).

These domains orientate and anchor the protein to the membrane (Figure 1.2). A conserved intron is also present in the 3' UTR of *Cf-2*, *Cf-9* and *Cf-4* (Dixon et al., 1996; Thomas et al., 1997). N-linked glycosylation sites are common in the Cf proteins and they may have a role in the tertiary structure of the protein e.g. 22 N-linked glycosylation sites domains B – E of Cf-9 at (Dixon et al., 1998; Van der Hoorn et al., 2005).

1.11.2 The *C. fulvum* effectors

The genes encoding *C. fulvum* effectors are, in general, race-specific and only expressed *in planta.* Effectors have been identified from the apoplastic fluids (AF) of tomato infected with *C. fulvum* (de wit and Spikman, 1982; Joosten and de Wit, 1988). The first *C. fulvum* Avr type effector cloned was *Avr9* (3.1 KDa and 28 amino acids mature protein), which is responsible for triggering *Cf-9*–dependent disease resistance (Marmeisse et al., 1993; Vankan et al., 1991). The identification of several other similar Avr genes followed; Avr4 (12 KDa and 106 amino acids), Avr4E (12 KDa and 121 amino acids) and Avr2 (12 KDa and 78 amino acids), which trigger *Cf-4*, *Cf-4E* and *Cf-2* dependent disease resistance, respectively (Joosten et al., 1994; Luderer et al., 2002b; Westerink et al., 2004).

Extracellular proteins (Ecp) are also present in the AF of compatible interactions between *C. fulvum* and tomato (Bolton et al., 2008; Joosten and de Wit, 1988; Wubben et al., 1994). These effectors do not seem to be race-specific. Ecp1 (7/9 KDa and 65 amino acids), Ecp2 (17 KDa and 142 amino acids), Ecp3 (19 KDa), Ecp4 (12 KDa), Ecp5 (7 KDa), Ecp6 (21 KDa and 199 amino acids) and Ecp7 (11 KDa and 100 amino acids) have been isolated (Bolton et al., 2008; Lauge et al., 2000; Lauge et al., 1998; Van den Ackerveken et al., 1993a). Resistances to Ecp1, Ecp2, Ecp3, Ecp4 and Ecp5 have been identified and designated as *Cf-Ecp1*, *Cf-Ecp2*, *Cf-Ecp3*, *Cf-Ecp4* and *Cf-Ecp5*, respectively (Haanstra et al., 1999; Haanstra et al., 2000; Soumpourou et al., 2007; Yuan et al., 2002).

In comparison to the allelic variation of Avr genes in *C. fulvum* populations (see section 1.10.5, 'Durability of *R* genes'), fewer mutations have been identified in *Ecp* genes. For example, Ecp6 variants contain a total of five single nucleotide polymorphism (SNPs), all of which cause no effect on virulence function (Bolton et al., 2008; Stergiopoulos et al., 2007). *Cf* genes, encoding Avr protein recognition, are widely deployed in cultivated tomato, while *Cf* genes encoding Ecp protein recognition are not. This may explain the apparent lack of sequence variation in Ecp genes. An alternative hypothesis is that the virulence function of the *Ecp* genes cannot be lost from the pathogen without causing a severe cost to virulence. The *Cf* genes recognising such effectors would be durable within a population, until the pathogen can evolve a new effector that can prevent the immunity triggered by the Ecp effector.

C. fulvum effectors are small, secreted and cysteine-rich proteins. They contain even numbers of cysteine molecules, which are thought to be involved in disulphide-bonds; eight cysteines in Avr2, Avr4, Ecp1, Ecp6, six cysteines in Avr9, Avr4E, Ecp5 and Ecp7, and four cysteines in Ecp2 (Bolton et al., 2008; Joosten et al., 1994; Luderer et al., 2002a; Luderer et al., 2002b; Van den Burg et al., 2003; Vervoort et al., 1997; Westerink et al., 2004). Such disulphide-bonds have the potential to confer structural support in the protease-rich apoplast (Luderer et al., 2002a; Van den Ackerveken et al., 1993b). The effectors also contain hydrophobic-signal peptides (ranging from 16 to 23 amino acids in length) at their N-terminus, dictating secretion from the fungus (Bolton et al., 2008; Joosten et al., 1994; Luderer et al., 2002b; Van den Ackerveken et al., 1993a; Van den Ackerveken et al., 1993b).

Some of the effectors undergo further cleavage by plant and fungal proteases in the apoplast, while others, such as Avr4E, do not (Westerink et al., 2004). The Avr4 117 amino acid-secreted peptide is further processed into a 106 amino acids mature protein and Ecp1 and Ecp2 lose a further eight and one amino acid in the apoplast, respectively (Joosten et al., 1994; Van den Ackerveken et al., 1993a). Once secreted, the 63 amino acid Avr9 is processed sequentially into 40, 34, 33 and 32 amino acid peptides. The final mature Avr9 peptide has a length of 28 amino acids. The latter four peptides are able to trigger *Cf-9*-dependent HR (Van den Ackerveken et al., 1993b).

Some of the effector genes contain small introns, typical of filamentous fungi. *Avr2* contains a 54 base pair (bp) intron, 159 bp downstream of the transcription start (Luderer et al., 2002b). Ecp1 and Ecp6 are interrupted by two introns and Ecp2 contains one 56 bp intron (Bolton et al., 2008; Van den Ackerveken et al., 1993a). These introns contain conserved splice junctions GT-AT and internal consensus sequences TACTAAC (Van den Ackerveken et al., 1993a). Furthermore, *Ecp1* and *Ecp2* both lack C-terminal polyadenylation sequences (AATAAA) (Van den Ackerveken et al., 1993a). The effectors are of single-copy number in *C. fulvum,* except for *Avr4* which occurs three times in the genome (Joosten et al., 1994).

1.11.3 Stable and transient assays to study *R* and Avr gene function

Agrobacterium-mediated transient expression has become an important tool for studying *R-Avr* gene interactions (Thomas et al., 2000; Van der Hoorn et al., 2000). Non-host tobacco plants exposed to Agrobacterium-mediated transient co-transformation of *Cf-4* or *Cf-9* with their corresponding effectors Avr4 or Avr9, respectively, have been shown to give rise to complete necrosis of the inoculated region (Thomas et al., 2000).

The potato virus X (PVX) genome can accommodate effector genes (Van der Hoorn et al., 2000). PVX inoculation on to plants enables the spread of an effector encoded within the PVX genome, since PVX is self-propagating and can move from cell to cell. Agroinfiltration of *Cf-4* or *Cf-9* can be combined with PVX infiltration of *Avr4* or *Avr9*, generating necrosis on tobacco (Van der Hoorn et al., 2000). A development of this was the insertion of the PVX genome between the left and right border of the Ti plasmid (Vleeshouwers et al., 2006). Furthermore, a strong seedling-lethal phenotype (SLP) can be scored in tomato, when a cloned *Cf* gene is expressed in tomato in a stable manner and crossed to another tomato stably expressing the corresponding *Avr* gene (Ade et al., 2007; Hammond-Kosack et al., 1994a; Soumpourou et al., 2007; Thomas et al., 1997). This enables the confirmation of the *Cf* gene phenotype. For example, *S. lycopersicum* Cf0 expressing *Cf-4* crossed to *S. lycopersicum* Cf0 expressing *Avr4* causes SLP (Hammond-Kosack et al., 1998). The assays described above recapitulate the gene-for-gene interaction in the absence of the *C. fulvum* pathogen, which simplifies the experimental design.

1.11.4 The core effector Ecp2

The Ecp2 effector is of unknown virulence function and has no homology to known sequences (Van den Ackerveken et al., 1993a). It is therefore unknown if the domain for virulence function and for recognition overlap. If so, loss of recognition would be more likely to result in loss of virulence function, which may not be beneficial for the pathogen (Lauge et al., 1998). The importance of the Ecp2 effector is represented in not only its presence in all 25 races of *C. fulvum* collected world-wide, but in many fungal species of the Dothidiomycete class, particularly in the Capnodiales (Lauge et al., 1998; Stergiopoulos et al., 2012; Stergiopoulos et al., 2010). The conservation of Ecp2 in *C. fulvum* may also be due to the lack of *Cf-Ecp2* deployment in tomato cultivars. Thus no selection pressure is on the fungus to lose this effector.

The *C. fulvum* Ecp2 protein is recognised when infiltrated into a plant containing *Cf-Ecp2* (Lauge et al., 1998). The level of Ecp2, which accumulates during infection, correlates with the aggressiveness of the *C. fulvum* strain (Lauge et al., 1998). Using immuno localisation, it was shown that Ecp2 localises mainly between the plant and fungal cell walls in relatively few hyphae (Wubben et al., 1994). Its expression, followed by an *Ecp2*-promoter beta-glucuronidase (GUS) fusion, was strongest during intercellular colonisation and highest in the hyphae surrounding the vascular tissue (Wubben et al., 1994). However, two strains of *C. fulvum*, in which *Ecp2* was either partially disrupted by the insertion of a hygromycin B resistance gene or completely knocked out by the insertion of a phleomycin-resistance gene, showed no difference in growth morphology compared to the wild type (WT) race 5 strain when infected on 14 day old Cf0 tomato (Marmeisse et al., 1994). Furthermore, a strain, which was over-expressing Ecp2 by more than 20-fold, showed identical infection compared to WT race 5 (Marmeisse et al., 1994).

Interestingly, in a separate study, an Ecp2 knock out strain inoculated on to 6-week-old Cf0 tomato showed poor colonisation of leaf tissue, the secretion of lower amounts of effectors (Ecp3, Avr4, Avr9) and generated few conidia compared to WT (Lauge et al., 1997). In addition, this Ecp2-deficient strain induced quicker PR protein accumulation and caused partial collapse of the guard cells in the host in comparison to WT (Lauge et al., 1997). Co-inoculation of mutant strains with WT enabled partial recovery of virulence (Lauge et al., 1997). These deficient strains showed no difference in growth morphology

or sporulation rate *in vitro*, attributing changes in virulence to be a direct cause of the effector mutation (Lauge et al., 1997). It should be mentioned that both of these studies were carried out with only one mutant strain (Lauge et al., 1997; Marmeisse et al., 1994).

1.11.5 Recognition of Ecp2 by *Cf-Ecp2*

The recognition of Ecp2 identified in *S. pimpinellifolium* demonstrates monogenic and dominant inheritance in genetic crosses (Haanstra et al., 1999; Lauge et al., 1998). Lines Cf18, Cf20, Cf23 and Cf24 all contain an *R* gene with Ecp2-recognition (Haanstra et al., 1999; Lauge et al., 1998). The lines were collectively renamed CfEcp2 (Haanstra et al., 1999; Lauge et al., 1998). Recognition was mapped to a new cluster of *Hcr9* genes, known as the *OR* cluster (Figure 1.1) (Haanstra et al., 1999). Furthermore, resistance co-segregates with *Hcr9* markers (de Kock et al., 2005). Thus, *Cf-Ecp2* is predicted to contain the *Hcr9* structure and to encode for an RLP. *Cf-Ecp2* maps 11.3 centiMogran (cM) +/- 1.8 cM distal to the *Cf-4* gene, flanked by the markers TG236 (7.7 cM distal to *Cf-Ecp2*) and TG184 (6.0 cM proximal to *Cf-Ecp2*) on the short arm of chromosome 1 (Haanstra et al., 1999). The *OR* cluster contains three characterised *Hcr9s* named *OR2A* (*2A*), *OR2B* (*2B*) and *OR2C* (*2C*) (de Kock et al., 2005).

The *OR* haplotypes from *S. pimpinellifolium* CfEcp3 and CfEcp5 and *S. lycopersicum* Cf0 also contain *Hcr9* genes (de Kock et al., 2005). However, most of these are pseudogenes (ψ); e.g. *ψOA* present in Cf0 shares 99.9% homology with *2C* but a 10 bp deletion causes a frame shift and early stop codon (de Kock et al., 2005). A large duplication present within the *Cf-Ecp2 OR* cluster of 2.6 kb, which is thought to result from a recent intragenic unequal crossing over, includes the promoter and initial 432 bp of the open reading frame (ORF) of *2A* and *2B* (de Kock et al., 2005). Of the duplicated region, the first 576 bp are shared with the potato cyst nematode *NB-LRR R* gene *hero*, which is located on chromosome 4 (de Kock et al., 2005; Ernst et al., 2002). This conserved element in the promoter region may indicate a common transcription factor acting to control expression of the respective genes (de Kock et al., 2005).

Analysis of the *2A*, *2B* and *2C* DNA sequences showed the presence of 27 LRRs in the extracellular domain C (Figure 1.2), with most variation in the solvent-exposed residues of the first 17 LRRs of the C and B domains (de Kock et al., 2005). Such variation is

similarly present in other *Hcr9s* and *Hcr2s* (de Kock et al., 2005). However, sequence comparison to *Cf-9* indicated only 50% homology (de Kock et al., 2005). In addition, variation in the 2A, 2B and 2C predicted proteins reaches to the C terminal LRRs 18 – 24, the loop out domain and the acidic E domain, usually conserved in *Hcr9s* and *Hcr2s* (de Kock et al., 2005; Dixon et al., 1996). Stable transformation of these three *OR* genes in combination or singularly into *S. lycopersicum* Cf0 with a 1 kb native 5' regulatory sequence, did not complement Ecp2-triggered HR. As a consequence, neither of these were designated the *Cf-Ecp2* gene (de Kock and colleagues, 2004).

1.12 *Nicotiana paniculata* CfEcp2

Ecp2 is also recognised by specific accessions in the non-hosts *N. paniculata*, *N. sylvestris, N. tobacum,* and *N. undulata* (de Kock et al., 2004; Lauge et al., 2000). This recognition was shown to be mediated by a single dominant gene, which, based on polymerase chain reaction (PCR) and hybridisation analysis, was concluded not to be homologous to *Hcr9* genes (de Kock et al., 2004). However, the basis of these findings relies on DNA homology, and DNA sequence is not always directly linked to protein structure. A protein similar in structure to those encoded for by these known *R* genes mediating this recognition cannot therefore be ruled out (de Kock et al., 2004).

Although Ecp2 was the only proteinaceous element of *C. fulvum* to induce HR on *Nicotiana sp.* of those tested, the absence of the Ecp2 protein does not enable growth of *C. fulvum* on *N. paniculata* (de Kock et al., 2004). This suggests that there is a role for another, perhaps toxic, component of *Nicotiana sp.,* which provides this growth inhibition, or there are other effectors of *C. fulvum*, as yet uncharacterised that are recognised by *R* genes in *N. paniculata* (de Kock et al., 2004; Takken et al., 2000).

1.13 The conservation of Ecp2 across the Dothidiomycetes and the potential for the utilisation of *Cf-Ecp2* encoded resistance

Fungal effectors generally have high sequence divergence, signifying their unique specificity. However, *Ecp2, Ecp6* and *Avr4* have been found in fungal species of the Dothidiomycetes class, in addition to *C. fulvum* (Bolton et al., 2008; de Wit et al., 2012; Stergiopoulos et al., 2010).

Ecp2 is present 153 times within 135 fungi investigated and within one genome, *Ecp2* may be present between 1 – 14 times (Stergiopoulos et al., 2012). Such homologs are termed *Homologs of C. fulvum Ecp2* (*Hce2*). *Ecp2* is not only present in phytopathogens, but also in human pathogens and saprophytes, and there is no correlation between the number of *Hce2s* present in the genome and the ecology of the species (Stergiopoulos et al., 2012). Therefore, the evolutionary driving force is uncertain and specific *Hce2s* may have experienced different functional divergence (Stergiopoulos et al., 2012).

Ecp2 is discontinuously distributed among 52 fungal species; 46 of these species are within Ascomycota, while the others are of the Basidiomycota phylum (Stergiopoulos et al., 2012). This suggests the presence of Ecp2 before the divergence of the Dikarya (Stergiopoulos et al., 2012).

Hce2s were grouped into three classes; class 1 containing 117 small secreted proteins ranging from 80 – 400 amino acids; class 2 consisting of eight proteins of up to 800 amino acids; and class 3 containing 38 proteins with complex architecture (Stergiopoulos et al., 2012). All three classes contain the conserved Ecp2 domain, which consists of four conserved cysteine residues.

Within *Mycosphaerella fijiensis,* the class 1 *Ecp2* sequence is present on three occasions (Stergiopoulos et al., 2010). Mf-Ecp2-1 (161 amino acids) has 57% amino acid sequence identity to *C. fulvum* Ecp2 and is able to induce HR in tomato containing *Cf-Ecp2* (Stergiopoulos et al., 2010). The two other homologs, Mf-Ecp2-2 (174 amino acids) and Mf-Ecp2-3 (236 amino acids), have lower sequence homology of 28% and 25%, respectively, to *C. fulvum* Ecp2 (Stergiopoulos et al., 2010). Both Mf-Ecp2-1 and Mf-Ecp2-2 have a 19 amino acid SP while Mf-Ecp2-3 has an 18 amino acid SP (Stergiopoulos et al., 2010). There is conservation of four cysteine residues present in Ecp2 and a conserved intron in position within the ORF (Stergiopoulos et al., 2010).

1.14 The need for genetic modification in bananas

Bananas, of the Zingiberales, are produced in hot countries surrounding the equator (FAOSTAT, 2015). The top producers of bananas are India, Brazil and Ecuador. All three

are developing countries whose methods of farming are vastly different from the western world (FAOSTAT, 2015; Foresight, 2011). Bananas are number four on the list of staple crops and yet represent a vulnerable agricultural industry (Koeppel, 2008; Shah, 2001). Although there are more than 300 species of banana, only one is grown commercially (Koeppel, 2008).

The Cavendish banana is grown as a monoculture, involving clones that are genetically identical (Koeppel, 2008). This increases the risk of disease, necessitating the application of up to 30 Kg pesticides per hectare per year. This pesticide application accounts for approximately 30% of the cost of bananas in supermarkets (Koeppel, 2008). In fact, Gros Michel, a banana grown previously, was wiped out by Panama disease in the 1950s, a disease caused by *Fusarium oxysporum* (Koeppel, 2008).

The application of pesticides on banana crops is harmful to the health of workers. A film produced by Fredrik Gertten, entitled 'Bananas!*' (http://www.bananasthemovie.com/) highlighted the health issues and economic imbalance associated with banana plantations. Vulnerable workers exposed to Nemagon, a chemical pesticide, became sterile, suffered respiratory problems and developed cancer (Boix and Bohme, 2012). This led to a legal case against Dole Food Company. Furthermore, such pesticides are expensive chemicals and many small holding farmers cannot afford them.

Since banana is sterile, containing the AAA triploid genome from the *Musa acuminata* ancestor (AA, 2n = 22 chromosomes), it does not lend itself to classical genetics. Thus, disease resistance cannot easily be bred in to the commercial banana from its wild relatives (Lescot et al., 2008). This calls for a GM approach to generating disease resistance in bananas.

1.15 The presence of Class 1 *Homologs of C. fulvum Ecp2* effectors within a range of crop pathogens

Class 1 *Hce2* effectors have been identified in many important crop pathogens, including;

- *Dothistroma septosporum*, the causal agent of needle blight on pine trees. This is present in Canadian forests and recent outbreaks may be linked to climate change (Stergiopoulos et al., 2012; Welsh et al., 2009).

- *Zymoseptoria tritici*, causes outbreaks of the devastating wheat disease Septoria leaf blotch and potentially reduces yields by 30 – 40% (reviewed in (Palmer and Skinner, 2002)).

- *V. dahliae*, the causal agent of Verticilum wilt, is capable of infecting tomatoes, potatoes and peppers along with many other dicots (Bhat and Subbarao, 1999).

- *F. graminearum*, the causal agent of wheat and barley head blight and diseases on many other crops, presents a challenge for food quality control due to the production of fungal mycotoxins harmful to humans (Goswami and Kistler, 2004; Logrieco et al., 2007).

- *Magnaporthe grisea*, the causal agent of rice blast, in addition to being the most devastating fungus of rice, also infects other crops (Talbot, 2003).

- *F. oxysporum,* induces vascular wilt of multiple crops including banana. Several *R* genes, effective against the fungus, have been identified in tomato (Michielse and Rep, 2009).

The presence of such effectors across a broad range of fungi, suggests that they provide basic virulence function enabling the infection of a wide range of hosts (Stergiopoulos et al., 2012). Class 1 *Hce2* effectors are key to the success of these fungi in terms of their inherent pathogenicity.

The identification of a gene capable of recognising a conserved effector (i.e. belonging to Class 1 *Hce2*) found across a variety of fungal pathogens, has the potential to be a powerful tool for generating broad spectrum resistance.

1.16 Outline of PhD

Within the research outlined in this PhD project, a candidate for *Cf-Ecp2* (*2A*) was identified in *S. pimpinellifolium* CfEcp2 (Chapter 3). This was completed by fine mapping of the *Cf-Ecp2* locus, subsequent sequencing of the locus and comparing it to two independently-derived deletion mutants that had lost the ability to recognise Ecp2. The ability of *S. pimpinellifolium* CfEcp2 to recognise *M. fijiensis* Ecp2 variants was also described (Chapter 3).

The *Cf-Ecp2* candidate (*2A*) was tested by stable transformation in *S. lycopersicum* Cf0 under control of the Cauliflower Mosaic Virus 35S (35S) promoter (Chapter 4). These stable transformants were characterised by infiltration of Ecp2 protein, Agrobacterium-mediated delivery of PVX:*Ecp2* and crossing with *S. lycopersicum* Cf0 which was overexpressing Ecp2 in a stable manner (Chapter 4).

Finally, the spectrum of recognition of Ecp2 homologs, encoded for by *N. paniculata Cf-Ecp2* was investigated (Chapter 5) and a genetic map of *N. paniculata* was generated.

Chapter 2

Materials and Methods

2.1 Media

Table 2.1 Media used in this study.

Media	Components in 1 Litre
Luria and Bertani (LB)	10 g tryptone, 5 g yeast extract, 10 g NaCl, pH 7.0 (15 g agar for solid media).
Murashige and Skoog (MS)	4.41 g of MS salts (440 mg $CaCl_2.H_2O$, 0.025 mg $CoCl_2.6H_2O$, 0.025 mg $CuSO_4.5H_2O$, 36.7 mg FeNa Ethylenediaminetetraacetic acid (EDTA), 6.2 mg H_3BO_3, 170 mg KH_2PO_4, 0.83 mg KI, 1900 mg $MgSO_4.7H_2O$, 22.3 mg $MnSO_4.H_2O$, 0.25 mg $Na_2MoO_4.2H_2O$, 1650 mg NH_4NO_3, 8.6 mg $ZnSO_4.7H_2O$, 2 mg glycine, 100 mg I-Inositol, 0.5 mg thiamine-HCl), 10 g glucose, 6 g agarose, adjusted to pH 5.8 with KOH.
Yeast Extract Peptone Dextrose Medium (YPD)	10 g yeast extract and 20 g peptone in 900 mL water (add 20 g agar if making plates), autoclaved, then added 100 mL 10X Dextrose.
Buffered-Glycerol/Methanol-complex Medium (BMGY/ BMMY)	10 g yeast extract and 20 g peptone in 1 litre of water, autoclave then added: 100 mL 1 M potassium phosphate pH 6.0 (23 g K_2HPO_4 and 118 g KH_2PO_4 in 1 litre of water adjusted to pH 6.0 with KOH), 100 mL 10x Yeast Nitrogen Base, 2 mL 500X biotin, 100 mL 10x glycerol (for BMGY) or 100 mL 10X methanol (for BMMY).
All media were autoclaved at 15 psi for 15 minutes.	

2.2 Antibiotics, IPTG and X-gal

Table 2.2 Antibiotics, IPTG and X-gal used in this study.

Antibiotic	**Stock**
Carbenicillin	100 mg/mL in water
Kanamycin	50 mg/mL in water
Rifampicin	25 mg/mL in DMSO
Gentamycin	25 mg/mL in water
Streptomycin	50 mg/mL in water
Tetracycline	10 mg/mL in 70% ethanol
Chloramphenicol	30 mg/mL in 100% ethanol
IPTG and X-gal	
IPTG	100 mM; 1.2 g in 50 µL water (100 µL spread on each plate)
X-gal	50 mg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactoside in N,N'-dimethylformamide (20 µL spread on each plate).

IPTG = Isopropyl-beta-D-thiogalactopyranoside , X-gal = 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, DMSO= dimethyl sulfoxide

2.3 Plant growth conditions

Under glasshouse conditions, plants were grown at 23 °C for 18 hours in the light and 6 hours in the dark. When in controlled environment rooms, the plants were grown at 25 °C for 18 hours in the light and 6 hours at 22 °C in the dark.

2.4 Crossing tomato plants

The male anthers from the tomato flower were removed using sterile forceps. The pollen on the surface of the anthers was tapped onto a glass slide. The stigma of another tomato flower was exposed, by removal of the anthers, and dipped in the pollen on the glass slide.

Table 2.3 Crosses completed and plant lines used in this study.

Plant ID
Solanum spp.
S. lycopersicum Cf0
[1]*S. lycopersicum* Cf0 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* Cf0
S. lycopersicum Cf0 x *S. pimpinellifolium* CfEcp2
S. pimpinellifolium CfEcp2 (Ontario 7518 accession)
[1]*S. pimpinellifolium* CfEcp2 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. pimpinellifolium* CfEcp2
S. pimpinellifolium 1178
S. pimpinellifolium 1179
S. lycopersicum Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p1
S. lycopersicum Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p1
S. lycopersicum Cf0 35S:*2A*p2
[1]*S. lycopersicum* Cf0 35S:*2A*p2 x *S. lycopersicum* Cf0 35S:*Ecp2* and *S. lycopersicum* Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p2
S. lycopersicum Cf0 35S:*2A*p3
[1]*S. lycopersicum* Cf0 35S:*2A*p3 x *S. lycopersicum* Cf0 35S:*Ecp2* and *S. lycopersicum* Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p3
S. lycopersicum Cf0 35S:*2A*p4
S. lycopersicum Cf0 35S:*2A*p4 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p5
[1]*S. lycopersicum* Cf0 35S:*2A*p5 x *S. lycopersicum* Cf0 35S:*Ecp2* and *S. lycopersicum* Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p5
S. lycopersicum Cf0 35S:*2A*p10
S. lycopersicum Cf0 35S:*2A*p10 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p13
S. lycopersicum Cf0 35S:*2A*p14
S. lycopersicum Cf0 35S:*2A*p14 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p16
S. lycopersicum Cf0 35S:*2A*p16 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p18
S. lycopersicum Cf0 35S:*2A*p18 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p24
S. lycopersicum Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p24
S. lycopersicum Cf0 35S:*2A*p26
S. lycopersicum Cf0 35S:*2A*p28
S. lycopersicum Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p28
S. lycopersicum Cf0 35S:*2A*p2.1

[1]Crosses were completed in either direction and, due to no observed differences, the results were combined. [2]A combination of data from myself and (Westergaard, 2012). [3]Data from (Harder, 2012).

Table 2.3 Continued. Crosses completed and plant lines used in this study.

Plant ID
Solanum spp.
S. lycopersicum Cf0 35S:*2A*p2.2
S. lycopersicum Cf0 35S:*2A*p2.3
S. lycopersicum Cf0 35S:*2A*p2.4
S. lycopersicum Cf0 35S:*2A*p2.5
S. lycopersicum 35S:*2A*p2.6
S. lycopersicum Cf0 35S:*2A*p3.1
S. lycopersicum Cf0 35S:*2A*p3.2
S. lycopersicum Cf0 35S:*2A*p3.3
S. lycopersicum Cf0 35S:*2A*p3.5
S. lycopersicum Cf0 35S:*2A*p3.6
S. lycopersicum Cf0 35s:*2A*p3.8
S. lycopersicum Cf0 35S:*2A*p3.8 x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p3.9
S. lycopersicum Cf0 35S:*2A*p3.10
S. lycopersicum Cf0 35S:*2A*p3.11
S. lycopersicum Cf0 35S:*2A*p4.1
S. lycopersicum Cf0 35S:*2A*p4.2
S. lycopersicum Cf0 35S:*2A*p5.1
S. lycopersicum Cf0 35S:*2A*p5.2
S. lycopersicum Cf0 35S:*2A*p5.3
S. lycopersicum Cf0 35S:*2A*p18.1
S. lycopersicum Cf0 35S:*2A*p24.1
S. lycopersicum Cf0 35S:*2A*p24.2
S. lycopersicum Cf0 35S:*2A*p24.3
S. lycopersicum Cf0 35S:*2A*p24.4
[1]*S. lycopersicum* Cf0 35S:*2A*p24.4 x *S. lycopersicum* Cf0 35S:*Ecp2 and S. lycopersicum* Cf0 35S:*Ecp2* x *S. lycopersicum* Cf0 35S:*2A*p24.4
S. lycopersicum Cf0 35S:*2A*p24.5
S. lycopersicum Cf0 35S:*2A*p24.4.2D
S. lycopersicum Cf0 35S:*2A*p24.4.2D x *S. lycopersicum* Cf0
S. lycopersicum Cf0 35S:*2A*p24.4.2D x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p24.4.4D
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0 35S:*Ecp2*
S. lycopersicum Cf0 35S:*2A*p24.5.5D
S. lycopersicum Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0
S. lycopersicum Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0 35S:*Ecp2*
S. pimpinellifolium 1179p15

[1]Crosses were completed in either direction and, due to no observed differences, the results were combined. [2]A combination of data from myself and (Westergaard, 2012). [3]Data from (Harder, 2012).

Table 2.3 Continued. Crosses completed and plant lines used in this study.

Plant ID
Solanum spp.
S. pimpinellifolium 1179p19
S. pimpinellifolium 1179p31
S. pimpinellifolium 1179p33
S. pimpinellifolium CfEcp2 x *S. pimpinellifolium* 1179p15
S. lycopersicum Cf0 35S:*2A*p3.8.1
S. lycopersicum Cf0 35S:*2A*p3.8.1 x *S. lycopersicum* Cf0
S. lycopersicum Cf0 35S:*2A*p3.8.1 x *S. pimpinellifolium* 1179p19
S. lycopersicum Cf0 35S:*2A*p24.4.2
S. pimpinellifolium 1179p31 x *S. lycopersicum* Cf0 35S:*2A*p24.4.2
S. lycopersicum Cf0 35S:*2A*p24.4.3
S. lycopersicum Cf0 35S:*2A*p24.4.3 x *S. pimpinellifolium* 1179p15
S. lycopersicum Cf0 35S:*2A*p24.4.4
S. lycopersicum Cf0 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19
S. lycopersicum Cf0 35S:*2A*p24.5.3
S. lycopersicum Cf0 x *S. lycopersicum* Cf0 35S:*2A*p24.5.3
S. lycopersicum Cf0 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19
S. lycopersicum Cf0 35S:*2A*p24.5.4
S. lycopersicum Cf0 35S:*2A*p24.5.4 x *S. pimpinellifolium* 1179p15
Nicotiana spp.
[2]*N. paniculata* TW99 CfEcp2 NP_00006
[2]*N. paniculata* TW102 cfecp2 NP_00009
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_1 NP_00019
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_1 NP_00022
[3]*N. paniculata* TW99 x *N. paniculata* TW102 F_2 NP_00036
[3]*N. paniculata* TW102 x *N. paniculata* TW99 F_2 NP_00038
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00145
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00157
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00171
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00176
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00179
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00187
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00191
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00196
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00202
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00208
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00137
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00140
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00148
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00152

[1]Crosses were completed in either direction and, due to no observed differences, the results were combined. [2]A combination of data from myself and (Westergaard, 2012). [3]Data from (Harder, 2012).

Table 2.3 Continued. Crosses completed and plant lines used in this study.

Plant ID
Solanum spp.
[2]*N. paniculata* TW99 x *N. paniculata* TW102 F_3 NP_00162
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00163
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00188
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00190
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00213
[2]*N. paniculata* TW102 x *N. paniculata* TW99 F_3 NP_00231

[1]Crosses were completed in either direction and, due to no observed differences, the results were combined. [2]A combination of data from myself and (Westergaard, 2012). [3]Data from (Harder, 2012).

2.5 Seed isolation

Tomato seeds were scooped out of the fruit, using a spatula. They were incubated in 50% hydrochloric acid for 10 minutes, rinsed, incubated in 10% alkali (38.01 g/L sodium phosphate tribasic dodecahydrate) for 20 – 30 minutes and rinsed. The seeds were air dried on filter paper, and stored in paper seed bags.

Nicotiana sp. seeds were isolated by holding partially-open seed pods with forceps and shaking the seeds into a paper seed bag. The seed pods were also placed into each seed bag.

2.6 Seed sterilisation

Seeds were sterilised in a fume hood for 2 minutes in 70% ethanol, followed by 30 minutes in 10% bleach. They were rinsed in water three times.

2.7 Kanamycin selection of seeds

Seeds were sterilised and grown in tubs of MS media with 300 mg/L kanamycin.

2.8 Plant transient transformation: heterologous expression in tomato and *Nicotiana sp.*

A glycerol stock of *Agrobacterium tumefaciens* GV3101, carrying the clone to be transformed (Table 2.4), was spread on a plate with LB media and selective antibiotics

and grown overnight at 28 °C. One colony-forming unit (CFU) was inoculated into 5 or 10 mL LB media with selective antibiotics and grown overnight at 28 °C on an orbital shaker at 220 rotations per minute (rpm). A 20 µL sample of the overnight culture was inoculated in 200 mL LB media with selective antibiotics and grown overnight at 28°C on an orbital shaker at 220 rpm. The cells were pelleted from solution (25 minutes at 3220 *g*), resuspended in infiltration solution (10 mM $MgCl_2$, 10 mM MES pH 5.6 with KOH, 150 µM acetosyringone in DMSO) and incubated at room temperature for 2 hours. Plant leaves were inoculated on their abaxial side using a sterile, blunt-end syringe. For tomato, both cotyledons were inoculated, while for *N. paniculata* and *N. benthamiana,* a section of a mature leaf was inoculated.

Table 2.4. Constructs used in this study.

Vector	Ab[1]	Description	Reference
Empty Vectors			
pSC-A-amp/kan	Carb	For blunt end cloning. Not suitable for plant transformation. T7 and T3 primers were used to amplify the insert from the vector. Blue white selection in StrataClone Solo cells, with IPTG and X-gal.	StrataClone
pICH86988	Kan	Blue white selection in *Escherichia coli* DH5α, with IPTG and X-gal. For Golden gate cloning. Suitable for plant transformation.	TSL SynBio vector
Clones received			
PVX:*Ecp2*	Kan	With *PR1a* SP in place of native SP.	(Soumpourou et al., 2007)
PVX:*Avr4*	Kan	With *PR1a* SP in place of native SP.	(Soumpourou et al., 2007)
pBIN:35S:*Ecp2*	Kan	With *PR1a* SP in place of native SP.	(Soumpourou et al., 2007)
pPIC9:*Ecp2*	N/A	For growth in *Pichia pastoris*, and secretion of Ecp2 protein. With *PR1a* SP in place of native SP.	(de Kock et al., 2004)
pSFINX:*MfEcp2*-1 i1	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-1 i2	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-1 i3	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-2 i1	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-2 i2	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-2 i3	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-2 i4	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i3	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i4	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i6	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i7	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i8	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i9	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i10	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)
pSFINX:*MfEcp2*-3 i11	Kan	*PR1a* SP in place of native SP.	(Stergiopoulos et al., 2012)

Ab[1]= Antibiotic selection, Carb = Carbenicillin 100 mg/L, Kan= Kanamycin 50 mg/L, SP = signal peptide, *2A* = *OR2A*

Table 2.4. Continued. Constructs used in this study

Vector	Ab[1]	Description	Reference
Clones generated			
pICH86988:35S:*2A*	Kan	*Bsa*I site removed from *2A* gene.	
pICH86988:35S:*Ecp2*	Kan	*PR1a* SP in place of native SP. Sequence of *PR1a*:*Ecp2* optimised by Genewiz.	
pICH86988:35S:*Avr4*	Kan		
pICH86988:35S:*DsEcp 2*	Kan	Ecp sequence from *Dothistroma septosporum. PR1a* SP in place of native SP. Sequence of *PR1a*:*Avr4* optimised by Genewiz.	
pICH86988:35S:*FgEcp 2*	Kan	Ecp sequence from *Fusarium graminearum. PR1a* SP in place of native SP. Sequence of *PR1a*:*FgEcp2* optimised by Genewiz.	
pICH86988:35S:*FoEcp 2*	Kan	Ecp sequence from *Fusarium oxysporum. PR1a* SP in place of native SP. Sequence of *PR1a*:*FoEcp2* optimised by Genewiz.	
pICH86988:35S:*MgrEc p2*	Kan	Ecp sequence from *Magnaporthe grisea. PR1a* SP in place of native SP. Sequence of *PR1a*:*MgrEcp2* optimised by Genewiz.	
pICH86988:35S:*MfEcp 2*	Kan	Ecp sequence from *Mycosphaerella fijiensis. PR1a* SP in place of native SP. Sequence of *PR1a*:*MfEcp2* optimised by Genewiz.	
pICH86988:35S:*MgEc p2*	Kan	Ecp sequence from *Mycosphaerella graminicola. PR1a* SP in place of native SP. Sequence of *PR1a*:*MgEcp2* optimised by Genewiz.	
pICH86988:35S:*SmEc p2*	Kan	Ecp sequence from *Septoria musiva. PR1a* SP in place of native SP. Sequence of *PR1a*:*SmEcp2* optimised by Genewiz.	

Ab[1]= Antibiotic selection, Carb = Carbenicillin 100 mg/L, Kan= Kanamycin 50 mg/L, SP = signal peptide, *2A* = *OR2A*

Table 2.4. Continued. Constructs used in this study

Vector	Ab[1]	Description	Reference
pICH86988:35S:*VdEcp2*	Kan	Ecp sequence from *Verticillium dahliae. PR1a* SP in place of native SP. Sequence of *PR1a*:*VdEcp2* optimised by Genewiz.	

Ab[1]= Antibiotic selection, Carb = Carbenicillin 100 mg/L, Kan= Kanamycin 50 mg/L, SP = signal peptide, *2A* = *OR2A*

2.9 Generation of electrocompetent *Agrobacterium tumefaciens* GV3101

A. tumefaciens GV3101 was streaked from a glycerol stock onto a plate with LB media supplemented with 50 µg/mL rifampicin and 25 µg/mL gentamycin and incubated for 48 hours at 28 °C. One CFU was incubated in 10 mL LB media with selective antibiotics over 48 hours at 28 °C with shaking, 5 mL of which were inoculated in 500 mL LB media with selective antibiotics and grown over 48 hours at 28 °C with shaking. The cells were placed on ice for 30 minutes, centrifuged at 3220 *g* for 15 – 20 minutes at 4 °C and the supernatant removed. The cells were resuspended in approximately 100 mL ice-cold 10% glycerol, and re-pelleted (3220 *g* 15 – 20 minutes at 4 °C). The process of washing in 10% glycerol was repeated four times, increasing from 15 – 30 minutes in the last two spins as the pellet became looser. The final pellet of cells was re-suspended in 3 mL 10% glycerol (v/v), split into 50 µL aliquots and stored at -80 °C. Great care was taken to keep the cells cold throughout.

2.10 Generation of chemically competent *Escherichia coli* DH5α

E. coli DH5α was streaked from glycerol stock on LB media and incubated for 24 hours at 37 °C. 5 mL LB media was inoculated with one CFU, and incubated for 24 hours at 37 °C with shaking, of which 100 µL was inoculated in 250 mL LB media and incubated for 24 hours at 37 °C with shaking. The culture was incubated on ice for 1 minute and the cells pelleted at 4500 *g* for 10 minutes at 4 °C. The cells were re-suspended in 100 mL ice-cold TFB1 buffer (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% (v/v) glycerol, pH 5.8 with 1 M glacial acetic acid and filter-sterilised), and incubated on ice for 15 minutes. The cells were centrifuged at 4500 *g* for 10 minutes at 4 °C. The supernatant was removed and the pellet was re-suspended in 10 mL TFB2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl and 15% (v/v) glycerol, pH to 6.5 with NaOH, and filter-

sterilised). The cells in TFB2 were incubated on ice for 30 – 60 minutes, and split into 100 µL aliquots, and stored at -80 °C. Great care was taken to keep the cells cold throughout.

2.11 Electroporation of *A. tumefaciens* GV3101 or Agl1

A 50 µL aliquot of *A. tumefaciens* GV3101 or Agl1 cells was thawed on ice for 5 - 20 minutes. A 1 – 10 µL aliquot of DNA was added to the cells. The tube was flicked to mix the contents, and incubated on ice for 30 minutes. The cells were transferred to a 1 mm cuvette and placed back on ice. The cells were subjected to electroporation (2.2 kV). 200 – 500 µL of LB media, was added to the cells and the total mix was transferred to a microcentrifuge tube. The cells were recovered by shaking-incubation at 28 °C for 1 – 2 hours. The cells were then plated on LB media plates containing 50 µg/mL rifampicin and 25 µg/mL gentamycin plus selective antibiotics to select for the clone. They were grown over two nights for 48 hours at 28 °C.

2.12 Heat shock transformation of *E. coli* DH5α

A 50 µL aliquot of *E. coli* DH5α was thawed on ice for 5 - 20 minutes. A 1 – 10 µL aliquot of DNA was then added to the cells. The tube was flicked to mix the contents and incubated on ice for 30 minutes. The cells were placed at 42 °C for 1 minute and placed back on ice immediately after for 2 minutes. 500 – 900 µL of LB media was added and the cells were recovered by shaking-incubation at 37 °C for 1 – 2 hours. The cells were plated on LB media plates containing selective antibiotics and grown overnight at 37 °C.

2.13 Protein production and analysis

2.13.1 Ecp2 production in *Pichia pastoris*

The construct pPIC9:PR1a:*Ecp2* was transformed into *Pichia pastoris* GS115 by Richard Hughes according to the Invitrogen *Pichia* expression kit user-guide (Invitrogen, 2014). The *P. pastoris* expressing Ecp2 was cultured according to the Invitrogen *Pichia* expression kit user-guide for secreted proteins (Invitrogen, 2014). The supernatant was cleared of any remaining cells by filtering through a 0.22 µM filter paper. The supernatant was stored at 4°C until processed for Ecp2 purification.

2.13.2 Protein gels

Pure proteins, protein mixes and supernatants containing proteins were run on 17% SDS-PAGE gels (5.7 mL 30% acrylamide solution, 2.5 mL 1.5 M Tris pH 8.8, 0.1 mL 10% SDS, 0.1 mL 10 % (w/v) ammonium persulfate, 0.004 mL TEMED, made up to 10 mL with water). 1x loading buffer (50 mM Tris-HCl pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% (v/v) glycerol) was added to 5 – 30 µL of protein sample and the mix was heated at 95 °C for 3 minutes. The gel was loaded with up to 10 µL of the protein-loading buffer mix and 4 µL of ladder (Expedeon RunBlue prestained dual colour marker) and run at 200 volts for 45 minutes. The gel was rotated in visualization buffer (Expedeon InstantBlue buffer) for 30 minutes and protein bands were observed.

2.13.3 Ecp2 protein isolation from *P. pastoris* supernatant

The ÄKTA-900 protein purifier (Box-900, UPC-900, P-900 and Frac-920) was used at room temperature with Unicorn software version 3.21. A 5 mL Histrap-FF column was loaded with 1 L of supernatant containing Ecp2 protein diluted 1x with A1 buffer, with a flow of 3 ml/minute and a pressure of 0.8 MPa. The column was cleaned using 20 mL of A1 buffer to remove unbound proteins, with the parameters outlined above, and the follow-through was collected. The protein was eluted with approximately 100 mL B1 buffer, with the parameters outlined above, or until the protein peak had been collected, and 1 mL fractions were collected. This was repeated for four separate litres of supernatant. The fractions collected across the protein peak were pooled and concentrated to 15 mL, using Vivaspin 500 (with 5000 Da molecular weight cut-off) according to the manufacturer's instructions.

The 15 mL protein sample was purified further using gel purification (HiLoad 26/600 and Superdex 75 pg column) on the ÄKTAxpress, and the fractions around the protein peak were collected and pooled. The protein was separated into 50 µL fractions of 500 µM and stored at -80 °C.

2.13.4 Protein infiltrations

A 500 µM fraction of protein in A4 buffer was diluted 33-fold to 15 µM with sterile water. The 15 µM Ecp2 was infiltrated into cotyledons or adult leaves of tomato or adult leaves of *N. paniculata* alongside A4 buffer diluted 33-fold with sterile water, with a blunt end syringe.

2.14 DNA extractions

2.14.1 DNA extraction from agarose gel

Selected DNA bands in the gel were cut out and extracted using the QIAquick gel extraction kit according to manufacturer's instructions (QIAgen).

2.14.2 Plasmid

Colonies were inoculated in 5 – 10 mL LB media with selective antibiotics and grown overnight at 28 °C (for *A. tumefaciens*) and 37 °C (for *E. coli*). Plasmid extractions were completed using QIAprep Spin Miniprep Kit according to the manufacturer's instructions (QIAgen).

2.14.3 BACs

BAC DNA extractions were completed using QIAGEN Large-Construct Kit according to the manufacturer's instructions (QIAgen).

2.14.4 Plant DNA: DNeasy Plant Mini Kit

Plant DNA extracted using the DNeasy Plant Mini Kit was completed according to the manufacturer's instructions (QIAgen).

2.14.5 RNase treatment of DNA

RNA was removed from a 30 -100 μL sample of DNA by the addition of 1 μL of 10 mg/mL RNase A. The reaction mix was heated for 30 minutes at 37°C. The RNase A was inactivated by heating the reaction mix at 65°C for 20 minutes.

2.14.6 Plant DNA: CTAB method

A 1 - 2 g sample of ground leaf tissue was added to 6 mL freshly-prepared nuclear extraction buffer (60 mL nuclear lysis buffer [200 mM TRIS-HCl pH 7.5, 50 mM EDTA, 2 M NaCl, 2% (w/v) CTAB, made up to 60 mL with water], 0.38 g sodium bisulphite and 40 mL water) and 1.3 mL 5% (v/v) N-lauryl sarkosine and 1 μL of 10 mg/ml RNase A. The sample was inverted six times to mix and incubated at 65 °C for 10 – 20 minutes, inverting the sample every 5 minutes to mix. An equal volume of phenol:chloroform:IAA (25:24:1) was added to the sample and inverted 30 times to mix. The sample was spun at >4000 rpm for 15 minutes and the supernatant retained. A 0.6 volume aliquot of 100% isopropanol was added to the supernatant and mixed by inversion three to four times. The DNA was pelleted by centrifugation at >4000 rpm for 5 minutes. The DNA pellet was washed in 80% ethanol, air dried for 30 minutes and suspended in 300 – 400 μL water. For smaller tissue samples the protocol was altered by reducing the volumes accordingly.

2.14.7 Plant DNA: Nuclear enrichment

Dry seedlings were ground in liquid Nitrogen to a fine powder and added to 30 mL Extraction buffer 1 (0.4 M sucrose, 10 mM Tris-HCl pH 8, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol). The components were mixed by inversion and filtered through double Miracloth into a clean tube. This was repeated to remove additional debris and the filtered solution was kept on ice. The filtered solution was pelleted by centrifugation 4000 rpm for 20 minutes at 4 °C and the supernatant was removed gently. The pellet was thoroughly resuspended in 1 mL Extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl pH8, 10 mM $MgCl_2$, 1% (w/v) Triton X-100 and 5 mM β-mercaptoethanol) and moved to a clean tube on ice. The nuclei were pelleted from solution by centrifugation at 12000 rpm for 10 minutes at 4 °C and the supernatant was removed gently. The inside of the tube was wiped to remove excess debris and the pellet was re-suspended in 300 μL Extraction

buffer 3 (1.7 M sucrose, 10 mM Tris-HCl pH 8, 0.15% (w/v) Triton X-100, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol). The solution carrying the re-suspended pellet was carefully pipetted on top of another 300 μL Extraction buffer 3 in a clean tube. This mixture was centrifuged for 30 minutes at 16,000 rpm at 4 °C. The pellet was re-suspended in 300 μL Extraction buffer 3 and the overlay step followed by centrifugation was repeated until all green (plastid DNA) was removed from the supernatant. The final clean pellet of chromatin DNA was re-suspended in 700 μL Nuclear Lysis buffer (200 mM Tris-HCl pH 7.5, 50 mM EDTA, 2 M NaCl, 2% CTAB titred to pH 7.5 with HCl). A further 1.5 mL Nuclear lysis buffer was added with 1 mL ice cold 5% (w/v) Sarkosyl. The final solution was incubated at 60°C for 20 minutes to lyse the nuclei. The solution was mixed with 8.5 mL chloroform:Isoamyl alcohol (IAA) (24:1) by inversion and centrifuged at 3000 rpm for 20 minutes. The aqueous top phase was placed into a new tube and two volumes of cold 100 % (v/v) isopropanol were added to it. This mixture was left to rest for 5 minutes inverting gently every minute. The DNA was pelleted by centrifugation at 4000 rpm for 20 minutes. The pellet was air dried and re-suspended in 100 μL of 10 mM Tris and 0.1 mM EDTA buffer pH 8.0 and 1 μL 10 mg/mL RNase.

2.15 Plant RNA extraction and conversion into cDNA

Plant RNA was extracted using RNeasy QIAgen kit according to the manufacturer's instructions. RNA was eluted in 50 μL nuclease-free water. The RNA was treated with Roche recombinant DNase according to the manufacturer's instructions and subjected to reverse transcription PCR with SuperScript™ II or III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. To show that DNA contamination did not influence later PCR reactions, each RNA sample was run with and without the SuperScript™ II or III Reverse Transcriptase to generate a negative control. The results of the negative controls were not discussed since products were not seen.

2.16 Cleaning DNA

If DNA was determined to be un-pure, or required further RNase treatment, it was cleaned using phenol-chloroform or a sodium acetate and isopropanol method.

2.16.1 Phenol chloroform purification

DNA was cleaned by the addition of 1 volume of phenol:chloroform:IAA (25:24:1). The reaction mix was vortexed and centrifuged at 13,000 rpm for 10 minutes. The supernatant was kept and 0.8 volumes of 100% (v/v) isopropanol was added and mixed by inverting the tube six times. The DNA was pelleted by centrifugation of the reaction mix at 13,000 rpm for 5 minutes. The supernatant was discarded and the DNA pellet was washed in 80% ethanol, air dried for 30 minutes and suspended in 30 - 100 µL water.

2.16.2 Sodium acetate and isopropanol purification

DNA was cleaned by the addition of 2.5 volumes of 100% isopropanol and 0.1 volume of 3M sodium acetate pH 5.2. The solution was stored at -20°C for 30 minutes. The DNA was pelleted by centrifugation at 13,000 rpm for 30 minutes. The supernatant was discarded and the DNA pellet was washed in 80% (v/v) ethanol, air dried for 30 minutes and suspended in 30 - 100 µL water.

2.17 Polymerase chain reaction

For homemade *Taq* polymerase, <100 ng DNA was added to the reaction mix (1x buffer [10 mM Tris pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl], 1 µM forward primer, 1 µM reverse primer and 5 units homemade *Taq* polymerase) and made up to 40 µL with sterile water. For the use of GoTaq® DNA Polymerase (Promega), reaction mixtures were made with <100 ng DNA according to the manufacturer's instructions. The *Taq* PCR were subjected to an initial denaturation at 94 °C for 2 – 3 minutes, followed by 35 – 40 cycles of denaturation at 94 °C for 30 seconds, annealing for 30 seconds (temperature primer-dependent) and elongation at 72 °C (time template dependent) (Appendix 1 Table 1.1). The final elongation was at 72 °C for 2 minutes.

For colony PCR, one colony was picked and added to the reaction mix in place of DNA. The rest of the protocol was followed, as outlined above.

Phusion® High-Fidelity DNA Polymerase reactions were run according to the manufacturer's instructions (New England Biolabs).

All PCR products were visualised using agarose gel electrophoresis, unless stated otherwise.

2.18 Restriction digestion

A 10 - 20 µL aliquot of PCR product was digested with specified restriction endonuclease in a solution of 1x recommended buffer and 1x BSA to a total reaction volume of 15 - 30 µL made up with sterile water.

All restriction digestion products were visualised using agarose gel electrophoresis, unless stated otherwise.

Table 2.5. Parameters for marker and BAC DNA digestion.

DNA	Restriction enzyme	Buffer	Temperature (°C)	Incubation time (hours)
TG236	*Nla*III [1]	4 or CutSmart (NEB)	37	12
SNPE	*Dde*I[1]	H (Roche)	37	12
CT116	*Hha*I[1]	4 or CutSmart (NEB)	37	12
SNPN	*Hinf*I[1]	4 or CutSmart (NEB)	37	12
TG184	*Bcl*I[1]	3.1 (NEB)	50	12
SNP-nn1	*Hyp*188I[1]	4 or CutSmart (NEB)	37	12
GJ32				
GJ44				
GJ43				
SNPQ	*Hinf*I[1]	4 or CutSmart (NEB)	37	12
BACS 7B, 4B, 11G	*Not*I[1]	H (Roche)	37	12
BAC 7B	*Eco*RI[1]	H (Roche)	37	12
BAC 7B	*Ssp*I[1]	*Ssp*1 (Invitrogen)	37	12

[1] 5 units of enzyme used, BAC = Bacterial artificial chromosome.

2.19 Electrophoresis

2.19.1 Agarose gel electrophoresis

A 1 – 25 µL sample of DNA or RNA was run with 1x loading buffer (6x loading buffer is 30% glycerol, 10 mM Tris pH 7.5, 1 mM EDTA, 0.04% w/v bromophenol blue, 0.04% w/v

Xylene Cyanol) on a 0.8, 1.0 or 2.0 % agarose gel made with 1x Tris Borate EDTA (TBE) against 5 µL of a DNA molecular weight marker, either 1 kb NEB, 100 bp NEB, 1 kb Invitrogen or 2log Invitrogen (10 µL 1 µg/µL marker, 10 mM pH 7.5 Tris, 1 mM Tris, 1 mM EDTA, 1x loading buffer, to 100 µL using sterile water).

2.19.2 Pulse field gel electrophoresis

A 20 – 25 µL sample of DNA was run with 1x loading buffer in a 1 % agarose gel made with 0.5x TBE against the markers 1 Kb+ (Invitrogen) and low range PFGE marker (Invitrogen), at 16 volts, lower time 5 minutes, upper time 15 minutes and ratio 1.0, for 17 hours.

2.20 Sepharose cleaning

A 25 mL Sepharose suspension was washed 2 – 3 times in 500 mL 10 mM Tris-HCl, 0.5 µM EDTA, and resuspended in 250 mL 10 mM Tris-HCl, 0.5 µM EDTA, 0.2 % sodium azide.

2.21 Sanger sequencing of PCR products

A 5 µL PCR product was sequenced with 5 µL of 10 µM of primer using Sanger sequencing by GATC biotech or The Genome Analysis Centre (TGAC). Sequencing products were compared to predicted sequences using Geneious mapping R8 software (Kearse et al., 2012), with default settings. Sequences were trimmed at the ends to remove low quality sequence, and sequencing errors were ignored.

2.22 Marker development in tomato

New Cleaved Amplified Polymorphic Sequence (CAPS) markers at the CfEcp2 locus (SNPE, SNPN, SNP-nn1, GJ32, GJ44, GJ43 and SNPQ) were developed by Gopaljee Jha (2011, unpublished). These were obtained by mapping short read data from *S. pimpinellifolium* CfEcp2 (Table 2.6) to the publicly available genome sequence of *S. lycopersicum* cv. Heinz (Sato et al., 2012). Further markers (TG58 and 60250) were identified between TG236 and SNPE and two markers (TG67 and TG24) between TG184 and SNPN, using publicly available markers from SolGenomics (https://solgenomics.net) (Appendix 1 Table 1.1).

2.23 Fine mapping *Cf-Ecp2*

To map the Cf-Ecp2 locus a population of 1700 F_2 plants from a cross between *S. pimpinellifolium* (CfEcp2) and *S. lycopersicum* (Cf0) were inoculated with *A. tumefaciens* GV3101 pBIN:PVX:*Ecp2* and their phenotypes scored 14 days post inoculation (d.p.i.) (Gopaljee Jha, unpublished). DNA from 450 survivors was extracted and genotyped with the markers TG236 and TG184 (Table 2.5) (Gopaljee Jha, 2011 unpublished). A total of 80 individuals were identified with recombinations between TG236 and TG184, and further genotyped with CAPS markers SNPE, SNPN, SNP-nn1, GJ32, GJ44, GJ43 and SNPQ (Gopaljee Jha, 2011 unpublished). Two recombinants, 4H09 and 2C10, were selected with recombination breakpoints between SNPE and SNPQ and were selfed to obtain F_3 families.

In addition, a PCR screen on 750 F_2 plants from the same cross was completed using the CAPs markers TG236 and TG184 (Table 2.5 and Appendix 1 Table 1.1) (Gopaljee Jha, 2011 unpublished). Out of 92 recombinants between the two markers, two recombinants (11E01 and 13D08) were identified that contained recombinations between the markers SNPE and SNPN. These two recombinants were further analysed with the CAPS markers SNP-nn1, GJ32, GJ44, GJ43 and SNPQ and selfed to obtain F_3 (Gopaljee Jha, 2011 unpublished).

From each F_3 family, of the recombinants 4H09, 2C10, 11E01 and 13D08, 21 to 38 15-day-old plants were inoculated with *A. tumefaciens* GV3101 pBIN:PVX:*Ecp2* and their phenotypes scored at 11 d.p.i. DNA was extracted from 12 survivors of each recombinant F_3 family. The recombinant family 4H09 was subjected to analysis using the TG236 marker and two F_3 plants were selected as homozygous for the CfEcp2 allele at the TG236 marker. The recombinant family 11E01 was subjected to analysis using the TG236 and CT116 markers and two F_3 recombinants were selected as homozygous for the Cf0 allele at both markers. The recombinant family 13D08 was analysed using the SNPQ marker and three F_3 plants were selected as homozygous for the CfEcp2 allele. The recombinant family 2C10 was subjected to the SNPQ marker and one F_3 plant was selected as homozygous for the CfEcp2 allele. The selected F_3 recombinants were selfed to obtain F_4 seed.

To further confirm the *Cf-Ecp2* map location, nine recombinants were selected from the 92 recombinants between the TG236 and TG184 markers described above. The nine recombinants contained the marker from the CfEcp2 parent at TG236 and SNPE or TG184 and SNPN, so they retained the CfEcp2 locus and the non-recombinant chromosome was from Cf0 (Gopaljee Jha, 2011 unpublished). These recombinants were selfed to obtain F_3 seed. To phenotype the F_3 plants, eight plants of each F_3 recombinant were inoculated with *A. tumefaciens* GV3101 pBIN:PVX:*Ecp2* and their phenotypes scored at 13 d.p.i., and DNA was extracted from six plants of each F_3 recombinant. The DNA samples were genotyped with the CAPS markers TG236, SNPE, CT116, SNPN and TG184. Plants homozygous for the recombinant chromosome were selected for further analysis. The markers TG58, 60250, TG67 and TG24 were tested on the recombinants to further define the recombination breakpoints. The PCR products were sequenced with Sanger sequencing to score their genotype.

2.24 Generation of BAC library

To aid in the cloning of *Cf-Ecp2*, a physical map of the locus encoding the absence of Ecp2 recognition was generated. A *S. pimpinelifolium* CfEcp2 BAC library of approximately 40,000 clones with an average insert size of 118 kb (equivalent to a five-fold coverage of the 950 Mb *S. pimpinelifolium* genome) was generated by BioS&T (Quebec, Canada), via partial digestion of the DNA using the restriction enzyme *Hind*III. The library was screened for clones containing SNPN, SNPE and *2A* (Appendix 1 Table 1.1).

2.25 BAC analysis by restriction endonuclease digestion

To confirm the insert size of the BACs, the BACs 11G and 4B were digested with *Not*I and run on a 0.8% agarose gel against the 1Kb+ Invitrogen marker. In addition, BAC 7B was digested with *Eco*RI and *Not*I and run on a 1% agarose gel under pulse field gel electrophoresis conditions. BAC 7B, digested with *Eco*R1, was also run on a 0.8% agarose gel alongside the 1Kb+ Invitrogen marker. *In silico* digestion of the BACs was performed using NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/).

2.26 Whole genome sequencing

The genomes of *S. pimpinellifolium* CfEcp2 and *S. pimpinellifolium* 1179 were sequenced on HiSeq2000 and HiSeq2500 (Table 2.6).

2.27 BAC sequencing

To determine the sequence of the *Cf-Ecp2 locus*, the BACs 11G, 7B and 4B were sequenced using a combination of short-read sequencing technologies, and PacBio and MinION single-molecule sequencing technologies (Table 2.6).

Table 2.6. Details of sequencing methods used to characterise the *Cf-Ecp2* locus.

Template	Template size	Sequencing platform and chemistry
S. pimpinellifolium CfEcp2	900 Mb	HiSeq2000 75 bp PE reads
BAC 4B	140 kb	454 PE reads
BAC 11G	110 kb	454 PE reads
BAC 7B	122 kb	MiSeq 250 bp PE reads; PacBio RSI, 10 kb library, 120 minute movie; PacBio RSII on two libraries; MinION on two libraries.

PE = Paired end, BAC = Bacterial artificial chromosome, kb = kilo base, Mb = Mega base, bp = base pair.

2.27.1 Short-read sequencing

The BACs 11G and 4B were sequenced on the 454 platform at TGAC (Norwich, UK) (Table 2.6). Paired End (PE) 454 data was assembled using MIRA v3.4.0 with standard parameter settings (-job = accurate, denovo, genome, 454). Prior to assembly, reads originating from the *E. coli* genome were discarded and parts of vector pIndigoBAC-5 were removed in order to break the circular structure of the BAC sequence. The protocol for MIRA assembly was essentially taken from Taudien (2011).

BAC 7B was sequenced using Illumina MiSeq short read technologies at TGAC (Norwich, UK) (Table 2.6). Prior to assembly, the reads were reduced to a 300-fold coverage. The reads from 7B were assembled into a 47 kb contig using MIRA (Taudien et al., 2011).

2.27.2 Single-molecule sequencing

BAC 7B was sequenced using PacBio RSI technology (Table 2.6). BAC 7B was also sequenced using PacBio RSII technology on two independent Single molecule real time sequencing (SMRT) cells. The BAC 7B PacBio runs were assembled independently into contigs using Hierarchical Genome Assembly Process (HGAP) V1 or V3, Celera Assembler V7.0 or V8.0 and Quiver.

Two independent Flow cell runs of MinION sequencing were conducted on 7B DNA (Kamil Witek, The Sainsbury Laboratory and Michael Giolai, (TGAC, Norwich, UK)).

2.28 Analyses of sequence homology

2.28.1 Read mapping

Reads were mapped using Burrow-Wheeler Aligner Mapper (bwa) (version bwa-0.7) (Li and Durbin, 2009). Mappings were filtered for reads mapping as a proper pair, sorted and, in case of Illumina data, filtered for redundancy using Samtools (version 0.1.19) (Li et al., 2009). See below for an exemplary mapping pipeline:

```
bwa index contig.fasta
bwa aln contig.fasta Read1.fastq > read1_sa.sai
bwa aln contig.fasta read2.fastq > read2_sa.sai
bwa sampe contig.fasta read1_sa.sai read2_sa.sai read1.fastq read2.fastq > aln-pe.sam
samtools view -b -f 2 -h -o aln-pe.bam -S -u aln-pe.sam
samtools sort aln-pe.bam aln-pe.sorted
samtools rmdup aln-pe.sorted.bam aln-pe.rmdup.bam
samtools index aln-pe.rmdup.bam
```

PacBio reads were mapped with mem:

```
bwa index contig.fasta
bwa mem -t 10 -L 1000 contig.fasta PacBio_reads.fastq > PacBio_reads.raw.sam
samtools view -Shub -o PacBio_reads.raw.bam -S -u PacBio_reads.raw.sam
samtools sort PacBio_reads.raw.bam PacBio_reads.sorted
samtools index PacBio_reads.sorted.bam
```

All outputs were visualised using Tablet version 1.14.10.20 (Milne et al., 2013) or Savant version 2.0.5 (Fiume et al., 2012).

2.28.2 Local alignments

NCBI Nucleotide Basic Local Alignment Search Tool (BLASTN) (version 2.2.28) (Zhang et al., 2000) was used to compute local alignments between nucleotide sequences.
Dotter (http://sonnhammer.sbc.su.se/Dotter.html) was used to generate dotplots for manual inspection of local alignments

2.28.3 Multiple alignments

Multiple sequence alignments were performed using Geneious, ClustalW and MUSCLE, using Geneious R8 software and the default parameters.

2.29 Analysis of Illumina and 454 generated sequences

The position of *Hcr9s* was identified via dotplot comparison of the contigs from BACs 4B, 11G and 7B, to each other and the published *OR* locus sequence (de Kock et al., 2005). The sequences of the *Hcr9s* were compared to each other using ClustalW. Furthermore, the marker GJ44 was identified on 4B, and the markers CT116 and SNPE were identified on 11G by mapping the known sequence of the markers onto the contigs using Geneious R8 software and default settings.

The 454 reads from 11G were mapped by Basic Local Alignment Search Tool (BLAST), onto the published *OR* cluster sequence. This enabled identification of reads that map to *2A* but with errors (de Kock et al., 2005). The assembled contigs from 11G, 7B and 4B, were analysed in NCBI, using BLASTN, to confirm their origin from *S. pimpinellifolium* (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

To determine the coverage of reads across the Illumina contig, the reads from the 7B BAC sequencing were re-mapped by BLAST. The reads were mapped back onto the assembled contig. In addition, the coverage and SNPs between the raw reads and assembly were

recorded. The 7B BAC 250 PE Illumina reads were also mapped back onto the 47 kb contig using BWA, and the BAM file was visualised in Savant version 2.0.5. This enabled closer inspection of the SNPs, and their percentage presence in the reads.

2.30 Analysis of PacBio assemblies and MinION reads

To resolve the repeat region of the *Cf-Ecp2* locus, which was collapsed in the assembly generated from Illumina 250 bp PE reads, BAC 7B was sequenced using single-molecule sequencing and assemblies were generated (Table 2.6). PacBio assemblies were compared to each other, to the 7B Illumina assembly and to *2A*/*B*/*C* homologs using Dotplots. This enabled positioning of the *Hcr9s* and identification of the 8.7 kb *2A*-containing repeats. The sequences of the *Hcr9s* were extracted from assemblies C and D and compared to each other using ClustalW.

All reads generated by MinION sequencing were aligned by BLAST against PacBio assembly D and the reads which were >9 kb and had BLAST hits in the repeat region were selected (nucleotide position 1,265 to 58,486). This resulted in 19 filtered reads. To determine the error rate of the MinION reads, the filtered reads were aligned to reference BAC sequence of PacBio assembly D using BLASTN. In this specific case, default option “megablast” was disabled setting parameter “-task” to "blastn". This reduces the word size of initial perfect match from 24 to 11 and thus increases sensitivity of the blast algorithm. Identity percentage was calculated as local alignment length / number of identical bases. Candidates from the local alignment between MinION reads and PacBio assemblies were further inspected using dotplots.

The positions of the three *2A* homologs on MinION read “20kb_run1” were identified using Dotter and Geneious. The sequence of each was extracted from the read and compared to the sequences of *2A*, *ψd2A4* and *ψ2A5* using Geneious. The SNPs between the read and the *2A* homologs were manually scored and points assigned to each homolog when they carried the same SNP as the MinION read but different from at least one of the other homologs.

2.31 Identification and manual resolution of errors in BAC assembly D

Errors in the PacBio assembly D were identified by mapping the MiSeq 250 bp paired-end reads and long (>15 kb) raw PacBio reads to assembly D using BWA and subsequent visual inspection using Savant. Each identified error was manually changed in the assembly according to the raw reads. The raw reads were then remapped onto the edited assembly and the procedure was re-iterated until no further inconsistencies between reads and assembly could be resolved. All homologs of *Cf-9* (i.e. *Hcr9s*) were annotated in the curated assembly and pairwise comparisons were made between them and their promoters using ClustalW.

2.32 *2A* homolog cloning and sequencing

2.32.1 Presence and absence of *OR2A*, *OR2B* and *OR2C* on BACs 7B, 11G and 4B

The BACs 7B, 4B and 11G were subjected to PCR to check for the presence or absence of *2A*, *2B* and *2C* (Appendix 1 Table 1.1). The BACs 4B and 11G were subjected to PCR for *2A* (Wulff 2012, unpublished). The products were run in a 1% agarose gel alongside the 1Kb+ Invitrogen Ladder, and presence/absence genotypes were scored.

2.32.2 Amplification and cloning of *2A* homologs from BAC 7B

A Phusion PCR was performed on BAC 7B to clone homologs of *2A*. A range of different amounts of template DNA including 0, 3, 12, 30 and 50 ng were used in six separate reactions. The primers 5’ ψOR2A57B and 3’ OR2A/B7B (set *ψ2A5*) were used to amplify homologs with the same promoter and terminator as *ψ2A5*, and the primers 5’ OR2A7B and 3’ OR2A/B7B were used to amplify homologs with the same promoter and terminator as *2A* (Appendix 1 Table 1.1). The PCR products were run in a 1% agarose gel and the smear of fragments between 2 – 3 kb was extracted and blunt-end cloned using the Strataclone vector pSC-A-amp/kan and Strataclone Solo Pack competent cells, as above (Table 2.4). A total of 47 white colonies from the set *ψ2A5*, 47 white colonies from the set *2A* and 1 blue colony were screened by PCR for inserts using T7 and T3 primers (Appendix 1 Table 1.1). A total of two PCR products derived from individual clones from set *ψ2A5* and six PCR products derived from individual clones from set *2A* were cleaned

using Sepharose and sequenced by Sanger sequencing. Plasmids were extracted from 17 colonies from set *ψ2A5* and 8 colonies from set *2A*. The inserts of 12 plasmids from set *ψ2A5* were sequenced directly using T3 and T7 primers (Appendix 1 Table 1.1). The rest of the plasmids were subject to PCR to amplify their inserts using T3 and T7 primers (Appendix 1 Table 1.1). All the inserts were sequenced using T3 and T7 primers and set *2A* plasmids were further sequenced using primers that bind to the internal sequence of *2A* (Appendix 1 Table 1.1). Set *ψ2A5* reads were mapped to *ψ2A5* ORF and promoter, whilst set *2A* were mapped to *2A, 2B, ψ2A4* and *ψ2A5*. Three clones were removed from set *2A*, because they only contained sequences of the cloning vector and 1 clone was removed from set *2A* because its origins could not be determined and had very short sequences.

The primers OR2A5' primer F and OR2A3' primer R (Appendix 1 Table 1.1), which bind in the coding region of *2A* from start codon (ATG) and the stop codon (TGA), respectively, were used to interrogate *2A* homologs for the presence of AAG at position 2360 as present in the sequence of d2A2. A Phusion PCR was performed on BAC 7B, and genomic DNA of *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 DNA. The PCR products from the 7B and CfEcp2 reactions were blunt-end cloned into the Stractaclone vector (see above). The transformation of Strataclone competent cells was repeated until all the reaction mixtures were used up. A total of 79 white and one blue colony from both the CfEcp2 and 7B amplifications were screened for cloned inserts with T3 and T7 primers (Appendix 1 Table 1.1). Sanger sequencing was performed using the OR2A3' primer R on 70 PCR products from the CfEcp2 reaction and 66 PCR products from the 7B reaction. The reads were mapped to *2A* and *d2A2* and the presence or absence of AAG at 2360 was scored.

2.32.3 Confirming the absence of *ψd2A4*

To specifically amplify *ψd2A4*, primers hcr9-5 fwd 530 and hcr9-5 rev 1020 were used in a PCR on BAC 7B, and genomic DNA of *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 DNA (Appendix 1 Table 1.1). The PCR product from 7B was Sepharose-cleaned and sequenced by Sanger sequencing using the hcr9-5 fwd 530 and hcr9-5 rev 1020 primers (Appendix 1 Table 1.1). Reads were mapped to *2A* and *d2A2.*

2.32.4 Confirming the presence of *ψ2A5*

To specifically amplify *ψ2A5*, primers OR2AF1 and OR2A4R were used in a PCR on 7B DNA and the product was sequenced by Sanger sequencing using the OR2AF1 and OR2A4R primers (Appendix 1 Table 1.1). Reads were mapped to *ψ2A5.*

2.33 Characterisation of CfEcp2- mutants

In an attempt to clone the gene encoding *Cf-Ecp2*, mutants of *S. pimpinellifolium* CfEcp2 were generated by Thomas et al. (2012, unpublished) (Figure 2.1). The two mutants, named *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, were the products of a transposon tagging-experiment, whereby a transgenic line of *S. lycopersicum* Cf0 which expresses *Ecp2* (35s:*Ecp2*) in a stable manner, was crossed to *S. pimpinellifolium* CfEcp2. This *S. pimpinellifolium* CfEcp2 carries a *Ds* element, *FT33*. *FT33* is genetically-linked to the *OR* locus, and unlinked to the *sAc* element (Jones et al., 1994; Takken et al., 1998).

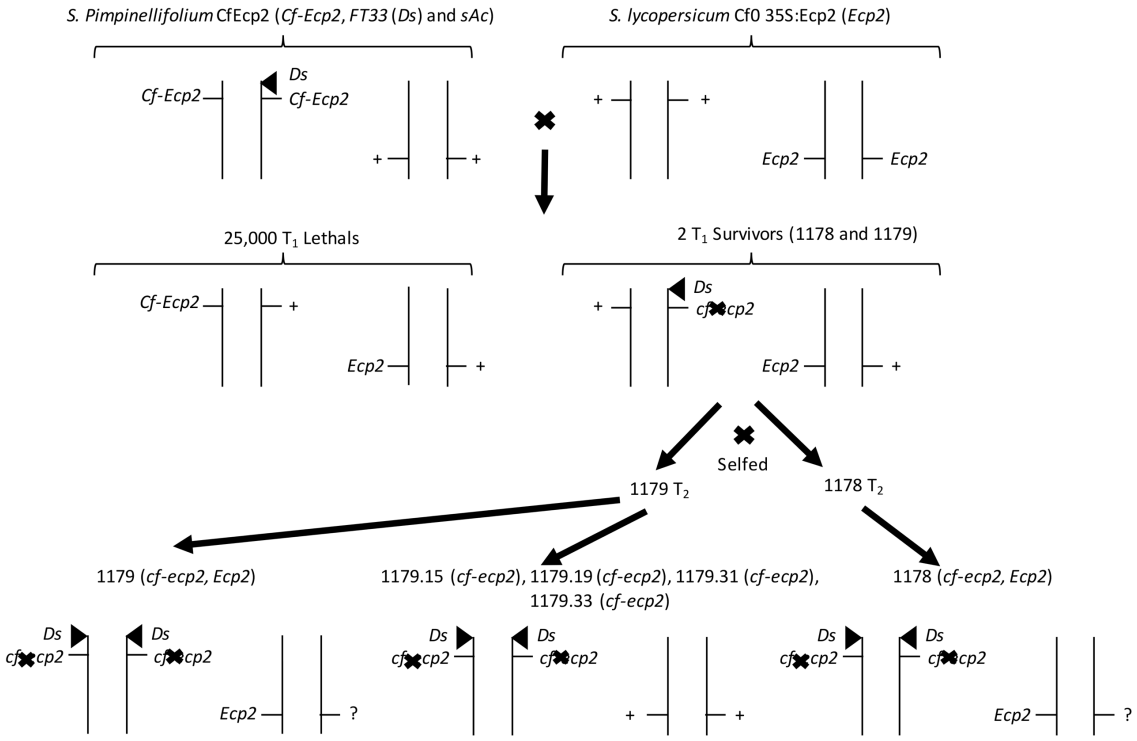


Figure 2.1. Crossing screen to select for mutants at the *Cf-Ecp2* locus.

S. pimpinellifolium CfEcp2, carrying *Cf-Ecp2*, the *Dissociation* element (*Ds*) *FT33* (black triangle) and *stabilised Activator* (*sAc*) element, was crossed to *S. lycopersicum* Cf0 homozygous for the stable transgene 35S:*Ecp2*. Two forms of T_1 progeny resulted; plants that carried a functional copy of *Cf-Ecp2* and *Ecp2* which died, and plants which survived that were heterozygous for the presence of Ecp2 and heterozygous for the loss of *Cf-Ecp2* (i.e. *cf-ecp2*). Two survivors were identified, where *Cf-Ecp2* was subsequently found to have been deleted from the genome, and named 1178 and 1179. These were selfed to generate the T_2. One T_2 plant from 1178 was selected that was homozygous for the presence of the deletion (*cf-ecp2*) and carried *Ecp2*. This plant was named 1178. A similar T_2 plant was selected from the T_2 of 1179 and named 1179. The 1179 T_2 progeny were also screened for plants that were homozygous for the deletion mutant yet lacked *Ecp2*. These four selected plants were named 1179p15, 1179p19, 1179p31 and 1179p33. + = wild type allele at locus, ? = unknown allele at locus.

The progeny generated by this cross were expected to die since they would carry the *Cf-Ecp2* resistance gene and the effector *Ecp2*. Recognition of Ecp2 by Cf-Ecp2 would trigger a HR throughout the plant. However, the progeny were expected to survive if the *Ds* element, activated by the *sAc*, jumped into *Cf-Ecp2* and inactivated it. Out of 25,000 progeny generated there were only two survivors, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179. These lines did not carry the *Ds* element but had lost a defined section of the short arm of chromosome 1.

PCR analysis was carried out on the DNA of *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, alongside *S. lycopersicum* Cf0 and *S. pimpinellifolium* CfEcp2, to determine the presence or absence of genes *2C*, *2B*, *2A*, *ψ2C* and markers CT116, SNP-nn1 and SNPN (Appendix 1 Table 1.1). The products were run on 1% agarose by gel electrophoresis (Gopal 2012 unpublished data). Furthermore, RNA was extracted from the plants *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, and converted to cDNA. PCR analysis was carried out on the cDNA to investigate the expression of *2B* in these plants using primers OR2A/BF1 and OR2BR1 (Appendix 1 Table 1.1).

2.34 MfEcp2 variant testing in CfEcp2 tomato

The ability of *S. pimpinellifolium* CfEcp2 to recognise variants of *M. fijiensis* MfEcp2 homologs was tested. Stocks of MfEcp2 variants were kindly received in *A. tumefaciens* GV3101, from Dr Ioannis Stergiopoulos (Stergiopoulos et al., 2012), and glycerol stocks were made (Table 2.4). The sequence of the clones were confirmed by colony PCR using primers OX10 and N31, followed by Sanger sequencing. For each construct and for the control *C. fulvum* PVX:*Ecp2*, four plants of *S. lycopersicum* Cf0, *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179 were inoculated. Photographs were taken and responses to the variants scored 14 d.p.i..

2.35 Golden Gate cloning of *2A*

To remove the *Bsa*1 site from *2A*, by conversion of A to G at 570 bp, and prepare it for Golden Gate cloning, two Phusion PCRs were completed on *2A* in the Strataclone vector pSC-A-amp/kan (colony 25 from set *2A* cloning of *2A* homologs) (Table 2.4). Primers GGOR2AF1 with GGOR2AR1, and GGOR2AF2 with GGOR2AR2 were used for this purpose (Appendix 1 Table 1.1). This procedure generated two products of 500 and 200 bp respectively, which were extracted from the agarose gel. The following reactions were performed into the destination vector pICH86988 (Table 2.7):

Table 2.7. Golden Gate cloning *OR2A.*

Reaction 1			
Component	**Initial amount**	**Final amount**	**Volume (µL)**
PCR product (2,000 bp)	50 ng/µL	50 ng	1.0
PCR product (500 bp)	50 ng/µL	50 ng	1.0
destination vector	50 ng/µL	50 ng	1.0
*Bsa*I	10,000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U µL	2.5 U	0.8
2x ligation buffer (Promega)	2x	1x	5.0
Water			0.8
			10.0 (Total)
Reaction 2 (Controls)			
Component	**Initial amount**	**Final amount**	**Volume (µL)**
PCR product (2,000 bp)	50 ng/µL	50 ng	1.0
PCR product (500 bp)	50 ng/µL	50 ng	1.0
*Bsa*I	10,000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U/µL	2.5 U	0.8
2x ligation buffer (Promega)	2x	1x	5.0
Water			1.8
			10.0 (Total)
Reaction 3 (Controls)			
Component	**Initial amount**	**Final amount**	**Volume (µL)**
Destination vector	50 ng/µL	50 ng	1.0
*Bsa*I	10000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U/µL	2.5 U	0.8
2x ligation buffer (Promega)	2x	1x	5.0
Water			2.8
			10.0 (Total)
Thermo-cycler	**Temperature (°C)**	**Time (min)**	
	37	30	
	50	5	
	80	5	

A 5 – 10 µL aliquot of each reaction was heat-shock transformed into *E. coli* DH5a cells. There were 73 colonies on the plate from reaction 1 and no colonies from other reactions. Colony PCR was performed on the 73 colonies, with primers GGOR2AF1 and GGOR2AR2 (annealing temperature 45°C, elongation time 3 minutes), and seven colony PCR products were sequenced using Sanger sequencing (Appendix 1 Table 1.1). One colony was chosen for plasmid extraction. The plasmid DNA was eluted in 30 µL water. To confirm that cloning was correct, the plasmid was subjected to PCR with primers GGOR2AF1 and GGOR2AR2, (annealing temperature 45 °C, elongation time 3 minutes), and the product was sequenced using Sanger sequencing (Appendix 1 Table 1.1). This process involved the use of primers, capable of binding across the *2A* sequence (i.e. the

same primers used in 'sequencing of *2A* homologs', Appendix 1 Table 1.1). The plasmid was subjected to PCR with primers pICH86988insetF and pICH86988R1 (annealing temperature 50 °C, elongation time 90 seconds) (Appendix 1 Table 1.1). The PCR product was sequenced using Sanger sequencing with primers pICH86988F and pICH86988R1, and digested overnight at 37 °C with *Sac*I (300 ng plasmid, 1x buffer A, 10 Units *Sac*I, up to 10 µL with water), to show that the insert was not *LacZ* (Appendix 1 Table 1.1).

The clone pICH86988:35S:*2A* was transformed into *A. tumefaciens* GV3101 and Agl1. To confirm the presence of the clones in *A. tumefaciens* GV3101, primers pICH86988F with OR2A670R and OR2A2700F with pICH86988R1, (annealing temperature 50 °C, elongation time 3 minutes), were used in a colony PCR on seven colonies, bridging across the two cloning sites (Appendix 1 Table 1.1). To confirm the presence of clones in *A. tumefaciens* Agl1, colony PCR was performed on three colonies and on a negative control (pICH86988 empty vector DNA). This involved the use of primers pICH86988F with OR2A670R and OR2A2700F with pICH86988R1 (annealing temperature 50 °C, elongation time 2 minutes) (Appendix 1 Table 1.1).

2.36 Transient transformation of 35S:*2A* and 35S:*Ecp2* into *N. paniculata*

Two leaves from one plant from each of the *N. paniculata* accessions; TW99, TW102 and TW99 x TW102 were transiently transformed with *A. tumefaciens* (Table 2.3). The bacterium was carrying 35S:*2A* or 35S:*Ecp2.* The *N. paniculata* was inoculated with each of these bacterial strains and with a 1:1 mixture of *A. tumefaciens* (one strain with 35S:*2A* and one strain with 35S:*Ecp2*). Photographs were taken 6 d.p.i., with and without ultraviolet exposure.

2.37 Stable transformation of 35S:*2A* into *Solanum lycopersicum*

Stable transformation of 35S:*2A* into *S. lycopersicum* Cf0, using *A. tumefaciens* Agl1, was completed by The Transformation Services at The Sainsbury Laboratory. Selection for positive transformation was completed by growing the products on MS media with 300 mg/L kanamycin. The 13 plants that were resistant to kanamycin were transferred to soil growth in containment. The plants were named as follows:

S. lycopersicum Cf0 35S:*2A*p1

S. lycopersicum Cf0 35S:*2A*p2

S. lycopersicum Cf0 35S:*2A*p3

S. lycopersicum Cf0 35S:*2A*p4

S. lycopersicum Cf0 35S:*2A*p5

S. lycopersicum Cf0 35S:*2A*p10

S. lycopersicum Cf0 35S:*2A*p13

S. lycopersicum Cf0 35S:*2A*p14

S. lycopersicum Cf0 35S:*2A*p16

S. lycopersicum Cf0 35S:*2A*p18

S. lycopersicum Cf0 35S:*2A*p24

S. lycopersicum Cf0 35S:*2A*p26

S. lycopersicum Cf0 35S:*2A*p28

2.38 Golden Gate cloning of 35S:*Ecp2* variants

The sequences of *Ecp2* variants were selected, alongside *C. fulvum Avr4* and *Ecp2*, if they met the following criteria (Table 2.8);

- contained the *Ecp2* domain class I,
- contained a predicted SP (determined using signal version 4.1 http://www.cbs.dtu.dk/services/SignalP/)
- came from an important crop pathogen

Table 2.8. Characteristics of selected Ecp2 variants for Golden Gate cloning.

Organism	Host	Protein ID (GenBank)	Class	Length	SP	% identity
Cladosporium fulvum Ecp2	Tomato	CAA78401.1	1	165	Y	100
Cladosporium fulvum Avr4	Tomato	CAA69643.1	N/A	135	Y	11.6
Dothistroma septosporum Ecp2	Conifers	Dotse1_158381	1	167	Y	59.8
Fusarium graminearum Ecp2	Grain cereals	Fgra_06106	1	189	Y	18.5
Fusarium oxysporum Ecp2	Various plants	Foxy_04770	1	151	Y	23.3
Magnaporthe grisea Ecp2	Rice and other important cereals	Mgri_03495T0	1	175	Y	12.5
Mycosphaerella fijiensis Ecp2	Banana	Mfij_52972	1	161	Y	57.7
Zymoseptoria tritici Ecp2	Wheat	Mgra_104404	1	179	Y	25.3
Septoria musiva Ecp2	Hybrid poplar plantations	Sepmu1_146583	1	182	Y	28.6
Verticillium dahliae Ecp2	Many including tomato	Vdah_05725T0	1	221	Y	15.8

Host = crop infected by the pathogen, Length = number of amino acids in protein, SP = signal protein, Y = yes, % identity = percentage identity to *C. fulvum* Ecp2 protein using ClustalW in Geneious R8 software.

The amino acid sequences of the Ecp2 homologs and *C. fulvum* Avr4 were compared using ClustalW. The distance tree was built from that multiple alignment in Geneious, with *C. fulvum* Avr4 as an outgroup, using Jukes cantor genetic distance model, Blosum45 and default bootstrapping options. This used the Jukes-Cantor genetic distance model, Blosum45 global alignment cost matrix and neighbour-joining tree building method. All other parameters were Geneious R8 default settings.

In order to direct the expressed protein to be secreted by the plant into the apoplast for each of the *Ecp2* sequences, the predicted SP was replaced with the PR1a SP from *N. tobaccum*. The additional sequence GGTCTCAAGGT was added to the 5' of the sequence and GCTTAGAGACC was added to the 3' of the sequence. The company Genewiz optimised the completed sequence for expression in *Nicotiana* sp. and synthesised the sequence into the pUC57-kan vector.

To alter the sequences for Golden Gate cloning, Phusion PCR was performed on each construct using specific primers (see Appendix 1 Table 1.1 of primers 'Altering *Ecp2* variants for Golden Gate cloning'). The PCR products were cleaned using Sepharose and glass beads. The PCR products were then cloned into pSC-A-amp/Kan, according to manufacturer's instructions (StrataClone). To determine the presence of an insert, colony PCR was performed on seven white and one blue colony per effector cloned. This involved the use of primers T3 and T7 (Appendix 1 Table 1.1). For those colonies, which gave PCR products of the correct size, the PCR products were sequenced using Sanger sequencing. For each effector in pSC-A-amp/Kan, one plasmid with the confirmed correct sequence was extracted.

The following Golden Gate cloning reactions were performed on each effector in pSC-A-amp/kan to move the effector into the destination vector pICH86988 (Table 2.9). Reaction 1 was performed on all effectors whilst reaction 2 was performed on the Ecp2 effector from *S. musiva* only as a control.

Table 2.9. Golden Gate cloning of Ecp2 effectors.

Reaction 1			
Component	**Initial concentration**	**Final amount**	**Volume (µL)**
Effector in pSC-A-amp/kan	100 ng/µL	50 ng	0.5
Destination vector (pICH86988)	20 ng/µL	50 ng	3.0
*Bsa*I	10,000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U/µL	2.5 U	0.8
5x ligation buffer (Promega)	5x	1x	2.0
Water		µ	3.3
			10.0 (Total)
Reaction 2 (Controls)			
Component	**Initial amount**	**Final amount**	**Volume (µL)**
Effector in pSC-A-amp/kan	100 ng/µL	50 ng	0.5
*Bsa*I	10,000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U/µL	2.5 U	0.8
5x ligation buffer (Promega)	5x	1x	2.0
Water			6.3
			10.0 (Total)
Reaction 3 (Controls)			
Component	**Initial amount**	**Final amount**	**Volume (µL)**
Destination vector (pICH86988)	20 ng/ µL	50 ng	3.0
*Bsa*I	10,000 U/mL	2.5 U	0.4
T4 Ligase (Promega)	3 U/ µL	2.5 U	0.8
5x ligation buffer (Promega)	5x	1x	2.0
Water			5.0
			10.0 (Total)
Thermo-cycler	**Temperature (°C)**	**Time (min)**	
	37	30	
	50	5	
	80	5	

A 5 – 10 µL aliquot of each reaction was heat-shock transformed into *E. coli* DH5a cells. To determine if the effectors had been cloned into pICH86988, colony PCR was performed on seven white colonies and one blue colony from each plate made from Reaction 1, using the primers PR1a (designed to bind to ambiguous sequences: the sequence of PR1a is different in each effector as a consequence of sequence optimisation performed by Genewiz) and pICH86988R1 (Appendix 1 Table 1.1). For each effector, one to three plasmids were extracted since they had given the predicted PCR product size. The plasmids were subjected to PCR with primers pICH86988F and pICH86988R1 (Appendix 1 Table 1.1). The PCR product was then sequenced with primer pICH86988F, using Sanger sequencing (Appendix 1 Table 1.1). One clone for each effector in pICH86988 was selected and the plasmid was transformed into *A. tumefaciens* GV3101. To confirm the presence of clones in *A. tumefaciens* GV3101, colony PCR was performed

on two colonies for each effector. This procedure used primers pICH86988F and pICH86988R1 and the PCR products were then sequenced using the primer pICH86988F with Sanger sequencing (Appendix 1 Table 1.1). Glycerol stocks were made for one colony from each effector showing the correct sequence.

The effectors were transformed into *N. paniculata* accessions TW99, TW102 and the F_1 of these accessions (NP_00021) using Agrobacterium transient transformation. This transformation was repeated three times for each of the effectors. Photographs were taken of the responses at 10 d.p.i..

2.39 Characterisation of 35S:*2A* transformants

In order to determine if *2A* encoded for *Cf-Ecp2* and thus recognised Ecp2, the ability of the transgenic plants to recognise Ecp2 was tested. This testing was completed via a number of different formats; the infiltration of Ecp2 protein, PVX mediated-delivery of Ecp2 and crossing the transgenic lines to *S. lycopersicum* Cf0 plants homozygous for 35S:*Ecp2*. Furthermore, since the 35S:*2A* construct was transformed into *S. lycopersicum* Cf0 in a cassette carrying kanamycin-resistance, it was expected that those lines that were resistant to kanamycin would also carry the 35S:*2A* transgene. The plants were therefore tested for the presence of *2A* by exposing them to kanamycin-selection.

2.39.1 Testing 35S:*2A* transformants with Ecp2 protein

Three months after transplanting in soil, the 13 kanamycin-resistant T_1 *S. lycopersicum* plants (candidates for stable transgenics expressing 35S:*2A*), alongside *S. lycopersicum* Cf0 and *N. paniculata* TW99, were infiltrated with 15 µM Ecp2 protein and 75-fold diluted A4 buffer (1 – 3 leaves per plant). Phenotypes were scored and photographs taken 10 or 17 d.p.i.

In addition, the cotyledons of the following T_2 plants were infiltrated with 15 µM Ecp2 protein and 33-fold diluted A4 buffer 11 days after sowing;

S. lycopersicum Cf0 35S:*2A*p3.1
S. lycopersicum Cf0 35S:*2A*p3.2
S. lycopersicum Cf035S:*2A*p3.3
S. lycopersicum Cf035S:*2A*p3.4
S. lycopersicum Cf035S:*2A*p3.5
S. lycopersicum Cf035S:*2A*p3.6
S. lycopersicum Cf035S:*2A*p3.7
S. lycopersicum Cf035S:*2A*p3.8
S. lycopersicum Cf035S:*2A*p4.1
S. lycopersicum Cf035S:*2A*p4.2
S. lycopersicum Cf035S:*2A*p24.1
S. lycopersicum Cf035S:*2A*p24.2
S. lycopersicum Cf035S:*2A*p24.3
S. lycopersicum Cf035S:*2A*p24.4

Two *S. pimpinellifolium* CfEcp2 plants, two *S. lycopersicum* Cf0 plants and two *S. lycopersicum* 35S:*Ecp2* plants, were infiltrated alongside those plants listed above. Phenotypes were scored and photographs were taken 6 d.p.i.

2.39.2 PVX:*Ecp2* and PVX:Avr4 infiltrations of 35S:*2A* transformants

To test for specific recognition of PVX:*Ecp2* in the candidate *S. lycopersicum* Cf0 35S:*2A* transformants, PVX:*Ecp2* and/or PVX:*Avr4* were inoculated into the cotyledons of the following lines (Table 2.10), and phenotypes were scored 14 or 24 d.p.i.

Table 2.10. Plant lines inoculated with PVX:*Ecp2* and/or PVX:*Avr4.*

	Total plants inoculated	
Plant line	PVX:*Ecp2*	PVX:*AVR4*
S. lycopersicum Cf0	22	4
S. pimpinellifolium CfEcp2	18	6
S. pimpinellifolium 1178	20	6
S. pimpinellifolium 1179	18	5
S. lycopersicum Cf0 35S:*2A*p1	2	0
S. lycopersicum Cf0 35S:*2A*p2	9	0
S. lycopersicum Cf0 35S:*2A*p3	2	0
S. lycopersicum Cf0 35S:*2A*p5	7	0
S. lycopersicum Cf0 35S:*2A*p13	1	0
S. lycopersicum Cf0 35S:*2A*p14	3	0
S. lycopersicum Cf0 35S:*2A*p16	4	0
S. lycopersicum Cf0 35S:*2A*p18	6	0
S. lycopersicum Cf0 35S:*2A*p22	4	0
S. lycopersicum Cf0 35S:*2A*p24	7	0
S. lycopersicum Cf0 35S:*2A*p28	4	0
S. lycopersicum Cf0 35S:*2A*p2.1	8	0
S. lycopersicum Cf0 35S:*2A*p2.2	6	0
S. lycopersicum Cf0 35S:*2A*p2.3	7	0
S. lycopersicum Cf0 35S:*2A*p2.4	7	0
S. lycopersicum Cf035S:*2A*p2.5	4	0
S. lycopersicum Cf035S:*2A*p2.6	8	0
S. lycopersicum Cf035S:*2A*p3.1	8	0
S. lycopersicum Cf035S:*2A*p3.2	8	0
S. lycopersicum Cf035S:*2A*p3.3	5	0
S. lycopersicum Cf035S:*2A*p3.5	5	0
S. lycopersicum Cf035S:*2A*p3.6	7	0
S. lycopersicum Cf035S:*2A*p3.8	18	10
S. lycopersicum Cf035S:*2A*p3.9	6	0
S. lycopersicum Cf035S:*2A*p3.10	8	0
S. lycopersicum Cf035S:*2A*p3.11	8	0
S. lycopersicum Cf035S:*2A*p4.1	4	0
S. lycopersicum Cf035S:*2A*p4.2	2	0
S. lycopersicum Cf035S:*2A*p5.1	8	0
S. lycopersicum Cf035S:*2A*p5.2	7	0
S. lycopersicum Cf035S:*2A*p5.3	8	0
S. lycopersicum Cf035S:*2A*p18.1	4	0
S. lycopersicum Cf035S:*2A*p24.1	2	0
S. lycopersicum Cf035S:*2A*p24.2	6	0
S. lycopersicum Cf035S:*2A*p24.3	7	0
S. lycopersicum Cf035S:*2A*p24.4	16	10
S. lycopersicum Cf035S:*2A*p24.5	2	0

PVX= Potato Virus X

2.39.3 Kanamycin selection of 35S:*2A* transformants

To determine the presence of 35S:*2A* in T_2 and T_3 families, plants were exposed to kanamycin-selection. *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* 1179 and T_2 seeds from the following T_1 lines were sown on 300 mg/mL kanamycin:

S. lycopersicum Cf0 35S:*2A*p1,
S. lycopersicum Cf0 35S:*2A* p2,
S. lycopersicum Cf0 35S:*2A* p3,
S. lycopersicum Cf0 35S:*2A* p5,
S. lycopersicum Cf0 35S:*2A* p14,
S. lycopersicum Cf0 35S:*2A* p16,
S. lycopersicum Cf0 35S:*2A* p18,
S. lycopersicum Cf0 35S:*2A* p22.

Twenty seeds were sown per plant line at 10 seeds per tub. From the seed sown, the following kanamycin-resistant T_2 plants were transplanted into soil 16 days post sowing (d.p.s.):

S. lycopersicum Cf0 35S:*2A* p2.1
S. lycopersicum Cf0 35S:*2A* p2.2
S. lycopersicum Cf0 35S:*2A* p2.3
S. lycopersicum Cf0 35S:*2A* p2.4
S. lycopersicum Cf0 35S:*2A* p2.5
S. lycopersicum Cf0 35S:*2A* p2.6
S. lycopersicum Cf0 35S:*2A* p3.9
S. lycopersicum Cf0 35S:*2A* p3.10
S. lycopersicum Cf0 35S:*2A* p3.11
S. lycopersicum Cf0 35S:*2A* p5.1
S. lycopersicum Cf0 35S:*2A* p5.2
S. lycopersicum Cf0 35S:*2A* p5.3
S. lycopersicum Cf0 35S:*2A* p18.1

These plants were then crossed with *S. lycopersicum* Cf0 35S:*Ecp2*, *S. lycopersicum* Cf0 and *S. pimpinellifolium* CfEcp2 plants (Table 2.3).

S. lycopersicum Cf0, *S. pimpinelifolium* 1179 and the T_3 self-seed plants from the T_2 lines *S. lycopersicum* Cf0 35S:*2A*p3.8 and *S. lycopersicum* Cf0 35S:*2A*p24.4 were sown on MS media, with and without 300 mg/mL kanamycin. Eight seeds were sown for each combination except for the T_2 lines on kanamycin where 16 seeds were sown (at 8 seeds per tub). Resistant or susceptible phenotypes were scored 24 d.p.s. and the percentage germination for each plant line was calculated.

2.39.4 Seedling lethal phenotype scoring of crosses between 35S:*2A* and 35S:*Ecp2* stable transgenic plants

The following lines were sown and phenotypes scored 15 to 30 d.p.s. (Table 2.11). For some of the transgenic lines (indicated in Table 2.11), crosses were made in both directions. Since there was no difference between the phenotypes in the progeny, the results for crosses in either direction were combined.

Table 2.11. Plant lines with phenotypes scored 15 to 30 d.p.s.

Plant line	Total seeds sown for scoring
S. lycopersicum Cf0	37
[1]*S. lycopersicum* Cf0 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* Cf0	42
S. lycopersicum Cf0 x *S. pimpinellifolium CfEcp2*	12
S. pimpinellifolium CfEcp2	37
[1]*S. pimpinellifolium* CfEcp2 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. pimpinellifolium* CfEcp2	21
S. lycopersicum 35S:*Ecp2*	12
S. lycopersicum 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p1	26
[1]*S. lycopersicum* 35S:*2A*p2 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* 35S:2Ap2	40
[1]*S. lycopersicum* 35S:*2A*p3 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p3	40
S. lycopersicum 35S:*2A*p4 x *S. lycopersicum* 35S:*Ecp2*	30
[1]*S. lycopersicum* 35S:*2A*p5 x *S. lycopersicum* 35S:*Ecp2* and *S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p5	50
S. lycopersicum 35S:*2A*p10 x *S. lycopersicum* 35S:*Ecp2*	30
S. lycopersicum 35S:*2A*p14 x *S. lycopersicum* 35S:*Ecp2*	25
S. lycopersicum 35S:*2A*p16 x *S. lycopersicum* 35S:*Ecp2*	40
S. lycopersicum 35S:*2A*p18 x *S. lycopersicum* 35S:*Ecp2*	40
S. lycopersicum 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p24	18
S. lycopersicum 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p28	13
S. lycopersicum 35s:*2A*p3.8	24
S. lycopersicum 35S:*2A*p3.8 x *S. lycopersicum* 35S:*Ecp2*	8
S. lycopersicum 35S:*2A*p24.4	24
[1]*S. lycopersicum* 35S:*2A*p24.4 x *S. lycopersicum* 35S:*Ecp2 and S. lycopersicum* 35S:*Ecp2* x *S. lycopersicum* 35S:*2A*p24.4	40
S. lycopersicum 35S:*2A*p24.4.2D	12
S. lycopersicum 35S:*2A*p24.4.2D x *S. lycopersicum* Cf0	24
S. lycopersicum 35S:*2A*p24.4.2D x *S. lycopersicum* 35S:*Ecp2*	24
S. lycopersicum 35S:*2A*p24.4.4D	24
S. lycopersicum 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0	24
S. lycopersicum 35S:*2A*p24.4.4D x *S. lycopersicum* 35S:*Ecp2*	24
S. lycopersicum 35S:*2A*p24.5.5D	24
S. lycopersicum 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0	24
S. lycopersicum 35S:*2A*p24.5.5D x *S. lycopersicum* 35S:Ecp2	24

[1]Crosses in either direction used, and due to no difference in results, the results were combined. d.p.s = days post sowing.

2.39.5 Integration of 35S:*2A* into a *S. pimpinelifolium* background

A CTAB DNA extraction was completed on 39 F_2 plants from the selfed *S. pimpinellifolium* 1179 F_1 (Figure 2.1). The DNA was eluted in 30 µL water, and those plants containing *Ecp2* were identified by PCR analysis and discarded (Appendix 1 Table 1.1). The remaining DNA samples were analysed for the presence of *2A*, and for the marker CT116 (Table 2.5 and Appendix 1 Table 1.1). Nine plants, which were homozygous for the CT116 marker from the *S. pimpinellifolium* CfEcp2 parent, were retained and a second DNA extraction was performed on them. The DNA was eluted in 40 µL water. The nine samples were analysed with the markers SNPN and TG236 (Table 2.5 and Appendix 1 Table 1.1). Four 1179 plants (1179p15, 1179p19, 1179p31, and 1179p33), were found to be homozygous for the *S. pimpinellifolium* CfEcp2 allele at both markers (Figure 2.1). To reconfirm that these plants were homozygous for the *S. pimpinellifolium* CfEcp2 parent allele, DNA from these plants was re-extracted analysed with the marker TG24 (Table 2.5 and Appendix 1 Table 1.1).

Plants 1179p15, 1179p19, 1179p31 and 1179p33 were crossed with T_3 lines from 35S:*2A*p3 and 35S:*2A*p24, and with controls (Table 2.3). The resulting plants were inoculated with PVX:*Ecp2* or not inoculated with PVX:*Ecp2* and phenotypes were scored 19 d.p.i. (Table 2.12).

Table 2.12. Plant lines inoculated with or without PVX:*Ecp2*.

	Number of plants inoculated (+) or not inoculated (-) with PVX:*Ecp2*	
Plant Line	+	-
S. lycopersicum Cf0	6	4
S. pimpinellifolium CfEcp2	6	6
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0	12	9
S. pimpinellifolium 1179p15	6	5
S. pimpinellifolium 1179p19	6	4
S. pimpinellifolium 1179p31	5	4
S. pimpinellifolium 1179p33	6	6
S. pimpinellifolium CfEcp2 x *S. pimpinellifolium* 1179p15	12	11
S. lycopersicum 35S:*2A*p3.8.1	5	2
S. lycopersicum 35S:*2A*p3.8.1x *S. lycopersicum* Cf0	9	5
S. lycopersicum 35S:*2A*p3.8.1x *S. pimpinellifolium* 1179p19	22	9
S. lycopersicum 35S:*2A*p24.4.2	12	6
S. pimpinellifolium 1179p31 x *S. lycopersicum* 35S:*2A*p24.4.2	12	9
S. lycopersicum 35S:*2A*p24.4.3	10	3
S. lycopersicum 35S:*2A*p24.4.3x *S. pimpinellifolium* 1179p15	12	9
S. lycopersicum 35S:*2A*p24.4.4	12	6
S. lycopersicum 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19	12	11
S. lycopersicum 35S:*2A*p24.5.3	5	0
S. lycopersicum Cf0 x *S. lycopersicum* 35S:*2A*p24.5.3	12	11
S. lycopersicum 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19	12	11
S. lycopersicum 35S:*2A*p24.5.4	12	9
S. lycopersicum 35S:*2A*p24.5.4x *S. pimpinellifolium* 1179p15	9	2

PVX = Potato Virus X

2.39.6 Analysis of DNA and RNA from stable transformants

2.39.6.i DNA and RNA from Ecp2 protein infiltrated plants

To confirm that the T_2 plants infiltrated with Ecp2 protein (see section 2.39.1) were segregating for the presence and expression of the transgene *2A*, the DNA and RNA of the following T_2 plants were extracted;

S. lycopersicum Cf0 35S:*2A*p3.1
S. lycopersicum Cf0 35S:*2A*p3.2
S. lycopersicum Cf0 35S:*2A*p3.3
S. lycopersicum Cf0 35S:*2A*p3.4
S. lycopersicum Cf0 35S:*2A*p3.5
S. lycopersicum Cf0 35S:*2A*p3.6
S. lycopersicum Cf0 35S:*2A*p3.7
S. lycopersicum Cf0 35S:*2A*p3.8
S. lycopersicum Cf0 35S:*2A*p4.1
S. lycopersicum Cf0 35S:*2A*p4.2
S. lycopersicum Cf0 35S:*2A*p24.1
S. lycopersicum Cf0 35S:*2A*p24.2
S. lycopersicum Cf0 35S:*2A*p24.3
S. lycopersicum Cf0 35S:*2A*p24.4

In addition, DNA and RNA were also extracted from two *S. pimpinellifolium* CfEcp2 plants, two *S. lycopersicum* Cf0 plants and two *S. lycopersicum* 35S:*Ecp2* plants (previously infiltrated with 15 µM Ecp2 protein and 33x diluted A4 buffer, see section 2.39.1). PCR was performed on the cDNA to amplify *2A*, using primers OR2A/BF1 and OR2AR1 (Appendix 1 Table 1.1). The presence of cDNA was confirmed by PCR for the house-keeping gene *EFa1* with primers EFa1F and EFa1R (Appendix 1 Table 1.1). The presence of DNA was confirmed by gel electrophoresis on 1 µL of each of the DNA samples. The DNA was subjected to PCR to amplify *2A* and 35S:*2A* using primers OR2A/BF1 with OR2AR1 and 35spromoterF with OR2AR1, respectively (Appendix 1 Table 1.1). The 35S:*2A* PCR products from 35S:*2A*p3.7 and 35S:*2A*p24.6 were subjected to Sanger sequencing using primers 35SpromoterF and OR2AR1 (Appendix 1 Table 1.1).

2.39.6.ii RNA analysis from PVX:*Ecp2* inoculated plants

RNA was extracted from the following plants inoculated with PVX:*Ecp2* (see section 2.39.2); three *S. lycopersicum* Cf0, three *S. pimpinellifolium* CfEcp2, three *S. lycopersicum* Cf0 35S:*2A*p24 with partial HR, three *S. lycopersicum* Cf0 35S:*2A*p24 with no HR and no vein clearing and four *S. lycopersicum* Cf0 35S:*2A*p16 with no HR and no vein clearing. The RNA was extracted from the tissue using RNeasy, treated with DNase and converted into cDNA. PCR was performed on the cDNA to test for the expression of *2A, Ecp2*, PVX replicase and PVX25K (Appendix 1 Table 1.1). This involved the use of primers OR2A/BF1 with OR2AR1, PR1aSPF with Ecp2R, PVX_replicaseF with PVX_replicaseR and PVX25KF with PVX25KR, respectively (Appendix 1 Table 1.1).

2.39.6.iii Analysis of DNA from plants the showing developmental phenotype

DNA was extracted from seven *S. lycopersicum* Cf0 35S:*2A*p2 x *S. lycopersicum* Cf0 35S:*Ecp2* showing the developmental phenotype. In addition, DNA was also extracted from four *S. lycopersicum* Cf0 35S:*2A*p2 x *S. lycopersicum* Cf0 35S:*Ecp2* and five *S. lycopersicum* Cf0 35S:*2A*p24 x *S. lycopersicum* Cf0 35S:*Ecp2* plants, showing normal phenotypes. PCR was performed to amplify 35S:*2A*, using 35SpromoterF and OR2AR1 primers (Appendix 1 Table 1.1).

2.39.6.iv DNA from crosses between lines stably expressing 35S:*2A* and 35S:*Ecp2*

DNA was extracted from the following plants:

Plant species/identification	Number of plants used for the procedure
S. lycopersicum Cf0	5
S. pimpinellifolium CfEcp2	5
S. lycopersicum 35S:*Ecp2*	3
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0	3
S. lycopersicum Cf0 x *S. lycopersicum* Cf0 35S:*Ecp2*	5
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2*	5
S. lycopersicum Cf0 35S:*2A*p3.8	9
S. lycopersicum Cf0 35S:*2A*p24.4.2D	3
S. lycopersicum Cf0 35S:*2A*p24.2D x *S. lycopersicum* Cf0	3
S. lycopersicum Cf0 35S:*2A*p24.2D x *S. lycopersicum* Cf0 35S:*Ecp2*	2
S. lycopersicum Cf0 35S:*2A*p24.4 D	12
S. lycopersicum Cf0 35S:*2A*p24.4 D x *S. lycopersicum* Cf0 35S:*Ecp2*	10
S. lycopersicum Cf0 35S:*2A*p24.4.4D	3
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0	3
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0 35S:*Ecp2*	16
S. lycopersicum Cf0 35S:*2A*p24.5.5D	2
S*. lycopersicum* Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0	2
S. lycopersicum Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0 35S:*Ecp2*.	3

PCR analysis was completed on the DNA to check for its presence via amplification of the house-keeping gene eEF1Art and for the presence and absence of 35S:*2A* (using primers 35SpromoterF and 35S:OR2AR) (Appendix 1 Table 1.1).

2.39.6.v DNA from plants with 35S:*2A* integrated into the *S. pimpinellifolium* 1179 line without Ecp2

DNA was extracted from the following plants which had no response to PVX:*Ecp2* infiltration:

Plant species/identification	Number of plants used for the procedure
S. lycopersicum Cf0	2
S. pimpinellifolium CfEcp2 (not inoculated with PVX:*Ecp2*, but of the same age)	2
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0	3
S. pimpinellifolium 1179p19	2
S. lycopersicum 35S:*2A*p3.8.1 (which were pooled)	2
S. lycopersicum 35S:*2A*p3.8.1 x *S. pimpinellifolium* 1179p19	3
S. lycopersicum 35S:*2A*p24.4.4	2
S. lycopersicum 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19	3
S. lycopersicum 35S:*2A*p24.5.3	2
S. lycopersicum 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19	3

PCR was completed to amplify 35S:*2A*, using primers 35SpromoterF and OR2AR1 (Appendix 1 Table 1.1). PCR was also completed to show the presence of DNA, by

amplification of the house-keeping gene eEF1Art, and to demonstrate the parental origin of the *OR* locus in each plant by amplification of the marker CT116 (Table 2.5 and Appendix 1 Table 1.1).

2.40 *N. paniculata* genetic map

2.40.1 Generating a mapping population segregating for the recognition of Ecp2 in *N. paniculata*

N. paniculata accession TW99 (original GRIN accession PI555550, syn. TW99) responds with a HR upon transient *Agrobacterium*-mediated delivery of 35S:*Ecp2* (de Kock et al., 2004; Lauge et al., 2000). On the other hand, the *N. paniculata* accession TW102 (original GRIN accession PI555545, syn. TW102) gives no response to 35S:*Ecp2* inoculation (de Kock et al., 2004; Lauge et al., 2000). To map the genetic component conferring the recognition of Ecp2 in the TW99 accession two F_1 families (NP_00018 and NP_00021) were generated from crosses in both directions between *N. paniculata* TW99 and *N. paniculata* TW102 (Harder, 2012). Each F_1 family was selfed, generating two F_2 populations (NP_00036 and NP_00038) (Harder, 2012). *Agrobacterium*-mediated transient transformation was used to deliver 35S:*Ecp2* into 44 – 48 individuals from each F_2 family, alongside six of each of the F_1 families and the parents TW99 and TW102 (Harder, 2012). Responses were scored 2 to 4 d.p.i. (Harder, 2012). The responses in the F_2 were scored as either no response (0), partial HR manifested as partial necrosis (+) or HR with confluent death of inoculated section (++) (Figure 2.2).

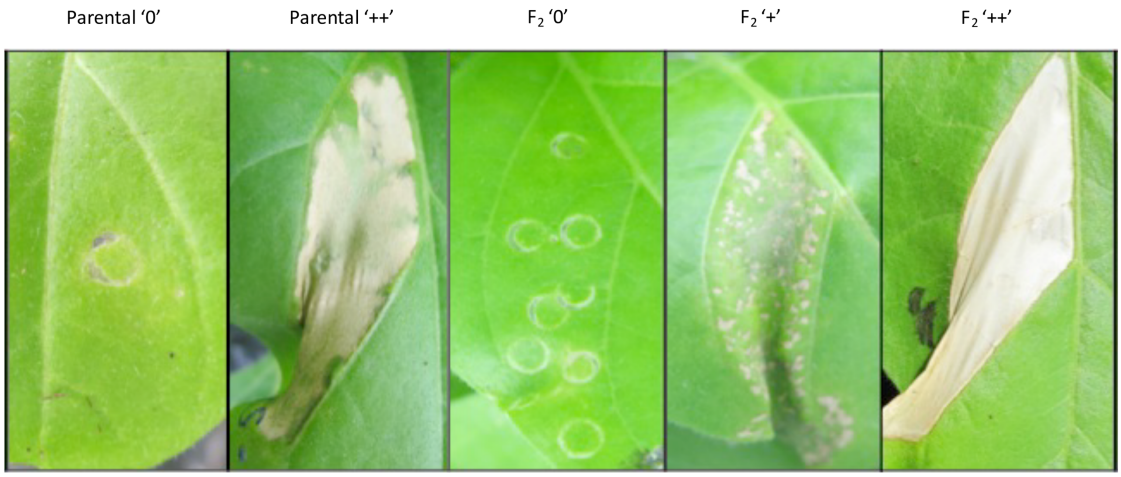


Figure 2.2. Scoring response to *Agrobacterium*-mediated delivery of 35S:*Ecp2* in *N. paniculata* (Harder, 2012). 0 = no response, + = partial hypersensitive response (HR) manifested as partial necrosis, ++ = HR with confluent death of inoculated section.

DNA was extracted from the accessions TW99, TW102, the F_1 (NP_00018 and NP_00021) and each of the F_2 plants that were phenotyped (Harder, 2012). The concentration of each DNA sample was quantified using the Picogreen® assay (Invitrogen), according to manufacturer's instructions (Westergaard, 2012).

2.40.2 Verification of observed F_2 phenotypes

The F_2 phenotypes that showed either 0 or ++ were predicted to be homozygous for the absence or presence of *Cf-Ecp2*, respectively. To confirm this a subset of F_2 phenotypes were validated by scoring the F_3 families which were derived from them. *Agrobacterium*-mediated transient transformation of 35S:*Ecp2* was performed on 8 to 14 individuals from 10 F_3 families originating from F_2 plants that scored 0 and 11 individuals from 10 F_3 families originating from F_2 plants that scored ++ (Westergaard, 2012).

2.40.3 Sequenom marker generation

A total of 96 single copy conserved orthologous set II (COSII) markers were randomly selected that spanned the genetic maps of *N. acuminata* and *N. tomentosiformis* (Wu et al., 2010). Primers for each of these markers were used (Appendix 1 Table 1.1) (Wu et al., 2010) in PCRs on the parental accessions TW99 and TW102. For those markers which gave distinct PCR products, the products were sequenced using Sanger sequencing (TGAC, Norwich, UK and GATC, Constance, Germany). SNPs were searched for between the

sequence of the PCR products from the two accessions using the SeqMan module of Lasergene (Madison, USA). For each COSII marker where SNPs were identified, one to two SNPs were used to generate Sequenom markers (Westergaard, 2012).

Sequenom markers were generated using the default settings of the Sequenom Assay Design Suit 1.0. A total of 75 Sequenom markers were generated that covered 47 COSII markers. These markers were then run on the parental (TW99 and TW102), F_1 (NP_00018 and NP_00021) and F_2 DNA using the MassArray Analyser at the Genomic Technologies Facility, Iowa State University, USA. This used Single Base Extension (SBE) with IPLEX chemistry. The resulting molecular weight, created by single base extension, depends on which allele the DNA contains. This molecular weight was measured by the MassArray Analyser 4. The results were curated by combining data from Sequenom markers that corresponded to one COSII marker, removing monomorphic or dominant markers and removing markers which completely lacked data. This resulted in 39 non-redundant markers which had high quality calls with few missing data points.

2.40.4 Generation of *N. paniculata* map

The 39 Sequenom markers were combined with 42 Sequenom markers from Westergaard (2012). These 81 markers were used to generate a genetic map for *N. paniculata* using Map Manager QTX v 0.30 with default settings (Manly et al., 2001). The linkage groups were orientated to each other and into a set of 12 chromosomes using information from *N. tomentosiformis* and *N. acuminata* genetic maps (Wu et al., 2010).

2.40.5 Linkage analysis of *N. paniculata* markers with Ecp2 recognition

The genotypes for the Sequenom markers, for those individuals that showed a 0 phenotype in response to 35S:*Ecp2,* were analysed with the Chi-Squared test. For each marker the observed number of individuals with each genotype was compared to the expected number of individuals with each genotype if the marker was segregating in a 1:2:1 ratio. This method identified markers that were not segregating in a 1:2:1 ratio and that were significantly different from the expected segregation. The direction in which the marker was significantly different from expected was recorded as loss of heterozygocity, linkage to *Cf-Ecp2* or linkage to *cf-ecp2*.

Quantitative trait locus (QTL) analysis was performed using the Rqtl (v1.33-7) package in R (https://www.r-project.org/)(Broman et al., 2003). First, genotype frequencies for missing data were simulated using sim.geno with the following parameters: the Kosambi map function, 128 simulation replications, a maximum distance of 2 cM between simulated genotype positions, and an error probability of 0.01. Next, interval mapping (scanone) was used to identify potential linkage between genetic loci and the ability to recognise Ecp2. Parameters used included expectation maximization (EM) function and a binary model. Experiment-wide threshold (EWT) were calculated using 1,000 permutations using a threshold of $\alpha < 0.05$.

Chapter 3

Genetic mapping, physical resolution and molecular characterisation of the *Cf-Ecp2* locus in currant tomato (*S. pimpinellifolium*)

3.1 Introduction

The application of sequencing methods to resolve complex repetitive regions of DNA has been transformed during recent years. The first entire genome to be sequenced was that of Bacteriophage phiX174, using the capillary-based Sanger sequencing method (Sanger et al., 1977). Since that time, sequencing technologies have evolved with a dramatic decrease in both the financial and time costs involved (Schatz, 2015). Successful application of sequencing methods requires an understanding of biology, data management and data manipulation (Schatz, 2015). Without such knowledge and related skills, technical artefacts and misinterpretation of data may arise.

Currently, short-read sequencing technologies such as Illumina and 454-pyrosequencing, generate 75-500 bp reads. These reads are clonally amplified and produce highly accurate sequences (Bentley et al., 2008; Reuter et al., 2015; Siqueira et al., 2012). However, these methods are not sufficient to resolve highly repetitive sequences in plant and animal genomes, including transposons, rRNA clusters, satellites (micro, mini and macro) and rapidly evolving gene families e.g. cancer gene loci, antibiotic gene clusters and complex plant *R* gene loci (Ashton et al., 2015; Jain et al., 2015; Snyder et al., 2010). In addition, a bias can be introduced as a result of the method of clonal PCR amplification applied to each read, prior to sequencing (Reuter et al., 2015).

For long repetitive sequences, the SMRT technology, PacBio, has proven useful. Hairpin adapters are ligated to each molecule of DNA. The sequence of this circular-capped template can then be read many times by a strand-displacing DNA polymerase via a method known as circular consensus sequencing (CCS) (Travers et al., 2010). The average read length is currently >14 kb. However, the average error rate of PacBio is approximately 11% (Reuter et al., 2015). Following the application of CCS, shorter reads can be read many more times, producing more accurate sequences (Koren et al., 2012).

These shorter, accurate sequences can be used to curate the longer, more error-prone, sequences in the HGAP pipeline (Chin et al., 2013).

A new single molecule sequencing platform, MinION, introduced in 2014 by Oxford Nanopore technologies, produces much longer reads of tens of kb (Jain et al., 2015; Quick et al., 2014). The length of the reads is limited only by the input length of DNA (Jain et al., 2015). The sequence of the DNA molecule is read as it passes through a bio-pore in a synthetic membrane. A 5-mer of DNA is within the pore at any one time and as the molecule transitions one base at a time through the pore, driven by a motor protein attached to the adaptor, changes in current are detected and translated into the DNA sequence. Following the annealing of a hairpin adaptor at the end of the DNA molecule, both strands of the molecule can be read (Quick et al., 2014). Such reads are called 2D reads.

Those reads that are read only on one strand (template without complement) are known as 1D reads. The MinION sequences are highly error-prone, 2D reads less so than 1D reads, and are most commonly used as a scaffold for determining repeats or for the assembly of Illumina sequences (Ashton et al., 2015; Jain et al., 2015; Quick et al., 2014). *De novo* assembly of the 4.6 Mb whole genome of Escherichia coli K-12 MG1655 has also been achieved using only MinION sequences (Loman et al., 2015). Continued advances in this technology are expected to improve the assembly of multiple repetitive sequences, for which it has not so far been possible to assemble.

Plant *R* genes enable the recognition of specific pathogen effectors (Jones and Dangl, 2006). This recognition results in a response that inhibits the growth of pathogens on the plant, often including a HR (Jones and Dangl, 2006). *R* genes often occur in clusters. Such clusters can carry multiple *R* genes of the same or different structural types (Andolfo et al., 2013). *R* gene clusters can encode for genes that work together to recognise an effector (Martin et al., 1993; Narusaka et al., 2009a; Narusaka et al., 2009b; Salmeron et al., 1996). In addition, clusters can include *R* genes that recognise different effectors from the same pathogen (Takken et al., 1998; Takken et al., 1999).

There are inherent problems associated with the sequencing of *R* genes. *R* genes of the same structural type evolve by tandem duplication, followed by sequence diversification

to generate clusters of *R* gene homologs with high sequence similarity (Botella et al., 1998; Huang et al., 2004; Huang et al., 2005; Li et al., 2011; Michelmore and Meyers, 1998; Parniske et al., 1999). As observed at the *Cf-4*/*Cf-9* locus, *R*-gene variation between sequence-related homologues is mainly found within the solvent exposed amino acid residues of the LRR domain, which is hypothesised to mediate protein-protein interactions (Parniske et al., 1997; Thomas et al., 1997).

The clustering of *R* genes and R gene analogs has been observed in genome wide analysis of many species including, for instance, Arabidopsis, potato, tomato, lettuce and rice (Andolfo et al., 2013; Jupe et al., 2012; McHale et al., 2009; Meyers et al., 2003; Monosi et al., 2004). Within the lettuce cultivar Diana, a 3.5 Mb locus encompasses 24 *NB-LRR* homologs (Meyers et al., 1998). Inside this, highly complex clusters of *NB-LRRs* is the *Dm3* gene, which encodes for resistance to the oomycete downy mildew fungus *Bremia lactucae* (Meyers et al., 1998).

R gene haplotypes often differ between different species belonging to the same family due to the lineage specific fast evolution of *R* gene loci (Michelmore and Meyers, 1998; Parniske et al., 1997; Parniske and Jones, 1999). For example, four distinct haplotypes have been identified at the tomato *MW* locus containing either one or five *Hcr9s*, depending on the species of origin. Even at the species level, *R* gene haplotypes may display significant variation between different accessions. For example, the *R* gene analogues at the *RPP5* locus of *A. thaliana* range from eight *NB-LRR* homologs in the accession Columbia (Col-0) to ten in the accession Landsberg *erecta* (L*er*) (Noel et al., 1999). The spectrum of *R* genes and *R* gene analogues in one plant accession is therefore not representative of the assortment of *R* genes throughout the plant species. Due to this copy number and sequence divergence, it can be difficult to establish clear orthogonal relationships between the *R* genes from the same species. Therefore, the exact composition of *R* genes of one accession cannot be inferred from a gold standard reference sequence of a species. Furthermore, due to intensive breeding of crop plants, many *R* genes have been lost from cultivated varieties (Doebley et al., 2006; Sato et al., 2012). *R* genes are, however, often identified in wild relatives where there is no reference genome sequence or even a genetic map.

Due to the complexity of *R* gene loci, methods such as re-sequencing of the locus within the specific resistant accession and transposon tagging have been employed after genetic fine-mapping in order to clone the *R* gene of interest. Cosmid or BAC libraries, containing DNA from the specific resistant parent, can be screened for flanking markers or homologs of the *R* gene to identify sections of DNA that originate from this locus. Such stretches of DNA can be sequenced and assembled into contigs, to generate a physical map of the locus. For example, three candidate *Cf-4* genes have been identified by generating recombinants between the haplotypes *Cf-4* and *Cf-9*, ruling out those genes resulting in disease sensitive plants (Thomas et al., 1997). Transformation of overlapping cosmids carrying combinations of the candidate *Cf-4* genes identified *Hcr9-4D* as *Cf-4* (Thomas et al., 1997). As another example, screening of cosmid and phage libraries of *A. thaliana* Wassilewskija (Ws-0) DNA with restriction fragment length polymorphism (RFLP) markers, enabled construction of a physical map of the *RPP1* locus and ultimately the cloning of three homologs of the *RPP1* gene (*RPP1*, *RPP10* and *RPP14*) (Botella et al., 1998). Each of the homologs had different resistance specificities to downy mildew fungus *Hyaloperonospora parasitica* isolates (Botella et al., 1998).

In the case of *Cf-9*, the tomato *R* gene encoding for *C. fulvum* Avr9 recognition was isolated by transposon tagging with a *Ds* element (Jones et al., 1994). Similarly, the maize *R* gene, *Rp1-D*, encoding rust resistance, was cloned by two independent transposon tagging experiments using *Ds* and *mutator* (Collins et al., 1999). The transposon tagging method involves crossing a plant with a *Ds* element (genetically linked to the locus encoding resistance) with a plant carrying a *sAc* transposase element (Jones et al., 1994). The *sAc* can activate the *Ds* which will then jump into genetically-linked locations and inactive genes in the F_1 generation (Jones et al., 1994). The F_1 plants can then be crossed with a plant which is over-expressing the corresponding effector induced by the 35S promoter (Jones et al., 1994). If the *R* gene is not inactivated by the *Ds* element then the progeny will show a SLP and die (Jones et al., 1994). If the *R* gene is inactivated and thus tagged by the *Ds* element jumping into it, the progeny will survive (Jones et al., 1994). This enables the identification of tagged *R* genes which can be confirmed by complementation experiments in a non-resistant background e.g. *S. lycopersicum* Cf0.

The fungal pathogen *C. fulvum* infects tomato and grows in a completely apoplastic manner (Thomma et al., 2005). All tomato *C. fulvum R* genes cloned to date encode RLPs.

Consistent with the apoplastic lifestyle of *C. fulvum*, these RLPs are located at the surface of the plant cell, where they monitor the apoplast (Wulff et al., 2009a). These RLPs have been grouped into two classes based on their sequence: *Hcr9s* and *Hcr2s* (Wulff et al., 2009a).

Hcr9s lie in clusters on the short arm of chromosome 1 and include genes encoding for *Cf-9*, *Cf-9B*, *Cf-4*, *Cf-Ecp2*, *Cf-Ecp5*, *Cf-Ecp1* and *Cf-Ecp4* (Parniske et al., 1997; Parniske et al., 1999; Thomas et al., 1997; Van der Hoorn et al., 2001a; Wulff et al., 2009a). *Hcr2s* lie in a cluster on the short arm of chromosome 6 and include genes encoding for *Cf-2* and *Cf-5* (Dixon et al., 1998; Dixon et al., 1996).

The *C. fulvum* effector Ecp2 was isolated from the apoplastic fluid of an infected tomato plant (de Wit et al., 1994). Ecp2 is recognised by a single dominant gene, *Cf-Ecp2*, in *S. pimpinellifolium* (Haanstra et al., 1999; Lauge et al., 1998). Recognition of Ecp2 results in HR (Lauge et al., 1998). *Cf-Ecp2* has been mapped to the *OR* locus on the short arm of chromosome 1 at 7.7 cM distal to the marker TG236 and 6.0 cM proximal to the marker TG184 (Haanstra et al., 1999). Within this interval, the marker CT116 was found to co-segregate with *Cf-Ecp2* in 564 products of meiosis (Haanstra et al., 1999). Having sequenced 29 kb around the CT116 marker, the *OR* locus was shown to contain three *Hcr9s*; *2A*, *2B* and *2C* (de Kock et al., 2005). However, these 29 kb did not include the flanking markers TG236 and TG184 genetically delimiting *Cf-Ecp2* (de Kock et al., 2005). Stable transformation of *2A*, *2B* and *2C* with 1 kb of native 5' regulatory sequence into *S. lycopersicum* did not confer recognition of Ecp2 in the transformants (de Kock and colleagues, 2004).

The Ecp2 effector is considered to be important to the fungus. Ecp2 is conserved across the fungal class Dothidiomycetes, and homologs of the *C. fulvum* Ecp2 sequence have been identified in not only important plant pathogens such as *M. fijiensis* and *Z. tritici*, but also human pathogens and saprophytic fungi (Stergiopoulos et al., 2012; Stergiopoulos et al., 2010). In addition, the *M. fijiensis* (MfEcp2) variant of Ecp2 has been found to be under-diversifying selection (Stergiopoulos et al., 2014). If an effector is not important then it can easily be lost by the pathogen, thus overcoming any *R* gene in the plant capable of recognising that specific effector (Van Kan et al., 1991). In the case of Ecp2, however, the propagation of its alleles in the population rather than deletion of the

gene altogether suggest the importance of the effector to the pathogen. These features highlight *Cf-Ecp2* as an important target for cloning. The *Cf-Ecp2* gene from *S. pimpinellifolium* , in addition to recognising Ecp2 from *C. fulvum*, is also able to recognise the *M. fijiensis* effector MfEcp2-1 (Stergiopoulos et al., 2012; Stergiopoulos et al., 2010). Similarly, the tomato *R* gene *Cf-4* can recognise the *M. fijiensis* effector MfAvr4 (Stergiopoulos et al., 2010). *M. fijiensis* causes destructive Black Sigatoka disease on banana (Churchill, 2011; Koeppel, 2008).

The most common banana plant grown is Cavendish (Koeppel, 2008). This banana plant is susceptible to *M. fijiensis* (Churchill, 2011; Koeppel, 2008). Since Cavendish banana is a sterile clone, classical genetics cannot be used to breed resistance into the crop from wild relatives. Genetic modification, however, provides an avenue for incorporating novel resistance to *M. fijiensis* into banana. If the resistance-signalling pathway in tomato is conserved in banana, then the genetic components coding for the recognition of MfEcp2 and MfAvr4 could be transferred from tomato into banana. This could generate a more durable resistance in banana against *M. fijiensis.*

Variants of MfEcp2-1, MfEcp2-2 and MfEcp2-3 have been identified in populations of *M. fijiensis* from South East Asia, Africa and Latin America (Stergiopoulos et al., 2014). Two of the four isoforms of Mf-Ecp2-1 from South East Asia are not recognised by *S. pimpinellifolium* CfEcp2 (Stergiopoulos et al., 2014). Furthermore, all isoforms of MfEcp2-2 and MfEcp2-3 are not recognised by *S. pimpinellifolium* CfEcp2 (Stergiopoulos et al., 2014).

The aim of this study was to determine whether the genetic element(s) that enable(s) the recognition of Ecp2 and MfEcp2 is/are encoded by the same locus in *S. pimpinellifolium* CfEcp2. This question was addressed by the generation and characterisation of mutants that no longer recognise Ecp2, and by attempting to clone *Cf-Ecp2*.

Furthermore, this study also aimed to determine the recognition spectrum of MfEcp2 isoforms in Ontario 7518, the *S. pimpinellifolium* accession used for cloning *Cf-Ecp2* in this study. Understanding the MfEcp2 recognition spectrum of Cf-Ecp2 is/would be important for deployment and maximising durability of the *R* gene in banana. In the current study, these isoforms will therefore be tested on the JIC Ontario 7518 of *S.*

pimpinellifolium CfEcp2 to confirm their similarity with that of the Stergiopoulos et al. (2014) Ontario 7518.

The aim of this current research is to clone the resistance gene *Cf-Ecp2* that encodes for recognition of the core effector Ecp2. The cloning of this gene may lead to its dissemination into other plant species that are also infected by pathogens carrying Ecp2, and thereby provide protection against these pathogens.

In this study, the *Cf-Ecp2* locus encoding for Ecp2-recognition in *S. pimpinellifolium* was mapped to a 40 kb region relative to the sequenced Heinz tomato genome. The locus was resolved using a combination of sequencing-by-synthesis (Illumina and 454-pyrosequencing), and single molecule sequencing (PacBio and MinION). The MinION sequencing platform was vital in generating long unbiased reads for determining the copy number of *R* gene homologues at the locus. The *Cf-Ecp2* locus was found to contain eight *Hcr9* genes, including four copies of *2A*, which are 100% identical to each other in their ORFs. Such significant conservation of gene sequence suggests recent duplication by illegitimus recombination. The use of a transposon tagging experiment in the current research enabled two independent deletion mutants to be generated. The mutants had lost the ability to recognise Ecp2. These deletion mutants were used to characterise the ability of *S. pimpinellifolium* CfEcp2 to recognise *M. fijiensis* isoforms of MfEcp2 and to determine whether MfEcp2 is recognised by the same genetic component as Ecp2. In addition, the deletion mutants identified *2A* as a candidate for *Cf-Ecp2*.

3.2 Results

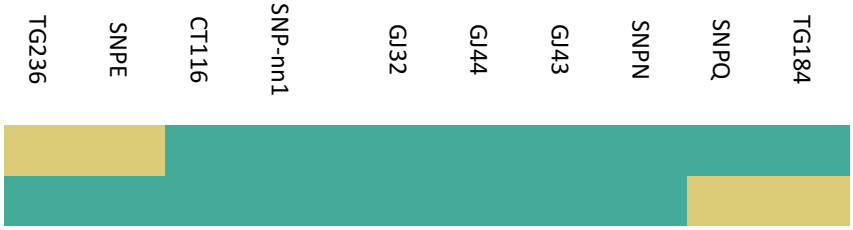
3.2.1 Genetic fine mapping of *Cf-Ecp2*

Note: The majority of work described in section 3.2.1 was performed by Gopaljee Jha. These unpublished results need to be described here as a necessary prerequisite for analysis described in this thesis.

Ecp2 recognition was previously mapped to the *OR* locus on the short arm of chromosome 1 in *S. pimpinellifolium,* at 7.7 cM distal to TG236 and 6.0 cM proximal to TG184, and completely linked to CT116 (<0.3 cM distance)(Haanstra et al., 1999).

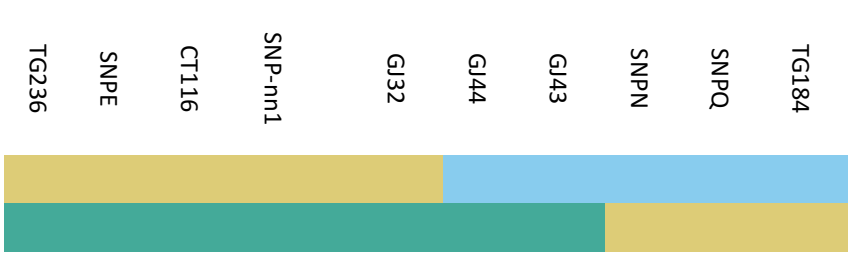
A)

F_2 recombinant	F_3 segregation ratio R:NR
4H09	0:27
2C10	0:21



B)

F_2 recombinant	F_3 segregation ratio R:NR
11E01	18:5
13D08	0:38



C)

F_2 recombinant	F_3 segregation ratio in F_3 R:NR
51	7:1
54	7:1
58	3:5
52	7:1
63	6:2
50	7:1
55	6:2
47	4:4
57	5:3
Cf0	0:8
CfEcp2	8:0

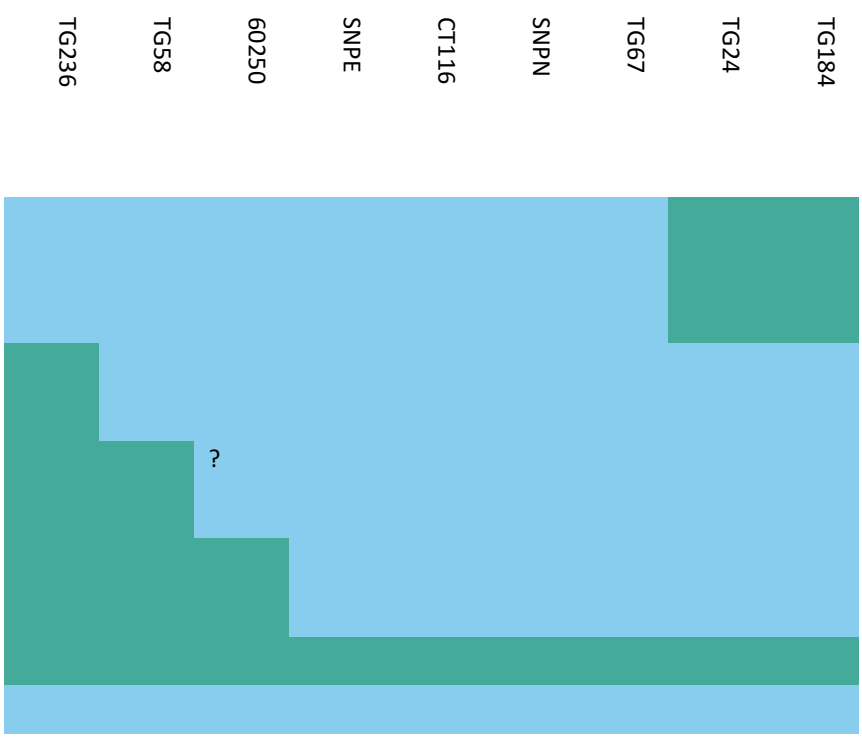


Figure 3.1. Genotypes and phenotypes of key recombinants delimiting *Cf-Ecp2*.

A) The genotypes of *S. pimpinellifolium* (CfEcp2) and *S. lycopersicum* (Cf0) F_2 recombinant plants, preselected for surviving an Potato Virus X (PVX):*Ecp2* screen i.e. not containing *Cf-Ecp2* B) The genotypes of the F_2 parents from a cross between *S. pimpinellifolium* (CfEcp2) and *S. lycopersicum* (Cf0). C) The response of eight F_3 individuals from each recombinant F_2 family (selected because the non-recombinant chromosome was Cf0 from a cross between *S. pimpinellifolium* [CfEcp2] and *S. lycopersicum* [Cf0]) to PVX:*Ecp2*. The genotypes of F_3 plants homozygous for the recombinant chromosome are shown. In A, B and C the response of F_3 individuals from each recombinant F_2 family to PVX:*Ecp2* are shown as a ratio of Response (R) to No response (NR). The marker from *S. pimpinellifolium* CfEcp2 parent is shown in blue and the marker from *S. lycopersicum* Cf0 parent in green. Those heterozygous for a marker are shown in yellow. Question marks indicate where the genotype is unknown.

To further delimit the genetic interval of the *Cf-Ecp2* locus in this study, as a prelude to cloning *Cf-Ecp2*, the progeny from a new cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 were analysed. One thousand seven hundred F_2 plants were inoculated with PVX:*Ecp2*, and 450 survivors were obtained (Gopaljee Jha, 2011 unpublished). This confirms the previous observation that *Cf-Ecp2* segregates as a single dominant gene in this cross.

To fine map *Cf-Ecp2*, the 450 survivors were genotyped with the markers TG236 and TG184 (Gopaljee Jha, 2011 unpublished). A total of 80 recombinants between TG236 and TG184 were identified, indicating that these markers are 8.9 cM apart, which is less, but comparable to the 13.7 cM interval obtained in the study by Haanstra et al (1999) (Appendix 2, Table 2.1) (Gopaljee Jha, 2011 unpublished). Seven new CAPS markers within this interval were obtained by comparing the published *S. lycopersicum* Heinz 1706 sequence (Sato et al., 2012) with Illumina reads from whole genome shotgun sequencing of *S. pimpinellifolium* CfEcp2 (Gopaljee Jha, 2011 unpublished). These markers were named SNPE, SNPN, SNP-nn1, GJ32, GJ44, GJ43 and SNPQ (Gopaljee Jha, 2011 unpublished). The recombinants were genotyped with these makers (See Appendix 2, Table 2.1). This allowed the identification of two key recombinants, 4H09 and 2C10, which delimit a 45 kb interval relative to the Heinz tomato genome, between the markers SNPE and SNPQ (0.22 cM) (Gopaljee Jha, 2011 unpublished). These recombinants were advanced to the F_3 generation in which their genotype and phenotype were confirmed (Figure 3.1A).

Because *Cf-Ecp2* segregates as a single dominant gene and the recombinants 4H09 and 2C10 do not recognise Ecp2, they define an interval required for *Cf-Ecp2* function. This in turn suggests that this interval is sufficient for *Cf-Ecp2* function and that *Cf-Ecp2* maps to this interval. To test this hypothesis, a new screen was performed for recombinants between TG236 and TG184 in the F_2 progeny from the cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicon* Cf0, but this time without preselecting for survival to PVX:*Ecp2* (Gopaljee Jha, 2011 unpublished). From 750 F_2 plants a total of 172 recombinants were identified (equivalent to a recombination distance of 11.5 cM) (Appendix 2, Table 2.2) (Gopaljee Jha, 2011 unpublished). These were further analysed with the CAPS markers SNPE and SNPN (Gopaljee Jha, 2011 unpublished). Two recombinants, 11E01 and 13D08, were identified that contained recombinantions between SNPE and SNPN (Gopaljee Jha, 2011 unpublished). These two recombinants were further analysed with the markers SNP-nn1, GJ32, GJ44 and GJ43 (Gopaljee Jha, 2011 unpublished). The recombinants were advanced to the F_3 generation and their genotype and Cf0 phenotype were confirmed. This further delimited the interval required for *Cf-Ecp2* function to a 40 kb interval between markers SNPE and GJ44 (0.18 cM). Nine recombinants were identified to contain the CfEcp2 parent allele at the markers SNPE and SNPN and where the non-recombinant chromosome was Cf0 (Appendix 2, table 2.2). These nine recombinants were advanced to the F_3 stage. For each selected F_2, eight F_3 plants were inoculated with PVX:*Ecp2* and phenotyped to determine the segregation for the presence and absence of *Cf-Ecp2* (Figure 3.1).

In order to refine the break points of the nine recombinants, F_3 individuals that were homozygous for the recombinant chromosome were identified and further analysed using publicly available markers from SolGenomics (https://solgenomics.net); TG58, TG24, TG67 and 60250 (Tables 2.2 and Figure 3.1).

This enabled mapping the presence of *Cf-Ecp2* to a 631 kb region relative to the publicly-available *S. lycopersicon* Heinz tomato genome sequence, between the markers 60250 and TG24 (0.69 cM) (Figure 3.1 and 3.2). This supports the interpretation that *Cf-Ecp2* resides within the 40 kb interval defined by the recombinants 4H09 and 11E01 that lack *Cf-Ecp2* function. However, it cannot be formally excluded that *Cf-Ecp2* is in fact comprised of two or more functional components that lie within the 631 kb region, and that one of these resides in the 40 kb region between SNPE and GJ44 (Figure 3.2).

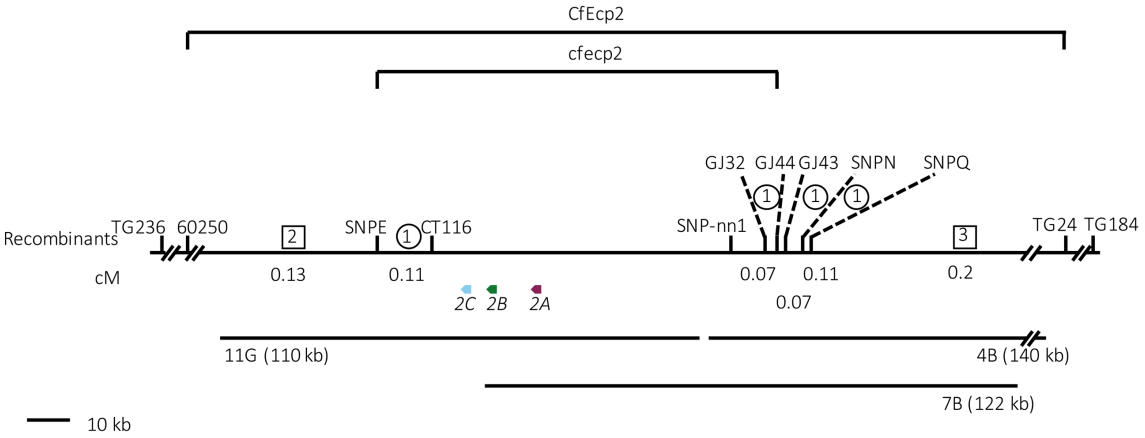


Figure 3.2. Physical and genetic map of the *Cf-Ecp2* locus in currant tomato.

Molecular markers are indicated above the tick marks. The number of recombination breakpoints between *S. pimpinellifolium* CfEcp2 (currant tomato) and *S. esculentum* Cf0 is indicated above the chromosome, while genetic distance (centiMorgan [cM]) is indicated below the chromosome. Numbers in squares indicate recombinants that retain *Cf-Ecp2* function whilst numbers in circles indicate recombinants that lack *Cf-Ecp2* function. The position of three homologues of Cladosporium resistance gene *Cf-9* (*Orion OR2C* [*2C*], *OR2B* [*2B*] and *OR2A* [*2A*]) is shown by pointed rectangles as identified by de Kock et al. (2005). The position and size of BACs 11G (110 kilo base [kb]), 7B (122 kb) and 4B (140 kb) is indicated.

3.2.2 Generation of a BAC minimal tiling path

To delimit a physical map of the locus, a BAC library was generated from the genome of *S. pimpinelifolium* CfEcp2. The molecular markers SNPE, *2A* and SNPN from the *OR* locus were used to identify three overlapping BACs; 11G, 7B and 4B. The three BACs were found to span the *Cf-Ecp2* map interval defined by recombinants 4H09 and 11E01. The BACs 11G, 7B and 4B were digested with *Not*I and analysed by PFGE, indicating the size of their inserts to be 110, 122 and 140 kb, respectively (Figure 3.3).

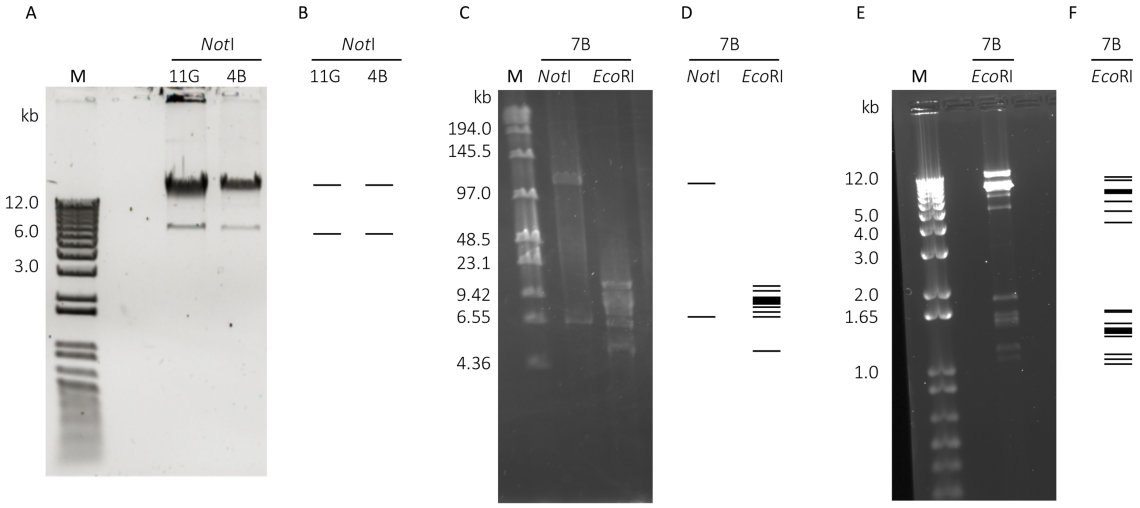


Figure 3.3. Restriction digestion of BACs 11G, 4B and 7B.

(A, B) BACs 11G and 4B digested with *Not*I and size-separated by electrophoresis in a 0.8% agarose gel or *in silico*.
(B, C) BAC 7B digested with *Not*I or *Eco*RI and size-separated by pulse-field gel electrophoresis in a 1% agarose gel or *in silico*.
(E, F) BAC 7B digested with *Eco*RI and size-separated by electrophoresis in a 0.8% agarose gel or *in silico*. Size of marker (M) fractions indicated in kilo bases (kb). BACs = Bacterial artificial chromosomes.

3.2.3 Sequence characterisation of the *Cf-Ecp2* locus

PCR analysis confirmed the presence of *2C* and *2B* on BAC 11G, *2B* and *2A* on BAC 7B and the absence of *2C*, *2B* and *2A* on BAC 4B (Figure 3.4). PCR for *2A* from CfEcp2 gave no visible amplification products in comparison to BAC 11G (Figure 3.4C). This is most likely due to the large template concentration in the BAC DNA sample in comparison to the whole genomic DNA sample. The 4B and 11G BACs were sequenced on the 454 pyrosequencing platform GS FLX+. The 454 reads from BAC 11G were assembled by the MIRA shotgun sequence assembler into three large contigs of 11,230 bp, 27,137 bp and 46,495 bp, and 84 smaller contigs (175 to 2,781 bp) (Table 3.1). The larger contigs had a 154–221-fold coverage. The 454 reads from BAC 4B were assembled into three large contigs of 12,986 bp, 53,994 bp, and 65,264 bp, and 69 shorter contigs (152 to 3,058 bp) (Table 3.1). The large contigs had a 94 to 98-fold coverage. The 7B BAC was sequenced on the sequencing-by-synthesis Illumina MiSeq platform with 250 bp PE reads. These reads were assembled into two large contigs of 47,096 bp and 35,488 bp and 69 shorter contigs (208 to 5,094 bp), using MIRA (Table 3.1). Since the shorter contigs, from the assemblies of 11G, 4B and 7B sequencing, had a lower average coverage, it is assumed

that they originated from contamination, sequencing errors or highly complex repetitive parts of the BAC.

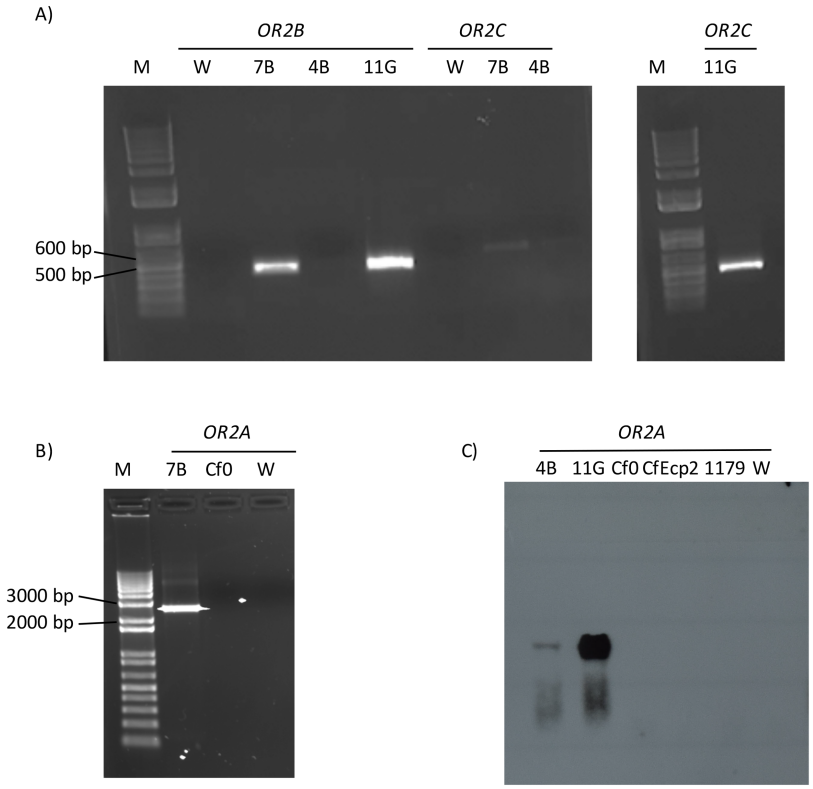


Figure 3.4. PCR analysis of BACs 11G, 4B and 7B for the presence and absence of *2A*, *2B* and *2C.*

(A) Screening by PCR for the presence of *OR2B* (*2B*) and *OR2C* (*2C*) in Bacterial artificial chromosomes (BACs) 7B, 4B and 11G.
(B) Screening by PCR for the presence of *OR2A* (*2A*) (3' primer and 5' primer) in BAC 7B, *S. lycopersicum* Cf0 and water.
(C) Screening by PCR for the presence of *2A* in BACs 11G and 4B, and in S. pimpinellifolium CfEcp2 and 1179 DNA.
PCR products, in (A) and (B) were run in 1% agarose gels and next to 1 Kb plus Invitrogen marker (M). Important size markers are indicated in base pairs (bp). W = water.

Table 3.1. Sequencing by 454 and Illumina and assembly of BACs 11G, 7B and 4B.

	BAC 11G	**BAC 7B**	**BAC 4B**
Sequencing:			
Sequencing platform	454	Illumina	454
Number of reads	25,736	17,188,756	17,306
Mean read length (bp)	534	250	538
MIRA assembly:			
Total number of contigs (bp)	87 (175 – 2781)	69 (208 – 5094)	72 (152 – 3058)
Size of contigs >10 kb (bp)	46,495 27,137 11,230	47,096 35,488	65,264 53,994 12,986
Fold coverage of >10 kb contigs	154-221	222-386	94-98

BAC = Bacterial artificial chromosome, bp = base pairs, kb = kilo bases.

To position the *Hcr9s* on BACs 11G, 7B and 4B, the assembled contigs that were >10 kb were aligned to each other, the individual genes *2A*, *2B* and *2C*, or the published *OR* locus sequence using a dotplot. The locations of the genetic markers were identified by mapping the known sequence of the markers onto the contigs, using Geneious R8 software. The 46,495 bp contig from BAC 11G contained SNPE, CT116, *2C*, *2B* and *2A*. The other two contigs from BAC 11G of 11,230 bp and 27,137 bp did not contain any *Hcr9s*. The contig of 12,986 bp from BAC 4B contained part of the vector sequence. A new gene, *Ψ2C*, which was closer in homology to *2C* than *2B* or *2A*, was determined by a pair-wise comparison of the gene sequences using Geneious (Figure 3.5). The 4B BAC contig of 53,994 bp contained the marker GJ44. The 4B BAC assembled contig of 65,264 bp contained no sequence homology to the *OR* locus or the published genome of *S. pimpinellifolium*, as determined by BLASTN. The 47,096 bp contig from BAC 7B was found to contain *2B*, a pseudo-gene of *2A* (*Ψ2A5*) and a pseudo-gene of *2C* (*Ψ2C*) (Figure 3.6A and B).

By aligning the 454 reads from BAC 11G onto the published *OR* locus, using BLASTn, a 4 to 5-fold increase in coverage was observed across the ORF of the *2A* gene. This indicated that there are multiple copies of *2A* present at the locus (Figure 3.7). Some of the BAC 11G 454 reads that contained part of the vector sequence also contained a sequence from a new *2A* gene, later identified as *Ψ2A5*. The 11G contigs also indicated that the locus reads in the opposite direction to that previously indicated by de Kock (de Kock et al., 2005).

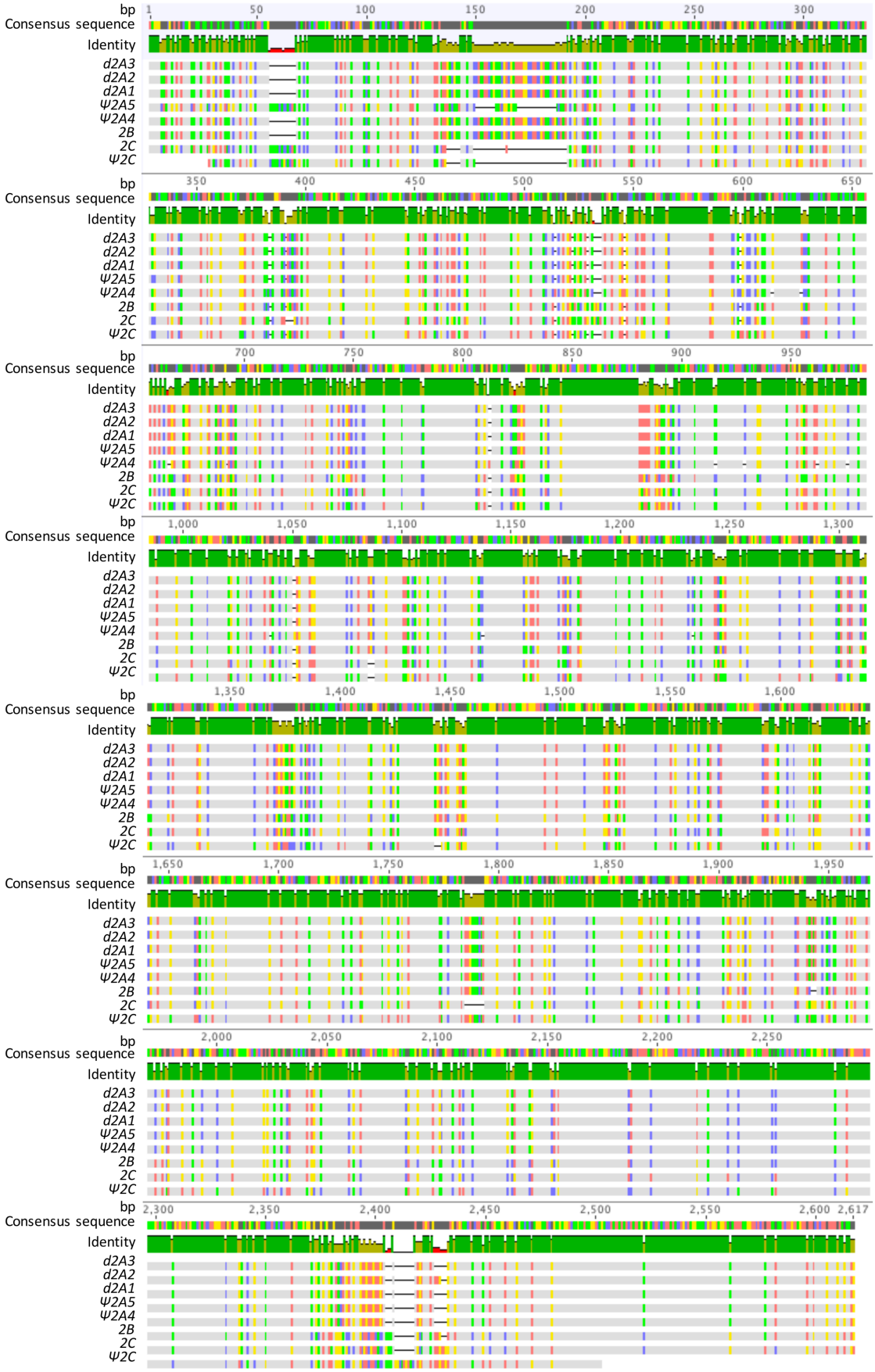

bp
Consensus sequence
Identity
d2A3
d2A2
d2A1
Ψ2A5
Ψ2A4
2B
2C
Ψ2C
1 50 100 150 200 250 300
350 400 450 500 550 600 650
700 750 800 850 900 950
1,000 1,050 1,100 1,150 1,200 1,250 1,300
1,350 1,400 1,450 1,500 1,550 1,600
1,650 1,700 1,750 1,800 1,850 1,900 1,950
2,000 2,050 2,100 2,150 2,200 2,250
2,300 2,350 2,400 2,450 2,500 2,550 2,600 2,617

Figure 3.5. Alignment of *Hcr9s* from PacBio assembly D.

The *Homologs of Cladosporium resitance gene Cf-9* (*Hcr9s*) from the PacBio assembly D, *dOR2A1* (*d2A1*), *dOR2A2* (*d2A*2), *dOR2A3* (*d2A3*), *ΨdOR2A4* (*Ψd2A4*), *ΨOR2A5* (*Ψ2A5*), *ΨOR2C* (*Ψ2C*), *OR2C* (*2C*) and *OR2B* (*2B*), were aligned to each other using ClustralW. The first row refers to the base pairs (bp) within the consensus, and the second row is the consensus sequence A (red), C (purple), G (yellow), T (Green). The third row is a graph of the % identity between the eight sequences, green is 100% identity and brown means less than 100% identity. This is followed by eight rows of the sequence of the homologs ordered according to their similarity. Similarity of sequence represented by grey bars. Disagreement between the eight sequences is highlighted Adenine (red), Cytosine (purple), Guanine (yellow), Thymine (green). Image produced in Geneious R8.

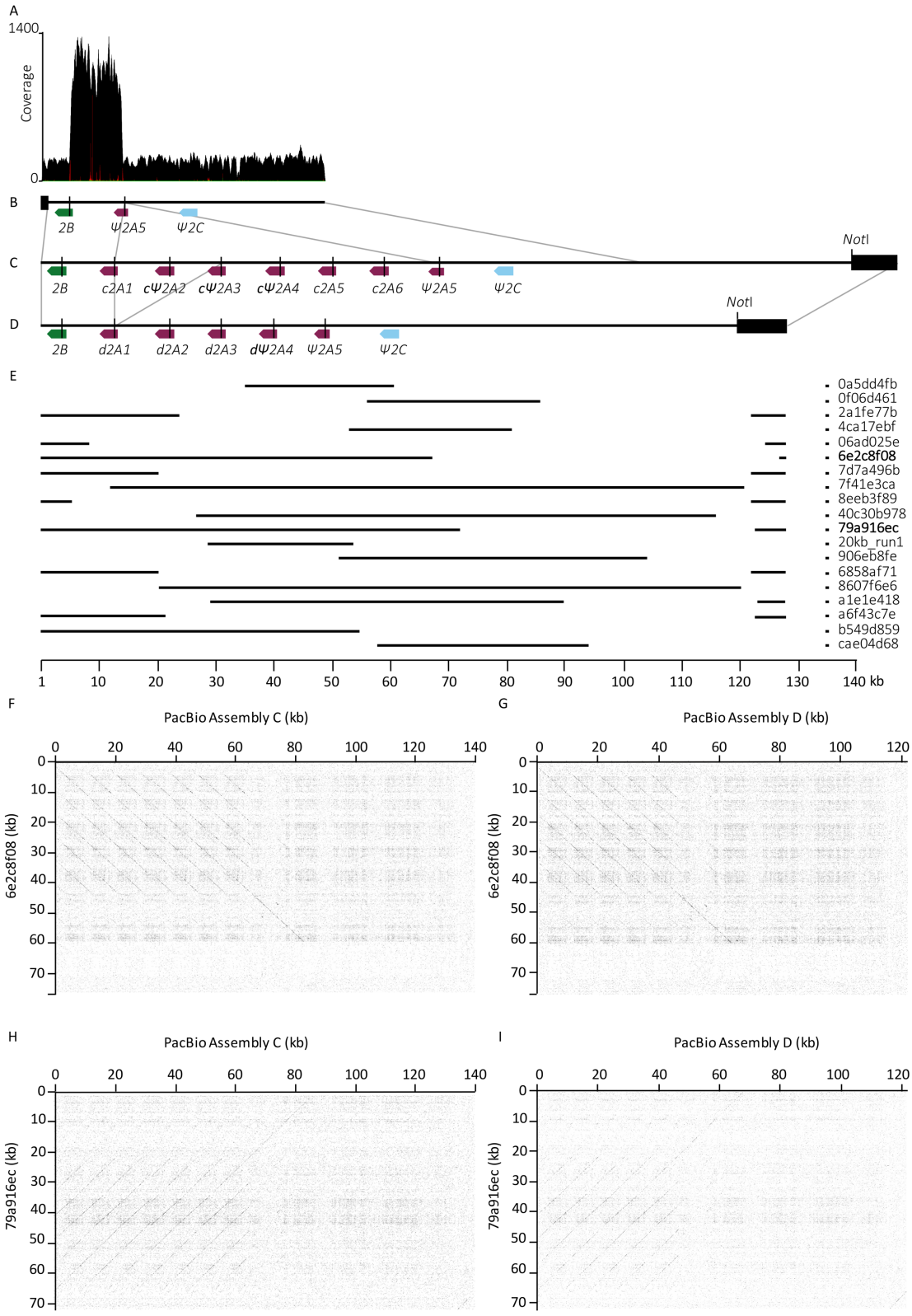

A
1400
Coverage
0
B
2B
Ψ2A5
Ψ2C
NotI
C
2B
c2A1
cΨ2A2
cΨ2A3
cΨ2A4
c2A5
c2A6
Ψ2A5
Ψ2C
NotI
D
2B
d2A1
d2A2
d2A3
dΨ2A4
Ψ2A5
Ψ2C
E
0a5dd4fb
0f06d461
2a1fe77b
4ca17ebf
06ad025e
6e2c8f08
7d7a496b
7f41e3ca
8eeb3f89
40c30b978
79a916ec
20kb_run1
906eb8fe
6858af71
8607f6e6
a1e1e418
a6f43c7e
b549d859
cae04d68
1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 kb
F
PacBio Assembly C (kb)
6e2c8f08 (kb)
G
PacBio Assembly D (kb)
6e2c8f08 (kb)
H
PacBio Assembly C (kb)
79a916ec (kb)
I
PacBio Assembly D (kb)
79a916ec (kb)

Figure 3.6. Assembly of BAC 7B and repeat copy number determination using Illumina, PacBio and MinIon sequencing technology.

(A) Illumina sequence reads from Bacterial artificial chromosome (BAC) 7B mapped by Basic Local Alignment Search Tool (BLAST) to the 47 kilo base (kb) contig of BAC 7B generated by MIRA assembly with the same reads. The red bars indicate the most prominent Single nucleotide polymorphisms (SNPs) in the reads compared to the published *Orion* sequence. The ~5-fold above average coverage denotes the boundaries of the 8.7 kb repeat.
(B) The 47 kb contig (horizontal black line) generated by MIRA assembly of 250 bp paired end (PE) reads from Illumina sequencing of BAC 7B.
(C) Assembly C (horizontal black line) generated using The Hierarchical Genome Assembly Process (HGAP) V3 with PacBio sequences from BAC 7B.
(D) Assembly D (horizontal black line) generated using HGAP V1 on PacBio sequences from 7B.
In B, C and D: *OR2B* (*2B*) homologs are indicated by green arrows, *OR2A* (*2A*) homologs are indicated by magenta arrows, and *OR2C* (*2C*) homologs are indicated by blue arrows. The vector is indicated by a black box and the position of the unique *Not*I restriction-site is shown. The 8.7 kb repeat is indicated by vertical black lines. Blocks of synteny between the three assemblies is indicated by interconnecting grey lines.
(E) MinION reads (black lines) positioned relative to Assembly D using dotplot. The name of each read is indicated on the right hand side. For the two reads in bold, '6e2c8f08' and '79a916ec', their dotplots against assemblies C and D are shown in F, G, H, and I.
The scale bar is relative to A, B, C, D, and E.
(F, G) Dotplots of MinION read '6e2c8f08' against PacBio assemblies C and D.
(H, I) Dotplots of MinION read '79a916ec' against PacBio assemblies C and D.
Note: Although very faint due to the high error rate in MinIon reads, the diagonals can clearly be identified.

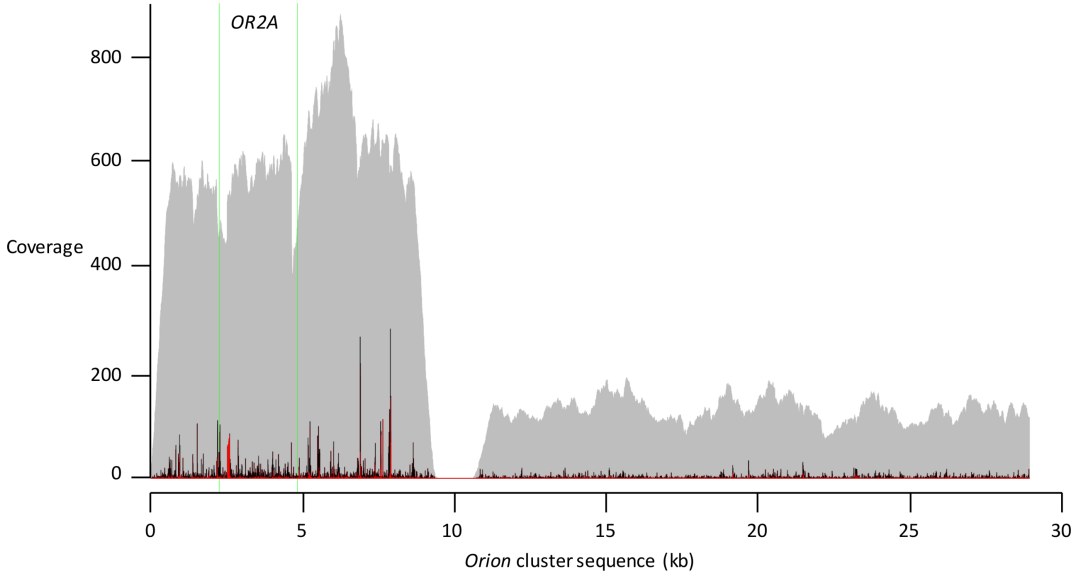


Figure 3.7. Mapping of 11G BAC sequencing reads onto the published *OR* locus.

11G Bacterial artificial chromosome (BAC) 454 reads mapped by Basic Local Alignment Search Tool (BLAST) onto the published *Orion* (*OR*) cluster sequence (de Kock et al., 2005). The position of *2A* is flanked by green lines. Black and red lines indicate the most prominent and next most prominent SNPs, respectively, in reads different to *OR*. The gap in coverage at 10,000 bp is due to the direct repeat in the promoter of *2A* and *2B*. The BLAST analysis will not count reads at two places if there are two perfect matches but will place them at the first position when scanning through the sequence from 0 to 29 kilo bases (kb).

BWA mapping of the BAC 7B 250 bp PE reads back onto the 7B 47,096 bp contig, showed a 5-fold increase in coverage from the 5' of *2B* and the 3' region of *Ψ2A5* (Figure 3.6). Since the 5' region of *2B* and 3' region of *Ψ2A5* are 100% identical to *2A* (Figure 3.5), this therefore also indicated that there are many copies of *2A* present at the locus. SNPs between the 47 kb contig and that of de Kock et al. (2005) were identified (Figure 3.6A red bars), indicating potential sequence variation between the *2A* copies at the locus. However, following inspection of the SNPs (identified by visual inspection in the Savant genome browser), the SNP with the most prominence within the coding region of *2A* was present only in 10% of the reads mapped (Table 3.2). If there are five copies of *2A* and the SNP is present in one of the genes it should be present in 20% of the reads.

Table 3.2. SNPs identified in Savant, between the raw Illumina reads and the MIRA-assembled contig.

	Position on Illumina contig	Assembled contig base	Presence in reads (%)	Alternative base in reads (SNP)	Presence in reads (%)
In *2A* ORF	12,714	G	96.7	C	3.2
	12,710	T	94.6	C/A	4.2/1.0
	12,708	A	93.5	C	6.4
	12,703	T	93.3	A	6.6
	12,689	C	95.8	T	4.1
	12,665	T	89.5	C	10.3
	12,647	A	98.8	C	1.2
	12,640	A	90.5	C	9.4
Not in *2A*	9,699	T	94.8	C	5.0
	9,694	A	91.5	C	7.7
	9,658	T	83.6	C	16.2
	9,633	A	85.6	C	14.1
	8,097	C	76.0	A	23.9
	8,087	T	94.6	A	5.4
	8,084	G	95.6	T	4.4
	8,067	T	92.7	A	7.2
	8,045	C	90.5	A	9.5

SNPs = Single nucleotide polymorphisms, *2A* = *OR2A,* ORF = Open reading frame.

To resolve the sequence spanning from *2B* to *Ψ2C*, two SMRT cells of single molecule PacBio sequencing were performed on BAC 7B. This involved the use of CCS with an average read length of 7,932 bp and 4,781 bp and with an output of 27,127 and 35,635 filtered reads (Table 3.3).

Table 3.3. PacBio sequencing and assembly of BAC 7B.

	Assembly C	**Assembly D**
Sequencing:		
Number of filtered reads	27,127	35,635
Average read length (bp)	7,932	4,781
Mean read quality	0.846	0.854
Assembler and parameters:		
Programs	HGAP V3; Celera Assembler V8.1; Quiver	HGAP V1; Celera Assembler V7.0; Quiver
xcoverage	15	15
default.fragMinLength	500	1000
minLongReadLength	6000	6000
minSubReadLength	500	500
filter.minLength	100	100
Output:		
Contig length (bp)	140720	123005
Number of *Hcr9s*	9	7

BAC = Bacterial artificial chromosome, HGAP = The Hierarchical Genome Assembly Process, bp = base pairs, *Hcr9s* = homologs of Cladosporium resistance gene *Cf-9*

Two different sets of parameters were used to assemble the resulting reads in HGAP, and each assembly differed in the number of homologs of *2A* present (Table 3.3). The assemblies were named C and D, and were 140,720 bp and 123,005 bp long, respectively. In order to identify the *Hcr9*s on each assembly, the assembly was aligned by dotplot to itself, to the 7B Illumina assembly and to the known genes *2A*, *2B* and *2C*. The identified homologs of *2A* were named according to their origin, either from assembly C or assembly D. Further on, genes from different assemblies are annotated with the prefixes c or d (e.g. *c2A*). In addition, their order within the assembly is denoted by a running number as suffix, e.g. *c2A1* being the first *2A* gene 5' to 2B within the assembly C. Assembly C contained a total of nine *Hcr9s* (*2B*, *Ψ2C* and seven homologs of *2A*). The *2A* homologs included three 100% identical copies of *2A* (*c2A1*, *c2A5* and *c2A6*) and four pseudogenes of *2A*: *cΨ2A2*, *cΨ2A3* and *cΨ2A4* (99.7, 99.7 and 99% identical, respectively) and *Ψ2A5* (94.6% identical to 2A). Homolog identity was determined via pair-wise comparisons (Figure 3.6C). Assembly D contained a total of seven *Hcr9s* (*2B*, *Ψ2C* and five homologs of *2A*). Two homologs were 100% identical to *2A* (*d2A1* and *d2A3*), one homolog, (*d2A2*) was 99.9% identical to *2A* and then there were two pseudogenes of *2A*

(*dΨ2A4* and *Ψ2A5*). Both assemblies contained the BAC vector sequence (Figure 3.6C and D).

Since the copy number of *2A* was not resolved by PacBio or Illumina short-read sequencing, two rounds of Nanopore sequencing on BAC 7B were performed using the Oxford Nanopore MinION platform, which has been reported to produce very long reads up to 67 kb (Ashton et al., 2015)(Table 3.4). The raw MinION reads were filtered and those (i) >9 kb, (ii) with BLAST hits in the region of the BAC known to contain the *Hcr9s* (1,265 – 58,486), and (iii) which could be positioned with certainty on the BAC using dotplot analysis, were selected for further analysis (Table 3.4 and Figure 3.6E). A total of 19 filtered reads were identified, ranging in sequence length from 10.2 kb to 89 kb and with a mean read length of 40 kb. An homology-based comparison of Nanopore reads against the BAC reference sequence revealed that best local alignments range between 65% and 82% identity. Since these values were taken from best alignments that do not necessarily include the entire read, the real identity for the global alignment of each read is far less. Due to this high error rate of the MinION reads, the reads could not be used to identify the specific sequence of the homologs. Nonetheless, they proved extremely powerful as a scaffold to determine the copy number of *2A* on BAC 7B.

Table 3.4. MinION sequencing output.

Run	Read type	Total number of reads	Length of reads		[1]Reads of use	[2] Average % identity
			Minimum	Maximum		
1	N/A	20,183	5	220,110	1	43.3
2	2D pass	146	492	24,153	18	61
	2D fail	320	182	33,191		
	Complement pass	146	529	23,875		
	Complement fail	904	12	89,076		
	Template pass	146	492	24,328		
	Template fail	3,655	5	226,723		

[1]Reads of use: >9 kilo bases (kb) in length and with Basic Local Alignment Search Tool (BLAST) hits for Bacterial artificial chromosome (BAC) 7B in the region 1,265-58,486 base pairs (bp) that can be positioned with certainty on the BAC.
[2]Average % identity for "Reads of use", when aligned to BAC 7B assembly. Global alignments between reads and predetermined mapping positions (using Dotter) in BAC were performed using MUSCLE.

To identify the copy number of *2A* on the BAC 7B, two MinION reads, '79a916ec' and '6e2c8f08', were anchored outside of the repeat region known to contain the *Hcr9s* (3' of *2B* and 5' of *2C*) on both sides. Dotplot alignments of these reads to each of the assemblies indicated that they contained five copies of *2A* (Figure 3.6F, G, H and I).

Assembly D was therefore taken forward for further analysis.

3.2.4 Manual editing of the 7B BAC sequence

Since several errors were identified between the 250 bp PE Illumina reads and BAC 7B assembly D (Figure 3.8A), the raw PacBio reads >15 kb were mapped to assembly D. The SNPs between the reads and the assembly were identified manually using Savant and Tablet as viewing platforms. By replacing the allele of the assembly with that of the raw reads, many of the disagreements between the Illumina reads and the PacBio assembly were resolved. The resulting manually-edited BAC 7B contained only three unresolved regions when the 250 bp PE reads were mapped to it (Figure 3.8B). The first of these was in a region where there were multiple homopolymers of adenine (Figure 3.9A). The second was in a region of adenine-thymine microsatellites (Figure 3.9B). The third was a homopolymer of 10 guanine bases followed by either another guanine or a cytosine (Figure 3.9C). Homopolymers are a known problem for many sequencing technologies including PacBio (Ross et al., 2013). The regions of homopolymers and microsatellites were not within the open reading frame of the *Hcr9s*. The final sequence of the *Cf-Ecp2* locus of 91,009 bp was reconstructed by combining assemblies of BAC 7B and 11G, prioritising the PacBio generated 7B assembly in the overlapping region. The locus spans the entire region between flanking markers SNPE to GJ44.

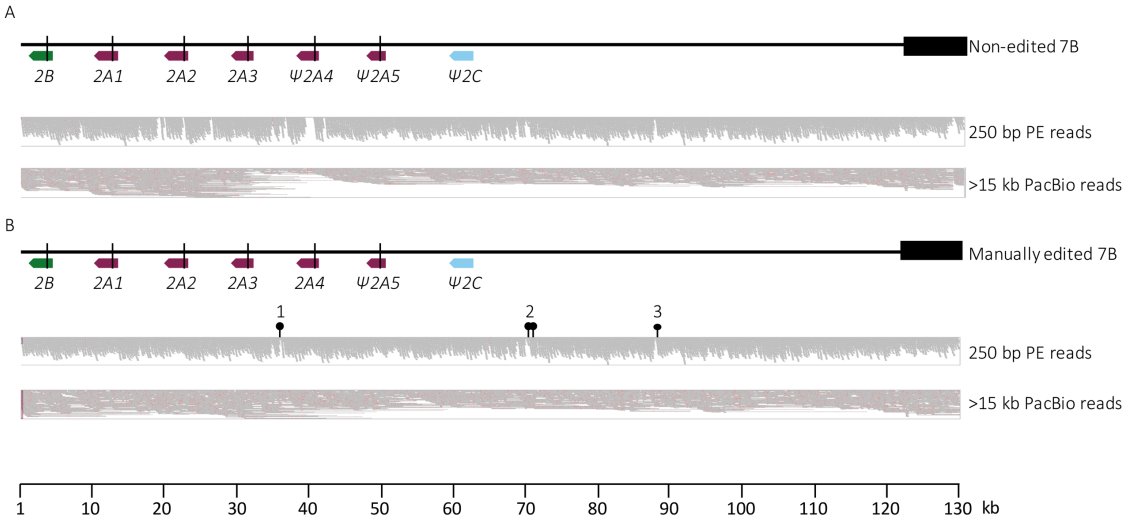


Figure 3.8. Manual editing of the BAC 7B Assembly D using mapped Illumina and PacBio raw reads.

(A) Non-edited PacBio Assembly D of Bacterial artificial chromosome (BAC) 7B (black horizontal line).
(B) Manually edited PacBio Assembly D of BAC 7B (black horizontal line).
The two tracks below non-edited and edited assembly contain the (i) 250 bp paired end (PE) Illumina reads, and (ii) the PacBio reads >15 kilo base (kb) mapped against the assembly using Burrow-Wheeler Aligner mapping (BWA). Unresolved microsatellites in the manually edited assembly are indicated by vertical lines and circles numbered 1, 2 and 3, and are detailed in Figures 3.9A, B and C, respectively. Homologs of *OR2B* (*2B*), *OR2A* (*2A*) and *OR2C* (*2C*) are indicated by green, magenta and blue arrows, respectively. The vector is indicated by a black rectangle. The 8.7 kb repeat is indicated by vertical black lines. The scale bar is relative to (A) and (B).

Figure 3.9. Unresolved positions in manually edited BAC 7B sequence.

Unresolved position 1 (A); adenosine microsatellite in Bacterial artificial chromosome (BAC) 7B. Unresolveposition 2 (B) adenine-thymine microsatellites. Unresolved position 3 (C) homopolymer of 10 guanine bases followed by either another guanine or a cytosine. For A, B and C, the first row is the section of the edited sequence of BAC 7B indicated in bold, with unresolved bases highlighted in pink. The following 12 rows are sequences from >15 kilo base (kb) reads from the PacBio sequencing mapped to the reference using Burrow-Wheeler Aligner mapping (BWA). SNPs are highlighted in green, insertions in blue, and deletions are denoted by asterisks. A= Adenine, C= Cytosine, G= Guanine, T = Thymine. Bp = base pairs.

3.2.5 Support for the 7B manually edited sequence

To confirm the physical map, BAC 7B was digested with *Eco*RI and analysed by PFGE. The electrophoresis pattern was compared to the predicated band separation obtained by *in silico* digestion of BAC 7B using NEBcutter and congruence between the digests was observed (Figure 3.3). Assembly D indicated the presence of *2B*, two copies of *2A* that were 100% identical and one copy of *2A* with an AAG insertion at 2360 bp from the start codon, two pseudogenes of *2A* (*Ψ2A4* and *Ψ2A5*) and a pseudogene of *2C* (*Ψ2C*) (Figure 3.5, 3.6D and 3.8). Pair-wise comparisons of *dΨ2A4* to the other *Hcr9s* at the locus indicated that it was composed of SNPs from *2A* and *2B* (Figure 3.5). Furthermore, the

insertion of AAG at 2360 bp was also present in *Ψ2C* and *2C*, but absent from *2A* and *2B* (Figure 3.5). The polished sequence of assembly D indicated the presence of *2B* and four copies of *2A* that were 100% identical in their ORF; with some SNPs in the promoters, mainly in homopolymer regions (Figure 3.8 and 3.10), *Ψ2A5* (94.6% identical to *2A*) and *Ψ2C* (85.1 identical to *2C*).

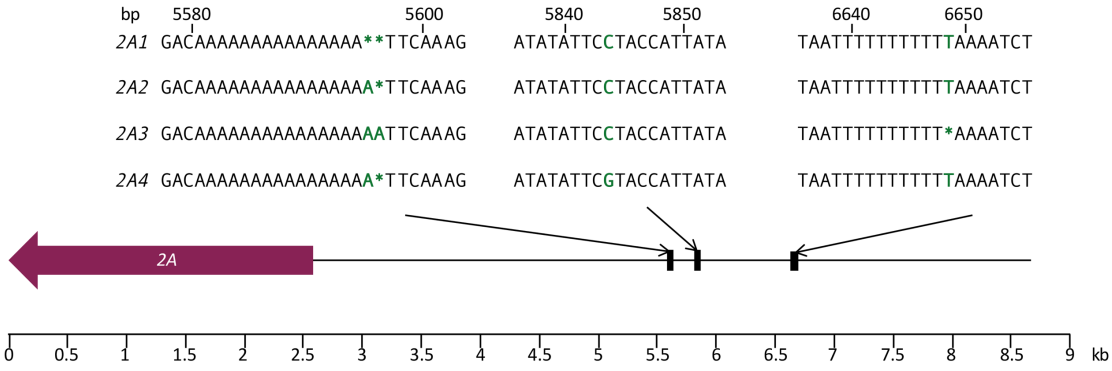


Figure 3.10. Polymorphisms in the *2A* containing repeat.

The four *OR2A* (*2A*) paralogues are 100% identical in their open reading frame. Polymorphisms between the 5' sequences of the four *2A* paralogs are highlighted in bold green and the sequence in this region is shown. Stars (*) are deletions. kb = kilo bases, bp = base pairs.

To confirm that there were no other sequence-distinct homologs of *2A* and that the manual editing of the BAC correctly removed *d2A2* and *dΨ2A4*, yet retained *Ψ2A5*, PCR amplifications and sequencing screens were set up. This procedure employed a high fidelity polymerase (Phusion) to amplify the *2A* and *Ψ2A* homologs from BAC 7B. Two sets of primers were designed to bind to the conserved parts of the promoter and terminator sequences of *Ψ2A5* and *2A,* respectively, thus amplifying the homologs with the same promoter and terminator sequences of *Ψ2A5* or *2A*. A gel purification of the range of products predicated to be 2 to 3 kb in length from the two sets of PCR were purified and cloned. The sequences of 19 clones from set *Ψ2A5* were compared to the *Ψ2A5* promoter and ORF. All 19 clones were 100% identical to *Ψ2A5* (including some gaps not encompassed by the sequences), and all contained the TAA stop codon at 49-51 bp (Table 3.5). A total of 15 out of the 19 clones were found not to contain the AAG insertion at 2,375 bp (2,389 bp in *2A*). In the remaining four clones position 2,375 could not be determined because their sequencing read ended before that position.

Sanger sequence reads of 11 clones from set *2A* were mapped to the ORF of *2A*, *2B*, *dΨ2A4* and *Ψ2A5*. The best match was consistently with *2A*. SNPs were examined between the reads and the reference *2A* sequence. There were many sequence errors, but most of these were accounted for by a separate read with the reference allele. Those reads with errors not accounted for by a separate read were trimmed. Eight out of 10 clones were shown not to carry the AAG (2,389 bp) insertion from *d2A2* (Table 3.5). The two remaining clones were not verified because their sequences did not cover that region. A total of 9 out of 10 clones were shown not to carry any of the *Ψ2A4* SNPs. The remaining clone was not verified because the sequence did not extend into that region (Table 3.5).

Table 3.5. Sequence analysis of clones from set *Ψ2A5* and set *2A.*

	Clone	AAG insertion (2,375 bp)	*Ψ2A5* TAA stop at 49-51 bp
set *Ψ2A5*	1	N	Y
	3	N	Y
	4	N	Y
	6	Unknown	Y
	7	N	Y
	8	N	Y
	9	Unknown	Y
	10	Unknown	Y
	11	N	Y
	14	N	Y
	15	N	Y
	16	N	Y
	17	N	Y
	18	N	Y
	19	N	Y
	20	N	Y
	21	Unknown	Y
	30	N	Y
	34	N	Y
	Clone	AAG insertion (2389 bp)	*dΨ2A4* SNPs from 370 to 1209 bp
set *2A*	25	N	N
	26	N	N
	33	N	N
	34	N	N
	36	N	N
	37	N	N
	40	N	N
	41	Unknown	N
	45	Unknown	N
	47	N	Unknown

Y = presence, N = absence, base pair (bp) of position analysed in brackets after description of sequence. A = Adenine, G = Guanine, T = Thymine, *2A* = *OR2A*, *dΨ2A4* = *dΨOR2A4, Ψ2A5* = *ΨOR2A5.*

Furthermore, PCR analysis using primers that bind to the unique sequence of *Ψ2A5* and sequencing of this product, confirmed the presence of *Ψ2A5* in BAC 7B (Figure 3.11). Attempted PCR amplification of *dΨ2A4*, using specific primers that bind to the unique

region of the gene, gave a product in 7B but not CfEcp2 and Cf0. However, sequencing of this product confirmed it was *2A* and not *dΨ2A4*.

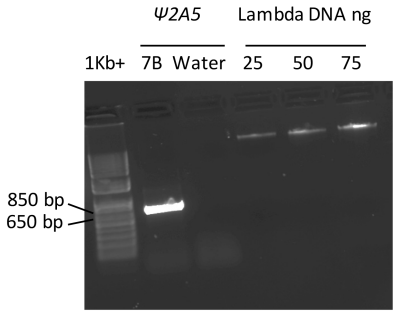


Figure 3.11. Agarose gel electrophoresis of PCR products from specific amplification of *Ψ2A5* from BAC 7B.

Products run next to the Invitrogen 1Kb+ marker (850 and 650 bp size markers indicated) and 25, 50 and 75 ng of Lambda phage DNA. *ΨOR2A5* (*Ψ2A5*), bp = base pairs.

To further confirm the absence of AAG at position 2,389 in *d2A2* in CfEcp2 and BAC 7B, primers that bind to the ATG of *2A*, and TGA of *2A*, were used to clone *2A* homologs from BAC 7B and CfEcp2 genomic DNA. Out of the 70 clones obtained for CfEcp2 and 66 clones obtained from BAC 7B (sequenced with the *OR2A* 3' primer) 47 and 29 mapped to *2A* with low sequence error, respectively. Those that did not map stemmed from error-prone sequencing. None of the 66 clones carried the AAG insert at 2,389 bp, indicating that *d2A2* was not present in either CfEcp2 or 7B. Since such inserts were sequenced with the OR2A 3'primer, none were checked for the presence of *dΨ2A4* SNPs.

The MinION read "20Kb_run1" contained three *Hcr9s*. Although it was not possible to determine the exact sequence of these, the read was positioned using dotplot to a location encompassing *ψ2A5*, *ψd2A4* and *d2A3*. By manually assigning points to each homolog with regards to their similarity to *Hcr9* on the MinION read, the results indicated that there were two copies of *2A* followed by a copy of *ψ2A5* (Table 3.6).

Table 3.6. Manually assigned points to each *2A* homolog from assembly D in relation to their similarity to *Hcr9s* A, B and C on the MinION read 20kb_run1.

2A homolog from Assembly D	MinION *2A* homolog A	B	C
2A	94	87	37
ψd2A4	69	59	17
ψ2A5	32	39	75

2A = OR2A, dΨ2A4 = dΨOR2A4, Ψ2A5 = ΨOR2A5, Hcr9s = homologs of Cladosporium resistance gene Cf-9.

3.2.6 Characterisation of *S. pimpinellifolium* CfEcp2 mutants

Note: The majority of work described in section 3.6 was performed by Colwyn Thomas and Michael Iakovidis. These unpublished results need to be described here as a necessary prerequisite for analysis described in this thesis.

To aid in the identification of the gene encoding CfEcp2, within the defined locus, two mutants of *S. pimpinellifolium* CfEcp2 were generated (Figure 2.1). The mutants were named *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179. Both mutants had lost the ability to recognise *C. fulvum* Ecp2 (Thomas et al., 2012, unpublised data). They were generated by crossing *S. pimpinellifolium* CfEcp2 (carrying the *Ds* element *FT33* genetically linked to *Cf-Ecp2*) with *S. lycopersicum* Cf0 stably overexpressing 35S:*Ecp2* (Jones et al., 1994; Takken et al., 1998). Out of 25,000 testcross progeny, only two survivors, 1178 and 1179, were identified. In both of these mutants the *Cf-Ecp2* gene was not tagged with the *Ds* element. It was therefore predicted that a deletion had occurred in the region of the genome encoding the *Cf-Ecp2*–mediated recognition of Ecp2.

To characterise the mutants further, the mutants and two controls (*S. lycopersicum* Cf0 and *S. pimpinellifolium* CfEcp2) were subjected to amplification of specific markers in the region where the presence and absence of recognition of Ecp2 had been previously mapped (Table 3.7 and Figure 3.2). As predicted, *S. lycopersicum* Cf0 generated amplification products for CT116, *2C*, *Ψ2C,* SNP-nn1 and SNPN (Table 3.7). However, no product was generated for *2B* or *2A* (Table 3.7). Furthermore, *S. pimpinellifolium* CfEcp2 generated amplification products for all markers and genes tested (Table 3.7). In comparison, the predicted deletion mutants, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, no longer gave amplification products for *2A* (Table 3.7). Since the *2A* gene appeared to be missing in the deletion mutants, the expression of *2B* was tested in the deletion mutants and *S. pimpinellifolium* CfEcp2 to ensure that the promoter of *2B* was not affected by the deletion. Expression of *2B* was confirmed in *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179 (Table 3.7).

Table 3.7. Presence and/or absence of genes and markers in *S. lycopersicum* Cf0, *S. pimpinellifolium* CfEcp2 and mutants of *S. pimpinellifolium*.

	The presence (✔) or absence (✖) of amplification products						
Plant ID	CT116	*2C*	*2B*	*2A*	*Ψ2C*	SNP-nn1	SNPN
DNA:							
Cf0	✔	✔	✖	✖	✔	✔	✔
CfEcp2	✔	✔	✔	✔	✔	✔	✔
1178	✔	✔	✔	✖	✔	✔	✔
1179	✔	✔	✔	✖	✔	✔	✔
Expression:							
CfEcp2		✖	✔				
1178			✔				
1179			✔				

Cf0 = *S. lycopersicum* Cf0, CfEcp2 = *S. pimpinellifolium* CfEcp2, 1178 = *S. pimpinellifolium* -/- cfecp2 1178, 1179 = *S. pimpinellifolium* -/- cfecp2 1179. CT116, SNP-nn1 and SNPN = Cleaved amplified polymorphic sequence markers, *2B* = *OR2B*, *2A* = *OR2A* and *2C* = *OR2C*.

3.2.7 Characterisation of responses to *M. fijiensis* variants in *S. pimpinellifolium*

In order to test the hypothesis that the recognition of the effector MfEcp2 from *M. fijiensis* and *C. fulvum* Ecp2 are recognised by the same locus in *S. pimpinellifolium*, the deletion mutants described in section 3.6 were used for recognition scoring.

A selection of MfEcp2 isoforms (MfEcp2-1, MfEcp2-2 and MfEcp2-3) were delivered by PVX-mediated transient expression into *S. lycopersicum* Cf0, *S. pimpinellifolium* CfEcp2 and the *S. pimpinellifolium* deletion mutants 1178 and 1179. As an additional control, *C. fulvum* Ecp2 was delivered by PVX-mediated transient expression to the same lines. *C. fulvum* PVX:*Ecp2* induced a hypersensitive response (HR^+) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant in *S. pimpinellifolium* CfEcp2 (Table 3.8 and Figure 3.12). As expected, *S. lycopersicum* Cf0, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179 failed to show recognition of *C. fulvum* PVX:*Ecp2* (Table 3.8). Whereas *C. fulvum* PVX:*Ecp2* resulted in a HR^+, a weaker HR response (HR^0) manifesting as patchy necrosis was observed with the PVX:*MfEcp2-1 isoform1 (i1)* variant in *S. pimpinellifolium* CfEcp2 and *S. pimpinellifolium* 1178 (Table 3.8 and Figure 3.12). However, no response of PVX:*MfEcp2-1 i1* was seen in *S. lycopersicum* Cf0 or *S. pimpinellifolium* 1179 (Table 3.8). Neither PVX:*MfEcp2-1 i2*, PVX:*MfEcp2-1 i3* or any of the PVX:*MfEcp2-2* and PVX:*MfEcp2-3* isoforms tested were recognised by any of the plant lines tested (Table 3.8 and Figure 3.12). *Note: So far, this experiment has been completed only once. Although a promising result, it is proposed that this is repeated via future studies to exclude random environmental effects.*

Table 3.8. Responses of *S. lycopersicum* and *S. pimpinellifolium* lines to isoforms of *C. fulvum* Ecp2 and *M. fijiensis* Ecp2.

	Plant ID			
Isoform	Cf0	CfEcp2	1178	1179
pBIN:PVX:*Ecp2*	NR	HR^{+}	NR	NR
pSFINX:PVX:*MfEcp2-1 i1*	NR	HR^{0}	HR^{0}	NR
pSFINX:PVX:*MfEcp2-1 i2*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-1 i3*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-2 i1*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-2 i2*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-2 i3*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-2 i4*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i2*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i4*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i7*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i8*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i9*	NR	NR	NR	NR
pSFINX:PVX:*MfEcp2-3 i11*	NR	NR	NR	NR

Cf0 = *S. lycopersicum* Cf0, CfEcp2 = *S. pimpinellifolium* CfEcp2, 1178 = *S. pimpinellifolium* -/- cfecp2 1178, 1179 = *S. pimpinellifolium* -/- cfecp2 1179, NR = no response, HR^{+} = hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, HR^{0} = HR manifesting as patchy necrosis. *Note: So far, this experiment has been completed only once. Although a promising result, it is proposed that this is repeated via future studies to exclude random environmental effects.*

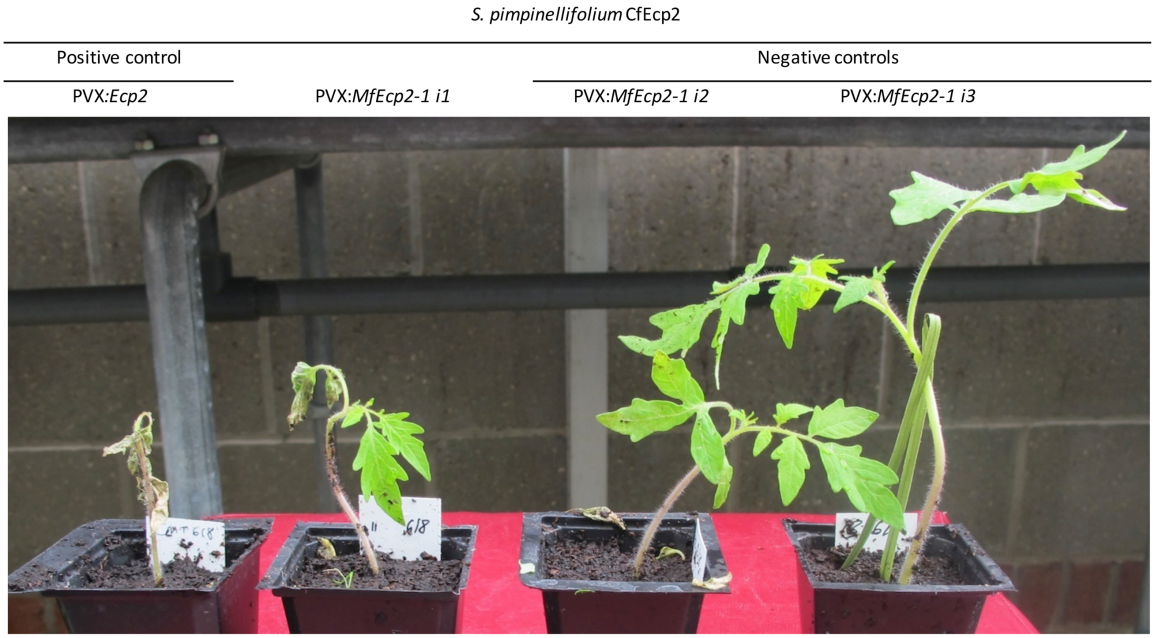


Figure 3.12. PVX mediated-delivery of Ecp2 variants from *C. fulvum* and *M. fijiensis* into *S. pimpinellifolium* CfEcp2.

S. pimpinellifolium inoculated with *C. fulvum Ecp2* and the *M. fijiensis* varietns MfEcp2; MfEcp2-1 i1, MfEcp2-1 i2 and MfEcp2-1 i3. Photographs taken 14 days post inoculation. PVX = Potato Virus X.

3.3 Discussion

Recognition of Ecp2 by CfEcp2 was previously shown to segregate in a 3:1 fashion (resistance to susceptibility) in an F_2 population for the cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 (Lauge et al., 1998). Furthermore, *Cf-Ecp2* was defined as 100% linked to the CAPS marker CT116 and 7.7 cM distal to TG236 and 6.0 cM proximal to TG184 (Haanstra et al., 1999). The data reported in this study supports this and moreover refines the mapping position. Absence of Ecp2 recognition was mapped to a 0.18 cM interval between flanking markers SNPE and GJ44. The interval corresponds to a 40 kb locus in the Heinz reference genome. In addition, presence of *Cf-Ecp2* was mapped to a 0.69 cM interval between flanking markers 60250 and TG24, corresponding to 631 kb in the Heinz genome (Figure 3.2).

Resistance can be genetically linked to a locus. However, defining the genetic component within the locus capable of determining resistance requires sequence analysis of clones e.g. BACs covering the locus followed by testing of candidate genes. The current study describes a *Hcr9* gene cluster comprising of *2C*, *2B*, four *2A* genes (with differences only in the promoter region), a pseudogene of *2A* and a pseudogene of 2C (Figures 3.8 and 3.10). Previously the *OR* locus was found to contain *2C*, *2B* and *2A* (de Kock et al., 2005). However, de Kock et al. (2005) did not generate a full physical sequence spanning a genetically defined map interval for *Cf-Ecp2*.

In order to determine the complete sequence of this locus, three BACs were selected that overlap and span the locus from flanking marker to flanking marker. BACs 11G and 4B were sequenced with 454-technology whilst BAC 7B was sequenced with Illumina 250 bp PE reads, two PacBio runs and two MinION runs. Each sequencing technology used contributed towards the determination of the sequence of the locus, yet each technology had its limitations.

By assembling the relatively short reads from Illumina and the 454 platforms, it was possible to identify the presence of specific markers on the BACs. Furthermore, the assemblies indicated that BAC 11G harboured *2C*, *2B* and *2A* whilst BAC 7B held *2B* and *2A*, and 4B contained neither *2C*, *2B* or *2A*. This was supported by the specific amplification of the homologs from each of the BACs (Figure 3.4). In addition, new *Hcr9s*

were identified from the Illumina and 454 sequence assemblies. *Ψ2A5* was identified on BACs 7B and 11G, whilst *Ψ2C* was identified on BACs 7B and 4B. These data were not previously identified by de Kock et al. (2005). The presence of *Ψ2C* on BACs 7B and 4B accounts for the background PCR products observed for *2C* on these BACs (Figure 3.4). Furthermore, it could be determined that the *OR* locus was in the opposite orientation to that previously indicated (de Kock et al., 2005).

Short-read sequences are potentially insufficient to resolve repetitive structures. This can cause lack of continuity or even errors in the assemblies, either leading to collapse of repeats and thus a mistake in copy number or chimeric contigs, consisting of sequences originating from different loci. In a study by Taudien et al. (2011), for example, four Barley BACs sequenced with 454 Titanium technology were assembled and compared to a manually curated reference. Barley is known for its highly complex repetitive structure, mainly caused by transposable elements (International Barley Genome Sequencing et al., 2012). In total, nine miss-assemblies and eight gaps were recorded.

Due to the similarity of the sequence between the *2A* repeats, and the fact that the reads are shorter than the repeat, the short read sequencing technologies collapsed the repeats into one gene representing the *2A* cluster on both 7B and 11G. The 3' sequence of *2A* is 100% identical to *Ψ2A5* and the 5' sequence of the *2A* gene and the *2A* promoter sequence is 100% identical to *2B* (Figure 3.5)(de Kock et al., 2005). On the BAC 11G, the sequence of *Ψ2A5* is not complete. The SNPs that define the *Ψ2A5* gene from the other homologs can be identified on reads from BAC 11G that contain vector sequence. This indicates that part of the *Ψ2A5* gene is located at the end of the BAC 11G and resulted in the 454 assembly of 11G collapsing the four *2A* and the partial *Ψ2A5* gene into one *2A* gene. The 7B Illumina sequencing assembly, which carries the full sequence of *Ψ2A5,* collapsed the four *2A* genes and one *Ψ2A5* gene into one *Ψ2A5* gene.

Mapping the 454 reads from 11G onto the *OR* locus and mapping the Illumina reads of BAC 7B onto the Illumina 47,096 bp 7B contig revealed a 4 to 5-fold increase in coverage across the *2A* gene (Figure 3.6A-B and 3.7). This indicated multiple copies of *2A* on the BACs 11G and 7B. Manual inspection of SNPs identified between the published *OR* cluster sequence and the BAC 7B reads showed that none of the SNPs were present in more than 10.3 % of the reads (Table 3.2). A single nucleotide variation in one of the four

copies of *2A* would have resulted in a ~25% allele frequency. A 10 % allele frequency, however, is more likely to be noise from miss-mapped reads. Therefore, no unique SNPs were identified between the four copies of *2A* in the Illumina reads.

This collapse of repetitive sequence by short read sequencers is common when the read length is shorter than that of the repeat (Ashton et al., 2015). Although the increase in coverage of sequence over the repeat segment can indicate the copy number of the repeat, it does not define the exact sequence of each of the repeats. Furthermore, an increase in coverage does not necessarily provide conclusive evidence to determine that there are many copies of one section because the process of Illumina and 454 sequencing inherently causes uneven sequence depths (Chin et al., 2014). A higher sequence depth would not resolve the *2A* repeat since the sequence depth from Illumina was >300x and was reduced to 300x for assembly. For 454 reads, it has been shown that approximately 20x coverage is required to assemble BAC sequences, and coverage >20x does not improve the quality of the assembly (Taudien et al., 2011). Therefore, longer read sequencing technology was required to resolve the *Cf-Ecp2* locus.

PacBio sequencing has been reported to be expedient for resolving repetitive sequences. For example, it was used to assemble 99% of the 245 Mb genome of *Oropetium thomaeum* into 625 contigs (Van Buren et al., 2015). Remarkably, the rRNA tandem arrays in this genome, spanning 51 kb and comprising five identical and one partial 9 kb repeat could be reconstructed. As another example, PacBio sequencing has been used to close or extend 55% of gaps present in the human GRCh37 reference genome, many of which included repetitive elements. This added a total of 1,119 kb of sequence to the human reference genome (Chaisson et al., 2015).

In the current study, two SMRT cells of PacBio sequencing on BAC 7B, followed by two independent HGAP assemblies, generated two contigs of 141 kb and 123 kb; named assembly C and assembly D, respectively. Based on pulse field gel electrophoresis, the insert size of BAC 7B was estimated to be 122 Kb (Figure 3.3). Assembly D therefore represented the length of the sequence more accurately (Figure 3.6D). The two PacBio assemblies differed in the number of *Hcr9s* present. Both assemblies confirmed the presence of *2B*, *Ψ2A5* and *Ψ2C* on the BAC (Figure 3.6C and D). However, the copy number of the 8.6 kb *2A* repeat differed between the two assemblies. In addition to

Ψ2A5, assembly C contained six *2A* repeats whilst assembly D contained four *2A* repeats (Figure 3.6C and D).

The nucleotide accuracy for single pass PacBio reads is between 82% and 85% (Koren et al., 2012). A reduced error rate is associated with shorter reads since they pass through the sequencer many times, generating a more accurate consensus read by CCS (Travers et al., 2010). In the assemblies generated for this study, an error correction step utilises short reads to correct the long reads that are used as a template. However, the large error rate of the long reads is far greater than the presence of SNPs between two of the *Hcr9s* in the locus. The genes *2A* and *Ψ2A5* are 94.6% identical (Figure 3.5). It is therefore likely that the error correction step in the PacBio assembly was hampered by the high number of large repeats in the *Hcr9* locus.

To complement the Illumina and PacBIO data, the MinION sequencer from Oxford Nanopore was used to determine the copy number of the 8.6 kb *2A* repeat in BAC 7B (Figure 3.6). High percentage error rates have previously been reported with MinION (Ashton et al., 2015). 2D reads have two-fold redundancy; template strand sequences can be corrected by the associated complement strand sequence. This suggests 2D reads to be preferable for data analysis, since they tend to have a lower percentage error rate than 1D reads. In the sequencing of the genome of *Salmonella* Typhi Haplotype 58 with MinION, the median percentage accuracy (derived from mapping) of template reads was 64.3%, while for complement reads it was 61.6% and for 2D reads it was 71.5% (Ashton et al., 2015). Due to the small percentage of reads that were (i) greater than the size of the *2A* repeat and (ii) mapped to the PacBio assembly, both 2D and 1D reads were used in the current analysis (Figure 3.6E). Because of the high error rates associated with these reads (best homology based local alignments ranging from 65% to 82% identity), dotplot analysis was required to align all MinION reads to the PacBio assemblies since mapping attempts with BWA were unsuccessful (Figure 3.6F-I). Although not attempted in the current study, others have used LAST to map MinION reads to published sequences (Jain et al., 2015). For the purpose of defining copy number of repeats in this study, however, the dotplot analysis was sufficient. Thus, even though MinION reads have extremely high error rates, they were shown to be invaluable for the current study by confirming the correct copy number of repeats. Two MinION reads, 6e2c8f08 and 79a916ec, both confirmed the presence of seven *Hcr9s* on the BAC 7B spanning the entire repeat interval

and matching non-repetitive parts of the locus on both sides. This therefore confirmed that PacBio assembly D reflected the true *Hcr9* copy number on the BAC (Figure 3.6F-I).

Raw MinION reads were previously used to determine repeat numbers on the human chromosome xq24 encompassing a 50 kb assembly gap (Jain et al., 2015). It was determined that eight copies of the cancer-testis gene (CT47) were present in this gap, with each repeat containing the CT47 gene encompassing 4.9 kb. Similarly, MinION has been used as a scaffold to determine the structure and insertion site of a highly repetitive chromosomal antibiotic resistance island in *S. typhi* H58 (Ashton et al., 2015). The Illumina sequencing of strain H125160566 reported one 7 kb region containing a *dihydropteroate synthase* gene. However, MinION sequencing of the same strain indicated that two copies of this gene had collapsed in the Illumina assembly.

Other methods have been cited to demonstrate the use of MinION reads as a scaffold for Illumina sequences. For example, to assemble a 5.18 Mb genome of *Bacteroides fragilis* strain BE1 (Risse et al., 2015). A similar method utilised MinION reads to recruit Illumina reads and perform a local assembly in order to generate Nanopore Synthetic Long (NaS) reads of up to 60 kb in length, which were then used in an assembly (Madoui et al., 2015). However, only 17 % of the original MinION reads became NaS when using this method.

Although the MinION data in the current research was helpful to scaffold the BAC 7B and determine the copy number of *Hcr9s* present, the reads were insufficient to identify the exact sequence for each of the repeats. To resolve large repeats, a sequencing technology would be required that produces reads greater than the length of the repeat and has sequencing errors occurring at a rate less than the informative polymorphisms between the repeats.

The sequencing of the locus carrying the *Hcr9* genes is challenging due to the clustered arrangement of such genes and the similarity of the gene sequences. It was determined in the current study using the MinION reads that PacBio assembly D provided the best representation of the sequence of BAC 7B. However, mapping the Illumina reads onto this assembly highlighted disagreements between Illumina and PacBio due to the formation of gaps where Illumina reads did not map onto the PacBio contig (Figure 3.8A). In the HGAP pipeline, the short reads (<1 kb) were used to correct the longer, more

error-prone reads. This feature is problematic if the repeat is >1 kb because it is not possible to determine which short reads originate from a specific repeat. Therefore, long PacBio reads >15 kb (greater in size than each of the individual 8.6 kb *2A* repeats) were required to manually-edit the PacBio assembly D. When this was completed, the Illumina reads were mapped back onto the edited PacBio assembly, indicating that many of the previously-identified gaps had been resolved (Figure 3.8B). However, three unresolved regions remain within the complete sequence (Figure 3.9). These are associated with homopolymers and microsatallites and do not reside within the *Hcr9* ORFs.

Manually-editing the BAC 7B sequence altered the sequence of the *2A* homologs. The original assembly D sequence contained two 100 % identical copies of *2A*, a copy of *2A* (d2A2) with a 3 bp AAG insertion (at 2,389 bp from the start of the gene) and two pseudo-genes of *2A Ψd2A4* and *Ψ2A5*. However, the manually-edited PacBio assembly indicated that, in addition to *Ψ2A5,* there are four copies of the *2A* gene that are 100% identical within their ORF, differing only at three positions in the promoter region (Figure 3.10). One of these positions is in a homopolymer of 15, 16 or 17 adenines. The other position is where the promoter of *2A4* carries a guanine at 5,844 bp (5' of ATG) in place of a cytosine relative to the other three *2A* genes. The final position is where the promoter of *2A3* carries a homopolymer of only 10 thiamines, whereas the other three *2A* promoters carry 11 thiamines. The promoter of *Ψ2A5* is distinct from the four copies of *2A*, which differ from each other by only a few bp. Interestingly, this conservation of intergenic regions is also seen within the Cf4 haplotype at the *Milky Way* locus in tomato (Parniske et al., 1997). The intergenic region between the genes *Hcr9-4C* and *Hcr9-4D* is almost identical to the region between *Hcr9-4A* and *Hcr9-4B*. In comparison, all intergenic regions within the Cf9 haplotype are distinct from each other (Parniske et al., 1997).

In the current research, further analysis was undertaken on BAC 7B to confirm that other alleles of *2A* and *Ψ2A5* were absent. *In silico* digestion of the manually-edited sequence with *Eco*R1 confirmed a match to the *Eco*R1 digested BAC (Figure 3.3). Furthermore, sequence analysis of multiple PCR clones of *2A* and *Ψ2A5* confirmed that PacBio assembly D had, indeed, incorrectly assembled two of the homologs of *2A* (*d2A2* and *Ψd2A4*). In fact, one MinION read also confirmed the absence of the *Ψ2A4* sequence from BAC 7B, since the read suggested that the two genes preceding *Ψ2A5* were more

similar to *2A* than *Ψd2A4* (Table 3.6). Besides, the AAG insertion at 2,389 bp, associated with *d2A2*, is actually present in the sequence of *Ψ2C*. This was identified when the homologs were aligned (Figure 3.5). In addition, when comparing the sequence of *2B*, *2A* and *Ψ2A4*, it can be seen that *Ψ2A4* is a combination of SNPs between *2B* and *2A* (Figure 3.5). These SNPs may, therefore, have been misplaced in the sequence due to inaccurate HGAP assembly and an error correction of the PacBio reads. PacBio assembly D did, however, correctly assemble *Ψ2A5* (Figure 3.5 and 3.6D).

In tomato, the arrangement in clusters is common for genes encoding for RLPs, including *Hcr9s* and *Hcr2s* (Andolfo et al., 2013). For example, a total of five *Hcr9s* can be present at the *MW* locus, on the short arm of chromosome 1, depending on the genotype (Kruijt et al., 2004; Parniske et al., 1997; Parniske and Jones, 1999). The tandemly arranged RLP-encoding genes in the Cf4 and Cf9 haplotypes sit in the same transcriptional orientation as each other (Thomas et al., 1998). This is also seen with the *Hcr9s* at the *OR* locus (Figure 3.8B). Similarly, many *R* genes in other species sit in clusters. This has been observed by the sequencing of whole *R* gene analogue complements in plant species (Andolfo et al., 2013; Jupe et al., 2012; McHale et al., 2009; Meyers et al., 2003; Monosi et al., 2004).

In addition to the complete *Hcr9s* identified at the *OR* locus, two pseudo-genes were also identified; *Ψ2A5* and *Ψ2C*. Similarly, at the *NL* locus in *S. lycopersicum*, two pseudo-genes are present (*ΨNL0A* and *ΨNL0B*) alongside three complete *Hcr9s* (*NL0C*, *NL0D* and *NL0E*)(Parniske and Jones, 1999). However, within both the Cf9 and Cf4 haplotypes at the *MW* locus pseudo-genes have not been identified (Parniske et al., 1997). Some truncated genes have been shown to be functional. For example, the *Xa21D* gene, which lacks the transmembrane and cytoplasmic domains of the receptor-like kinase *Xa21* gene, is able to confer partial resistance to *Xanthomonas oryzae* pv *oryzae* (Wang et al., 1998). However, according to the amino-acid prediction, pseudo-genes *Ψ2A5* and *Ψ2C* would produce severely truncated peptides, implying that they most likely do not contribute to resistance.

The sequence conservation of the 8.6 kb repeat sequence suggests recent duplication by illegitimus recombination of the *2A* gene. It has been proposed that *R* genes evolve by duplication followed by diversification (Michelmore and Meyers, 1998). There are two

100% identical copies of the gene *9DC* lying in tandem within a cluster of five *Hcr9s* (Kruijt et al., 2004). In addition, there are two almost-identical copies of *Cf-2* sitting in tandem on the short arm of chromosome 6 within a cluster of three *Hcr2s (Dixon et al., 1996)*. However, the presence of four tandem-duplicated 100% identical copies of an *Hcr9*, as outlined in the current study, is considered to be unique at this present time.

To aid in the identification of candidate genes for *Cf-Ecp2*, two deletion mutants, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, were generated in the current study, using the transposon *Ds FT33.* This transposon is genetically linked to the *Cf-Ecp2* locus. It is also located 3 cM distal to *Cf-9*, used previously to tag and clone the *Cf-9* and *Cf-4E* genes (Jones et al., 1994; Takken et al., 1998).

To clone *Cf-9,* 118 survivors were recovered from 160,000 testcross progeny derived from the cross between plants heterozygous for *Ds* and *sAC* element (required for activation of *Ds*) but homozygous for *Cf-9* with plants homozygous for the *Avr9* transgene but lacking *Cf-9*. The survivors included 37 independent *Ds* insertions into the gene encoding for *Cf-9.* This enabled isolation of the gene (Jones et al., 1994). In the current study, only two survivors were identified in a cross between a *S. pimpinellifolium* plant heterozygous for *Ds* and *sAc* but homozygous for *Cf-Ecp2* with a *S. lycospersicum* plant homozygous for the *Ecp2* transgene but lacking *Cf-Ecp2*. A 10-fold difference in survival was therefore seen between the *Cf-9* and *Cf-Ecp2* transposon tagging experiments (0.07% survivors compared to 0.008% survivors). This may be attributed to either the *FT33* gene being in a closer genetic location to *Cf-9* than *Cf-Ecp2* or due to the presence of more than one active copy of the *Cf-Ecp2* gene. The latter proposal is supported by the fact that the two survivors generated in this experiment were deletion mutants and no genes were tagged with *Ds*, in contrast to the transposon tagging experiment of *Cf-9* (Jones et al., 1994).

In the current study, the *S. pimpinellifolium cf-ecp2* deletion mutants, that had lost the ability to recognise Ecp2, had also lost the presence of *2A* within the genome. However, the deletion mutants retained the presence of *2C* and *2B* (Table 3.7). The expression of *2B* was also retained in these two mutants, thus the promoter of *2B* is still functional (Table 3.7). *2B* is therefore not a candidate for *Cf-Ecp2*.

Furthermore, *2C* is known not to be expressed in *S. pimpinellifolium* CfEcp2 whilst *2B* and *2A* are expressed (de Kock and colleagues, 2004). *2C* is therefore not a candidate for *Cf-Ecp2*. In this study it was established that there are four copies of *2A* present in tandem at the *OR* locus of *S. pimpinellifolium* CfEcp2. Therefore, deletion of all copies of *2A* from the genome, in both deletion mutants generated in the transposon tagging experiment, led to the loss of recognition of Ecp2. This observation supports the role of *2A* as a requirement for *Cf-Ecp2*–mediated recognition of Ecp2.

Stable transformation of *2A*, *2B* or *2C* with 1 kb of their native promoter into *S. lycopersicum* Cf0 does not enable recognition of Ecp2 (de Kock and colleagues, 2004). However, a conserved region of 576 bp in the promoter of *2A* and *2B* (1.5 kb upstream of the translation start codon) is also present in the promoter region of the tomato *R* gene *Hero* (de Kock and colleagues, 2004; Ernst et al., 2002). This gene confers resistance to potato cyst nematodes (Ernst et al., 2002). This conserved element may be important for coordinated transcriptional activity of the *R* genes. This conserved element would be missing from such transgenic lines (de Kock et al., 2005). In addition, mRNA was not detected for *2C*, *2B* and *2A* in any of these transgenic lines (de Kock and colleagues, 2004).

In general, genes are not randomly distributed throughout a genome, but reside in 'functional neighbourhoods' (Al-Shahrour et al., 2010; Andolfo et al., 2013). Some resistance loci in tomato have been shown to encompass a mixture of different types of resistance genes (Andolfo et al., 2013). One locus can carry different resistance gene structures, which are both required for the recognition of a specific effector (Martin et al., 1993; Salmeron et al., 1996). For example, the *Prf NB-LRR* gene is located within a cluster of five protein kinases, including the *Pto* gene encoding for a cytoplasmic serine-threonine kinase (Martin et al., 1993; Salmeron et al., 1996). *Prf* is required for the recognition of the *P. syringae* effector *AvrPto* by *Pto* (Salmeron et al., 1996). Many *NB-LRR* genes are located head-to-head within an *R* gene locus and are both required for recognition of an effector. For example, the *NB-LRRs RRS1* and *RPS4* in *A. thaliana* work together to recognise effectors from three separate pathogens: *Collectotrichum higginsianum*, *P. syringae* pv tomato strain DC3000 and *Ralstonia solanacearum* 1 (Narusaka et al., 2009a; Narusaka et al., 2009b). Genes that have evolved as a functional

unit are less likely to be separated by recombination if they are closely linked, thus tight linkage of genes required for resistance ensuring joint inheritance.

It is possible that more than one of the *Hcr9s* present at the *OR* locus is required for recognition of Ecp2. In such a case, deletion of one of these copies i.e. *2A*, in the mutants 1178 and 1179, would still lead to loss of recognition as both genes are required for the response. The *R* gene candidates 35S:*2C,* 35S:*2B* and 35S:*2A* were individually or in combinations transiently co-expressed in *Nicotiana* sp. together with 35S:*Ecp2* (de Kock and colleagues, 2004). However, none of these combinations gave the HR observed when 35S:*Cf-4* was transiently-transformed alongside 35S:*Avr4* in the same *Nicotiana* sp (de Kock and colleagues, 2004). This suggests that either *2A* is not *Cf-Ecp2*, or that the *Nicotiana* species used are not compatible for the resistance signalling encoded for by *2A*. The latter would mean that other components of the signalling pathway need to be transferred alongside *2A* for functionality in *Nicotiana* sp.. Indeed, transient expression of the *A. thaliana NB-LRR* gene *RPM1* requires the co-expression of RIN4 in *N. benthaminana* to mediate recognition of *P. syrinage* effectors AvrB and AvrRpm1 and instigate a HR (Chung et al., 2011).

In addition to recognition of Ecp2, the *OR* locus *Hcr9* genes may encode for recognition of other effectors. The tandem arrangement of repeat elements promotes recombination between the homologous sequences, generating new specificities (Parniske et al., 1997). Some *R* gene loci encode for the recognition of multiple effectors from the same pathogen. For example, recognition of the *Phytophthora infestans* effectors Avr3a and Avr3b is encoded for by the *NB-LRR*s *R3a* and *R3b,* respectively. These two genes are located in close proximity on chromosome 11 in the potato genome (Huang et al., 2004; Huang et al., 2005; Li et al., 2011). The Cf9 haplotype of the *MW* locus carries the *Cf-9* and *Cf9-B* genes, capable of recognising the effectors Avr9 and Avr9B, respectively (Jones et al., 1994; Panter et al., 2002). Similarly, the Cf4 haplotype of the *MW* locus carries the *Cf-4* gene and *Cf-4E* gene, with the capacity to recognise the effectors Avr4 and Avr4E, respectively (Takken et al., 1998; Takken et al., 1999; Thomas et al., 1997). Although, recognition of Ecp3 encoded by the *Cf-Ecp3* gene has been mapped to the *OR* locus, recognition of Ecp3 is not conferred by *S. pimpinellifolium* CfEcp2 (Lauge et al., 2000). Therefore, *Cf-Ecp3* is found within a different haplotype of the *OR* locus (de Kock et al.,

2005). However, this does not exclude the principle that other effectors, from either *C. fulvum* or other pathogens of tomato, are recognised by the RLPs positioned within the *OR* locus as defined in the current research.

Stergiopoulos et al. (2010) identified three homologs of the *C. fulvum* Ecp2 effector present in *M. fijiensis,* named MfEcp2-1, MfEcp2-2 and MfEcp2-3. The fungus, *M. fijiensis,* causes the devastating Black Sigatoka disease in banana (Churchill, 2011). The common banana grown for human consumption is the Cavendish banana, which is susceptible to Black Sigatoka (Churchill, 2011). It has been confirmed that *S. pimpinellifolium* CfEcp2 is able to recognise the effector MfEcp2-1 (Stergiopoulos et al., 2010). However, *S. pimpinellifolium* CfEcp2 did not recognise the *M. fijiensis* homologs MfEcp2-2 or MfEcp2-3 (Stergiopoulos et al., 2010).

Furthermore, exploration of the variation in amino acid sequences of MfEcp2 from 34 strains of *M. fijiensis* (from South East Asia, Africa and Latin America) enabled the identification of 4 isoforms of MfEcp2-1 (MfEcp2-1 i1-4) (all originating in strains from South East Asia), four isoforms of MfEcp2-2 (MfEcp2-2 i1-4) (all originating in strains from South East Asia) and 11 isoforms of MfEcp2-3 (MfEcp2-3 i1-11) (Stergiopoulos et al., 2014). It was shown that *S. pimpinellifolium* CfEcp2 could not recognise isoforms of MfEcp2-2 and MfEcp2-3. In addition, *S. pimpinellifolium* CfEcp2 failed to recognise MfEcp2-1 i2 and MfEcp2-1 i2 but there was a HR-associated recognition of MfEcp2-1 i1, as observed previously by Stergiopoulos et al. (2010). There was also a weaker form of HR associated with the recognition of MfEcp2-1 i4 (Stergiopoulos et al., 2014).

The results of the current study, in respect of *S. pimpinellifolium* CfEcp2-recognition or absence of recognition of the MfEcp2-1, MfEcp2-2 and MfEcp2-3 isoforms tested, support earlier results (Stergiopoulos et al., 2014) (Table 3.8 and Figure 3.12). In addition, the deletion mutant *S. pimpinellifolium* 1179 had lost the ability to recognise the variant MfEcp2-1 (Table 3.8). This supports the hypothesis that the genetic component required for both Ecp2-recognition and MfEcp2-recognition is lost in the deletion mutant 1179. Furthermore, this indicates that the genetic component required for recognition of Ecp2 and MfEcp2 in *S. pimpinellifolium* CfEcp2 is encoded by the same genetic locus. However, the deletion mutant *S. pimpinellifolium* 1178 also showed weak necrosis when inoculated with MfEcp2-1, in comparison to a lack of response when inoculated with Ecp2. The

recognition of these two variants may therefore be encoded by different sections of this genetic locus. However, it is considered that repeats of this experiment are required to give further support to these conclusions. Therefore, it is still assumed that the same gene encodes for recognition of Ecp2 and MfEcp2.

The *S. pimpinellifolium Cf-Ecp2 R* gene conferring recognition of the effector MfEcp2-1 can potentially be transformed into Cavendish banana. To utilise MfEcp2-recognition by *S. pimpinelllifolium* CfEcp2 in Cavendish banana, the population dynamics of *M. fijiensis* in the geographical area where the transgenic banana will be grown must first be understood. Individuals of the fungus in the population that lack those isoforms of MfEcp2 recognised by the *Cf-Ecp2* gene will be able to prosper in the population and the durability of the resistance will be short. However, if all individuals carry the isoforms of MfEcp2 recognised by the *Cf-Ecp2* gene, then the durability of the resistance will be as long as it takes for the pathogen to mutate or delete the recognised MfEcp2 isoform. The use of individual *R* genes in a crop variety are generally not durable (Rommens and Kishore, 2000). For example, monogenic Tobacco Mosaic virus (TMV) resistant varieties of tomato were cultivated from 1966 onwards in Britain (Pelham et al., 1970). Sampling of TMV strains before (1966) and after (1967 and 1968) this introduction, showed the increase in Strain 1 after 1966, which was able to infest the resistant varieties that had been introduced (Pelham et al., 1970).

It is known that Cf4 of tomato can recognise *M. fijiensis* effector MfAvr4 (Stergiopoulos et al., 2010). Co-transformation of CfEcp2 and Cf4 into the Cavendish banana would be beneficial because it would increase the durability of resistance against *M. fijiensis*. This is because the pathogen would have to mutate or lose both effectors (McDonald and Linde, 2002a, b). The loss of both effectors may affect the fitness of the pathogen. Deletions of Ecp2 from *C. fulvum*, cause reduced pathogenicity of the fungus on tomato since the fungus fails to produce spores (Lauge et al., 1997). On the other hand, *Avr4* can be mutated in *C. fulvum*, overcoming Cf4 mediated resistance in tomato, with little effect on the fitness of the pathogen (Joosten et al., 1994).

It was observed in the current study that recognition of MfEcp2 by *S. pimpinellifolium* CfEcp2 is weaker than that seen in recognition of Ecp2 (Table 3.8 and Figure 3.12). Since the amino acid identity between MfEcp2 and Ecp2 is only 57%, it could mean that *S.*

pimpinellifolium CfEcp2 has evolved to specifically recognise *C. fulvum* Ecp2 (Stergiopoulos et al., 2010). This needs to be taken into consideration if *Cf-Ecp2* is to be utilised for resistance to *M. fijiensis*. The optimisation of recognition by mutation of the gene may first be required. Such an approach has been shown before. For example, artificial evolution of the plant resistance gene *R3a* enabled recognition of a virulent form of AVR3a that was not recognised by the wild-type *R* gene (Chapman et al., 2014).

In the current study, a locus required for the recognition of Ecp2 was mapped relative to a 40 kb region in the sequenced Heinz tomato genome. Sequence analysis of three BACs covering this region identified eight *Hcr9s*; *2C*, *2B*, four 100% identical copies of *2A*, *Ψ2A5* and *Ψ2C*. As a consequence of present inadequacies of current sequencing technologies, a combination of Illumina, PacBio and MinION technologies were required to resolve this complex sequence. In the future, if the error rate linked to MinION technologies is reduced or if the length of PacBio reads increases, MinION and PacBio will become increasingly exploited to resolve super-complex repetitive loci, including plant *R* gene loci. In addition, analysis of two independent deletion mutants, which have lost the ability to recognise Ecp2, indicated that *2A* is *Cf-Ecp2*. Furthermore, *Cf-Ecp2* recognises the MfEcp2-1 isoform 1 but not isoforms 2 and 3. It is considered that we are now one step nearer to cloning the resistance gene encoding *Cf-Ecp2*, which may provide resistance to multiple crop pathogens carrying the effector Ecp2.

Chapter 4

A receptor-like protein confers partial recognition of Ecp2

4.1 Introduction

Plants harbour many different species of microorganisms. Few of the microorganisms living on and within plants are pathogenic (Jones and Dangl, 2006). Pathogenic species are able to overcome plant defences and colonise plant tissues. To enable their pathogenicity, microbes secrete effectors (Jones and Dangl, 2006). Some plants, however, have evolved *R* genes which recognise specific pathogenic effectors and trigger a defence response within the plant. This response allows the plant to be resistant to the attacking pathogen (Jones and Dangl, 2006).

Flor (1971) proposed that *R* genes and Avr genes work in a gene-for-gene manner. This model suggests that a single dominant *R* gene recognises one effector from a pathogen. Although, most *R* gene-mediated resistance does seem to be inherited in a single dominant manner, in many cases it is becoming apparent that the underlying mechanism is more complicated than a simple gene-for-gene model.

Some proteins encoded by *R* genes directly interact with their corresponding effector (Jones and Dangl, 2006). The flax rust fungus AvrL567 and AvrM effectors directly interact in yeast-two-hybrid assays with the NB-LRR proteins encoded by the Flax *L5*/*L6*/*L7* and *M* genes, respectively (Dodds et al., 2006). Likewise, the rice blast fungus (*Magnaporthe oryzae*) effector Avr-Pita also directly interacts with the protein encoded for by the rice *NB-LRR* gene *Pi-ta* (Jia et al., 2000).

Effectors target host processes. The rice blast effector AvrPiz-t, targets APIP6 in rice (Park et al., 2012). APIP6 is a RING E3 ubiquitin ligase involved in ROS production, in response to PAMP recognition (Park et al., 2012). Targeting of APIP6 by AvrPiz-t acts to suppress PTI in plants (Park et al., 2012). Some effectors are secreted during infection function to protect the fungal hyphae from degradation. This was observed for the tomato pathogen *C. fulvum* (de Jonge et al., 2010; Sanchez-Vallet et al., 2013; Westerink et al., 2002). Avr4, for example, binds to the chitin within the fungal cell wall and protects it from degradation by plant chitinases (Westerink et al., 2002). Another function, as

hypothesised for Ecp6, is to sequester pieces of fungal chitin (de Jonge et al., 2010). The pieces are released by plant chitinase breakdown of the fungal cell wall within the apoplastic space. This sequestering stops the fungus from triggering a defence response in plants (de Jonge et al., 2010; Sanchez-Vallet et al., 2013). Other effectors work to target specific plant processes. Avr2 is a cysteine protease inhibitor, which binds and inhibits the chitin proteases Rcr3 and Pip1 (Van Esse et al., 2008). Furthermore, Avr2 is a competitive inhibitor of plant cysteine protease inhibitor e67 which binds the active site of Rcr3 (Rooney et al., 2005). Similarly, the Oomycete pathogen *Phytopthora infestans* secretes the effectors EPIC1 and EPIC2B which also target Rcr3 in tomato (Song et al., 2009).

Host effector targets can be monitored by *R* genes products. This is known as the guard hypothesis (Dangl and Jones, 2001; Van der Biezen and Jones, 1998). The tomato *R* gene *Cf-2* encodes an RLP (Dixon et al., 1996). Cf-2 recognises the modification of the effector target, Rcr3, by the *C. fulvum* effector Avr2 (Rooney et al., 2005). In the absence of Rcr3, Cf-2 is not activated. In addition, some effector targets may be genetically linked to the *R* gene, which monitors them and are thus inherited as a unit (Salmeron et al., 1996).

Some *R* genes work in pairs to recognise multiple effectors (Cesari et al., 2014). These pairs of *R* genes can sit side by side in the genome and are therefore inherited as a single dominant gene in a Mendelian fashion (Cesari et al., 2014).

Other components may be required to interact with *R* gene products to trigger a resistance response upon recognition of the effector. RLPs are composed of a leucine-rich-repeat domain followed by a transmembrane domain, which is flanked by an acidic domain (extracellular) and basic domain (intracellular). The presence of this short cytoplasmic C-terminal tail in RLPs prompted the expectation that another component was required for signalling (Jones et al., 1994). For Cf4, Suppressor Of BIR1-1/Evershed (SOBIR1) is required for stabilisation of the *R* gene product (Liebrand et al., 2013; Postma et al., 2015). Furthermore, BRASSINOSTEROID INSENSITIVE 1 – associated receptor kinase 1 (BAK1) is required for Avr4–Cf-4 mediated resistance (Postma et al., 2015).

It is believed that the signalling components of *R* genes, post effector/elicitor perception, are conserved across plant species. This enables movement of functional *R* genes (Wulff

et al., 2011). For example, the Arabidopsis RLK EFR recognises bacterial EF-TU and signals recognition via its kinase domain (Lin et al., 2014; Zhang et al., 2010; Zipfel et al., 2006). When the EFR extracellular LRR domain is bound to the kinase domain of rice RLK *Xa21*, signalling of recognition of EF-TU still occurs in the Arabidopsis plant (Holton et al., 2015). The Xa21 kinase domain of rice is therefore functional in Arabidopsis. This indicates that the downstream-signalling components of *R* genes encoding RLKs in both monocots and dicots is conserved.

C. fulvum is a pathogen of the tomato plant (Joosten and de Wit, 1999). The fungus enters the plant via the stomata and its hyphae grow between the cells (Joosten and de Wit, 1999). *C. fulvum* does not form haustoria or any other intracellular invasions. Whilst in the apoplastic space, the fungus secretes proteinaceous effectors, which enable its pathogenicity (Joosten and de Wit, 1999). *R* genes have been identified in species of wild tomato that enable recognition of specific *C. fulvum* effectors (de Wit et al., 1994; Lauge et al., 1998). All *R* genes that are so far cloned from tomato and capable of recognising effectors from *C. fulvum* are RLPs. These RLPs sit on the surface of the plant cell and monitor the external apoplastic space (de Wit et al., 1994).

In currant tomato (*S. pimpinellifolium*), recognition of the *C. fulvum* effector Ecp2, which results in a HR, is encoded by a single dominant gene, *Cf-Ecp2* (Lauge et al., 1998). The non-host *N. paniculata* is also able to recognise the effector Ecp2 and trigger a HR (Lauge et al., 2000). The F_2 progeny from the cross between *N. paniculata* TW99 (recognises Ecp2) and *N. paniculta* TW102 (no recognition of Ecp2) segregate for recognition:no recognition of Ecp2 in a 3:1 dependent manner (de Kock et al., 2004). This indicates that *N. paniculta*-recognition of Ecp2 is also mediated by a single dominant gene (de Kock et al., 2004).

Cf-Ecp2-mediated recognition of Ecp2 was previously mapped to the *OR* locus on the short arm of chromosome 1 in *S. pimpinellifolium* CfEcp2 (Haanstra et al., 1999). The *Cf-Ecp2* locus was further genetically delimited to a 0.18 cM region between markers SNPE and GJ44 (Chapter 3) corresponding to a 40 kb region relative to the sequenced Heinz tomato genome. Generation of a physical sequenced contig across this 0.18 cM region revealed an 91 kb region harbouring eight *Hcr9*s including *2C*, *2B*, four copies of *2A* that are 100% identical, *Ψ2A5* and *Ψ2C* (Chapter 3). Two independent deletion mutants were

generated by a transposon tagging experiment in the *S. pimpinellifolium* CfEcp2 genetic background (Chapter 3). These deletion mutants, named *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179, have lost the ability to recognise Ecp2. In addition, they were found to retain *2C* and *2B*, and yet have lost *2A* from their genomes. These deletion mutants support *2A* as a candidate for *Cf-Ecp2*. Although, loss of recognition of Ecp2 indicates that a component is required for Ecp2 recongition, it does not mean it is the only factor required. The map position for the presence of Ecp2-recognition (a 631 kb region relative to the publicly- available *S. lycopersicon* Heinz tomato genome sequence) indicates, however, that all components required for recognition of Ecp2 lie in this region.

The Ecp2 effector is conserved across the fungal class *Dothidiomycetes* (Stergiopoulos et al., 2012; Stergiopoulos et al., 2010). Ecp2 is present not only in many fungal plant pathogens but human pathogens and saprophytic fungi (Stergiopoulos et al., 2012). One of the important crop pathogens that contain Ecp2 homologs is *M. fijiensis*, which causes Black Sigatoka disease in banana. Cloning *Cf-Ecp2* may therefore enable the transformation of banana with *Cf-Ecp2* to generate resistance to *M. fijiensis*. In addition, it was found that Ecp2 is under-diversifying selection within *M. fijiensis*, suggesting that evading recognition via loss of function would incur a virulence penalty (Stergiopoulos et al., 2014).

If an effector is important to the fungus it is assumed that it will not be lost easily. Other effectors, e.g. *Avr9* and *Avr4,* have been lost or mutated by the pathogen, enabling the pathogen to overcome the resistance encoded by plants that recognise these specific effectors (Joosten et al., 1994; Marmeisse et al., 1993). Deletion of Ecp2 from the genome of *C. fulvum* causes a reduction in virulence of the pathogen (Lauge et al., 1997). The role of Ecp2 as a virulence factor, its conservation across multiple Dothiodimycetes and the fact it is under-diversifying selection suggests the role of Ecp2 as a core effector. This implies that resistance encoded by *Cf-Ecp2* may be more durable than resistance encoded by an *R* gene that recognises an effector that can easily be lost by the pathogen.

To move resistance into other species, the underlying mechanism of resistance encoded for by the *R* gene must be understood. Furthermore, it must be determined if this mechanism is conserved in the receiving plant species. In the context of the current study,

the requirement of additional factors for recognition of Ecp2 must be taken into account if *Cf-Ecp2* is to be transferred from tomato into another crop species such as banana.

When *R* gene candidates against *C. fulvum* are identified in tomato, their function can be determined by their transgenic expression in *S. lycopersicum* Cf0, which lacks functional *R* genes against *C. fulvum* (de Kock and colleagues, 2004; Thomas et al., 1997). These transgenic plants can then be characterised by determining their ability to recognise the corresponding effector. With RLPs, this characterisation can be completed by infiltration of pure effector protein in the leaves of the transgenic lines since RLPs can recognise effectors in the apoplast (de Kock and colleagues, 2004; Joosten et al., 1994). Infiltration of Ecp2 into *S. pimpinellifolim* CfEcp2 triggers a necrosis in the infiltrated section (de Kock and colleagues, 2004). This is not observed when Ecp2 protein is infiltrated into *S. lycopersicum* Cf0 (de Kock and colleagues, 2004). In addition, effectors can be expressed in a transient manner in tomato, using the PVX as a vector (Chapman et al., 1992). The PVX is able to spread the effector systemically. When the *R* gene is present, systemic recognition and HR occurs, causing death (Chapman et al., 1992; Joosten et al., 1997; Thomas et al., 1997). When *S. pimpinellifolium* CfEcp2 plants are inoculated with PVX:*Ecp2* a systemic HR is observed (de Kock and colleagues, 2004; Haanstra et al., 1999; Soumpourou et al., 2007). In contrast, when *S. lycopersicum* Cf0 plants are inoculated with PVX:*Ecp2* a mosaic phenotype is seen, characteristic of PVX infection (de Kock and colleagues, 2004; Haanstra et al., 1999).

The sustained delivery of effectors into tomato plants can be ensured by crossing the plant carrying the *R* gene into a tomato plant which is over-expressing the corresponding effector (Hammond-Kosack et al., 1994a; Thomas et al., 1997). If there is recognition of the effector by the *R* gene, an SLP is observed. When *S. pimpinellifolium* CfEcp2 is crossed to *S. lycopersicum* Cf0 overexpressing Ecp2, via the 35S promoter, SLP is observed in all progeny (Soumpourou et al., 2007).

Previously, stable transgenic *S. lycopersicum* Cf0 plants were generated containing the *Cf-Ecp2* candidate gene *2A* with 1 kb of 5' native regulatory sequence (de Kock and colleagues, 2004). However, these transgenic lines did not respond to Ecp2 protein infiltration or PVX-mediated delivery of Ecp2 (de Kock and colleagues, 2004). This lack of response may be explained by the presence of a conserved promoter element 1.5 kb 5'

of *2A* and *2B*, which is hypothesised to have a role in concerted control of *R* gene expression (de Kock and colleagues, 2004). The use of only 1 kb of native 5' regulatory sequence of *2A* excludes this element and may therefore affect the functional expression of the *R* gene. Indeed, no mRNA transcripts for *2A* were identified in these stable transformants (de Kock and colleagues, 2004). Consequently, in the current study *S. lycopersicum* Cf0 plants were stably transformed with a construct, in which *2A* expression is driven by the Cauliflower Mosaic Virus (CaMV) constitutive 35S RNA promoter. Subsequently, the transgenics were characterised for their ability to recognise Ecp2. The transgenics did not appear to respond to Ecp2 protein infiltration into leaves or cotyledons, while PVX-mediated delivery of Ecp2 lacked penetrance. However, sustained delivery of Ecp2, by crossing the 35S:*2A* stable transgenic lines to transgenic lines over-expressing Ecp2, resulted in SLP. This correlated with the presence of the 35S:*2A* transgene. As a result, it is proposed that the over-expression of *2A* via the 35S promoter interferes with its function.

4.2 Results

4.2.1 Characterisation of *S. lycopersicum* Cf0 candidates for stable over- expression of *2A*

The candidate gene encoding *Cf-Ecp2* mediated recognition of Ecp2, *2A*, was stably transformed into *S. lycopersicum* Cf0. A binary construct was generated by Golden Gate cloning, in which *2A* expression is driven by the CaMV 35S promoter to induce constitutive over-expression of the gene in the transformed plant. The clone was transformed into *A. tumefaciens* strain Agl1 for stable transformation of *S. lycopersicum* Cf0. A total of 13 kanamycin-resistant plants were selected as candidates for *S. lycopersicum* Cf0 stably over expressing *2A*. In addition, the 35S:*2A* clone was transformed into *A. tumefaciens* GV3101 for transient assays in *N. paniculata*.

4.2.1.i Response of 35S:*2A* candidates to Ecp2 protein

To test the hypothesis that *2A* is *Cf-Ecp2*, the ability of the transgenic plants to recognise Ecp2 was tested. Thirteen candidate T_1 plants (*S. lycopersicum* Cf0 carrying 35S:*2A*), *N. paniculata* TW99 (known to recognise Ecp2) and *S. lycopersicum* Cf0 were infiltrated with 15 µM Ecp2 protein (in A4 buffer) alongside 75-fold diluted A4 buffer alone (Table 4.1). Adult leaves were inflitrated (Table 4.1). *N. paniculata* TW99 responded to the Ecp2 protein with a strong HR^+ in the infiltrated leaf section, but not to A4 buffer (Table 4.1). This HR was termed HR^+ to represent necrosis with complete death. The negative control, *S. lycopersicum* Cf0, did not respond in three out of the four sections infiltrated with Ecp2 protein (Table 4.1). However, in each of these lines one of the sections infiltrated with Ecp2 protein showed a response in the form of HR^0 (Table 4.1). *S. lycopersicum* Cf0 did not respond to A4 buffer.

Out of the 13 candidate T_1 plants infiltrated, eight showed a HR^0 to Ecp2 protein-infiltration in at least one of the leaf sections infiltrated (Table 4.1). This was scored as a HR^0 response rather than HR^+ since it is only necrosis and no death was observed. HR^0 was observed to develop slower than HR^+. However, the response was inconsistent since six out of these eight plants failed to respond to Ecp2 protein-infiltration in other leaf sections that were infiltrated (Table 4.1).

No response to A4 buffer was observed in 11 out of the 13 T_1 candidate plants (Table 4.1). The two remaining plants showed a response to A4 buffer in at least one of their infiltrated leaf sections. In the plants *S. lycopersicum* Cf0 35S:*2A*p3, 35S:*2A*p4, and 35S:*2A*p24 four out of five, four out of four, or five out of five of the sections infiltrated, respectively, with Ecp2 protein showed a HR^0. However, none of the three to four sections infiltrated with A4 buffer showed a response in these plants (Table 4.1). These three lines (*2A*p3, *2A*p4 and *2A*p24) were selected for further detailed analysis, including in some cases consideration of some additional lines to encompass extended responses.

Table 4.1. *S. lycopersicum* Cf0 T_1 candidates for stable expression of 35S:*2A*, infiltrated with Ecp2 protein in their adult leaves.

	15 µM Ecp2			A4 buffer		
Plant line	HR	NR	Total[1]	HR	NR	Total[1]
N. paniculata TW99 (CfEcp2)	1[2]	0	1	0	1	1
S. lycopersicum Cf0	1[3]	3	4	0	2	2
S. lycopersicum Cf0 35S:*2A*p1	0	4	4	0	3	3
S. lycopersicum Cf0 35S:*2A*p2	0	4	4	0	2	2
S. lycopersicum Cf0 35S:*2A*p3	4[3]	1	5	0	4	4
S. lycopersicum Cf0 35S:*2A*p4	4[3]	0	4	0	3	3
S. lycopersicum Cf0 35S:*2A*p5	1[3]	4	5	0	4	4
S. lycopersicum Cf0 35S:*2A*p10	2[3]	3	5	2[3]	2	4
S. lycopersicum Cf0 35S:*2A*p13	0	5	5	0	4	4
S. lycopersicum Cf0 35S:*2A*p14	3[3]	2	5	1[3]	3	4
S. lycopersicum Cf0 35S:*2A*p16	1[3]	3	4	0	3	3
S. lycopersicum Cf0 35S:*2A*p18	1[3]	3	4	0	3	3
S. lycopersicum Cf0 35S:*2A*p24	5[3]	0	5	0	4	4
S. lycopersicum Cf0 35S:*2A*p26	0	4	4	0	4	4
S. lycopersicum Cf0 35S:*2A*p28	0	4	4	0	3	3

Total[1] = Total number of leaf sections infiltrated, NR = no response, [2] =HR^+ = hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [3]= HR^0 = HR manifesting as patchy necrosis, *2A* = *OR2A*.

In addition to the infiltration of Ecp2 protein into T_1 adult leaves, the cotyledons of 11 day old selected T_2 families (progeny of T_1 plants *S. lycopersicum* 35S:2Ap3, *S. lycopersicum* 35S:*2A*p4 and *S. lycopersicum* 35S:*2A*p24) were infiltrated with 15 µM Ecp2 protein and 33-fold diluted A4 buffer. The negative controls *S. lycopersicum* Cf0 and *S. lycopersicum* 35S:*Ecp2,* and the positive control *S. pimpinellifolium* CfEcp2 were also infiltrated. The negative controls and the T_2 stable transgenic plants did not respond to Ecp2 protein (Table 4.2 and Figure 4.1). The positive control *S. pimpinellifolium* CfEcp2

responded to Ecp2 protein infiltration with a rapid, confluent HR^+ (Table 4.2 and Figure 4.1). None of the plants tested responded to A4 buffer infiltration (Table 4.2, Figure 4.1).

Table. 4.2. *S. lycopersicum* Cf0 T_2 35S:*2A* families infiltrated with Ecp2 protein in 11 day old cotyledons.

	Phenotype 6 d.p.i.				
	15 uM Ecp2 protein		33x diluted A4 buffer		Total number of plants infiltrated
Plant line	HR	NR	HR	NR	
S. lycopersicum Cf0	0	2	0	2	2
S. pimpinellifolium CfEcp2	2[1]	0	0	2	2
S. lycopersicum Cf0 35S:*Ecp2*	0	1	0	1	1
S. lycopersicum Cf0 35S:*2A*p3	0	8	0	8	8
S. lycopersicum Cf0 35S:*2A*p4	0	2	0	2	2
S. lycopersicum Cf0 35S:*2A*p24	0	6	0	6	6

NR = no response, [1]HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, d.p.i = days post inoculation, *2A* = *OR2A*.

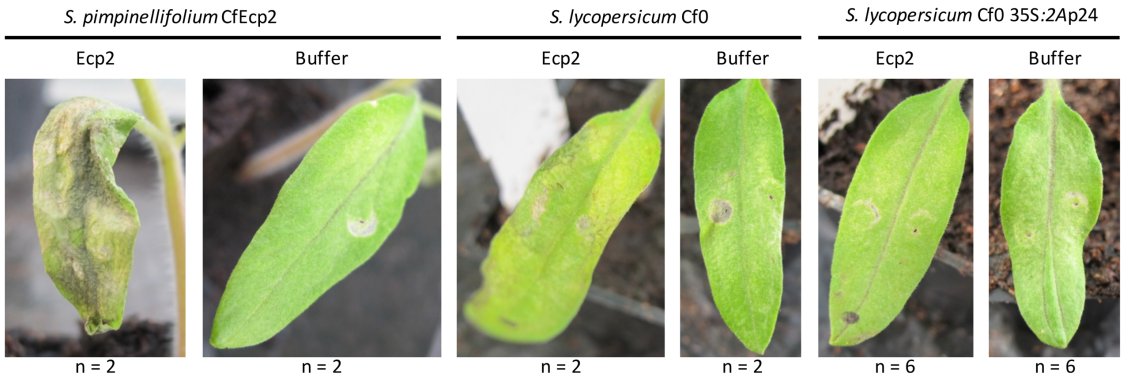


Figure 4.1. Ecp2 protein infiltrated into the cotyledons of 11-day-old tomato seedlings.

15 µM Ecp2 protein (Ecp2) or 33-fold diluted A4 buffer (buffer) was infiltrated into *S. pimpinellifolium* CfEcp2, *S. lycopersicum* Cf0 and *S. lycopersicum* Cf0 35S:*2A*p24. n= number of plants infiltrated per line. Photographs were taken 6-days post infiltration.

The T_2 plants were expected to segregate for the presence and expression of *2A*. However, by chance, those plants infiltrated with Ecp2 protein may not have contained the construct 35S:*2A* or may have failed to express *2A*. To test whether the T_2 plants were segregating for the presence and expression of the transgene *2A*, the RNA and DNA of the T_2 plants and the control plants was extracted and analysed.

The DNA was analysed for the presence of *2A* and 35S:*2A*, whilst the RNA (cDNA) was analysed for the presence of transcripts of *2A* (Figure 4.2). The cDNA was also subjected to PCR for the housekeeping gene *EFa1*, to confirm the presence of amplifiable cDNA in the sample (Figure 4.2). All cDNA samples, except for *S. lycopersicum* Cf0 35S:*2A*p3.5, gave amplification products for *EFa1*. All of the samples therefore contained cDNA except for *S. lycopersicum* Cf0 35S:*2A*p3.5, which will not be considered further. The presence of DNA was confirmed in all samples by gel electrophoresis of each DNA sample (data not shown). In the negative control *S. lycopersicum* Cf0, there was no amplification product for *2A* in the DNA or cDNA, or 35S:*2A* in the DNA. In the negative control *S. lycopersicum* 35S:*Ecp2,* there was no amplification product for *2A* in the cDNA or 35S:*2A* in the DNA. However, there was a weak background signal observed with the primers for *2A* in the DNA, (Figure 4.2). The positive control *S. pimpinellifolium* CfEcp2 gave an amplification product for *2A* in the DNA but not for 35S:*2A* (Figure 4.2). This was expected since *2A* is located downstream of the native promoter in this line. However, there was only weak amplification of *2A* in the cDNA from these plants (Figure 4.2). This may have been due to low levels of transcript under the native promoter, in comparison to the over-expressing transgenic lines.

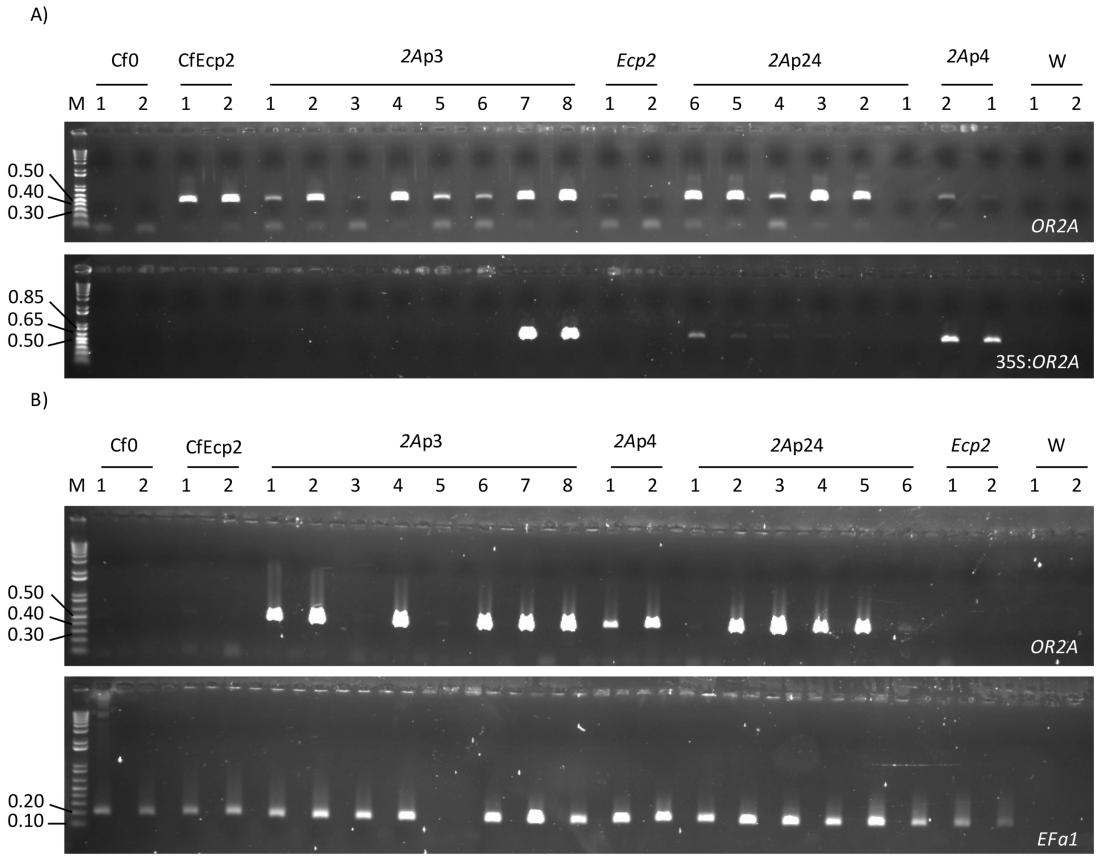


Figure 4.2. Agarose gel electrophoresis of amplification products of *2A*, 35S:*2A* or the housekeeping gene *EFa1* from DNA (A) and cDNA (B).

A, *OR2A* (*2A*) and 35S:*2A* were specifically amplified from the DNA of two *S. lycopersicum* Cf0 plants (Cf0), two *S. pimpinellifolium* CfEcp2 plants (CfEcp2), two *S. lycopersicum* Cf0 35S:*Ecp2* plants (*Ecp2*), eight *S. lycopersicum* Cf0 35S:*2A3*p3 plants (*2A*p3), *two S. lycopersicum* Cf0 35S:*2A3*p4 plants (*2A*p4) and six *S. lycopersicum* Cf0 35S:*2A3*p3 plants (*2A*p24). **B,** cDNA from the same plants used in A, were subjected to specific amplification for *2A* and *EFa1*. Two water controls were used per amplification reaction. All amplification products were run on a 1% agarose gel alongside a 1 Kb+ Invitrogen marker (M). Sizes of key marker bands are indicated in kilo bases.

For the T_2 family *S. lycopersicum* Cf0 35S:*2A*p3, seven out of the eight plants tested showed the presence of *2A*, although one of the plants tested showed only weak amplification of *2A* (Figure 4.2). In addition, two out of the eight plants tested (plant numbers 7 and 8) showed the presence of 35S:*2A* (Figure 4.2). Expression of *2A* was observed in six out of the seven plants tested in this line (Figure 4.2).

For the T_2 family *S. lycopersicum* 35S:*2A*p4, one plant showed a product for *2A* and the other showed a weaker amplification product for *2A* (Figure 4.2). However, both plants showed presence of 35S:*2A*, and both plants showed the presence of *2A* transcripts (Figure 4.2). For the T_2 family *S. lycopersicum* 35S:*2A*p24, five out of the six plants tested showed the presence of *2A* within the DNA but only one showed an amplification product for 35S:*2A* with two of the samples giving weaker products (Figure 4.2). In addition, four out of the 6 plants showed expression of *2A* whilst the other two plants had a weaker amplification product for *2A* transcripts (Figure 4.2). No amplification products were found in the water control for *2A* (in both reactions), 35S:*2A* and *EFa1* (Figure 4.2). Therefore, the lack of response to Ecp2 protein infiltration in the cotyledons was not correlated to the absence or lack of expression of *2A*.

4.2.1.ii Response of *S. lycopersicum* Cf0 35S:*2A* T_2 and T_3 families to PVX:*Ecp2* and PVX:*Avr4*

A total of 12 T_2 families and 23 T_3 families were selected and inoculated with PVX:*Ecp2* (Appendix 3 Table 3.1). Selection of the families was based on the fact that some leaves of the T_1 plant gave a response to Ecp2 protein infiltration and enough seed was collected from the parent plant for the experiment. Furthermore, the controls *S. lycopersicum* Cf0, *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179 were also inoculated with PVX:*Ecp2*. HR^0 or a HR^+ was scored in comparison to no response in each of the lines inoculated at 14 or 21 d.p.i..

In the negative controls (*S. lycopersicum* Cf0, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179) none of the plants tested responded to PVX:*Ecp2* inoculation with a HR (Table 4.3 and Figure 4.3). In the positive control, *S. pimpinellifolium* CfEcp2, all 13 plants tested responded with a HR^+ to PVX:*Ecp2* (Table 4.3 and Figure 4.3). In the T_2 families of *S. lycopersicum* Cf0 35S:*2A*p2, *S. lycopersicum* Cf0 35S:*2A*p3 and *S. lycopersicum* Cf0 35S:*2A*p24, segregation occurred in the T_2 plants for response to the inoculated PVX:*Ecp2* (by a HR^0 phenotype) to no response in 1:8, 1:1 and 3:4 ratios, respectively. None of the other T_2 families tested showed a response to PVX:*Ecp2* infiltration e.g. *S. lycopersicum* Cf0 35S:*2A*p16 (Table 4.3 and Appendix 3, Table 3.1).

Table 4.3. Response of *S. lycopersicum* 35S:*2A* transgenic lines to PVX:*Ecp2.*

Plant line	Phenotype 14 or 21 d.p.i. HR	NR	Total plants inoculated
Controls			
S. lycopersicum Cf0	0	18	18
S. pimpinellifolium CfEcp2	13[1]	0	13
S. pimpinellifolium 1178	0	15	15
S. pimpinellifolium 1179	0	13	13
T_2 families			
S. lycopersicum Cf0 35S:*2A*p2	1[2]	8	9
S. lycopersicum Cf0 35S:*2A*p3	1[2]	1	2
S. lycopersicum Cf0 35S:*2A*p16	0	4	4
S. lycopersicum Cf0 35S:*2A*p24	3[2]	4	7
T_3 families			
S. lycopersicum Cf0 35S:*2A*p3.1	7[2]	1	8
S. lycopersicum Cf0 35S:*2A*p3.2	0	8	8
S. lycopersicum Cf0 35S:*2A*p3.3	0	5	5
S. lycopersicum Cf0 35S:*2A*p3.5	2[2]	3	5
S. lycopersicum Cf0 35S:*2A*p3.6	3[2]	4	7
S. lycopersicum Cf0 35S:*2A*p3.8	2[2]	6	8
S. lycopersicum Cf0 35S:*2A*p3.9	5[2]	1	6
S. lycopersicum Cf0 35S:*2A*p3.10	0	8	8
S. lycopersicum Cf0 35S:*2A*p3.11	0	8	8
S. lycopersicum Cf0 35S:*2A*p24.1	0	2	2
S. lycopersicum Cf0 35S:*2A*p24.2	2[2]	4	6
S. lycopersicum Cf0 35S:*2A*p24.3	3[2]	4	7
S. lycopersicum Cf0 35S:*2A*p24.4	4[2]	3	7
S. lycopersicum Cf0 35S:*2A*p24.5	1[2]	1	2

NR = no response, [1]=HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [2]= HR^0 = HR manifesting as patchy necrosis, d.p.i = days post inoculation, PVX = Potato virus X, *2A* = *OR2A*.
Green = cDNA analysed for presence of transcripts of *2A*, *Ecp2* and *PVX.*

S. lycopersicum Cf0 (18/18)

S. lycopersicum Cf0 35S:*2A*p24.4 (3/7)

S. lycopersicum Cf0 35S:*2A* p24.4 (4/7)

S. pimpinellifolium CfEcp2 (13/13)

Figure 4.3. PVX-mediated delivery of Ecp2 into tomato plants.

S. lycopersicum Cf0, *S. lycopersicum* Cf0 35S:*2A*p24.4 and *S. pimpinellifolium* CfEcp2 were inoculated with Potato Virus X PVX:*Ecp2* and photographs were taken 14 days post infiltration. The number of plants of the lines which responded as shown in the photograph are indicated in brackets. *2A* = *OR2A*.

Three out of four of the T_2 families tested, which responded to Ecp2 protein infiltration in the T_1 parent, also responded to PVX-mediated delivery of Ecp2 in the T_2. However, the response observed in the T_2 families to PVX:*Ecp2* was expressed as a necrosis and stunted growth (HR^0). This is different to the rapid developing confluent necrosis (HR^+) observed in the positive control *S. pimpinelifolium* CfEcp2, in which all of the inoculated plants invariably died (Figure 4.3). To investigate this further, three plants, which responded, were analysed by RT-PCR. The same analysis was performed on three further plants, which did not respond to PVX:*Ecp2* inoculation from the *S. lycopersicum* Cf0 35S:*2A*24 T_2 family, and four plants, which did not respond to PVX:*Ecp2* infiltration from the *S. lycopersicum* Cf0 35S:*2A*p16 T_2 family. They were examined for the presence of transcripts for *2A*, *Ecp2* and PVX.

As expected, all plants tested were positive for Ecp2 and PVX transcripts (Table 4.4). This leads to the conclusion that lack of response to PVX:*Ecp2* in respective plants was not due to a lack of PVX-mediated Ecp2 spread. In the context of the negative control *S. lycopersicum* Cf0, as expected, transcripts for *2A* were absent. For the positive control *S. pimpinelifolium* CfEcp2 *2A,* transcripts were only identified in one out of the three plants tested (Table 4.4). This may have been due to low levels of transcript generated by the native promoter in comparison to the over-expressing transgenic lines (as seen earlier in section 4.3.1). As expected, *2A* transcripts were present in all three *S. lycopersicum* Cf0

35S:*2A*p24 T_2 plants which responded to PVX:*Ecp2* infiltration (Table 4.4). However, *2A* transcripts were present in all four *S. lycopersicum* Cf0 35S:*2A*p16 T_2 plants and in one out of three of the *S. lycopersicum* Cf0 35S:*2A*p24 T_2 plants tested that did not respond to PVX:*Ecp2* infiltration (Table 4.4). The presence of the *2A* gene transcripts did therefore not correlate with response to PVX:*Ecp2.* However, all T_2 plants showing a response to PVX:*Ecp2* infiltration had *2A* transcripts present.

Table 4.4. RT-PCR analysis of 35S:*2A* stable transgenic lines inoculated with PVX:*Ecp2*.

Plant line	PVX:*Ecp2*	Number	Presence (+) or absence (-) of transcripts					
			Expected			Observed		
			2A	*Ecp2*	PVX	*2A*	*Ecp2*	PVX
S. lycopersicum Cf0	NR	3	-	+	+	-	+	+
S. pimpinellifolium CfEcp2	HR^+	1	+	+	+	+ (1/3)	+	+
S. lycopersicum Cf0 35S:*2A*p24	HR^0	3	+	+	+	+	+	+
S. lycopersicum Cf0 35S:*2A*p24	NR	3	-	+	+	+ (1/3)	+	+
S. lycopersicum Cf0 35S:*2A*p16	NR	4	-	+	+	+	+	+

NR = no response, HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, HR^0 = HR manifesting as patchy necrosis, PVX = Potato virus X, *2A* = *OR2A*.

Between the nine T_3 families derived from the T_1 plant *S. lycopersicum* Cf0 35S:*2A*p3 and the five T_3 families derived from the T_1 plant *S. lycopersicum* Cf0 35S:*2A*p24, there was segregation for response to PVX:*Ecp2*, *i.e.* some T_3 families contained individuals, which responded, whilst some families did not contain a single individual that responded. Similar phenotypes were seen in other T_3 families tested (Appendix 3, Table 3.1). Furthermore, where there was no response in the T_2 family to PVX:*Ecp2*, there was also no response to PVX:*Ecp2* in the T_3 plants tested (Appendix 3, Table 3.1).

To confirm that responses observed in relation to PVX:*Ecp2* were specific to Ecp2, individual plants from the T_3 families *S. lycopersicum* Cf0 35S:*2A*p3.8 and *S. lycopersicum* Cf0 35S:*2A*p24.4 were inoculated with PVX:*Ecp2* or PVX:*Avr4* (Table 4.5). The plants *S. lycopersicum* Cf0, *S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179 were inoculated as negative controls and *S. pimpinellifolium* CfEcp2 was inoculated as a positive control for response to PVX:*Ecp2* (Table 4.5). Responses were scored 30 d.p.i.. All the controls behaved as expected; none of the lines responded to PVX:*Avr4*, the negative controls did not respond to PVX:*Ecp2*, whereas the positive control did respond with a HR^+ when inoculated with PVX:*Ecp2* (Table 4.5). In the T_3 familiy *S. lycopersicum* Cf0 35S:*2A*p3.8,

one out of ten of the plants developed a HR^0 to PVX:*Ecp2* infiltration (Table 4.5). In the T_3 famliy *S. lycopersicum* Cf0 35S:*2A*p24.4, three out of nine plants showed a similar HR^0 to PVX:*Ecp2* infiltration (Table 4.5).

Table 4.5. Response of *S. lycopersicum* 35S:*2A* transgenic lines to PVX:*Ecp2* and PVX:*Avr4.*

Plant line	Response 30 d.p.i				Total number of plants inoculated
	PVX:*Ecp2*		PVX:Avr4		
	HR	NR	HR	NR	
S. lycopersicum Cf0	0	4	0	4	8
S. pimpinellifolium CfEcp2	5[1]	0	0	6	11
S. pimpinellifolium 1178	0	5	0	6	11
S. pimpinellifolium 1179	0	5	0	5	10
S. lycopersicum Cf0 35S:*2A*p3.8	1[2]	9	0	10	20
S. lycopersicum Cf0 35S:*2A*p24.4	3[2]	6	0	10	19

NR = no response, [1]=HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [2]= HR^0 = HR manifesting as patchy necrosis, d.p.i = days post inoculation, PVX = Potato virus X, *2A* = *OR2A*.

4.2.1.iii Response of *S. lycopersicum* Cf0 35S:*2A* T_2 and T_3 families to kanamycin

Since 35S:*2A* was transformed into *S. lycopersicum* Cf0 in a cassette carrying the *neophosphotransferase* (*npt*) gene conditioning resistance to the antibiotic selectable marker kanamycin, it was expected that those lines that were resistant to kanamycin would also carry the 35S:*2A* transgene. Consequently, to determine the presence of 35S:*2A* in T_2 and T_3 families, their seed was sown on Murashige and Skoog media supplemented with 300 mg/L kanamycin. Plants resistant to kanamycin form a normal root system, cotyledons and leaves (Figure 4.4). Plants susceptible to kanamycin show an accumulation of anthocyanins in the stem and cotyledons. They also have a stunted root system and do not form true leaves (Figure 4.4).

Figure 4.4. Response of *S. lycopersicum* Cf0 35S:*2A*p24.4 to kanamycin.

Seedlings were germinated on MS media supplemented with 300 mg/L kanamycin. The resistant (left) and susceptible (right) phenotype is shown. Photographs were taken 24-days post sowing. *2A* = *OR2A*.

Out of the eight T_2 families tested, five of them had equal to or less than the percentage germination (20%) seen with the non-kanamycin resistant *S. pimpinellifolium* CfEcp2 and kanamycin-resistant *S. pimpinellifolium* 1179 (Appendix 3, Table 3.2). The highest percentage germination was 40% for the T_2 family *S. Lycopersicum* Cf0 35S:*2A*p2 (Appendix 3, Table 3.2). The data suggested that the T_2 family *S. lycopersicum* Cf0 35S:*2A*p3 was segregating for resistance to kanamycin and thus segregating for the presence of the 35S:*2A* transgene (Table 4.6). Those T_2 plants resistant to kanamycin were transplanted into soil and grown to establish crosses with *S. lycopersicum* Cf0 35S:*Ecp2* (Appendix 3, Table 3.2).

Table 4.6. Response of *S. lycopersicum* 35S:*2A* T_2 families grown on kanamycin.

	Phenotype 16 d.p.s.		
	300 mg/L kanamycin		Percentage
Plant line	Resistant	Sensitive	germination (%)
S. pimpinellifolium CfEcp2	0	4	20
S. pimpinellifolium 1179	4	0	20
S. lycopersicum 35S:*2A*p3	3	1	20

d.p.s. = days post sowing, *2A* = *OR2A.*

To further investigate if kanamycin affects the percentage germination of transgenic plants, T_3 families of *S. lycospersicum* Cf0 35S:*2A*p3.8 and *S. lycospersicum* Cf0 35S:*2A*p24.4 and the control plants *S. lycopersicum* Cf0 and *S. pimpinellifolium* 1179, were sown on MS media with or without kanamycin. kanamycin had no effect on the germination rate of *S. lycopersicum* Cf0 or *S. pimpinellifolium* 1179 (Table 4.7). However,

there was increased percentage germination in T_3 transgenic families *S. lycospersicum* Cf0 35S:*2A*p3.8 and *S. lycospersicum* Cf0 35S:*2A*p24.4, when in the presence of kanamycin (Table 4.7).

Table 4.7. Response of *S. lycopersicum* 35S:*2A* T_3 families to kanamycin.

Plant line	Phenotype 24 d.p.s.					
	300 mg/L kanamycin			No kanamycin		
	Resistant	Sensitive	Germination (%)	Resistant	Sensitive	Germination (%)
S. lycopersicum Cf0	0	7	87.5	7	0	87.5
S. pimpinellifolium 1179	3	0	37.5	3	0	37.5
S. lycopersicum Cf0 35S:*2A*p3.8	8	2	62.5	6	0	37.5
S. lycopersicum Cf0 35S:*2A*p24.4	5	1	37.5	2	0	12.5

d.p.s. = days post sowing, *2A* = *OR2A.*

Those plants, which were transgenic (i.e. *S. pimpinellifolium* 1179, *S. lycopersicum* Cf0 35S:*2A*p3.8 and *S. lycopersicum* Cf0 35S:*2A*p24.4) and were homozygous for or segregating for kanamycin-resistance, had a lower perecentage germination in comparison to the non-transgenic *S. lycopersicum* Cf0 (Table 4.7). In addition, the data suggested that *S. lycospersicum* Cf0 35S:*2A*p3.8 and *S. lycospersicum* Cf0 35S:*2A*p24.4 lines were both segregating for resistance to kanamycin. In order to confirm the conclusions from these data, it is proposed that more repeats would need to be undertaken.

4.2.1.iv 35S-mediated stable expression of *Ecp2* in *S. lycopersicum* Cf0 35S:*2A* results in SLP

Protein infiltration and PVX-mediated delivery of Ecp2 into the transgenic plants generated weaker phenotypes to those seen in the resistant parent *S. pimpinellifolium* CfEcp2. Furthermore, infiltration of Ecp2 protein into the cotyledons of T_2 families did not result in a visible response in comparison to the HR^+ seen in *S. pimpinellifolium* CfEcp2. Therefore, the response of *2A* to Ecp2 appeared to depend on the method of delivery of the effector. In an attempt to achieve a more robust response to Ecp2, selected T_1, T_2 and T_3 plants were crossed to *S. lycopersicum* Cf0 plants homozygous for the presence of 35S:*Ecp2*. Here, the Ecp2 would be present from germination of the seed, and sustained throughout plant growth.

In the situation where *S. pimpinellifolium* CfEcp2 was crossed with *S. lycopersicum* Cf0 35S:*Ecp2,* all progeny showed an SLP characterised as death of the whole plant soon after emergence of the first true leaves (Table 4.8). This is because there was recognition of the Ecp2 protein by the product of the *Cf-Ecp2* gene throughout the whole plant, caused by the presumed over expression of *Ecp2* in every cell type. In contrast, when *S. lycopersicum* Cf0 was crossed with *S. lycopersicum* Cf0 35S:*Ecp2,* all progeny had a normal phenotype because there was no CfEcp2 and so no recognition of Ecp2 (Table 4.8). If *2A* was *Cf-Ecp2*, it would be predicted that the progeny from T_1 lines crossed with *S. lycopersicum* Cf0 35S:*Ecp2* would segregate for SLP to normal growth phenotype. This was observed (Table 4.8, Figure 4.5 and Appendix 3, Table 3.3). In addition, other intermediate phenotypes were observed in the progeny. The intermediate phenotypes included anthocyanin-accumulation (A), developmental phenotype (D) or stunted growth (St) (Figures 4.5 and 4.6 and Appendix 3, Table 3.3). The SLP was observed from the onset of germination in the progeny from crosses between T_1 plants and *S. lycopersicum* Cf0 35S:*Ecp2.* In comparison, the progeny from *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2* grew into seedlings with developed cotyledons, appearing healthy before death occurred.

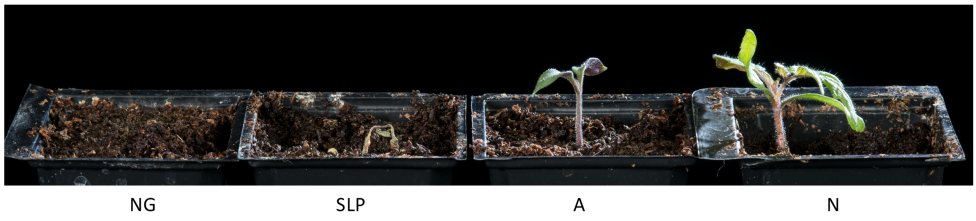


Figure 4.5. Phenotypes of progeny from the cross *S. lycopersicum* Cf0 35S:*2A*p2 x *S. lycopersicum* Cf0 35S:*Ecp2*.

The phenotypes were scored from left to right as; no germination, seedling lethal phenotype, anthocyanin accumulation, and normal growth. Photographs were taken 15-days post sowing. *2A* = *OR2A.*

The SLP and intermediate phenotype responses observed were specific for crosses of *2A* transgenics or *S. pimpinellifolium* CfEcp2 to plants expressing 35S:*Ecp2*. In the T_3 families *S. lycopersicon* Cf0 35S:*2A*p3.8 and *S. lycopersicon* Cf0 35S:*2A*p24.4, none of the plants showed SLP or any of the intermediate phenotypes (A, D or St) (Table 4.8). In addition, none of the transgenic lines showed SLP or any of the intermediate phenotypes (A, D or St) when they were selfed or crossed to *S. lycopersicon* Cf0 (Table 4.8).

Table 4.8. Phenotypes of progeny from crosses between *S. lycopersicon* Cf0 stably expressing 35S:*2A* x *S. lycopersicon* Cf0 35S:*Ecp2.*

	Phenotype				
Plant line	SLP	A or D or St	NG	N	Total
Controls					
S. lycopersicon Cf0	0	0	13	24	37
[1]*S. lycopersicon* Cf0 x *S. lycopersicon* 35S:*Ecp2*	0	0	19	23	42
S. lycopersicon Cf0 x *S. pimpinellifolium CfEcp2*	0	0	0	12	12
S. pimpinellifolium CfEcp2	0	0	16	21	37
[1]*S. pimpinellifolium* CfEcp2 x *S. lycopersicon* 35S:*Ecp2*	15	0	5	1	21
S. lycopersicon 35S:*Ecp2*	0	0	2	10	12
T_1 crosses					
[1]*S. lycopersicon* Cf0 35S:*2A*p3 x *S. lycopersicon* Cf0 35S:*Ecp2*	14	0	26	0	40
S. lycopersicon Cf0 35S:*Ecp2 x S. lycopersicon* Cf0 35S:*2A*p24	6	0	6	6	18
T_3 families					
S. lycopersicon Cf0 35S:*2A*p3.8	0	0	7	17	24
S. lycopersicon Cf0 35S:*2A*p24.4	0	0	11	13	24
T_2 crosses					
S. lycopersicon Cf0 35S:*2A*p3.8 x *S. lycopersicon* Cf0 35S:*Ecp2*	8	0	0	0	8
[1]*S. lycopersicon* Cf0 35S:*2A*p24.4 x *S. lycopersicon* Cf0 35S:*Ecp2*	6	1	15	18	40
T_4 families					
S. lycopersicon Cf0 35S:*2A*p24.4.2D	0	0	1	11	12
S. lycopersicon Cf0 35S:*2A*p24.4.4D	0	0	9	15	24
S. lycopersicon Cf0 35S:*2A*p24.5-5D	0	0	16	8	24
T_3 crosses					
S. lycopersicon Cf0 35S:*2A*p24.4.2D x *S. lycopersicon* Cf0	0	0	3	21	24
S. lycopersicon Cf0 35S:*2A*p24.4.2D x *S. lycopersicon* Cf0 35S:*Ecp2*	13	0	11	0	24
S. lycopersicon Cf0 35S:*2A*p24.4.4D x *S. lycopersicon* Cf0	0	0	20	4	24
S. lycopersicon Cf0 35S:*2A*p24.4.4D x *S. lycopersicon* Cf0 35S:*Ecp2*	4	0	6	16	24
S. lycopersicon Cf0 35S:*2A*p24.5.5D x *S. lycopersicon* Cf0	0	0	14	10	24
S. lycopersicon Cf0 35S:*2A*p24.5.5D x *S. lycopersicon* Cf0 35S:*Ecp2*	0	0	10	14	24

[1]Crosses in either direction used, and due to no difference in results, the results were combined.

Phenotypes scored 15 – 26 days post sowing.

SLP = Seedling lethal phenotype, A = Anthocyanin accumulation, D = Developmentally-different, NG = No germination, N = Normal, Total = Total plants sown for scoring, *2A* = *OR2A.*

The developmental phenotype was observed in some progeny that also showed accumulation of anthocyanin (Figure 4.6). The presence of 35S:*2A* was correlated with the presence of the developmental phenotype in progeny from the cross between *S. lycopersicon* Cf0 35S:*2A*p2 and *S. lycopersicon* Cf0 35S:*Ecp2* (Table 4.9). Plants showing the developmental phenotype gave an amplification product for 35S:*2A*, whilst those plants with normal growth phenotype gave no amplification product for 35S:*2A* (Table

4.9). This is in comparison to the control plants tested, in which progeny from the cross *S. lycopersicon* Cf0 35S:*2A*p24 x *S. lycopersicon* Cf0 35S:*Ecp2* with normal growth phenotype gave no amplification product for 35S:*2A*.

Figure 4.6. Ecp2-induced developmental phenotype in the transgenic line 35S:*2A*p2.

Normal growth (N) and developmental phenotype (D) of progeny from the cross *S. lycopersicum* 35S:*2A*p2 x *S. lycopersicum* 35s:*Ecp2.* Phenotypes were scored 30 days post sowing and photographs were taken 40 days post sowing. *2A* = *OR2A*.

Table 4.9. PCR of 35S:*2A* stable transgenic plants with developmental phenotype.

Plant line	Number of plants tested	Phenotype	35S:*2A*
S. lycopersicon Cf0 35S:*2A*p2 x *S. lycopersicon* Cf0 35S:*Ecp2*	7	D	+
[1] *S. lycopersicon* Cf0 35S:*2A*p2 x *S. lycopersicon* Cf0 35S:*Ecp2*	4	N	_
S. lycopersicon Cf0 35S:*2A*p24 x *S. lycopersicon* Cf0 35S:*Ecp2*	5	N	_

[1]Crosses in either direction used, and due to no difference in results, the results were combined.
Phenotypes scored 15 – 26 days post sowing, D = developmental phenotype, N = Normal phenotype, *2A* = *OR2A*.

It was anticipated that the low level of germination observed in some crosses was linked to the presence of the transgene. This was reasoned, because a cross of *N. tobacum* stably expressing Cf-9 with *N. tobacum* stably expressing Avr9 may not give rise to germination, depending on the direction of the cross (Hammond-Kosack et al., 1998). However, since the controls also had high levels of non-germinated seed, this link cannot be assumed (Table 4.8). Furthermore, there was no difference seen in response and

germination rate when completing crosses in either direction (data not shown). As a consequence, the results for crosses in either direction were combined.

In the cross *S. lycopersicon* Cf0 35S:*2A*p3 x *S. lycopersicon* Cf0 35S:*Ecp2,* 14 of the progeny showed SLP, whilst 26 of the seeds failed to germinate. In contrast, in the cross between *S. lycopersicon* Cf0 35S:*Ecp2* and *S. lycopersicon* Cf0 35S:*2A*p24 six of the plants showed SLP whilst six of the plants showed a normal phenotype. The final six progeny from this cross did not germinate.

These data suggested that *S. lycopersicon* Cf0 35S:*2A*p3 was homozygous for the presence of the transgene 35S:*2A*, whilst *S. lycopersicon* Cf0 35S:*2A*p24 was segregating for the presence of the transgene. This was supported by the results observed, when T_2 plants were crossed to *S. lycopersicon* Cf0 35S:*Ecp2* (Table 4.8). When the T_2 plant *S. lycopersicon* Cf0 35S:*2A*p3.8 was crossed to *S. lycopersicon* Cf0 35s:*Ecp2,* all eight progeny tested showed SLP (Table 4.8). Furthermore, when the T_2 plant *S. lycopersicon* Cf0 35S:*2A*p24.4 was crossed to *S. lycopersicon* Cf0 35s:*Ecp2,* six of the progeny showed SLP, one of the progeny showed anthocyanin accumulation, 15 of the seed did not germinate and 18 of the progeny showed a normal phenotype (Table 4.8).

When T_3 plants of *S. lycopersicon* Cf0 35S:*2A*p24.4.2D and *S. lycopersicon* Cf0 35S:*2A*p24.5.3D were crossed to *S. lycopersicon* Cf0 35s:*Ecp2,* fixation for the presence or absence of SLP was observed in the progeny, respectively. However, in the cross *S. lycopersicon* Cf0 35S:*2A*p24.4.4D x *S. lycopersicon* 35S:*Ecp2,* segregation for SLP and normal plant growth was demonstrated (Table 4.8).

To correlate the presence of the 35S:*2A* construct alongside the SLP, DNA from selected plants from the above analysis was analysed for the presence of the construct 35S:*2A* (Table 4.10). All samples tested were confirmed to contain PCR-amplifiable DNA because all gave amplification products for the housekeeping gene *eEF1Art* (Table 4.10 and Figure 4.7). No amplification products were found for 35S:*2A* in any of the control lines tested, including *S. pimpinellifolium* CfEcp2 and progeny from the cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.10 and Figure 4.7). This was expected since, although *S. pimpinellifolium* CfEcp2 carried *2A*, it did not contain 35S but a native promoter.

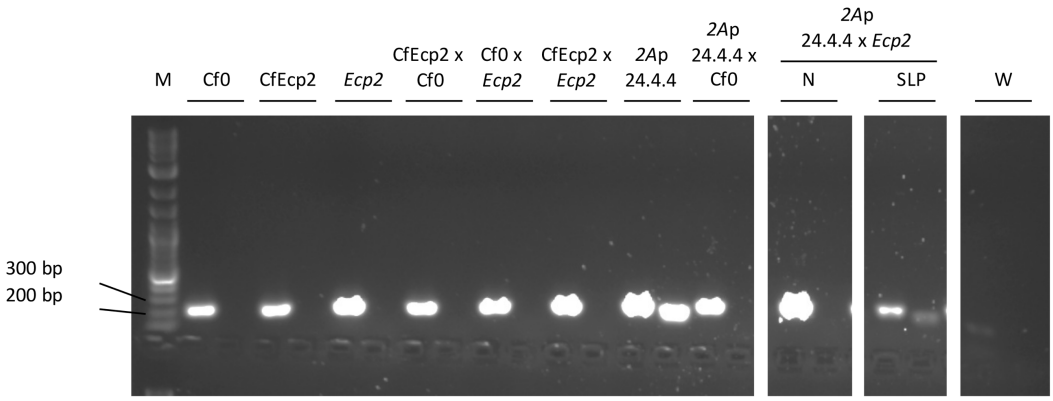


Figure 4.7. Agarose gel electrophoresis of amplification products of 35S:*2A* and the housekeeping gene *EFArt* and from tomato lines.

DNA from the tomato lines *S. lycopersicum* Cf0 (Cf0), *S. pimpinellifolium* CfEcp2 (CfEcp2), *S. lycopersicum* Cf0 35S:Ecp2 (*Ecp2*), *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 (CfEcp2 x Cf0), *S. lycopersicum* Cf0 x *S. lycopersicum* Cf0 35S:*Ecp2* (Cf0 x *Ecp2*), *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2* (CfEcp2 x *Ecp2*), *S. lycopersicum* Cf0 35S:*2A*p24.4.4 (*2A*p24.4.4), *S. lycopersicum* Cf0 35S:*2A*p24.4.4 x *S. lycopersicum* Cf0 (*2A*p24.4.4 x Cf0) and *S. lycopersicum* Cf0 35S:*2A*p24.4.4 x *S. lycopersicum* Cf0 35S:*Ecp2* (*2A*p24.4.4 x *Ecp2*) were subjected to specific amplification of *EFArt* (left hand lane) and 35S:*2A* (right hand lane) for each sample. One sample is shown from each line except for *2A*p24.4.4 x Ecp2, whereby one sample is shown from a plant that had a normal phenotype (N) and one with a seedling lethal phenotype (SLP). All samples were run on a 1% agarose gel next to a 2log ladder (M) and important size markers are indicated. *2A* = *OR2A.*

The lines *S. lycopersicum* Cf0 35S:*2A*p3 and *S. lycopersicum* Cf0 35S:*2A*p24.4.2D were considered to carry more than one copy of 35S:*2A* or to be homozygous for the presence of 35S:*2A,* respectively. This was concluded because all progeny from the cross of these plants to *S. lycopersicum* Cf0 35S*:Ecp2* showed SLP (Table 4.8). All plants tested from the lines *S. lycopersicum* Cf0 35S:*2A*p3 and *S. lycopersicum* Cf0 35S:*2A*p24.4.2D showed the presence of 35S:*2A* (Table 4.10). However, *S. lycopersicum* Cf0 35S:*2A*p24.4.2D may have been segregating for the presence of 35S:*2A* since one of the plants from the cross between *S. lycopersicum* Cf0 35S:*2A*p24.4.2D and *S. lycopersicum* Cf0 gave no amplification product for 35S:*2A* (Table 4.10). This could be resolved by increasing the population size for testing. Conversely, all plants from the line *S. lycopersicum* Cf0 35S:*2A*p24.5.5D showed no amplification product for 35S:*2A* (Table 4.10). This was

expected because none of the plants from the cross *S. lycopersicum* Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0 35S:*Ecp2* gave SLP or intermediate phenotypes (Table 4.8).

Table 4.10. PCR analysis of 35S:*2A* stable transgenic lines x 35S:*Ecp2* stable transgenic lines.

Plant line	Phenotype	Number of plants containing		Total number of plants tested
		35S:*2A*	*EFArt*	
Controls				
S. lycopersicum Cf0	N	0	5	5
S. pimpinellifolium CfEcp2	N	0	5	5
S. lycopersicum Cf0 35S:*Ecp2*	N	0	3	3
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0	N	0	3	3
S. lycopersicum Cf0 x *S. lycopersicum* Cf0 35S:*Ecp2*	N	0	5	5
S. pimpinellifolium CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2*	SLP	0	5	5
T_3 and T_4 families				
S. lycopersicum Cf0 35S:*2A*p3.8	N	9	9	9
S. lycopersicum Cf0 35S:*2A*p24.4	N	9	12	12
S. lycopersicum Cf0 35S:*2A*p24.4.2D	N	3	3	3
S. lycopersicum Cf0 35S:*2A*p24.4.4D	N	3	3	3
S. lycopersicum Cf0 35S:*2A*p24.5.5D	N	0	2	2
T_3 crossed to Cf0				
S. lycopersicum Cf0 35S:*2A*p24.4.2D x *S. lycopersicum* Cf0	N	2	3	3
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0	N	0	3	3
S. lycopersicum Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0	N	0	2	2
T_3 and T_4 crossed to Ecp2				
S. lycopersicum Cf0 35S:*2A*p24.4 x *S. lycopersicum* Cf0 35S:*Ecp2*	N	0	8	8
S. lycopersicum Cf0 35S:*2A*p24.4 x *S. lycopersicum* Cf0 35S:*Ecp2*	SLP	2	2	2
S. lycopersicum Cf0 35S:*2A*p24.4.2D x *S. lycopersicum* Cf0 35S:*Ecp2*	SLP	2	2	2
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0 35S:*Ecp2*	N	0	14	14
S. lycopersicum Cf0 35S:*2A*p24.4.4D x *S. lycopersicum* Cf0 35S:*Ecp2*	SLP	2	2	2
S. lycopersicum Cf0 35S:*2A*p24.5.5D x *S. lycopersicum* Cf0 35S:*Ecp2*	N	0	3	3

[1]Crosses in either direction used, and due to no difference in results, the results were combined.
Phenotypes were scored 15 – 26 says post sowing. N = Normal, SLP = seedling lethal phenotype *2A* = *OR2A*.

The T_3 family *S. lycopersicum* Cf0 35S:*2A*p24.4 and T_4 family *S. lycopersicum* Cf0 35S:*2A*p24.4.4D were segregating for the presence of 35S:*2A* (Table 4.10 and Figure 4.7). Furthermore, those progeny from crosses of these lines with *S. lycopersicum* Cf0 35S:*Ecp2,* which gave an SLP, carried 35S:*2A*, whilst those plants that had a normal growth phenotype lacked the presence of 35S:*2A* (Table 4.10 and Figure 4.7). Therefore,

correlation between the presence of 35S:*2A* and an SLP in the presence of 35S mediated stable expression of Ecp2 was confirmed.

4.2.1.v Integration of 35S:*2A* into *S. pimpinellifolium* 1179

There was a lack of penetrance of the PVX:*Ecp2* phenotype in the *S. lycopersicum* Cf0 35S:*2A* transgenic lines. This may be due to an additional component within *S. pimpinellifolium* CfEcp2, which is required for *Cf-Ecp2* function but is missing in S. *lycopersicum* Cf0. The F_2 progeny from *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 segregates in a 3:1 manner of resistance to susceptibility (Haanstra et al., 1999; Lauge et al., 1998). Therefore, a potential component required for *Cf-Ecp2* function that is missing in *S. lycopersicum* Cf0 would be genetically linked to *Cf-Ecp2* and map within the 631 Kb interval where the presence of *Cf-Ecp2* function has been mapped (Chapter 3).

To test this hypothesis, 35S:*2A* T_3 transgenic plants were crossed to *S. pimpinellifolium* 1179 plants, known to be lacking 35S:*Ecp2* (*S. pimpinellifolium* 1179p15, *S. pimpinellifolium* 1179p19, *S. pimpinellifolium* 1179p31 and *S. pimpinellifolium* 1179p33) (Figure 2.1). This placed the 35S:*2A* transgene into a genetic background heterozygous for *S. pimpinellifolium* and *S. lycopersicum*. The testcross progeny were then inoculated with PVX:*Ecp2* and responses scored at 19 d.p.i. As seen before with the 35S:*2A* T_3 plants, some of the *S. pimpinellifolium* 35S:*2A* plants developed a HR^0 (e.g. in the cross *S. lycopersicon* Cf0 35S:*2A*p3.8.1 x *S. pimpinellifolum* 1179p19) (Table 4.11). However, none of the testcross progeny developed the strong HR^+ leading to death of the plant as seen in all positive controls of *S. pimpinellifolium* CfEcp2 (Table 4.11). Changing the genetic background of the 35S:*2A* transgene did therefore not confer the full-penetrance of the HR^+ phenotype as seen in *S. pimpinellifolium* Cf-Ecp2–mediated recognition of Ecp2 delivered by PVX.

Table 4.11. Response of *S. pimpinellifolium* 1179 carrying 35S:*2A* to PVX:*Ecp2*.

Plant line	PVX:*Ecp2*		
	HR	NR	Total[1]
S. lycopersicon Cf0	0	6	6
S. pimpinellifolium CfEcp2	6[2]	0	6
S. pimpinellifolium CfEcp2 x *S. lycopersicon* Cf0	12[2]	0	12
S. pimpinellifolium 1179p15	0	6	6
S. pimpinellifolium 1179p19	0	6	6
S. pimpinellifolium 1179p31	0	5	5
S. pimpinellifolium 1179p33	0	6	6
S. pimpinellifolium CfEcp2 x *S. pimpinellifolium* 1179p15	12[2]	0	12
S. lycopersicon 35S:*2A*p3.8.1	0	5	5
S. lycopersicon 35S:*2A*p3.8.1x *S. lycopersicon* Cf0	0	9	9
S. lycopersicon 35S:*2A*3.8.1x1179p19	2[3]	20	22
S. lycopersicon 35S:*2A*p24.4.4	0	12	12
S. lycopersicon 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19	1[3]	11	12

Total[1] = Total number of leaf sections inoculated, NR = no response, [2]=HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [3]= HR^0 = HR manifesting as patchy necrosis, PVX = Potato virus X, *2A* = *OR2A*.

To determine whether the lack of response to PVX:*Ecp2*-mediated inoculation correlated with the absence of 35S:*2A*, DNA from those plants failing to respond to PVX:*Ecp2*-inoculation was subjected to PCR amplification of 35S:*2A*. In addition, the plants were genotyped with the CT116 marker to confirm that they were *bona fide S. lycopersicum* by *S. pimpinellifolium* heterozygotes. As expected, the negative controls *S. lycopersicum* Cf0, *S. pimpinellifolium* CfEcp2, *S. pimpinellifolium* CfEcp2 x *S. lycopersicon* Cf0 and *S. pimpinellifolium* 1179p19 did not contain 35S:*2A* (Table 4.12). On the other hand, 35S:*2A* was found to be present in all three plants tested from the cross between *S. lycopersicon* 35S:*2A*p3.8.1 and *S. pimpinellifolum* 1179p19 (Table 4.12).

In all plants tested from the crosses *S. lycopersicon* Cf0 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19 and *S. lycopersicon* 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19, no 35S:*2A* was present (Table 4.12). However, a weak amplification of 35S:2*A* was detected in one of the two plants tested from each of the T_4 families *S. lycopersicon* Cf0 35S:*2A*p24.4.4 and *S. lycopersicon* Cf0 35S:*2A*p24.5.3 (Table 4.12). As expected, the plants tested from the T_4 families *S. lycopersicon* Cf0 35S:*2A*p24.4.4 and *S. lycopersicon* Cf0 35S:*2A*p24.5.3 were homozygous for the CT116 marker from *S. lycopersicum* Cf0 (Table 4.12).

All progeny from crosses, whereby 35S:*2A* transgenic lines were crossed with *S. pimpinellifolium* 1179 lines, were heterozygous for the CT116 marker (Table 4.12). This indicated that the crosses were successful in combining the *S. pimpinellifolium* 1179 and *S. lycopersicum* Cf0 genomes.

Table 4.12. Presence of 35S:*2A* and origin of CT116 marker in *S. pimpinellifolium* 1179 plants carrying 35S:*2A* that failed to respond to PVX:*Ecp2.*

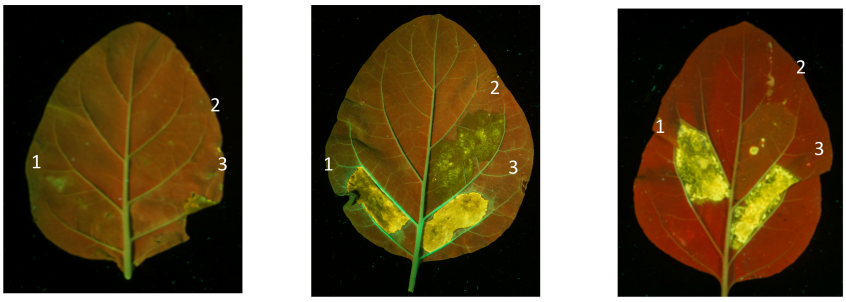
Plant ID	No. plants 35S:*2A* genotyped		CT116
S. lycopersicon Cf0	2	-	Cf0
S. pimpinellifolium CfEcp2	2	-	CfEcp2
S. pimpinellifolium CfEcp2 x *S. lycopersicon* Cf0	3	-	Heterozygous
S. pimpinellifolium 1179p19	2	-	CfEcp2
S. lycopersicon 35S:*2A*p3.8.1	2[1]	+	Cf0
S. lycopersicon 35S:*2A*3.8.1 x *S. pimpinellifolium* 1179p19	3	+	Heterozygous
S. lycopersicon 35S:*2A*p24.4.4	2	+/-[2]	Cf0
S. lycopersicon 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19	2	-	Heterozygous
S. lycopersicon 35S:*2A*p24.5.3	2	+/-[2]	Cf0
S. lycopersicon 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19	3	-	Heterozygous

[1] Pooled two samples, [2] One plant tested positive for 35S:2A, the other tested negative, PVX = Potato virus X, *2A* = *OR2A*

4.2.2 Transient Co-transformation of 35S:*2A* and 35S:*Ecp2* into *N. paniculata*

In a previous study, de Kock and colleagues (2004) used Agrobacterium to transiently co-express *2A* and *Ecp2* in five species of *Nicotiana* (*N. benthamiana*, *N. glutinosa*, *N. kawakamii*, *N. langsdorfii* and *N. tomentosiformis*). This co-expression did not result in the same HR^+ response, as observed for the positive controls (co-expression of *Cf-4* and *Cf-9* with *Avr4* and *Avr9*, respectively). None of the *Nicotiana* species inoculated by de Kock and colleagues (2004) contained accessions that responded to *Ecp2* protein infiltration or Agrobacterium-mediated transient expression of *Ecp2*. In contrast, *N. paniculata* contains accessions that do respond (TW99) and accessions that do not respond (TW102) to Ecp2. There may therefore be another component in addition to *Cf-Ecp2* in *N. paniculata* that is required for recognition of Ecp2. This component might be missing in *N. benthamiana*, *N. glutinosa*, *N. kawakamii*, *N. langsdorfii* and *N. tomentosiformis*. Therefore, presuming *2A* is *Cf-Ecp2*, a transient expression in *N. paniculata* might be successful.

To test this hypothesis, Agrobacterium was used to transiently express *2A* or *Ecp2* alone or in combination in the *N. paniculata* accessions TW99, TW102 and the F_1 from the cross TW99 x TW102 (Figure 4.8). As expected, TW99 and the F_1 plants responded to expression of *Ecp2* and *Ecp2* with *2A*. However, they did not respond to the expression of *2A* alone (Figure 4.8). Furthermore, TW102 did not respond to the co-expression of *2A* and *Ecp2* or to the expression of these genes individually (Figure 4.8). Co-expression of *2A* and *Ecp2* in TW102, did not therefore recapitulate the HR^+ response phenotype seen in TW99 when expressing *Ecp2.*



N. paniculta TW102 | *N. paniculta* TW99 x *N. paniculta* TW102 | *N. paniculta* TW99

Figure 4.8. Co–expression of 35S:*2A* and 35S:*Ecp2* in *N. paniculata*.

Two leaves from one plant of each of the *N. paniculata* accessions TW102, TW99 and the F_1 from the cross TW99 x TW102 were co-infiltrated with Agrobacterium containing the binary constructs (1) 35S:*2A* and 35S:*Ecp2*, (2) 35S:*2A*, and (3) 35S:*Ecp2*. Photographs were taken under UV light exposure at 6 days post inoculation. *2A* = *OR2A*.

4.3 Discussion

Thirteen stable transgenic *S. lycopersicum* Cf0 plants expressing the *Cf-Ecp2* candidate gene *2A* under the constitutive 35S CaMV promoter were generated in the current study. The 35S promoter was used to ensure expression of the transgene in the transgenic lines since no *2A* transgene expression had been obtained in a previous study using 1.5 kb of native 5' regulatory sequence (de Kock and colleagues, 2004).

To test the role of the transgene 35S:*2A* in the recognition of Ecp2, Ecp2 protein was infiltrated into either adult plant leaves of T_1 transgenic lines or cotyledons of T_2 transgenic lines. When infiltrated into adult plant leaves of T_1 plants, 15 μM Ecp2 protein induced a weak necrotic response (HR^0), which appeared after 10 days in eight out of the 13 transgenic lines tested (Table 4.1). However, this response was not consistent since not all leaves on the same plant responded to the infiltrated Ecp2 protein (Table 4.1). Furthermore, the negative control line *S. lycopersicum* Cf0 also responded in one out of four leaves infiltrated with Ecp2 protein (Table 4.1). This may have been a response to the A4 buffer used to deliver the Ecp2 protein because two of the T_1 plants tested contained leaves which responded to the infiltration of A4 buffer without Ecp2 protein (Table 4.1). Due to the lack of consistency and specificity in response to Ecp2 protein, no firm conclusions regarding the role of the *2A* transgene in the recognition of Ecp2 could be drawn based on these experiments.

When the cotyledons of T_2 plants (generated from transgenic lines *S. lycopersicum* Cf0 35S:*2A*p3, *S. lycopersicum* Cf0 35S:*2A*p4 and *S. lycopersicum* Cf0 35S:*2A*p24) were infiltrated with 15 μM Ecp2 protein, no response was observed (Table 4.2 and Figure 4.1). In comparison, a strong, confluent and rapid necrotic response was seen in *S. pimpinellifolium* CfEcp2 infiltrated with 15 μM Ecp2 protein (Table 4.2 and Figure 4.1). The presence and expression of *2A* was observed in a number of the T_2 plants tested (Figure 4.2). Therefore, the lack of Ecp2 protein-recognition in the cotyledons was not due to the absence or lack of expression of *2A*.

The *NB-LRR* resistance gene, *Mi*, encoding for recognition of aphids in tomato plants, is developmentally regulated in plants, independent of *Mi1.2* transcript abundance and ineffective until 5 weeks post-germination (Goggin et al., 2004; Kaloshian et al., 1995).

Similarly, the RLP Cf-9B, recognising the effector Avr9B (Panter et al., 2002) is apparent only in adult plants and is linked to the onset of flowering (Panter et al., 2002). This late-onset resistance is not constrained by the *Cf-9B* promoter because the Cf9B promoter when fused to the *Cf-9* gene drives seedling-mediated resistance conferred by Cf-9 (Panter et al., 2002). Furthermore, both *Cf9* and *Cf9B* mRNA is present in tomato seedlings (Panter et al., 2002). However, in the current study *S. pimpinellifolium* CfEcp2 cotyledons did respond to Ecp2 protein infiltration. The lack of response to the infiltration of Ecp2 protein into the cotyledons of transgenic lines expressing *2A* is therefore not due to late-onset of resistance conferred by *2A*. This phenomenon may be either due to the fact that *2A* is not *Cf-Ecp2* or that another factor is interfering with the phenotype in the transgenic lines.

In the current study, the lines *S. lycopersicum* Cf0 35S:*2A*p3 and *S. lycopersicum* Cf0 35S:*2A*p24 were the most robust in their response to Ecp2 protein-infiltration in adult leaves (Table 4.1). In these plants, all leaves infiltrated with Ecp2 protein, except for one leaf on the *S. lycopersicum* Cf0 35S:*2A*p3 plant, developed a necrotic response (Table 4.1). In addition, all leaves infiltrated with A4 buffer failed to respond (Table 4.1). These findings therefore warranted further investigation of the 35S:*2A* transgenic plants, in particular the *2A*p3 and *2A*p24 lines.

PVX-mediated delivery of effector proteins is often used to characterise plants transformed in a stable manner with RLP encoding genes (Thomas et al., 1997). For example, when *S. lycopersicum* Cf0 expressed the *Cf-4* gene in a stable manner, Cf-4 was able to induce a severe systemic necrotic response to PVX-mediated delivery of its corresponding effector Avr4. This is analogous to the response seen in the native Cf4 parent plant (Thomas et al., 1997).

In the current study, the response of T_2 and T_3 *S. lycopersicum* Cf0 35S:*2A* transgenic families to PVX:*Ecp2*-inoculation was weaker than the HR^+ seen in *S. pimpinellifolium* CfEcp2 to PVX:*Ecp2* inoculation (Table 4.3 and Figure 4.3). In addition, there were inconsistencies between the response to Ecp2 protein and PVX:*Ecp2* in some of the transgenic 35S:*2A* lines. Even though Ecp2 protein-infiltration into adult leaves of the T_1 plants gave a response, only a few of the resulting T_2 families tested responded to the PVX-mediated delivery of Ecp2 (Table 4.1, Table 4.3 and appendix 3 Table 3.1). This may

be because the Ecp2 protein infiltrations gave unspecific responses in the transgenic lines (Table 4.1). However, the response to PVX:*Ecp2* in the 35S:*2A* transgenic lines was specific. There was no response to PVX:*Ecp2* in *S. lycopersicum* Cf0. Furthermore, no response was seen in the T_3 transgenic families *S. lycopersicum* Cf0 35S:*2A*p3.8 and *S. lycopersicum* Cf0 35S:*2A*p24.4 to PVX:*Avr4* (Table 4.5).

The response to PVX:*Ecp2* was inherited from T_2 families into the T_3 families (Table 4.3). In the T_2 families *S. lycopersicum* Cf0 35S:*2A*p3 and *S. lycopersicum* Cf0 35S:*2A*p24, the response to PVX:*Ecp2* segregated (Table 4.3). This was further reflected in the derived T_3 families: some families retained the segregation for response to PVX:*Ecp2*, whilst others lost the ability to recognise PVX:*Ecp2* (Table 4.3 and Appendix 3, Table 3.1). However, failure to respond to PVX:*Ecp2* did not correlate with the absence of *2A*, PVX or Ecp2 transcripts (Table 4.4). Nevertheless, all T_2 plants, which showed a response to PVX:*Ecp2*-inoculation, had *2A* transcripts present (Table 4.4). Therefore, lack of response to PVX:Ecp2 despite presence of *2A* may be due to lack of penetrance of the PVX:*Ecp2* phenotype in the transgenic lines *S. lycopersicum* Cf0 35S:*2A*.

The genotyping of transgenic lines with PCR was problematic. Due to homology between *Hcr9s*, and the presence of *OR20A* in *S. lycopersicum* Cf0 (de Kock et al., 2005), weaker background-bands were occasionally observed within those lines not expected to carry *2A*, e.g. *S. lycopersicum* Cf0 35S:*Ecp2* (Figure 4.2). Such observations made it difficult to confirm the presence and expression of *2A* in the transgenic lines.

Since genotyping with PCR proved challenging, kanamycin-selection was explored to select for the presence of the transgene and to determine its copy number. Since the selectable marker on the 35S:*2A* binary construct contained the neophosphotransferase gene conferring kanamycin resistance, it was expected that 35S:*2A* would be inherited along with kanamycin resistance. Previously, stable transgenic tobacco plants expressing *Cf-9* were generated with a similar cassette carrying kanamycin-resistance (Hammond-Kosack et al., 1998). Resistance to kanamycin in these plants correlated with the presence of the *Cf-9* transgene, as determined by PCR analysis (Hammond-Kosack et al., 1998).

In the current study, kanamycin resistance selection was applied to a number of T_2 and T_3 families of *S. lycopersicum* Cf0 35S:*2A* (Table 4.6, Table 4.7, Figure 4.4 and Appendix 3, Table 3.2). However, low germination rates were observed in tissue culture (Table 4.6 and Table 4.7). This was neither attributed to the transgenic plants nor to the presence of kanamycin (Table 4.6 and Table 4.7), because (i) *S. pimpinellifolium* CfEcp2 plants had similar germination rates to the transgenic lines (Table 4.6) and (ii) low germination rates of the transgenic lines were also seen under conditions where kanamycin was absent. Due to the low germination rate in these experiments, the number of independently segregating 35S:*2A* transgene copies could not be ascertained. However, kanamycin-selection made it possible to select plants for generation advancement and test crossing that were assumed to carry 35S:*2A*.

There were some inconsistences in the response to Ecp2 protein-infiltrations and lack of penetrance of the PVX-mediated delivery of Ecp2 in the transgenic 35S:*2A* plants. Furthermore, a correlation between the response to Ecp2 and the presence of *2A* was not observed, when Ecp2 was delivered to the 35S:*2A* transgenic lines by protein-infiltration or PVX-mediated delivery. To establish a more robust phenotype in the transgenic 35S:*2A* lines to Ecp2 delivery, specific T_1, T_2 and T_3 35S:*2A* transgenic plants were crossed with *S. lycopersicum* Cf0 homozygous for the transgene 35S:*Ecp2* (Table 4.8). This enabled sustained delivery of Ecp2 from germination of the seed.

When *S. pimpinellifolium* CfEcp2 was crossed with *S. lycopersicum* Cf0 35S:*Ecp2*, all progeny showed an SLP and died (Table 4.8). This was also observed by Soumpourou et al. (2007). This is due to the expression of the *Cf-Ecp2* gene and its corresponding effector within the same plant. Similar results are seen when crossing plants expressing *Cf-9* or *Cf4* with plants expressing *Avr9* or *Avr4*, respectively (Hammond-Kosack et al., 1994b; Thomas et al., 1997). On the other hand, when *S. lycopersicum* Cf0 was crossed with *S. lycopersicum* Cf0 35S:*Ecp2* in the current study, SLP was not observed due to the lack of *Cf-Ecp2*-presence (Table 4.8). This was also found by Soumpourou et al. (2007).

The SLP observed in the progeny from *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2* was different to that observed in the progeny from *S. lycopersicum* Cf0 35S:2A x *S. lycopersicum* Cf0 35S:*Ecp2.* In the case of *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2*, the progeny developed cotyledons and appeared healthy before death

occurred. This phenotype was also reported earlier for the Cf9 and Avr9, and the Cf4 and Avr4 combinations (Hammond-Kosack et al., 1994b; Thomas et al., 1997). In contrast, the SLP in progeny from *S. lycopersicum* Cf0 35S:*2A* x *S. lycopersicum* Cf0 35S:*Ecp2* was observed from germination. This was manifested as seedlings with small and poorly expanding cotyledons followed by death of the whole seedling (Figure 4.5). In addition, intermediate phenotypes were observed in the progeny from *S. lycopersicum* Cf0 35S:*2A* x *S. lycopersicum* Cf0 35S:*Ecp2.* This was not seen in the progeny from *S. pimpinellifolium* CfEcp2 x *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.8 and Figure 4.5).

In three of the 35S:*2A* lines, the 35S:*Ecp2* test-cross progeny showed a pronounced abnormal development with long spindly petioles with small rounded leaflets (Figure 4.6). In the line *S. lycopersicum* Cf0 35S:*2A*p2 this phenotype was unequivocally correlated with the presence of the transgene (Table 4.9 and Figure 4.6).

Low levels of germination were observed in the testcross experiments (Table 4.8). Previously, low germination rates in crosses between a plant expressing the *R* gene and a plant expressing the corresponding effector, were associated with the direction of the cross (Hammond-Kosack et al., 1998). When tobacco plants stably expressing *Cf-9* were crossed to tobacco stably expressing *Avr9*, a low germination rate was seen when *Avr9* was delivered by the paternal parent (Hammond-Kosack et al., 1998). The small size of Avr9 enables it to cross plant cell walls and to diffuse into all tissues of the seed which are over expressing *Cf-9*. This means that the seedling will encounter the toxic products generated by the response associated with the recognition of Avr9 by Cf-9. This spread of toxic products throughout the seed results in the termination of the seed and no germination (Hammond-Kosack et al., 1998). However, in the current study low germination was not associated with the presence of the transgene since *S. lycopersicum* Cf0 also experienced low germination rates (Table 4.8). Furthermore, the direction of the cross did not affect the germination rate (data not shown).

Although there was a difference in the SLP observed if *S. lycopersicum* Cf0 35S:*Ecp2* was crossed to *S. pimpinellifolium* CfEcp2 or crossed to the transgenic 35S:*2A* lines, the response in the progeny was specific to crossing Ecp2 into these lines. When transgenic lines were selfed or when they were crossed to *S. lycopersicum* Cf0, SLP was not observed (Table 4.8).

The presence of the transgene 35S:*2A* was correlated with the SLP response in the progeny of crosses between *S. lycopersicum* Cf0 35S:*2A* and *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.10 and Figure 4.7). In the cross between the T_1 plant *S. lycopersicum* Cf0 35S:*2A*p3 with *S. lycopersicum* Cf0 35S:*Ecp2*, all of the plants that germinated showed SLP (Table 4.8). This phenotype was inherited and could be recapitulated in the T_2 testcrosses of this line with *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.8). All the testcross progeny displayed the SLP. This suggests that *S. lycopersicum* Cf0 35S:*2A*p3 carries more than one copy of 35S:*2A*. In the T_2, the selected line may have been homozygous for one of the copies. This observation is supported by the fact that all plants tested from the T_3 family *S. lycopersicum* Cf0 35S:*2A*p3.8 contained the 35S:*2A* transgene (Table 4.10 and Figure 4.7). On the other hand, progeny from the cross between the T_1 plant *S. lycopersicum* Cf0 35S:*2A*p24 with *S. lycopersicum* Cf0 35S:*Ecp2* segregated for SLP versus normal phenotype (Table 4.8). This was also seen in the T_2 crosses of this line with *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.8). However, in the T_3 crosses the presence of the transgene 35S:*2A* became fixed in the *S. lycopersicum* Cf0 35S:*2A*p24.4.2D T_3 family, as did the presence of SLP when this it was crossed to *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.10).

On the other hand, the absence of the transgene 35S:*2A* became fixed in the *S. lycopersicum* Cf0 35S:*2A*p24.5.5D T_3 family and all the progeny from the cross between this T_3 family and *S. lycopersicum* Cf0 35S:*Ecp2* showed a normal phenotype (Table 4.8 and Table 4.10). In the *S. lycopersicum* Cf0 35S:*2A*p24.4.4D, segregation for the presence of 35S:*2A* continued and correlated with the SLP response in the progeny from the cross between *S. lycopersicum* Cf0 35S:*2A*p24.4.4D with *S. lycopersicum* Cf0 35S:*Ecp2* (Table 4.10). The presence of *2A* therefore correlated with the presence of SLP when the transgenic line was crossed to a line homozygous for the expression of 35S:*Ecp2*.

To summarise, transient delivery of Ecp2 via protein infiltration or PVX-mediated delivery into 35S:*2A* transgenic lines did not give consistent responses. However, sustained delivery of Ecp2 from germination into the 35S:*2A* transgenic plant resulted in a strong and consistent SLP. This in turn could be unambiguously correlated with the presence of *2A*.

Three hypotheses are therefore proposed to explain this phenomenon. Firstly, an error was made in the assembly of the BAC 7B sequence or there was a deletion in the BAC 7B before it was sequenced and another variant of *2A* exists within the locus, which is not yet identified. It is well known that direct repeats in constructs can lead to genetic instability of a BAC in *E. coli*. Secondly, another component required for Ecp2 recognition lies within the genetically defined Ecp2-recognition locus in *S. pimpinellifolium* CfEcp2, and this component is absent or functionally diverged in *S. lycopersicum* Cf0. Thirdly, the 35S promoter interferes with the function of the *2A* gene.

With regards to the first hypothesis, the copy number of *2A* homologs is supported by MinION data (Chapter 3). In addition, many PCR clones were analysed for the presence of alternative *2A* alleles in both 7B and CfEcp2 genomic DNA but none were found (Chapter 3). The presence of another unidentified allele of *2A* in BAC 7B or the genome of *S. pimpinellifolium* CfEcp2 is therefore unlikely.

The second hypothesis indicates the requirement of a second component in the system, needed for *CfEcp2*-mediated recognition of Ecp2. This component may be the plant target of Ecp2 or it may be involved in the signal transduction of Ecp2 perception by *2A*. The precedence for such a model is provided by the Avr2, Rcr3, Cf-2 interaction. Avr2 is a cysteine protease inhibitor and is secreted by *C. fulvum* to inhibit the action of the cysteine proteases Rcr3 and Pip1, plant targets for Avr2 (Rooney et al., 2005). Rcr3 is required by Cf-2 for the recognition of *C. fulvum* effector Avr2 (Kruger et al., 2002; Rooney et al., 2005). Infiltration of Rcr3 or Avr2 proteins alone into Cf2 tomato plants do not trigger a response, whilst co-infiltration of Rcr3 and Avr2 proteins into Cf2 tomato plants triggers *Cf-2*–dependent HR (Rooney et al., 2005). Cf-2 therefore monitors Rcr3 and detects Avr2 when it binds Rcr3, resulting in the initiation of defence responses (Rooney et al., 2005). Similarly, the NB-LRR RPS5 detects cleavage of the kinase PBS1 at the apex of its activation loop by the *P. syrinage* effector AvrPphB, which is a cysteine protease (Ade et al., 2007; Qi et al., 2014; Zhu et al., 2004). Activation of RPS5-dependent resistance by AvrPphB is dependent on the presence of the effector target PBS1 (Ade et al., 2007; Swiderski and Innes, 2001).

In addition, it is predicted that the Cf-9–Avr9 interaction is indirect and mediated via an Avr9 plant target since no direct interaction has been identified between Cf-9 and Avr9

(Luderer et al., 2001). Avr9 is a cysteine knot protein and has no known function (Vervoort et al., 1997). Although no known inhibitor function for Avr9 has been identified, it is structurally homologous to carboxypeptidase inhibitors (Vervoort et al., 1997). A potential target for Avr9 has been identified in the plasma membrane of solanaceous plants (Kooman-Gersmann et al., 1996). Specific, saturated and reversible binding of radioactively labelled Avr9 has been identified on the membranes of tomato plants independent of the presence of Cf-9 (Kooman-Gersmann et al., 1996). Binding-affinity for this High Affinity Binding Site (HABS) on the surface of plant cell membranes by Avr9 has been correlated with *Cf-9*–dependent necrosis (Kooman-Gersmann et al., 1996; Kooman-Gersmann et al., 1998). The HABS of Avr9 has yet to be cloned. There is no known function for Ecp2 and no known plant target (Lauge et al., 1997; Van den Ackerveken et al., 1993a). This plant target may, as in the case of Avr2/Rcr3/Cf-2, be monitored by 2A (Cf-Ecp2) and its interaction with Ecp2 may be required to trigger *2A*-dependent resistance.

The small intracellular domain of RLPs contains no obvious signalling capacity (Jones et al., 1994). Therefore, in addition to the effector targets they guard, RLPs are likely to require other partners to transmit signals to the inside of the plant cell. The RLP Cf-4, has been shown to interact with the RLK SOBIR1 both in the presence and absence of Avr4 (Liebrand et al., 2013; Postma et al., 2015). SOBIR1 is involved in the stability of Cf-4 at the surface of the cell (Liebrand et al., 2013; Postma et al., 2015). When Avr4 is detected, Cf-4 interacts with BAK1 (Postma et al., 2015). This interaction is required for Cf-4–mediated Avr4 perception and resistance to *C. fulvum* (Postma et al., 2015). SOBIR1 and BAK1 homologs are also present in *Nicotiana* species. This explains why *Cf-4* is still functional when expressed in these species (Liebrand et al., 2014a; Postma et al., 2015). SOBIR1 and BAK1 have been shown to be required for multiple RLPs, which are more sequence-divergent to Cf-4 than 2A is to Cf-4. For example, tomato Ve1 requires SOBIR1 and BAK1 to mediate resistance to *V. dahlia*, and *A. thaliana* RLP30 requires SOBIR1 and BAK1 to function against the necrotrophic pathogen *Sclerotina sclerotiorum* (Liebrand et al., 2013; Zhang et al., 2013). Not all RLPs require both these RLKs to function. For example, RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASE-1 (RBPG1) only requires SOBIR1 but not BAK1 for functionality (Liebrand et al., 2014b). Therefore, it is probable that 2A requires SOBIR1 for functionality and potentially also BAK1.

It is possible that *2B* or *2C* are also required for *2A* function. *R* gene pairs have been identified as being required for effector recognition (Narusaka et al., 2009a; Narusaka et al., 2009b). RPS4 and RRS1 sit head to head in the genome of *A. thaliana*. These NB-LRR encoding genes are both required for the recognition of effectors from three pathogens, namely (i) AvrRps4 from *P. syringae*, (ii) PopP2 from *Ralstonia solanacearum* and (iii) an unkown component from *Colletotrichum higginsianum* (Narusaka et al., 2009a; Narusaka et al., 2009b). Similarly, the wheat *Lr10* and *RGA2* genes are required for recognition of *Puccinia triticina* (Loutre et al., 2009).

To test whether more than one of the RLPs sitting at the defined *Cf-Ecp2* locus are required for recognition of Ecp2, large sections of the BAC, carrying *2A*, *2B* and *2C*, could be transformed into *S. lycopersicum* Cf0. Previously, it was possible to identify *Cf-4* and *Cf-2* by transformation of overlapping binary-vector cosmid clones carrying many *Cf-4* or *Cf-2* candidates into tomato (Dixon et al., 1996; Thomas et al., 1997). However, it was not possible to move cosmids carrying multiple RLPs of the *OR* locus into *Agrobacterium*, since recombination occurred between the highly homologous genes (de Kock et al., 2005). This recombination between homologous sequences was also observed when transferring binary cosmid-vectors carrying multiple *Cf-5* candidate genes (Dixon et al., 1998). Furthermore, transient co-expression of *2A* and Ecp2, with either *2B* or *2C* within *Nicotiana* spp., did not elicit a response (de Kock and colleagues, 2004).

In the current study, other component candidates required for *2A* function were not identified in the sequence of the locus encoding for the absence of Ecp2-recognition (Chapter 3). Therefore, any other component required for *2A* function would reside within the locus encoding for the presence of Ecp2-recognition. This observation is analogous to the tomato *Prf* gene, a *NB-LRR* gene, which is genetically linked to *Pto*, a serine threonine protein kinase (Martin et al., 1993; Salmeron et al., 1996). *Prf* is required for *Pto*-dependent recognition of the effector AvrPto (Salmeron et al., 1996).

To explore this further in the current study, the transgene 35S:*2A* from T_3 plants *S. lycopersicum* Cf0 35S:*2A*p3.8.1 and *S. lycopersicum* Cf0 35S:*2A*p24.4.4 (along with others) was moved into the deletion mutant *S. pimpinellifolium* 1179 which lacks *2A* yet retains the flanking-markers for the presence of the Ecp2-recognition locus from *S. pimpinellifolium* CfEcp2. However, the presence of 35S:*2A* in the *S. pimpinellifolium*

background did not reconstitute the full HR^+ to PVX:*Ecp2*. When a response was observed, it was a weak necrosis (HR^0) as seen when the 35S:*2A* transgene was in the *S. lycopersicum* background (Table 4.11).

The lack of a response to PVX:*Ecp2* did not correlate with the absence of the transgene 35S:*2A* (Table 4.12). Lack of penetrance of the PVX:*Ecp2* phenotype therefore persisted, whether the 35S:*2A* transgene was in a heterozygous genetic background of *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0 or in a homozygous genetic background of *S. lycopersicum* Cf0.

Transient co-expression of *Cf-4* and *Avr4* or *Cf-9* and *Avr9*, respectively in *N. benthamiana* or *N. tabacum*, induces a HR (Thomas et al., 1997; Thomas et al., 2000; Van der Hoorn et al., 2001a). However, when *2A* and *Ecp2* were transiently co-expressed in tobacco (and other *Nicotiana* spp), this did not induce a HR (de Kock and colleagues, 2004). If another component is required for the function of *2A*, and it is not present in *S. lycopersicum* Cf0, it is unlikely to be present in the more evolutionary diverged *Nicotiana* species. The Nicotiana species tested so far for 2A functionality do not contain accessions capable of recognising Ecp2 (de Kock et al., 2004). A response would therefore not be expected when *2A* is transiently co-transformed with 35S:*Ecp2* into lines lacking the other component.

Within *N. paniculata*, some accessions respond (TW99) whereas others do not respond (TW102) to 35S:*Ecp2* transient transformation (de Kock et al., 2004; Lauge et al., 2000). The *N. paniculata* accessions TW102, TW99 and the F_1 products from a cross between these two accessions, were transiently co-transformed with 35S:*2A* and 35S:*Ecp2*. In addition, each of the constructs were transformed independently into each of the accessions. If another component was required for Ecp2-recognition in *N. paniculata* TW99 alongside *2A,* this would suggest that this component is also likely to be present in *N. paniculata* TW102. However, when *N. paniculata* TW102 was co-transformed with 35S:*2A* and 35S:*Ecp2* no response was observed (Figure 4.8). This lack of response may be due to the gene encoding for Ecp2-recognition in *N. paniculata* TW99 working in an entirely different manner to *2A* and the signalling apparatus required by *2A* is missing in *Nicotiana* spp.

The third hypothesis presumes that sensitivity of *2A* protein levels plays a role in the function of *2A*. The presence of mRNA transcripts of *2A* in *S. pimpinellifolium* CfEcp2 has previously been observed (de Kock and colleagues, 2004). The strength of bands generated with amplification of *2A* from cDNA was found in the current study to be much greater in the T_2 transgenic lines (expressing 35S:*2A*) than in the native *S. pimpinellifolium* CfEcp2 lines tested (Figure 4.2). The native promoter in the *S. pimpinellifolium* CfEpc2 plants therefore induces much lower transcription rates of *2A* compared to those in the transgenic 35S:*2A* lines.

Over-expression of *2A* may cause saturation of the RLP at the surface of the cell. Another component required by 2A for the perception of Ecp2 or downstream signalling is unlikely to be present at an equally high abundance when *2A* is over-expressed. There may therefore be competition for recognition of Ecp2 by 2A alone vs 2A in conjunction with another component. Recognition of Ecp2 by 2A alone would not trigger a response. However, recognition of Ecp2 by 2A and the second component would result in HR within the cell. It is therefore conceivable that the over-expression of 2A at the surface of the cell sequesters Ecp2 from active 2A-recognition/signalling component complexes.

This component, is not likely to be required by Cf-4 because over-expression of Cf-4 in tomato does not affect the function of Cf-4–mediated recognition of Avr4 (Joosten et al., 1994; Thomas et al., 1997). This second component is therefore more likely to be specific to Cf-Ecp2 function, such as would be the effector target of Ecp2.

It is proposed that the third hypothesis could be further analysed by stable transformation of *S. lycopersicum* Cf0 and *S. pimpinellifolium* 1179p19 with *2A*, driven by the full-length native promoter of 6 kb upstream from the ATG start codon. This is in contrast to the 1 kb promoter upstream of *2A*, used by de Kock and colleagues (2004) in the generation of *2A* stable transformants in *S. lycopersicum* Cf0, which did not recognise Ecp2 and lacked the presence of *2A* mRNA.

In the current study, the response of *S. lycopersicum* Cf0 35S:*2A* stable transformants to Ecp2 was found to depend on the delivery mechanism of Ecp2, with sustained delivery from germination generating a consistent and measurable recognition response. It is possible that *2A* requires a second component for signalling the recognition of Ecp2.

Furthermore, it is likely that 35S promoter affects the functioning of *2A*. For the future, transmission of *Cf-Ecp2* mediated resistance into other plant species, resolving the problems associated with the function of the 35S:*2A* transgene, will be paramount.

Chapter 5

Characterisation of Ecp2 recognition in *Nicotiana paniculata*

5.1 Introduction

To generate disease resistant crops, *R* genes have been introgressed from wild relatives into the genomes of cultivated varieties (Hajjar and Hodgkin, 2007). However, *R* gene introgression requires intensive back-crossing and the plant species involved must be sexually-compatible (Hajjar and Hodgkin, 2007). Furthermore, despite back-crossing, *R* gene introgression often leads to the co-introduction of genetically linked deleterious alleles (linkage drag) (Fukuoka et al., 2009; Hajjar and Hodgkin, 2007; Parniske et al., 1997; Parniske et al., 1999). Genetic modification enables the transfer of *R* genes between distant relatives, breaking down all species barriers (Wulff et al., 2011) and also avoids linkage drag. Furthermore, methods of genome editing enables the targeted alteration of a gene in a plant genome, thereby conferring resistance to a target pathogen (Liu et al., 2016).

Genetic modification and genome editing require the cloning of the *R* gene. This process can present a number of challenges (see Chapter 3). To aid the process of cloning *R* genes, genetic maps are generated from crosses between a resistant accession and susceptible accession. This enables the genetic positioning of *R* genes within the genome.

In addition to cloning *R* genes from a plant host species, *R* genes and other forms of resistance can be found in plant species that are non-hosts to a given pathogen (Borhan et al., 2008; Bos et al., 2006; Heath, 2000; Kamoun et al., 1998; Kanzaki et al., 2008). Non-hosts are those plants, in which the entire species is resistant to all isolates of a pathogen species (Heath, 2000; Schulze-Lefert and Panstruga, 2011). Non-host resistance can be conferred by preformed defences or inducible defence responses (Heath, 2000). Preformed defences include peptides, proteins and non-proteinaceous metabolites (Heath, 2000). Non-host inducible defence responses are believed to have similarities with mechanisms of host-mediated pathogen resistance. They rely on molecular mechanisms triggered by PRRs and *R* gene products (Heath, 2000; Schulze-Lefert and Panstruga, 2011). For example, the NB-LRR encoding *R* gene, *WRR4,* of *Arabidopsis*

thaliana, confers non-host resistance against subspecies of *Albugo candida* (Borhan et al., 2008). Similarly, the lectin-like receptor kinase (LRK) encoding *R* gene, from *N. benthaminana*, confers recognition of the INF1 elicitor from *Phytophthora infestans*, resulting in resistance to this Oomycete in most *Nicotiana* species (Kamoun et al., 1998; Kanzaki et al., 2008).

Nicotiana species are non-hosts of the fungal pathogen *C. fulvum* (Bond, 1938; Lauge et al., 2000). However, some *Nicotiana* species have been shown to recognise the *C. fulvum* effector Ecp2 (de Kock et al., 2004; Lauge et al., 2000). Out of a total of 71 accessions tested across 38 *Nicotiana* species, all accessions tested from *N. sylvestris*, *N. undulata* and *N. tobaccum* species responded with a HR upon delivery of PVX:*Ecp2* (de Kock et al., 2004). Interestingly, whilst the *N. paniculata* accessions TW99 and TW101 responded with HR following inoculation with PVX:*Ecp2*, no response was observed in the accessions TW100 (de Kock et al., 2004). Furthermore, the closely-related *N. ramondii* accession TW102 failed to respond to the infiltration of PVX:*Ecp2* (de Kock et al., 2004). The F_2 from crosses between the accessions TW99/TW101 and TW100 or TW99 and TW102 showed a 3:1 segregation for recognition to no recognition (de Kock et al., 2004; Harder, 2012). This segregation suggests the presence of a single dominant gene in the accession TW99, controlling recognition of Ecp2. However, some F_2 plants recognising Ecp2, showed a weaker HR phenotype (+) than the strong HR (++) seen in the parent accession (Figure 2.2) (Harder, 2012). In fact, a 1:2:1 segregation was observed in the F_2 for no recognition to partial response to strong HR (Harder, 2012). These results suggest that recognition of Ecp2 may be encoded by a semi-dominant gene in *N. paniculata* TW99 (Harder, 2012). The gene encoding recognition of Ecp2 in *N. paniculata* was termed *Cf-Ecp2* (Harder, 2012).

A high rate of polymorphism was identified between the accessions TW99 and TW102, via the analysis of 8.044 kb of genetic sequence from 14 conserved orthologous set (COS) markers (Harder, 2012). These accessions should therefore be ideal for the generation of a genetic linkage map of *N. paniculata* and the possibility of mapping the *Cf-Ecp2* gene encoding for recognition of Ecp2.

Tomato is a host plant for *C. fulvum*. Recognition of Ecp2 has been identified in wild currant tomato *S. pimpinellifolium* CfEcp2 (Lauge et al., 1998). Recognition of Ecp2, is

conferred by a single dominant gene, *Cf-Ecp2*, and results in a HR (Haanstra et al., 1999; Lauge et al., 1998). *Cf-Ecp2* has been mapped to a sequenced and assembled 91 kb locus on the short arm of chromosome 1, in the genome of *S. pimpinellifolium* CfEcp2 (Chapter 3). The candidate for *Cf-Ecp2*, *2A*, encodes a predicted RLP (Chapter 3 and 4)(de Kock and colleagues, 2004).

Recognition of Ecp2 could have evolved in *Nicotiana* spp. in response to two events. Firstly, Ecp2-recognition in *Nicotiana* spp. could enable *Nicotiana* to exist as a non-host of *C. fulvum*. Secondly, Ecp2 may exist as an effector in a pathogen of *Nicotiana* spp. Recognition of Ecp2 conferred by an *R* gene may therefore have been positively selected over the course of evolution to enable ETI in *Nicotiana* to such a pathogen.

Indeed, the Ecp2 effector is conserved across the fungal class Dothidiomycetes and homologs of the *C. fulvum* Ecp2 sequence have been identified in not only important plant pathogens but also human pathogens and saprophytic fungi (Stergiopoulos et al., 2012; Stergiopoulos et al., 2010). Important crop pathogens such as *F. graminearum*, *F. oxysporum*, *Magnaporthe grisea*, *M. fijiensis*, *Z. tritici*, *Septoria musiva* and *V. dahliae*, infect a wide variety of plants, yet all contain homologs of Ecp2 (Table 5.1)(Stergiopoulos et al., 2012). Interestingly, *Nicotiana* spp. are hosts for the pathogens *F. oxysporium* and *V. dahliea* (Mace et al., 1981). Moreover, the Ecp2 homolog from *M. fijiensis* (MfEcp2-1) and *D. septosporum* are recognised by *S. pimpinellifolium* CfEcp2 (de Wit et al., 2012; Stergiopoulos et al., 2010). In addition, there are homologs of the *C. fulvum* effector Avr4 in both *M. fijiensis* (MfAvr4) and *D. septosporum* (DsAvr4)(de Wit et al., 2012; Stergiopoulos et al., 2010). Both MfAvr4 and DsAvr4 are recognised by the tomato *R* gene encoded RLP, Cf-4 (de Wit et al., 2012; Stergiopoulos et al., 2010).

Table 5.1. Important plant pathogens carrying homologs of Ecp2.

Species of pathogen	Disease	Host species	Pathogen lifestyle
Cladosporium fulvum	Tomato leaf mould	Tomato	Hemibiotroph
Dothistroma septosporum	Needle blight	Conifers	
Fusarium graminearum	Fusarium head blight	Grain cereals	
Fusarium oxysporum	Fusarium wilt	Various plants	Saprophytic
Magnaporthe grisea	Rice blast	Rice and other cereals	
Mycosphaerella fijiensis	Black sigatoka	Banana	
Zymoseptoria tritici	Septoria tritici blotch	Wheat	
Septoria musiva		Hybrid poplar plantations	
Verticillium dahliae	Verticillium wilt	Many including tomato and *Nicotiana*	

If Ecp2 homologs from major crop pathogens are recognised by *N. paniculata* CfEcp2 there is the potential to transfer this gene to crop varieties by genetic modification, which could in turn generate resistance to these pathogens. This is of particular interest for *M. fijiensis*, which infects banana, causing Black Sigatoka disease. The commonly-cultivated banana, Cavendish, is a clone and is sterile (Koeppel, 2008). It is therefore not amenable to the introgression of resistance from wild relatives. As a solution, genetic modification of banana has the potential to enable the transfer of well-studied plant resistances, such as *Cf-Ecp2* from *N. paniculata* or tomato and *Cf-4* from tomato, into banana.

However, the functional transfer of an RLP (or other *R* gene products) from a *Solanaceous* species into banana requires interacting factors and components of the downstream signaling cascade, leading to resistance responses, to be conserved across monocots and dicots (Wulff et al., 2011). Any component not present in the plant receiving the *R* gene, but required for *R* gene function, must be co-transferred with the *R* gene. This will then enable the *R* gene to function in the new host (Ade et al., 2007; Wulff et al., 2011).

The recognition spectrum of *N. paniculata* TW99 CfEcp2 may be determined for many Ecp2 homologs from important crop pathogens. The cloning of CfEcp2 from *N. paniculata* is important due to its potential use in generating resistance in a number of crops. Furthermore, it is an interesting feature to consider when comparing the *R* genes conferring recognition of Ecp2 in the *C. fulvum* non-host *N. paniculata* with those in the host tomato. In the present study, a genetic map of *N. paniculata* was generated to facilitate the genetic positioning of *Cf-Ecp2*.

5.2 Results

5.2.1 Characterisation of *N. paniculata* CfEcp2

The effector Ecp2 is conserved across the fungal class Dothidiomycetes. Many fungal crop pathogens carry homologs of Ecp2. Previously, homologs of Ecp2 (containing the Ecp2 domain class I) have been selected from fungal species that are important crop pathogens (Stergiopoulos et al., 2012). In the current study, a multiple alignment was used to determine the relationship between the amino acid sequences of various Ecp2 homologs (Table 5.2 and Figure 5.1).

Table 5.2. Pairwize comparisons of Ecp2 homolog amino acid sequences using ClustalW.

		% identity to protein									
#	**Description of protein**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	*Cladosporium fulvum* Avr4	100	16	16	16	12	15	17	10	7	10
2	*Zymoseptoria tritici* Ecp2	16	100	20	15	26	25	27	8	10	10
3	*Septoria musiva* Ecp2	16	20	100	20	25	27	26	22	9	14
4	*Fusarium oxysporum* Ecp2	16	15	20	100	23	23	21	13	13	17
5	*Cladosporium fulvum* Ecp2	12	26	25	23	100	59	57	15	15	11
6	*Dothistroma septosporum* Ecp2	15	25	27	23	59	100	69	17	15	16
7	*Mycosapharella fijiensis* Ecp2	17	27	26	21	57	69	100	15	16	14
8	*Verticillium dahliae* Ecp2	10	8	22	13	15	17	15	100	14	14
9	*Fusarium graminearum* Ecp2	7	10	9	13	15	15	16	14	100	14
10	*Magnaporthe grisea* Ecp2	10	10	14	17	11	16	14	14	14	100

= assigned protein number. % identity = % identity between proteins calculated using ClustalW.

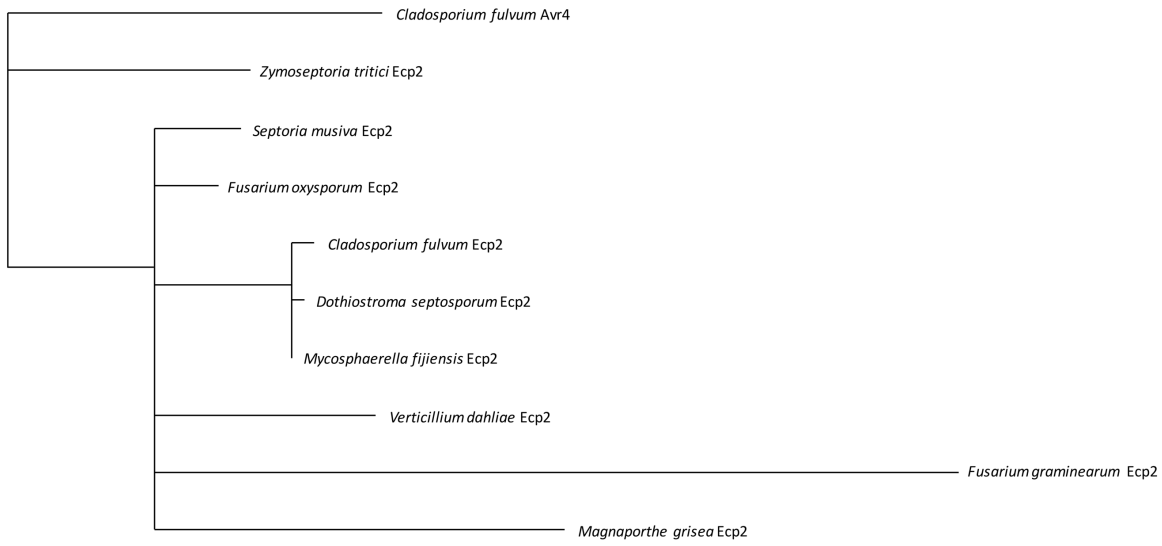


Figure 5.1. Sequence-based phylogeny of Ecp2 from crop pathogens.

Distance tree of amino acid sequences of Ecp2 effector variants from different crop pathogens calculated from ClustalW pairwise comparisons in Geneious (Table 5.2).

Ecp2 homologs are very different from each other, with pairwise identities ranging between 11 and 69 % (Table 5.2). These effector sequences were cloned into a binary vector for Agrobacterium-mediated transient expression in *Nicotiana* sp. To determine the recognition spectrum of *N. paniculata Cf-Ecp2*, the Ecp2 clones were transiently expressed in the *N. paniculata* accessions TW99 (CfEcp2), TW102 and the F_1 alongside the *C. fulvum* Ecp2 construct from Soumpourou et al. (2007) (Figure 5.2). Responses were scored and photographs taken 10 d.p.i. (Table 5.3, Figure 5.2 and Appendix 4, Table 4.1).

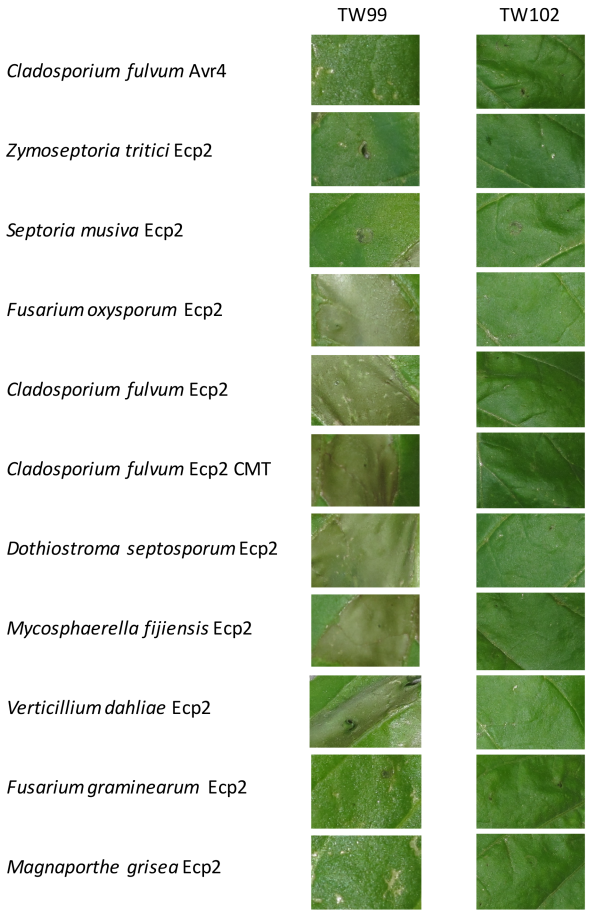


Figure 5.2. Characterisation of *N. paniculata* CfEcp2 by infiltration of 35S:*Ecp2* homologs.

Homologs of Ecp2 from different crop species were delivered by Agrobacterium-mediated transient transformation into *N. paniculata* accessions TW99 (CfEcp2) and TW102 (cfecp2). Responses were scored and photographs taken 10 days post inoculation.

None of the Ecp2 effectors activated a HR response when expressed in TW102 (Table 5.3 and Figure 5.2). *Agroabacterium* mediated expression of *C. fulvum* Ecp2 in TW99 and the F_1 plants resulted in a ++ phenotype (Table 5.3, Figure 5.2 and Appendix 4, Table 4.1). In addition, *Agrobacterium*-mediated expression of Ecp2 homologs from the crop pathogens *D. septosporum, V. dahliae, F. oxysporum* and *M. fijiensis* gave either a HR manifested as a + or ++ phenotype in a high percentage of the TW99 and F_1 leaves (Table 5.3 and Figure 5.2). However, when the Ecp2 homologs from *V. dahliae, F. oxysporum* and *M. fijiensis* were transiently expressed in F_1 plants, a lower percentage of the inoculated leaves responded with + or ++ compared to transient expression in TW99 (Table 5.3). This indicates that a dilution of the resistance gene in the heterozygous background affects the recognition of these Ecp2 homologs. Although Ecp2 from additional crop pathogens, *F. graminearum*, *Magnaporthe grisea* and *M. fijiensis*, also resulted in a + or ++ phenotype in some instances (when transiently expressed in the leaves of TW99 and F_1 plants), very few of the leaves responded in this way (Table 5.3

and Figure 5.2). This was also seen when the negative control *C. fulvum* Avr4 was transiently expressed in TW99. In this case, 5% of the inoculated leaves responded with + phenotype (Table 5.3 and Figure 5.2).

Table 5.3. Response of *N. paniculata* accessions to Agrobacterium-mediated transient expression of Ecp2 homologs from pathogens of important crop species.

Fungal pathogen	Effector	*Nicotiana paniculata* accession					
		TW99		TW102		F_1	
		% ++/+	total	% ++/+	total	% ++/+	total
Cadosporium fulvum	Avr4	5	19	0	11	0	10
Zymoseptoria tritici	Ecp2	14	14	0	7	33	3
Septoria musiva	Ecp2	62	13	0	11	22	9
Fusarium oxysporum	Ecp2	85	13	0	11	67	9
Cladosporium fulvum	Ecp2	100	13	0	11	100	9
[1]*Cladosporium fulvum*	Ecp2	100	29	0	22	100	20
Dothistroma septosporum	Ecp2	100	13	0	11	100	9
Mycosphaerella fijiensis	Ecp2	85	13	0	11	67	9
Verticillium dahliae	Ecp2	69	13	0	11	33	9
Fusarium graminearum	Ecp2	54	13	0	11	11	9
Magnaporthe grisea	Ecp2	31	13	0	11	11	9

[1]Construct from Soumpourou et al. (2007), total = total number of leaf sections scored, + = partial hypersensitive response (HR) manifested as partial necrosis, ++ = HR with confluent death of inoculated section. Blue = high percentage of leaves showed response in both TW99 and F_1.

5.2.2 Phenotyping *N. paniculata* F_2 populations segregating for *Cf-Ecp2*

Note: The majority of work described in section 5.2.2 was performed by Stuart Harder and Marie Wulff. These unpublished results need to be described here as a necessary prerequisite for analysis described in this thesis.

To generate a population segregating for the presence of *Cf-Ecp2*, F_2 populations were generated by crossing *N. paniculata* accession TW99 (CfEcp2) with *N. paniculata* accession TW102 (cfecp2) in either direction. For each of the F_2 populations, 46 to 48 individuals were characterised for their ability to recognise *Ecp2* by *Agrobacterium*-mediated transient transformation of 35S:*Ecp2* into their leaves. The responses, scored 2 to 4 d.p.i., were identified as either 0, + or ++ (Table 5.4 and Figure 2.2)(Harder, 2012).

It was hypothesised that plants, which were scored as 0, were homozygous for the absence of *Cf-Ecp2*, those scored as ++ were homozygous for the presence of *Cf-Ecp2*, whilst those scored as + were heterozygous for *Cf-Ecp2*. The F_2 progeny segregated in a 1:2:1 manner (0:+:++). *Cf-Ecp2* was therefore classed as a semi-dominant gene (Harder, 2012) (Table 5.4).

To investigate this hypothesis further, 20 F_3 families from F_2 plants with the Ecp2 response phenotype 0 or ++, were tested for their response to Agrobacterium-mediated expression of *Ecp2* (Table 5.4). Ten F_3 families derived from F_2 plants with a ++ phenotype were tested. Seven out of those were found to segregate for Ecp2 response. Nine out of 10 F_3 families that were derived from F_2 plants with a 0 phenotype also showed a 0 phenotype. One out of the 10 F_3 families derived from F_2 plants with a 0 phenotype contained one plant, which showed a partial necrosis in response to 35S:*Ecp2* expression. However, this was classed as a non-specific necrosis and this family (NP_00171) was scored as homozygous (Table 5.4).

Table 5.4. Phenotype of *N. paniculata* germplasm in response to Agrobacterium-mediated transient expression of Ecp2.

N. paniculata germplasm	Family number	F_2 phenotype	Number tested	Segregation			F_2 model
				0	+	++	
[2]*N. paniculata* TW99 (CfEcp2)	NP_00006	N/A	5	0	0	5	Hom
[2]*N. paniculata* TW102 (cfecp2)	NP_00009	N/A	5	5	0	0	Hom
[2]*N. paniculata* TW99 x TW102 F_1	NP_00019	N/A	5	0	0	5	Het
[2]*N. paniculata* TW102 x TW99 F_1	NP_00022	N/A	5	0	0	5	Het
[1]*N. paniculata* TW99 x TW102 F_2	NP_00036	N/A	44	11	20	13	Seg
[1]*N. paniculata* TW102 x TW99 F_2	NP_00038	N/A	48	15	19	14	Seg
[2]*N. paniculata* TW99 x TW102 F_3	NP_00145	0	14	14	0	0	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00157	0	11	11	0	0	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00171	0	14	12	1	0	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00176	0	8	8	0	0	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00179	0	14	14	0	0	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00187	0	14	14	0	0	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00191	0	14	14	0	0	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00196	0	14	14	0	0	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00202	0	14	14	0	0	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00208	0	14	14	0	0	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00137	++	11	0	2	9	Hom
[2]*N. paniculata* TW99 x TW102 F_3	NP_00140	++	11	1	3	7	Seg
[2]*N. paniculata* TW99 x TW102 F_3	NP_00148	++	11	3	0	8	Seg
[2]*N. paniculata* TW99 x TW102 F_3	NP_00152	++	11	2	1	8	Seg
[2]*N. paniculata* TW99 x TW102 F_3	NP_00162	++	11	1	0	10	Seg
[2]*N. paniculata* TW102 x TW99 F_3	NP_00163	++	10	0	0	10	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00188	++	11	1	1	9	Seg
[2]*N. paniculata* TW102 x TW99 F_3	NP_00190	++	11	0	2	9	Hom
[2]*N. paniculata* TW102 x TW99 F_3	NP_00213	++	11	2	2	7	Seg
[2]*N. paniculata* TW102 x TW99 F_3	NP_00231	++	11	3	1	7	Seg

[1] Data from (Harder, 2012), [2]Combined data from the current study and (Westergaard, 2012). 0 = no response, + = partial hypersensitive response (HR) manifested as partial necrosis, ++ = HR with confluent death of inoculated section, N/A = not applicable, Hom= homozygous for presence or absence of Cf-Ecp2, Het = heterozygous in F_1 for Cf-Ecp2, Seg = segregating in F_2 populations for presence of *Cf-Ecp2*.

5.2.3 Construction of a genetic map for *N. paniculata*

A total of 75 Sequenom markers were designed from 47 COSII markers (Wu et al., 2010) and run on the MassAnalyser. The MassAnalyser scored the genotypes for each marker in TW99, TW102, the F_1, 46 individuals from the NP_00036 F_3 population and 46 individuals from the NP_00038 F_3 population, previously phenotyped with 35S:*Ecp2*. The output was curated to combine redundant genotype calls, resulting in genotypes covering 39 COSII markers. These markers were combined with 42 non-redundant markers generated by Westergaard (2012). A genetic map was generated for *N. paniculata* from the 81 Sequenom markers using Map Manager QTX v 0.30.

The *N. paniculata* genetic map contained 24 linkage groups, covering 620 cM of the genome (Figure 5.3). It was assumed that the markers at the end of linkage groups, but not at the telomere, were able to detect linkage of up to 10 cM beyond the linkage group. Consequently, there was the potential to identify linkage in approximately 920 cM of the *N. paniculata* genome. This is equivalent to approximately 87% of the genome. A recombination fraction plot was generated, which showed linkage between markers within linkage groups but not between linkage groups (Figure 5.4).

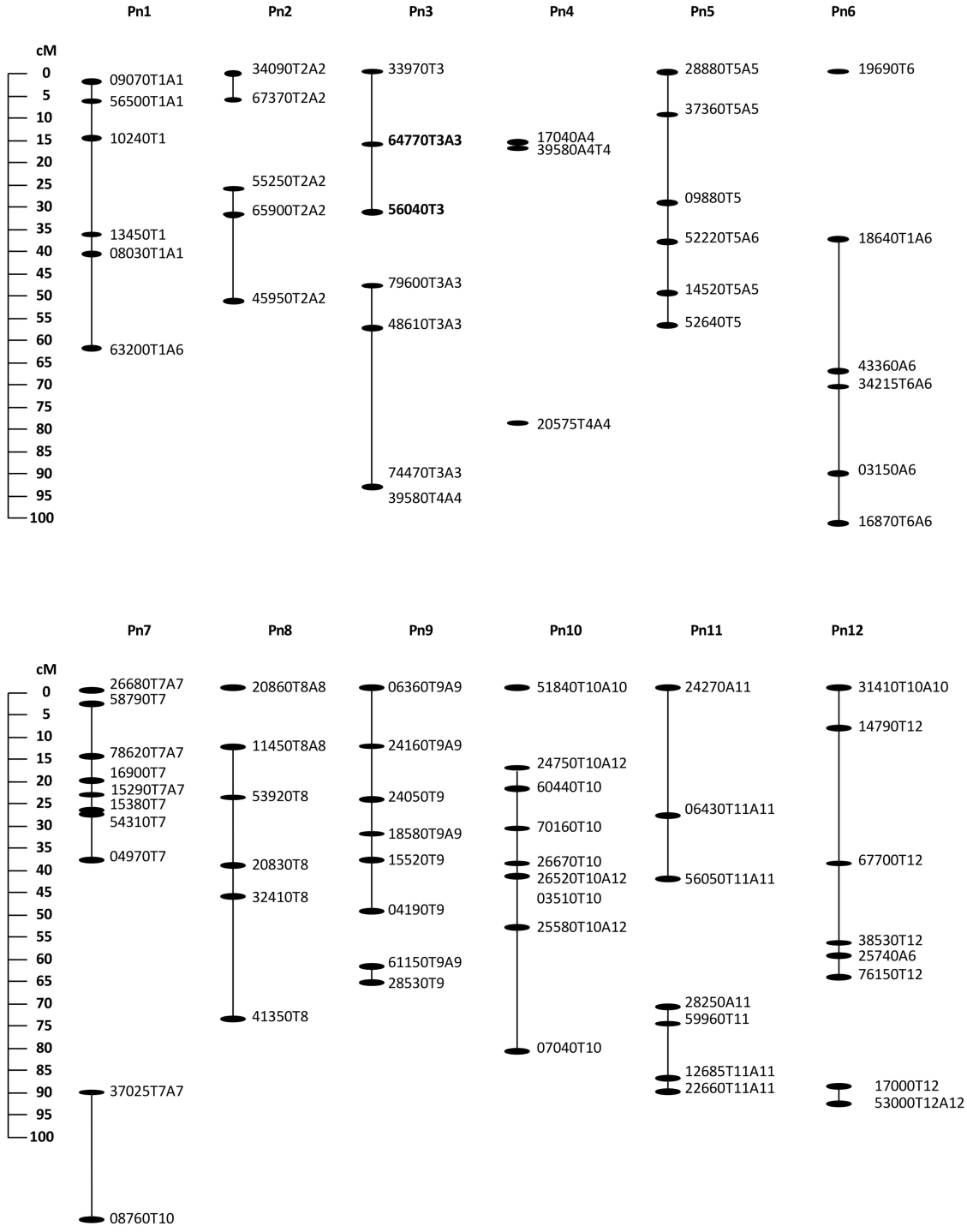


Figure 5.3. Genetic map of *N. paniculata*.

The genetic map of *N. paniculata was* generated using Map Manager QTX v0.30. Linkage is indicated by black lines between markers. Linkage groups are orientated to each other and assigned to chromosomes based on synteny to *N. acuminata* and/or *N. tomentosiformis* maps. Genetic linkage was identified between markers in **bold** and *cf-ecp2* using Chi-squared analysis.

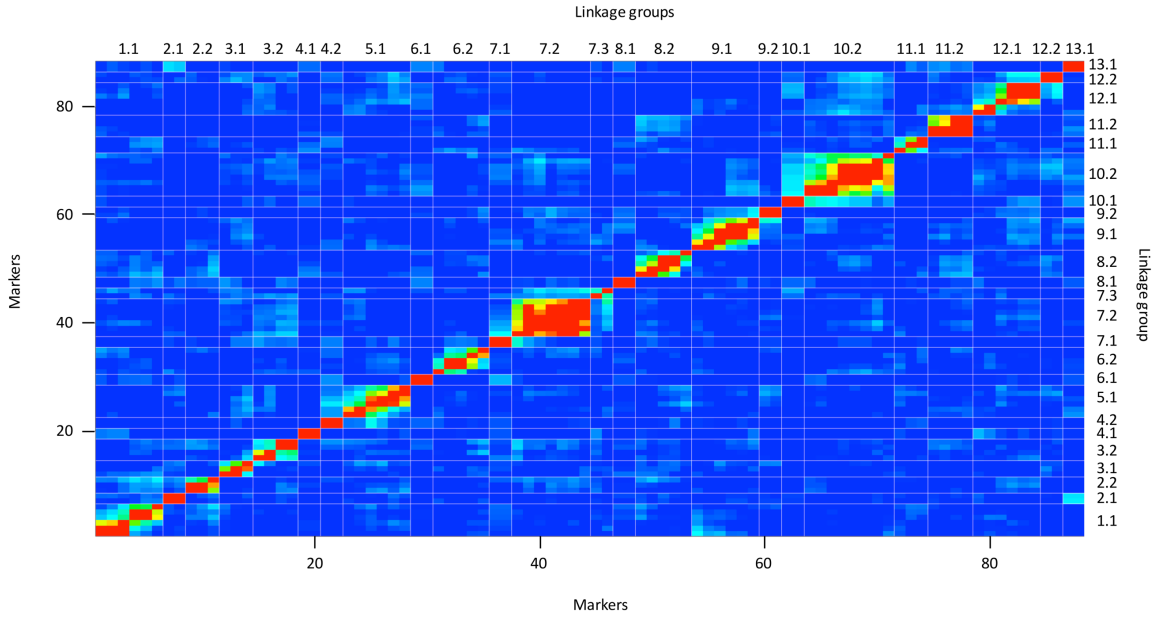


Figure 5.4. Heat map representation of recombination fraction plot of markers in *N. paniculata* linkage groups.

The recombination fractions are in the upper left triangle and Logarithm of the odds (LOD) scores are in the lower right triangle. Scores are represented in a heat map; red indicates a high LOD score or low recombination fraction. Blue indicates a low LOD score or high recombination fraction.

The 24 linkage groups of the *N. paniculata* genetic map were further ordered into 12 chromosomes, by virtue of synteny or inferred synteny with the *N. tomentosiformis* and *N. acuminata* maps (Figure 5.3). The genetic map of *N. paniculata* showed a high degree of synteny to that of *N. tomentosiformis* (Figure 5.3). For example, chromosome 8 of *N. tomentosiformis* had conserved marker order with linkage groups 8.1 and 8.2 of *N. paniculata* (Appendix 4, Table 4.2). The same is true for chromosome 9 of *N. tomentosiformis* and linkage groups 9.1 and 9.2 of *N. paniculata* (Appendix 4, Table 4.2).

Some inversions were also identified. For example, the first four markers of linkage group 1 of *N. paniculata* were inverted in contrast to those markers on chromosome 1 of *N. tomentosiformis* (Appendix 4, Table 4.2). In addition, translocations of markers were identified. For example, chromosome 7 of *N. tomentosiformis* shares marker order with *N. paniculata* linkage groups 7.1, 7.2 and 7.3. However, the final marker of linkage group 7.3 is present on *N. tomentosiformis* chromosome 10 (Figure 5.3 and Appendix 4, Table 4.2). There is also homology between the *N. acuminata* and *N. paniculata* marker order (Figure 5.3 and Appendix 4, Table 4.2).

5.2.4 Marker association analysis for Ecp2 recognition

Associations between the segregation of Sequenom marker genotypes and the segregation of Ecp2 recognition were investigated using the ChiSquared test on the F_2 individuals that did not show recognition of 35S:*Ecp2* (i.e. they showed a 0 or cfecp2 phenotype when inoculated with 35S:*Ecp2*). This identified the markers that were significantly different from a 1:2:1 segregation ratio and therefore either showed loss of heterozygocity, linkage of the cfecp2 phenotype to the *Cf-Ecp2* genotype or linkage of the cfecp2 phenotype to the *cfecp2* genotype.

Markers within linkage groups 3.1 and 3.2 (namely markers 64770, 56040, 79600 and 48610) showed linkage of the cfecp2 phenotype to the *CfEcp2* genotype (Table 5.5). Other markers in linkage groups showed linkage between the cfecp2 phenotype and the *CfEcp2* genotype. They were either in a linkage group in the absence of other markers, at the end of linkage groups or within a linkage group surrounded by markers that did not support the linkage observed (Table 5.5). If linkage group 10.2 was orientated in the opposite direction to *N. tomentosiformis*, between linkage groups 10.1 and 10.2, there is a possible linkage between the cfecp2 phenotype and the *CfEcp2* genotype (Table 5.5). In addition, loss of herterozygocity was identified in markers 55250, 03150, 16870 and 60440. However, no linkage was identified between the cfecp2 phenotype and the *cf-ecp2* genotype.

Table 5.5. Chi-squared (χ^2) analysis of genotypes of F_2 individuals that do not recognise Ecp2.

Marker	linkage group	χ^2	≥0.05	≥0.09	Direction of linkage	Position of markers
55250	2.1	0.03	*	*	loss of het	End of linkage group
64770	3.1	0.05		*	CfEcp2	At the end of two linkage groups.
56040	3.1	0.005	*	*		
79600	3.2	0.09		*		
48610	3.2	0.05		*		
52220	5.1	0.06		*	CfEcp2	Within linkage group. Distal and proximal markers do not show linkage.
3150	6.2	0.001	*	*	loss of het	Lie next to each other at end of linkage group 6.2.
16870	6.2	0.05	*	*		
24750	10.1	0.06		*	CfEcp2	End of linkage group
60440	10.2	0.09		*	loss of het	Within linkage group. Distal and proximal markers do not show linkage.
7040	10.2	9.80E-09	*	*	CfEcp2	End of linkage group.
26680	7.1	0.003	*	*	CfEcp2	Not in linkage group with other markers.
37025	7.3	0.04	*	*	CfEcp2	End of linkage group.

Het = heterozygocity.

5.2.5 QTL analysis indicates no linkage between markers and Ecp2 recognition

An analysis of linkage between markers within linkage groups and recognition of Ecp2 in *N. paniculata* was performed using QTL analysis (Figure 5.5). In order for a marker to show linkage to Ecp2 recognition, and for this linkage not to be by chance alone in the population of 94 F_2 individuals, the makers would have an LOD score >3.4. Although peaks were observed on linkage groups 3.1 and 9.1, the LOD score was approximately 2.2 and thus <3.4. This QTL analysis showed no significant linkage between markers and the recognition of Ecp2 in the *N. paniculata* genetic map.

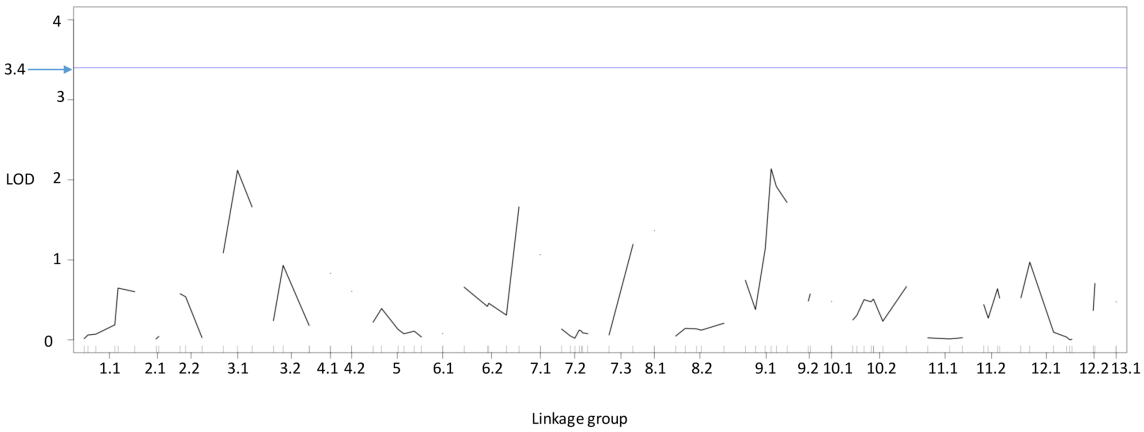


Figure 5.5 QTL analysis for linkage to recognition of Ecp2 in *N. paniculata*.

Logarithm of the odds (LOD) score >3.4 (blue arrow and blue line) indicates threshold for linkage between marker and recognition of Ecp2 in *N. paniculata*. QTL = quantitative trait loci.

5.3 Discussion

There are homologs of Ecp2 present in a number of plant pathogens. Multiple alignment of the amino acid sequence of these homologs showed that the sequence of the homologs was not conserved across the pathogens (Table 5.2 and Figure 5.1). The non-host *N. paniculata* (accession TW99) recognised the effector Ecp2 from *C. fulvum* (Figure 5.2). Recognition of Ecp2 homologs from the plant pathogens *D. septosporum*, *V. dahliae*, *F. oxysporum* and *M. fijiensis* was also identified in TW99 in comparison to no response in *N. paniculata* TW102 (Table 5.3 and Figure 5.2). However, there was no recognition of *M. graminicola*, *F. graminearum* and *Magnaporthe grisea* Ecp2 homologs or *C. fulvum* Avr4 in both *N. paniculata* TW99 and TW102 (Table 5.3 and Figure 5.2).

Both *D. septosporum* and *M. fijiensis* Ecp2 homologs are the most closely related to *C. fulvum* Ecp2 within the range of pathogens tested (Figure 5.1). However, it is assumed that a structural recognition rather than sequence recognition of Ecp2 is conferred by *Cf-Ecp2* because the very distant homologs of *C. fulvum*, *V. dahlia* and *F. oxysporum* Ecp2 were also recognised by *N. paniculata* CfEcp2 (Table 5.2, Table 5.3, Figure 5.1 and Figure 5.2). As an extension of the current study, it may be of interest to compare the recognition spectrum of *N. paniculata* CfEcp2 with that of the *C. fulvum* host *S. pimpinellifolium* CfEcp2 and to look for similarities and discrepancies.

If *N. paniculata* Cf-Ecp2 can be cloned, there is the potential to move this gene into other plant species to generate resistance against the pathogens *D. septosporum*, *V. dahliae*, *F. oxysporum* and *M. fijiensis*. *V. dahliae* causes vascular wilt in more than 200 dicotyledonous species (Fradin and Thomma, 2006). *Verticillium* wilt causes wilting in *Nicotiana* plants and during the process of flower-development the plants collapse (Mace et al., 1981). In tomato, *Verticillium* wilt causes wilting, the yellowing of leaves, stunting of the plant and reduced fruit size (Mace et al., 1981). The transcriptome of *V. dahlia*-infected *N. benthaminana* has been sequenced, and transcripts from many genes related to stress were identified in the infected plant (Faino et al., 2012). However, the only gene so far cloned for resistance to *V. dahlia* is *Ve1* from tomato (Fradin et al., 2009).

F. oxysporum also infects a wide variety of hosts. *F. oxysporum formae speciales* have been characterised according to their host range (Mace et al., 1981). For example, *F. oxysporum* f. sp. *lycopersici* is pathogenic on tomato species, whilst *F. oxysporum* f. sp.

nicotianae is pathogenic on *Nicotiana* species (Mace et al., 1981). *F. oxysporum* is devastating for both *Nicotiana* and tomato, causing stunting of the plant and yellowing of leaves (Mace et al., 1981). Some *Nicotiana* species have been identified that are resistant to *F. oxysporum* f. sp. *nicotianae* (Mace et al., 1981). The *R* gene *I2* conferring resistance to *F. oxysporum* f. sp. *lycopersici* has been cloned from tomato (Simons et al., 1998). However, an *R* gene has not been cloned against *M. fijiensis. M. fijiensis* causes Black Sigatoka disease on Banana. Banana is the fourth most economically important food crop (Churchill, 2011). The most popular cultivated banana, Cavendish, is sterile and a clone. As a result, the incorporation of *R* genes by traditional crossing to wild relatives is not an option. Here, genetic modification provides an approach, whereby *R* genes can be transferred into banana to generate resistance (Wulff et al., 2011). In addition, *R* gene analogues already present within the banana genome can be modified by genome editing to become *Cf-Ecp2*. Genome editing allows for precise modification of the genome without the negative connotations associated with genetic modification (Gaj et al., 2013; Kuzma, 2016).

Although some *R* genes show long durability, e.g. *Sr31*, many *R* gene-encoded resistances have a short period of efficacy against the pathogen (Ayliffe et al., 2008; McDonald and Linde, 2002b). This lack of durability is due to the ability of the pathogen to jettison effectors that are not vital to their pathogenicity, or to accumulate mutations that evade R protein recognition, whilst still maintaining effector function (Joosten et al., 1997; Van Kan et al., 1991). However, Ecp2 is considered to be a conserved core effector. It is present in many fungal species from the fungal class Dothidiomycetes and deletion of Ecp2 from the genome of *C. fulvum* reduces the pathogenicity of the fungus on tomato (Lauge et al., 1997; Stergiopoulos et al., 2012).

In addition, when there is only one *R* gene encoding for recognition of a pathogen in a host plant, mutation of the effector recognised by the *R* gene product within the pathogen enables the pathogen to overcome the resistance (McDonald and Linde, 2002a). However, the presence of more than *R* gene encoding for recognition of several effectors from the pathogen means that all recognised effectors must mutate or be lost at the same time in order for the pathogen to overcome the resistance. Thus the presence of more than one *R* gene encoding for recognition of a pathogen in a host species has the potential to increase the durability of resistance (McDonald and Linde,

2002a). It is more likely that one effector rather than two effectors will be mutated by chance. The feature, whereby several *R* genes are present in a host and the *R* genes are genetically-linked, is known as pyramiding *R* genes (McDonald and Linde, 2002a). The tomato *R* gene product Cf-4 recognises the *M. fijiensis* effector Avr4 (Stergiopoulos et al., 2010). Therefore, if *Cf-Ecp2* can be cloned from *N. paniculata* then it may be transformed alongside *Cf-4* into banana to maximise the potential for durable resistance against *M. fijiensis*.

The effectiveness of *R* gene-encoded resistance in the host also depends on the lifestyle of the pathogen. *Cf-Ecp2* and *Cf-4* encoded-recognition of their corresponding effectors results in a HR (Lauge et al., 1998; Thomas et al., 1997). *M. fijiensis* is a hemibiotroph. It begins its lifecycle on the host as a biotroph then switches to a necrotrophic lifestyle (Churchill, 2011). The necrotrophic lifestyle includes the killing of plant cells, on which the pathogen then feeds. The HR must therefore be triggered by the *R* gene-encoded recognition of the effector during the biotrophic phase of the lifecycle. If recognition of the effector and the triggering of HR occurred after the switch to necrosis, it will only aid the pathogen by killing more plant cells. Fortunately, extended periods of biotrophy have been observed in *M. fijiensis* when infecting banana plants (Churchill, 2011). This extended biotrophy persist for several weeks (Churchill, 2011).

Other components required for recognition of Ecp2 and Avr4, and the downstream signalling of this recognition must either be conserved in banana or transformed into banana along with the *R* genes for functional resistance. The NB-LRR encoded by the *R* gene *Rps5* confers recognition of the *P. syringae* effector AvrPphB (Ade et al., 2007). However, recognition of AvrPphB by RPS5 requires the protein kinase PBS1 (Ade et al., 2007). The *Agrobacterium*-mediated transient co-transformation of *N. benthamiana* with *Rps5* and *AvrPphB* results in no response (Ade et al., 2007). However, when *Rps5* and *AvrPphB* are co-transformed with *PBS1*, a HR occurs in the section inoculated on the *N. benthamiana* leaf (Ade et al., 2007). Similarly, if recognition of Ecp2 or Avr4 encoded by *Cf-Ecp2* and *Cf-4* (respectively) is indirect, these factors must be co-transformed with the *R* gene to enable effectiveness of the *R* gene-encoded resistance. It is believed that recognition of Avr4 by Cf-4 is direct, since Avr4 binds to the chitin in the fungal cell wall and thus protects it from plant chitinase-mediated degradation (Van den Burg et al., 2006). For this virulence function, it is believed that Avr4 does not require a plant target.

Furthermore, Agrobacterium-mediated transient co-transformation of *N. benthamiana* with *Cf-4* and Avr4 results in a HR (Thomas et al., 2000). However, the interaction between Cf-Ecp2 and Ecp2 is still to be determined as either direct or indirect.

Following the initial recognition of a pathogen effector, the downstream signalling and activation of defence responses are thought to be conserved across many plant species. *R* genes have been transferred from monocots to dicots and still remain functional in the recipient species (Mendes et al., 2010). The Xa21 gene in rice encodes for an RLK that confers resistance to *Xanthomonas oryzae* pv. *oryzae* (Song et al., 1995). Sweet oranges (*Citrus sinensis*) with a stable expression of the transgene *Xa21* confer higher levels of resistance to *X. axonopodis* pv. *citri* compared to non-transgenic plants (Mendes et al., 2010).

Although there are issues associated with the movement of *Cf-Ecp2* and *Cf-4* into banana, as outlined above, it is of great interest to clone *Cf-Ecp2* in order to generate stable transgenic banana plants that are resistant to *M. fijiensis*. With this in mind, the inheritance of *C. fulvum* Ecp2 recognition in *N. paniculata* TW99 was investigated further in the current study (Table 5.4).

The inheritance of *C. fulvum* Ecp2-recognition conferred by *N. paniculata* TW99 was investigated. The F_2 populations NP_00036 and NP_00038 were generated from crosses between *N. paniculata* accessions TW99 and TW102. These F_2 populations were segregating for recognition of Ecp2 in a 1:2:1 manner for no response (0): a weak HR manifested by patchy necrosis (+): a strong HR manifested by confluent necrosis (++) (Table 5.4 and Figure 2.2) (Harder, 2012). This would suggest that *Cf-Ecp2* is a semi-dominant gene and that the phenotype 0 characterises the absence of *Cf-Ecp2*, the phenotype + characterises the heterozygosity for the presence of the *Cf-Ecp2* gene, whilst the ++ phenotype characterises the homozygosity for the presence of *Cf-Ecp2*. If these hypotheses were true, all F_3 plants derived from F_2 with a 0 phenotype should also have a 0 phenotype upon infiltration with 35S:*Ecp2*. In accordance with this, the 0 phenotype in the F_2 plants bred true in the F_3 (Table 5.4). Only one out of 10 F_3 families, derived from F_2 families with a 0 response, showed one plant with a + response (Table 5.4). This + response may have been due to a stray seed.

At the same time, the hypothesis states that those F_3 plants derived from F_2 plants with a ++ phenotype should all also have a ++ phenotype. However, when 10 F_3 families, derived from F_2 individuals, showing a response with a ++ phenotype, were inoculated with 35S:*Ecp2* seven of these families were segregating for recognition of 35S:*Ecp2* (Table 5.4). Therefore, a ++ phenotype in the F_2 does not mean the plant is homozygous for the presence of *Cf-Ecp2*.

As a consequence, it is concluded that *N. paniculata* carrying *Cf-Ecp2* responds with either a + or ++ phenotype upon recognition of Ecp2. To map the position of the gene within the genome, both phenotypes were pooled and determined to represent the presence of *Cf-Ecp2*.

Following an unbiased approach, in an attempt to identify a genetic location of *Cf-Ecp2*, a genetic map was generated for *N. paniculata*. This was based on 92 individuals from the F_2 populations (NP_00036 and NP_00038), previously scored for recognition of Ecp2 (Table 5.4). COSII markers from the genetic maps of *N. acuminata* and *N. tomentosiformis* were converted into Sequenom markers (Wu et al., 2010). Based on genotyping, 81 markers could be ordered into 24 linkage groups spanning 620 cM of the genome (Figure 5.3). The linkage groups were supported by a heat map representing a recombination fraction (Figure 5.4). Following the assumption that the markers at the ends of the linkage groups could detect linkage in up to 10 cM beyond their position, linkage within the genetic map should be detected in approximately 87% of the *N. paniculata* genome. Due to the high conservation of marker order between *N. tomentosiformis*, *N. acuminata* and *N. paniculata* linkage groups, the linkage groups of *N. paniculata* could be ordered into the 12 *N. paniculata* chromosomes by inferring information from the *N. acuminate* and *N. tomentosiformis* genetic maps (Figure 5.3 and Appendix 4, Table 4.2).

Although Chi-Squared analysis detected linkage between the 0 phenotype and CfEcp2 genotype on chromosome 3, no significant linkage was detected between phenotype and genotype using QTL analysis (Table 5.5 and Figure 5.5). This is because Chi-Squared analysis was too sensitive and identified false positives. In order to detect linkage and map the *Cf-Ecp2* gene, it is considered that more markers would need to be tested within the population.

Other methods of *R* gene detection that could be utilised to clone *Cf-Ecp2* from *N. paniculata* TW99 include the use of Ethyl methanesulfonate (EMS) mutagenesis followed by RNA-bait selection of *R* gene homologs and sequencing of mutants. This approach is termed MutRenSeq (Steuernagel et al., 2016). Within this approach, EMS mutagenesis would be performed on seed of *N. paniculata* TW99. Mutant plants would be grown and tested for the ability to recognise Ecp2. Those individual mutants that had lost the ability to recognise Ecp2 would be selected for further analysis. Targeted enrichment of DNA from mutants as well as the wildtype *N. paniculata* TW99 is based on an RNA bait library. Such a bait library can be generated from known *R* gene homologs from *N. paniculata* and close relatives of *N. paniculata* i.e. tobacco, pepper, tomato and potato (Andolfo et al., 2013; Jupe et al., 2012). Since *R* genes are highly homologous and the RNA baits can bind to DNA targets with ≥80% identity, it is likely that *Cf-Ecp2* would be selected by the bait library. Bait library selected DNA would then be sequenced for each of the mutants and for the *N. paniculta* TW99 parent. By generating a *de novo* assembly of the *N. paniculata* TW99 reads and mapping the mutant reads to it, contigs can be identified where all mutant lines have a mutation in the same. This gene would be a candidate for *N. paniculata Cf-Ecp2*.

The MutRenSeq approach requires the assumption that *N. paniculata Cf-Ecp2* is an *R* gene, and homologous to other *R* genes so far identified in *Nicotiana* and *Solanum* species (Steuernagel et al., 2016). Ecp2 recognition may have evolved in *N. paniculata* as a mechanism of non-host resistance. In fact Schulze-Lefert and Panstruga (2011), proposed a model for non-host resistance inducible defence responses. This model states that the closer phylogenetically-related a plant is to the host of the pathogen, the higher is the contribution of *R*-gene product triggered immunity and the lower is the contribution of PRR triggered immunity (Schulze-Lefert and Panstruga, 2011). Since *N. paniculata* is a close relative of tomato (with the last common ancestor estimated to be 30 Million years ago) and recognition of Ecp2 is encoded by a single dominant gene, this model would predict that *N. paniculata* Cf-Ecp2 is probably an *R* gene (Schulze-Lefert and Panstruga, 2011; Wang et al., 2008). However, Ecp2 recognition in *Nicotiana* species does not appear to be the only factor determining the non-host status of *Nicotiana* to *C. fulvum*. because there is no difference in *C. fulvum* growth on *Nicotiana* accessions that can or cannot recognise Ecp2 (de Kock et al., 2004). When the fungus was observed to

penetrate the stomata it did not get beyond the sub-stomatal cavity before inhibition of growth (de Kock et al., 2004). *Nicotiana* may possess preformed defences, of which *C. fulvum* fails to penetrate, rendering it a non-host.

On the other hand, recognition of Ecp2 may have evolved in *Nicotiana* species in relation to Ecp2 from another fungal pathogen. Indeed, in the current study Ecp2 from the Nicotiana pathogens *V. dahlia* and *F. oxysporum* were recognised by *N. paniculata* TW99 (Table 5.3 and Figure 5.2). *Cf-Ecp2*, may therefore be an *R* gene that has evolved recognition of Ecp2 from *V. dahlia* and *F. oxysporum*.

Other effectors have also been shown to be recognised by more than one plant species. Recognition of the effector conferred by the *R* gene can evolve from a common ancestor or independently in both plant species. The *R* gene *Mla* has been conserved for over 12 million years across barley and wheat (Jordan et al., 2011). *Mla* confers resistance to *Blumeria graminis* f. sp. *hordei* in barley and *B. graminis* f. sp. *tritici* in wheat (Jordan et al., 2011).

If the *Cf-Ecp2* genes in *N. paniculata* and *S. pimpinellifolium* share a common ancestor, then it may be predicted that *N. paniculata* Cf-Ecp2 shares homologs with *2A*.

In Arabidopsis, the *P. syringae* effectors AvrB and AvrRpm1 are recognised by the protein encoded by the same *R* gene *Rpm1*, whilst in soybean, they are recognised by proteins encoded by two independent *R* genes, *Rpg1b* and *Rpg1r*, respectively (Ashfield et al., 2003; Ashfield et al., 1995; Ashfield et al., 2004; Ashfield et al., 2014; Bisgrove et al., 1994; Grant et al., 1995). The homologous *R* genes *Rpg1b* and *Rpg1r* sit within an *R* gene cluster on chromosome 13 of the soybean genome and share a close evolutionary relationship (Ashfield et al., 2003; Ashfield et al., 1995; Ashfield et al., 2014). However, they are not orthologous to RPM1 (Ashfield et al., 2004; Ashfield et al., 2014). These results suggest an independent evolution for the recognition of the effectors AvrB and AvrRpm1 in Arabidopsis and soybean (Ashfield et al., 2014). If Ecp2 recognition has evolved independently in *N. paniculata* and *S. pimpinellifolium*, it may be assumed that *N. paniculata Cf-Ecp2* is not homologous to *2A*. Indeed, the presence or absence of the *PR1a* SP in PVX:*Ecp2* clones delivered into *N. paniculata* TW99 does not affect the

response to the effector. The effector may therefore be recognised both intracellularly and extracellularly in *Nicotiana* spp.

Recognition of Ecp2 may therefore be conferred by another *R* gene structural class or a completely different mechanism. Some *R* gene products recognise intracellular effectors and reside within the plant cell (e.g. NB-LRRs and serine/threonine protein kinases), whilst others recognise extracellular effectors and sit on the surface of the cell (e.g. RLKs and RLPs).

Infiltration of the proteinaceous effector ToxA from the necrotrophic pathogens *Stagonospora nodorum* or *Pyrenophora tritici-repentis* into the apoplastic space of wheat leaves results in necrosis (Friesen et al., 2006; Tuori et al., 1995). Sensitivity to ToxA, and thus susceptibility to the necrotic pathogen, is conferred by the serine/threonine-protein kinase–NB-LRR (S/T PK–NB-LRR) gene *Tsn1* (Faris et al., 2010; Friesen et al., 2006; Lamari and Bernier, 1989). The *Tsn1* gene does not carry any predicted transmembrane domains (Faris et al., 2010). It is not known how the 13 KDa protein ToxA is recognised by the intracellular protein encoded by *Tsn1,* when infiltrated into the apoplast of wheat leaves (Faris et al., 2010; Tuori et al., 1995). However, yeast-two-hybrid experiments suggest no direct interaction between Tsn1 and ToxA (Faris et al., 2010). Interaction between ToxA and an unknown transmembrane protein may therefore enable transmission of this signal across the plant membrane for recognition by the protein encoded by *Tsn1*. A similar mechanism may occur within *N. paniculata* to confer recognition of Ecp2.

As a modification of the MutRenSeq approach, the entire genome of *N. paniculata* mutants could be sequenced and compared to that of the parent *N. paniculata* TW99. This would represent an unbiased approach and may be possible in *N. paniculata* due to the small size of the genome. The drawback with EMS mutagenesis-based methods is that resistance must be conferred by a single dominant gene (Steuernagel et al., 2016). In tomato, four 100% identical homologs of *2A*, believed to encode for recognition of Ecp2, lie side-by-side (Chapter 3 and Chapter 4). Unless EMS mutagenesis deleted this entire section of the genome, incorporating the four functional homologs, it would fail to identify *Cf-Ecp2* in tomato. If *N. paniculata Cf-Ecp2* existed in a similar manner in the *N. paniculata* genome, then this method of *R* gene cloning would not be appropriate for identification of *N. paniculata* Cf-Ecp2.

The recognition of *C. fulvum*, *D. septosporum*, *V. dahliae*, *F. oxysporum* and *M. fijiensis* Ecp2 homologs is conferred by *N. paniculata* TW99. Recognition of *C. fulvum* Ecp2 is conferred by a single dominant gene, *Cf-Ecp2*. It is proposed that *Cf-Ecp2* of *N. paniculata* also encodes for recognition of the Ecp2 homologs of these other plant pathogens. This is encouraging, since the cloning of *Cf-Ecp2* from *N. paniculata* and its transfer into crop species such as banana has the potential to generate host resistance to a wide variety of fungal pathogens. A genetic map of *N. paniculata* was created in the present study. However, the genetic position for *Cf-Ecp2* was not identified. Other methods of cloning *Cf-Ecp2* from *N. paniculata* rely on it being a homolog of known *Solanaceae R* genes. There is an abundance of evidence for *N. paniculata Cf-Ecp2* to be an *R* gene, yet there continues to be some uncertainty as its structural class. Cloning of *Cf-Ecp2* from *N. paniculata* may therefore require an unbiased approach.

Chapter 6

General discussion: The pathway to impact

6.1 What have we learnt for the future cloning of *R* genes?

To clone a plant *R* gene, the gene is typically fine-mapped within the genome. Dense genetic maps are available for many plant species. However, *R* genes are often identified in wild relatives where no genetic map is available. In the current research, the tomato *Cf-Ecp2* locus, which contains a component required for the recognition of the effector Ecp2, was genetically delimited to an 0.18 cM interval spanning 91 kb (Chapter 3).

Due to the fast evolution of *R* genes, there is a high level of variation in gene copy number and sequence at *R* gene loci at the species and subspecies level (Michelmore and Meyers, 1998; Noel et al., 1999; Parniske et al., 1997). As a consequence, to determine the sequence of an *R* gene locus, the resistant parent must also be sequenced. To do this in a cost- and labour-effective manner, a process called genome complexity reduction can be employed. During this process, a library of clones of the parental resistant genome is generated and a minimal tiling path is selected across the locus containing the *R* gene.

The sequencing of three BACs (11G, 7B and 4B), covering the tomato *Cf-Ecp2* locus, highlighted the complexity of *R* gene loci (Chapter 3). Eight *Hcr9s* named *2C*, *2B*, *2A1*, *2A2*, *2A3*, *2A4*, *Ψ2A5* and *Ψ2C* were located at the *Cf-Ecp2* locus (Figure 3.8). The sequences of genes *2A1*, *2A2*, *2A3* and *2A4* were 100% identical throughout the entire ORF and shared regions of homology with the other genes at this locus, including across the intergenic regions (Figure 3.10).

Short read technologies are known to generate highly accurate sequences. However, due to the length of the reads (at 75 – 500 bp), they readily collapse upon assembly of repetitive elements (Reuter et al., 2015; Taudien et al., 2011). Indeed, due to the high level of sequence homology at the *Cf-Ecp2* locus, homologous regions collapsed when assembling the output from sequencing with short read technologies (Figure 3.6 and

Figure 3.7). As a result, the sequence of the short-read assembled contig was found not to truly represent the sequence of the locus (Figure 3.6).

It has previously been demonstrated that long read sequencing technologies such as PacBio and Oxford Nanopore generate long raw reads (Ashton et al., 2015; Chaisson et al., 2015; Jain et al., 2015; Van Buren et al., 2015). However, these reads carry high sequencing errors (approximately 11% for PacBio and 29% for MinION 2D reads) (Ashton et al., 2015; Reuter et al., 2015). Even with such high error rates, biologically relevant data can be extracted from such reads. For example, 1D and 2D MinION reads were used in the current study to scaffold the *Cf-Ecp2* locus and to accurately determine the copy number of the 8.7 kb repeat element containing the *2A* homologs (Figure 3.6). However, due to the high error rates they could not be used to determine the sequence of the locus.

Following the successful determination of the copy number of the 8.7 kb repeat using the MinION reads, the PacBio assembly of BAC 7B was manually curated using the Illumina reads and the >15 kb PacBio reads (Figure 3.8). Sequencing the clones with many technologies was therefore required to determine the sequence of this complex *R* gene locus.

It has previously been shown that artefacts can be incorporated into the sequence output when utilising PacBio sequencing assemblies (Schatz, 2015). It is understood that such artefacts were generated due to the way the assembly algorithms manage the sequencing data. This can lead to unreliable assemblies. Without an understanding of the biology behind the sequence and the use of multiple sequencing techniques, conclusions may be based on unreliable evidence (Schatz, 2015). To validate the curated sequence in the current study, this locus was tested using many cloning methods to demonstrate that no other variant of *2A* existed on BAC 7B (Chapter 3).

After the sequence of a locus is determined, candidate genes can be tested generating stable transformants in a close relative known to lack the *R* gene (Ernst et al., 2002; Song et al., 1995; Thomas et al., 1997). These transgenic plants can then be characterised for their ability to recognise the corresponding effector(s) of the *R* gene and/or for their ability to confer disease resistance. This analysis has enabled others to identify multiple *R*

genes (Ernst et al., 2002; Song et al., 1995; Thomas et al., 1997). In time, it is expected that more whole genomes will be resolved using these long read technologies as in the recent case of the 245 Mb diploid *Oropetium thomaeum* (Van Buren et al., 2015). The main barriers to the widespread adoption of the technology is the cost per bp and the relatively high error rate of PacBio (11%) compared to Illumina (<1%) (Reuter et al., 2015).

In addition to genome complexity reduction followed by sequencing, transposon-tagging experiments can be employed to clone *R* genes (Collins et al., 1999; Jones et al., 1994; Takken et al., 1998; Takken et al., 1999). However, in the case of DNA transposons such as *Ds* or *Mutator*, this requires that the transposon element is genetically linked to the gene of interest (Collins et al., 1999; Jones et al., 1994; Takken et al., 1998; Takken et al., 1999). Furthermore, the current study showed that, in order to clone the gene of interest by tagging it with a transposon element, only one copy of the gene must exist at the locus (Chapter 3). If more than one functional copy of the gene does exist at the locus, only mutants that carry deletions generated by the transposon jumping will be identified (Chapter 3).

Such deletion mutants may remove an entire section of the locus including more than one gene. This means that further research must be undertaken to identify the candidate gene. At the *Cf-Ecp2* locus, two deletion mutants (*S. pimpinellifolium* 1178 and *S. pimpinellifolium* 1179) were generated using a transposon tagging experiment (Chapter 3). These two mutants had lost the ability to recognise the effector Ecp2. Molecular characterisation of these mutants indicated that they retained *2C* and *2B* yet had lost *2A* from the genome (Chapter 3). Furthermore, expression of *2B*, the gene closest to the deletion, was retained. As a consequence of these findings, *2A* became the candidate for *Cf-Ecp2*.

The future of *R* gene cloning now looks to the application of methods generating loss-of-function mutants combined with complexity reduced sequencing. This was exemplarily demonstrated recently with MutRenSeq, combining EMS mutagenesis with *R* gene homolog enrichment sequencing (Steuernagel et al., 2016). Sequencing data is generated and through mutational genomics, a candidate gene can be identified by its sequence having a mutation in each susceptible mutant. However, MutRenSeq does rely on the assumption that resistance is conferred by one *R* gene, and that this *R* gene has sequence

homology to known *R* gene analogues whose sequences were represented in the bait library.

It is expected that research will drive the development of new enabling technologies that will aid not only in the cloning of *R* genes but also other agriculturally important genes. This is required to enhance the process of *R* gene cloning in the fight against ever evolving pathogens and to gain food security.

6.2 The future of food security

The bulk of human food comes from a few main crops including wheat, rice and potato (FAOSTAT, 2015). Plants are often grown as monocultures with farmers choosing the best crops identified for their high yield output. Unfortunately, monocultures can be easily wiped out by disease. As a result, plant disease causes devastation to not only plants but also to humans (Koeppel, 2008; Yoshida et al., 2013). In the 1840s, the Irish potato famine claimed the lives of 1 million people (Yoshida et al., 2013). This famine was caused by the Oomycete fungus *Phytophthora infestans*, which depleted the potato crops in Ireland (Yoshida et al., 2013). This disease is still a significant problem today due to its ability to quickly overcome resistances that are deployed in the crop (Fry, 2008).

Similarly, the banana cultivar Gros Michel was wiped out by Panama disease caused by the fungal pathogen *F. oxysporum* f. sp. *cubense* (Koeppel, 2008). This banana cultivar was quickly replaced by another cultivar, called Cavendish. Cavendish was resistant to the strain of fungus causing Panama disease (Koeppel, 2008). However, Cavendish is now under threat from Panama disease since a new strain of the fungus has emerged, which is able to cause disease symptoms in Cavendish (Ploetz, 2006).

As is the case for many other crop plants, bananas are susceptible to different pathogens and therefore the diseases they cause. *M. fijiensis* is the causal agent of Black Sigatoka disease on bananas. This fungus started to appear from 1935 onwards (Koeppel, 2008). In contrast to Panama disease, which is spread by contaminated soil, water and humans, *M. fijiensis* spreads via airborne spores (Koeppel, 2008) and thus even faster.

The control of Black Sigatoka disease was attempted by spraying plantations with Bordeaux mixture (containing copper sulphate, lime and oil) 20 to 30 times per year with 250 gallons per acre (Koeppel, 2008). However, this method of control was not economically available to all banana farmers and caused severe health issues for plantation workers (Koeppel, 2008).

To control these devastating diseases, it is widely accepted that genetic resistance needs to be introduced into bananas. The use of conventional breeding methods to move *R* genes from wild relatives into banana is not feasible (Koeppel, 2008). The length of time it takes to selectively-breed bananas that have resistance to Black Sigatoka and/or Panama Disease, plus other agronomically important traits, is not feasible in relation to the speed at which these diseases are spreading (Koeppel, 2008). Traditional cross-breeding is constrained further by the fact that Cavendish banana is sterile and does not produce seeds. The absence of seeds is a desirable trait for human consumption but makes it impractical for breeding.

A biotechnological approach is therefore required to fight these diseases. The introduction of a cloned resistance gene into the Cavendish banana genome via genome editing or genetic modification is desirable.

In the current study it was confirmed that *Cf-Ecp2* of the tomato *S. pimpinellifolium* CfEcp2 and *N. paniculata* TW99 had the ability to recognise the effector MfEcp2 from the fungal pathogen *M. fijiensis* (Chapter 3 and 5). In the case of tomato, this had been observed previously alongside the feature that tomato lines carrying *Cf-4* recognise the *M. fijiensis* effector MfAvr4 (Stergiopoulos et al., 2010). This outcome is promising since the transfer of a construct carrying two or more *R* genes is preferable over one *R* gene (McDonald and Linde, 2002a).

The deployment of multiple *R* genes in one species is known as *R* gene pyramiding (Joshi and Nayak, 2010). The presence of >1 *R* gene with distinct specificities would mean that mutations would have to occur in >1 effector of the fungus for the fungus to overcome the resistance presented by the pyramid. In crops that reproduce sexually, the *R* genes should be genetically linked to each other to be inherited as a single unit. As a result, the

resistance pyramid would not inadvertently be separated during breeding or accidental outcrossing.

Transfer of *R* genes from tomato into banana is based on the principle that *R* genes function the same way in both dicots and monocots, which diverged 140 to 150 million years ago (Chaw et al., 2004). The NB-LRR pair *RPS4* and *RRS1* from *A. thaliana* recognises effectors from three different pathogens; *P. syringae* pv. tomato DC3000 (specifically the effector AvrRps4), *Ralstonia solanacearum* (specifically the effector PopP2) and *Colletotrichum higginsianum* (Narusaka et al., 2009a; Narusaka et al., 2009b). *N. benthamiana*, expressing *RPS4* and *RRS1* in a stable manner, induced a HR following *Agrobactrium*-mediated transient expression of AvrRps4 or PopP2 (Narusaka et al., 2013). In addition, *S. lycopersicum*, expressing *RPS4* and *RRS1* in a stable manner, were resistant to *P. syringae* pv. tomato DC3000 and *R. solanacearum* (Narusaka et al., 2013). Similarly, cucumber plants expressing *RPS4* and *RRS1* in a stable manner were resistant to *C. orbiculare*, usually infecting cucumber (Narusaka et al., 2013). The NB-LRR pair was therefore able to function in four relatively unrelated species of plants from three different orders.

Some *R* genes of monocot origin are functional in dicots. For example, resistance to the powdery mildew fungus *Blumeria graminis* f. sp. *hordei* in barely is conferred by the *Mla* locus, carrying multiple *NB-LRR* genes including the *Mla1* gene (Zhou et al., 2001). Immunocompromised *A. thaliana* (*pen2*, *pad4*, *sa101*) plants, stably expressing *Mla*, confer resistance to *B. graminis* f.sp *hordei* in comparison to those immunocompromised plants not transgenic for the *Mla* gene (Maekawa et al., 2012).

Such examples of functional *R* gene transfer between distantly related plant species suggest the conservation of a downstream signalling cascade leading to defence. However, in some cases other signalling components that are required by the *R* gene to function need to be transferred along with the *R* gene for recognition of the effector to be triggered in the transgenic plant (Ade et al., 2007). This other component is usually the effector target and therefore presumably lies upstream of the *R* gene product in the resistance signalling cascade. For example, the *P. syringae* effector AvrPphB targets and cleaves the *A. thaliana* protein kinase PBS1 (Shao et al., 2003). The *A. thaliana* NB-LRR *Rps5* confers recognition of AvrPphB, when transiently expressed in *N. benthamiana* but

it must be co-delivered with *PBS1* (Ade et al., 2007). By extrapolation, if the gene *Cf-Ecp2* confers recognition of Ecp2 via an effector target then the gene coding for this specific target will likely also have to be identified and transformed alongside *Cf-Ecp2* into banana for the recognition to be functional.

Unfortunately, the GM of plants evokes negative connotations for many (Prakash, 2001). These are associated with the fear of the introduction of foreign DNA into crops and what this will mean for human and plant health (Prakash, 2001). However, it is understood that such anxieties have not been fuelled by scientific evidence but by the media and anti-GM campaign groups (Prakash, 2001). Genome editing, which enables modification of the plant genome with the introduction of minimal foreign DNA, holds great promise. This method will hopefully not attract the level of unfounded disapproval associated with GM (Gaj et al., 2013; Kuzma, 2016).

One potential use of genome editing is for the manipulation of plant genomes to enable the recognition of new pathogen effectors. It has been proposed that the LRR domain encoded for by *R* genes confers recognition of the effector (Kobe and Deisenhofer, 1994). Thus, to transfer the functional recognition of an effector from one species to another may only require the discrete editing of an LRR domain in an *R* gene analogue that pre-exists in the recipient species. To start this process, an *R* gene could be cloned in one plant species (species A) and the resistance conferred by this *R* gene (by recognition of a specific effector) would be transferred to another plant species (species B). The genome of species B can be searched for an *R* gene analogue with close sequence identity. The LRR domain of the gene in species B may be modified by genome editing to become that of the target *R* gene. The modified gene in species B would then code for a protein that contained an LRR domain, conferring recognition of the effector. This *R* gene product can plug into the pre-established plant resistance-signalling pathway in species B. Repeating this process for more than one *R* gene would represent a way of *R* gene pyramiding.

R gene expression requires fine-tuning. This was observed in the current research whereby the use of a promoter (35S), causing over expression of the *R* gene, was believed to have impeded the function of the *R* gene (Chapter 4). That said, cases without these difficulties are known. For example, transgenic *N. tabacum* plants,

expressing the tomato *R* gene *Cf-4* in a stable manner and driven by the 35S promoter, were able to recognise the corresponding effector Avr4 (Thomas et al., 2000).

Native promoters may be required for expression of *R* genes in target plants to acquire the right level of *R* gene product. However, these promoters may not work in more distantly-related plant species (Potrykus et al., 1998). The use of genome editing instead of GM is likely to be beneficial, since genome editing would rely on the use of the promoter from the recipient species. In contrast, GM would require the transformation of an *R* gene with either the native *R* gene promoter or another promoter.

The lifecycle of the target pathogen can determine the functionality of an *R* gene. The ability of *S. pimpinellifolium* CfEcp2 and *N. paniculata* TW99 to recognise MfEcp2 does not necessarily imply that *Cf-Ecp2* can confer resistance to *M. fijiensis*. Recognition of MfEcp2 by Cf-Ecp2 triggers a HR. *M. fijiensis* is a hemibiotrophic pathogen (Churchill, 2011). This means that the fungus begins its infection life cycle as a biotroph on the plant and then switches to a necrotrophic lifestyle. As a biotroph the HR and other responses triggered by the recognition of MfEcp2 by Cf-Ecp2 may confer resistance to the pathogen. However, when the pathogen switches to necrotrophy, a HR will only act to enhance the susceptibility of the plant to the pathogen. Therefore, it depends on when the *Cf-Ecp2* gene is active during the lifecycle of the pathogen. In addition, plant tissues displaying *R* gene expression must also be colonised by the pathogen at the same time in order to establish effective resistance.

N. paniculata TW99 is able to recognise an Ecp2 homolog from *V. dahliae* (Figure 5.2). *V. dahliae* infects a wide range of plant species including tomato, potato and cotton (Mace et al., 1981). *V. dahlia* is a vascular pathogen, infecting the plant via the root and colonising the root cortex and endodermis before entering the xylem (Mace et al., 1981). This is a very different infection system in comparison to that of *C. fulvum*, which penetrates host tissues via the stomata on the abaxial side of the leaf (Joosten and de Wit, 1999). *C. fulvum* hyphae grow between cells in the spongy parenchyma of the leaf (Joosten and de Wit, 1999). It is unknown if *Cf-Ecp2* is expressed in the root and vascular system and whether it is able to confer resistance to *V. dahlia*.

The deployment of *R* genes in crops has not always led to durable resistance. For example, *Cf-4* was crossed into cultivated tomato plants to provide resistance against *C. fulvum* (Kerr and Bailey, 1964; Stevens and Rick, 1988). However, *C. fulvum* circumvented the resistance encoded by this *R* gene via the selection for mutation of the recognised effectors (Joosten et al., 1997). Durable resistance was therefore not achieved under such circumstances. As a result, it may be questioned whether it is right to deploy *R* genes to confer resistance since they may lack durability.

Fortunately, some *R* genes may be more durable than others. Certain effectors could be more important than other effectors to the pathogen (Lauge et al., 1997). Loss of such effectors may reduce the virulence of the pathogen (Lauge et al., 1997). Therefore, an *R* gene coding for the recognition of such an effector may have prolonged effectiveness against the pathogen.

It is believed that Ecp2 is a core effector. It is present in many pathogen species, independent of their lifestyle and is undergoing diversifying selection in *M. fijiensis* (Stergiopoulos et al., 2014; Stergiopoulos et al., 2012). However, if the domain of the effector that is recognised by the *R* gene product is spatially separated from the domain which codes for virulence, loss of recognition by mutation of this effector does not necessarily mean loss of pathogen virulence.

Not only *R* genes can be deployed in crop species to generate resistance against certain pathogens. PRRs confer recognition of conserved pathogen molecules (PAMPS) and trigger a resistance response in the plant (Zipfel, 2009; Zipfel and Robatzek, 2010). The *A. thaliana* PRR EFR confers recognition of the peptide efl18 from the bacterial protein EF-TU (Kunze et al., 2004; Zipfel et al., 2006). Transgenic expression of *EFR* in wheat conferred recognition of efl18 and increased resistance to *P. syringae* p.v. *oryzae* (Schoonbeek et al., 2015).

The current study contributes to the knowledge required for further *R* gene cloning and the deployment of *R* genes in cultivated plant species. *R* gene loci are complex and challenge current technology to correctly reconstruct their sequence. However, decoding such loci is important for within them lies the *R* genes that may provide important disease resistance for future farmer 'tool kits'. The ability to transfer *R* genes, encoding

for resistance to important crop pathogens, across plant species provides exciting opportunities for resistance breeding in the future. It will require a combined approach of *R* gene pyramids, genes encoding PRRs and other such resistance encoding genes to enable future global food security. GM and genome editing platforms are promising technologies to enable their efficient deployment in crop plants. However, pathogen lifestyle and effector populations must be taken into consideration in order to enhance the durability and effectiveness of *R* genes against pathogens.

Bibliography

Ade, J., de Young, B.J., Golstein, C., and Innes, R.W. (2007). Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. Proceedings of the National Academy of Science U S A *104*, 2531-2536.

Al-Shahrour, F., Minguez, P., Marques-Bonet, T., Gazave, E., Navarro, A., and Dopazo, J. (2010). Selection upon genome architecture: conservation of functional neighborhoods with changing genes. PLoS Computational Biology *6*, e1000953.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997). Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell *9*, 641-651.

Andolfo, G., Sanseverino, W., Rombauts, S., Van de Peer, Y., Bradeen, J.M., Carputo, D., Frusciante, L., and Ercolano, M.R. (2013). Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important *Solanum R* locus dynamics. New Phytologist *197*, 223-237.

Armstrong, M.R., Whisson, S.C., Pritchard, L., Bos, J.I., Venter, E., Avrova, A.O., Rehmany, A.P., Bohme, U., Brooks, K., Cherevach, I.*, et al.* (2005). An ancestral oomycete locus contains late blight Avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proceedings of the National Academy of Science U S A *102*, 7766-7771.

Arneson, P., and Durbin, R.D. (1968). Studies on the mode of action of tomatine as a fungitoxic agent. Plant Physiology *43*, 683-686.

Ashfield, T., Bocian, A., Held, D., Henk, A.D., Marek, L.F., Danesh, D., Penuela, S., Meksem, K., Lightfoot, D.A., Young, N.D.*, et al.* (2003). Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of *NBS-LRR* genes. Molecular Plant Microbe Interactions *16*, 817-826.

Ashfield, T., Keen, N.T., Buzzell, R.I., and Innes, R.W. (1995). Soybean Resistance genes specific for different *Pseudomonas syringae* Avirulence genes are allelic, or closely linked, at the *RPG1* locus. Genetics *141*, 1597-1604.

Ashfield, T., Ong, L.E., Nobuta, K., Schneider, C.M., and Innes, R.W. (2004). Convergent evolution of disease resistance gene specificity in two flowering plant families. Plant Cell *16*, 309-318.

Ashfield, T., Redditt, T., Russell, A., Kessens, R., Rodibaugh, N., Galloway, L., Kang, Q., Podicheti, R., and Innes, R.W. (2014). Evolutionary relationship of disease resistance genes in soybean and *Arabidopsis* specific for the *Pseudomonas syringae* effectors AvrB and AvrRpm1. Plant Physiology *166*, 235-251.

Ashton, P.M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J., and O'Grady, J. (2015). MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nature Biotechnol *33*, 296-300.

Ayliffe, M., Singh, R., and Lagudah, E. (2008). Durable resistance to wheat stem rust needed. Current Opinion in Plant Biology *11*, 187-192.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods *9*, 39.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R*., et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53-59.

Bhat, R.G., and Subbarao, K.V. (1999). Host range specificity in *Verticillium dahliae*. Phytopathology *89*, 1218-1225.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and Innes, R.W. (1994). A disease resistance gene in *Arabidopsis* with specificity for two different pathogen Avirulence genes. Plant Cell *6*, 927-933.

Bogdanove, A.J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. Current Opinion in Plant Biology *13*, 394-401.

Boix, V., and Bohme, S.R. (2012). Secrecy and justice in the ongoing saga of DBCP litigation. Internation Journal of Occupational and Environmental Health *18*, 154-161.

Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annual Review of Plant Biology *60*, 379-406.

Bolton, M.D., Van Esse, H.P., Vossen, J.H., de Jonge, R., Stergiopoulos, I., Stulemeijer, I.J.E., van den Berg, G.C.M., Borras-Hidalgo, O., Dekker, H.L., de Koster, C.G.*, 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.

Bond, T., E, T., (1938). Infection experiments with *Cladosporium fulvum* Cooke and related species. Annual of Applied Biology *25*, 277-307.

Borhan, M.H., Gunn, N., Cooper, A., Gulden, S., Tor, M., Rimmer, S.R., and Holub, E.B. (2008). *WRR4* encodes a TIR-NB-LRR protein that confers broad-spectrum white rust resistance in *Arabidopsis thaliana* to four physiological races of *Albugo candida*. Molecular Plant Microbe Interactions *21*, 757-768.

Bos, J.I., Kanneganti, T.D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R., Birch, P.R., and Kamoun, S. (2006). The C-terminal half of *Phytophthora infestans* RXLR effector Avr3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. Plant Journal *48*, 165-176.

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D. (1998). Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* Avirulence determinants. Plant Cell *10*, 1847-1860.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889-890.

Browns, T., A., (2002). Genomes, 2nd edn (Oxford: Wiley-Liss).

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proceedings of the National Academy of Science U S A *107*, 9452-9457.

Cardenas, M., Grajales, A., Sierra, R., Rojas, A., Gonzalez-Almario, A., Vargas, A., Marin, M., Fermin, G., Lagos, L.E., Grunwald, N.J.*, et al.* (2011). Genetic diversity of *Phytophthora infestans* in the Northern Andean region. BMC Genetics *12*, 23.

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014). A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. Frontiers in Plant Science *5*, 606.

Chaisson, M.J., Huddleston, J., Dennis, M.Y., Sudmant, P.H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M.*, et al.* (2015). Resolving the complexity of the human genome using single-molecule sequencing. Nature *517*, 608-611.

Chapman, S., Kavanagh, T., and Baulcombe, D. (1992). Potato Virus X as a vector for gene expression in plants. Plant Journal *2*, 549-557.

Chapman, S., Stevens, L.J., Boevink, P.C., Engelhardt, S., Alexander, C.J., Harrower, B., Champouret, N., McGeachy, K., Van Weymers, P.S., Chen, X.*, et al.* (2014). Detection of

the virulent form of Avr3a from P*hytophthora infestans* following artificial evolution of potato Resistance gene *R3a*. PLoS One *9*, e110158.

Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. Journal of Molecular Evolution *58*, 424-441.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E*., et al.* (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods *10*, 563-569.

Chin, F.Y., Leung, H.C., and Yiu, S.M. (2014). Sequence assembly using next generation sequencing data-challenges and solutions. Science China Life Sciences *57*, 1140-1148.

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The *Arabidopsis* receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell *18*, 465-476.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nurnberger, T., Jones, J.D., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature *448*, 497-500.

Chung, E.H., da Cunha, L., Wu, A.J., Gao, Z., Cherkis, K., Afzal, A.J., Mackey, D., and Dangl, J.L. (2011). Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants. Cell Host Microbe *9*, 125-136.

Churchill, A.C.L. (2011). *Mycosphaerella fijiensis,* the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. Molecular Plant Pathology *12*, 307-328.

Collins, N., Drake, J., Ayliffe, M., Sun, Q., Ellis, J., Hulbert, S., and Pryor, T. (1999). Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell *11*, 1365-1376.

Dangl, J.L., and Jones, J.D. (2001). Plant pathogens and integrated defence responses to infection. Nature *411*, 826-833.

Dangl, J.L., and McDowell, J.M. (2006). Two modes of pathogen recognition by plants. Proceedings of the National Academy of Science U S A *103*, 8575-8576.

de Jonge, R., Van Esse, H.P., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., Van der Krol, S., Shibuya, N., Joosten, M.H.A.J., and Thomma, B.P.H.J. (2010). Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science *329*, 953-955.

de Kock, M.J.D., Brandwagt, B.F., Bonnema, G., de Wit, P.J.G.M., and Lindhout, P. (2005). The tomato *Orion* locus comprises a unique class of *Hcr9* genes. Molecular Breeding *15*, 409-422.

de Kock, M.J.D., and colleagues (2004). Recognition of the *Cladosporium fulvum* Ecp2 elicitor in tomato and non-host plants. Thesis Wageningen University, The Netherlands

de Kock, M.J.D., Iskander, H.M., Brandwagt, B.F., Lauge, R., de Wit, P.J.G.M., and Lindhout, P. (2004). Recognition of *Cladosporium fulvum* Ecp2 elicitor by non-host *Nicotiana* spp. is mediated by a single dominant gene that is not homologous to known *Cf*-genes. Molecular Plant Pathology *5*, 397-408.

de Wit, P.J., Joosten, M.H., Honee, G., Wubben, J.P., Van den Ackerveken, G.F., and Van den Broek, H.W. (1994). Molecular communication between host plant and the fungal tomato pathogen *Cladosporium fulvum*. Antonie Van Leeuwenhoek *65*, 257-262.

de Wit, P.J., Van der Burgt, A., Okmen, B., Stergiopoulos, I., Abd-Elsalam, K.A., Aerts, A.L., Bahkali, A.H., Beenen, H.G., Chettri, P., Cox, M.P.*, 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.

de wit, P.J.G.M., and Spikman, G. (1982). Evidence for the occurrence of race and cultivar-specific elicitors of necrosis in inter-cellular fluids of compatible interactions of *Cladosporium fulvum* and tomato. Physiological Plant Pathology *21*, 1-&.

Defago, G., and Kern, H. (1983). Induction fo *Fusarium solani* mutants insensitive to tomatine, their pathogenicity and aggressiveness to tomato fruits and pea plants. Physiological plant pathology *22*, 29 - 37.

Dixon, M.S., Golstein, C., Thomas, C.M., Van der Biezen, E.A., and Jones, J.D.G. (2000). Genetic complexity of pathogen perception by plants: the example of *Rcr3*, a tomato gene required specifically by *Cf-2*. Proceedings of the National Academy of Science U S A *97*, 8807-+.

Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K., and Jones, J.D.G. (1998). The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell *10*, 1915-1925.

Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K., and Jones, J.D.G. (1996). The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell *84*, 451-459.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I., Ayliffe, M.A., Kobe, B., and Ellis, J.G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax Resistance genes and flax rust Avirulence genes. Proceedings of the National Academy of Science U S A *103*, 8888-8893.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell *127*, 1309-1321.

Ernst, K., Kumar, A., Kriseleit, D., Kloos, D.U., Phillips, M.S., and Ganal, M.W. (2002). The broad-spectrum potato cyst nematode Resistance gene (*Hero*) from tomato is the only member of a large gene family of *NBS-LRR* genes with an unusual amino acid repeat in the LRR region. Plant Journal *31*, 127-136.

Exposito-Rodriguez, M., Borges, A.A., Borges-Perez, A., and Perez, J.A. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biology *8*, 131.

Faino, L., de Jonge, R., and Thomma, B.P. (2012). The transcriptome of *Verticillium dahliae*-infected *Nicotiana benthamiana* determined by deep RNA sequencing. Plant Signaling and Behaviour *7*, 1065-1069.

FAOSTAT (2015). http://faostat3.fao.org/home/index.html - VISUALIZE.

Faris, J.D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., Fellers, J.P., Meinhardt, S.W., Rasmussen, J.B., Xu, S.S*., et al.* (2010). A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proceedings of the National Academy of Science U S A *107*, 13544-13549.

Fiume, M., Smith, E.J., Brook, A., Strbenac, D., Turner, B., Mezlini, A.M., Robinson, M.D., Wodak, S.J., and Brudno, M. (2012). Savant Genome Browser 2: visualization and analysis for population-scale genomics. Nucleic Acids Research *40*, W615-621.

Flor, H.H. (1971). Current status of Gene-for-Gene concept. Annual Review of Phytopathology *9*, 275-+.

Foresight (2011). The future of food and farming. In Final Project report (The Govenment Office for Science London).

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

Fradin, E.F., Zhang, Z., Juarez Ayala, J.C., Castroverde, C.D., Nazar, R.N., Robb, J., Liu, C.M., and Thomma, B.P. (2009). Genetic dissection of *Verticillium* wilt resistance mediated by tomato *Ve1*. Plant Physiology *150*, 320-332.

Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A., and Oliver, R.P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics *38*, 953-956.

Fry, W. (2008). *Phytophthora infestans*: the plant (and *R* gene) destroyer. Molecular Plant Pathology *9*, 385-402.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K.*, et al.* (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. Science *325*, 998-1001.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology *31*, 397-405.

Gimenez-Ibanez, S., Hann, D.R., Ntoukakis, V., Petutschnig, E., Lipka, V., and Rathjen, J.P. (2009). AvrPtoB targets the LysM receptor kinase CERK1 to promote bacterial virulence on plants. Current Biology *19*, 423-429.

Goggin, F.L., Shah, G., Williamson, V.M., and Ullman, D.E. (2004). Developmental regulation of Mi-mediated aphid resistance is independent of *Mi-1.2* transcript levels. Molecular Plant Microbe Interactions *17*, 532-536.

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. Molecular Cell *5*, 1003-1011.

Goswami, R.S., and Kistler, H.C. (2004). Heading for disaster: *Fusarium graminearum* on cereal crops. Molecular Plant Pathology *5*, 515-525.

Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. (1995). Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science *269*, 843-846.

Greenberg, J.T., and Yao, N. (2004). The role and regulation of programmed cell death in plant-pathogen interactions. Cell Microbiology *6*, 201-211.

Haanstra, J.P.W., Lauge, R., Meijer-Dekens, F., Bonnema, G., de Wit, P.J.G.M., and Lindhout, P. (1999). The *Cf-Ecp2* gene is linked to, but not part of, the *Cf-4*/*Cf-9* cluster on the short arm of chromosome 1 in tomato. Molecular and General Genetics *262*, 839-845.

Haanstra, J.P.W., Meijer-Dekens, F., Lauge, R., Seetanah, D.C., Joosten, M.H.A.J., de Wit, P.J.G.M., and Lindhout, P. (2000). Mapping strategy for resistance genes against Cladosporium fulvum on the short arm of Chromosome 1 of tomato: *Cf-Ecp5* near the *Hcr9 Milky Way* cluster. Theoretical and Applied Genetics *101*, 661-668.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica *156*, 1-13.

Hammond-Kosack, K.E., Harrison, K., and Jones, J.D.G. (1994a). Developmentally-regulated cell-death on expression of the fungal Avirulence gene Avr9 in tomato seedlings carrying the disease-resistance gene *Cf-9*. Proceedings of the National Academy of Science U S A *91*, 10445-10449.

Hammond-Kosack, K.E., Jones, D.A., and Jones, J. (1994b). Identification of two genes required in tomato for full *Cf-9*-dependent resistance to *Cladosporium fulvum*. Plant Cell *6*, 361-374.

Hammond-Kosack, K.E., Tang, S., Harrison, K., and Jones, J.D. (1998). The tomato *Cf-9* disease resistance gene functions in tobacco and potato to confer responsiveness to the fungal Avirulence gene product Avr9. Plant Cell *10*, 1251-1266.

Harder, S. (2012). Mapping the Resistance gene of the *Cladosporium fulvum* effector Ecp2 in *Nicotiana paniculata*. MSc Plant Genetics and Crop Improvement

Heath, M.C. (2000). Nonhost resistance and nonspecific plant defenses. Current Opinion in Plant Biology *3*, 315-319.

Holton, N., Nekrasov, V., Ronald, P.C., and Zipfel, C. (2015). The phylogenetically-related pattern recognition receptors EFR and XA21 recruit similar immune signaling components in monocots and dicots. PLoS Pathogens *11*, e1004602.

Howard, R.J., and Valent, B. (1996). Breaking and entering: host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. Annual Review of Microbiology *50*, 491-512.

Huang, S., Vleeshouwers, V.G., Werij, J.S., Hutten, R.C., Van Eck, H.J., Visser, R.G., and Jacobsen, E. (2004). The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked *R* genes with distinct specificities. Molecular Plant Microbe Interactions *17*, 428-435.

Huang, S.W., Van der Vossen, E.A.G., Kuang, H.H., Vleeshouwers, V., Zhang, N.W., Borm, T.J.A., Van Eck, H.J., Baker, B., Jacobsen, E., and Visser, R.G.F. (2005). Comparative genomics enabled the isolation of the *R3a* late blight Resistance gene in potato. Plant Journal *42*, 251-261.

International Barley Genome Sequencing, C., Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K.*, et al.* (2012). A physical, genetic and functional sequence assembly of the barley genome. Nature *491*, 711-716.

Invitrogen (2014). User Guide; Pichia Expression Kit, For expression of recombinant proteins in *Pichia pastoris*. Invitrogen by Life Technologies *MAN0000012*

Ito, S., Ihara, T., Tamura, H., Tanaka, S., Ikeda, T., Kajihara, H., Dissanayake, C., Abdel-Motaal, F.F., and El-Sayed, M.A. (2007). alpha-Tomatine, the major saponin in tomato,

induces programmed cell death mediated by reactive oxygen species in the fungal pathogen *Fusarium oxysporum*. FEBS Lett *581*, 3217-3222.

Jain, M., Fiddes, I.T., Miga, K.H., Olsen, H.E., Paten, B., and Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. Nature Methods *12*, 351-356.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of Resistance gene and Avirulence gene products confers rice blast resistance. EMBO Journal *19*, 4004-4014.

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balintkurti, P.J., and Jones, J.D.G. (1994). Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by Transposon Tagging. Science *266*, 789-793.

Jones, J.D., and Dangl, J.L. (2006). The plant immune system. Nature *444*, 323-329.

Jones, J.D., Witek, K., Verweij, W., Jupe, F., Cooke, D., Dorling, S., Tomlinson, L., Smoker, M., Perkins, S., and Foster, S. (2014). Elevating crop disease resistance with cloned genes. Philosophical Transactions of the Royal Society of London B Biology Sci *369*, 20130087.

Joosten, M., and de Wit, P. (1999). The tomato-C*ladosporium Fulvum* interaction: a versatile experimental system to study plant-pathogen interactions. Annual Review of Phytopathology *37*, 335-367.

Joosten, M.H., Vogelsang, R., Cozijnsen, T.J., Verberne, M.C., and de Wit, P.J. (1997). The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable Avr4 elicitors. Plant Cell *9*, 367-379.

Joosten, M.H.A.J., Cozijnsen, T.J., and de wit, P.J.G.M. (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.A.J., and de Wit, P.J.G.M. (1988). Isolation, purification and preliminary characterization of a protein specific for compatible *Cladosporium fulvum* (syn. *Fulvia fulva*)-tomato interactions. Physiological and Molecular Plant Pathology *33*, 241 - 253.

Jordan, T., Seeholzer, S., Schwizer, S., Toller, A., Somssich, I.E., and Keller, B. (2011). The wheat *Mla* homologue *TmMla1* exhibits an evolutionarily conserved function against powdery mildew in both wheat and barley. Plant Journal *65*, 610-621.

Joshi, R., K.,, and Nayak, S. (2010). Gene pyramiding-a broad spectrum technique for developing durable stress resistance in crops. Biotechnology and Molecular Biology Review *5*, 51-60.

Jupe, F., Pritchard, L., Etherington, G.J., Mackenzie, K., Cock, P.J., Wright, F., Sharma, S.K., Bolser, D., Bryan, G.J., Jones, J.D*., et al.* (2012). Identification and localisation of the *NB-LRR* gene family within the potato genome. BMC Genomics *13*, 75.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., and Shibuya, N. (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proceedings of the National Academy of Science U S A *103*, 11086-11091.

Kaloshian, I., Lange, W.H., and Williamson, V.M. (1995). An aphid-resistance locus is tightly linked to the nematode-Resistance gene, *Mi*, in tomato. Proceedings of the National Academy of Science U S A *92*, 622-625.

Kamoun, S., Van West, P., Vleeshouwers, V.G., de Groot, K.E., and Govers, F. (1998). Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. Plant Cell *10*, 1413-1426.

Kanzaki, H., Saitoh, H., Takahashi, Y., Berberich, T., Ito, A., Kamoun, S., and Terauchi, R. (2008). NbLRK1, a lectin-like receptor kinase protein of *Nicotiana benthamiana*, interacts with *Phytophthora infestans* INF1 elicitin and mediates INF1-induced cell death. Planta *228*, 977-987.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C*., et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kerr, E., A.,, and Bailey, D., L., (1964). Resistance to *Cladosporium fulvum* CKE. obtained from wild species of tomato. Canadian Journal of Botany/Revue Canadienne de Botanique *42*, 1541 - 1554.

Kimura, S., and Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A model fruit-bearing crop. CSH Protoc *2008*, pdb emo105.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature *366*, 751-756.

Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. Trends in Biochemical Sciences *19*, 415-421.

Kobe, B., and Deisenhofer, J. (1995). A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature *374*, 183-186.

Koeppel, D. (2008). Banana: The fate of the fruit that changed the world (Hudson Street Press; Penguin Books).

Kooman-Gersmann, M., Honee, G., Bonnema, G., and de Wit, P.J.G.M. (1996). A high-affinity binding site for the Avr9 peptide elicitor of *Cladosporium fulvum* is present on plasma membranes of tomato and other solanaceous plants. Plant Cell *8*, 929-938.

Kooman-Gersmann, M., Vogelsang, R., Vossen, P., Van den Hooven, H.W., Mahe, E., Honee, G., and de Wit, P.J. (1998). Correlation between binding affinity and necrosis-inducing activity of mutant Avr9 peptide elicitors. Plant Physiology *117*, 609-618.

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D.*, et al.* (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nature Biotechnol *30*, 693-700.

Kruger, J., Thomas, C.M., Golstein, C., Dixon, M.S., Smoker, M., Tang, S., Mulder, L., and Jones, J.D. (2002). A tomato cysteine protease required for *Cf-2*-dependent disease resistance and suppression of autonecrosis. Science *296*, 744-747.

Kruijt, M., Brandwat, B.F., and de Wit, P.J.G.M. (2004). Rearrangements in the *Cf-9* disease resistance gene cluster of wild tomato have resulted in three genes that mediate Avr9 responsiveness. Genetics *168*, 1655-1663.

Kubat, Z. (2007). Chromosome walking with BAC clones as a method of genome mapping. Plant Soil and Environment *53*, 447-450.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in *Arabidopsis* plants. Plant Cell *16*, 3496-3507.

Kuzma, J. (2016). Policy: reboot the debate on genetic engineering. Nature *531*, 165-167.

Lairini, K., Perez-Espinosa, A., Pineda, M., and Ruiz-Rubio, M. (1996). Purification and characterization of tomatinase from *Fusarium oxysporum* f. sp. lycopersici. Applied and Environmental Microbiology *62*, 1604-1609.

Lamari, L., and Bernier, C.C. (1989). Toxin of *Pyrenophora tritici repentis* - host-specificity, significance in disease, and inheritance of host-reaction. Phytopathology *79*, 740-744.

Lauge, R., Goodwin, P.H., de Wit, P.J.G.M., and Joosten, M.H.A.J. (2000). Specific HR-associated recognition of secreted proteins from *Cladosporium fulvum* occurs in both host and non-host plants. Plant Journal *23*, 735-745.

Lauge, R., Joosten, M., Van den Ackerveken, G., Van den Broek, H.W.J., and de Wit, P. (1997). The in planta-produced extracellular proteins Ecp1 and Ecp2 of *Cladosporium fulvum* are virulence factors. Molecular Plant-Microbe Interactions *10*, 725-734.

Lauge, R., Joosten, M.H.A.J., Haanstra, J.P.W., Goodwin, P.H., Lindhout, P., and de Wit, P.J.G.M. (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.

Lawrence, G.J., Finnegan, E.J., Ayliffe, M.A., and Ellis, J.G. (1995). The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell *7*, 1195-1206.

Lescot, M., Piffanelli, P., Ciampi, A.Y., Ruiz, M., Blanc, G., Leebens-Mack, J., da Silva, F.R., Santos, C.M.R., D'Hont, A., Garsmeur, O*., et al.* (2008). Insights into the *Musa* genome: Syntenic relationships to rice and between *Musa* species. BMC Genomics *9*.

Li, G., Huang, S., Guo, X., Li, Y., Yang, Y., Guo, Z., Kuang, H., Rietman, H., Bergervoet, M., Vleeshouwers, V.G*., et al.* (2011). Cloning and characterization of *R3b*; members of the *R3* superfamily of late blight Resistance genes show sequence and functional divergence. Molecular Plant-Microbe Interactions *24*, 1132-1142.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotechnology *30*, 390-392.

Liebrand, T.W., Kombrink, A., Zhang, Z., Sklenar, J., Jones, A.M., Robatzek, S., Thomma, B.P., and Joosten, M.H. (2014a). Chaperones of the endoplasmic reticulum are required for *Ve1*-mediated resistance to *Verticillium*. Molecular Plant Pathology *15*, 109-117.

Liebrand, T.W., Van den Berg, G.C., Zhang, Z., Smit, P., Cordewener, J.H., America, A.H., Sklenar, J., Jones, A.M., Tameling, W.I., Robatzek, S.*, et al.* (2013). Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. Proceedings of the National Academy of Science U S A *110*, 10010-10015.

Liebrand, T.W., Van den Burg, H.A., and Joosten, M.H. (2014b). Two for all: receptor-associated kinases SOBIR1 and BAK1. Trends in Plant Science *19*, 123-132.

Lin, W., Li, B., Lu, D., Chen, S., Zhu, N., He, P., and Shan, L. (2014). Tyrosine phosphorylation of protein kinase complex BAK1/BIK1 mediates *Arabidopsis* innate immunity. Proceedings of the National Academy of Science U S A *111*, 3632-3637.

Liu, D., Hu, R., Palla, K.J., Tuskan, G.A., and Yang, X. (2016). Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. Current Opinion in Plant Biology *30*, 70-77.

Logrieco, A., Moretti, A., Perrone, G., and Mule, G. (2007). Biodiversity of complexes of mycotoxigenic fungal species associated with *Fusarium* ear rot of maize and *Aspergillus* rot of grape. International Journal of Food Microbiology *119*, 11-16.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nature Methods *12*, 733-735.

Loutre, C., Wicker, T., Travella, S., Galli, P., Scofield, S., Fahima, T., Feuillet, C., and Keller, B. (2009). Two different *CC-NBS-LRR* genes are required for *Lr10*-mediated leaf rust resistance in tetraploid and hexaploid wheat. Plant Journal *60*, 1043-1054.

Luderer, R., de Kock, M.J.D., Dees, R.H.L., de Wit, P.J.G.M., and Joosten, M.H.A.J. (2002a). Functional analysis of cysteine residues of ECP elicitor proteins of the fungal tomato pathogen *Cladosporium fulvum*. Molecular Plant Pathology *3*, 91-95.

Luderer, R., Rivas, S., Nurnberger, T., Mattei, B., Van den Hooven, H.W., Van der Hoorn, R.A., Romeis, T., Wehrfritz, J.M., Blume, B., Nennstiel, D*., et al.* (2001). No evidence for binding between Resistance gene product Cf-9 of tomato and avirulence gene product Avr9 of *Cladosporium fulvum*. Molecular Plant-Microbe Interactions *14*, 867-876.

Luderer, R., Takken, F.L., de Wit, P.J., and Joosten, M.H. (2002b). *Cladosporium fulvum* overcomes *Cf-2*-mediated resistance by producing truncated Avr2 elicitor proteins. Molecular Microbiology *45*, 875-884.

Mace, M., E., Bell, A., A., and Beckman, C., H. (1981). Fungal Wilt Diseases of Plants (Academic Press).

Mackey, D., Belkhadir, Y., Alonso, J.M., Ecker, J.R., and Dangl, J.L. (2003). *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates *RPS2*-mediated resistance. Cell *112*, 379-389.

Mackey, D., Holt, B.F., 3rd, Wiig, A., and Dangl, J.L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for *RPM1*-mediated resistance in *Arabidopsis*. Cell *108*, 743-754.

Madoui, M.A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P., and Aury, J.M. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. BMC Genomics *16*, 327.

Maekawa, T., Kracher, B., Vernaldi, S., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2012). Conservation of NLR-triggered immunity across plant lineages. Proceedings of the National Acadamty of Science U S A *109*, 20119-20123.

Manly, K.F., Cudmore, R.H., Jr., and Meer, J.M. (2001). Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome *12*, 930-932.

Marmeisse, R., Van den Ackerveken, G.F., Goosen, T., de Wit, P.J., and Van den Broek, H.W. (1994). The in-planta induced Ecp2 gene of the tomato pathogen *Cladosporium fulvum* is not essential for pathogenicity. Current Genetics *26*, 245-250.

Marmeisse, R., Van den ackerveken, G.F.J.M., Goosen, T., de wit, P.J.G.M., and Van den Broek, H.W.J. (1993). Disruption of the Avirulence gene Avr9 in two races of the tomato pathogen *Cladosporium fulvum* causes virulence on tomato genotypes with the complementary Resistance gene *Cf9*. Molecular Plant-Microbe Interactions *6*, 412-417.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science *262*, 1432-1436.

McDonald, B.A., and Linde, C. (2002a). Pathogen population genetics, evolutionary potential, and durable resistance. Annual Review of Phytopathology *40*, 349-+.

McDonald, B.A., and Linde, C. (2002b). The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica *124*, 163-180.

McHale, L.K., Truco, M.J., Kozik, A., Wroblewski, T., Ochoa, O.E., Lahre, K.A., Knapp, S.J., and Michelmore, R.W. (2009). The genomic architecture of disease resistance in lettuce. Theoretical Applied Genetics *118*, 565-580.

Medzhitov, R., and Janeway, C.A., Jr. (1997). Innate immunity: impact on the adaptive immune response. Current Opinion in Immunology *9*, 4-9.

Mendes, B.M.J., Cardoso, S.C., Boscariol-Camargo, R.L., Cruz, R.B., Mourao, F.A.A., and Bergamin, A. (2010). Reduction in susceptibility to *Xanthomonas axonopodis* pv. *citri* in transgenic *Citrus sinensis* expressing the rice *Xa21* gene. Plant Pathology *59*, 68-75.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z., and Michelmore, R.W. (1998). The major Resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell *10*, 1817-1832.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell *15*, 809-834.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of Resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research *8*, 1113-1130.

Michielse, C.B., and Rep, M. (2009). Pathogen profile update: *Fusarium oxysporum*. Mol Plant Pathology *10*, 311-324.

Milne, I., Stephen, G., Bayer, M., Cock, P.J., Pritchard, L., Cardle, L., Shaw, P.D., and Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics *14*, 193-202.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. Proceedings of the National Acadamy of Science U S A *104*, 19613-19618.

Monosi, B., Wisser, R.J., Pennill, L., and Hulbert, S.H. (2004). Full-genome analysis of Resistance gene homologues in rice. Theoretical Applied Genetics *109*, 1434-1447.

Narusaka, M., Kubo, Y., Hatakeyama, K., Imamura, J., Ezura, H., Nanasato, Y., Tabei, Y., Takano, Y., Shirasu, K., and Narusaka, Y. (2013). Breaking restricted taxonomic functionality by dual Resistance genes. Plant Signaling and Behaviour *8*, e24244.

Narusaka, M., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y. (2009a). A dual Resistance gene system prevents infection by three distinct pathogens. Plant Signaling and Behaviour *4*, 954-955.

Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y. (2009b). *RRS1* and *RPS4* provide a dual Resistance-gene system against fungal and bacterial pathogens. Plant Journal *60*, 218-226.

Noel, L., Moores, T.L., Van Der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E., and Jones, J.D. (1999). Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell *11*, 2099-2112.

Oerke, E.C. (2006). Crop losses to animal pests, plant pathogens, and weeds. Encyclopedia of pest management.1-5.

Palmer, C.L., and Skinner, W. (2002). *Mycosphaerella graminicola*: latent infection, crop devastation and genomics. Molecular Plant Pathology *3*, 63-70.

Panter, S.N., Hammond-Kosack, K.E., Harrison, K., Jones, J.D.G., and Jones, D.A. (2002). Developmental control of promoter activity is not responsible for mature onset of *Cf-9B*-mediated resistance to leaf mold in tomato. Molecular Plant-Microbe Interactions *15*, 1099-1107.

Park, C.H., Chen, S., Shirsekar, G., Zhou, B., Khang, C.H., Songkumarn, P., Afzal, A.J., Ning, Y., Wang, R., Bellizzi, M.*, et al.* (2012). The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. Plant Cell *24*, 4748-4762.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B.H., and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell *91*, 821-832.

Parniske, M., and Jones, J.D. (1999). Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proceedings of the National Acadamy of Science U S A *96*, 5850-5855.

Parniske, M., Wulff, B.B., Bonnema, G., Thomas, C.M., Jones, D.A., and Jones, J.D. (1999). *Homologues of the Cf-9 disease resistance gene* (*Hcr9s*) are present at multiple loci on the short arm of tomato chromosome 1. Molecular Plant Microbe Interactions *12*, 93-102.

Pelham, J., Fletcher, J.T., and Hawkins, J.H. (1970). Establishment of a new strain of Tobacco Mosaic Virus resulting from use of resistant varieties of tomato. Annual of Applied Biology *65*, 293-&.

Peralta, I.E., Knapp, S.K., and Spooner, D.M. (2005). New species of wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae) from Northern Peru. Systemic Botany *30*, 424-434.

Pitblado, R.E., and McNeill, B.H. (1983). Genetic basis of resistance to *Psudomonas syringae* pv. tomato in field tomatoes. Canadian Journal of Plant Pathology *5*, 251 - 255.

Ploetz, R.C. (2006). *Fusarium* Wilt of banana is caused by several pathogens referred to as *Fusarium oxysporum* f. sp. *cubense*. Phytopathology *96*, 653-656.

Postma, J., Liebrand, T., W, H.,, Bi, G., Evrard, A., Bye, R., R.,, Mbengue, M., Joosten, M., H, A, J.,, and Robatzek, S. (2015). The Cf-4 receptor-like protein associates with the BAK1 receptor-like kinase to initiate receptor endocytosis and plant immunity. bioRxiv, 1-37.

Potrykus, I., Bilang, R., Futterer, J., Sautter, C., Schrott, M., and Spangenberg, G. (1998). Genetic engineering of crop plants (New York: Marcel Dekker, Inc.).

Prakash, C.S. (2001). The genetically modified crop debate in the context of agricultural evolution. Plant Physiology *126*, 8-15.

Qi, D., Dubiella, U., Kim, S.H., Sloss, D.I., Dowen, R.H., Dixon, J.E., and Innes, R.W. (2014). Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 is dependent on s-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. Plant Physiology *164*, 340-351.

Quick, J., Quinlan, A.R., and Loman, N.J. (2014). A reference bacterial genome dataset generated on the MinION portable single-molecule nanopore sequencer. Gigascience *3*, 22.

Reuter, J.A., Spacek, D.V., and Snyder, M.P. (2015). High-throughput sequencing technologies. Mol Cell *58*, 586-597.

Risse, J., Thomson, M., Patrick, S., Blakely, G., Koutsovoulos, G., Blaxter, M., and Watson, M. (2015). A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. Gigascience *4*, 60.

Rommens, C.M., and Kishore, G.M. (2000). Exploiting the full potential of disease-resistance genes for agricultural use. Curr Opinion in Biotechnology *11*, 120-125.

Rommens, C.M., Rudenko, G.N., Dijkwel, P.P., van Haaren, M.J., Ouwerkerk, P.B., Blok, K.M., Nijkamp, H.J., and Hille, J. (1992). Characterization of the *Ac*/*Ds* behaviour in transgenic tomato plants using plasmid rescue. Plant Molecular Biology *20*, 61-70.

Ron, M., and Avni, A. (2004). The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. Plant Cell *16*, 1604-1615.

Rooney, H.C., Van't Klooster, J.W., Van der Hoorn, R.A., Joosten, M.H., Jones, J.D., and de Wit, P.J. (2005). *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for *Cf-2*-dependent disease resistance. Science *308*, 1783-1786.

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., and Jaffe, D.B. (2013). Characterizing and measuring bias in sequence data. Genome Biology *14*, R51.

Salmeron, J.M., Oldroyd, G.E., Rommens, C.M., Scofield, S.R., Kim, H.S., Lavelle, D.T., Dahlbeck, D., and Staskawicz, B.J. (1996). Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell *86*, 123-133.

Sanchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., Hansen, G., Valkenburg, D.J., Thomma, B.P., and Mesters, J.R. (2013). Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. Elife *2*, e00790.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. Nature *265*, 687-695.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko, T., Nakamura, Y., Shibata, D., Aoki, K*., et al.* (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635-641.

Schatz, M.C. (2015). Biological data sciences in genome research. Genome Research *25*, 1417-1422.

Schoonbeek, H.J., Wang, H.H., Stefanato, F.L., Craze, M., Bowden, S., Wallington, E., Zipfel, C., and Ridout, C.J. (2015). *Arabidopsis* EF-Tu receptor enhances bacterial disease resistance in transgenic wheat. New Phytologist *206*, 606-613.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science *16*, 117-125.

Shabab, M., Shindo, T., Gu, C., Kaschani, F., Pansuriya, T., Chintha, R., Harzen, A., Colby, T., Kamoun, S., and Van der Hoorn, R.A. (2008). Fungal effector protein Avr2 targets diversifying defense-related cys proteases of tomato. Plant Cell *20*, 1169-1183.

Shah, A. (2001). Bananas. In Global issues Social, political, economic and environmental issues that effect us all (http://www.globalissues.org/article/241/bananas).

Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J.E., and Innes, R.W. (2003). Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. Science *301*, 1230-1233.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., de Both, M.*, et al.* (1998). Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell *10*, 1055-1068.

Siqueira, J.F., Fouad, A.F., and Rocas, I.N. (2012). Pyrosequencing as a tool for better understanding of human microbiomes. Journal of Oral Microbiology *4*.

Snyder, M., Du, J., and Gerstein, M. (2010). Personal genome sequencing: current approaches and challenges. Genes and Development *24*, 423-431.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., Van der Hoorn, R.A., and Kamoun, S. (2009). Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. Proceedings of the Natlional Acadamy of Science U S A *106*, 1654-1659.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H.*, et al.* (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science *270*, 1804-1806.

Soumpourou, E., Iakovidis, M., Chartrain, L., Lyall, V., and Thomas, C.M. (2007). The *Solanum pimpinellifolium Cf-Ecp1* and *Cf-Ecp4* genes for resistance to *Cladosporium fulvum* are located at the *Milky Way* locus on the short arm of chromosome 1. Theoretical Applied Genetics *115*, 1127-1136.

Stergiopoulos, I., Cordovez, V., Okmen, B., Beenen, H.G., Kema, G.H., and de Wit, P.J. (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. Moleculat Plant Pathology *15*, 447-460.

Stergiopoulos, I., de Kock, M.J.D., Lindhout, P., and de Wit, P.J.G.M. (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., Kourmpetis, Y.A., Slot, J.C., Bakker, F.T., de Wit, P.J., and Rokas, A. (2012). *In silico* characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Molecular Biology an Evolution *29*, 3371-3384.

Stergiopoulos, I., Van den Burg, H.A., Okmen, B., Beenen, H.G., Van Liere, S., Kema, G.H.J., and de Wit, P.J.G.M. (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 USA *107*, 7610-7615.

Steuernagel, B., Periyannan, S.K., Hernandez-Pinzon, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D*., et al.* (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nature Biotechnology.

Stevens, M., A.,, and Rick, C., M., (1988). Genetics and breeding (London: Chapman and Hall).

Swiderski, M.R., and Innes, R.W. (2001). The *Arabidopsis PBS1* Resistance gene encodes a member of a novel protein kinase subfamily. Plant Journal *26*, 101-112.

Szczesny, R., Buttner, D., Escolar, L., Schulze, S., Seiferth, A., and Bonas, U. (2010). Suppression of the AvrBs1-specific hypersensitive response by the YopJ effector homolog AvrBsT from Xanthomonas depends on a SNF1-related kinase. New Phytologist *187*, 1058-1074.

Takken, F.L., Luderer, R., Gabriels, S.H., Westerink, N., Lu, R., de Wit, P.J., and Joosten, M.H. (2000). A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. Plant Journal *24*, 275-283.

Takken, F.L.W., Schipper, D., Nijkamp, H.J.J., and Hille, J. (1998). Identification and *Ds*-tagged isolation of a new gene at the *Cf-4* locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. Plant Journal *14*, 401-411.

Takken, F.L.W., Thomas, C.M., Joosten, M.H.A.J., Golstein, C., Westerink, N., Hille, J., Nijkamp, H.J.J., de Wit, P.J.G.M., and Jones, J.D.G. (1999). A second gene at the tomato *Cf-4* locus confers resistance to *Cladosporium fulvum* through recognition of a novel Avirulence determinant. Plant Journal *20*, 279-288.

Talbot, N.J. (2003). On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. Annual Reviews of Microbiology *57*, 177-202.

Taudien, S., Steuernagel, B., Ariyadasa, R., Schulte, D., Schmutzer, T., Groth, M., Felder, M., Petzold, A., Scholz, U., Mayer, K.F.*, et al.* (2011). Sequencing of BAC pools by different next generation sequencing platforms and strategies. BMC Research Notes *4*.

Thomas, C.M., Dixon, M.S., Parniske, M., Golstein, C., and Jones, J.D. (1998). Genetic and molecular analysis of tomato *Cf* genes for resistance to *Cladosporium fulvum*. Philosophical Transactions Royal Society London B Biological Sciences *353*, 1413-1424.

Thomas, C.M., Jones, D.A., Parniske, M., Harrison, K., Balint-Kurti, P.J., Hatzixanthis, K., and Jones, J.D.G. (1997). Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell *9*, 2209-2224.

Thomas, C.M., Tang, S., Hammond-Kosack, K., and Jones, J.D. (2000). Comparison of the hypersensitive response induced by the tomato *Cf-4* and *Cf-9* genes in *Nicotiana* spp. Molecular Plant-Microbe Interactions *13*, 465-469.

Thomma, B.P.H.J., Van Esse, H.P., Crous, P.W., and de Wit, P.J.G.M. (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.

Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research *38*, e159.

Tsuda, K., and Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Current Opinions in Plant Biology *13*, 459-465.

Tuori, R.P., Wolpert, T.J., and Ciuffetti, L.M. (1995). Purification and immunological characterization of toxic components from cultures of *Pyrenophora tritici-repentis*. Molecular Plant-Microbe Interactions *8*, 41-48.

Van Buren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E*., et al.* (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature *527*, 508-511.

Van den Ackerveken, G.F.J.M., Van Kan, J.A.L., Joosten, M.H.A.J., Muisers, J.M., Verbakel, H.M., and de wit, P.J.G.M. (1993a). Characterization of two putative pathogenicity genes of the fungal tomato pathogen *Cladosporium fulvum*. Molecular Plant-Microbe Interactions *6*, 210-215.

Van den Ackerveken, G.F.J.M., Vossen, P., and de Wit, P.J.G.M. (1993b). The Avr9 race-specific elicitor of *Cladosporium fulvum* is processed by endogenous and plant proteases. Plant Physiology *103*, 91-96.

Van den Burg, H.A., Harrison, S.J., Joosten, M.H., Vervoort, J., and de Wit, P.J. (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., Westerink, N., Francoijs, K.J., Roth, R., Woestenenk, E., Boeren, S., de Wit, P.J., Joosten, M.H., and Vervoort, J. (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 Biological Chemistry *278*, 27340-27346.

Van der Biezen, E.A., and Jones, J.D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci *23*, 454-456.

Van der Hoorn, R.A., Kruijt, M., Roth, R., Brandwagt, B.F., Joosten, M.H., and de Wit, P.J. (2001a). Intragenic recombination generated two distinct *Cf* genes that mediate Avr9 recognition in the natural population of *Lycopersicon pimpinellifolium*. Proceedings of the National Acadadamy of Science U S A *98*, 10493-10498.

Van der Hoorn, R.A., Roth, R., and de Wit, P.J. (2001b). Identification of distinct specificity determinants in resistance protein Cf-4 allows construction of a Cf-9 mutant that confers recognition of Avirulence protein Avr4. Plant Cell *13*, 273-285.

Van der Hoorn, R.A.L., Laurent, F., Roth, R., and de Wit, P.J.G.M. (2000). Agroinfiltration is a versatile tool that facilitates comparative analyses of *Avr9/Cf-9*-induced and *Avr4/Cf-4*-induced necrosis. Molecular Plant-Microbe Interactions *13*, 439-446.

Van der Hoorn, R.A.L., Wulff, B.B.H., Rivas, S., Durrant, M.C., Van der Ploeg, A., de Wit, P.J.G.M., and Jones, J.D.G. (2005). Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. Plant Cell *17*, 1000-1015.

Van Esse, H.P., Van't Klooster, J.W., Bolton, M.D., Yadeta, K.A., Van Baarlen, P., Boeren, S., Vervoort, J., de Wit, P.J., and Thomma, B.P. (2008). The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. Plant Cell *20*, 1948-1963.

Van Kan, J.A.L., Van den Ackerveken, G.F.J.M., and de Wit, P.J.G.M. (1991). Cloning and characterization of cDNA of Avirulence gene Avr9 of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Molecular Plant-Microbe Interactions *4*, 52-59.

Van't Slot, K., A, E.,, and Knogge, W. (2002). A dual role for microbial pathogen-derived effector proteins in plant disease and resistance. Critical Reviews in Plant Sciences *21*, 229-271.

Van Kan, J.A.L., Van den Ackerveken, G.F.J.M., and de wit, P.J.G.M. (1991). Cloning and Characterization of Cdna of Avirulence Gene Avr9 of the Fungal Pathogen *Cladosporium fulvum,* Causal Agent of Tomato Leaf Mold. Molecular Plant-Microbe Interactions *4*, 52-59.

Vervoort, J., Van den Hooven, H.W., Berg, A., Vossen, P., Vogelsang, R., Joosten, M.H.A.J., and de Wit, P.J.G.M. (1997). The race-specific elicitor Avr9 of the tomato pathogen *Cladosporium fulvum:* A cystine knot protein-sequence-specific H-1 NMR assignments, secondary structure and global fold of the protein. FEBS Letters *404*, 153-158.

Vleeshouwers, V.G., Driesprong, J.D., Kamphuis, L.G., Torto-Alalibo, T., Van't Slot, K.A., Govers, F., Visser, R.G., Jacobsen, E., and Kamoun, S. (2006). Agroinfection-based high-throughput screening reveals specific recognition of INF elicitins in *Solanum*. Molecular Plant Pathology *7*, 499-510.

Wang, G.L., Ruan, D.L., Song, W.Y., Sideris, S., Chen, L., Pi, L.Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S*., et al.* (1998). *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell *10*, 765-779.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., and Qiu, J.L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotechnology *32*, 947-951.

Wang, Y., Diehl, A., Wu, F., Vrebalov, J., Giovannoni, J., Siepel, A., and Tanksley, S.D. (2008). Sequencing and comparative analysis of a conserved syntenic segment in the *Solanaceae*. Genetics *180*, 391-408.

Weinthal, D., Tovkach, A., Zeevi, V., and Tzfira, T. (2010). Genome editing in plant cells by zinc finger nucleases. Trends in Plant Science *15*, 308-321.

Welsh, C., Lewis, K., and Woods, A. (2009). The outbreak history of *Dothistroma* needle blight: an emerging forest disease in North Western British Columbia, Canada. Canadian Journal of Forest research *39*, 2505 - 2519.

Westergaard, M., C, W., (2012). Mapping the Resistance gene of the *Cladosporium fulvum* effector Ecp2 in *Nicotiana paniculata*. MSc, Faculty of Science, University of Copenhagen.

Westerink, N., Brandwagt, B.F., de Wit, P.J., and Joosten, M.H. (2004). *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the *Cf-4* locus (*Hcr9-4E*) by secretion of a stable Avr4E isoform. Molecular Microbiology *54*, 533-545.

Westerink, N., Roth, R., Van den Burg, H.A., de Wit, P.J., and Joosten, M.H. (2002). The Avr4 elicitor protein of *Cladosporium fulvum* binds to fungal components with high affinity. Molecular Plant-Microbe Interact *15*, 1219-1227.

White, F.F., Yang, B., and Johnson, L.B. (2000). Prospects for understanding Avirulence gene function. Current Opinions in Plant Biology *3*, 291-298.

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the Tobacco Mosaic Virus Resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell *78*, 1101-1115.

Wu, F., Eannetta, N.T., Xu, Y., Plieske, J., Ganal, M., Pozzi, C., Bakaher, N., and Tanksley, S.D. (2010). COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theoretical Applied Genetics *120*, 809-827.

Wubben, J.P., Joosten, M.H., and de Wit, P.J. (1994). Expression and localization of two in planta induced extracellular proteins of the fungal tomato pathogen *Cladosporium fulvum*. Molecular Plant-Microbe Interactions *7*, 516-524.

Wulff, B.B.H., Chakrabarti, A., and Jones, D.A. (2009a). Recognitional specificity and evolution in the tomato-*Cladosporium fulvum* pathosystem. Molecular Plant-Microbe Interactions *22*, 1191-1202.

Wulff, B.B.H., Heese, A., Tomlinson-Buhot, L., Jones, D.A., de la Pena, M., and Jones, J.D.G. (2009b). The major specificity-determining amino acids of the tomato Cf-9 disease resistance protein are at hypervariable solvent-exposed positions in the central leucine-rich repeats. Molecular Plant-Microbe Interactions *22*, 1203-1213.

Wulff, B.B.H., Horvath, D.M., and Ward, E.R. (2011). Improving immunity in crops: new tactics in an old game. Current Opinion in Plant Biology *14*, 468-476.

Xiao, S., Charoenwattana, P., Holcombe, L., and Turner, J.G. (2003). The *Arabidopsis* genes *RPW8.1* and *RPW8.2* confer induced resistance to powdery mildew diseases in tobacco. Molecular Plant-Microbe Interactions *16*, 289-294.

Yoshida, K., Schuenemann, V.J., Cano, L.M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F.N., Kamoun, S., Krause, J.*, et al.* (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. Elife *2*, e00731.

Young, N.D., and Tanksley, S.D. (1989). RFLP analysis of the size of chromosomal segments retained around the *TM-2* locus of tomato during backcross breeding Theoretical and Applied Genetics *77*, 353-359.

Yuan, Y.N., Haanstra, J., Lindhout, P., and Bonnema, G. (2002). The *Cladosporium fulvum* Resistance gene *Cf-Ecp3* is part of the *Orion* cluster on the short arm of tomato chromosome 1. Molecular Breeding *10*, 45-50.

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., Zou, Y., Gao, M., Zhang, X., Chen, S.*, et al.* (2010). Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. Cell Host and Microbe *7*, 290-301.

Zhang, W., Fraiture, M., Kolb, D., Loffelhardt, B., Desaki, Y., Boutrot, F.F., Tor, M., Zipfel, C., Gust, A.A., and Brunner, F. (2013). *Arabidopsis* receptor-like protein30 and receptor-like kinase suppressor of BIR1-1/EVERSHED mediate innate immunity to necrotrophic fungi. Plant Cell *25*, 4227-4241.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. Journal of Computational Biology *7*, 203-214.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R., and Schulze-Lefert, P. (2001). Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a Rar1-independent signaling pathway. Plant Cell *13*, 337-350.

Zhu, M., Shao, F., Innes, R.W., Dixon, J.E., and Xu, Z. (2004). The crystal structure of *Pseudomonas* Avirulence protein AvrPphB: a papain-like fold with a distinct substrate-binding site. Proceedings of the National Acadamy of Sciences U S A *101*, 302-307.

Zipfel, C. (2009). Early molecular events in PAMP-triggered immunity. Current Opinion Plant Biology *12*, 414-420.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. Cell *125*, 749-760.

Zipfel, C., and Robatzek, S. (2010). Pathogen-associated molecular pattern-triggered immunity: veni, vidi...? Plant Physiology *154*, 551-554.

Appendix 1: Materials and methods (data linked to Chapter 2)

Appendix Table 1.1. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
TG236	TG236F2	GGAAGGAAAAGAGCAGTTCTAGTT	50	60
	TG236R2	GCACCAGATTTTGAGGCTGAGGCT		
TG184	TG184F1	TCTTCATGTGGTAAGTTGCTCTTTG	50	60
	TG184R1	CCACTATTCATCTCTTCCAAAGGTC		
SNPE	SNPEF1	GAGGAAAGGTGGTCTCTAGG	50	60
	SNPER1	CTGGCAAAGACCATACTAGC		
CT116	CT116F1	AATATCTTCGAGGCCGATTGA	50-55	60
	CT116R1	TAAAAGCCCATGAATGTTGAGG		
SNP-nn1	2493200-bN1F1	AAGTCTTTCGAGATTGGTGA	50-55[1]	60
	2493200-bN1R1	CCGTTAGAGGAAAAGGGTAT		
GJ32	GJ32F1	AAGTGTTCAACTGACCATCC	50-55	60
	GJ32R1	CAACCAAGTGCTTCTACCTC		
GJ44	GJ44F	ATTGGTACGACACAAAGGAT	50-55	60
	GJ44R	TATTTCTTTGTGTTGCCTCA		
GJ43	GJ43F	TGACGTAGGAAAAGGGAATA	50-55	60
	GJ43R	TACACGTAGGACACCACGTA		
SNPN	SNPNF1	AAGGCTTAATCAAACACGTC	50	60
	SNPNR1	TACCTTGTCCCGATCCTAC		
SNPQ	QF1	CCATTAGTAGAGGTCGGGTA	51	60
	QR1	TGGGATGTTTACCATTAACC		
TG58	TG58F	CTGAGTAGATCCTGTGTGATACGGAA	50	60
	TG58R	CCCAATTCCCTTTTCATGTATGGTTAC		
60250	60250F	TGACATGAGAAGGTACATCTTGTC	50	60
	60250R	GGATGAACTATGGAGAAATACTGTAGTG		
TG67	TG67F	TGAAGAATCCAGAGCTTATAAATATG	50	60
	TG67R	CTTGCATAAATATGTTAACATATTCTTG		
TG24	TG24F	ATGATGATATCTTGCAAGAATTTTTTCTTAG	50	60
	TG24R	TTGTGATGTTCTATTCCTCTATTCTGT		
2A	OR2A/BF1	CTCTGTCAACTTGCTTCATC	50 - 58	60 - 90
	OR2AR1	TGAGGCTCGAATCTTAATTGAC		
2B	OR2A/BF1	CTCTGTCAACTTGCTTCATC	55	60 - 90
	OR2BR1	CTCACCAAATTTAGGCGAAATGTA		
2C	OR2CF1	CTCTTTCAACTTGTTTTCTCCTC	55	60
	OR2CR1	TGAGGCTCGAATCTTAATTGAC		
T7 and T3	T7	GTAATACGACTCACTATAGGG	51 - 55	60 to 180
	T3	CAATTAACCCTCACTAAAGGG		
Set *ψ2A5*	5' ψOR2A57B	TAATTGGGTCAGAAGATAATATTCT	61[1]	90
	3' OR2A/B7B	ATTGATAAACGAAAACTAATGCTTTT		
Set *2A*	5' OR2A7B	TCACTTTCAACCCGATGATTGC	64[1]	90
	3' OR2A/B7B	ATTGATAAACGAAAACTAATGCTTTT		
***2A* homolog sequencing**	OR2A607F	CAAATCTAAGACTTTGGAACACG	N/A	N/A
	OR2A607R	TGGAAAGGTGGAAAACTCCTTC	N/A	N/A
	OR2A1300F	CCATTTGAATGGGACTATACCATC	N/A	N/A
	OR2A1300R	AATAGGACCTTGCAGCTGATTTT	N/A	N/A
	OR2A2000F	CTCTCATCCAATGGATTTAGTGG	N/A	N/A
	OR2A2000R	CAAAACTCGAGGAAGTTCAAGG	N/A	N/A
	OR2A2700F	TCAATATCCAGCATGGTTTTCGA	N/A	N/A

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
***2A* homolog sequencing**	OR2A2700R	TAAATCCATCCTCGAAAACCATG	N/A	N/A
	OR2A504-482 R	CTGGAGACGAAGAACATGTAACT	N/A	N/A
***2A* homolog cloning**	OR2A5' primer F OR2A3' primer R	ATGGGTTACGTAAAACTTGT CTAACATCTTTTCTTGTGTTT	53[1]	90
ψOR2A4	hcr9-5 F 530 hcr9-5 R 1020	TCATCCAGGAATAGTCATACTTAGATTCGAAG GCCACTCAAAATTGTTATTTCAAGTGA	50[1]	60
ψOR2A5	OR2AF1 OR2A5R	CAAGAGTTGGAAAACTTACTTAGT CTCCTTCGGGCAATGTCCCG	59[1]	120
ψOR2C	OR2B-H pimp3 prime R	TTATCCTGTCAGGCATTCTT	50	60
	OR2B-H pimp5 prime F	ATGGAATAGCAGTGCATCAC	50	60
Golden gate cloning *2A*	GGOR2AF1 GGOR2AR1	AAGGTCTCAAATGATGGGTTACGTAAAACTT AAGGTCTCACCCTTAATTGGGTCAAGTTCTT	51[1]	90
	GGOR2AF2 GGOR2AR2	AAGGTCTCAAGGGACCTTGACCTTCGCTTTG AAGGTCTCAAAGCCTAACATCTTTTCTTGTG	47[1]	90
pICH86988 insert	pICH86988F	TTGGAGAGGACACGCTCGAG	50	60-120
	pICH86988R1	TTCTCGCATATCTCATTAAAGCAGG		
	PR1a F	ATGGGATTTGTTCTCTTTTCACAAT	50	60
	pICH86988R1	TTCTCGCATATCTCATTAAAGCAGG		
	pICH86988F OR2A670R	TTGGAGAGGACACGCTCGAG TGGAAAGGTGGAAAACTCCTTC	50	180
	OR2A2700F	TCAATATCCAGCATGGTTTTCGA	50	180
	pCIH86988R1	TTCTCGCATATCTCATTAAAGCAGG		
Altering *Ecp2* variants for Golden Gate cloning	*Cladosporium fulvum* Ecp2 F	AAGGTCTCaAATGATGGGATTTGTTCTCTTTTCAC	61[1]	67
	Cladosporium fulvum Ecp2 R	AAGGTCTCaAAGCCTAGTCATCG		
	Cladosporium fulvum Avr4 F	AAGGTCTCaAATGATGGGATTTGTTCTCTTTTCAC	61[1]	67
	Cladosporium fulvum Avr4 R	AAGGTCTCaAAGCTCAATAGCCAG		
	Dothistroma septosporum Ecp2 F	AAGGTCTCaAATGATGGGTTTCGTGTTGTTCTC	61[1]	67
	Dothistroma septosporum Ecp2 R	AAGGTCTCaAAGCTCATTCGTCG		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
Altering *Ecp2* variants for Golden Gate cloning	*Fusarium graminearum* Ecp2 F	AAGGTCTCaAATGATGGGTTTCGTGCTTTTCT	61[1]	67
	Fusarium graminearum Ecp2 R	AAGGTCTCaAAGCTCAGTAACCTGTAGA		
	Fusarium oxysporum Ecp2 F	AAGGTCTCaAATGATGGGTTTTGTGTTGTTTTCA	61[1]	67
	Fusarium oxysporum Ecp2 R	AAGGTCTCaAAGCTCATGTATGG		
	Magnaporthe grisea Ecp2 F	AAGGTCTCaAATGATGGGATTTGTGTTGTTCTCA	61[1]	67
	Magnaporthe grisea Ecp2 R	AAGGTCTCaAAGCTCACCTCTTTG		
	Mycosphaerella fijiensis Ecp2 F	AAGGTCTCaAATGATGGGTTTCGTGTTGTTCTC	61[1]	67
	Mycosphaerella fijiensis Ecp2 R	AAGGTCTCaAAGCTCAGTTAGGTGG		
	Mycosphaerella graminicola Ecp2 F	AAGGTCTCaAATGATGGGTTTTGTTTTGTTCTCA	61[1]	67
	Mycosphaerella graminicola Ecp2 R	AAGGTCTCaAAGCTCAGTTGTGT		
	Septoria musiva Ecp2 F	AAGGTCTCaAATGATGGGTTTCGTGCTTTTCT	61[1]	67
	Septoria musiva Ecp2 R	AAGGTCTCaAAGCTCACAAGTC		
	Verticillium dahlia Ecp2 F	AAGGTCTCaAATGATGGGTTTCGTGTTGTTCTC	61[1]	67
	Verticillium dahlia Ecp2 R	AAGGTCTCaAAGCTCACCTGTACTTACC		
eEF1Art	EFArtF[2]	AGTCAACTACCACTGGTCAC		
	EFArtR[2]	GTGCAGTAGTACTTAGTGGTC		
PVX replicase	PVXreplicase F	CTAAACCTGTCCCTGCAAGG	50	90
	PVX:replicase R	TGATGTCCGAGCCCTTGC		
PVX 25K	PVX 25K F	AGTAGTTTGAAAAGTTTAGGT	50	90
	PVX 25K R	TGATGTTTCCAAGTAGAAGT		
35S:*OR2A*	35S promoter F	AAGACCCTTCCTCTATATAAGG	50	90
	OR2AR1	TGAGGCTCGAATCTTAATTGAC		
	35S promoter F	AAGACCCTTCCTCTATATAAGG	50	90
	35S:OR2AR	AAGTTTTACGTAACCCATCATT		
EFa1	EFa1F[3]	TACTGGTGGTTTTGAAGCTG		
	EFa1R[3]	AACTTCCTTCACGATTTCATCATA	50	90
	C2At1g30580R	TCTCTCCACAGCAGCACTGAAAGG		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
COSII markers[4]	C2At5g13450F	AGGTTTGCTGAGTTAACCATGGC	50-55	60
	C2At5g13450R	TCTGTTCAATGTTCACCTTCTTCCC		
	C2At5g08430F	TGAAAGAATAAAAGCTGGTTTGCT	50-55	60
	C2At5g08430R	AACTTGTGCTTTTGACTCAAGCTCCA		
	C2At1g10240F	AGCCGGCTAGCACATTTTGTGG	50-55	60
	C2At1g10240R	TCTGGTATTTGAAAGCAATTTCCTG		
	C2At4g01210F	ACTATAGAGACGTTCTTGGTGAATATGGA	50-55	60
	Ntd4g01210R	TTTCTTATTAGTCTGTTTCAAAAAGTAT		
	C2At3g19230F	TCCTCCTTTGGCTTGTTAGCATTCC	50-55	60
	C2At3g19230R	TTGTTGGATTTTTTCATTTCCACCAC		
	Ntd4g03200F	GAGATTGGCCTCTATGGTTGCTGTT	50-55	60
	4g03200R2	GCCAGCATATCTGCTGCACAGCAC		
	C2At4g37130F	TTACAGCAAACTGTAGCAAGATTTGAG	50-55	60
	C2At4g37130R	TGCTGTTTTCATTGATTCAATGTACTG		
	C2At1g30580F	TTCTGCCGAAGATTCATGCATGG	50-55	60
	C2At1g30580R	TCTCTCCACAGCAGCACTGAAAGG		
	C2At5g37260F	AGAGCACAAAAAGTTCCTTGAAGC	50-55	60
	Ntd5g37260R	AGATTTGGCTCTTAATAATGACGGCAG		
	C2At1g74730F	AATAATGGCTTTAGTTCCAAGCCC	50-55	60
	C2At1g74730R	AACCATTGCTTTTTCCAATTGTCATC		
	C2At1g33970F	TGGAAGTGCAATAAGTGATGATTGGG	50-55	60
	C2At1g33970R	TCGTTTCCAACAAATTCAGGTTCAG		
	C2At2g14260F	AGGATCTATACCCCTCTATAGAGCC	50-55	60
	2g14260R1	CGAAGGTTAAATATATACAACATTTGA		
	C2At3g56040F	TCGCTATTGGATATAATGCGTAATGC	50-55	60
	C2At3g56040R	AACTCAGCAACCTCTATTAGCAACTC		
	C2At1g80460F	TCTTTATAACGCCATTGTTTGGATG	50-55	60
	C2At1g80460R	AATCTCAGAATTGCTAATAATCTTTGG		
	3g51010F1	TAATTTGCAAAAGGCCGTTCAA	50-55	60
	3g51010R1	AGGGAACCAAGACAAGGCAATA		
	C2At1g44790F	TCGGTTTTATCAAAGGCTATCGTC	50-55	60
	C2At1g44790R	TGTTACTGTTCTACCTGGGAATTCTGG		
	C2At1g21690F	ATGCAGAGCTCTCAGCCATGGG	50-55	60
	1g21690R1	AACAACTAATGCGGGCACTACAA		
	C2At1g77250F	AATCAAAGAAAGTCCCATAAGAAGTTGG	50-55	60
	C2At1g77250R	AGCCGTCACAAAGTACGATTTTGTCATC		
	C2At1g78230F	ACTCTGGGATCACTACCAAAGGGTC	50-55	60
	C2At1g78230R	AGAGAATTGTAGTTTGCCACAAGCTG		
	C2At5g25900F	TGCTAATTGGGCTGAAACTTATGG	50-55	60
	C2At5g25900R	TGTTAGCTTTCTAGTTGAGATGGATG		
	Ntd2g46580F	TGACCCAGATAAACTTAAGGTAATTCG	50-55	60
	2g46580R1	ACTGAACATATGGGCTGTGCAGAT		
	3g55360F2	TGTTAGTGATCTCCAGATGAAGATTGG	50-55	60
	3g55360R2	GTCTTCCTTCCTTTCCATCAAAGAG		
	C2At3g54860F	TGTTGATGGAAGGCGTTCTTTGGT	50-55	60
	C2At3g54860R	TGTAGTTGCTACAATTATGTCATATGCC		
	C2At2g01720F	ACAAATTGGTACATGCTGGTGCTC		
	C2At2g01720R	TGGCCTGTTAGACTGATATTCAAC	50-55	60
	C2At5g22620R	TCCCTGAGGGTCCATCAGGAAAATC		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
COSII markers[4]	C2At5g09880F	AGGGTCAAGCATATCCATGTGGAC	50-55	60
	C2At5g09880R	TCTCTTGTGCATGGCTTGTTGAGC		
	C2At5g64730F	TGAAGTCCGCGATGTCCATGTCAC	50-55	60
	5g64730R1	TCCGCCACTGGGTGTAGGTTCTT		
	C2At3g52220F	TGCTCGGGTGGATGGTCTTGG	50-55	60
	C2At3g52220R	TGATGGTGAACTTGGTTCTTCCC		
	C2At5g50720F	TTGGATGATGAACAATGGCTTGC	50-55	60
	5g50720R1	CGGGAGTGATGAAGTCCACAAA		
	C2At3g52640F	TACCTTGGCAGTAGAAGATTTCTTCTTG	50-55	60
	C2At3g52640R	AACCCTTTCCAACTGATCCAATTTC		
	C2At5g19690F	AGAGAGGCTTATGCATGGTTGAGC	50-55	60
	C2At5g19690R	TAGTTTGATAGCCATAGTCCCACC		
	C2At1g24340F	ACAAATATTTGAAGAAGGACAGAGCCT	50-55	60
	C2At1g24340R	ACGTCATCTTCAGACACCAATACTCCTTT		
	C2At2g43770F	AATGGAGTTGTCTGTTGTGCTTCC	50-55	60
	C2At2g43770R	TGGCGCATATCCCAAAGTTTAGC		
	C2At4g27700F	AGCTACTAAACCTGCTAAATCACC	50-55	60
	4g27700R1	CCAAAAAAGGCAAATGCAGC		
	C2At5g22620F	TCAAGAAGCTGCTGACTGCCTTGG	50-55	60
	C2At5g22620R	TCCCTGAGGGTCCATCAGGAAAATC		
	C2At5g56940F	TAGCAGATCTCCGAGAGACGGC	50-55	60
	C2At5g56940R	TACTTCAACCTATCAAAATTAAGACC		
	C2At5g46630F	TGGCGCCTTTGATGAAGATGC	50-55	60
	C2At5g46630R	AGATTTTGAGGGTAACCAAAGTCC		
	C2At4g10030F	TCGCGTCAGCTGCTCTTGATGTC	50-55	60
	C2At4g10030R	TTTCCTCCAAAGCTGTGACCAAC		
	C2At4g28530F	TGGAAATTGATCTTCACACTTGTGAGC	50-55	60
	C2At4g28530R	TCAAGTCGAAATTCATGCATGATCC		
	Ntd2g42750F	CATGCTCTTTTACAAGGCTAGAAAAT	50-55	60
	C2At2g42750R	ACTCTAGCTCTTCCAAAGTCTTCCTC		
	C2At3g58790F	ATCATCTAAGTTGGATTATGATCGTTG	50-55	60
	C2At3g58790R	TGGCTTGGCTGGTCCACTGAAATG		
	C2At2g38020F	TGCAGCTTTGCTTTATGATGCC	50-55	60
	C2At2g38020R	AAAGGCTTGGCCGTAGCTTGC		
	C2At1g78600F	ATGAGAAGGTTCATGCTGCTAATAAG	50-55	60
	C2At1g78600R	TAATGCCCGATCCTCAAGACAAAAG		
	C2At1g16900F	TGTCTCCCACGCTCGGACATTTTC	50-55	60
	C2At1g16900R	AAGAAGGGAACCGGTGCCACTCAC		
	C2At3g15380F	TTGTTTGGCGGCTATTGGGC	50-55	60
	C2At3g15380R	AGCATTACGATTCACAGATTTGATGG		
	C2At5g54310F	AGAAGCCTGGGGGTGCACATATC	50-55	60
	C2At5g54310R	AACTTGTTCAGGAAGCCATGTGTC		
	C2At1g55840F	TATATTCAATCACATATCCAAATGAATG	50-55	60
	C2At1g55840R	ACAATGTAATATGTATCTGTCTTCTCAGG		
	C2At5g56130F	ACATATAGCTGTTGGGAACAGGG	50-55	60
	Ntd5g56130R	CAGGCAATAGAAAGAGACGAAGAGTTT		
	C2At3g53580F	AGGGTTGACATGGGTGAGCCCATTC	50-55	60
	C2At3g53580R	TGTTCAAACTTTGGACCAATGTCTGC		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
COSII markers[4]	C2At3g52610F	ACTGCCGTTCAATTCTTGATCCAG	50-55	60
	C2At3g52610R	AGGAAATTCAGGTCTCCCATGTCTG		
	C2At4g03400F	TGCTATGGAATCACTTTATGCTACTATGG	50-55	60
	C2At4g03400R	ATTTCTGTCTTCCCTCTGTGGTTCCAG		
	C2At3g13235F	AGGTCGTATTCATGTTGCTCCAATC	50-55	60
	C2At3g13235R	AGCATATCAAGTCCAAAGAGAAATTCC		
	4g19003F1	AGGTCGTGCAGAAGGGATGGA	50-55	60
	C2At4g19003R	ACCCCAGAACGTATTTCTTCAACTGTC		
	5g11480F1	GAAAGACACAATGCATCAACCATTT	50-55	60
	Ntd5g11480R	ATGTTTTAACAATAGTCTCTAATCCAATA		
	C2At3g53920F	AAATTGCCAAGTTCACGGGTCTATCC	50-55	60
	C2At3g53920R	ATGACAGTTTCCTCTGGGCTTGTAATTG		
	C2At1g20830F	TCCGCACCGAGTTATGGTTGTCAAGTC	50-55	60
	C2At1g20830R	TGCAATTGCTAAACTAGCACCAGAATG		
	C2At1g32410F	TGTTAGTGTCTGGAGGGATTGTATTG	50-55	60
	C2At1g32410R	AGATTCGGTGTAGAGACTGGAAGTATC		
	C2At5g41350F	ACCTCCAATGCCAATGCCTTATG	50-55	60
	C2At5g41350R	ATGAAACTGATGCTCGCATTTTG		
	C2At4g23840F	AGGGGAGAGAGTTATGTGGATGCAG	50-55	60
	C2At4g23840R	ACCATCAGCTGTGACACCTGTTTCTG		
	C2At1g16590F	ATTCTTCTGTCAATGGACTTCTTCC	50-55	60
	C2At1g16590R	TTACATTTATCTTGAAAACAAATCTCTC		
	C2At4g14570F	AAAGCTTCTTGTAGTTAGAAATCCTGAA	50-55	60
	C2At4g14570R	TTTCCAGCTCCCACAGTCCTTATCTG		
	C2At3g16840F	ACGTCCATTGCAGCTTTACGCC	50-55	60
	C2At3g16840R	AGTGAATGACAGTTCGAACACCAGG		
	C2At3g24050F	ATGATGAAGATGACCTTCTCAACTTCTC	50-55	60
	3g24050R2	TCCCTACTTTGAGCCAGTCAATCC		
	C2At5g18580F	TGCCACATTGCCTCTGTATGTACAGAAC	50-55	60
	C2At5g18580R	ATGTCAATTCGGGCTTGAGTAAGTG		
	C2At4g15520F	TCCGTTTCCGCCACTTCCAC	50-55	60
	C2At4g15520R	ATTCAATGAAGCAGTACCACACCC		
	C2At3g23590F	TCGAAGCTGTCCCCTTTGTGGTTG	50-55	60
	C2At3g23590R	AGGCTTGGAACATCTACACCAGTGGC		
	C2At1g04190F	TCATTTCTTCGACAGTATGCTGAAGATT	50-55	60
	C2At1g04190R	ATTCCATCATTTTGTCCATGCTTCC		
	C2At1g28530F	ATTATGAAGATGTCTATACACTTCCCTAC	50-55	60
	C2At1g28530R	AGAGATTGCTTTTGACATAGAAATGCTT		
	C2At3g08760F	TCTCCAGAACGTTGTGTGTCAGAAGG	50-55	60
	C2At3g08760R	TCCTCATGTAGAAATGTAAGACCTTG		
	C2At1g70160F	ATTTGGACTTGTATGGTATTCTTGCTG	50-55	60
	C2At1g70160R	TGGAGGTAAAGAAGGACAATTCTCATTC		
	C2At1g60440F	TGCCCGGTCCCTCTTAAGGATG	50-55	60
	C2At1g60440R	TCCGCTTGAGCCCAAAACGAAG		
	C2At1g07040F	TGCTGATGATGATCAAATTTGCTATG	50-55	60
	C2At1g07040R	TCAACAAGAAGTTCGGATATAAATATTC		
	C2At1g14270F	TAGGCATTACTGGTGTACTTGCTCC	50-55	60
	C2At1g14270R	ACCGCTGCAAGCGGTGTAGG		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4] (Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix Table 1.1. Continued. List of primers used in this study.

Target	Primers	Sequence	°C[a]	S[b]
COSII markers[4]	C2At2g03510F	TGATACCCTGCTGAATTATGGGGTC	50-55	60
	C2At2g03510R	TGGTGCGCTCCTGTTCCATGTTCTC		
	C2At1g26520F	ACTGTGAAGCATAGTTTGGTTCAAGC	50-55	60
	C2At1g26520R	ACATCTGATTCCAGCTGATCATCCAAC		
	5g60990F2	AATTGCTGAACAGTTTGAAGCTTT	50-55	60
	5g60990R1	TGATCCAAGAGGCGACCAGG		
	C2At1g26670F	AAAATGGATCTCGAGGCAAGAAG	50-55	60
	C2At1g26670R	AAGCGTTCCCTTTGACTTGCAGAAG		
	C2At1g25580F	ATTCCAACTGTTAATGAGGATGATGG	50-55	60
	C2At1g25580R	AAGATAACAGGTTTTGTGCGACCAGTC		
	C2At2g18710F	TAAAAGCACAGCGGCATTCATC	50-55	60
	C2At2g18710R	AGAACCAAGAACTGATATCCGGC		
	C2At2g24090F	AGGAGGAGAGCTGGGAAGCAG	50-55	60
	C2At2g24090R	AATGCACCGATTACATTGTTG		
	C2At5g58200F	TATTGATGGGGGTGACCATGGTG	50-55	60
	C2At5g58200R	TCCAGGTTTCTGCTATCTTCTCCACTC		
	Ntd3g44600F	TTCTTCGACTTTCATTTCTGATTTTCG	50-55	60
	C2At3g44600R	AGATTCTATGTTTCTTGAAAGCACAGC		
	C2At3g53400F	AGGCAACTTGATAAGGTCTGCCAC	50-55	60
	C2At3g53400R	CCACCATTGATGAGCAAACACACC		
	C2At2g28490F	ACGGAGTATTCTCCATTGAAACACTCTG	50-55	60
	2g28490R1	GGAACCCTTCATCACACTACATCTTA		
	C2At5g59960F	TCCGATACTCATCAGCTCTTGTTC	50-55	60
	C2At5g59960R	ACGCCTTGTGTTTGTTTGGATGTC		
	C2At5g12200F	TCCTGACTTCCATTTTGCAGCAAG	50-55	60
	C2At5g12200R	TGCTCTGATTGGTGGGCTCATGAC		
	C2At2g24580F	TCATTGGAGCTGGTATAATGGGTAG	50-55	60
	C2At2g24580R	AGCCATTTCCAATAGTGAATTTTGACTC		
	C2At1g14790F	TGGAGAGAACACTAATATTCTCAAGG	50-55	60
	C2At1g14790R	ATGTCTTGATCCCAGCAAACAAAG		
	C2At1g30360F	GCCTTCCACAGAATGCAACTTTCTTC	50-55	60
	C2At1g30360R	AAGCCTCCTTTATCTCAGCTTCAG		
	C2At1g22860F	AGAGACTGTCACCAGATATGCCCC	50-55	60
	C2At1g22860R	ATCTTTCCTCAAACCTTGCCAAGC		
	C2At1g67700F	AAGAGGAAATTGTTAGTGGTTGAAGC	50-55	60
	C2At1g67700R	ACTGCTGCGAGATTCCTAGCTAGAG		
	C2At5g38530F	TCGCACAAGCCCAACTCTGCTG	50-55	60
	C2At5g38530R	ACTCCTCACCGATAACGGTCTGATG		
	C2At4g16580F	TGTTACCTGCCTCATCCTGATAAAG	50-55	60
	C2At4g16580R	ATTTTGAAGACCTCTCCAGAACTTGG		
	1g76150F1	AGTATTGTCGGATTGCATGATAAAG	50-55	60
	1g76150R1	GGATCTGAATGCAATGGATTGTAAT		
	C2At1g17410F	ATTGGCTATAATAAAGCCAGATGG	50-55	60
	C2At1g17410R	AATTAATGCACGCCAATCAGC		
	3g17000F1	TGTGCACATGTAGATGTATGTAGGGTG	50-55	60
	C2At3g17000R	TCGTGGACCCCTGATTGCAAATTG		
	C2At1g48300F	AAGAAGATGAAATTACTTAAGGGTTTG	50-55	60
	C2At1g48300R	TTTAGTGTTGCATTCTCAAGTGCTCG		

[a]Annealing temperature, [b]Elongation time (seconds), [1]Annealing temperature for Phusion, [2](Szczesny et al., 2010), [3](Exposito-Rodriguez et al., 2008)[4](Wu et al., 2010), *2A* = *OR2A, 2B* = *OR2B, 2C* = *OR2C*, N/A = not applicable.

Appendix 2: Data linked to Chapter 3

Appendix Table 2.1. Genotypes of recombinants identified in a PVX:*Ecp2* screen for survivors in the F_2 from a cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0.

	Marker					
Recombinant	**TG236**	**SNPE**	**CT116**	**SNPN**	**SNPQ**	**TG184**
1E06	0	0/?	0	0	0	H
4E10	0	0	0	0/?	0	?
1A08	0	0	0	0	0	H
1F07	0	0	0	0	0	H
1F10	0	?	0	0	0	H
1H04	0	0	0	0	0	H
2B08	0	0	0	0	0	H
2C10	0	0		0	H	H
2C09	0	0	0	0	0	H
2D03	0	0	0	0	0	H
2D05	0	0	0	0	0	H
2D09	0	0	0	0	0	H
2F01	0	0	0	0	0	H
2F02	0	0	0	0	0	H
2G 08	0	0		0	0	H
2H 02	0	0/?		0	0	H
2H03	0	0	0	0	0	H
2H11	0	0	0	0	0	H
3E03	0	0	0	0	0	H
3E 05	0	?	0	0	0	H
3G03	0	0	0	0	0	H
3B05	0	0	0	0	0	H
3B02	0	0	0	0	0	H
3C10	0	0	0	0	0	H
3D09	0	0/?	0	0	0	H
3G05	0	0	0	0	0	H
4E07	0	0/?	0	0	0	H
4E08	0	?	0	0	0	H
4A02	0	0	0	0	0	H
4A06	0	0	0	0	0	H
4D05	0	0	0	0	0	H
4G07	0	2/H	0	0	0	H
5D01	0	0	0	0	0	H
5H 05	0	0	0	0	0	H
3G10	0	0	0	0	0	H
2C12	0	0	0	0	0	H/?
2D04	0	0	0	0	0	H/?
2E11	2/?	0	0	0	0	0
2F10	2/?	0	0	0	0	0
1F03	2/H	0		0	0	0
1G11	2/H	0	0	0	0	?
1C02	H	0	0	0	0	0
1G08	H	0	0	0	0	0
1G10	H	0	0	0	0	0
2E01	H	0	0	0	0	0
2C11	H	0	0	0	0	0
2D12	H	0	0	0	0	0
2E06	H	0		0	0	0
2F05	H	0		0	0	O
2F04	H	0				0

0 = Cf0 allele, 2 = CfEcp2 allele, H = heterozygous for Cf0 and CfEcp2 alleles, blank = not tested, ? = unknown, PVX = Potato Virus X .

Appendix Table 2.1. Continued. Genotypes of recombinants identified in a PVX:*Ecp2* screen for survivors in the F_2 from a cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0.

	Marker					
Recombinant	**TG236**	**SNPE**	**CT116**	**SNPN**	**SNPQ**	**TG184**
2H05	H	0	0	0	0	H
3E04	H	0	0	0	0	0
3A01	H	0	0	0	0	0
3C06	H	0	0	0	0	0
3D12	H	0	0	0	0	0
3H05	H	0	0	0	0	H
3H01	H	0	0	0	0	0
4B06	H	0	0	0	0	0
4C08	H	0	0	0	0	0
4D07	H	0	0	0	0	?
4G05	H	0	0	0	0	0
4H09	H	H/2	0	0	0	0
4H12	H	0	0	0	0	0
5D03	H	0	0	0	0	0

0 = Cf0 allele, 2 = CfEcp2 allele, H = heterozygous for Cf0 and CfEcp2 alleles, blank = not tested, ? = unknown, PVX = Potato Virus X.

Appendix Table 2.2. Genotypes of recombinants identified in a genotype screen in the F_2 from a cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0.

	Marker				
Recombinant	**TG236**	**SNP-E**	**CT116**	**SNP-N**	**TG184**
6A04	0/H	0		0	0
6E09	0	0		0	H
6F01	H	0		0	0
6G01	H	0		0	0
7B04	0	0		0	H
7E02	0	0		0	H
8E01	0	0		0	H
9A11	0	0		0	H
9E03	H	0		0	0
9H11	0	0		0	H
10H06	H	0		0	0
11C02	0	0		0	H
11C12	0	0		0	H
11D02	0	0		0	H
11E07	0	0		0	H
11F02	0	0		0	H
13D06	0	0		0	H
13D08	0	0		H	H
13E11	0	0		0	H
13G06	0	0		0	H
14C01	0	0		0	2
14C03	H	0		0	0
14E02	H	0		0	0
6D03	2	2		2	0
6E03	H	2		2	2
6E06	2	2		2	H
6E07	2	2		2	0
6E08	2	2		2	0
6F08	2	2		2	0
6G07	2	2		2	0
7C01	2	2		2	H
7H05	2	2		2	H
8A04	H	2		2	2
8G06	2	2		2	0
9D04	2	2		2	H
10A04	2	2		2	H
10B07	2	2		2	H
10F12	H	2		2	2
11A03	H	2		2	2
11C01	2	2		2	H
11C04	2	2		2	H
11C10	H	2		2	2
11D09	H	2		2	2
11G08	2	2		2	H
13A08	H	2		0	2
13B02	H	2		0	2
13B10	H	2		2	2
13D01	2	2		2	H
10B02	H	?		?	0
6A11	2	2/H		H	H
10E02	H	2/H		?	2

0= Cf0 allele, 2 = CfEcp2 allele, H = heterozygous for Cf0 and CfEcp2 allele, blank = not tested, ? = unknown allele. Bold red = those recombinants taken forward for further analysis.

Appendix Table 2.2. Continued. Genotypes of recombinants identified in a genotype screen in the F_2 from a cross between *S. pimpinellifolium* CfEcp2 and *S. lycopersicum* Cf0.

	Marker				
Recombinant	**TG236**	**SNP-E**	**CT116**	**SNP-N**	**TG184**
13A02	H	2/H		0	2
14F02	2	2/H		H	H
14F08	H	2/H		H	2
7C11	0	H		H	H
7F02	0	H		H	H
7H09	0	H		H	H
11A05	0	H		H	H
11G01 (**47**)	0	H		H	H
12C03 (**57**)	0	H		H	2
13B03 (**50**)	0	H		H	H
13D04 (**52**)	0	H		H	H
13F09 (**55**)	0	H		H	2
14G01 (**63**)	0	H		H	H
6C08	2	H		H	H
6F11	2	H		H	H
7B02	2	H		H	0
7D09	2	H		H	H
7E01	2	H		H	H
7F05	2	H		H	H
8G03	2	H		H	H
8H03	2	H		H	H
9B01	2	H		H	H
10D06	2	H		H	H
11B03	2	H		H	H
11G02	2	H		H	H
13E05 (**58**)	2	H		H	0
14A05	2	H		H	H
14D05	2/?	H		H	H
6C07	H	H		H	2
6H10	H	H		H	2
7G04	H	H		H	0
9E07	H	H		H	0
11C11	H	H		H	2
11E01	H	H		2	2
11E05	H	H		H	0
11F10	H	H		H	0
13B04 (**51**)	H	H		H	0
13E07 (**54**)	H	H		H	0
13H03	H	H		H	2
14B02	H	H		H	2
7E08	H	H		H	2
Cf0	0	0		0	
CfEcp2	2	2		2	
F_1	H	H		H	

0= Cf0 allele, 2 = CfEcp2 allele, H = heterozygous for Cf0 and CfEcp2 allele, blank = not tested, ? = unknown allele. Bold red = those recombinants taken forward for further analysis.

Appendix 3: Data linked to Chapter 4

Appendix Table 3.1. Response of *S. lycopersicum* 35S:*2A* transgenic lines to PVX:*Ecp2.*

Plant line	Phenotype 14 or 21 d.p.i. HR	NR	Total plants inoculated
Controls			
S. lycopersicum Cf0	0	18	18
S. pimpinellifolium CfEcp2	13[1]	0	13
S. pimpinellifolium 1178	0	15	15
S. pimpinellifolium 1179	0	13	13
T_2 families			
S. lycopersicum 35S:*2A*p1	0	2	2
S. lycopersicum 35S:*2A*p2	1[2]	8	9
S. lycopersicum 35S:*2A*p3	1[2]	1	2
S. lycopersicum 35S:*2A*p5	0	7	7
S. lycopersicum 35S:*2A*p13	0	1	1
S. lycopersicum 35S:*2A*p14	0	3	3
S. lycopersicum 35S:*2A*p16	0	4	4
S. lycopersicum 35S:*2A*p18	0	6	6
S. lycopersicum 35S:*2A*p22	0	4	4
S. lycopersicum 35S:*2A*p24	3[2]	4	7
S. lycopersicum 35S:*2A*p28	0	4	4
T_3 families			
S. lycopersicum 35S:*2A*p2.1	0	8	8
S. lycopersicum 35S:*2A*p2.2	1[2]	5	6
S. lycopersicum 35S:*2A*p2.3	0	7	7
S. lycopersicum 35S:*2A*p2.4	2[2]	5	7
S. lycopersicum 35S:*2A*p2.5	2[2]	2	4
S. lycopersicum 35S:*2A*p2.6	0	8	8
S. lycopersicum 35S:*2A*p3.1	7[2]	1	9
S. lycopersicum 35S:*2A*p3.2	0	8	8
S. lycopersicum 35S:*2A*p3.3	0	5	5
S. lycopersicum 35S:*2A*p3.5	2[2]	3	5
S. lycopersicum 35S:*2A*p3.6	3[2]	4	7
S. lycopersicum 35S:*2A*p3.8	2[2]	6	8
S. lycopersicum 35S:*2A*p3.9	5[2]	1	6
S. lycopersicum 35S:*2A*p3.10	0	8	8
S. lycopersicum 35S:*2A*p3.11	0	8	8
S. lycopersicum 35S:*2A*p4.1	4[2]	0	4
S. lycopersicum 35S:*2A*p4.2	0	2	2
S. lycopersicum 35S:*2A*p5.1	0	8	8
S. lycopersicum 35S:*2A*p5.2	0	7	7
S. lycopersicum 35S:*2A*p5.3	0	8	8
S. lycopersicum 35S:*2A*p18.1	0	4	4
S. lycopersicum 35S:*2A*p24.1	0	2	2
S. lycopersicum 35S:*2A*p24.2	2[2]	4	6
S. lycopersicum 35S:*2A*p24.3	3[2]	4	7
S. lycopersicum 35S:*2A*p24.4	4[2]	3	8
S. lycopersicum 35S:*2A*p24.5	1[2]	1	2

NR = no response, [1]=HR^+= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [2]= HR^0 = HR manifesting as patchy necrosis, d.p.i = days post inoculation, PVX = Potato Virus X, *2A* = *OR2A*. Green = cDNA analysed for presence of transcripts of *2A*, *Ecp2* and *PVX.*

Appendix Table. 3.2. Response of *S. lycopersicum* 35S:*2A* T_2 families grown on media supplemented with kanamycin.

	Phenotype 16 d.p.s.		
	300 mg/L kanamycin		Percentage
Plant line	Resistant	Sensitive	germination (%)
S. pimpinellifolium CfEcp2	0	4	20
S. pimpinellifolium 1179	4	0	20
S. Lycopersicum 35S:*2A*p1	0	5	25
S. Lycopersicum 35S:*2A*p2	6	2	40
S. Lycopersicum 35S:*2A*p3	3	1	20
S. Lycopersicum 35S:*2A*p5	3	3	30
S. Lycopersicum 35S:*2A*p14	0	3	15
S. Lycopersicum 35S:*2A*p16	0	1	5
S. Lycopersicum 35S:*2A*p18	1	1	10
S. Lycopersicum 35S:*2A*p22	0	1	5

d.p.s = days post sowing, *2A* = *OR2A.*

Appendix Table 3.3. Phenotypes of progeny from crosses between *S. lycopersicon* Cf0 stably expressing 35S:2A x *S. lycopersicon* Cf0 35S:*Ecp2.*

Plant line	Phenotype				Total
	SLP	A or D or St	NG	N	
Controls					
S. lycopersicon Cf0	0	0	13	24	37
[1]*S. lycopersicon* Cf0 x *S. lycopersicon* 35S:*Ecp2*	0	0	19	23	42
S. lycopersicon Cf0 x *S. pimpinellifolium CfEcp2*	0	0	0	12	12
S. pimpinellifolium CfEcp2	0	0	16	21	37
[1]*S. pimpinellifolium* CfEcp2 x *S. lycopersicon* 35S:*Ecp2*	15	0	5	1	21
S. lycopersicon 35S:*Ecp2*	0	0	2	10	12
T_1 crosses					
S. lycopersicon 35S:*Ecp2 x S. lycopersicon* 35S:*2A*p1	3	8	12	3	26
[1]*S. lycopersicon* 35S:*2A*p2 x *S. lycopersicon* 35S:*Ecp2*	2	13	14	11	40
[1]*S. lycopersicon* 35S:*2A*p3 x *S. lycopersicon* 35S:*Ecp2*	14	0	26	0	40
S. lycopersicon 35S:*2A*p4 x *S. lycopersicon* 35S:*Ecp2*	10	4	8	8	30
[1]*S. lycopersicon* 35S:*2A*p5 x *S. lycopersicon* 35S:*Ecp2*	4	7	15	24	50
S. lycopersicon 35S:*2A*p10 x *S. lycopersicon* 35S:*Ecp2*	0	1	13	16	30
S. lycopersicon 35S:*2A*p14 x *S. lycopersicon* 35S:*Ecp2*	0	3	14	8	25
S. lycopersicon 35S:*2A*p16 x *S. lycopersicon* 35S:*Ecp2*	7	3	21	9	40
S. lycopersicon 35S:*2A*p18 x *S. lycopersicon* 35S:*Ecp2*	1	2	21	16	40
S. lycopersicon 35S:*Ecp2 x S. lycopersicon* 35S:*2A*p24	6	0	6	6	18
S. lycopersicon 35S:*Ecp2 x S. lycopersicon* 35S:*2A*p28	2	0	10	1	13
T_2 selfs					
S. lycopersicon 35S:*2A*p3.8	0	0	7	17	24
S. lycopersicon 35S:*2A*p24.4	0	0	11	13	24
T_2 crosses					
S. lycopersicon 35S:*2A*p3.8 x *S. lycopersicon* 35S:*Ecp2*	8	0	0	0	8
[1]*S. lycopersicon* 35S:*2A*p24.4 x *S. lycopersicon* 35S:*Ecp2*	6	1	15	18	40
T_3 selfs					
S. lycopersicon 35S:*2A*p24.4.2D	0	0	1	11	12
S. lycopersicon 35S:*2A*p24.4.4D	0	0	9	15	24
S. lycopersicon 35S:*2A*p24.5-5D	0	0	16	8	24
T_3 crosses					
S. lycopersicon 35S:*2A*p24.4.2D x *S. lycopersicon* Cf0	0	0	3	21	24
S. lycopersicon 35S:*2A*p24.4.2D x *S. lycopersicon* 35S:*Ecp2*	13	0	11	0	24
S. lycopersicon 35S:*2A*p24.4.4D x *S. lycopersicon* Cf0	0	0	20	4	24
S. lycopersicon 35S:*2A*p24.4.4D x *S. lycopersicon* 35S:*Ecp2*	4	0	6	16	24
S. lycopersicon 35S:*2A*p24.5.5D x *S. lycopersicon* Cf0	0	0	14	10	24
S. lycopersicon 35S:*2A*p24.5.5D x *S. lycopersicon* 35S:*Ecp2*	0	0	10	14	24

[1]Crosses in either direction used, and due to no difference in results, the results were combined.
Phenotypes scored 15 – 26 says post sowing.
SLP = Seedling lethal phenotype, A = Anthocyanin accumulation, D = Developmentally-different, NG = No germination, N = Normal, Total = Total plants sown for scoring, *2A* = *OR2A.*

Appendix Table 3.4. Response of *S. pimpinellifolium* 1179 carrying 35S:2A to PVX:*Ecp2*

Plant ID	PVX:*Ecp2*			No PVX:*Ecp2*		
	HR	NR	Total	HR	NR	Total[1]
S. lycopersicon Cf0	0	6	6	0	4	4
S. pimpinellifolium CfEcp2	6[2]	0	6	0	6	6
S. pimpinellifolium CfEcp2 x *S. lycopersicon* Cf0	12[2]	0	12	0	9	9
S. pimpinellifolium 1179p15	0	6	6	0	5	5
S. pimpinellifolium 1179p19	0	6	6	0	4	4
S. pimpinellifolium 1179p31	0	5	5	0	4	4
S. pimpinellifolium 1179p33	0	6	6	0	6	6
S. pimpinellifolium CfEcp2 x *S. pimpinellifolium* 1179p15	12[2]	0	12	0	11	11
S. lycopersicon 35S:*2A*p3.8.1	0	5	5	0	2	2
S. lycopersicon 35S:*2A*p3.8.1 x *S. lycopersicon* Cf0	0	9	9	0	5	5
S. lycopersicon 35S:*2A*3.8.1 x *S. pimpinellifolium* 1179p19	2[3]	20	22	0	9	9
S. lycopersicon 35S:*2A*p24.4.2	2[3]	10	12	0	6	6
S. pimpinellifolium 1179p31x *S. lycopersicon* 35S:*2A*p24.4.2	1[3]	11	12	0	9	9
S. lycopersicon 35S:*2A*p24.4.3	4[3]	6	10	0	3	3
S. lycopersicon 35S:*2A*p24.4.3 x *S. pimpinellifolium* 1179p15	3[3]	9	12	0	9	9
S. lycopersicon 35S:*2A*p24.4.4	0	12	12	0	6	6
S. lycopersicon 35S:*2A*p24.4.4 x *S. pimpinellifolium* 1179p19	1[3]	11	12	0	11	11
S. lycopersicon 35S:*2A*p24.5.3	0	5	5	0	0	0
S. lycopersicon Cf0 x *S. lycopersicon* 35S:*2A*p24.5.3	0	12	12	0	11	11
S. lycopersicon 35S:*2A*p24.5.3 x *S. pimpinellifolium* 1179p19	0	12	12	0	11	11
S. lycopersicon 35S:*2A*p24.5.4	0	12	12	0	9	9
S. lycopersicon 35S:*2A*p24.5.4 x *S. pimpinellifolium* 1179p15	0	9	9	0	2	2

Total[1] = Total number of leaf sections inoculated, NR = no response, [2]=HR[+]= hypersensitive response (HR) manifesting as confluent necrosis associated with tissue collapse and death of the whole plant, [3]= HR[0] = HR manifesting as patchy necrosis, PVX = Potato Virus X, *2A* = *OR2A*.

Appendix 4: Data linked to Chapter 5

Appendix Table 4.1. Response of *N. paniculata* accessions infiltrated with Ecp2 homologs from important crop species.

		Nicotiana paniculata accession											
		TW99				TW102				F1			
Species of fungus	Effector	++	+	0	total	++	+	0	total	++	+	0	total
Septoria musiva	Ecp2	6	2	5	13	0	0	11	11	1	1	7	9
Dothiostroma septosporum	Ecp2	13	0	0	13	0	0	11	11	9	0	0	9
Zymoseptoria tritici	Ecp2	0	2	12	14	0	0	7	7	0	1	2	3
verticillium dahliae	Ecp2	7	2	4	13	0	0	11	11	0	3	6	9
Fusarium oxysporium	Ecp2	11	0	2	13	0	0	11	11	1	5	3	9
Fusarium graminearium	Ecp2	7	0	6	13	0	0	11	11	1	0	8	9
Magnaporthe grisea	Ecp2	4	0	9	13	0	0	11	11	1	0	8	9
Mycosapharella fijiensis	Ecp2	11	1	1	13	0	0	11	11	2	4	3	9
Cladosporium fulvum	Ecp2	13	0	0	13	0	0	11	11	9	0	0	9
Cadosporium fulvum	Avr4	0	1	18	19	0	0	11	11	0	0	10	10
[1]*Cladosporium fulvum*	Ecp2	29	0	0	29	0	0	22	22	20	0	0	20

[1]Construct from Soumpourou et al. (2007), total = total number of leaf sections scored, + = partial hypersensitive response (HR) manifested as partial necrosis, ++ = HR with confluent death of inoculated section. Blue = high percentage of leaves showing response in both TW99 and F_1.

Appendix Table 4.2. Orientation of markers in the linkage groups of *N. paniculata* compared to the genetic maps of *N. tomentosiformis* and *N. acuminate.*

[1]Linkage group in Pn	Marker orientation		
	[2]Acm	[3]Pn	[4]Tmf
1	09070	09070	13450
	56500	56500	10240
	-	10240	56500
	-	13450	09070
	08030	08030	08030
	x(6)	63200	63200
2.1	65900	34090	45950
	55250	67370	65900
	45950	55250	34090
	67370	65900	67370
2.2	34090	45950	55250
3.1	-	33970	33970
	67440	64770	64770
	-	56040	56040
3.2	79600	79600	79600
	48610	48610	74470
	74470	74470	48610
	x(4)	39580	x(4)
4.1	17040	17040	-
13	39580	39580	39580
4.2	20575	20575	20575

[1]Linkage group number of *N. paniculata* markers. Direction of linkage group inferred from *N. tomentosiformis* and *N. acuminate* genetic maps and may be opposite to that shown here. [2] Acm= order of markers in *N. acuminate* genetic map. [3]Pn= order of markers in *N. paniculata* genetic map. [4]Tmf= prder of markers in *N. tomentosiformis* genetic map. The chromosome first marker in the genetic map of *N. paniculta* is coloured blue followed by, turquoise, purple, green, brown, yellow, pink and magenta. The same markers are then coloured accordingly in *N. acuminate* and *N. tomentosiformis* to show inversions and movement of marker positions.

Appendix Table 4.2. Continued. Orientation of markers in the linkage groups of *N. paniculata* compared to the genetic maps of *N. tomentosiformis* and *N. acuminate.*

[1]Linkage group in Pn	Marker orientation		
	[2]Acm	[3]Pn	[4]Tmf
5.1	28880	28880	28880
	37360	37360	37360
	-	09880	09880
	x(6)	52220	52220
	-	14520	52640
	14520	52640	14520
6.1	-	19690	19690
6.2	18640	18640	x(1)
	34215	43360	-
	43360	34215	34215
	03150	03150	-
	16870	16870	16870
7.1	26680	26680	26680
	-	58790	58790
	78620	78620	78620
	-	16900	16900
7.2	15290	15290	15290
	-	15380	15380
	-	54310	54310
	-	04970	04970
7.3	37025	37025	37025
	-	08760	x(10)
8.1	20860	20860	20860

[1]Linkage group number of *N. paniculata* markers. Direction of linkage group inferred from *N. tomentosiformis* and *N. acuminate* genetic maps and may be opposite to that shown here. [2] Acm= order of markers in *N. acuminate* genetic map. [3]Pn= order of markers in *N. paniculata* genetic map. [4]Tmf= prder of markers in *N. tomentosiformis* genetic map. The chromosome first marker in the genetic map of *N. paniculta* is coloured blue followed by, turquoise, purple, green, brown, yellow, pink and magenta. The same markers are then coloured accordingly in *N. acuminate* and *N. tomentosiformis* to show inversions and movement of marker positions.

Appendix Table 4.2. Continued. Orientation of markers in the linkage groups of *N. paniculata* compared to the genetic maps of *N. tomentosiformis* and *N. acuminate.*

[1]Linkage group in Pn	Marker orientation		
	[2]Acm	[3]Pn	[4]Tmf
8.2	11450	11450	11450
	-	53920	53920
	-	20830	20830
	-	32410	32410
	-	41350	41350
9.1	06360	06360	06360
	24160	24160	24160
	-	24050	24050
	18580	18580	18580
	-	15520	15520
	-	04190	04190
9.2	61150	61150	61150
	-	28530	28530
10.1	51840	51840	51840
10.2	x(12)	24750	24750
		60440	70160
		70160	60440
		26670	07040
	x(12)	26520	03510
		03510	26520
	x(12)	25580	26670
		07040	25580
11.1	24270	24270	-
	06430	06430	06430
	56050	56050	56050

[1]Linkage group number of *N. paniculata* markers. Direction of linkage group inferred from *N. tomentosiformis* and *N. acuminate* genetic maps and may be opposite to that shown here. [2] Acm= order of markers in *N. acuminate* genetic map. [3]Pn= order of markers in *N. paniculata* genetic map. [4]Tmf= prder of markers in *N. tomentosiformis* genetic map. The chromosome first marker in the genetic map of *N. paniculta* is coloured blue followed by, turquoise, purple, green, brown, yellow, pink and magenta. The same markers are then coloured accordingly in *N. acuminate* and *N. tomentosiformis* to show inversions and movement of marker positions.

Appendix Table 4.2. Continued. Orientation of markers in the linkage groups of *N. paniculata* compared to the genetic maps of *N. tomentosiformis* and *N. acuminate.*

[1]Linkage group in Pn	Marker orientation		
	[2]Acm	[3]Pn	[4]Tmf
11.2	28250	28250	-
	-	59960	59960
	22660	12685	22660
	12685	22660	12685
12.1	x(10)	31410	x(10)
	-	14790	14790
	-	67700	67700
	-	38530	38530
	x(6)	25740	-
	-	76150	76150
12.2	-	17000	17000
	53000	53000	53000

[1]Linkage group number of *N. paniculata* markers. Direction of linkage group inferred from *N. tomentosiformis* and *N. acuminate* genetic maps and may be opposite to that shown here. [2] Acm= order of markers in *N. acuminate* genetic map. [3]Pn= order of markers in *N. paniculata* genetic map. [4]Tmf= prder of markers in *N. tomentosiformis* genetic map. The chromosome first marker in the genetic map of *N. paniculta* is coloured blue followed by, turquoise, purple, green, brown, yellow, pink and magenta. The same markers are then coloured accordingly in *N. acuminate* and *N. tomentosiformis* to show inversions and movement of marker positions.