# Plant pathogen effector proteins and their host targets: Functional studies and manipulation to generate enhanced resistance

Benjamin Kent Hall

Thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

2016

Banfield Laboratory

Department of Biological Chemistry

John Innes Centre

Norwich Research Park

# Abstract

As global population expands, so does its demand for food. This will require vast amounts of plant-derived calories. This increased pressure means it is vital that we do more with less; less land, less water, less chemical input and less labour. One major threat to crop production is posed by plant pests and pathogens. In particular, filamentous plant pathogens – oomycetes and fungi – are among the most devastating organisms known to agriculture. With their rapidly evolving genomes which specialise in breaking plant immune systems and chemical control methods, control of these pathogens is becoming extremely problematic. Especially as they often encounter crops in monoculture. An often overlooked potential strategy for developing pathogen resistance is manipulation of the host targets of plant pathogen effector proteins. In order to manipulate such targets to condition enhanced resistance, one must first develop a sound understanding of the interaction between effector and target, the target's function *in planta* and which processes are being perturbed by the interaction.

A previously discovered interaction between a *P. infestans* RXLR effector and a host MAPK known to be a positive regulator of immunity, where the effector acts to perturb positive regulation of immunity, is used here to investigate the possibility of manipulating a host target for enhanced immunity. Effector-insensitive variants were produced and characterised, prior to expression in CRISPR-edited tomato plants. We tentatively suggest that at least one variant may condition enhanced resistance. Another interactor of the same effector protein was investigated for its role in immunity and evidence suggesting it may function as a helper of the effector is presented.

Additionally, the expression, purification and crystallisation of native and heavy atom derivatives of a candidate effector protein from *B. graminis* f.sp. *hordeii* is demonstrated.

# Acknowledgements

This has been the single most difficult, yet most worthwhile experience of my life. Thankyou first – and foremost – to Professor Mark Banfield, for taking me on, guiding me and helping me produce a PhD thesis. I'm both proud and fortunate to have had you as a supervisor. Thanks also to Dr. Silke Robatzek for her valuable support, for helping me keep focused on the key questions and key experiments, and for always telling me when something is not good enough. Not forgetting Dr. Hannah Kuhn for stepping in during Silke's leave.

Banfield lab members past and present, you've all helped me in some small (or big) way but special mentions must go to Stuart (for setting this all up with an excellent thesis and paper) and to Richard, Abbas, Marina, Mim, John and Lennart (for their invaluable help, advice and suggestions during my four years). I'd also like to thank members of the Kamoun and Jones labs at TSL, especially Joe, Angela, Vlad and Kamil. To all the brilliant scientists and support staff who drive the work done at JIC and TSL, I am so grateful. In particular; Dave Lawson and Clare Stevenson (Crystallography), Gerhard Saalbach (Proteomics), Nicola Patron and Mark Youles (SynBio) and Matthew Smoker and Jodie Pike (Plant Transformation). Not to mention the media kitchen and the horticultural services team.

Daniel, Tom, Jon, Alex, Matt, Izzy and Lizzie – thanks for sharing lunchtimes, 'Pub Thursdays', kayaking adventures and board game nights. You've made the last four years a blast. Friends and family from Sunderland, thanks for keeping me grounded and reminding me of where I come from.

I feel exceptionally lucky to have the parents I do, and the wonderful step-families that life has brought to me. Your love and support has been incredibly helpful during this process, particularly through the writing up period.

Elizabeth, thank you for sharing in this challenge with me, for teaching me to believe in myself and for helping me to understand that it's ok to 'switch off'. I look forward to entering the next stage of our lives, together. Hopefully with a lot less stress!

# List of abbreviations

The present thesis uses one letter code for nucleic acids (ATCGU) and one and three letter codes for amino acids (e.g. R or Arg). Standard SI units are used, unless explicitly stated.

| Abbreviation | Meaning |
|---|---|
| **3-AT** | 3-amino-1, 2, 4-triazole |
| **$A_{280}$** | Absorbance at 280nm |
| **AP** | Alkaline phosphatase |
| **ASR** | Asian Soybean Rust |
| **ATG8** | Autophagy-related protein 8 |
| **ATP** | Adenosine Triphosphate |
| **AVR** | Avirulence protein |
| **BAK1** | BRI1-associated kinase 1 |
| **BCIP** | 5-bromo-4-chloro-3-indolyl phosphate |
| **BEC** | *Blumeria* effector candidate |
| **BiFC** | Bimolecular fluorescence complementation |
| **Bis-Tris** | 2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1, 3-diol |
| **BLAST** | Basic local alignment search tool |
| **Cas9** | CRISPR-associated protein 9 |
| **CD** | Circular Dichroism |
| **CDS** | Coding Sequence |
| **CERK1** | Chitin elicitor receptor kinase 1 |
| **CID** | Collision induced dissociation |
| **CRISPR** | Clustered regularly interperspaced short palindromic repeats |
| **CRN** | Crinkler effector |
| **CSEP** | Candidate secreted effector protein |
| **CTAB** | Cetyl trimethylammonium bromide |
| **DAMP** | Damage Associated Molecular Pattern |

| | |
|---|---|
| **DCL** | Dicer-Like |
| **DLS** | Diamond light source |
| **DMF** | Dimethyl formamide |
| **DMSO** | Dimethyl sulfphoxide |
| **dNTP** | Deoxyribose nucleotide triphosphate |
| **DSB** | Double stranded break |
| **DTT** | Dithiothreitol |
| **ECL** | Enhanced chemiluminesence |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **EFR** | Ef-Tu receptor |
| **Ef-Tu** | Bacterial elongation factor 'Tu' |
| **EHM** | Extrahaustorial membrane |
| **EST** | Expressed Sequence Tag |
| **ETI** | Effector Triggered Immunity |
| **ETS** | Effector Triggered Susceptibility |
| **flg22** | Flagellin 22 |
| **FLS2** | Flagellin Sensitive 2 |
| **HA** | Human influenza hemagglutinin tag |
| **HCD** | Higher-energy collision dissociation |
| **HDR** | Homology-directed repair |
| **HEPES** | 2-[4-(2-hydroxyethyl)piperazin-1-y]ethanesulfonic acid |
| **HMA** | Heavy metal-associated |
| **HR** | Hypersensitive Response |
| **HRP** | Horseradish peroxidase |
| **HRV** | Human rhinovirus |
| **Hsp** | Heat shock protein |
| **IMAC** | Immobilised metal ion affinity chromatography |
| **IP** | Invasion Pattern |
| **IPR** | Invasion Pattern Receptor |
| **IPTG** | Isopropyl β-D-1-thiogalactopyranoside |
| **IPTR** | Invasion Pattern Triggered Immunity |
| **JCSG** | Joint Centre for Structural Genomics screen |

| | |
|---|---|
| **KISS** | Keep it simple screen |
| **LB** | Luria-bertani broth |
| **LC-MS/MS** | Tandem liquid chromatography mass spectrometry |
| **LysM** | Lysine-rich motif |
| **MAD** | Multi-wavelength anomalous dispersion |
| **MAMP** | Microbe Associated Molecular Pattern |
| **MAPK** | Mitogen-activated Protein Kinase |
| **MEKK** | MAPK/ERK kinase |
| **MES** | 2-(N-morpholino)ethanesulfonic acid |
| **MQ** | Milli-Q |
| **MR** | Molecular replacement |
| **MS** | Mass Spectrometry |
| **MW** | Molecular Weight |
| **NBT** | Nitro blue terazolium chloride |
| **NHEJ** | Non-homologous end joining |
| **$Ni^{2+}$-NTA** | Nickel- nitrilotriacetic acid column |
| **NLP** | Nep1-like protein |
| **NLR** | Nucleotide-binding Leucine-rich repeat protein |
| **$OD_{600}$** | Optical density at 600nm |
| **OG** | Oligogalacturonic acid |
| **ORF** | Open Reading Frame |
| **PAGE** | Polyacrylamide gel electrophoresis |
| **PAMP** | Pathogen Associated Molecular Pattern |
| **PCD** | Programmed Cell Death |
| **PCR** | Polymerase chain reaction |
| **PCR-RE** | PCR-Restriction enzyme assay |
| **PEG** | Polyethylene glycol |
| **PexRD** | Predicted Extracellular with RXLR dEER motif protein |
| **PI3P** | Phosphatidylinositol 3-phosphate |
| **PP1c** | Protein phosphatase 1c |
| **PRR** | Pattern Recognition Receptor |
| **PTI** | PAMP Triggered Immunity |

| | |
|---|---|
| **PVDF** | Polyvinylidene fluoride |
| **R protein** | Resistance protein |
| **RIN4** | RPM1-interacting protein 4 |
| **ROS** | Reactive oxygen species |
| **RPM1** | Resistance to *Pseudomonas synringae pv. maculicola* 1 |
| **RPS2** | Resistance to *P. syringae* 2 |
| **RSA** | Rye sucrose agar |
| **RT-PCR** | Reverse transcriptase-PCR |
| **SA** | Salicylic acid |
| **SAD** | Single-wavelength anomalous dispersion |
| **SAR** | Systemic Acquired Resistance |
| **SC -LW** | Synthetic complete medium lacking Leucine and Tryptophan |
| **SC –LWH** | Synthetic complete medium lacking Leucine, Tryptophan and Histidine |
| **SC -LWU** | Synthetic complete medium lacking Leucine, Tryptophan and Uracil |
| **SDS** | Sodium dodecyl sulphate |
| **SEC** | Size exclusion chromatography |
| **SeMet** | L-selenomethionine |
| **SERK3** | Somatic embryogenesis receptor kinase 3 |
| **SFI** | Suppressor of early Flg22-induced Immune response protein |
| **sgRNA** | Small guide RNA |
| **SUMO** | Small Ubiquitin-like modifier |
| **T3SS** | Type 3 secretion system |
| **TALEN** | Transcriptional activator-like endonuclease |
| **TBS-T** | Tris-buffered saline (with Tween20) |
| **TE** | Transposable Element |
| **TFB1/2** | Transformation buffer 1/2 |
| **TMBSV** | Tomato bushy stunt virus |
| **Tris** | 2-amino-2-hyrdoxymethyl-propane-1, 3-diol |
| **USD** | United States Dollars |
| **X-GAL** | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |

| | |
|---|---|
| **Y2H** | Yeast 2-hybrid |
| **YPAD** | Yeast-extract, peptone, adenine hemisulphate, dextrose medium. |
| **ZFN** | Zinc Finger Nuclease |

# Contents

**List of Figures**

**List of Tables**

# 1

# General introduction

# Chapter 1 General Introduction

## 1.1. The plant immune system

As humans, we often use our own immune system to conceptualise host response to pathogens. The fine detail of the vertebrate immune system lies beyond the scope of the present thesis but one key difference between it and the plant immune system is the existence of specialised, mobile immune cells. Phagocytes patrol the circulatory system and ingest foreign agents like toxins and bacteria (Dale et al., 2008) whilst cytotoxic t-lymphocytes detect and destroy infected cells (Iannacone et al., 2006). Plants do not have the luxury of mobile immune cells capable of responding to invasion anywhere in the plant and, in general, must respond to invading pathogens on a cell by cell basis (with the exception of pre-formed defences and systemic signals which arise from pathogen perception, which will be discussed later). That is not to say that the plant immune system is not fit-for-purpose. Indeed, plants are constantly exposed to a variety of microbes and viruses yet the vast majority of plant-microbe interactions do not result in the death of the plant, or long term colonisation in the case of a biotrophic pathogen. Only certain combinations of plant and pathogen genotypes result in successful infection and propagation of the pathogen, whereas some plant-microbe interactions may have beneficial results for both parties.

The most obvious layer of defence, which repels the majority of would-be invaders, is contributed by pre-formed defences. Most plant organs have a waxy cuticle covering their epidermis. The plant cuticle first evolved some 450 million years ago during the colonisation of land by plants as a means of conserving water and its specific properties have led to it becoming an effective barrier to most pathogenic microbes. Plant cell walls provide both static and dynamic defences. The cell wall acts as a simple, structural barrier to

invasion but, upon attempted penetration by a pathogen, the cell wall can be locally strengthened by the development of localised appositions known as papillae (Underwood, 2012). In addition to these physical barriers, there are pre-formed chemical defences which includes the inhospitable, low pH of the apoplast, a class of pre-formed plant antimicrobials termed phytoanticipins (VanEtten et al., 1994) and a class of peptides known as plant defensins which are highly conserved across animals and plants and appear to function as broad spectrum anti-fungal agents (Thomma et al., 2002).

Whilst these pre-formed barriers are generally extremely successful at warding off would-be attackers, specialised pathogens are capable of breaching them. This is often through the production of cell-wall degrading glycohydrolases or through the development of specialised infection structures such as appressoria or haustoria. In the event of successful penetration the plant must mount a response. Until recently, this has been conceptualised as a sequential response whereby the host recognises a Microbe/Pathogen Associated Molecular Pattern (M/PAMPs[1]), or signals produced by pathogen attack known as Damage Associated Molecular Patterns (DAMPs), and mounts a response to this known as PAMP Triggered Immunity (PTI). Perception of M/PAMPs is mediated by a class of receptors termed Pattern Recognition Receptors (PRR). Successful pathogens can overcome this through secretion of cytoplasmic effector proteins to suppress PTI, resulting in Effector Triggered Susceptibility (ETS) whilst hosts possess intracellular Nucleotide-binding Leucine-rich Repeat receptors (NLRs, encoded by resistance - or *R* - genes) capable of recognising effector proteins, or their activities, and eliciting an accelerated defence response, often resulting in a variety of programmed cell death (PCD) known as the Hypersensitive Response (HR). This zig-zag model of plant defence responses was formalised in (Jones and Dangl, 2006).

[1] **M/PAMPs** shall henceforth be referred to as 'Microbe Associated Molecular Patterns' in order to reflect their conservation across pathogenic and non-pathogenic microbes.

Intracellular NLR immune receptors may either bind directly to a secreted effector protein in an allele specific manner, as with the interaction between Pi-ta and AVR-Pita in the rice blast-rice pathosystems (Jia et al., 2000). Direct interactions may also involve NLR integrated domains as in the AVRPik-Pikp interaction from the same pathosystem (Maqbool et al., 2015). Further to this, NLRs may recognise 'modified-self' where the activity of the pathogen leads to an effect on some host molecule which is guarded by the receptors. A good example supporting this 'guard hypothesis' is the NLR RPM1 guarding the host protein RIN4, which will be discussed in greater detail in section 1.3.2. Another emerging theme in the study of effector-NLR interactions is that of the integrated decoy model (Cesari et al., 2014) mentioned above which proposes that many NLR proteins have evolved a domain which mimics host-targets of effector proteins and serves as a recognition domain, often in the context of paired NLR proteins. It is anticipated that direct interaction of the effector with such domains will initiate conformational change in the receptor or receptor pair and activate defence signalling (Cesari et al., 2013, Williams et al., 2014, Maqbool et al., 2015). A model summarising our current understanding of the plant immune system is presented in Figure 1.

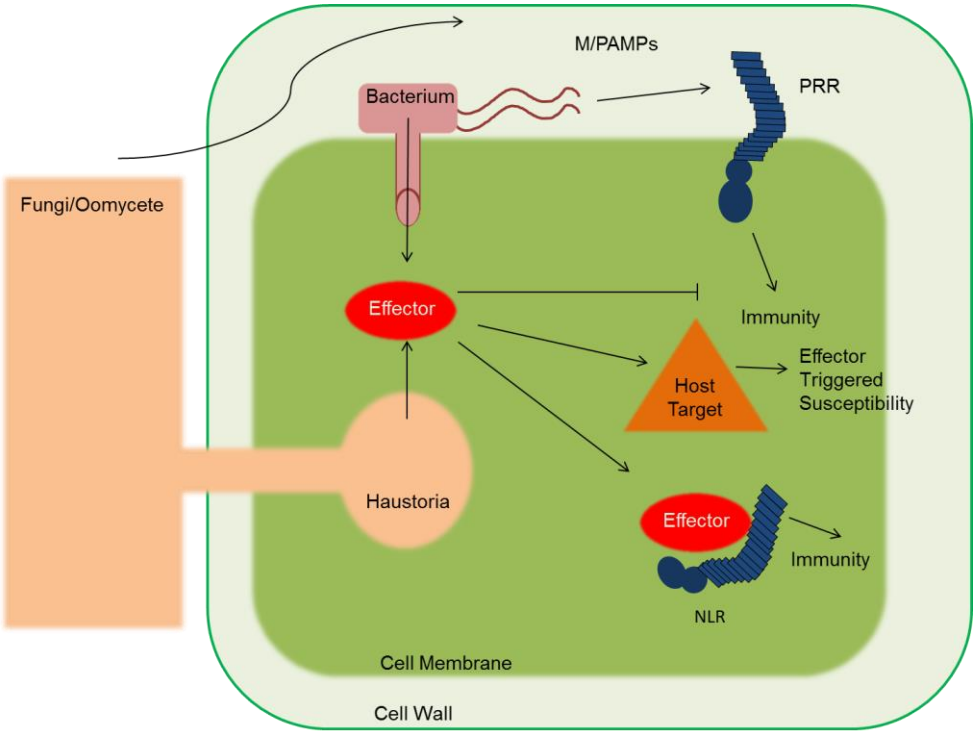



**Figure 1| General model for the secretion of effector proteins into the host cell and the outcomes for resistance/colonisation.**

Filamentous pathogens and bacteria secrete effectors *via* their haustoria or T3SS. These act to condition effector triggered susceptibility (ETS) through repressing PTI signalling from PRRs or by targeting host factors to modulate host cell biology to the benefit of the pathogen, or are recognised by NLR proteins, leading to an immune response. Adapted from (Dodds and Rathjen, 2010).

Over the last ten years, our understanding of the roles of, and interplay between, molecular players in the battlefield of plant-microbe interactions have demonstrated that the zig-zag model (Jones and Dangl, 2006) may too rigidly define the various molecules perceived by the plant as either MAMPS or effectors. A compelling argument is made in (Cook et al., 2015) where an 'invasion model' is proposed. This model defines invasion patterns (IPs) forming a continuum from lineage specific effectors to highly conserved MAMPs. This rethink of how molecular plant-microbe interactions are defined is predicated on the discovery of highly conserved regions in 'effector proteins' which elicit immune response and blur the lines between MAMPs

and effectors. For example, the NLP[2] family of effectors have been demonstrated, through complementation assays, to be critical virulence factors in plant pathogenic bacteria (Ottmann et al., 2009) and may be legitimately called effectors – a term which was adopted to reflect their inherent virulence function from the point of view of the pathogen, as opposed to avirulence proteins, which reflects recognition from the plant's point of view. However, a 20-24 amino acid region (nlp20 or nlp24), conserved across NLPs has now been identified which elicits immune responses (Böhm et al., 2014, Oome et al., 2014). Additionally, a *Botrytis cinerea* effector known as BcSp11 contains two conserved regions in the space of 40 amino acids which trigger host immunity. BcSp11 is a cerato-platanin effector which forms a family only recently identified and whose functions are still largely unknown (Chen et al., 2013). Members of this effector family are expressed by many taxonomically diverse fungi and all of them possess these two, conserved immune response-inducing regions. With this in mind, it may be prudent to define pathogen derived ligands as 'invasion patterns' when describing an overarching general model of immunity.

Another often misunderstood dichotomy set up by the 'zig-zag' model is that which exists between the receptors of invasion patterns. Because MAMPs are typically defined as rigidly invariant, slowly evolving molecules conserved across all species which possess them whilst effectors are encoded by rapidly evolving genes which can be lost or gained over one generation, it is often assumed that their cognate ligands must be slowly evolving (MAMP perceiving PRRs) or rapidly evolving (NLR proteins). Indeed, the data available quite clearly show a great deal of diversity in both the sequence of MAMPs and the strength of recognition by their cognate PRRs. This implies that, like effector genes, genes encoding MAMPs are under strong diversifying selection to evade host perception. This has been observed in *Xanthamonas campestri* pv. *campestris* where a correlation was found between natural variation in the flg22 (a 22 amino acid peptide derived from bacterial flagellin) encoding region of flagellin genes and seedling growth inhibition in *A.*

[2] **NLP** – Nep1-like protein (Nep1 – necrosis and ethelyne-inducing peptide 1).

*thaliana* (Sun et al., 2006), which was taken to be a phenotype associated with defence response (Greenberg and Ausubel, 1993). Similar observations have been made in other bacterial species (Pfund et al., 2004, Clarke et al., 2013) and in other PRR-MAMP interactions, such as the EF-Tu[3]-EFR[4] interaction (Furukawa et al., 2014).

In essence, a compelling argument is made that the current paradigm of plant immune response should be simplified to consider immunogenic ligands like MAMPs, DAMPs, modified 'self' molecules and effector proteins as IPs and the plant derived molecules which perceive these IPs as Invasion Pattern Receptors (IPRs) capable of signalling an Invasion Pattern Triggered Response (IPTR). Invading organisms may have strategies to suppress this response which involve effectors as IPs. The outcome of these interactions determines the success or failure of the invading organism.

Additionally, perception of many of these IPs induces Systemic Acquired Resistance (SAR) which is defined as a systemic response to local pathogen perception. SAR can be induced by activation of both PRR and NLR mediated immune signalling pathways and the systemic resistance produced is often effective against multiple pathogens, not just the one which has elicited the response (Shah, 2009).

[3] **EF-Tu** – Bacterial elongation factor Tu.
[4] **EFR** – EF-Tu receptor.

## 1.2. Filamentous plant pathogens

Many of the most devastating pathogens of our staple and high value food crops are filamentous in nature. 'Filamentous' here refers to a common lifestyle shared between two evolutionarily divergent lineages of pathogens: the oomycetes and the fungi. Both are classes of microbes which form filamentous feeding and sexual structures during different phases of their life cycles. Despite their largely shared infection strategies, key biological differences exist between them and these are summarised in Table 1.

| Characteristic Feature | Fungi | Oomycetes |
|---|---|---|
| **Cell Wall** | Chitin, proline. | Cellulose and glycans, hydroxyproline. |
| **Zoospore Flagella** | Single, posterior 'whiplash' flagella when present. | Biflagellated: 'whiplash' and 'tinsel' flagella which beat in opposite directions. |
| **Vegetative Tissue Genome** | Haploid. | Diploid. |
| **Hyphal Architecture** | Septated. | Rarely septated. |
| **Mitochondrial Ultrastructure** | Flattened christae. | Tubular christae. |
| **Lysine Metabolism** | α-aminoadipic pathway. | α-Ɛ-diaminopimelic acid pathway. |

**Table 1| A description of key physiological differences between Fungi and Oomycetes.**

### 1.2.1. Oomycetes

The Oomycota are a class of filamentous, eukaryotic micro-organisms. Similar to - and often mistaken for - fungi, the oomycetes share a closer evolutionary history with diatoms, brown algae and the apicomplexa than they do with fungi (Baldauf et al., 2000). Nevertheless, oomycetes and fungi are extremely similar in terms of morphology and lifestyle. The lineage belongs to the stramenophiles, a group believed to have arisen from an ancient symbiosis event, where a common ancestor

of all extant stramenophiles presumably appropriated an organism not dissimilar to the red algae we know today (Cavalier-Smith and Chao, 2006, Harper et al., 2005, Patron et al., 2004). The genomes of extant oomycetes still bear the vestiges of this event as photosynthesis related genes can be identified, although no known oomycetes possess plastid like organelles. Within the oomycetes, two major clades exist; the saproglenia, which are predominantly saprotrophic, freshwater organisms, and the peronosporales, which includes several of the most devastating plant pathogens, including the *Phytophthora* genus (Beakes et al., 2012).

The genus *Phytophthora* contains many of the most devastating pathogens known to agriculture and are typically characterised by a large, repeat region-rich genome. One member is especially notable for its cultural and economic significance: the causative agent of late blight in potato and tomato, *Phytophthora infestans*. At least partly responsible for a series of famines across Europe in the middle of the 19$^{th}$ century, it's emergence in Ireland at a politically turbulent time during which the populace were reliant on potatoes for sustenance led to the death of ~1 million people and the emigration of ~1 million more. It remains a significant challenge to global agriculture today, with estimated costs associated with yield loss and control measures totalling some $6.7 billion (USD) *per annum* (Haverkort et al., 2008).

An obligate hemibiotroph[5], *P. infestans* can complete its life cycle both sexually and asexually. Its similarity to fungi is most apparent in its lifestyle. The most immediately obvious shared strategy is the formation of specialised infection structures at the tips of penetrative hyphae known as 'haustoria'. The haustoria invaginate the host cell membrane (presumably through turgor pressure) and form a close association with the membrane which is known as the 'extrahaustorial membrane' (EHM). It is evident that the composition of the EHM differs from the rest of the host cell membrane although the relative contributions of *de novo* EHM membrane protein production and selective exclusion of host membrane proteins like transporters, aquaporins and PRRs remain unclear (Lu et al., 2012). The extent to which invading pathogens can control the composition of the extrahaustorial

[5] Hemibiotroph – An organism which requires living tissue for initial colonisation and parasitism but either requires, or can tolerate, dead tissue for the completion of its infection cycle.

membrane remains an open and thoroughly interesting question. Upon infection, the first macroscopic symptoms appear after as little as 2 days post infection. Infected plants present with patches of spreading tissue necrosis, the *modus operandi* for hemibiotrophic pathogens. Given favourable temperatures (15-24˚C) and sufficient humidity, sporangia can be produced asexually in as little as 4 days. 300,000 sporangia per sporulating lesion may be produced which detach readily and disperse aerially (Aylor et al., 2001). These sporangia germinate on host tissue at temperatures of 20-25˚C or release motile, flagellated zoospores at lower temperatures which will encyst in host tissue. Asexual reproduction is the primary means by which *P. infestans* proliferates but it can also undergo sexual reproduction where mating types A1 and A2 co-exist. The development of sexual structures is reliant on the presence of a bipartite system of sex hormones and serves as a source of genetic variation in the field (Qi et al., 2005). The oospores produced in sexual reproduction have been observed to survive for years in the field under a range of different conditions (Mayton et al., 2000) although upon infection of a susceptible host they appear to be less virulent than sporangia.

As mentioned previously, the *Phytophthora* are characterised by their extremely large genomes rich in repetitive elements. Members of the 1C clade of *Phytophthora* (which includes *P. infestans*) have genomes ranging from 220-280Mb (Tyler et al., 2006, Haas et al., 2009), this is extremely large compared to the smut fungi with ~20Mb genomes (Schirawski et al., 2010, Kamper et al., 2006) and the nearest, non-disease causing relatives of *Phytophthora*, the diatoms, whose genomes never exceed 56Mb (Gobler et al., 2011). One lifestyle trait which has been linked to these expanded genomes is biotrophy. Species of *Phytophthora* and *Hyaloperonospora arabidopsidis* both require living tissue for part, or all, of their association with the host plant. Additionally, their host range tends to be limited. A potential explanation for the increase in genome size is the emergence of asexuality as a consequence of host domestication, leading to polyploidy. However, this does not account for the expanded genomes of *Phytophthora spp.* with natural hosts and their inherent variation, suggesting that genome size may not be influenced by host domestication.

As previously stated, the genome of *P. infestans* contains an extremely large amount of repetitive DNA. Indeed, three quarters of its DNA is made up of repeated

regions and a large amount of this is accounted for by transposable elements (TEs) (Haas et al., 2009). Furthermore, this tendency toward repetition in the genome appears to be associated with biotrophy (Raffaele et al., 2010). The architecture of the *P. infestans* genome reinforces the contribution made by TEs. Some 2000 gene sparse regions, containing fewer than 10 genes per region, have been identified and are characterised by their richness in repetitive DNA elements. The similarity in the GC content in these regions with the overall *P. infestans* genome, added to the similarity in GC content and codon usage between *P. infestans* genes and TEs, also lends support to this hypothesis. Indeed, many other members of the *Phytophthora* genus share this genetic architecture, suggesting that the prevalence of TE derived DNA in their genomes may have been the result of entry into the genome of a last common ancestor by a TE before the divergence of extant *Phytophthora spp.* Indeed, these TE rich, repetitive regions may have driven divergence and diversity among *Phytophthora spp.* as the ratio of dN/dS mutations in genes residing in these regions are far higher than their counterparts located in gene rich regions, suggesting a 'two-speed genome'. The evolution of the genomes of *Phytophthora spp.* will be discussed further in section **1.3.2**.

### 1.2.2. Fungi

The fungi are a kingdom of heterotrophic eukaryotes which range from single celled microbes - like yeasts - through to multicellular organisms which produce fruiting bodies known as mushrooms or toadstools. It is difficult to overstate the importance of fungi in almost every terrestrial ecosystem on the planet given their role as prime decomposers of organic matter. Further to this, the positive impact of fungi on human culture has been enormous, being of critical importance in bread, beer and wine making.

Despite the benefits provided by some fungal species, others are capable of causing mass devastation to crops. The rice blast fungus, *Magnoportha oryzae*, is an Ascomycete fungus which is endemic in every rice growing region of the world and the lost yield attributed to it is estimated to be enough to feed 60 million people annually (Fisher et al., 2012). Symptoms present as white-grey elliptical lesions

which can spread and destroy whole leaves, as well as neck blast which prevents grain maturation in panicles. One sporulating lesion can produce up to 1,000 spores per night and can continue doing this for up to 20 days, rapidly infecting and destroying whole fields of rice. A Basidiomycete, *Phakopsora pachyrhizi*, is responsible for Asian Soybean Rust (ASR). *P. pachyrhizi* causes yellowing of leaves, severe lesions on the abaxial face of leaves and necrotic lesions with pustules and visible ASR spores. The ultimate result is defoliation, abrogation of pod production and poor pod fill. Yield losses can run to 80%. ASR spores are windborne and are thought to travel vast distances in storm systems and upon infection they can complete their life cycles within 9-10 days. Another Basidiomycete of economic importance is the Corn Smut Fungus *Ustiligo maydis*. Yield losses associated with smut are relatively minor, particularly in the USA where resistant cultivars are grown. Yield loss, however, sometimes comes as a double edged sword for farmers as the dramatic looking fruiting body from which *U. maydis* takes its common name 'smut' has been eaten since before European colonisation of the Americas in the 16$^{th}$ century. Indeed, market value of smut sometimes exceeds that of the maize on which it grows.

One member of the Ascomycetes - *Blumeria graminis* – is the causative agent of powdery mildew on cereals and will form part of the focus of the present work. An obligate biotroph, symptoms are invariably small, powdery white spots on both faces of the leaf in addition to chlorotic lesions around the site of infection. Despite not being visibly destructive through symptoms like defoliation and neck blasts, as with previously discussed diseases, yield losses of up to 45% have been observed in favourable conditions for the pathogen through reduction of viable photosynthetic area. Furthermore, liberal application of nitrogenous fertiliser favours fungal growth and can act to reduce yield.

The infection cycle of *B. graminis* is described in (Both et al., 2005). Briefly, conidia land on a susceptible surface and, given favourable conditions germinate. 15 hpi a fully functional appressorium and penetration peg form which invade the host's intracellular space and permit the formation of specialised infection structures called haustoria. Following penetration, epiphytic mycelia which produce further infectious

condidia and the typical white, powdery spots visible on the leaf surface appear 3-4 days post infection.

Consistent with observations made in oomycetes, biotrophy seems to be associated with expanded genomes. *Blumeria graminis* f.sp. *hordei* has a genome of around 120M, more than four times the median size of other Ascomycete genomes (Spanu et al., 2010). Despite this large genome size, the number of predicted genes is actually toward the lower end of the range for fungi whose genomes have been annotated. Predictably, the genes missing in comparison to other fungi are those not required for an obligately biotrophic lifestyle, like genes involved in inorganic nitrogen fixation, secondary metabolite synthetic enzymes, toxin transporters and genes which may be involved in the degradation of plant cell walls. Further to this, it was found that 64% of the *B. graminis* f.sp. *hordei* genome was made up of TEs, compared to ~75% in *P. infestans* (See section 1.2.1). It has been hypothesised that frequent transposition events may account for the high rates of gene loss evident in the *B. graminis* f.sp. *hordei* genome.

## 1.3. *Phytophthora* effectors and their targets

Many *Pyhtophthora* effector proteins are secreted into the apoplastic space where their primary function seems to be inhibition of host hydrolytic enzymes and proteases. However, their direct relevance to the work presented in this thesis is limited, so they are not discussed in detail here. The effector proteins which are secreted by the pathogen and taken up by the host by poorly understood mechanisms – known as cytoplasmic effectors – are the focus of the present thesis and discussed here.

### 1.3.1. Cytoplasmic effectors

To date, two major classes of Oomycete effector proteins have been described; the RXLR effectors – which form much of the basis of the work presented in this thesis - will be discussed in depth in section 1.3.1.1, and the Crinkler (CRN) effectors. Unlike RXLR effectors, CRN effectors are ubiquitous in known plant

pathogenic oomycete genomes. The *P. infestans* genome encodes a significantly expanded complement of CRN proteins relative to other *Phytophthora* spp. with 196, compared to 100 in *P. sojae* and only 19 in *P. ramorum*. CRN effectors are typically large proteins with a highly conserved domain architecture comprising an N terminal signal peptide, followed by the 'LFLAK' motif (Win et al., 2006) which is thought to be involved in translocation of the effector domain into the host cell (Schornack et al., 2010) and a C terminal effector domain which accounts for the diversity of functions in CRN effectors. Overexpression of some CRN effectors in *S. lycopersicum* and *N. benthamiana* results in 'crinkling' of leaf tissue and severe necrosis (Torto et al., 2003). This induction of necrosis has led to the hypothesis that some CRNs may function as regulators of colonisation later in the pathogen lifecycle, during or following the switch to necrotrophy (Qutob et al., 2002). Knowledge of the exact *in planta* function of CRN effectors is still in its infancy, but one well studied example – the nuclear-localised CRN8 – functions as an RD kinase in the model host *N. benthamiana* where it has been shown to autophosphorylate (Van Damme et al., 2012) and its kinase activity appears to be important for *P. infestans* virulence. However, CRN8's position in any endogenous MAPK signalling pathway in the nucleus remains unknown.

Localisation studies of CRN effectors have revealed a diversity of sub-nuclear localisations and patterns, in addition to a diversity of associated responses when transiently expressed in *N. benthamiana*. (Stam et al., 2013) observed that three separate CRN effectors (CRN79_188, CRN83_152 and CRN20_624) induced delayed, rapid and intermediate cell death symptoms, respectively. Distinct localisation within the nucleus and differential co-localisation with chromatin and Histone complex sub-units was also observed, indicating that cell death phenotypes may correlate with the nature of effector association with DNA.

#### 1.3.1.1 **RXLR effectors**

As alluded to in section 1.3.1, one of the two known classes of oomycete effector proteins is the RXLR effectors. 'RXLR' refers to a conserved amino acid motif

toward the N terminus of the protein comprising arginine, any amino acid, leucine and arginine and is often followed by a dEER motif which comprises a poorly conserved aspartate residue, two glutamates and an arginine (Dou et al., 2008). The RXLR family in *Phytophthora* spp. is comprised of >300 proteins in *P. sojae* and *P. ramorum*, respectively and >500 in *P. infestans* (Jiang et al., 2006, Haas et al., 2009). The RXLR effectors share a conserved domain architecture with an N terminal signal peptide followed by the canonical RXLR-dEER motif and a C terminal 'effector domain' which encodes the effector's biochemical function and its NLR recognition specificity. Significant sequence diversity is observed in the C terminal regions of RXLR effectors, consistent with their identification as a rapidly evolving gene family as well as with their role in virulence. In addition to this, new members of the RXLR family are rapidly evolving, through processes such as pseudogenisation, and members deleterious to fitness are deleted or pseudogenised when exposed to specific cultivars (Haas et al., 2009)

It has been hypothesised that the RXLR motif is required for translocation of effector proteins into the host cell (Whisson et al., 2007, Dou et al., 2008) and it is known that it is not required for virulence activity (Bos et al., 2006). This translocation mechanism may share similarities with the translocation of effector proteins from the apicomplexan malaria parasite *Plasmodium falciparum* into host cells (Bhattacharjee et al., 2006). *P. falciparum* effectors are translocated across a tight association between host and pathogen membranes but pathogen independent entry into the host cell has not yet been observed (Marti et al., 2004). Whilst the oomycete RXLR motif is demonstrably required for translocation, it remains unclear whether or not it is sufficient for pathogen independent entry into the host cell. The evidence for pathogen independent entry was based on recombinant, fluorescently tagged effectors being applied to plant roots in the absence of a pathogen, and their uptake observed (Dou et al., 2008). However, the strength of this type of assay has been called into question by the observation that non-specific uptake of fluorescent proteins by plant roots occurs at a similar rate to fluorescently labelled effector proteins (Wawra et al., 2013).

The mechanism by which RXLR proteins enter the host cell independently of the pathogen remains a topic of considerable debate among the oomycete community. In

2010 it was found that RXLR motifs bind to phosphatidylinositol-3-phosphate (PI3P) outside of the host cell before being endocytosed (Kale et al., 2010). However, since then evidence from the *P. falciparum* pathosystem has suggested effector proteins may bind PI3Ps inside of the host cell, rather than outside, calling into question its functional relevance for effector translocation (Bhattacharjee et al., 2012). Additionally, the sufficiency of the RXLR motif in isolation for PI3P interaction has been questioned, with reports suggesting that the C terminal domain of RXLR effectors may have a greater affinity for PI3P than the putative translocation motif (Sun et al., 2012). Further to this, it was shown that denatured Avr3a could still interact with PI3Ps, raising doubts over any proposed relevance of an RXLR-PI3P interaction in effector translocation (Wawra et al., 2013).

### 1.3.2. Host Targets

As previously mentioned, the primary role of cytoplasmic effector proteins is to enter the host cell and modulate host biology in some way which is beneficial to the invading organism; this normally manifests itself as suppression of immune responses. However it may also take the form of redirection of nutrients to the pathogen (Yang et al., 2006, Antony et al., 2010).

The first host-targets of plant-pathogen effector proteins to be identified were of bacterial effectors delivered *via* the type-three secretion system (T3SS). One particularly notable example of a host target is the *A. thaliana* RPM1-interacting protein RIN4. The activity of at least seven *P. syringae* T3E proteins converge on RIN4. Examples of these include; AvrPt2 (Axtell and Staskawicz, 2003), AvrB (Mackey et al., 2002), AvrPm1 (Mackey et al., 2003) and AvrPto (Luo et al., 2009). Reduction in RIN4 levels leads to enhanced resistance of *A. thaliana* to several pathogens and AvrB and AvrPm1 phosphorylate RIN4 in order to stabilise it and promote its negative regulation of basal defence. However, this is only the case if the plant under attack does not express functional RPM1 as the RIN4-RPM1 association is one of the classical examples of ‘the guard hypothesis’, whereby an effector does not interact directly with an NLR protein to trigger ETI. Instead, the NLR protein ‘guards’ targets of effectors, surveying the cell for ‘modified self’. Phosphorylation

of RIN4 by one of these T3Es then triggers RPM1 mediated immunity. Likewise, RIN4 also physically associates with the NLR Resistance to *Pseudomonas syringae* 2 (RPS2). AvrPt2 directly interacts with RIN4 and reduces its levels through its activity as a cysteine protease, resulting in RPS2 activation and ETI (Axtell and Staskawicz, 2003). AvrPto also reduces RIN4 levels in a Pto/Prf dependent manner (Luo et al., 2009). Degradation of a negative regulator of immunity is a difficult observation to explain but it has been suggested that this may have evolved as a means to limit recognition of AvrB and AvRpm1 mediated RIN4 modification.

Oomycetes do not possess a T3SS, or other bacterial style secretion system with which to deliver effector proteins to the host cytoplasm. As outlined in section 1.2.1, oomycetes form specialised infection structures known as haustoria, across which they appropriate nutrients from the host and, it is thought, secrete effectors *via* a mechanism likely to involve the RXLR motif and other, as yet uncharacterised, translocation motifs.

With the availability of high quality genome sequences for numerous plant pathogens, effector discovery and characterisation has increased in recent years. This has been made possible through the advancement of next generation sequencing technologies (NGS) and improved bioinformatics pipelines for the large amounts of data generated by the short reads inherent to NGS methods. The identification of putative effector complements in plant pathogens has allowed the screening of effector proteins for host interactors and subsequent validation of the interaction. Once this is complete, experiments can be designed to ascertain a biologically relevant outcome of a given effector-target interaction. A non-exhaustive list of *Phytophthora* spp. effector proteins and their identified host targets, with a brief note on the consequences of each interaction in the context of infection, is presented in Table 2 so as to give an overview of the diversity of processes oomycete effectors as capable of manipulating.

| Target | Effector | Organism | Biological Function | Reference |
|---|---|---|---|---|
| **Sec5** | **Avr1** | *P. infestans* | Avr1 binds an exocyst complex protein perturbing callose deposition and PR1 secretion. May also disturb vesicle trafficking. | (Du et al., 2015) |
| **BSL1** | **Avr2** | *P. infestans* | BLS2 is closely related to the Brassinosteroid response regulator BSU1, linking it to hormone signalling. The exact consequences of this interaction remain unknown. | (Saunders et al., 2012) |
| **CMPG1** | **Avr3a** | *P. infestans* | Avr3a stabilises CMPG1, a host E3 Ligase. CMPG1 is degraded during INF1-induced cell death and Avr3a mediated stabilisation suppresses this response. | (Bos et al., 2010) |
| **C14** | **Avrblb2** | *P. infestans* | Avrblb2 prevents C14 accumulation in the apoplast. | (Bozkurt et al., 2011) |
| **C14** | **EPIC1/EPIC2B** | *P. infestans* | Apoplastic effectors EPIC1 and EPIC2B physically associate with C14 and inhibit its protease activity. | (Kaschani et al., 2010) |
| **LecRK-I.9** | **IPI-O** | *P. infestans* | LecRKI.9 maintains CW-PM integrity. IPI-O directly binds it and disrupts the CW-PM continuum. | (Bouwmeester et al., 2011) |
| **MAPKKK$_\varepsilon$** | **PexRD2** | *P. infestans* | PexRD2 suppresses MAPKKK$_\varepsilon$ mediated cell death response to promote virulence. | (King et al., 2014) |
| **ATG8** | **PexRD54** | *P. infestans* | PexRD54 associates with host autophagy related protein 8 (ATG8) in order | (Dagdas et al., 2016) |

| | | | | |
|---|---|---|---|---|
| | | | to perturb host selective autophagy. | |
| **KRBP1** | **Pi04089** | *P. infestans* | KRBP1 accumulation is promoted by Pi04089 to promote infection. | (Wang et al., 2015) |
| **PP1c** | **Pi04314** | *P. infestans* | Pi04314 mimics a PIP through its KVTF motif and directly interacts with PP1c to promote infection through formation of a holoenzyme. | (Boevink et al., 2016) |

**Table 2| A none-exhaustive list of *Phytopthora infestans* effector-target interactions.**

The effector-target interaction that will form the focus of one chapter in the present thesis is shown in Table 2. (King et al., 2014) showed that a *P. infestans* RXLR effector – PexRD2 – directly interacted with a solanaceous MAPKKK$_{\varepsilon}$ in order to suppress its function as a positive regulator of immunity to biotrophic and hemibiotrophic pathogens. The interaction between PexRD2 and MAPKKKK$_{\varepsilon}$ is discussed in greater detail in Chapter 3.

## 1.4. Mitogen activated protein kinase signalling in plant immunity

As key regulators of a wide variety of processes, including immune signalling, MAP Kinases represent a class of molecules that one may predict to be targeted by plant pathogen effectors. Examples of host kinases targeted by microbial effector proteins are abundant in the literature. The *Oryza sativa* Receptor-like cytoplasmic kinase OsRLCK185 is a protein kinase which directly interacts with – and is phosphorylated by – the chitin and peptidoglycan receptor OsCERK1 in order to upregulate the expression of defence response genes. The *Xanthomonas oryzae* effector protein Xoo1488 physically associates with OsRLCK185 and suppresses its

phosphorylation by OsCERK1, preventing any downstream defence signalling (Yamaguchi et al., 2013). A further example links back to the guard hypothesis, in which NLR proteins monitor the cell for effector activity – the *P. synringae* T3SS effector HopAI1 physically associates with, and inhibits, the *A. thaliana* kinase MPK4 with the in order to suppress expression of defence related genes. However, the NLR protein SUMM2 detects suppression of the MEKK1-MKK1/MKK2-MPK4 signalling cascade and mounts an effector-triggered immune response, highlighting the sophistication of the plant intracellular immune system (Zhang et al., 2012b). Further to this, 33 RXLR effectors from *P. infestans* were screened in transient, protoplast based assays for their ability to suppress Flg22 mediated defence signalling. In tomato protoplasts, 8 out of 33 of the screened effectors were capable of suppressing Flg22 signalling at the MAP Kinase level whilst in protoplasts from the non-host *A. thaliana*, 3 of the effectors had the same effect and were termed Suppressor of early Flg22-induced Immune response (SFI) proteins (Zheng et al., 2014). The ability of *P. infestans* effectors to suppress MAP Kinase mediated immune signalling in a non-host organisms suggests that targeting of MAP Kinase mediated signalling may be an evolutionarily conserved virulence strategy, common to many plant-associated microbes.

Upon perception of modified self, direct detection of an effector protein or MAMP perception, the resulting signal must be amplified and translated into some tangible response. Clearly there is a spatial disconnect between, for example, the highly-conserved, plasma membrane-bound flagellin receptor FLS2 and the transcriptional changes and ROS burst associated with flagellin perception which presumably require a signal to be carried from the plasma membrane, through the cytoplasm (where stimuli perception may also occur through intracellular NLRs) to the nucleus. This disconnect is, at least in part, bridged by a class of signalling proteins conserved across eukaryotes: the mitogen activated protein kinases (MAPKs) (sometimes known as MAPK/ERK kinases (MEKKs)). In general, MAPK signalling cascades incorporate three functionally similar proteins in which downstream kinases are sequentially phosphorylated by their upstream kinase, before the terminal MAPK phosphorylates further signal transduction components – including ribosomal sub-units and transcription factors - which induce the responses outlined in Figure 2. MAPK activation is typically a rapid response and can be

detected as little as five minutes after perception of an avirulence effector. The best understood of these three tiered signalling cascades is the one already alluded to in the present section in which flg22 is detected by FLS2. This recognition event leads to the rapid formation of a complex between flg22 bound FLS2 and the receptor-like kinase SERK3/BAK1 and subsequent activation of two MAPK signalling cascades, which were introduced briefly earlier in this section. The first of these – which promotes defence response - involves MEKK1, MKK4/5 and MPK6/3. The second also includes MEKK1 which in turn activates MKK1/2 and MPK4. The latter pathway had been thought to antagonise the immune response but recent evidence suggests that this may not be the case. Plants deficient in the MPK4 signalling cascade exhibit constitutive resistance phenotypes (Petersen et al., 2000). However, it turns out that MPK4 is a *guardee* of SUMM2, an intracellular NLR (Zhang et al., 2012b). *Summ2* mutants no longer exhibit the constitutive immune response associated with *mpk4* mutants and SUMM2 is required for defence response triggered by HopAI1. It was proposed that MPK4 negatively regulates MEKK2 by phosphorylation and that activation of MEKK2 in the absence of MPK4 leads to constitutive SUMM2 activation.

Adding further weight to the Invasion Model proposed in (Cook et al., 2015), multiple invasion patterns may converge upon the same MAPK cascade suggesting that these are not separate responses but immune signalling hubs. DAMPs, such as oligogalacturonic[6] acids (OGs), for example, are perceived by wall-associated kinases (WAKs) WAK1 and WAK2 which activate the same downstream effectors as flg22.

Several of the outcomes of MAPK signalling cascades as plant immune responses are outlined in Figure 2. Of those, cell death is of particular importance to the present thesis. The hypersensitive response is a form of programmed cell death which is the result of SA and ROS accumulation and numerous metabolic changes which essentially act to prevent the spread of biotrophic or hemibiotrophic pathogens by killing the cells in which the pathogen has a foothold. Also, a more robust form of cell death can be elicited by MAP Kinases involved in plant immune signalling, such

[6] Oligogalacturonic acid – Breakdown product of plant cell walls denatured by fungal galacturonases.

as by MAPKKK$_\varepsilon$ in response to Cf-4 or AvrPto (Melech-Bonfil and Sessa, 2010b, King, 2013). A combination of silencing and overexpression studies have demonstrated that solanaceous MAPKKK$_\alpha$ and MAPKKK$_\varepsilon$ act upstream of these signalling modules and transduce perception of effector proteins by intracellular NLRs (Pozo et al., 2004, Melech-Bonfil and Sessa, 2010b). Our understanding of the roles of these two MAPKKKs will be discussed in greater depth in Chapter 3.

It is clear that MAP Kinases represent a generally conserved class of virulence targets for plant-associated pathogens from all kingdoms of life. This is supported by our emerging understanding of NLR-integrated domains, which are hypothesised to be derived from effector-target proteins. In a number of bioinformatic searches of crop plant reference genomes, MAP Kinase domains have been the best represented class of putative integrated domain (Sarris et al., 2016, Kroj et al., 2016) suggesting selection pressure favouring the accumulation of NLRs containing domains homologous to common targets of plant pathogen effectors – presumably to serve as 'decoy domains'.

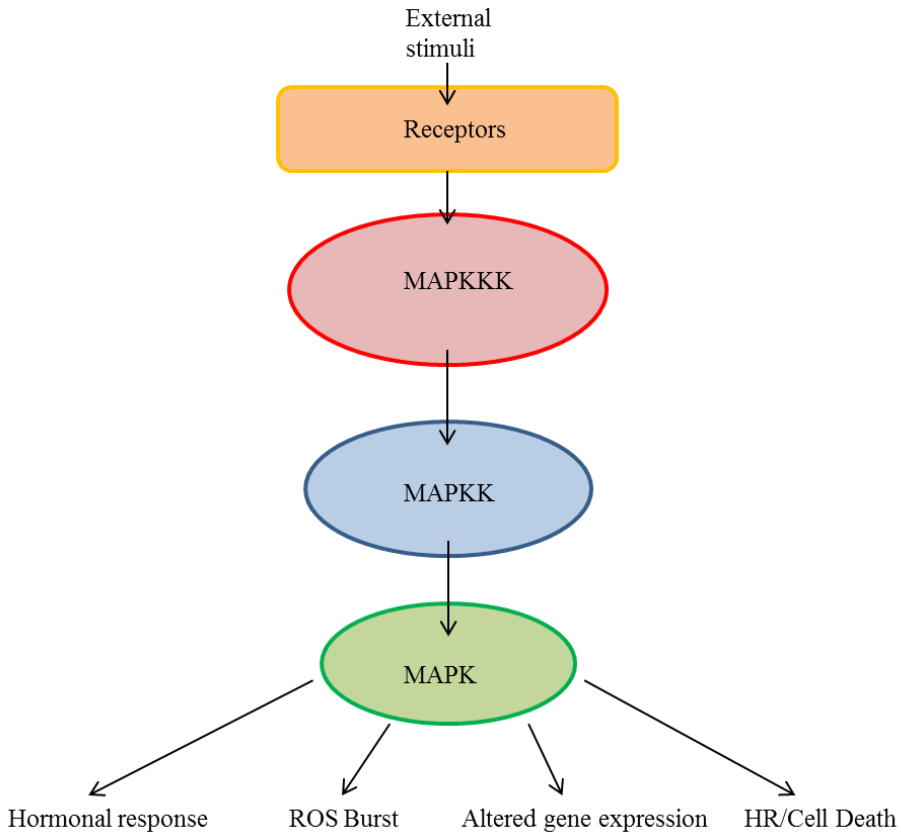



**Figure 2| Generalised representation of the core MAPK signalling module and the responses they may elicit upon pathogen perception.**

At the base of a given cascade, a MAPK is activated by phosphorylation of a Thr-x-Tyr motif in its activation loop by the activity of a MAPKK, which is in turn phosphorylated by a MAPKKK. MAPKKK's may also be phosphorylated by further module known as a MAPKKKK. Diversity in known MAP Kinases reduces down the cascade, with at least 80 putative MAPKKKs, 20 MAPKKs and 10 MAPKs annotated in the *A. thaliana* genome (Ichimura et al., 2002, Colcombet and Hirt, 2008). Four responses are generally associated with MAPK cascade activation. Accumulation of defence hormones like Salicylic acid and Methyl jasmonate may be promoted. Accumulation of ROS such as superoxide and hydrogen peroxide may occur in order to strengthen cell walls and limit pathogen spread. Additionally, gene expression is altered so as to upregulate defence related genes and downregulate susceptibility genes. Lastly, a localised HR or cell death response may be induced in order to limit the spread of biotrophic pathogens.

## 1.5. CSEPs from Blumeria graminis *f.sp.* hordei

### 1.5.1. Identification

The Candidate Secreted Effector Proteins (CSEPs) of *Blumeria graminis f. sp. Hordei* (*Bgh*) were initially identified when the *Bgh* genome was published (Spanu et al., 2010). In this paper, 248 proteins with an N-terminal signal peptide, no predicted transmembrane domains and no homology to proteins outside of the mildews were identified bioinformatically and predicted to be effector proteins. Consistent with this, expression analysis found that 79% of these genes were preferentially expressed in haustoria. A more comprehensive analysis, incorporating genomic, transcriptomic and proteomic techniques also identified the aforementioned 248 proteins. A further 243 candidate secreted effector proteins were added through iterative rounds of BLAST searches using previously identified CSEPs as queries and signal peptide analysis. The approach also employed BLAST searches against an EST database using putative CSEPs as queries, RNA expression analysis as well as proteomic analysis of extracts from sporulating *Bgh* hyphae and barley epidermal tissue containing *Bgh* haustoria (Pedersen et al., 2012). In all, ~500 proteins have been identified as CSEPs. Interestingly, a Y/F/WxC motif has been identified in a sub-set of CSEPs, where the first residue in the triplet motif is always aromatic, the second is any amino acid and the last is invariably a cysteine (Godfrey et al., 2010). However, the functional significance of this motif is not yet clear.

### 1.5.2. Functional knowledge

Despite such a large number of predicted CSEPs, confirmation of effector function and more detailed characterisation has proved challenging. One major hurdle has been an inability to transform *Bgh*, which is largely a product of its obligately biotrophic lifestyle (making it difficult to culture in the laboratory) and hydrophobic cell surface. Therefore, all functional characterisation of *Bgh* CSEPs to date has been done through Host Induced Gene Silencing (HIGS). Effector function can be assigned to CSEPs in this way by knocking them down and assaying for

compromised virulence. Around half of the ~500 described CSEPs resulted in compromised virulence when silenced (Nowara et al., 2010).

At least two CSEPs - CSEP0105 and CSEP0162 - resulted in a 40% lower haustorial formation rate when silenced and were characterised further. Almost exclusive haustorial expression of both during the process of infection was reconfirmed through qPCR (Ahmed et al., 2015) and interaction partners were identified through a yeast 2-hybrid screen against a prey library constructed from infected Barley derived cDNA (Zhang et al., 2012a). It was found that both CSEPs under investigation interacted with a pair of small heat shock proteins; Hsp16.9 and Hsp 17.5. Furthermore, co-expression of fluorescently tagged CSEPs were found to localise exclusively to the cytoplasm when co-expressed with the Hsps, the known cellular location of these Hsps. Despite strong evidence for a physical interaction between these effectors and Hsps, as well as for a virulence function of both CSEP0105 and CSEP0162, whether these CSEPs are actually targeting Hsps in order to promote virulence remains unclear. CSEP0105 was shown to disrupt the chaperone activity of Hsp16.9 whilst the same result could not be obtained for CSEP0162 (Ahmed et al., 2015). A direct role for the perturbation of function in *Bgh* virulence has not been established although it is known that Hsps can play an important role in defence response. It remains a possibility that the observed interaction with Hsp16.9 and Hsp 17.5 is an artefact. Indeed, Hsps are known 'sticky hubs' which are prone to generating false positive results in protein-protein interaction studies.

Another CSEP, CSEP0055, was found to be important for virulence, again through HIGS (Zhang et al., 2012a) and was subsequently found to interact with several proteins known to accumulate in the apoplast in response to *Bgh* invasion. The *in planta* roles of these putative CSEP0055 targets remain unclear so the relevance of the interaction for disease outcome is unknown, although it has been suggested that CSEP0055 interacting proteins may function as proteases in the apoplast, based on sequence similarity to known proteases. This would suggest that CSEP0055 may function as an inhibitor of protease activity in the apoplast.

It is clear that our understanding of effector proteins from this economically important pathogen of cereal crops is in its infancy and that work directed at

understanding the diversity of roles these effectors fulfil in host-colonisation will be crucial to developing crop protection strategies.

## 1.6. The potential of genome editing technologies for delivering enhanced resistance

In the coming years we may expect crop manipulation for a range of traits – including resistance to pathogens – to increase in pace. The development of three technologies in particular, each based around endonuclease proteins, will permit this.

Zinc Finger Nucleases (ZFNs) are an artificial class of two-component endonucleases, in which a zinc-finger DNA binding domain is fused to the non-specific cleavage domain of a Type II restriction enzyme, typically FokI (Kim et al., 1996). Target specificity is encoded by the DNA binding domain which often contains an array of zinc-finger repeats, and efforts to engineer greater target specificity have focused on building arrays of individual zinc-fingers – recognising 3bp each – which will recognise sequences ranging from 9-18bp (Ramirez et al., 2008). ZFNs have been successfully employed in economically important crop plants, such as Soybean, where they were deployed to disrupt Dicer-like proteins (DCLs) and other components of the endogenous RNA silencing pathway *via* hairy roots transformation (Curtin et al., 2011).

Transcriptional activator-like effectors (TALs) are a class of effector proteins discovered in members of the *Xanthomonas* genus of plant pathogenic bacteria, and are now known to be expressed in *Ralstonia spp.* and *Burkholderia spp.*. Their function, in the context of natural plant microbe interactions, appears to be to enter the host nucleus and bind to specific promoter sequences, inducing transcription of the gene under that promoter's control in order to benefit the pathogen. Examples of this include susceptibility genes *Os*8N3, which is upregulated by pthXo1 (Yang et al., 2006), and *Os*-11n3, which is upregulated by AvrXa7 (Antony et al., 2010). Sequence specificity is encoded by variable, paired amino acids in a 34 amino acid region of the TAL effector. The paired amino acids bind individual nucleotides in a

2:1 ratio and binding specificity is predictable, enabling researchers to design TALENs to target specific DNA sequences (Moscou and Bogdanove, 2009, Boch et al., 2009). Like ZFNs, TAL effectors require fusion of the cleavage domain of restriction enzyme (often FokI) in order to become functional genome editing tools known as Transcriptional activator-like effector nucleases (TALENs). These artificial TALENs have been successfully deployed in rice in order to edit a susceptibility gene appropriated by *Xanthomonas oryzae* in order to divert sugars to the pathogen (Li et al., 2012) resulting in almost fully resistant rice plants.

CRISPR/Cas9 is now the most widely adopted strategy for genome editing in plants and animals due to its efficiency and ease of programmability, relying on specific nucleic acid guide sequences to guide the Cas9 nuclease to the intended target. The discovery of CRISPR/Cas9, its appropriation for genome editing in all kingdoms of life and the underlying biology behind the system are discussed in greater detail in section 3.2.6. Briefly, CRISPR/Cas9 was derived from a type of bacterial immune system which has evolved to defend bacterial cells from viral invasion. It relies on the expression of artificial nucleic acids which specify DNA binding, Cas9 association as well as the expression of a Cas9 nuclease to induce DSBs at target sites. Its ease of programmability and use has seen it become widely adopted by researchers across almost all experimental systems.

## 1.7. The role of structural biology in understanding molecular plant-microbe interactions

In recent decades the number of deposited structures in the Protein Data Bank (PDB) has increased dramatically in general, but so to have the number of 3D structures of molecules relevant to plant-microbe interactions. This is the result of a number of key technical advances. One historical limitation had been the production of large quantities of pure protein for crystallisation. Protein had to be obtained through expression in their native organism, often at quite low yields. Today, advances in cloning methods have allowed the rapid development of suites of vectors for expression in a number of heterologous hosts, notably *Escherichia coli*, yeast, insect cells, mammalian cells and plants (Berrow et al., 2007). In some of these

heterologous expression systems, notably *E. coli*, there is a great diversity in available strains dependent on the properties of the protein of interest. Bioinformatic progress has enabled accurate prediction of disordered amino acid sequences, enabling rapid design of constructs for expression and crystallisation, where protein order is a major concern. Purification methods have also seen improvements with multi-step, automated systems enabling the rapid purification of single proteins from complex cell lysates through methods such as Immobilised Metal Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC). The process of crystallisation is now automated and carried out on a nano-litre scale, enabling the screening of 1000s of conditions in commercially available, sparse matrix screens with small volumes of protein prior to optimisation around successful screening conditions. Access to high powered, multi-user X-ray diffraction facilities with tuneable beamlines has improved for many researchers enabling the collection of high quality data. This often results in higher resolution 3D structures, whilst improved software pipelines now mean that automated structure solution without additional user input is becoming increasingly commonplace. Additionally, the deposition of numerous 3D structures has allowed modern researchers to utilise *a priori* knowledge in order to solve some 3D structures through molecular replacement methods, rather than through isomorphous replacement or anomalous dispersion experiments (see Chapter 5 for further discussion).

Structural information can provide many key insights which further the development of hypotheses on protein function and inform experiments with which to test these ideas. Often, function is not clear from an amino acid sequence or conserved structural features cannot be determined, obtaining a high quality 3D structure can allow functional predictions to be made rationally and reveal conserved secondary structural elements between related but sequence divergent proteins, allowing unique evolutionary insights.

Co-crystallisation of effector proteins and their host targets or cognate NLR proteins has resulted in great excitement in recent years, with the molecular basis of Harold Flor's gene-for-gene hypothesis being elucidated. Specific insights into plant-microbe interactions derived from structural studies are beyond the scope of the

present chapter but an excellent review on the structural biology of plant-microbe interactions is provided by (Wirthmueller et al., 2013).

Perhaps the most exciting conclusion to emerge from studying the structural biology of plant-microbe pathosystems to date has been an elucidation of the structural basis of the 'gene-for-gene' model of plant immunity/susceptibility proposed by Howard Flor (Flor, 1942, Flor, 1947, Flor, 1955). The recognition of the Rice Blast fungus (*Magnaporthae oryzae*) translocated effector protein AVR-Pik by the Rice intracellular NLR immune receptor Pikp had previously been described. As had natural polymorphisms in both effector and receptor contributing to different disease outcomes. Pikp functions as a heterodimer *in planta* with one interacting partner (Pikp-1) possessing a HMA domain between its CC and NB-ARC domains as a putative integrated decoy domain (Cesari et al., 2014). Previous studies had demonstrated a physical interaction between the Pikp-1 HMA domain and AVR-PikD (Kanzaki et al., 2012). This interaction was further confirmed through surface plasmon resonance (SPR) studies which found that AVR-PikD binds the Pikp-1 HMA domain with nanomolar affinity (Maqbool et al., 2015). Further to this, they were able to co-express the Pikp-1 HMA and AVR-PikD in *E. coli*, co-crystallise the interaction partners and solve the crystal structure providing unprecedented insights into the interface between a plant pathogen effector protein and its cognate NLR. Through structurally informed mutation studies, an ancestral Histidine residue at position 46 in AVR-PikD, which is shown to be buried in the Pikp-1 HMA domain in complex, was shown to be crucial for interaction with Pikp-1. AVR-PikD is the only natural variant of AVR-Pik to have a Histidine at position 46 and, tellingly, is the only AVR-Pik variant to be recognised by paired Pikp receptors, supporting the idea that $His^{46}$ is ancestral and the Asparagine polymorphism has evolved to help AVR-Pik evade recognition and limit immune response to *M. oryzae*.

Another utility of structural biology is assignment of function to proteins whose function would not otherwise be predicted based on their amino acid sequence. The *P. syringae* effector AvrPtoB is a multi-domain protein with two domains functioning as kinase inhibitors of Pto and BAK1 (Cheng et al., 2011, Mathieu et al., 2014) whilst a third, C-terminal domain from amino acid 436-553 was found, through determining its 3D structure, to share homology with the eukaryotic U-box

E3 RING-finger ubiquitin ligase (Abramovitch et al., 2006). This was not apparent from its amino acid sequence and it was found that AvrPtoB does indeed tag components of *A. thaliana* immune complexes for proteasomal degradation, thus promoting *P. syringae* virulence.

In a similar vein, insights into effector evolution can be gained through structural knowledge where sequence homology lets us down. A highly conserved, alpha-helical fold was identified in effectors from the peronosporales order of oomycetes. Initially, the fold was discovered in PexRD2, Avr3a4, Avr3a11 from *Phytophthora* spp. and ATR1 (Boutemy et al., 2011, Win et al., 2012) from *Hyaloperonospora arabidopsidis*. The only signature of this secondary structure feature in the amino acid sequence is a pair of conserved Tryptophan and Tyrosine residues (these are repeated in ATR1 which contains an additional helix). Structural analysis revealed that these conserved residues interact to form a stable hydrophobic core around which the 3-helical fold can form. Around 44% of known *Phytophthora* spp. RXLR effectors (some 500 proteins, in total) are predicted to contain this fold based on the presence of one or more WY repeats. Given that RXLR effectors typically share <20% sequence homology it has been proposed that the WY motif provides an evolutionarily conserved stable core around which functional diversification can evolve, permitting new functionality, novel target specificity and evasion of NLR recognition (Boutemy et al., 2011, Win et al., 2012). It would be interesting to obtain a crystal structure for an RXLR protein which does not contain a WY domain in order to see if there is a second conserved fold present in RXLR effectors or if the remaining effectors exhibit a great degree of structural diversity as well as sequence diversity.

## 1.8. Aims & objectives

Effector proteins and their host targets represent a potentially untapped resource for developing novel resistance strategies to pests and pathogens in crop plants and with the proliferation of multiple, tractable genome editing technologies, manipulating crops for enhanced resistance has never been more achievable. Here, we aim to add to our knowledge of the structure and function of effector proteins and

their targets from two economically significant, filamentous pathogens of important crop species, namely *P. infestans* and *B. graminis f. sp. hordei*. In addition to this, we aim to exploit existing knowledge of a host target by engineering a variant of it which is insensitive to its cognate effector protein, thus generating resistance in the host plant.

In order to do this, we address three main topics:

- Exploring the possibility of generating an effector-insensitive variant of a well characterised solanaceous host target of the *P. infestans* effector protein PexRD2 and assessing the utility of any effector-insensitive variants in conditioning disease resistance.
- Exploring the role of a PexRD2 interacting kinase – *St*MAPKKKK – in *P. infestans*-plant interactions and attempting to determine whether it functions as a target or helper of PexRD2.
- The expression, purification, crystallisation and preliminary X-ray analysis of a barley powdery milder effector protein – CSEP0162 – the first steps toward elucidating its crystal structure.

# 2

# Materials & Methods

# Chapter 2 Materials & Methods

## 2.1. Bacterial strains and plasmids

Table 3 describes the *Escherichia coli* strains used in the present work. DH5α and Stellar™ cells were used for routine DNA work and general cloning, *ccd*B Survival™ 2 cells were used for propagation of plasmids containing the *ccd*B lethal gene and SHuffle® T7 Express cells were used for heterologous production of disulphide bond-containing proteins.

| Strain | Genotype | Reference/Source |
|---|---|---|
| DH5α | F– Φ80***lac***ZΔM15 Δ(***lac***ZYA-***arg***F) U169 ***rec***A1 ***end***A1 ***hsd***R17 (rK–, mK+) ***pho***A ***sup***E44 λ– ***thi***-1***gyr***A96 ***rel***A1 | (Hanahan, 1983) |
| Stellar™ | F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ (lacZYA - argF) U169, Δ (mrr - hsdRMS - mcrBC), ΔmcrA, λ– | Clontech |
| SHuffle® T7 Express | fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(*mcr-73::miniTn10--TetS*)2 [dcm] R(*zgb-210::Tn10 --TetS*) endA1 Δgor Δ(*mcrC-mrr*)114::IS10 | New England Biolabs |
| *ccd*B Survival™ 2 | F- ***mcr***A Δ(***mrr-hsd***RMS-***mcr***BC) Φ80***lac***ZΔM15 Δ***lac***X74 ***rec***A1 ***ara***Δ139 Δ(***ara-leu***)7697 ***gal***U ***gal***K***rps***L (StrR) ***end***A1 ***nup***G ***fhuA::IS2*** | Invitrogen |

**Table 3| *E. coli* strains used in the present study.**

Table 4 describes the *Agrobacterium tumefaciens* strains used in the present work. GV3101::pMP90 (Larebeke et al., 1974, Koncz and Schell) was used for routine transient infiltration of *N. benthamiana* leaves and AGL1 (Jin et al., 1987) was used for transient expression of sgRNA-Cas9 constructs in *S. lycopersicum* as well as stable transformation of *S. lycopersicum*.

Plasmid vectors and constructs associated with the present work, whether constructed during the course of the present work or obtained from elsewhere, are listed in Appendix Tables A-D.

| Strain | Background | Marker | Ti | Marker | Opine |
|---|---|---|---|---|---|
| GV3101::pMP90 | C58 | Rif | pMP90 (pTiC58DT-DNA) | Gent | Nopaline |
| AGL1 | C58, *rec*A | Rif, Carb | pTiBo542DT-DNA | - | Succinomapine |

**Table 4| *A. tumefaciens* strains used in the present study.**

## 2.2. General chemicals

Chemical reagents were purchased from Sigma-Aldrich Company Ltd (Poole, UK) unless otherwise stated.

## 2.3. Growth media

All media was prepared with deionised water and autoclave sterilisation (30 min, 15psi, 121˚C). Liquid media was cooled to room temperature, and media containing agar cooled to ~50˚C, before addition of temperature sensitive supplements. All supplements added following autoclave sterilisation were filter sterilised using Minisart® (Sartorius) single use filter units with 0.2µm pores.

*E. coli* was grown in Lysogenic Broth Medium (LB) [1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) sodium chloride, pH 7.0] (Bertani, 1951). Where salt-sensitive antibiotics like Spectinomycin were required for selection, L Medium was used instead [1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride, 1.0% (w/v) glucose, pH 7.0]. Solid LB and L were prepared as described above with addition of 1.0% (w/v) microbiology grade agar.

Antibiotics were added to media where necessary at the following concentrations; carbenicillin (100 $\mu g.mL^{-1}$), kanamycin (50 $\mu g.mL^{-1}$), gentamicin (25 – 30 $\mu g.mL^{-1}$), rifampicin (50 $\mu g.mL^{-1}$), spectinomycin (100 $\mu g.mL^{-1}$), streptomycin (100 $\mu g.mL^{-1}$), and chloramphenicol (34 $\mu g.mL^{-1}$). Refer to Appendix B for details of resistance for individual constructs.

## 2.4. Preparation of glycerol stocks

Aliquots of cultures grown overnight at 28˚C (*S. cerevisae* and *A. tumefaciens*) or 37˚C (*E. coli*) from single, well-isolated colonies were diluted to a final concentration of 30% (v/v) glycerol in sterile, 1.5mL Eppendorf tubes and flash frozen in liquid $N_2$ before storage at -80˚C.

## 2.5. General DNA procedures

### 2.5.1. Isolation of plasmid DNA from bacteria

Following transformation or streaking from glycerol stocks, single, well-isolated colonies were used to inoculate small (5-10mL) aliquots of LB or L medium which were grown overnight (~16 h) at a temperature appropriate for the bacteria being used (see section **2.4**). Cells were harvested by centrifugation at 4,000 x g for 7 min at 4˚C in 15mL polypropylene centrifuge tubes (Corning). Cell pellets were washed by resuspension in 1x volume of $ddH_2O$ and centrifuged as before. Washed cell pellets were dried and stored at -20˚C or immediately re-suspended using re-suspension buffer supplied with the QIAprep Spin Mini-prep Kit (Qiagen) or the Isolate Plasmid Mini Kit (Bioline) and DNA was isolated in accordance with the manufacturer's guidelines. Broadly speaking, both kits use an alkali lysis method (Birnboim and Doly, 1979), neutralisation of cell lysate, clarification of cell lysate and adsorption to a silica membrane in the presence of chaotropic salts at high concentration. Plasmid DNA was washed with ethanol to remove contaminants and eluted in $dH_2O$. Plasmid DNA was stored at 4˚C in the short term to avoid multiple freeze-thaw cycles and at -20˚C for longer term storage.

For isolation of DNA from *A. tumefaciens*, the above procedure was followed but re-suspension buffer was supplemented with 100 µL of a freshly prepared stock of 10 $mg.mL^{-1}$ lysozyme to make cell lysis more efficient.

### 2.5.2. Transformation of chemically competent *E. coli*.

Chemically competent *E. coli* cells were prepared by inoculation of 5mL of LB medium with a freshly streaked, single, well isolated colony grown on LB agar. Cells were incubated ~16 h at 37˚C with shaking at 220 rpm. This overnight culture was used to inoculate 200mL LB medium which was grown until cells reached an $OD_{600}$ reading of 0.3. Cells were then chilled at 4˚C for 10 min before being harvested by centrifugation at 1,600 x g for 10 min at 4˚C. Supernatant was discarded and cells resuspended in 80 mL of filter sterilised TFB1 [30 mM potassium acetate, 10 mM calcium chloride, 100 mM rubidium chloride, 50 mM manganese chloride, 15% (v/v) glycerol, pH 5.8]. Cells were once more harvested as described above and resuspended in 8 mL of filter sterilised TFB2 [10 mM MOPS, 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol, pH 6.5]. Resuspended cells were then aliquoted out at volumes of 50-100µl into pre-chilled 1.5mL microcentrifuge tubes on ice and flash frozen in liquid $N_2$ before storage at -80˚C.

Frozen cells were thawed on ice. Immediately upon thawing, plasmid DNA was added to cell pellets which were incubated on ice for a further 20-30 min. Cells were heats-shocked in a water bath set to 42˚C for 30s before cooling on ice for 1 minute. 250-500µl SOC broth [2.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 21.3 mM magnesium chloride, 10 mM magnesium sulphate, 20 mM glucose] was added before cells were allowed to recover at 37˚C for 1 hour with shaking at 220 rpm. 70-100µl of this transformation mixture was then spread on pre-warmed LB agar plates supplemented with appropriate antibiotic and incubated at 37˚C in a static incubator for ~16 h.

This above method is based on that described in (Hanahan, 1983).

### 2.5.3. Transformation of electrocompetent *A. tumefaciens*.

All strains of electrocompetent *A. tumefaciens* used in the present study were prepared as follows: a 50mL aliquot of LB medium were inoculated with a freshly

streaked, single, well isolated colony from an LB agar plate grown at 28˚C until colonies appeared (usually 24-48hrs) and grown for ~24 h at 28˚C with shaking at 220rpm. This starter culture was then diluted 1:200 in a 1 litre culture of L medium for a further 16 h at 28˚C or until an $OD_{600}$ reading of 1.5 was achieved. This culture was split 4 ways between 4 x 1,000mL Nalgene® centrifuge bottles pre-chilled at 4˚C. These bottles were placed in a pre-chilled (4˚C) Fibrelite® F9S-4x1000y rotor and centrifuged at 5,500 x g for 15 min at 4˚C in an RC 6+ free-standing centrifuge (Thermo Scientific®/Sorvall®). Pelleted cells were resuspended and washed in a total volume of 1L pre-chilled (4˚C), sterile water before centrifugation as described previously. Cells were then resuspended in a total volume of 1L pre-chilled (4˚C), sterile water before centrifugation as before. Cell pellets were now resuspended in 40mL sterile, pre-chilled (4˚C) 10% (v/v) glycerol before being pelleted in pre-chilled (4˚C) 40mL Oakridge centrifuge tubes in a pre-chilled (4˚C) SS34 rotor (Sorvall) with the force and duration described previously in an RC 6+ free-standing centrifuge (Thermo Scientific®/Sorvall®). Pellets were resuspended in 10mL sterile, pre-chilled (4˚C) 10% (v/v) glycerol. 50µl aliquots were pipetted into pre-chilled 1.5mL microcentrifuge tubes and flash frozen in liquid $N_2$ before storage at -80˚C.

Frozen cells were thawed on ice and plasmid DNA was added immediately upon thawing. Cells and DNA were transferred to pre-chilled (on ice) electroporation cuvettes (Cell Projects). These were dried and placed in a MicroPulser Electroporator (Bio-Rad Laboratories Ltd) where they were subjected to an electrical pulse at 2.20 kV. Electroporated transformation mix was resuspended in 250-500µl of SOC broth (see section 2.5.2) and allowed to recover for 1 hour at 28˚C with shaking at 220 rpm. 70-100µl of recovered cells were then plated on pre-warmed L agar plates supplemented with appropriate antibiotic and incubated in a static incubator at 28˚C until single colonies appeared (usually 36-48 h).

### 2.5.4. Polymerase Chain Reaction

Standard PCR was performed using PfuTurbo® DNA polymerase (Stratagene), VELOCITY® DNA polymerase (Bioline) or Phusion® DNA polymerase (New England Biolabs) using reagents and buffers provided by the manufacturers. Typical

reaction mixtures contained 1x reaction buffer, 0.2mM of each dNTP, 0.1-0.5µM of each oligonucleotide primer and 1U of DNA polymerase. Template DNA was added at an appropriate amount dependent on its source and made up to 20µl, 25µl or 50µl dependent on application. Additives were used to enhance the efficiency of difficult to optimise PCRs such as 1M betaine or 5% (v/v) sterile DMSO. Components were assembled in an *n*+1x master mix on ice before being placed in a pre-heated Biometra T3000 Thermocycler (Biometra, Göttingen, Germany). Typical amplification proceeded according to the conditions shown in Table 6. For random mutagenesis PCR, manganese sulphate concentration was altered in the reaction mix as per the guidelines in the Diversify® (Clontech) user manual.

#### 2.5.4.1 **Oligonucleotide synthesis**

Oligonucleotide primers were synthesised by Integrated DNA Technologies, Inc. (Belgium). Primers were supplied lyophilised and were resuspended in sterile water upon delivery to a final concentration of 100 µM, prior to dilution to a working concentration 3 – 20 µM, and stored at -20 ºC. All oligonucleotide primers used in the present work are listed in Appendix A alongside their Banfield lab ID number with a brief description of their use.

2.5.4.2 **Colony PCR**

In order to confirm the presence of a construct of interest in transformed bacterial colonies, small amounts of single, well isolated colonies were picked with pipette tips and transferred to PCR reaction tubes with reagents already added. Cells were resuspended by gentle pipetting and lysed with an initial lysis step of 95˚C for 10 min. Colony PCR proceeded according to the thermocycler shown in Table 5.

| | | | |
|---|---|---|---|
| **Activation/ Cell lysis** | 95 ºC | 10 min | 1 x |
| **Denaturation** | 95 ºC | 30 s | 35 x |
| **Primer annealing*** | 45 - 55 ºC | 60 s | |
| **Primer extension** | 60 ºC | 2 min | |
| **Final extension** | 68 ºC | 5 min | 1 x |
| **Holding step** | 8 ºC | ∞ | 1 x |

**Table 5| General colony PCR cycle.**

| | PfuTurbo® DNA polymerase | | | VELOCITY DNA polymerase | | | Phusion® DNA polymerase | | |
|---|---|---|---|---|---|---|---|---|---|
| **Amount of template DNA per 50 µL reaction** | | | | | | | | | |
| **genomic DNA / cDNA** | 50 – 100 ng | | | 5 - 200 ng | | | 50 - 250 ng | | |
| **plasmid DNA** | | | | 50 pg- 50 ng | | | 1 pg- 10 ng | | |
| **PCR reaction conditions** | | | | | | | | | |
| **Activation** | 95 ºC | 2 min | 1 x | 98 ºC | 4 min | 1 x | 98 ºC | 30 s | 1 x |
| **Denaturation** | 95 ºC | 60 s | 30 x | 98 ºC | 30 s | 40 x | 98 ºC | 5 s | 35 x |
| **Primer annealing*** | 45 - 69 ºC | 60 s | | 50 - 68 ºC | 30 s | | 45-72˚C | 5 s | |
| **Primer extension** | 72 ºC | 60 s per kb | | 72 º | 15 – 30 s per kb | | 72 ºC | 15-30 s per kb | |
| **Final extension** | 72 ºC | 10 min | 1 x | 72 ºC | 4 min | 1 x | 72 ºC | 5-10 min | 1 x |
| **Holding step** | 4 ºC | ∞ | 1 x | 10 ºC | ∞ | 1 x | 4 ºC | ∞ | 1 x |

**Table 6| General PCR conditions.**

* Primer annealing temperature was 2 – 5 ºC below the predicted melting temperature (Tm) for the primer with the lowest value in a given reaction. Tm values were calculated using the OligoAnalyzer 3.1 tool on the IDT) website (http://eu.idtdna.com/analyzer/applications/oligoanalyzer/).

### 2.5.5. cDNA synthesis

Total RNA was extracted from tomato leaves using the Qiagen RNeasy Plant Mini Kit as per the manufacturer's guidance and an on-column DNaseI digest was carried out using RNase-free DNaseI (Qiagen) to remove contaminating gDNA. All gloves, pipettes, tip boxes, reagent bottles and surfaces were wiped down with RNaseZap (Ambion) prior to starting the procedure and regularly throughout.

Total RNA (1µg), gene specific RT primer (3µM), 0.2mM each dNTP and $dH_2O$ to a final volume of 12µL were incubated at 65˚C for 5 min and chilled on ice. 4µL of Superscript II first strand buffer and 2µL DTT were added to each reaction before incubation at 42˚C for 2 min. 1µL Superscript II reverse-transcriptase was added to each reaction in addition to $dH_2O$ before incubation at 42˚C for 15 min and at 75˚C for 10 min. cDNA was used in standard PCR using gene specific primers.

### 2.5.6. Agarose gel electrophoresis

Typically, gels of 1% (w/v) molecular biology grade agarose (Melford Biolaboratories Ltd.) in TAE buffer [40 mM Tris-acetate, 1.0 mM EDTA, pH 8.0] supplemented with 0.5µg.mL$^{-1}$ ethidium bromide were cast and allowed to set before submersion in TAE buffer. DNA samples were premixed 3:1 (v:v) with 4x 'FOG' DNA loading buffer [12% (w/v) Ficoll 400 and 0.25% (w/v) Orange G]. Samples were loaded alongside a molecular weight marker appropriate for the predicted product size [1 kb Plus DNA ladder (Invitrogen), or Low Molecular Weight DNA Ladder (New England Biolabs)]. Gels were run at a constant voltage of 90-100V until tracker dye had run almost to the edge of the gel. DNA was visualised using a transilluminator emitting UV light at $\lambda = 365$nm.

For detection of CRISPR/Cas9 mediated deletions in tomato gDNA by the PCR-RE assay described in section 2.5.10, 4% (w/v) Agarose gels were used.

#### 2.5.6.1 **PCR cleanup**

PCR cleanup was performed using the QIAquick PCR Purification Kit (Qiagen) or NucleoSpin® Extract II Kit (MACHEREY-NAGEL) as per the manufacturer's protocol. The protocols rely on similar principles to extraction of DNA from bacterial cell lysate: adsorbtion to a silica membrane in the presence of chaotropic salts followed by elution in sterile water following multiple washes with ethanol to remove contaminants.

#### 2.5.6.2 **Gel extraction**

Purification was carried out using the Zymoclean™ Gel DNA Recovery kit (Zymo Research) as per the manufacturer's protocol following agarose gel electrophoresis as described in section 2.5.6 and excision of a well separated band with a sterile razor blade. The protocol relies on similar principles to those outlined for isolation of plasmid DNA and PCR clean up.

### 2.5.7. Site-directed mutagenesis

All mutant constructs used in the present study were either obtained from previous publications and are listed in Appendix B or made to order by Genscript (USA) or Genewiz (USA) by mutagenesis of a user supplied DNA template or synthesis of a codon optimised sequence flanked by *att*B recombination sites. Mutants were supplied in pUC57 vector backbones except for mutant MAPKKK$_{\varepsilon}$ sequences which were sent to Genscript (USA) in the bespoke pERCH vector featuring a β-estradiol inducible promoter.

### 2.5.8. DNA sequencing

From commencement of the present work until January 2013, DNA sequencing was submitted to Genome Enterprise Ltd (Norwich Research Park). ~100ng of DNA was mixed with 2 μL of 5 x sequencing buffer [350 mM Tris-HCl pH 8.8, 2.5 mM

magnesium chloride], 1 μL ABI Big Dye Terminator Ready Reaction Mix Ver 3.1 (Invitrogen), 1 μL of 30% (v/v) DMSO, and 1 μL of sequencing primer (5 – 10 μM) in a total volume of 10 μL. Reactions were mixed and incubated in a Thermocycler using the conditions described in Table 6 for Velocity polymerase. Reactions were analysed on an ABI 3730xl sequencer. Upon cessation of the above locally provided service in January 2013, DNA sequencing work was submitted to GATC Biotech. Ltd. (The London BioScience Innovation Centre, London, UK). 400-500ng of plasmid DNA was sent with custom sequencing primer at a final concentration of 10pmol.$\mu L^{-1}$ in 10μL total final volume.

## 2.5.9. Cloning procedures

### 2.5.9.1 **TOPO® cloning**

Purified PCR products encoding *CACC* at their 5' terminus derived from PCR primers were cloned into the pENTR™/D-TOPO® vector as per the manufacturer's protocol.

### 2.5.9.2 **Gateway cloning**

Gateway cloning was performed using Gateway® BP Clonase™ II Enzyme mix (Invitrogen) to generate Entry (ENTR) constructs or Gateway® LR Clonase™ II Enzyme mix (Invitrogen) to generate Destination/Expression (DEST/EXP) constructs. All reactions were performed as per the manufacturer's protocol.

### 2.5.9.3 **Golden Gateway assembly**

Preparation of *S. lycopersicum* editing and complementation constructs was performed by Mr Mark Youles (The Sainsbury Laboratory, Norwich, UK) with guidance from Dr Nicola Patron (The Sainsbury Laboratory, Norwich, UK) as part of The Sainsbury Laboratory's Synthetic Biology pipeline. Complementation cassettes were synthesised by Genewiz (USA) with sgRNA recognition sites re-coded – where appropriate - with synonymous mutations to abolish recognition and

were supplied in pUC57 with flanking BsaI sites for direct use in Golden Gateway assembly reactions.

Components were assembled in a one pot reaction containing 100ng of the acceptor plasmid and each individual module in a 1:2 ratio. In addition, 1.5µL of T4 ligase buffer (New England Biolabs), 1.5µL of Bovine Serum Albumin (New England Biolabs) (at 10x concentration), 200U[7] of T4 DNA Ligase (New England Biolabs) and 5U[8] of BsaI (Eco31) (ThermoFisher Scientific). Assembled components were incubated in a pre-heated Biometra T3000 Thermocycler (Biometra, Göttingen, Germany) and subjected to the protocol outlined in Table 7.

#### 2.5.9.4 **In-Fusion® PCR cloning**

pOPINF:CSEP0162 was constructed by Miss Agnieszka Siwoszek (Univ. Copenhagen) using In-Fusion® PCR cloning (Clontech) following PCR with forward and reverse primers designed to generate 5' and 3' flanking regions homologous to the KpnI/HindIII linearised pOPINF vector. 3'-5' proof reading exonuclease activity of the proprietary In-Fusion® enzyme creates single stranded overhangs in the linearised vector with homology to the newly generated flanking sequences of the coding sequence allowing joining of the linearised vector and PCR product.

[7] 1 T4 Ligase U defined as "the amount of enzyme required to give 50% ligation of HindIII fragments of λ DNA (5´ DNA termini concentration of 0.12 µM, 300- µg/ml) in a total reaction volume of 20 µl in 30 min at 16°C in 1X T4 DNA Ligase Reaction Buffer." (Direct quote from NEB website).

[8] 1 BsaI U defined as "the amount of Eco31I required to digest 1 µg of lambda DNA dcm–-HindIII fragments in 1 hour at 37°C in 50 µL of recommended reaction buffer." (Direct quote from Thermo Fisher Scientific website).

| | | |
|---|---|---|
| 30 ºC | 20s | 1 x |
| 37 ºC | 3 min | 26 x |
| 16 ºC | 4 min | |
| 50 ºC | 5 min | 1 x |
| 80 ºC | 5 min | 1 x |
| 16 ºC | ∞ | 1 x |

**Table 7| Thermocycle for one-pot Golden Gate assembly.**

## 2.5.10. PCR-Restriction Enzyme Assay

To enrich for edited DNA when transiently testing sgRNAs and screening putative edited tomato plants described in Chapter 3, a PCR-RE enzyme assay approach was used. Tomato gDNA extracted as described in section 2.10.5 was used. gDNA concentration was normalised to 100ng/µL. 5µL of normalised gDNA was mixed with SuRE Cut Buffer B (Roche) at 1x concentration, 5 units of EcoRV (Roche) and $dH_2O$ to a total reaction volume of 25µL. Reactions were incubated at 37˚C for 1 hour and taken straight to PCR.

1µL of each reaction was used for PCR. To this, 0.2mM of each dNTP, 0.5µM forward and reverse primer (see Appendix A), 5% DMSO, 1U[9] of Q5-HF DNA polymerase (New England Biolabs) and 1x Q5 reaction buffer were added. Components were assembled on ice before transfer to a Biometra T3000 Thermocycler (Biometra, Göttingen, Germany) where the cycle shown in Table 6 was used.[10]

[9] 1 Q5-HF polymerase U defined as "the amount of enzyme that will incorporate 10 nmol of dNTP into acid insoluble material in 30 min at 74°C." (Direct quote from NEB website).

[10] Whilst Q5-HF polymerase was used for this reaction, the same thermocycler used for Phusion polymerase was used.

PCR products were run on a 4% (w/v) agarose gel as described in section 2.5.6 for 1 hour.

## 2.6. Protein procedures

### 2.6.1. Discontinuous sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for the separation of proteins

Separation of proteins by discontinuous SDS-PAGE has previously been described in (Laemmli, 1970). Home-made gels used in the present study contained 17% (w/v) acrylamide solution unless explicitly stated. Gels were prepared by dilution of the appropriate volume of a 30% (w/v) stock solution of acylamide/bis-acrylamide (37.5:1) in 375 mM Tris-HCl pH 8.8 and 0.1% (w/v) SDS, supplemented with 0.1% (w/v) ammonium persulphate, and 0.04% (v/v) N,N,N',N'-tetramethylethylenediamine and mixing thoroughly in a conical flask (which was not added until immediately before pouring). The resulting resolving gel solution was poured between Bio-Rad Mini-PROTEAN® system glass plates until they were filled to ~2cm below the top line of the short plate. A small amount of water-saturated butanol was pipetted over the surface of the unset gel in order to create a smooth boundary between the resolving gel and stacking gel. Once set, a 5% (w/v) stacking gel was prepared by diluting the appropriate volume of a 30% (w/v) stock solution of acylamide/bis-acrylamide in 63 mM Tris-HCl pH 6.8 and 0.1% (w/v) SDS, supplemented with 0.1% (w/v) ammonium persulphate, and 0.1% (v/v) N,N,N',N'-tetramethylethylenediamine and mixing thoroughly in a conical flask (once again, added only immediately prior to pouring). This stacking gel solution was poured over the set resolving gel (having removed the water-saturated butanol). Plastic combs were inserted into the unset stacking gel solution in order to create wells for sample loading. Gels were used immediately or stored at 4˚C individually wrapped in cling film, inside a Tupperware box sandwiched between two sheets of damp

paper towels. In addition, pre-cast Mini-PROTEAN® TGX™ gels (Bio-Rad) were also used.

Cast gels were run in a Mini-PROTEAN® Tetra Cell system (Bio-Rad) within Mini-PROTEAN® Tetra electrode assembly units or Tetra companion running modules (Bio-Rad) and plastic combs were removed. A Mini Cell buffer dam (Bio-Rad) was used if necessary. The Tetra Cell tank was filled up to the recommended line for the number of gels being run and gel chambers were filled in order to totally cover the gels with with SDS-running buffer [25 mM Tris-HCl, 250 mM glycine, and 0.1% (w/v) SDS]. Protein samples were prepared by incubation at 95˚C for 5-10 min in 4x SDS gel-loading buffer [0.2 M Tris-HCl pH 6.8, 0.4 M DDT, 8 % (w/v) SDS, 0.4% bromophenol blue, 40% glycerol] before being loaded on stacking gels alongside an appropriate molecular weight marker [SeeBlue® Plus2 Pre-Stained Standard (Novex), PageRuler™ Prestained Protein Ladder (Fermentas) or RunBlue™ (Expedeon) Prestained Protein Ladder]. Gels were run at a constant voltage of 100-150V for 45-120 min.

Protein was visualised using InstantBlue™ (Expedeon) in accordance with the manufacturer's guidelines.

### 2.6.2. Expression of native recombinant proteins in *E. coli* and harvesting of cells.

Overnight starter cultures were set up using a single, freshly-transformed or streaked, well-isolated colony of *E. coli* (SHuffle® T7 Express) harbouring pOPINF:CSEP0162 from an LB agar plate supplemented with carbenicillin and incubated overnight at 37˚C in a static incubator to inoculate 10mL of LB medium supplemented with carbenicillin. Starter cultures were grown for ~16 h at 30˚C with shaking at 220rpm and used to inoculate 1L cultures of LB supplemented with carbenicillin in 2L flasks. 1L cultures were grown at 30˚C with shaking at 220rpm until an $OD_{600}$ reading of 0.6 was reached. At this point, cultures were induced with 1mM IPTG and incubated at 18˚C for 18 h to allow protein to accumulate.

Cells were harvested in pre-chilled (4˚C) 1000mL Nalgene® centrifuge bottles which were spun in a pre-chilled (4˚C) Fibrelite® F9S-4x1000y (Thermo Scientific) rotor at 5,500 x g for 7 min at 4˚C in an RC6+ Centrifuge (Thermo Scientific/Sorvall). Harvested cell pellets were stored at -80˚C or immediately moved forward to the protein purification procedure described in section 2.6.4.

### 2.6.3. Expression of Selenomethionine-derivative protein in *E. coli* using metabolic inhibition of methionine biosynthesis

Starter cultures were initiated as described in section 2.5.2. 500mL aliquots of metabolic inhibition media [1x M9 minimal salts, 0.2% (w/v) glucose, 2mM $MgSO_4$, 0.1mM $CaCl_2$, 0.0001% (w/v) Thiamine, 1x Carbenicillin, 20mg of the following amino acids; valine, phenylalanine, isoleucine, leucine, aspartic acid, glutamic acid, lysine, arginine, serine, threonine, tyrosine, histidine, glutamine and tryptophan made up to 500mL with Milli-Q (MQ) water] were prepared. Starter cultures were centrifuged at 4000 x g for 7 min and washed in one volume of the previously described metabolic inhibition media. This was repeated a further two times. Washed, resuspended cells were then used to inoculate 500mL of media and grown at 30˚C until an $OD_{600}$ reading of 0.4-0.6 was reached. A 3mL sample of un-induced culture was taken, centrifuged at 16,000 x g for 10 min, re-suspended in SDS buffer and prepared for SDS-PAGE analysis as described in section 2.6.1. 50mg each of Threonine, Lysine and Phenylalanine, 25mg each of Leucine, Isoleucine and Valine in addition to 30mg of L-selenomethionine were added per 500mL of culture. Cultures were incubated at 30˚C with shaking at 220rpm for a further 45 min before induction with 1mM IPTG. Cultures were then incubated at 18˚C for 18 h to allow accumulation of Selenomethionine-derivative CSEP0162.

Cells were harvested as described in section 2.6.2.

## 2.6.4. Purification of recombinant proteins from *E. coli*

### 2.6.4.1 **Extraction of protein from cells and purification of soluble fraction**

Cells were thawed (if required) and re-suspended in 25mL ice cold buffer A1 [50 mM Tris-HCl pH 8.0, 50 mM glycine, 5% (v/v) glycerol, 500 mM sodium chloride, 20 mM imidazole] per 1L of original culture, supplemented with one EDTA free protease inhibitor tablet per 50mL buffer on ice. Cells were lysed by sonication with a Vibra-Cell™ 750 Watt ultrasonic processor, VC 750, (Sonics & Materials, Inc., Newtown, CT , USA) at a maximum 40% amplitude with sonication on for 2s followed by 3s of no sonication to avoid overheating of cell lysate. Total time spent with sonication 'on' was 5 min and cells were kept on ice to reduce overheating. A 3µL sample of unclarified lysate was retained and prepared for SDS-PAGE analysis as described in section 2.6.1. Cell debris was pelleted by centrifugation in a 50mL Oak Ridge centrifuge tube (Nalgene) at 18,000 x g for 15 min, at 4˚C in a pre-chilled SS34 rotor (Sorvall) in an RC6+ free-standing centrifuge (Thermo Scientific/Sorvall). Semi-clarified lysate was carefully poured off into another Oakridge centrifuge tube and subjected to the same centrifugation step to further clarify cell lysate. A 3µL sample of this clarified lysate was retained and prepared for SDS-PAGE analysis as described in section 2.6.1.

### 2.6.4.2 **Purification of hexa-histidine tagged protein from soluble fraction and size exclusion chromatography**

Clarified lysate was loaded onto a 5mL Nickel (II) ($Ni^{2+}$) charged HisTrap FF column (GE Healthcare) which had been pre-equilibrated with buffer A1 using an automated ÄKTAxpress (GE Healthcare) purification system. Columns were washed with 10 column volumes of buffer A1 prior to step elution with buffer B1 [50 mM Tris-HCl pH 8.0, 50 mM glycine, 5% (v/v) glycerol, 500 mM sodium chloride, 500 mM imidazole]. Eluate was immediately loaded onto a Hi-Load 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare) pre-equlibrated and run in A4 buffer [50 mM HEPES, 150 mM sodium chloride, pH 7.5] on the ÄKTAxpress (GE Healthcare) system for purification by size exclusion chromatography (SEC).

Columns were run at a flow rate of ~3mL.min$^{-1}$ with total flow through the column of ~2 column volumes. Absorbance at 280nm was continuously monitored and fractions spanning peak absorbance were prepared for analysis by SDS-PAGE as described in section 2.6.1 alongside pre-induction, whole cell lysate and cleared lysate samples taken earlier.

### 2.6.5. Cleavage of hexa-histidine affinity tag with 3C protease

Fractions containing soluble protein were pooled and concentrated using an ultrafiltration centrifuge cell (Sartorius) of the appropriate molecular weight cut off and spun according to the manufacturer's recommendation to a concentration of 1-2mg/mL. A 3µL sample was retained and prepared for SDS-PAGE analysis. 12µg of recombinant human rhinovirus (HRV) 3C protease per mg of tagged protein was added to protein in A4 buffer (see section 2.6.4.2) and incubated at 4˚C for ~16 h. HRV 3C protease cleaves at the following site: LEVLFQ▼GP leaving the short glycine-proline dipeptide attached to the protein of interest. The protease is also hexa-histidine tagged so as to allow separation from the solution containing protein of interest post-cleavage.

Following incubation, a 3µL sample was retained and prepared for SDS-PAGE analysis, before the solution was manually applied to a 5mL Nickel (II) (Ni2+) charged HisTrap FF column (GE Healthcare) pre-equilibrated with buffer A1, followed by washing with 5 column volumes of buffer A1. Flow through, containing cleaved protein, was collected. Cleaved hexa-histidine tag and 3C protease were eluted by washing with 5 column volumes of buffer B1. Solution containing protein of interest was applied manually using a 10mL syringe (Terumo) and all washes were applied using a peristaltic pump (GE Healthcare).

Flow through containing cleaved protein of interest was concentrated as described previously to a total volume of <10mL and injected manually onto a Hi Load 26/60 Superdex 75 gel filtration column (GE Healthcare) as described in section 2.6.4.2.

In the case of SeMet-derivative CSEP0162, cleaved protein was not subjected to this final, preparative gel filtration step and was instead concentrated and buffer exchanged (A1 to A4) by concentrating and replacing with 1 original volume of buffer A4. This was repeated 5 times in order to fully exchange buffers.

## 2.6.6. LC-MS/MS analysis of FLAG-Tagged effector proteins extracted from *N. benthamiana* for post-translational modification.

### 2.6.6.1 **Preparation of gel slices for trypsin digest.**

Gels were stained as described in section 2.6.1 for ~20 min or until the band of interest became clearly visible. Bands were excised from gels using a sterile scalpel blade. Gel slices were de-stained in 30% (v/v) ethanol for 30 min at 65˚C with shaking. Ethanol was removed and gel slices were de-stained further in 30% (v/v) ethanol until staining was totally removed. Ethanol was removed and gel slices were washed in 50% (v/v) acetonitrile in 50mM TEAB [Triethylammonium bicarbonate pH 8.5] for 15 min at room temperature. Previous buffer was removed and gel slices were washed in 10mM DTT in 50mM TEAB for 30 min at 55˚C. Previous buffer was removed and gel slices were incubated with 30mM Iodacetamide in 50mM TEAB at room temperature for 30 min, covered with aluminium foil. Gel slices were vortexed in IAA/TEAB buffer and buffer was removed. Gel slices were washed with 50% (v/v) acetonitrile in 50mM TEAB for 15 min at room temperature. Buffer was removed and gel slices were washed in 50mM TEAB for 15 min at room temperature. Buffer was removed and gel slices were cut into ~1mm$^3$ pieces in sterile petri dishes before transfer to 1.5mL, low bind eppendorf style tubes. Gel slice pieces were washed in 50% (v/v) acetonitrile in 50mM TEAB for 15 min at room temperature. Buffer was removed and gel slice pieces were washed in 100% (v/v) acetonitrile for 15 min at room temperature. Buffer was removed and gel slice pieces were dried in a speed vac (manufacturer) for 30 min prior to submission to the JIC proteomics platform for trypsin digest.

#### 2.6.6.2 **Orbitrap mass spectrometry of in-gel digested peptides**

Protein extracts were prepared as described in 2.10.3 and epitope tagged proteins were enriched using the immunoprecipitation procedure described in 2.10.5.

Samples were analysed by nanoLC-MS-MS on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer with an UltiMate® 3000 RSLCnano LC system (Thermo Scientific). Sample was separated on a PepMap™ 100 C18 LC Column (Thermo Scientific) and infused directly into the mass spectrometer which was set to positive ion mode with quadrupole isolation. Quadrupole isolation was set to 120K resolution with a mass-to-charge ratio of 350-1550 (m/z) for scanning of precursor ions. Precursors were fragmented by collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) with an isolation window of 1.6Da. Spectra were analysed with MaxQuant 1.5.1.2 and database search – using merged CID and HCD peaklists – performed with Mascot 2.4. Database search was performed on a *P. infestans* protein database (Uniprot) with a tolerance of 6ppm for precursors and for peptides of 0.6Da. Identification probabilities of 99% and 95% for proteins and peptides were used when search results were imported into Scaffold 4.4.1.1.

## 2.6.7. Western blotting

All western blotting in the present work was carried out using a wet blotting protocol in a Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad). Two pieces, per gel, of 3mm filter paper (Whatman) were cut to the size of SDS-PAGE gels and pre-soaked in Bjerrum transfer buffer [48 mM Tris, 39 mM glycine, pH 9.3, 20% methanol] with the addition of 0.0125% (w/v) SDS (Bjerrum et al., 1987) alongside two foam pads. One piece, per gel, of Immobilon-P PVDF membrane (Merck Millipore) was cut to the size of SDS-PAGE gels and incubated with pure methanol for 10-15 min prior to blotting in order to activate the hydrophobic PVDF membrane and allow aqueous transfer buffer to fully saturate it, facilitating efficient protein transfer. At this point, the transfer module was assembled between a gel holder in the following order: one foam pad, one piece of blotting paper, gel containing proteins of interest, PVDF membrane, second piece of blotting paper, second foam pad. A blot roller was used to remove air bubbles between gel and membrane. The assembled blotting module was placed in a Mini Trans-Blot® cell, placed inside a Mini PROTEAN® Tetra Cell system tank along with an ice pack and magnetic stirring bar. The tank was filled with ice cold Bjerrum buffer until buffer was level with the top of PVDF membranes. Transfer was conducted on a magnetic stirring platform over the course of one hour with a current of 250mA.

Transfer and loading was assessed by staining with Ponceau S. Immediately following transfer, membranes were incubated with pure methanol for ~30s. Membranes were then incubated with Ponceau S stain [0.1% (w/v) Ponceau S in 5% (v/v) acetic acid] for ~5 min and repeatedly rinsed with MQ $H_2O$ until clear bands were observable with little to no background staining. Staining was reversed by washing with 0.1M NaOH and MQ $H_2O$.

### 2.6.7.1 **Immuno-detection of proteins immobilised on nitrocellulose membranes following western blotting**

Subsequent to transfer and Ponceau S staining, membranes were incubated ~16 h at 4˚C with shaking at ~100rpm in blocking solution [5% (w/v) dried skimmed milk

powder in 0.1% TBS-T buffer (50 mM Tris-HCl, 200 mM sodium chloride, pH 7.4, supplemented with 0.1% (v/v) Tween®-20)]. Blocking solution was rinsed away with three short washes in ~10mL 0.1% TBS-T prior to 3 more rigorous wash steps of 5 min each, in TBS-T. Membranes were then probed with a primary antibody for the protein to be detected with incubation at 4˚C for 1-2 h in 0.1% TBS-T. For one-step detection, membranes were washed as previously described and appropriate detection agents applied. For two-step detections, membranes were washed and probed with the secondary antibody in 0.1% TBS-T at 4˚C for 1-2 h. See Table 8 for details of antibodies used in the present study.

Horseradish peroxidase conjugated antibodies were detected by application of SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) to membranes, supplemented with 10% (v/v) SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) to increase sensitivity according to manufacturer's instructions. Typically this involved 1 min incubation at room temperature followed by removal of excess substrate and detection with an Image Quant LAS500 (GE Healthcare) set to automatically detect chemiluminescent signal at an appropriate exposure time. Occasionally, detection was carried out by exposing Fuji Medical X-Ray Film (Fuji) to membranes and incubating for an appropriate length of time inside a light blocking film cassette, in a dark room before development.

Alkaline phosphatase (AP) conjugated anti-FLAG antibodies were detected by treatment with BCIP/NBT colour development substrate (Promega) diluted in Alkaline Phosphatase Buffer [100mM Tris-HCl pH 9.0, 150mM NaCl, 1mM $MgCl_2$] in accordance with the manufacturer's guidelines. In this assay, colour develops on the membrane so no detection step is necessary but the reaction is irreversible so this procedure was invariably carried out after the detection of any other antibodies.

| Antibody | Supplier | Origin | Working dilution | Buffer | Incubation conditions | Detection reagents |
|---|---|---|---|---|---|---|
| **Primary antibodies** | | | | | | |
| **α-HA** | (3F10) Roche | rat monoclonal | 1:6,000 | 0.1% TBS-T | RT for 1 h | n/a |
| **α-FLAG-AP*** | (A9469) Sigma | mouse monoclonal | 1:1,000 | 0.05% TBS-T | RT for 1 h | AP |
| **α-GAL4$_{DBD}$-HRP*** | (RK5C1) Santa Cruz Biotechnology | mouse monoclonal | 1:1,000 | 0.1% TBS-T | RT for 2 h | Pico |
| **α-GAL4$_{AD}$-HRP*** | Santa Cruz Biotechnology | mouse monoclonal | 1:1,000 | 0.1% TBS-T | RT for 2 h | Femto |
| **Secondary antibodies** | | | | | | |
| **α-mouse-HRP** | (A4416) Sigma | goat | 1:20,000 | 0.1% TBS-T | RT for 1 h | Pico |
| **α-rat-HRP** | (A9037) Sigma | goat | 1:20,000 | 0.1% TBS-T | RT for 1 h | Pico or Pico/Femto |

**Table 8| Antibodies used in the present study.**

(Note – '*' indicates antibodies suitable for one-step detection. All other antibodies were used in two-step detections).

### 2.6.8. Circular Dichroism Spectroscopy

Circular Dichroism (CD) spectroscopy was carried out as described (Walden et al., 2014) in a Chirascan-plus CD spectrophotometer (Applied Photophysics). Protein of interest was diluted to ~0.4-0.8mg/mL in 20mM $K_2HPO_4$. Sample was pipetted into a 0.5mm quartz glass CD cell (Hellma) and data was collected at a constant temperature of 20˚C. Instrumentation was blanked with a buffer only control in triplicate, at wavelengths from 180nm to 260nm at intervals of 1nm. The mean absorbance of this blank was subtracted from averaged CD spectra collected from protein of interest. CD spectra of protein of interest were measured by the same strategy, with measurements taken in triplicate and averaged. As described, raw data was corrected by subtracting the baseline derived from a buffer only sample and

converted to Mean Residue Ellipticity (MRE) using the equation described in section 5.2.2.

In order to assign secondary structure elements, the averaged CD spectrum was submitted to the DichroWeb server [11]where the Cdsstr method – using reference set 7 – was applied to the data (Whitmore and Wallace, 2008, Sreerama et al., 1999, Sreerama and Woody, 2000).

## 2.7. Crystallographic methods.

### 2.7.1. Crystallisation trials

Initial screening was conducted with protein in buffer A4 - purified and concentrated as described in section 2.6.4. Prior to use in crystallographic trials, protein solution was centrifuged at 17,000 x *g* for 20 min in order to sediment any insoluble material which may interfere with crystallisation. Protein concentration was determined by measuring $A_{280nm}$ using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and predicting the protein's percent solution extinction coefficient using ProtParam on the ExPASy proteomics server (Gasteiger et al., 2005) which were used in equation 2.1.

**Equation 2.1**

$$\frac{\text{A280nm of sample}}{\text{percent solution extinction coefficient}} \times \frac{1}{\text{dilution factor}}$$

Trials were performed in a 96 well format with two protein drops per condition allowing two protein concentrations per condition to be screened simultaneously. MRC 2-well crystallization plates (Molecular Dimensions) were used. Precipitant wells were filled with 40µL of precipitant solution from the screens described in

[11] http://dichroweb.cryst.bbk.ac.uk/html/home.shtml

Table 9**Error! Reference source not found.** which was dispensed using a Liquidator™ 96 (RainIn). 0.3µL aliquots of concentrated protein were mixed with equal volumes of each precipitant and dispensed into protein wells by an Oryx Nano robot (Douglas Instruments) under a rolling evaporation shield. Plates were sealed with adhesive film and stored in a Minstrel Plate Hotel (Rigaku), at 18˚C, which records visible and UV light images of

protein wells immediately upon loading and at user-defined intervals indefinitely.

| Screen | Supplier |
|---|---|
| **JCSG-plus™** | Molecular Dimensions (MD1-37) |
| **Keep It Simple Screen (KISS)** | Designed by Dr Clare Stevenson (JIC) and supplied by Molecular Dimensions (Custom Screen Service) |

**Table 9| Crystallographic screens used in the present study.**

## 2.7.2. Optimisation

Initial hits from trials were invariably harvested for data collection but in order to grow more, or even better, crystals, conditions yielding crystals were optimised. Optimisation screens were carried out in a 48 well format using MRC Maxi plates (SwissSci) with a typical strategy of varying two conditions from the initial hit along the *x* and and *y* axes of the plate. For example, if crystallisation was achieved in 1M ammonium sulphate, 5% (w/v) PEG 3350 and 0.1M Bis-Tris (pH 5.5) then Ammonium Sulphate concentration may be varied from 0.8M to 1.2M along one row of wells with PEG3350 concentration being varied from 0% to 10% (w/v) along the other, in equal increments. Buffer and overall pH were kept constant. Each precipitant well contained 200µL precipitant and each protein drop well contained 0.3µL protein solution mixed with an equal volume of precipitant solution. Precipitant wells were prepared using an Oryx Nano robot (Douglas Instruments) as

were protein drop wells under a rolling evaporation shield. Plates were sealed and monitored in a plate hotel as described in section 2.7.1.

### 2.7.3. Harvesting, heavy atom soaking and cryoprotection

Crystals were physically harvested and mounted on loops by Dr. Miriam Walden and Prof. Mark Banfield.

Crystals must be cryoprotected in order to protect against the formation of crystalline ice. Ice affects crystallographic experiments in two ways; firstly, it disrupts the internal order of protein crystals through expansion into solvent channels and, secondly, it produces confounding ice rings in diffraction patterns, obfuscating any potential solution. In the present study, 25% (w/v) ethylene glycol was added to the mother liquor and heavy atom solution. Ethylene glycol acts as a cryprotectant by occupying the solvent channels of the protein crystal, preventing crystalline ice from disrupting the crystal's order. Crystals were soaked in the solution containing mother liquor, cryoprotectant (and heavy atom if soaking) for ~30s before being frozen by immersion in liquid $N_2$.

### 2.7.4. Data collection

All datasets were collected at the Diamond Light Source (DLS) synchrotron facility in Oxfordshire. Native and iodine derivative datasets were collected on the i02 beamline whilst SeMet derivative datasets were collected on the i04-1 beamline. Crystals were cryo-cooled during data collection.

Native data was collected at λ=0.9795Å, iodine derivative data was collected at λ=1.8Å and SeMet data was collected at λ=0.9282Å. 5 test images per crystal were collected at a maximum resolution of 2.2Å, 2.6Å and 3.2Å for native, iodine and SeMet datsets, respectively, with Ω=0.15˚ rotation angle separated by 45˚ rotation. These test images were used to allow the EDNA auto-processing software in the

DLS pipeline to determine likely diffraction resolutions and strategies for data collection.

### 2.7.5. Data processing

Datasets were processed through the XIA2 pipeline, using XDS, POINTLESS and XSCALE to index, reduce and scale data, in addition to assigning a likely space group.

XDS indexes data by identifying diffraction spots in diffraction images through identification of pixel intensities and labelling them with miller indices (h k l). XDS also makes inferences about crystal geometry (i.e. dimensions of unit cell, orientation and point group) by predicting where Bragg's law is satisfied in reciprocal space and where one may predict a reflection to appear. These are used by POINTLESS to set up constraints for assignment of Laue groups, which is done through assessment of symmetry-related reflections and systemic absences (where diffraction spots do not appear where predicted by XDS) in the unmerged data produced by XDS. XSCALE scales data from multiple images over multiple sweeps from one – or more – crystals into one set of structure factors normalised according to their symmetry.

### 2.7.6. Attempts to obtain phase information

Solving a crystal structure is essentially synonymous with solving 'the phase problem' which refers to lost information when diffraction spots are recorded. The intensities of diffraction spots can be measured providing information on the amplitude of diffracted x-rays but the phases are lost. In order to produce an electron density map through an inverse Fourier transform, phase information is required. Without it, a vector map of the atoms in a crystal would be produced but the 3D structure would not be solved.

Two methods are generally employed to solve crystal structures. Molecular Replacement (MR) may be used where *a priori* knowledge of a structure sharing

significant homology with the protein of unknown structure is available. The known structure can be positioned in the unit cell of the crystal of unknown structure so as to match the experimentally determined diffraction pattern. Following this, phases can be determined by bootstrapping phase information from the known structure. A severe limitation of this technique is that it can severely bias the structure of the unknown protein toward similarity to the search model.

Where a suitable search model is unavailable, several experimental methods can be used to obtain phase information *de novo*. In general, this is done by producing derivatives of native crystals through the incorporation of heavy atoms by soaking or by incorporation into the nascent protein during expression. The incorporation of these heavy atoms creates differences in the measured intensities of diffracted x-ray radiation which can be used to produce a smaller sub-structure. Structure factors for this sub-structure can be calculated based on the differences in measured intensities between native and derivative crystals, which in turn allow phase information for the sub-structure to be calculated which can be bootstrapped onto the whole structure allowing calculation of an initial electron density map.

Specific phasing methods attempted in the present thesis are discussed in greater detail in Chapter 5.

## 2.8. Yeast 2-hybrid procedures

### 2.8.1. Strains and media

All yeast work in the present study was carried out with *Saccharomyces cerevisiae* strain Mav203. Mav203 contains deletions in its endogenous GAL4 and GAL80 transcription factors, rendering it suitable for use in yeast 2-hybrid systems which utilise GAL-4 based reporter assays. In addition to this, Mav203 contains three Gal-4 inducible reporter genes expressing *HIS3*, *LacZ* and *URA3*. *LacZ* ecodes the enzyme β-galactosidase, which cleaves lactose. *HIS3* and *URA3* encode enzymes

involved in the biosynthesis of the amino acids histidine and uracil. Further to this, Mav203 is auxotrophic for leucine and tryptophan enabling selection of cells successfully transformed with two plasmids.

YPAD agar [YPAD broth = 1% (w/v) yeast extract, 2% (w/v) bactopeptone, 2% (w/v) glucose, 0.003% (w/v) adenine hemisulphate, pH 6.0; supplemented with 2% (w/v) microbiology grade agar] was used to culture untransformed Mav203. Mav203 was maintained by streaking <1 month old colonies onto fresh YPAD plates, incubation at 28˚C for ~4 days before storage at 4˚C. If colonies were left for longer than 1 month at 4˚C, a fresh culture was prepared from glycerol stocks stored at -80˚C.

The GAL-4 based assay employed in the present study was based on the ProQuest system (Invitrogen) which makes use of the pDEST32 bait vector and pDEST22 prey vector. These vectors carry *LEU2* and *TRP1* selectable markers, respectively, enabling selection of yeast co-transformed with both bait and prey constructs on plates deficient in leucine and tryptophan.

Three control strains were used and these were obtained from Dr Miles Armstrong (James Hutton Institute). These all expressed Krev1 fused to the GAL4-DNA binding domain and either wild type RalGDS (which strongly interacts with Krev1), $RalGDS^{I77T}$ (which weakly interacts with Krev1) or $RalGDS^{L65P}$ (which does not interact with Krev1) fused to the GAL4-Actication domain (Herrmann et al., 1996, Serebriiskii and Kotova, 2004).

For positive selection of Mav203 co-transformed with bait and prey plasmids, synthetic complement, double-dropout (SC-LW) medium [0.67% (w/v) Yeast Nitrogen Base without Amino Acids (Y0626), 0.16% (w/v) Yeast Synthetic Drop-out Medium Supplements (Y2001), 0.0075% histidine, 0.0075% uracil, 2% (w/v) glucose, pH 5.6; supplemented with 2% (w/v) microbiology grade agar] was used. As described previously, Mav203 is auxotrophic for leucine and tryptophan so both bait and prey plasmids are required to complement auxotrophy when yeast is grown on media lacking these amino acids. Similar to this, for reporter assays (section **2.9.2**), triple drop out media, lacking either histidine or uracil in addition to leucine and tryptophan, was used. In order to suppress ‘leaky’ auto-activation of *HIS3*

expression and possible detection of false positives, 10 – 30 mM 3-amino-1,2,4-triazole (3AT) was added to molten agar immediately before plates were poured.

### 2.8.2. Yeast transformation by the lithium acetate/single-stranded carrier DNA/polyethylene glycol method

The following procedure is based on the ProQuest™ Two-Hybrid system (Invitrogen) manual with slight modifications. Competent cells were prepared by inoculating 10mL of YPAD medium with a single, well-isolated Mav203 colony no more than 7 days old using a sterile, plastic inoculating loop (VWR). A separate, un-inoculated, 40mL aliquot of YPAD was set up. Both were sealed with parafilm and incubated ~16 h at 28˚C for shaking at 220rpm. After 16 h, the un-inoculated aliquot of YPAD was checked for infection and the inoculated aliquot was diluted in fresh YPAD media 1:10 before an $OD_{600}$ reading was taken. This was used to set up two, 20mL cultures of this starter culture with an $OD_{600}$ of 0.4. These two cultures were then grown for 3-5 h at 28˚C with shaking at 220rpm. Yeast was centrifuged at 1,000 x *g* for 5 min at room temperature in a Legend RT (Sorvall) benchtop centrifuge. Supernatant was carefully removed and cell pellets were washed in one original volume of MQ $H_2O$ by re-suspending with gentle agitation. Washed cells were pelleted once more by centrifugation as described previously.

Whilst yeast cultures were growing, 200ng relevant bait and prey constructs were dispensed into sterile, polypropylene, 0.2mL 8-Tube PCR strips (Thermo Scientific). In addition, an aliquot of Salmon Sperm DNA, previously sheared, to act as a carrier for plasmid DNA, was boiled at 95˚C for 10 min and rapidly cooled on ice. 10µg per reaction was dispensed into tubes containing bait and prey plasmid DNA.

Washed cell pellets were re-suspended in 250µL Yeast Resuspension Buffer [100 mM lithium acetate and 5 mM Tris-HCl, 0.5 mM EDTA, pH 7.5] by gentle agitation. Re-suspended yeast was then left to incubate at RT for 5-10 min. 10µL of yeast suspension per reaction was added to bait/prey plasmids and carrier DNA using a multi-channel pipette and mixed by stirring with tips. 70µL of Yeast Transformation Buffer [40% PEG 3350, 100 mM lithium acetate and 10 mM Tris-HCl, 1 mM

EDTA, pH 7.5] was dispensed into each tube with a multi-channel pipette and gently mixed with tips. Reaction mixtures were incubated at 28˚C for 30 min. Following this, 8.8µL sterile DMSO was added to each reaction and mixed by stirring with tips. Reactions were heat shocked at 42˚C for 7 min in a thermocycler. Cells were pelleted by centrifugation for 1 minute at 1,000 x *g* in a Legend RT (Sorvall) benchtop centrifuge. Supernatant was removed and each pellet was re-suspended in 100µL MQ $H_2O$ and spread on dry, sterile, double-dropout synthetic medium (SC-LW) agar and incubated for 3-4 days at 28˚C. All transformation procedures were carried out in a laminar flow hood sterilised by an internal UV light for ~20 min and washing with 70% Ethanol.

### 2.8.3. Yeast 2-hybrid reporter assays

Following the development of multiple, well separated colonies indicating successful co-transformation of Mav203 yeast, individual colonies were selected and replica plated, in triplicate, on reporter assay plates. Pipette tips were used to pick equal amounts of yeast from three individual colonies per transformation reaction and cells were re-suspended in 100µL MQ $H_2O$. 2µL per transformation was spotted onto 14cm diameter reporter assay plates (YPAD with a nylon membrane overlaid, SC-LWH + 10-30mM 3 AT or SC-LTU) using a multichannel repeater pipette. In addition, cells were spotted onto 14cm diameter SC-LT agar for maintenance and subsequent expression testing. Reporter assay plates were incubated for 24-48 h at 28˚C in a static incubator and growth of yeast colonies was scored.

Colonies grown on nylon membranes overlaid on YPAD agar were used to assess activation of the *LacZ* reporter gene. Membranes were removed using sterile forceps and immersed in liquid $N_2$ for 30 seconds in order to lyse cells. Two pieces of filter paper slightly larger than the area of the nylon membrane were cut and pre-soaked in an X-Gal solution prepared by dissolving 20 mg of X-gal in 100 µL of N,N-dimethyl formamide (DMF), adding 60 µL of β-mercaptoethanol and 10 mL of Z-buffer [60 mM disodium hydrogen phosphate (Na2HPO4), 40 mM sodium dihydrogen phosphate ($NaH_2PO_4$), 10 mM potassium chloride, 1 mM magnesium sulphate, pH 7.0]. The nylon membrane was sandwiched between pre-equilibrated

blotting paper and incubated for 24 h at 37˚C on a slant to avoid accumulation of excess X-Gal solution, resulting in smearing of blue colouration. Following incubation, membranes were assessed for development of blue colouration indicating a positive interaction.

### 2.8.4. Yeast protein extraction for western blotting and expression testing

Strains of transformed yeast to be assayed for protein expression were used to inoculate 5mL aliquots of SC-LW medium (section 2.9.1, without agar) whilst an untransformed colony of MaV203 was used to inoculate a 10mL aliquot of YPAD medium. Media was vortexed briefly to ensure dispersal of yeast cells and then incubated for ~16 h at 28˚C with shaking at 220rpm. Cultures were vortexed again to disperse yeast cells and subsequently used to inoculate 50mL aliquots of YPAD medium, which were incubated at 28˚C with shaking at 220rpm until $OD_{600}$ reached 0.4-0.6. $OD_{600}$ values were recorded and multiplied by overall culture volume to determine $OD_{600}$ units per culture. Cultures were poured into pre-chilled 250mL centrifuge bottles (Nalgene) and centrifuged at 1,000 x *g* for 5 min in a prechilled Fibrelite® F14S-6x250y rotor (Thermo Scientific) and an RC6+ free-standing centrifuge (Thermo Scientific/Sorvall). Supernatant was discarded and pellets were washed in 50mL ice cold MQ $H_2O$ before being centrifuged as before. Supernatant was discarded and pellets were flash frozen in liquid $N_2$ prior to storage at -80˚C.

Cell pellets were thawed and re-suspended in Cracking Buffer [8 M urea, 5% (w/v) SDS, 40 mM Tris-HCl pH 6.8, 0.1 mM EDTA, 0.4 $mg.mL^{-1}$ bromophenol blue] prewarmed to 60 ºC and supplemented with Yeast and Fungus Protease Inhibitor Cocktail (50µL/g of yeast cells), 125.7mM β-mercaptoethanol and 4.4 mM phenylmethanesulfonyl fluoride (PMSF). 100µL supplemented cracking buffer per 7.5 $OD_{600}$ units of culture was added to pellets along with 80 uL of glass beads (425 – 600 µm; Sigma Cat No. G-8772) per 7.5 $OD_{600}$ units of cells in 2mL screw cap centrifuge tubes. Cells were incubated for 10 min at 70˚C with shaking at 800rpm before being vortex mixed for 1 minute. Samples were spun down at 14,000 x *g* for 5 min at 4˚C and supernatants transferred to fresh, 1.5mL Eppendorf tubes. Pellets

containing cellular debris and glass beads were incubated for a further 5 min at 99˚C before vortexing and centrifugation as before. The supernatant from this stage was added to the supernatant from the previous stage and stored at -80˚C prior to SDS-PAGE analysis as described in section 2.6.1.

### 2.8.5. Probing western blots with α-GAL4$_{DBD}$ and α-GAL4$_{AD}$ antibodies to detect protein expression

Yeast protein extracts prepared as described in section 2.9.4 were loaded onto 17% SDS-PAGE gels and separated as described in section 2.6.1. Proteins were transferred to nitrocellulose membranes as described in section 2.6.7. Ponceau S staining was used as a loading and transfer control prior to detection using the α-GAL4$_{DBD}$ and α-GAL4$_{AD}$ HRP-conjugated antibodies and ECL reagents as described in section 2.6.7.1.

## 2.9. *In planta* procedures

### 2.9.1. Growth conditions

*N. benthamiana* and *S. lycopersicum* cv. Moneymaker plants were grown and maintained by the John Innes Centre horticultural services staff in controlled environment rooms or glasshouses at 22 ºC with 55% humidity and a 16h/8h light/dark photoperiod.

### 2.9.2. Stable transformation of *S. lycopersicum* cv. Moneymaker

(Note – Tomato plants were transformed by Mr. Matthew Smoker and Ms. Jodie Pike (TSL Tissue Culture & Transformation team)).

Tomato seeds were sterilised with ethanol and $H_2O_2$, plated on germination media and vernalised at 4˚C for 2 weeks in a cold room with all light blocked out. Seeds were removed from the cold room and allowed to germinate with a 16h photoperiod and selected for transformation 7-10 days post germination.

2mL of a fine tobacco cell suspension was spread on MS medium plates [1x MS salts, 1% glucose and 0.6% agar]. These plates were stored overnight under low light. On the same day, *A. tumefaciens* strain AGL1 cultures containing binary plasmids were set up in L medium with appropriate antibiotic supplementation and incubated at 28˚C for ~16hrs.

A Whatman No. 1 filter paper was placed on top of the MS medium plates described previously. Explants were prepared from seedlings by removing cotyledons and making transverse cuts across them to generate two explants of ~0.5cm long each (the tip was discarded). Explants were stored in sterile $H_2O$ whilst preparation of further explants proceeded. *A. tumefaciens* cultures were centrifuged at 4,000 x g, washed in 1x volume of sterile $H_2O$, spun down as before and resuspended in MS medium + 3% sucrose to an $OD_{600}$ reading of 0.4-0.5. Cut explants were immersed in the *A. tumefaciens* suspension, dried on sterile filter paper and placed abaxial face up on the plates described in the previous paragraph. Plates were incubated under low light for 48hrs.

Explants co-incubated with *A. tumefaciens* were transferred to regeneration plates [1x MS salts, 0.01% myo-inositol, 1x Nitsch's vitamins, 2% sucrose, 0.4% AgarGel, 0.002% zeatin riboside, 0.032% timentin and 0.01% kanamycin, pH 6.0] abaxial face down. Plates were incubated under low light and transferred to fresh regeneration medium every 2-3 weeks. Regenerating becoming too large for petri dishes were transferred to large glass jars containing regeneration medium. Following 6 weeks in glass jars, regenerating material was transferred to glass jars containing rooting medium [0.5x MS salts, 0.5% sucrose, 0.225% Gelright, 0.032% timentin and 0.005% kanamycin].

Plants producing roots in rooting medium were transferred to autoclaved, water saturated jiffy plugs (plugs of compost contained within a permeable membrane) and grown in plastic cases to ensure high humidity. Once roots were visibly protruding from jiffy plugs, they were transferred to the glasshouse and potted out. Potted out plants were kept under plastic incubator covers for 4-5 days so as to reduce humidity gradually to normal glasshouse levels.

### 2.9.3. *In planta* gene expression using transient infiltration of *N. benthamiana* and *S. lycopersicum* with *A. tumefaciens*.

Transformation was carried out as described in (Kapila et al., 1997, Van der Hoorn et al., 2000). Strains of *A. tumefaciens* harbouring binary plasmids were grown in LB or L media supplemented with antibiotics appropriate for the strain and plasmid being used (see Table 4 for bacterial strains and Appendix Table C for plasmids). Cultures were typically grown for 16-36 h with shaking at 200rpm at a temperature of 28˚C. Cultures were centrifuged at 1,500 x *g* for 10 min at room temperature, supernatant was discarded and pellets were washed by resuspension in MQ $H_2O$ before being centrifuged as described once more. Washed pellets were resuspended in 4-5mL of MMA buffer [10 mM magnesium chloride, 10 mM MES pH 5.6, 150 – 200 μM acetosyringone] in the case of GV3101 destined for *N. benthamiana* or, for AGL1 destined for *S. lycopersicum*, the same buffer lacking MES. $OD_{600}$ values of the cultures were recorded and cultures were diluted to an appropriate $OD_{600}$ value for a given infiltration experiment (typically 0.1-0.8). Suspensions were incubated for 1-2h at room temperature prior to infiltration into plant leaves. For *N. benthamiana* this involved using a blunt syringe to force cell suspensions into the intracellular spaces of the leaves *via* the abaxial face of the leaf. For *S. lycopersicum*, mechanical wounding of the abaxial face of the leaf with a hypodermic needle was required to aid infiltration as tomato leaves are particularly recalcitrant to this procedure.

Most binary vectors used in the present study featured a constitutive CaMV 35S promoter. However, where β-estradiol inducible promoters were used, gene expression was induced by 10 μM β-estradiol sprayed onto both the axial and abaxial

sides of the leaf from a plastic, atomiser spray bottle. First application was performed 48 h post infiltration to allow recovery. Subsequent inductions were administered following this as required, typically every 48 h.

### 2.9.4. Extraction of protein from *N. benthamiana* leaf tissue

*N. benthamiana* leaves were harvested at an appropriate time-point post infiltration/induction and were wrapped in aluminium foil before being frozen in liquid $N_2$ for transport or prior to storage at -80˚C. Leaves had their mid-veins removed prior to freezing. Leaves were then ground to a homogenous powder using a pestle and mortar pre-chilled in liquid $N_2$ or with plastic micropestles in Eppendorf tubes also pre-chilled with liquid $N_2$. Tubes were secured in a rack pre-chilled with liquid $N_2$. Each sample had a volume of GTEN buffer [10% (v/v) glycerol, 150 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM sodium chloride] supplemented with 10 mM DTT, 2% (w/v) polyvinyl polypyrrolidone (PVPP), 1% (v/v) Protease Inhibitor Cocktail (P 9599, Sigma), 0.1% (v/v) Tween®-20 equal to 2 times the sample weight in grams (2mL:1g) added to it prior to 30 seconds of vortex mixing, incubation of ice for 10 min, a further 30 seconds of vortex mixing and centrifugation at 11,500 x g for 10 min at 4 ºC. The resulting supernatant was transferred to a pre-chilled 15mL centrifuge tube and spun as described previously. The resultant supernatant was dispensed into pre-chilled 1.5mL Eppendorf tubes and used in co-immunoprecipitation experiments, prepared for SDS-PAGE analysis as described in section **2.6.1** or flash frozen in liquid $N_2$ prior to storage at -80˚C.

### 2.9.5. Extraction of gDNA from *S. lycopersicum* leaf tissue

Tomato gDNA was extracted from leaf tissue using the method described in (Murray and Thompson, 1980). The method uses Cetyltrimethyleneammonium bromide (CTAB), a positively charged surfactant which improves purity of isolated DNA through more efficiently denaturing glycoproteins and polysaccharides often co-precipitated with DNA in other extraction methods. Tomato leaf tissue was harvested at an appropriate time point post infiltration and flash frozen in liquid $N_2$. Leaf tissue was stored at -80˚C. Approximately ¼ of a 6 week old *S. lycopersicum*

leaf was used for DNA extraction. Leaf tissue was ground in an Eppendorf tube with a plastic micropestle in 200µL of CTAB buffer [100mM Tris-HCl pH 7.5, 20mM EDTA, 1.4M NaCl, 2% (w/v) CTAB, 1% (w/v) polyvinylpolypyrrolidone 40,000] on ice until leaf tissue was well homogenised. Leaf tissue was incubated in CTAB buffer at 65˚C for 10 min. 200µL of chloroform and isoamyl alcohol (in a 24:1 ratio) were added to each tube and vortex mixed for 15 seconds. Tubes were spun at 4˚C for 3 min at 13,000rpm. The upper phase of the resultant supernatant was carefully collected and transferred to a fresh 1.5mL Eppendorf tube. 200µL of isopropanol was added to each sample and incubated at room temperature for 5 min. Samples were centrifuged at 4˚C for 15 min at 13,000 rpm and supernatant was discarded. DNA pellet was washed in ice cold 70% ethanol and centrifuged at 4˚C for 3 min at 13,000 rpm, supernatant was discarded. DNA pellets were re-suspended in 20µL TE [10mM Tris-HCl pH 7.5, 1mM EDTA, pH 8.0]. Extracted gDNA was run on agarose gels as described in section 2.5.6 to confirm extraction and integrity before storage at -20˚C.

### 2.9.6. Immunoprecipitation from plant extracts

Total soluble extract from plant leaf tissue prepared as described in section 2.10.4 was used in immuno-precipitation experiments to enrich for a protein of interest in mass spectrometry experiments or in expression testing.

Immunoprecipitations using the HA epitope tag were performed using magnetic anti-HA affinity matrix (Roche). 75µL of well homogenised matrix was washed three times in 5x volume of ice cold IP buffer [GTEN buffer + 0.1% (v/v) Tween®-20]. Affinity matrix was then re-suspended to its original volume in IP buffer and mixed with 250µL total extract from leaf tissue in a 1.5mL Eppendorf tube and mixed on a rotary mixer for 30 min at 4˚C. Affinity matrix bound to HA-tagged

protein was separated on a magnetic Eppendorf tube rack (Roche) and supernatant was discarded. Matrix was washed in 1mL IP buffer and re-suspended by inversion. This was repeated a further 4 times as described. After the final wash step and removal of supernatant, affinity matrix was resuspended in 30µL IP buffer and 10µL of SDS-PAGE loading buffer (stock concentration 4x working concentration) and boiled at 95˚C for 10-15 min to elute bound protein from affinity matrix. Affinity matrix was separated from protein in a magnetic rack as described and supernatant containing protein of interest was collected and loaded onto SDS-PAGE gels as described in section 2.6.1.

### 2.9.7. Infection of leaves with *P. infestans*

Whole leaves were harvested from plants and laid in plastic trays on top of damp paper towelling abaxial face up. *P. infestans* strain 88069 was grown as described in (Kamoun et al., 1998) on Rye Sucrose Agar (RSA) plates and incubation at 18˚C for 11-14 days. 5mL of sterile, pre-chilled (4˚C) $dH_2O$ was dispensed onto these plates which were incubated at 4˚C for 3 h. Water containing motile zoospores was then collected and analysed under a light microscope to assess number of zoospores/mL. Suspensions were diluted to 100 zoospores/mL and 10µL droplets of zoospore suspensions were spotted onto abaxial faces of leaves. Lids were placed on top of trays containing leaves creating an environment of 100% relative humidity and infections on *N. benthamiana* leaves were incubated at room temperature on the bench with daily monitoring of pathogen growth. *S. lycopersicum* infection assays were conducted as above except for incubation, which was at 18˚C in a constant temperature cabinet with controlled light/dark cycles.

White light images were recorded with Nikon COOLPIX L24 digital camera. Lesion area was determined using GIMP (v2.8) software using the free select tool in pixels, which were converted to $mm^2$.

### 2.9.8. Cell death suppression assays

*A. tumefaciens* strains harbouring relevant constructs were grown and prepared as described in section 2.9.3. For initial mutant screening described in Chapter 3, MAPKKK$_\varepsilon$ mutant candidates were grown in sealed 96 deep well blocks under the same conditions as larger cultures. Infiltration proceeded as described in section 2.9.3 although rather than whole leaves, ~25mm diameter spots were infiltrated with six spots per leaf comprising a positive control, a negative control and 4 candidates. Effectors were diluted to a final $OD_{600}$ of 0.6 and kinases were diluted to a final $OD_{600}$ of 0.4, the Tomato Bushy Stunt Virus RNA silencing suppressor protein p19 (Voinnet et al., 2003)[12] was included at an $OD_{600}$ of 0.1.

Expression of kinases under the control of β-estradiol inducible promoters was induced as described in section 2.9.3. Cell death was monitored daily up to 8 days post induction and images were recorded with a Nikon COOLPIX L24 digital camera or by Mr. Andrew Davies (JIC Photography).

### 2.9.9. Ion leakage assays

Ion leakage assays were, broadly speaking, carried out as described in (Melech-Bonfil and Sessa, 2010a). For each combination of constructs, three leaf discs of 8mm diameter were harvested and floated in 15mL of MQ water in a 50mL centrifuge tube (Corning). In addition to this, samples were also taken from untransformed tissue to control for ion leakage due to mechanical wounding associated with harvest. Leaf disks were incubated at room temperature for >2h with gentle shaking. A B-173 Compact Twin Conductivity Meter (HORIBA) was used to record conductivity in micro-Siemens per cm ($\mu S.cm^{-1}$) which was taken to be a proxy for ion leakage from leaf tissue into the medium and, so, a proxy for cell death.

[12] Whilst the paper cited here has been retracted following allegations of misconduct in its production, the efficacy of the p19 RNA silencing suppressor protein in suppressing endogenous RNA silencing pathways to boost expression in transient infiltration assays has been observed repeatedly since its publication and is in no doubt.

# 3

# Engineering a Solanaceous MAPKKK for insensitivity to PexRD2

# Chapter 3 Engineering a Solanaceous MAPKKK for insensitivity to PexRD2.

## 3.1. Introduction.

### 3.1.1. Crop pathogen management strategies

Manipulation of crop plants in order to avoid pre-harvest yield loss to pests and pathogens has been a long standing concern in crop breeding. Whilst the allele specific, gene for gene basis for immunity was not understood until Howard Flor's series of studies through the 1940s and 50s on the Flax rust pathogen (*Melamspora lini*) and Flax (*Linum usitatissimum*) (Flor, 1942, Flor, 1947, Flor, 1955), with the first NLR-encoding Resistance genes not being cloned and characterised until many years later (Staskawicz et al., 1995), humans have probably been selecting crops for disease resistance since the dawn of agriculture. More recent efforts made to enhance resistance of crops to their pathogens utilising fine knowledge of effector-NLR interactions have been effective, as exemplified by the introgression of *Solanum venturii Rpi-Vnt1.1* into potato and tomato crops, resulting in complete, broad spectrum resistance to *P. infestans* (Foster et al., 2009). This approach shows great promise, as NLR based resistance is often highly effective. However, oomycetes – in particular the *Phytophthora* genus - are well poised to adapt to changes in the NLR complement of their hosts. Their large, effector rich genomes (Haas et al., 2009), render them well adapted to evade host recognition with redundant effectors being lost and gained at high frequencies. Exposure to newly deployed NLRs over a limited number of growing seasons has frequently led to the emergence of strains which are virulent on these cultivars (Malcolmson, 1969).

One strategy which has, thus far, been somewhat ignored, is the manipulation of endogenous host targets, susceptibility factors or NLR proteins and their recognition

mechanisms. Most efforts on this have been focused on expanding the recognition specificity of NLRs and their guardees. One elegant study employed the *Arabidopsis thaliana* NLR RPS5 to expand recognition to other *Pseudomonas syringae* effectors (Kim et al., 2016). Normally, RPS5 is activated upon AvrPhB mediated cleavage of PBS1 as opposed to direct recognition (Ade et al., 2007). AvrPhB functions as a protease *in planta* and its cleavage site is well defined (DeYoung et al., 2012). In a similar mechanism from the same pathosystem, the effector protein AvrPt2 functions as a protease which cleaves RIN4, activating the NLR, RPS2 (Axtell and Staskawicz, 2003, Mackey et al., 2003). Again, the AvrPt2 cleavage site of RIN4 is well defined and was swapped with the AvrPhB cleavage site in PBS1, resulting in *A. thaliana* lines capable of recognising AvrPt2 and eliciting an immune response in the absence of both RIN4 and RPS2. The authors also found that replacement with protease sites cleaved by viral proteases could condition immune response to viral proteases and significantly limit the virulence 'live' virus *in planta*.

Another approach has been direct manipulation of NLRs. The potato NLR R3a recognises the *P. infestans* RXLR effector Avr3a. Two alleles of Avr3a are prevalent in *P. infestans* populations; $Avr3a^{EM}$ and $Avr3a^{KI}$. The KI allele is recognised by R3a, leading to activation, whilst the EM allele is only recognised weakly, resulting in a weak cell death response only observable as autofluoresence and successful host colonisation (Armstrong et al., 2005, Bos et al., 2009, Bos et al., 2010). (Segretin et al., 2014) used a random mutagenesis approach to expand the recognition specificity of R3a to elicit a strong immune response to $Avr3a^{EM}$. Further, the same mutations were transferred to the tomato NLR I2 which mediates resistance to *Fusarium oxysporum* f. sp. *Lycopersici* (I2 is a homologue of R3a which responds weakly to Avr3a in transient assays in *N. benthamiana* leaves). One of the mutations in I2 enhanced response to Avr3a and also expanded the response of I2 to multiple *F. oxysporum* f. sp. *Lycopersici* strains (Giannakopoulou et al., 2015).

To date, no examples of the manipulation of *bona fide* virulence targets of plant pathogen effector proteins exist in the literature. With high quality reference genomes available for many economically important crop pathogens and pests, identification and characterisation of putative effector targets has increased. Furthermore, with clues from recently gained understanding that many plant NLRs

contain ‘integrated’ domains which likely derive from virulence targets of effector proteins (Sarris et al., 2016, Kroj et al., 2016) this trend of increasingly rapid effector identification is likely to continue through deployment of integrated domains as ‘baits’ with which to identify putative effectors. A sound, basic understanding of an effector-target interaction, its outcome for virulence and which processes are being perturbed should enable the design of assays and mutagenesis strategies – either targeted or random – with which to capture variants of the target which are refractory or insensitive to the effector’s virulence function. As the pathogen will no longer be able to perturb the function of these effector-insensitive targets, this presents an alternative strategy to engineer disease resistance. When combined – or ‘stacked’ – with NLR-mediated resistance strategies this may provide more robust, durable resistance.

### 3.1.2. The *P. infestans* RXLR effector PexRD2 interacts with a host MAPKKK to suppress cell death signalling

As discussed in section 1.3, *P. infestans* secretes effector proteins into the cytoplasm of its host to perturb normal cellular function, and facilitate colonisation. One such effector is PexRD2. PexRD2 has a typical, RXLR protein domain organisation with an N terminal signal peptide, central RXLR-dEER motif and C terminal effector domain. The effector domain confers the effector’s virulence function *in planta*. Structural studies of PexRD2, revealed the effector domain forms a homo-dimer (and oligomerisation is also observed *in planta*) with each monomer adopting a highly conserved α-helical fold, termed the ‘WY fold’. This fold is found in many sequence-divergent RXLR effectors and may underpin the rapid diversification of these proteins by providing a stable ‘scaffold’ around which new functions and/or mutations conditioning evasion of recognition can evolve (Boutemy et al., 2011).

The PexRD2 effector domain was shown to be necessary and sufficient for interaction with a *Solanum tuberosum* Mitogen Activated Protein Kinase: MAPKKK$_{\varepsilon}$ (King et al., 2014). The *S. lycopersicum* homologue of MAPKKK$_{\varepsilon}$

confers enhanced immunity to two bacterial pathogens (*P. syringae* and *Xanthomonas campestri*) by positively regulating immunity-related cell death signalling (Melech-Bonfil and Sessa, 2010b). Heterologous expression of MAPKKKε in the model solanaceous crop *N. benthamiana* leads to macroscopic cell death, which is hypothesised to be a defence strategy against biotrophic or hemibiotrophic pathogens such as *P. infestans*. Indeed, virus induced gene silencing (VIGS) of *Nb*MAPKKKε and subsequent challenge with *P. infestans* results in enhanced pathogen growth (King et al., 2014).

When MAPKKKε and PexRD2 are co-expressed in *N. benthamiana* leaves, the cell death initiated by MAPKKKε is suppressed. This suggests a function for PexRD2, that it is secreted into the host cell to suppress MAPKKKε-mediated cell death signalling, maintaining living tissue and promoting pathogen colonisation. A mutation in the hydrophobic interface of the PexRD2 homodimer, Leucine 109 to Aspartate (L109D), prevents effector-homodimerisation and abolishes suppression of MAPKKKε mediated cell death (Boutemy et al., 2011, King et al., 2014). This mutant provides a useful tool for assaying PexRD2 activity.

Here, MAPKKKε was subjected to random mutagenesis PCR and screened in *N. benthamiana* for insensitivity to PexRD2's cell death suppression activity. The generation of such a variant would be attractive to (a) enable further study of how PexRD2 suppresses MAPKKKε-mediated cell death (and its contribution to resistance to *P. infestans*) and (b) an effector insensitive kinase may have utility in generating resistance to *P. infestans* in crops. Effector insensitive variants were identified and re-tested. Their ability to interact with PexRD2 was investigated. Additionally, CRISPR/Cas9 constructs designed to knock out endogenous MAPKKKε were generated and used to transform tomato plants alongside a re-coded complementation cassette comprised of one of the MAPKKKε mutants developed here simultaneously. Plants expressing this construct were assayed for resistance to *P. infestans*.

### 3.1.3. Introduction to CRISPR/Cas9 genome editing.

Recent advances in the field of targeted genome editing have revolutionised the fields of molecular biology and biotechnology. Researchers now have unprecedented ability to introduce deletions, insertions and point mutations at user-specified regions of the genome thanks to three technologies; Zinc Finger Nucleases (ZFNs) (Bibikova et al., 2002, Urnov et al., 2005), Transcription Activator-Like Effector Nucleases (TALENS) (Zhang et al., 2011, Miller et al., 2011) and Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated (CRISPR/Cas) mediated genome editing (Cong et al., 2013, Mali et al., 2013). ZFNs and TALENs were not used in the present study so will not be described here. A brief discussion on these can be found in section 1.5.

The first genetic element to be termed a 'CRISPR' locus was characterised in the archaeobacterium *Haloferax mediterrani* (Mojica et al., 1993). Subsequent research continued to reveal stretches of these 'Clustered Regularly Interspaced Short Palindromic Repeat' loci, interspersed with unique, 33-39bp sequences in other bacterial species (Mojica et al., 1995). Indeed, today it is known that CRISPR loci are widely distributed among bacteria and near-ubiquitous among archaeobacteria whose genomes have been sequenced (Grissa et al., 2007, Chylinski et al., 2014). Understanding of these loci continued to grow leading to the discovery that the unique, interspacing sequences were derived from bacteriophages and the hypothesis that these unique genetic elements may provide the prokaryotes whose genomes they reside in some form of adaptive immunity against bacteria (Mojica et al., 2005, Pourcel et al., 2005). This hypothesis proved to be correct when it was demonstrated that *Streptococcus thermophillus* integrated phage-derived spacers following phage challenge and that manipulation of these phage-derived spacer loci led to abrogation, or acquisition, of immunity to bacteriophage in a spacer sequence dependent manner (Barrangou et al., 2007).

The two other components required for CRISPR/Cas mediated genome editing were discovered through analysis of the *S. thermophillus* genome (Bolotin et al., 2004). A number of so-called 'CRISPR associated' (Cas) genes had previously been identified but a 2005 study revealed a family of Cas genes (then designated Cas5,

now known as Cas9) containing a HNH motif known to be present in several site-specific endonucleases (Bolotin et al., 2005). The aforementioned study also became the first to note the existence of the proto-spacer adjacent motif (PAM) which was found to be critical for target recognition.

It had been predicted that RNA would be the target of crRNAs. However, an elegant experiment in which a self-splicing intron was inserted into a *Staphylococcus epidermidis* nickase gene critical for conjugation and plasmid transformation showed that DNA was the target. Under normal circumstances, a CRISPR spacer with complementarity to the nickase gene interferes with these processes. However, insertion of the self-splicing intron abolished this interference by rendering crRNAs incapable of target sequence recognition indicating that CRISPR acts on DNA as opposed to mRNA (Marraffini and Sontheimer, 2008).

The precision with which the Cas9 endonuclease cleaves its target DNA was confirmed when it was found that double stranded breaks occurred 3bp upstream of the PAM identified in (Bolotin et al., 2005), demonstrating that the endonuclease cuts within the region complementary to the spacer (Garneau et al., 2010). It was also shown that a characteristic of Cas9 – as well as other type II CRISPR/Cas systems – was the lack of requirement for other cleavage complex components. Cas9 plus guide RNA is sufficient for targeted DNA cleavage.

The final piece of basic mechanistic understanding of type II CRISPR/Cas systems to be found was trans-activating CRISPR RNA (tracrRNA). tracrRNAs are RNA molecules complementary to CRISPR repeat sequences which, upon transcription, base pair with their complementary sequence forming a dsRNA which is cleaved by RNAse III in order to form a mature crRNA molecule (Deltcheva et al., 2011).

In order to demonstrate CRISPR/Cas9's potential utility as a general tool for genome editing across a range of organisms, it was necessary to demonstrate its efficacy in organisms other than the organism in which a given Cas9 protein originates. This was achieved when a CRISPR/Cas locus from *S. thermophillus* was expressed in *E. coli* and conditioned resistance to plasmid transformation and bacteriophage infection, again in a sequence specific manner (Sapranauskas et al.,

2011). Mutagenesis experiments in the same study also demonstrated the importance of the aforementioned HNH endonuclease motif, as well as a RuvC motif, in Cas9 for cleavage of target DNA loci. The bipartite specificity of these two motifs in the mode of action of CRISPR/Cas9 complexes was also demonstrated when characterising how CRISPR works at the target sequence level. HNH is responsible for cleavage of the DNA strand complementary to the CRISPR loci whilst the RuvC motif cleaves the non-complementary strand (Gasiunas et al., 2012). Further to this, it was demonstrated that the minimal number of nucleotides required for efficient target recognition is 20nt. Additionally, it was shown that crRNA and tracrRNA could be combined on a single transcript and still perform CRISPR/Cas mediated target cleavage (Jinek et al., 2012).

Evidence of efficacy in eukaryotes was demonstrated when CRISPR/Cas9 constructs were used to cleave targets in *Homo sapiens* and *Mus musculus*. It was also shown that multiple loci can be targeted at once and that Cas9 can drive homology directed repair of double stranded DNA breaks, allowing DNA with flanking regions homologous to the regions either side of the cleavage site to be substituted for native DNA (Cong et al., 2013, Mali et al., 2013).

DNA repair damage mechanisms are constantly at work in cells where UV and radiation damage are ever present threats causing DNA lesions. It is the repair mechanisms which are engaged by damage to dsDNA where both strands of the dsDNA molecule are cut which permits CRISPSR/Cas9 mediated genome editing. Non-homologous end-joining (NHEJ) (Moore and Haber, 1996) relies on microhomology, or no homology at the ends of the strands to be ligated, and allows the introduction of deletions where paired sgRNAs have been used to excise a portion of a target gene or smaller INDELs. Alternatively, homologous recombination (HR) involves the exchange of nucleotides between two identical DNA sequences, allowing whole sections of genes to be replaced with sequences harbouring one, or more, mutations (Sung and Klein, 2006).

## 3.2. Results.

### 3.2.1. Generating a library of MAPKKK$_\varepsilon$ mutants.

In the absence of structural data for MAPKKK$_\varepsilon$ to guide rational design of mutations that may confer insensitivity to PexRD2, a random mutagenesis approach was taken. Using this approach, potentially thousands of putative variants could be screened in a medium-throughput manner using *N. benthamiana* as a model system.

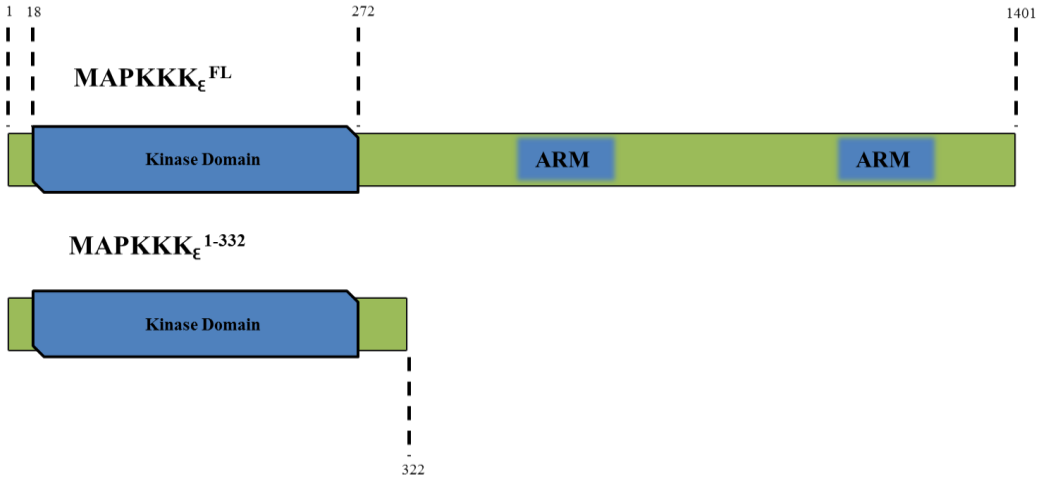



**Figure 3| Schematic representation of MAPKKK$_\varepsilon$ construct used for random mutagenesis.**

Shown above is full length MAPKKK$_\varepsilon$ with C-terminal ARM repeats. Kinase domain is shown as a blue box between residues 18 and 272. Below is the construct used for random mutagenesis spanning residues 1-332.

Mutagenesis was achieved using the Diversify® Random Mutagenesis PCR kit supplied by Clontech. Error prone PCR relies on the fidelity of the DNA polymerase being lowered by the buffer conditions. In this case, raising the concentration of Manganese is sufficient to lower the polymerase's fidelity in order to achieve a desirable rate of mutations/base. A 996bp construct spanning MAPKKK$_\varepsilon$'s N-terminus to its 332$^{nd}$ residue (Figure 3), which is necessary and sufficient for both its

function *in planta* and its interaction with PexRDX2 (King et al., 2014) was subjected to error prone PCR in conditions predicted to yield ~3 mutations per kilobase. This level of fidelity should lead to 2-3 mutations – on average - in the truncated MAPKKK$_{\varepsilon}$ construct. A subset of putative mutants were sequenced in order to confirm the desired rate of mutagenesis. Putative variants were then placed under the control of a β-estradiol inducible promoter and fused to 2 copies of the HA epitope in a bespoke vector derived from pER8 (Zuo et al., 2000).

### 3.2.2. Screening putative MAPKKK$_{\varepsilon}$ variants *in planta*.

*A. tumefaciens* inoculation concentration and β-estradiol concentration were optimised for screening. Final $OD_{600}$ values of *A. tumefaciens* harbouring pERCH:MAPKKK$_{\varepsilon}^{Wt}$ ranging from 0.1-0.6 were tested with inducer concentrations of 5μM, 7.5μM and 10μM. Cell death was scored as >50% of the infiltrated area exhibiting fully confluent cell death. From this analysis it was clear that $OD_{600}$ values of 0.4-0.6 all produced a clear cell death phenotype and that there was little difference between the three inducer concentrations tested (Figure 4). Therefore, a final $OD_{600}$ of 0.4 and a β-estradiol concentration of 10μM were chosen for screening.

The ability of MAPKKK$_{\varepsilon}$ variants to evade cell death suppression by PexRD2 was then assayed by transient infiltration of *N. benthamiana* leaves with *A. tumefaciens* GV3101 harbouring plant expression vectors carrying MAPKKK$_{\varepsilon}$ and PexRD2 at an $OD_{600}$ of 0.4 and 0.6 respectively. All infiltration mixtures included *A. tumefaciens* GV3101 harbouring the tomato bushy stunt virus (TBSV) p19 suppressor of RNA silencing (Voinnet et al., 2003) in order to enhance expression. Cultures harbouring p19 were delivered at an $OD_{600}$ of 0.1. A generic layout for infiltrations in each leaf is depicted in Figure 5a. Initially 288 kinase variants were screened in total. Kinase expression was induced by spraying the abaxial and adaxial faces of leaves with β-estradiol - 24-48 h post infiltration - at a concentration of 10μm. 6-8 days post induction images were taken and cell death symptoms scored. Initial symptom scoring was done coarsely with throughput of greater importance than fine detail. Leaf images were analysed in GIMP 2.0. Area of tissue exhibiting

severe cell death symptoms (in pixels) was divided by total infiltrated area (in pixels) in order to give a percentage of infiltrated area exhibiting severe cell death symptoms. Infiltration sites exhibiting >50% of total infiltrated area exhibiting cell death symptoms were treated as putative positive candidates whilst infiltration sites exhibiting ≤50% of total infiltrated area exhibiting cell death symptoms were discarded. The initial screen of 288 kinase variants yielded 8 positive candidates that evaded cell death suppression according to the previously described criteria (Figure 5b). Positive candidates were sequenced and sequence alignments are shown in Figure 6.

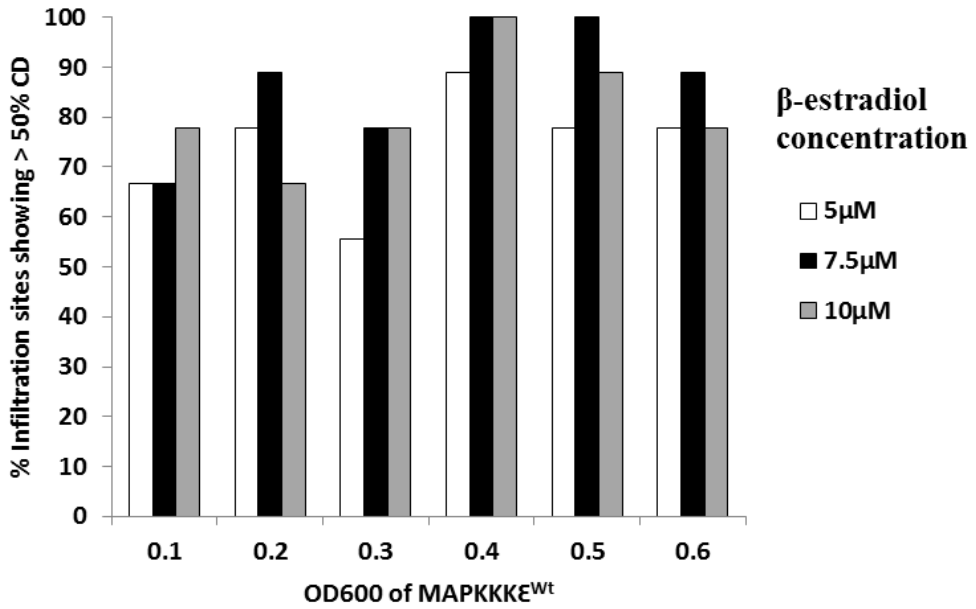
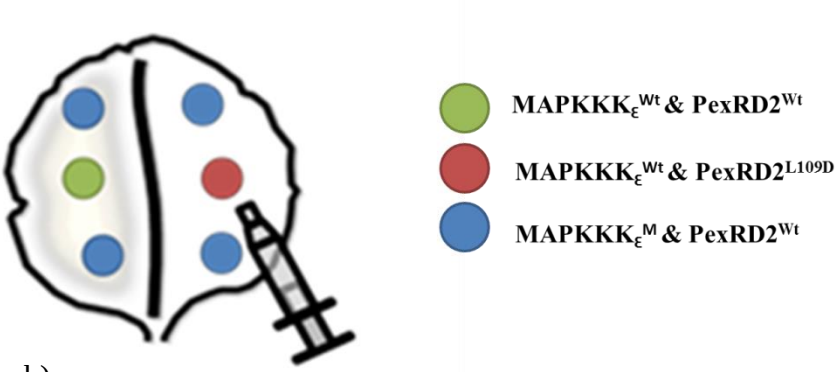



**Figure 4| The effect of *A. tumefaciens* cell density and β-estradiol concentration on pERCH:MAPKKK$_{\varepsilon}$Wt mediated cell death in *N. benthamiana*.**

9 replicates per cell density and inducer concentration value were tested. Plants were infiltrated as described in section 2.9.3, allowed to recover for 24-48 h before β-estradiol was applied. B-estradiol was applied every 48 h until cell death was recorded at 6 days post induction. 'Cell death' was scored as >50% of infiltrated area exhibiting fully confluent cell death and bars represent % of the 9 replicates for each combination exhibiting >50% fully confluent cell death.

a)




b)

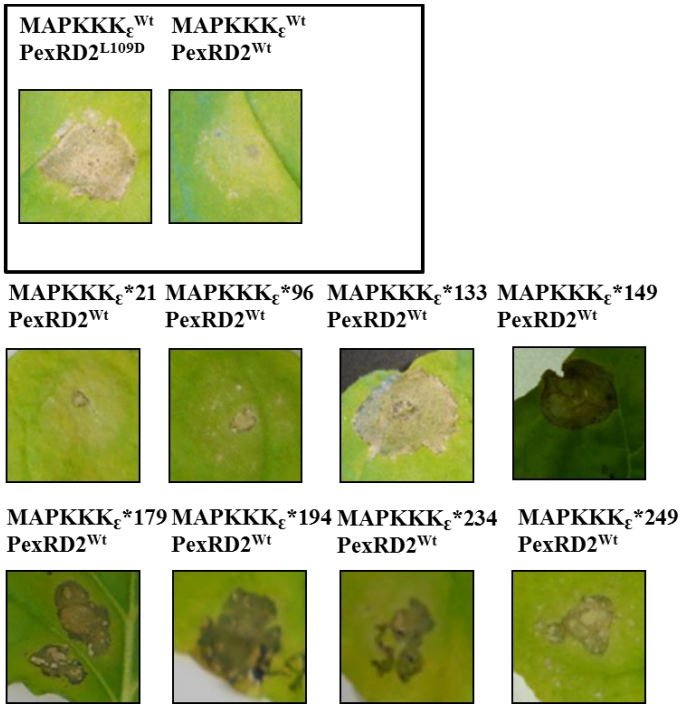



**Figure 5| Images of positive candidates from effector-insensitive MAPKKK$_\varepsilon$ screen**

Positive and negative controls are shown boxed, with non-interacting PexRD2$^{L109D}$ failing to suppress MAPKKK$_\varepsilon$ mediated cell death representing positive control and PexRD2$^{Wt}$ almost completely suppressing MAPKKK$_\varepsilon$ mediated cell death. Positive candidates are shown sequentially below. Numbers correspond to the order in which single *A. tumefaciens* colonies were selected from selective plates.

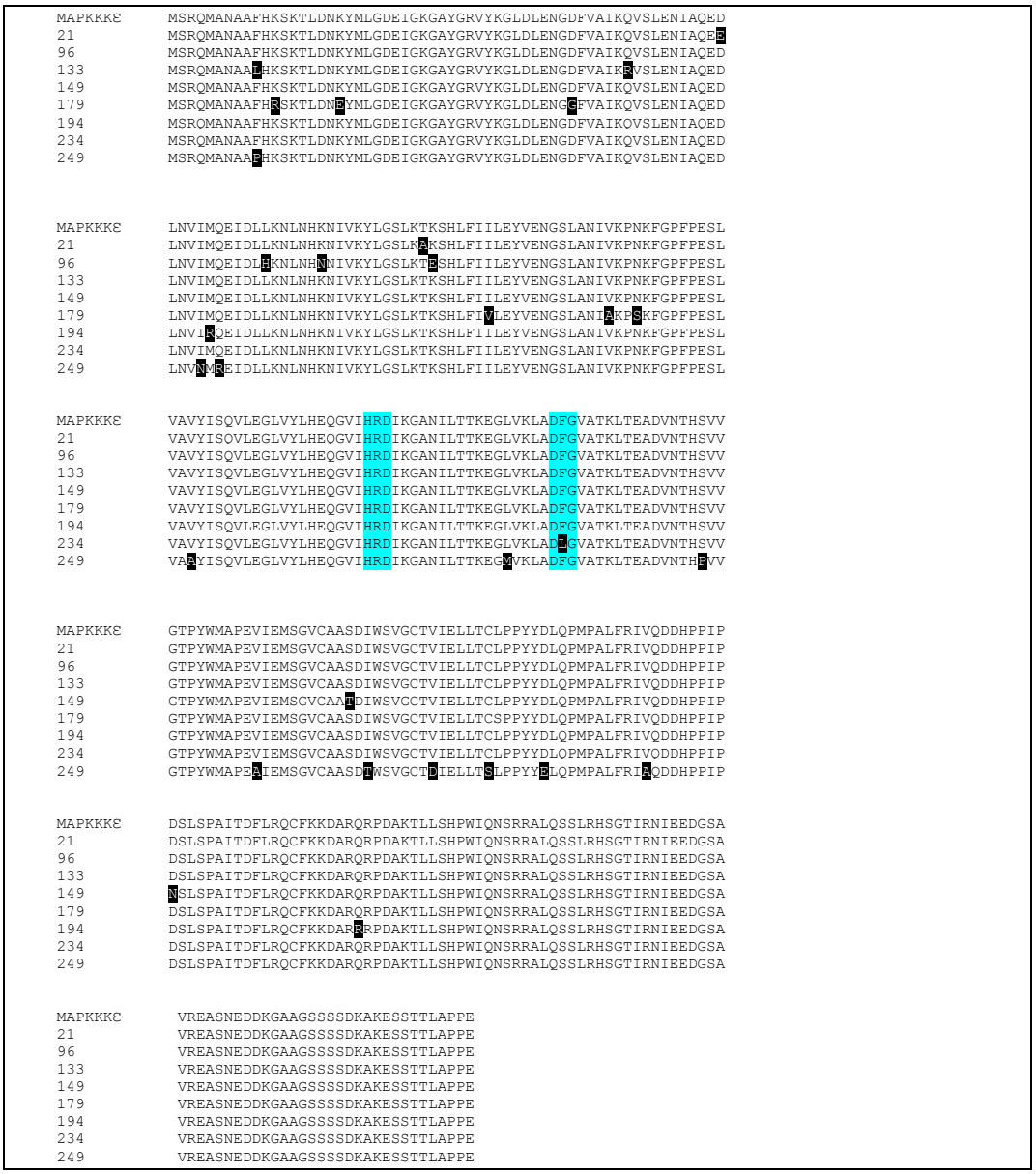



**Figure 6| Amino acid sequences of putative effector insensitive MAPKKK$_\varepsilon$ mutants identified in primary screening.**

Amino acid sequences are displayed aligned with the wild type sequence of MAPKKK$_\varepsilon$ from aa 1 to 332. Amino acid substitutions relative to wild type sequence are highlighted in black. Catalytically active HRD and DFG motifs of the kinase domain are highlighted in light blue.

The distribution of mutations in the sequence was analysed for any potential clustering that could suggest whether any particular region of MAPKKK$_{\varepsilon}$ conditioned insensitivity to PexRD2 (Figure 7). From this analysis, some clustering at the N terminus – residues 1 – 60 - was evident. Only two mutations were evident near the HRD-DFG motif. It is to be expected that in a screen which is intended to explore variants which retain activity, that little to no mutations in or around the active site would be obtained. However, it is known that certain mutations at these residues can be tolerated with minimal alteration to activity (Strong et al., 2011). The Phenylalanine to Leucine mutation evident in candidate 234 in the DFG motif itself has been identified in a kinase variant involved in cancer development in humans – BRAF$^{F595L}$ – and is known not to abolish kinase activity (Kordes et al., 2015). This is presumably due to the role of the F in the DFG motif normally being the occupation of the ATP binding pocket, preventing aberrant kinase activity (Jauch et al., 2005). C terminal to this region, further mutations were evident but were not as highly concentrated as those at the extreme N terminus of the protein.

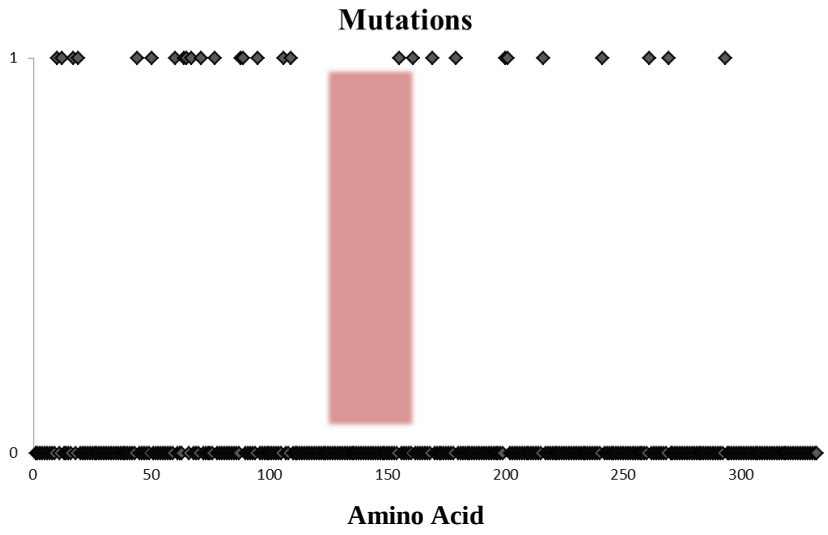



**Figure 7| Amino acid substitutions plotted against amino acid position.**

Black data points at 0 on the *y* axis indicative of no amino acid substitution at a given residue. Blue data points at position 1 on the *y* axis indicative of an amino acid substitution at a given residue. Pink bar represents location of HRD-DFG motif.

In order to be certain of the robustness of positive candidates identified through primary screening, we thoroughly re-tested all positive candidates in a secondary screen (Figure 8).

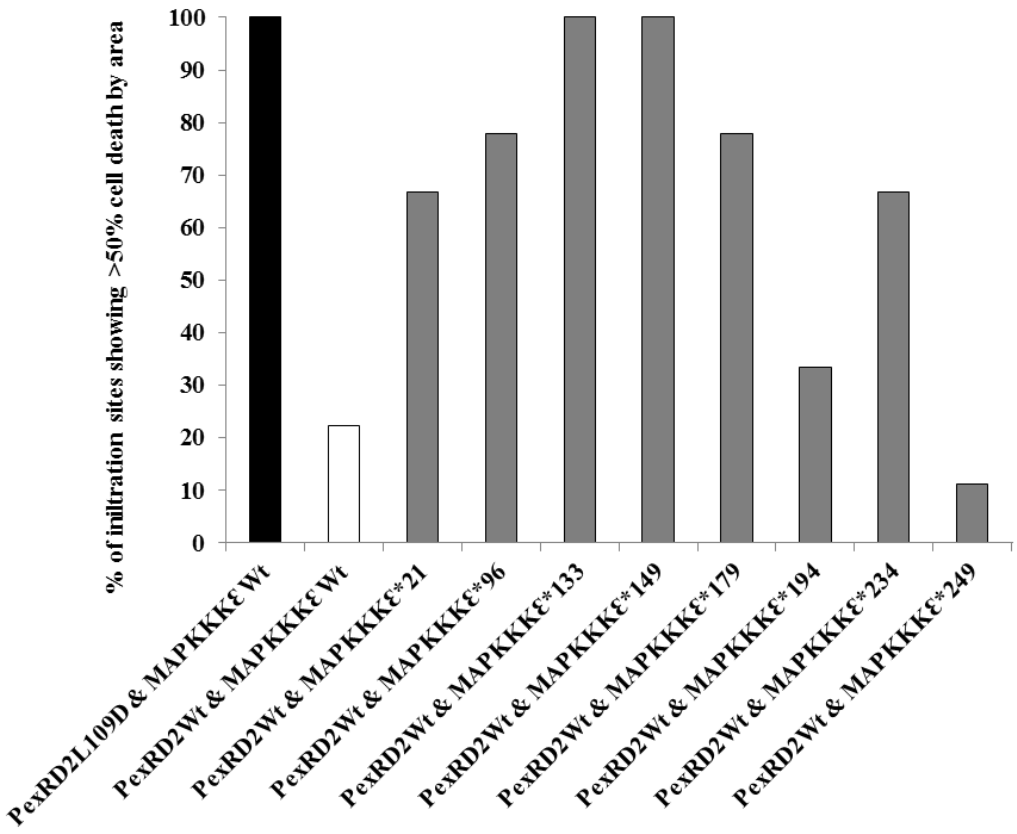



**Figure 8| Initial re-testing of positive candidates from screen reveals that candidates MAPKKK$_{\varepsilon}^{*133}$ and MAPKKK$_{\varepsilon}^{*149}$ reproducibly evade PexRD2 mediated suppression of cell death.**

Bars represent percentage of infiltrated sites exhibiting >50% fully confluent cell death relative to total infiltrated area. Data representative of 9 repeats per variant.

The 8 positive candidates from the initial screen were subjected to more intense scrutiny and repeated on 3 leaves per plant x 3 plants, giving a total of 9 replicates per variant. Infiltration procedure, experimental layout and timescales remained as described previously. From this analysis, candidates 133 and 149 robustly and reproducibly evaded suppression of cell death (Figure 8) to the same level as the positive control of MAPKKK$_{\varepsilon}^{Wt}$ and PexRD2$^{L109D}$. Candidates 21, 96, 179, and 234 also showed a clear difference in their ability to evade PexRD2 mediated cell death

relative to MAPKKK$_{\varepsilon}^{Wt}$. Sequencing of candidates 133 and 149 (Figure 6) revealed 2 amino acid substitutions per variant; Phe10Leu and Gln50Arg in 133 and Ser200Thr and Asp241Asn in 149.

### 3.2.3. De-convoluting the relative contributions of amino acid substitutions to insensitivity to PexRD2 in two positive candidates.

In order to deconvolute the contribution of individual mutations to the evasion of cell death activity, DNA encoding each mutation was synthesised (Genscript USA) and prepared in the pERCH vector described in section 3.2.1.

Infiltrations were performed as described previously. Analysis of images was initially carried out as outlined in 3.2.2 with the 4 candidates; Phe10Leu, Glu50Arg, Ser200Thr, Asp241Asn. Each of the candidates consistently evaded PexRD2 mediated suppression of cell death induction (Figure 9d). The data generated here was re-analysed using a scoring system based on an arbitrary scale of cell death shown in Figure 9c. Symptoms were scored and subjected to one-way ANOVA with Tukey's *post hoc* t-test in order to determine significant differences between means. This analysis revealed a similar picture to the more coarse analysis, with the 4 MAPKKK$_{\varepsilon}$ variants robustly and reproducibly evading PexRD2 mediated suppression of cell death induction, and differing from the negative control Figure 9e. The 4 MAPKKK$_{\varepsilon}$ mutants were tested in two different environments; a controlled environment room (CER) and a temperature controlled glasshouse in order to account for any variation between the two environments even though each environment was, in theory, set to the same day/night cycle, temperature and humidity. Cell death was slightly enhanced in CER assays relative to glasshouse assays in both controls and mutants but the overall trend remained consistent between the two environments. Data presented here is representative of Glasshouse assays.

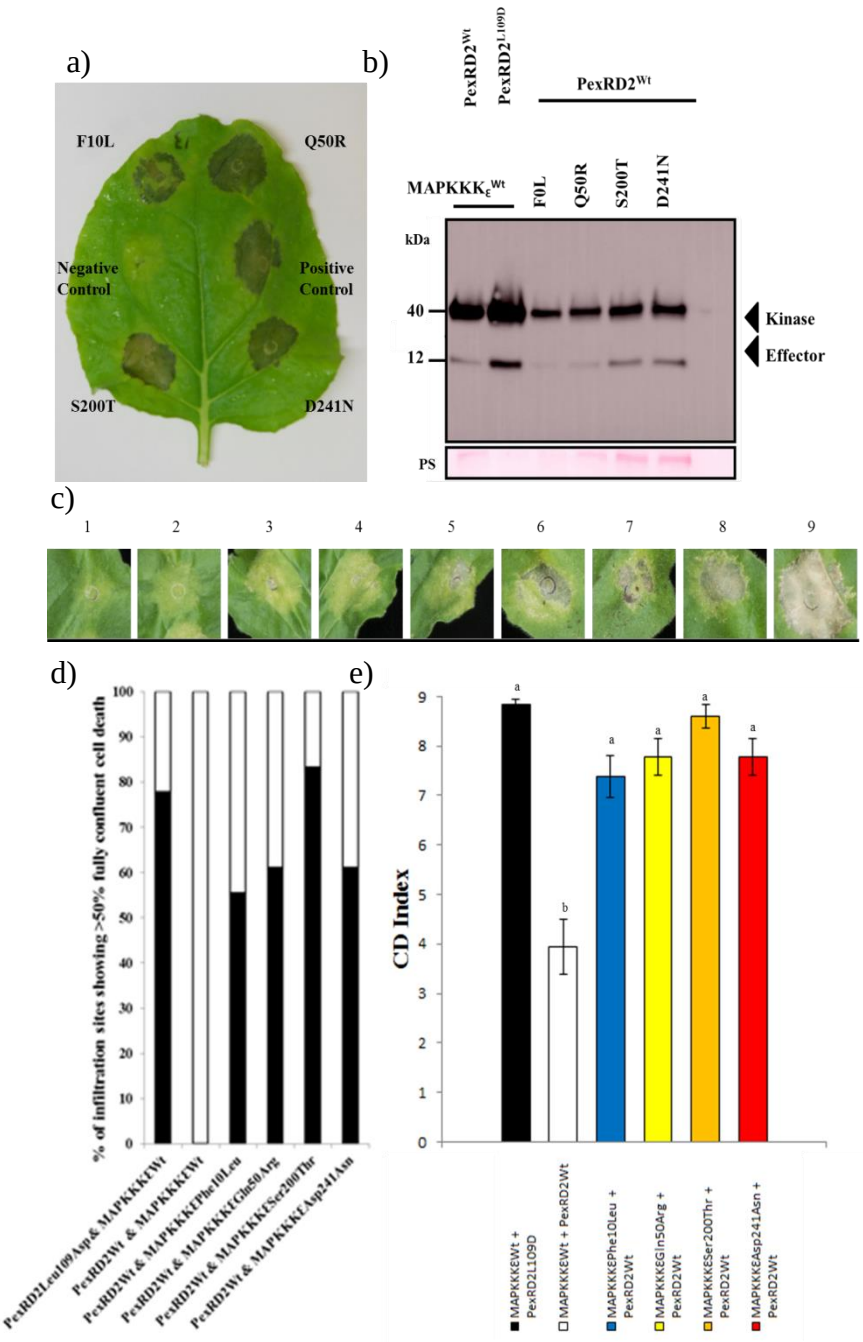



**Figure 9| Analysis of single amino acid polymorphisms for effector insensitivity reveals all four contribute to effector insensitivity.**

a) Representative leaf image showing each amino acid polymorphism tested against PexRD2$^{Wt}$ for effector-insensitivity. b) Western blot from crude N. benthamiana extract probed with anti-HA antibodies demonstrating expression of HA-tagged kinase and effector in effector-insensitivity assays. c) Reference for arbitrary scale of cell death severity ranging from scores of 1 (no response other than mild yellow of leaf from A. tumefaciens infiltration and mechanical wounding) to 9 (fully confluent cell death in 100% of infiltrated area). d) Bar chart showing overall results of single polymorphism effector-insensitivity assay. Bars represent % of infiltration sites exhibiting greater than 50% fully confluent cell death. Bars are split between; black (representing the percentage showing >50% cell death) and white (representing the percentage showing <50% cell death). Results representative of 9 repeats per combination of effector and kinase variant. d) Bar chart showing mean cell death index rating. Based on 9 repeats per construct. Significance was determined by one-way ANOVA and Tukey's HSD test.

### 3.2.4. Attempt to quantify evasion of cell death suppression through ion leakage assays.

As plant tissue senesces, or begins to die, the integrity of the membranes in affected cells is compromised. Therefore, the magnitude of cell death response can be quantified by measuring electrolyte leakage into a conductive medium. The infiltration procedure described in section 3.2.3 was repeated and $MAPKKK_{\varepsilon}$ expression induced at 24-48 h post infiltration. 8mm leaf disks were harvested ~36 h post induction and floated in de-ionised water. Following 1-2 h of incubation at room temperature conductivity readings were taken.

Conductivity readings were taken for 10 replicates per combination and are presented in Figure 10. A mean conductivity reading of 52.6µs/cm ± 2.697 was observed for $MAPKKK_{\varepsilon}^{Wt}$ & $PexRD2^{L109D}$. This differed from the negative control, $MAPKKK_{\varepsilon}^{Wt}$ & $PexRD2^{Wt}$, at the $P < 0.01$ level, with a conductivity reading of 40.3µs/cm ± 1.707. Indicating that, in principle, the assay is sensitive enough to detect differences between plant tissue where PexRD2 is suppressing cell death and tissue where PexRD2 mediated cell death suppression is being evaded. However, conductivity readings in the lower to mid-40µs/cm range for all four effector-insensitive variants of $MAPKKK_{\varepsilon}$ (Phe10Leu = 43.4µs/cm ± 1.845, Glu50Arg = 44.4µs/cm ± 2.5, Ser200Thr = 44.2µs/cm ± 2.337 and Asp241Asn = 41.3µs/cm ± 1.828) suggested that PexRD2 was still capable of suppressing their cell death induction activity, under the present conditions.

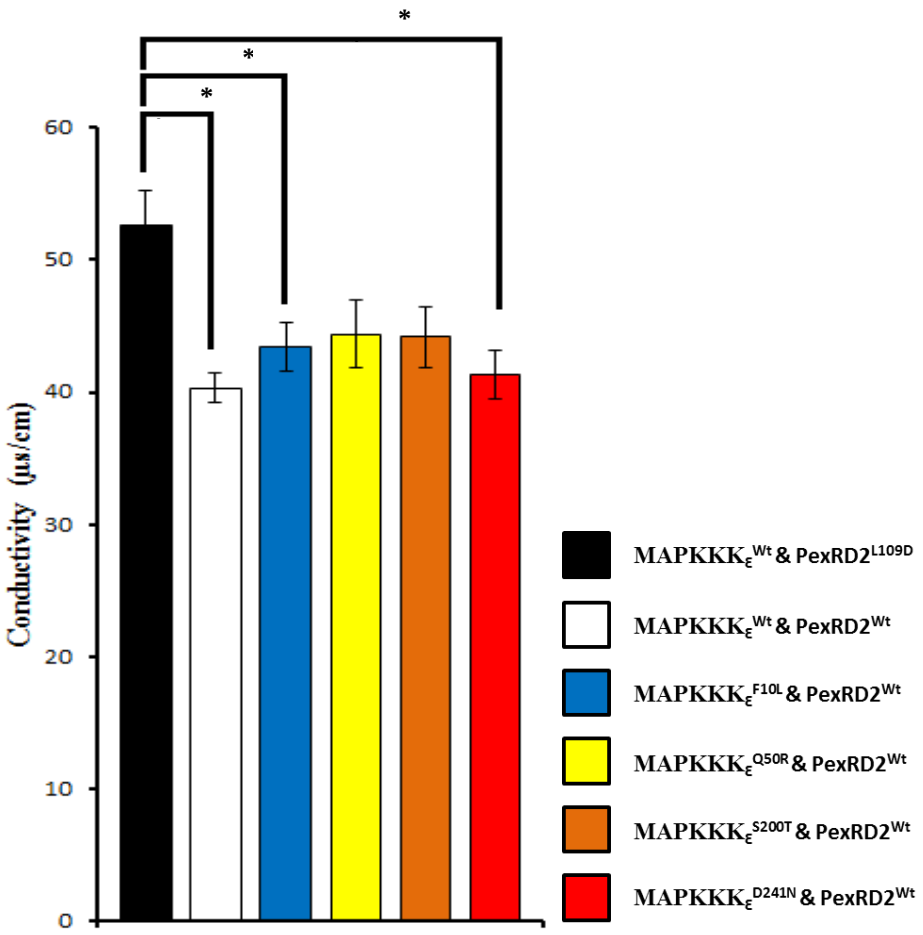



**Figure 10| Electrolyte leakage assays are unsuitable for quantifying insensitivity to PexRD2.**

Bar chart showing conductivity of medium in microsiemens/cm. Individual bars representative of data from 10 replicates per combination of effector and kinase. Significance determined by one-way ANOVA and Tukey's HSD test. * = significant difference at the $P < 0.05$ level. ** = significant difference at the $P < 0.01$ level.

However, previous observations indicate that PexRD2$^{L109D}$ accelerates MAPKKK$_{\varepsilon}$ mediated cell death response when co-expressed with MAPKKK$_{\varepsilon}$ relative to MAPKKK$_{\varepsilon}$ overexpressed alone (King, 2013). This results in a dominant negative phenotype where even the cell death suppressing effects of PexRD2$^{Wt}$ are masked by the accelerated cell death associated with PexRD2$^{L109D}$. The consequence of this dominant negative phenotype in the present assay is that the positive control (MAPKKK$_{\varepsilon}^{Wt}$ co-expressed with PexRD2$^{L109D}$) undergoes a very rapid cell death response, meaning that electrolyte leakage readings must be taken before visible cell death symptoms become too severe, as the assay measures dying cells, rather than dead cells. At this time point (24-48 h post induction) it has been observed that % of infiltrated area exhibiting cell death can be >20% greater than in tissue overexpressing MAPKKK$_{\varepsilon}$ alone. At this point, cell death associated with mutant variants of MAPKKK$_{\varepsilon}$ co-expressed with PexRD2$^{Wt}$ has not significantly exceeded the baseline electrolyte leakage levels attributable to agroinfiltration and mechanical tissue damage from excision of leaf disks.

## 3.2.5. Probing the mechanisms behind effector insensitivity

### 3.2.5.1 Homology modelling of MAPKKK$_\varepsilon$ allows mapping of effector-insensitive mutations

Efforts to purify MAPKKK$_\varepsilon$ by Dr Abbas Maqbool (JIC) repeatedly failed to yield sufficiently high yields of soluble, correctly folded protein for biochemical/structural studies. Therefore, to gain insight into the positions of the MAPKKK$_\varepsilon$ mutations that allow the kinase to evade suppression of cell death signalling, we performed homology modelling using the Phyre2 server (Kelley et al., 2015). Submission of the wild type MAPKKK$_\varepsilon$1-332 sequence returned the Bovine G protein-coupled receptor kinase 5 (Homan et al., 2015) as the top scoring template, with a confidence value of 100%. The final model generated is shown in Figure 11. The position of amino acids conditioning insensitivity to PexRD2 are highlighted as well as key residues in the active site.

Mutations did not cluster to a particular spatial location on the homology model making predictions on how each mutation may perturb PexRD2 mediated suppression of cell death difficult. Indeed, F10L and D241 appear to be particularly distant from the active site. However, it is possible that mutations distal to an enzyme's active/binding site can still alter its conformation enough to perturb activity/binding. It is interesting to note that Q50R and S200T map closer to the active site, and that these two mutations conditioned enhanced insensitivity relative to F10L and D241N in transient cell death suppression assays.

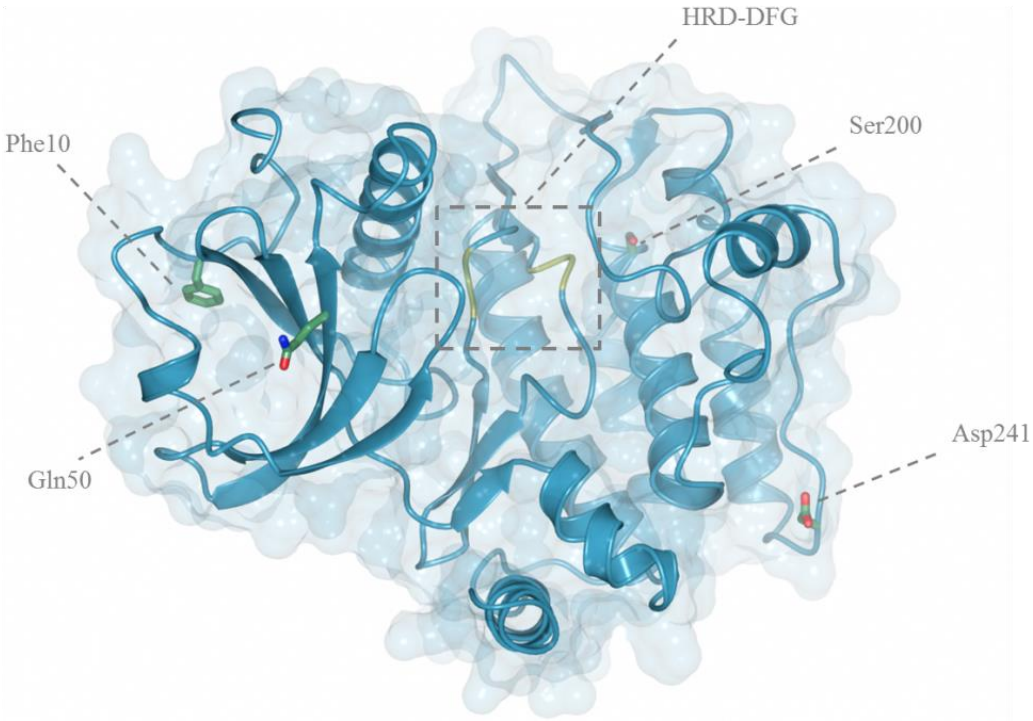



**Figure 11| Homology model of MAPKKKƐWt (1-332).**

Phyre2 3D structure prediction for MAPKKK$_{\varepsilon}$, rendered in ccp4mg. Location of mutations conditioning insensitivity to PexRD2 are indicated, as well as the location of the catalytically active HRD-DFG motif. Surface map of MAPKKK$_{\varepsilon}$ homology model is superposed on top of cartoon structure in order to indicate whether or not key residues are solvent exposed or buried. Ser 200 is the only residue not solvent exposed.

Rendering of the homology model to show surface exposed residues revealed that three of the mutated residues; 10, 50 and 241, are surface exposed, whilst residue 200 is buried (Figure 11). The phenylalanine to leucine mutation at position 10 consists of the replacement of an aromatic, hydrophobic side chain with an aliphatic, non-polar one. Replacement of Phe with another aromatic amino acid – such as tyrosine – and with another simple, aliphatic non-polar amino acid like valine may shed light on whether or not there is a requirement for an aromatic side-chain at position 10 for interaction with PexRD2.

Glutamine to arginine at position 50 consists of the replacement of an uncharged, carboxamide group presented at the end of two methylenes as a side chain with a positively charged guanidinium group presented at the end of a long side chain. This could indeed perturb any interaction with PexRD2. Much of the solvent exposed

surface of the PexRD2 dimer is positively charged, with a large number of lysine residues being evident. Conversion of an uncharged residue (such as glutamine) to a positively charged residue could potentially repel interactions with the positively charged residues exposed on the surface of PexRD2. Mutagenesis studies involving substituting glutamine50 for another uncharged residue, such as asparagine, and for another positively charged residue, such as lysine, may reveal whether or not the presence of a positively charged residue at this position generally interferes with effector-kinase binding.

Aspartate presents a hydrophilic, negatively charged (at physiological pH) carboxylate group as its side chain. Mutation of this residue to the similar, but uncharged (the only difference is that aspartate's carboxylate group is replaced with a carboxamide group in asparagine) asparagine may perturb interaction by replacing a negatively charged residue which may be important for mediating interactions with the positively charged surface of PexRD2. Experiments substituting glutamate (also negatively charged) for Aspartate should reveal whether or not a negatively charged residue at position 241 is required for interaction with PexRD2.

The relevance of a serine to threonine mutation at position 200 is more challenging to explain in terms of its direct relevance for interaction with PexRD2 as (a) it is not surface exposed and (b) Ser and Thr are extremely similar amino acids, with Thr possessing just one extra methyl group relative to Ser. However, it has previously been observed that mutations to buried residues can alter protein-protein interaction (McLaughlin et al., 2007). Replacement of Ser with amino acids containing far bulkier side chains may provide insights into whether or not mutations at this residue are somehow altering overall protein conformation and interfering with MAPKKK$_\varepsilon$'s interaction with PexRD2.

#### 3.2.5.2 **Yeast two-hybrid assays reveal weakened interactions between PexRD2 and effector-insensitive kinase variants**

Whilst the ultimate goal in generating effector-insensitive MAPKKK$_\varepsilon$ variants is to explore their utility in generating disease resistance in crops, the nature of insensitivity is an interesting and relevant biological question. The simplest

hypothesis is that MAPKKK$_\varepsilon$ variants are able to evade cell death suppression activity of PexRD2 by preventing the effector from binding directly to the kinase.

As Y2H is an established assay for monitoring the interaction between PexRD2 and MAPKKK$_\varepsilon$, we employed this technique to investigate the interactions between the effector and PexRD2-insensitive MAPKKK$_\varepsilon$ variants. To confirm previous results, PexRD2$^{Wt}$ and PexRD2$^{L109D}$ were fused to the GAL4-DBD in pEXP32™ and used to co-transform MaV203 *S. cerevisae* alongside either *St*MAPKKK$_\varepsilon^{Wt\ 1\text{-}300}$, *St*MAPKKK$_\varepsilon^{(M)\ 1\text{-}332}$, *St*MAPKKK$_\alpha$ fused to the GAL4-AD in pEXP22™. Co-transformants were selected on SC-LW plates and three individual co-transformants were selected and plated onto SC-LWH + 3-AT plates for –histidine auxotrophy assays, onto SC-LWU plates for –uracil auxotrophy assays or onto nylon membranes overlaying YPAD plates for X-Gal assays. The results of yeast 2-hybrid reporter assays are shown in Figure 12a, confirmation of expression of GAL4-DBD fused constructs by western blot is shown in Figure 12b and confirmation of expression of GAL4-AD fused constructs is shown in figure Figure 12c. In –histidine (+10mM 3-AT) auxotrophy assays, no clear differential between MAPKKK$_\varepsilon^{Wt}$ and the four MAPKKK$_\varepsilon$ variants was observable, confirming that they are capable of interaction in yeast. The –histidine assay was also performed with concentrations of 20mM and 30mM 3-AT with similar results (data not shown). However, in the most stringent auxotrophy assay - -uracil – a clear differential can be observed after 5 days growth. MAPKKK$_\varepsilon^{Wt}$ shows clear growth on –LWU plates relative to PexRD2 insensitive mutants when co-expressed with PexRD2$^{Wt}$. This observation was also recapitulated in colonies grown on nylon/YPAD plates and subjected to the X-GAL assay. These results clearly indicate a weakened, but not fully abolished interaction between PexRD2 and insensitive MAPKKK$_\varepsilon$ variants. However, a weakened interaction in a heterologous host such as *S. cerevisiae* may be sufficient to account for insensitivity to PexRD2 *in planta* and establishment of an *in planta* assay with which to test MAPKKK$_\varepsilon$ mutants may reveal a totally abolished interaction. A probable false positive interaction was observed in the –uracil auxotrophy assay where PexRD2$^{L109D}$ and MAPKKK$_\varepsilon^{F10L}$ were co-expressed. This was not observed in the other two reporter assays (including the less stringent –histidine assay). Furthermore, it was not observed in any previous repeat of the assay so is unlikely to represent a

*bona fide* interaction. Confirmation of expression of all constructs was vital to confirm that negative results from reporter assays were the result of non-interaction, as opposed to lack of protein expression.

We attempted to establish *in planta* Co-IP assays with which to test the interaction of the PexRD2 insensitive mutants with PexRD2. However, even the interaction between wild type effector and kinase has proved difficult to establish despite extensive attempts at optimisation.

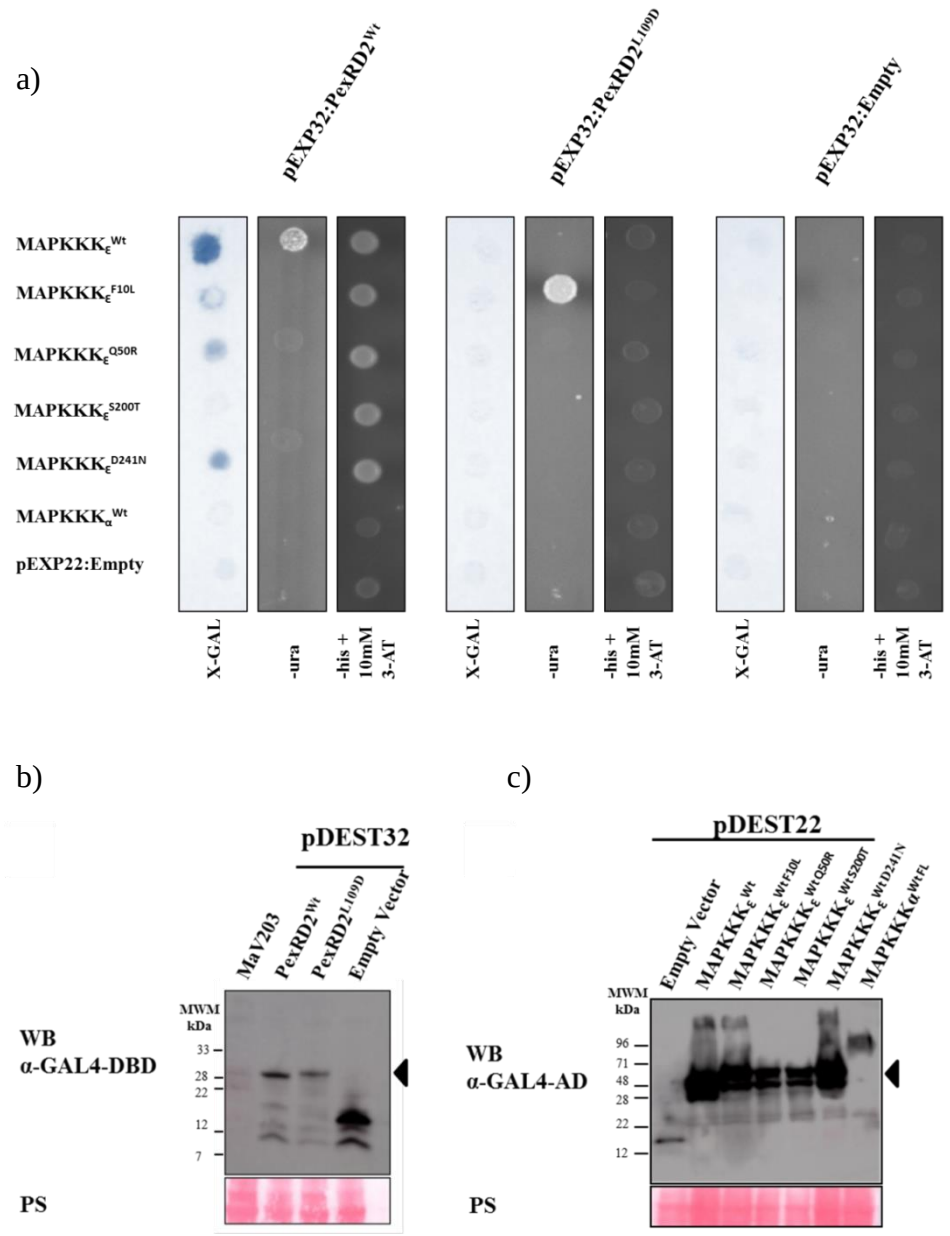



**Figure 12| PexRD2-insensitive kinase variants exhibit weakened interaction with PexRD2 in yeast 2-hybrid Uracil auxotrophy assay but not Histidine auxotrophy assay.**

a) –LWU and –LWH (+ 10mM 3-AT) reporter assays. All interactions were tested in triplicate with replicates derived from independent MaV203 co-transformant colonies selected on –LW plates. –LWU plates were incubated for 5 days at 28˚C and –LWH plates incubated for 24hrs at 28˚C followed by 16hrs at room temperature. –LWH assays were conducted with a range of 3-AT concentrations from 10-30mM with the same results. b) Western blot confirming expression of GAL4-DNA Binding Domain tagged proteins. Membrane probed with α-GAL4-DBD antibodies conjugated to HRP for 2hrs at room temperature with shaking. c) Western blot confirming expression of GAL4-Activation Domain tagged proteins. Membrane probed with α-GAL4-AD antibodies conjugated to HRP for 2hrs at room temperature with shaking.

## 3.2.6. CRISPR/Cas9 mediated editing of *Solanum lycopersicum* MAPKKK$_{\varepsilon}$

### 3.2.6.1 **Design of sgRNAs and complementation constructs**

*(Note – sgRNAs were designed with Dr Nicola Patron and constructs were prepared by Mr Mark Youles)*

The importance, and regulation, of eukaryotic protein kinases has been discussed at length in section 1.4. The amino acid residues required for kinase activity are well defined and highly conserved. Two motifs known to be crucial for activity are the HXD motif and the DFG motif. The DFG motif is known to be important for the docking of ATP whilst the HXD motif is involved in establishing and maintaining the correct orientation of substrates at the P-site, correct orientation of the activation segment (in RD kinases where X = R in the HXD motif) and as a structural co-ordinator of several other key regulatory elements (Zhang et al., 2015, Kannan and Neuwald, 2005). Mutations in many of these residues can abrogate or abolish kinase activity. Therefore, knockout of the entire region will clearly abolish kinase activity, whilst INDELS which induce frameshift mutations at individual sgRNA sites should also abolish kinase activity. We, therefore, attempted to knock out endogenous MAPKKK$_{\varepsilon}$ and complement knockout tomato plants with PexRD2-insensitive MAPKKK$_{\varepsilon}$ variants in one stable transformation experiment incorporating sgRNAs against MAPKKK$_{\varepsilon}$, Cas9 and a complementation cassette.

Pairs of sgRNAs were designed to guide Cas9 to sequences flanking either the HRD motif alone (sgRNA1 and sgRNA3) or the HRD and DFG motif together (sgRNA2 and sgRNA4) in order to target MAPKKK$_{\varepsilon}$. Sequences for each sgRNA are shown in Figure 13 as well as a schematic representation of the loci targeted by each sgRNA. Each sgRNA targets a sequence with a 3' protospacer adjacent motif (PAM), required for Cas9 mediated cleavage of the target sequence.

sgRNA1-3

1) *AAATTTTGATACTTGTTTAA*
3) *ATCTATATCTGAATTCACAC*

sgRNA 2-4

2) *CTTCTGTTGTAGGTATTGGA*
4) *TGTTGCAACTAAATTAACGG*

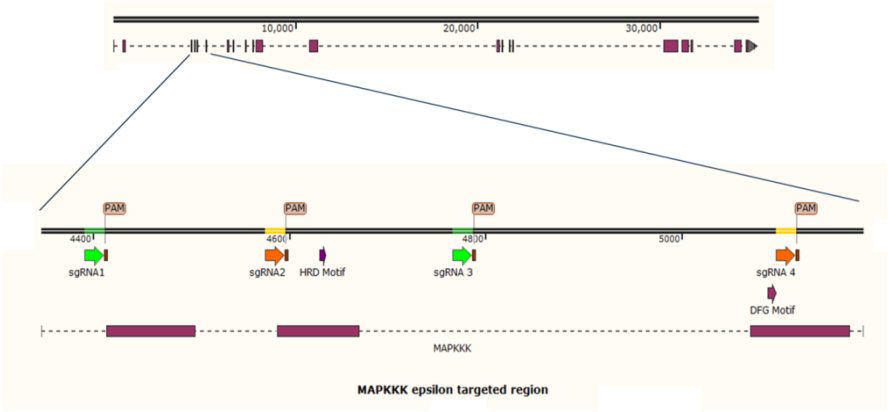



**Figure 13| sgRNAs targeting a genomic region containing MAPKKKε catalytically active residues.**

Guide RNAs were paired so as to induce large deletions through error prone non-homologous end joining (NHEJ) or to induce smaller indels upstream of PAM sites by the same DSB repair mechanism. Sequences are shown above in the 5' to 3' orientation. sgRNA1 and sgRNA3 target sites are shown by green arrows upstream of an NGG PAM whilst sgRNA2 and sgRNA4 target sites are shown by orange arrows. The catalytically active motifs – HRD and DFG – targeted by these guide RNAs are shown as purple arrows. CDS are represented by purple boxes. The targeted region is shown magnified from its overall genomic context on chromosome 11 in the *S. lycopersicum* genome.

An inherent problem in using stably expressed sgRNAs to guide targeted double stranded breaks or indels in a gene of interest, where the goal is to then express a variant of the gene of interest which differs by only a single amino acid residue is that the sgRNAs will also target the complementation sequence. To avoid this, the redundancy of the genetic code was exploited to re-code the MAPKKK$_\varepsilon$ sequence targeted by sgRNA4. sgRNAs 1,2 & 3 had recognition sequences which partly, or completely, lay in introns. As the complementation cassette was comprised of only the CDS of MAPKKK$_\varepsilon$ no further re-coding was required other than to remove any BpiI or BsaI sites which may interfere with Golden Gate assembly, a process known as domestication. This re-coding of DNA within complementation cassettes also facilitated the design of oligonucleotides able to differentiate between mRNA derived from transgenic complementation cassettes and mRNA derived from endogenous MAPKKK$_\varepsilon$. This was critical in order to confirm expression of the transgene complementation cassettes and correlate any enhanced resistance with expression of MAPKKK$_\varepsilon$ mutants.

#### 3.2.6.2 **Transient screening of paired sgRNAs**

In order to ensure the efficacy of sgRNA pairs prior to stable transformation, each sgRNA pair was tested in transient assays having been delivered into leaf tissue *via A. tumefaciens* infiltration. Each sgRNA pair was expressed from the same binary vector also encoding the Cas9 nuclease under the control of a double CaMV 35S promoter. Plasmids carrying sgRNAs, under the control of a synthetic U6 RNA polymerase III promoter, were transformed into *A. tumefaciens* strain AGL1. Positive transformants were grown overnight in liquid media, incubated with $MgCl_2$ and acetosyringone before delivery into the leaves of 3-4 week old tomato plants through a blunt syringe at the site of small wounds generated by hypodermic needles on the abaxial face of the leaf. *A. tumefaciens* cells carrying sgRNAs were co-delivered with bacteria harbouring the TBSV p19 RNA silencing suppressor. Leaf

tissue was harvested 48-72 h post infiltration and flash frozen in liquid $N_2$. Genomic DNA was extracted as described in section 2.10.5.

In order to enrich for successfully edited DNA, total gDNA was digested with EcoRV (Roche). EcoRV's cut site occurs just 3' of the HRD motif of MAPKKK$_\varepsilon$, between each pair of sgRNAs. Successfully edited DNA should be resistant to EcoRV mediated cleavage and be amplified by primers flanking sgRNA target sequences whereas unedited DNA should be cleaved by the restriction enzyme. EcoRV treated gDNA was then amplified by PCR with primers flanking the target sequences of sgRNA 1 and sgRNA 4, before being run on a 3% agarose gel (Figure 14). Band shifts of ~390bp and ~540 bp if sgRNA 1 & 3 and sgRNA 2 & 4 were active were predicted, respectively. Wild type, unedited leaf material shows two clear bands at ~850bp corresponding to the expected amplified region in the previously described PCR. sgRNA 1-3 showed no observable editing activity whilst sgRNA 2 & 4 yielded a band of 250-300bp, corresponding to the expected product size if sgRNA 2-4 were both active. A wild type band of ~790bp is visible in the sgRNA 2-4 edited, EcoRV treated lane, probably due to incomplete digestion by EcoRV. In addition, an unexpected high molecular weight band is visible above the 790bp band. Anecdotally, this is frequently observed where CRISPR/Cas9 mediated editing activity is high. The causes behind this are currently unclear although it has been proposed that it could be the result of various genetic elements from binary plasmids inserting between the double stranded breaks (Nicola Patron & Oleg Raitskin, personal communication). Sequencing of the band at around 250bp confirmed it was MAPKKK$_\varepsilon$ although the read stopped short of the first predicted DSB site, meaning determination of the exact nature of the knockout was not possible. However, we were confident from this that sgRNA2-4 were active and induced a large deletion in the MAPKKK$_\varepsilon$ gene.

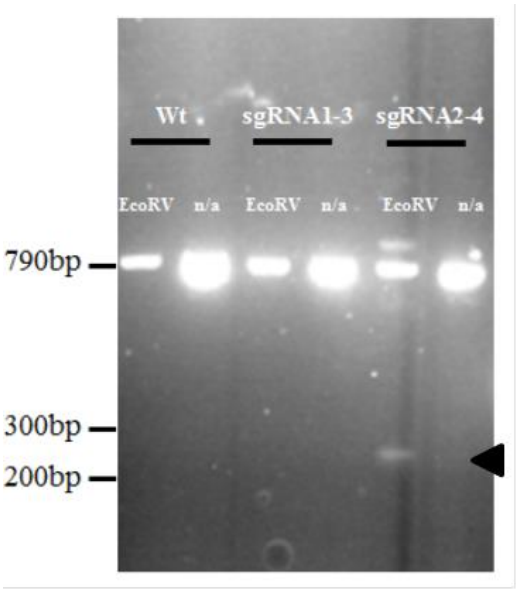



**Figure 14| Gel electrophoresis analysis of undigested and digested gDNA from *S. lycopersicum* leaf tissue transiently infiltrated with *A. tumefaciens* harbouring Cas9 nuclease and sgRNAs against tomato DNA.**

Untransformed, Wt tomato tissue is used as a negative control and shows expected 790bp band in both digested and undigested samples. Tissue infiltrated with sgRNA1-3 does not exhibit any editing activity whilst EcoRV treated gDNA from tissue infiltrated with sgRNA2-4 shows expected 250bp band corresponding to a large deletion between the two sgRNA target sites (indicated by black arrowhead) as well as a larger band indicative of a large insertion between the sgRNA target sites. Samples were analysed on a 3% (w/v) Agarose TAE gel.

### 3.2.6.3 **Screening of stable transgenic tomato plants for CRISPR/Cas9 activity in *S. lycopersicum***

Transgenic plants were prepared by Matthew Smoker and Jodie Pike (TSL Tissue Culture & Transformation team) as described in section 2.10.2**.**

Primary transformants ($T_0$) were screened for the large deletion as described in section 3.5.2.3. Leaves were harvested and flash frozen in liquid $N_2$ before being finely ground with a pre-chilled micro-pestle. gDNA was extracted using the CTAB method described in section 2.10.5. Untreated gDNA was taken to PCR using the primer pairs described in section 3.5.2.3 and PCR products were run on a 2% agarose gel (Figure 15).

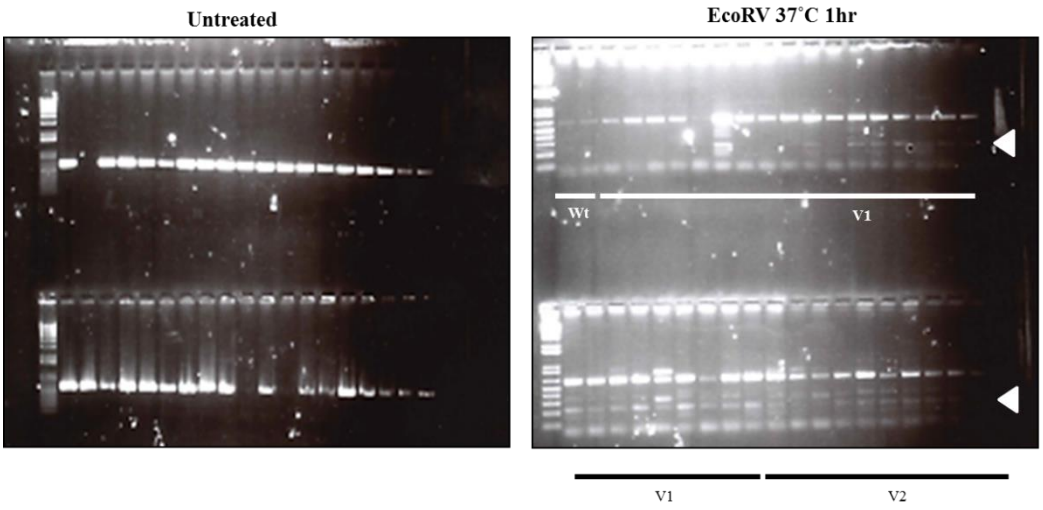



**Figure 15| EcoRV treatment of gDNA from plants stably transformed with CRISPR/Cas9 machinery directed against MAPKKKε reveals presence of desired large deletions at low frequency.**

Left hand panel shows PCR products from untreated gDNA, right hand panel shows PCR products from gDNA pre-digested with EcoRV for 1hr at 37˚C. White arrowheads indicate expected size (~250bp) of large deletion between sgRNA sites. Presence of strong 'wild type' band in both treated and untreated samples indicates that large deletions are relatively rare events in the tissue sampled.

Additionally, gDNA was subjected to EcoRV treatment for 1hr at 37˚C as described previously before being used as a template in the PCR, using the same primer pair. PCR products were run on a 2% agarose gel (Figure 15). Large deletions between sgRNA 2 and sgRNA4 are detected where gDNA is pre-treated with EcoRV. However, the fact that this large deletion is not detected where gDNA is untreated prior to PCR suggests that the transgenic plants are chimeric in terms of CRISPR/Cas9 expression and activity. If this deletion was prevalent in most cells in the plant, we would expect to amplify the characteristic 250bp band even where gDNA was not pre-digested with EcoRV. Indeed, amplification of the large, 790bp band where gDNA is pre-digested is indicative that the large deletion is not especially prevalent. This is not surprising as larger deletions between two, individual guide RNAs are much rarer events than smaller indel mutations at individual guide RNA sites. Nevertheless, the small band corresponding to a large deletion between the paired guide RNAs was cloned into pENTR for sequencing and revealed a 516bp deletion, which was the product of DNA cleavage exactly 3/4bp upstream of PAM for each sgRNA site. The sequence of this large deletion, aligned to the wild type sequence, is shown in Figure 16.

Editing activity of individual guide RNAs was assessed by direct sequencing of the cleaned-up, 790bp PCR product amplified from digested, and undigested, gDNA. This revealed a number of indel mutations close to the PAM from deletions to insertions (Figure 17). In order to determine the zygosity of these indels, sequencing chromatograms were assessed (Figure 17). Where a mutation is detected, clean chromatogram peaks indicate the mutation is homozygous in the cells analysed, a peak with another peak under it indicates the mutation is heterozygous whilst a clean series of peaks which degenerates into essentially unreadable sequence immediately before the PAM is indicative that the tissue sampled was highly chimeric and contains a wide range of mutations in addition to unedited cells. DNA yields from gel extractions were typically poor, therefore only a small sample of indels could be detected as sequencing quality DNA was difficult to obtain and time was limited at this stage of the project. DNA obtained through cloning of bands into plasmids could have revealed more about the range of indels obtained but would not have yielded any information on their zygosity in the tissue sampled.

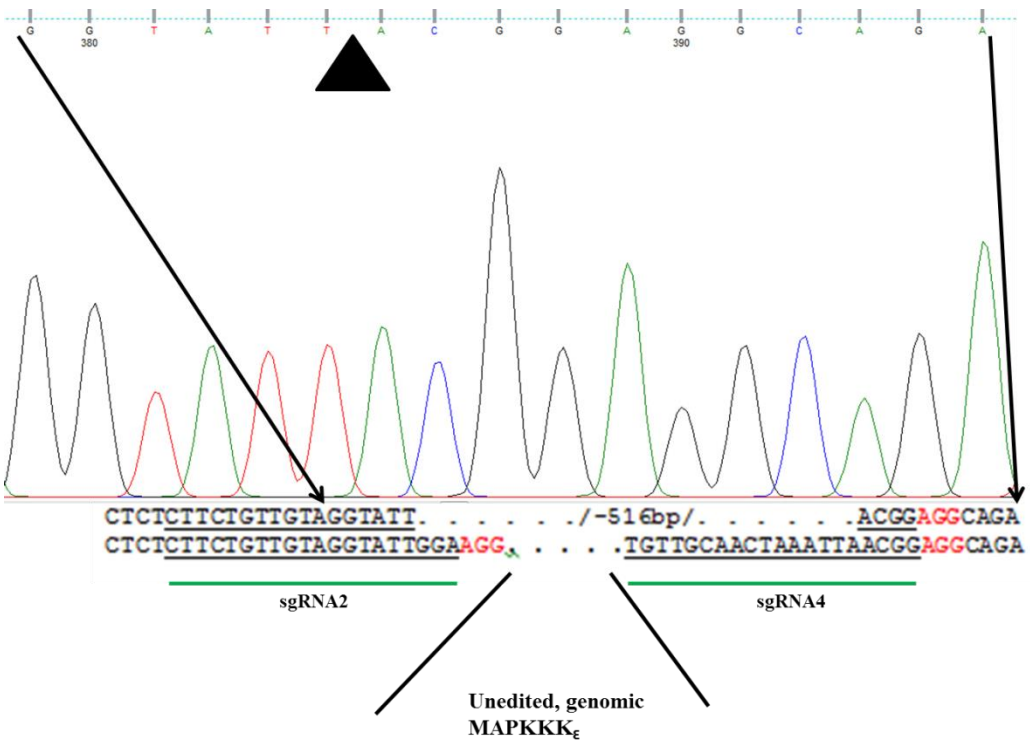



**Figure 16| Sequencing of CRISPR/Cas9 – sgRNA2-4 mediated large deletion.**

Upper sequence in alignment corresponds to deletion band, lower sequence corresponds to wild type, unedited band. Sequence complementary to sgRNAs is underlined and PAM sites are highlighted in red. Large black triangle in chromatogram image indicates border between gDNA joined by NHEJ following Cas9 induced DSB and chromatogram indicates clean border between re-joined DNA strands.

Further to this, the spatial and temporal control of sgRNA expression during regeneration of whole plants is unlikely to be uniform. It is more likely that a range of indels at individual target sequences and larger deletions between sgRNA target sequences exist in each individual plant. Thus, any effort to screen CRISPR/Cas9 activity in the $T_0$ generation was always likely to simply sample the potential activity of sgRNA2 and sgRNA4 guided Cas9 nucleases in stably transformed plants. The stable inheritance of CRISPR/Cas9 mediated deletions or indels will depend on the presence of edited DNA in the germLine of primary transformants. Therefore, all progeny of self-fertilised $T_0$ plants will need to be collected, germinated and screened for CRISPR/Cas9 mediated editing.

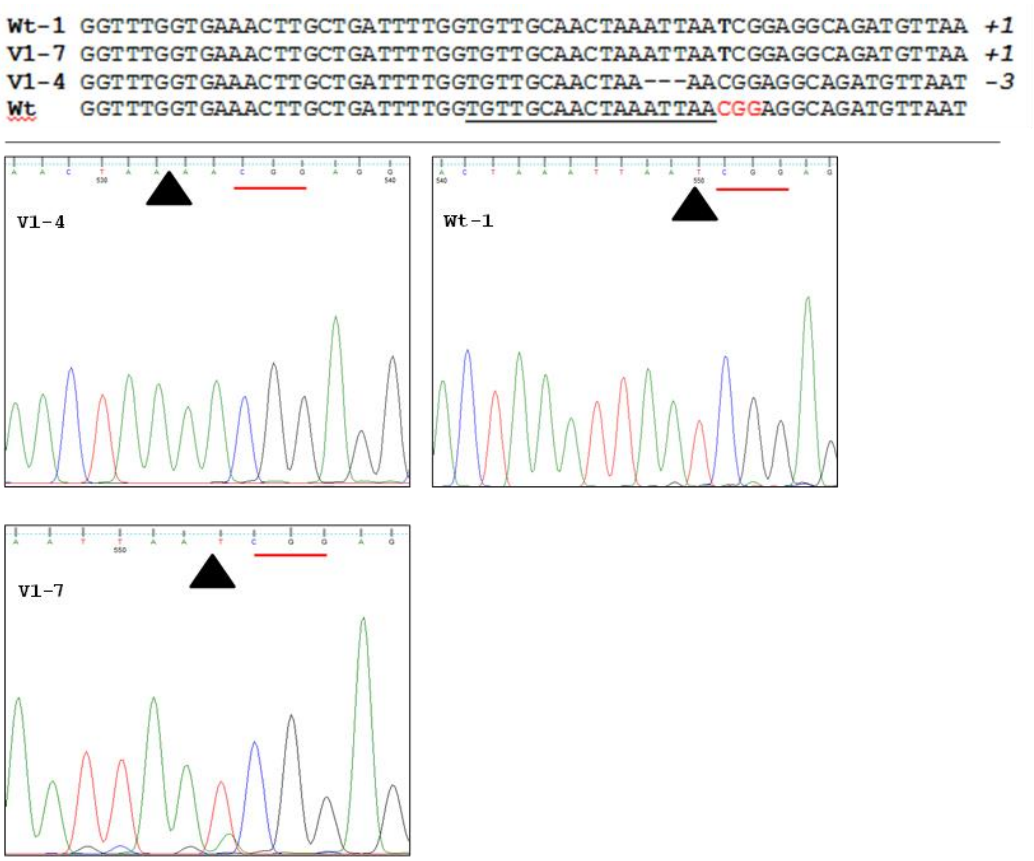



**Figure 17| Sequencing reveals indels at individual sgRNA target sequences.**

Representative sample of indels detected at sgRNA4 target sequence. *Wt-1* and *V1-7* both showed a T insertion 1bp upstream of the PAM. This results in a frameshift causing a premature stop codon. Sequencing chromatogram reveals that *Wt-1* is homozygous for this mutation and *V1-7* is heterozygous. *V1-4* had a 3bp deletion 2bp upstream of the PAM. This results in the deletion of one residue and does not cause any frameshift in the rest of the protein. Sequence chromatogram reveals that *V1-4* is homozygous for this mutation. sgRNA4 target sequence is underlined in the 'Wt' refererence sequence and 'CGG' PAM is highlighted in red. PAM location is underlined in chromatogram images and locations of insertions or deletions are indicated by black arrowheads.

3.2.6.4 **Screening of transgenic tomato plants for transgene expression**

RNA extraction from leaves of $T_0$ plants was performed using RNeasy Plant Mini columns (Qiagen) as per the manufacturer's protocol, outlined in section 2.5.5 using pre-ground tissue. One modification was made in that an on-column DNase digest, using the RNase-free DNaseI kit (Qiagen) was performed to remove residual gDNA. Purity of RNA extracts was determined on a NanoVue (GE Healthcare) and all samples returned A260/A280 ratios of 2.0 ± 0.1, indicating good RNA purity.

Gene specific (GS) primers capable of distinguishing between transgenic mRNA and endogenous mRNA (based on 3bp mismatches at their 3' ends, two of which are A:G mismatches known to have a severe impact on primer efficiency even when present singly (Stadhouders et al., 2010) were used to generate 1$^{st}$ strand cDNA using the SuperScript-RT enzyme with the procedure outlined in section 2.5.5. cDNA was then amplified using the GS primers from cDNA synthesis as reverse primers and a GS forward primer to detect expression of transgenic mRNA. Efficacy of GS primers in distinguishing between transgene and endogenous mRNA was confirmed by the mutant specific primer's inability to amplify MAPKKK$_\varepsilon$ from gDNA extracted from untransformed *S. lycopersicum* cv. Moneymaker (see Supplementary figure **1**). Expression of transgene in the primary transformants tested is shown in Figure 18.

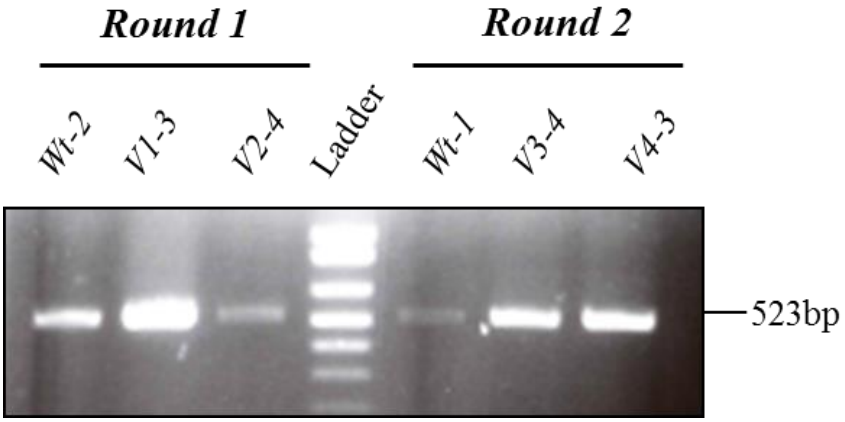



**Figure 18| Reverse-transcription PCR shows expression of complementation transgene mRNA.**

Gel shows bands of 523bp corresponding to the predicted amplified product from PCR with mutant specific primers described in section 2.3.6.5. 'Round 1' and 'Round 2' refer to plants made in successive rounds of transformation. *Wt* plants expressing a wild type complementation cassette. *V1*, *V2*, *V3* and *V4* refer to plants expressing F10L, Q50R, S200T and D241N complementation cassettes, respectively.

### 3.2.7. Infection of detached leaves with *P. infestans*

Leaves of *S. lycopersicum* primary transformants 5 weeks post transplanting were detached and placed in trays at 100% relative humidity as described in section 2.10.7. Infection was allowed to proceed at 18˚C with images taken of detached leaves at approximately 24 hour intervals up to 6dpi. Individual lesion areas were measured in pixels and converted into $mm^2$. Representative images are shown in Figure 19a and Figure 20a and mean lesion sizes from 1dpi to 6dpi are shown in Figure 19b and Figure 20b.

As early as 3dpi, a greater than twofold reduction in mean lesion size is observed in plants expressing $MAPKKK\varepsilon^{Q50R}$ relative to wild type and $MAPKKK\varepsilon^{F10L}$ (which do not differ significantly from each other in terms of mean lesion size). Pathogen growth remains markedly restricted in Q50R plants throughout the time course, with mean lesion size always being significantly lower than that observed in the wild type. This indicates that $MAPKKK\varepsilon^{Q50R}$ may confer enhanced resistance to *P. infestans* 88069 in the tested line.

In separate assays including $MAPKKK\varepsilon^{S200T}$ and $MAPKKK\varepsilon^{D241N}$ (V3 and V4, respectively) compared to wild type, the effector-insensitive variants appeared to condition enhanced susceptibility to *P. infestans* over a 6 day time course. Representative leaf images are shown in Figure 20 and mean lesion size from 1dpi to 6dpi is shown in Figure 20b.

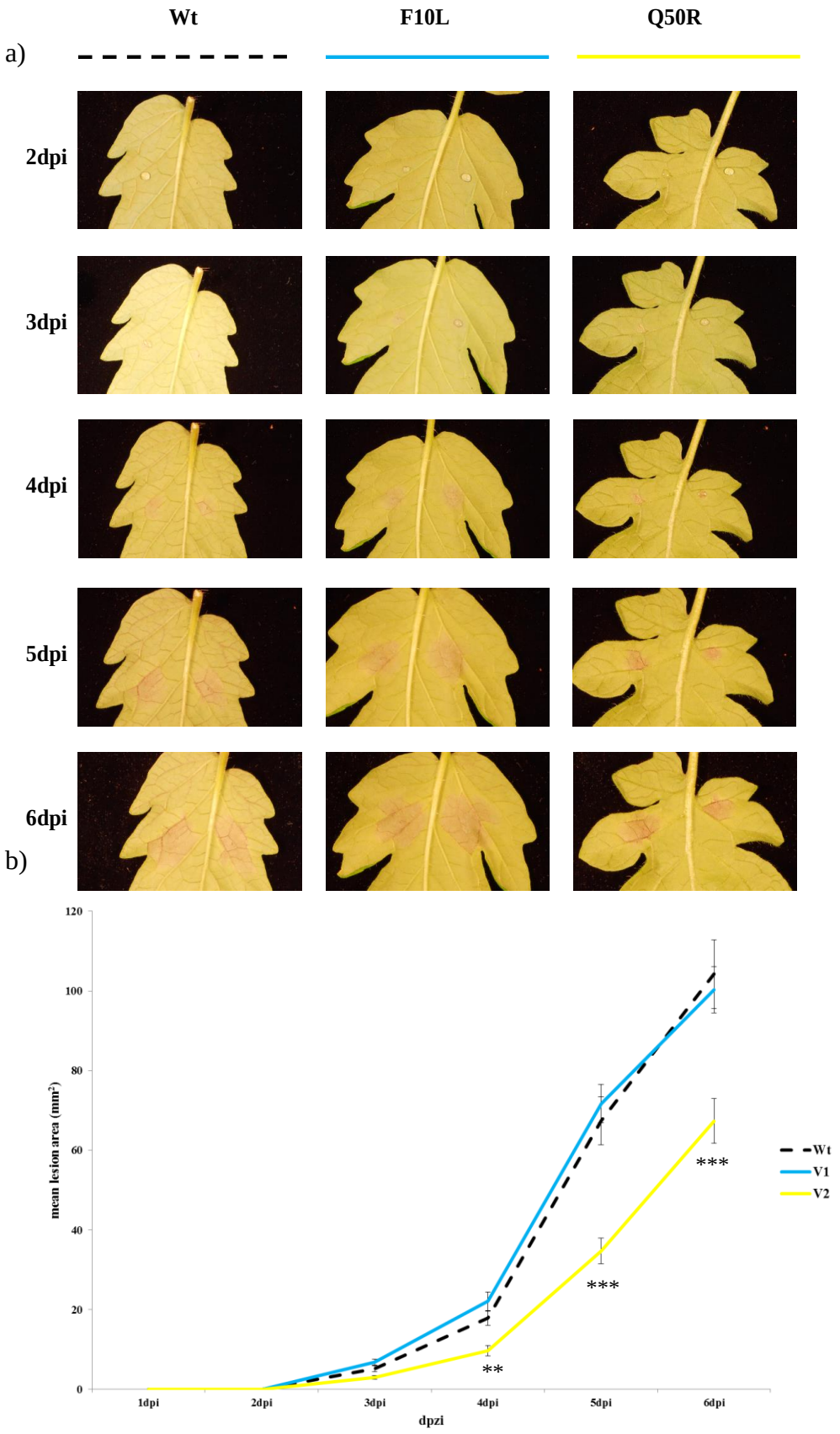



**Figure 19| Growth of *P. infestans* 88069 is visibly restricted on tomato plants expressing MAPKKK$_{\varepsilon}^{Q50R}$.**

(a) Representative images following infection sites from 2-6dpi are shown. (b) Each data point represents mean ($n$ =20) lesion area (mm$^2$) at a given time point in 24hr intervals post inoculation. Error bars = ± standard error. Significant differences determined by one-way ANOVA and Tukey's HSD test. ** = significantly different from the wild type mean at the $p<0.05$ level, *** = different from the wild type mean at the $p<0.01$ level.

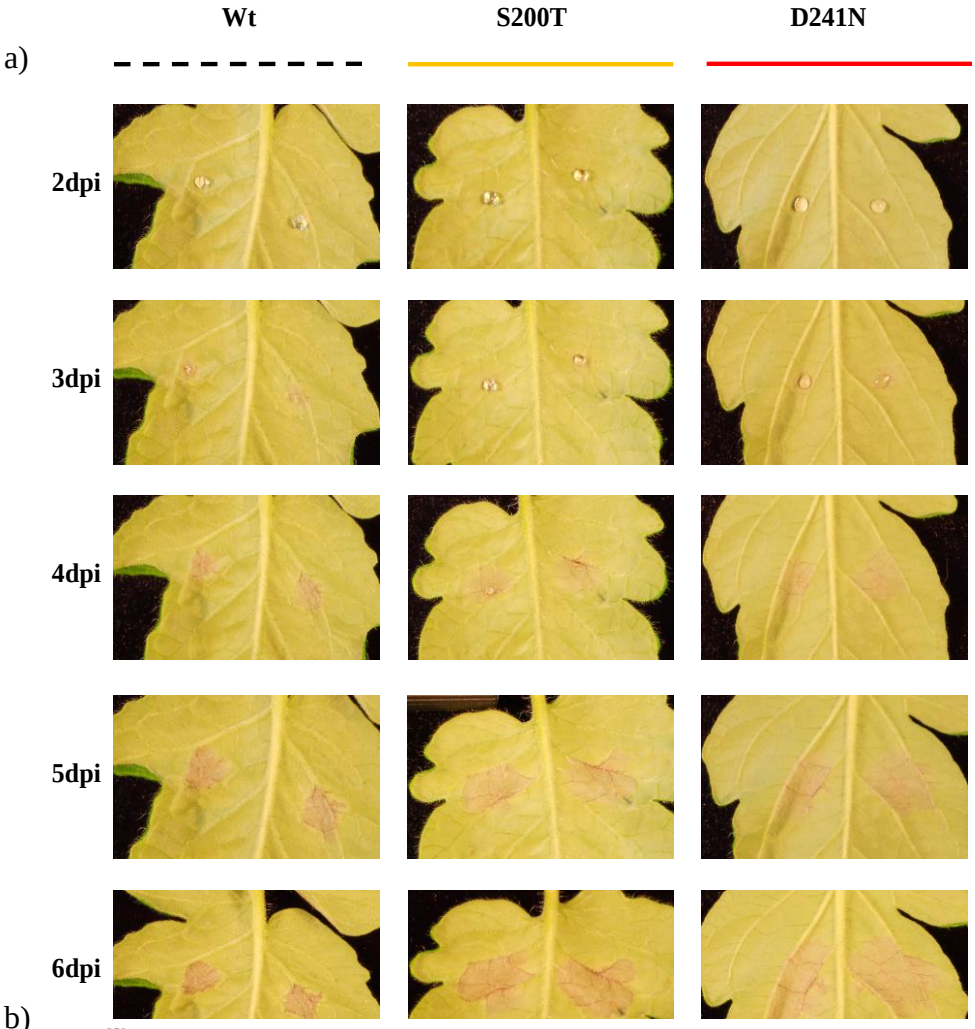



**Figure 20| Growth of *P. infestans* 88069 is visibly enhanced on tomato plants expressing MAPKKK$_{\varepsilon}^{S200T}$ and $^{D241N}$.**

(a) Representative images following infection sites from 2-6dpi are shown. (b) Each data point represents mean ($n$ =20) lesion area (mm$^2$) at a given time point in 24hr intervals post inoculation. Error bars = ± standard error. Significant differences determined by one-way ANOVA and Tukey's HSD test. ** = significantly different from the wild type mean at the $p<0.05$ level, *** = different from the wild type mean at the $p<0.01$ level. ***a/b = means that differ significantly at the $p<0.01$ level.

### 3.2.8. Discussion

In an increasingly globalised world, with crops and produce moving freely between borders and an ever increasing population to feed, strategies to ensure our continued food security will become increasingly vital over the coming years. Since humans first began cultivating crops, resistance has been informally selected for and since the identification of gene-for-gene resistance it has been formally selected for by breeders. However, in the face of intensifying agriculture and a preponderance for growing most staple crops in monoculture, populations of adaptable plant pathogens have been quick to evolve resistance to NLRs when exposed to them for only a few growing seasons (Rouxel et al., 2003). The manipulation of host-targets and susceptibility factors in order to generate resistance to pests and pathogens has been an overlooked strategy until recently. A reluctance to investigate this strategy may stem from concerns that manipulation of targets of effector proteins may interfere with their normal cellular function, given that a unifying feature of these targets is their importance in host cell biology.

MAP Kinases and PRRs possessing a kinase domain have emerged as a common target for many plant pathogen effectors given their critical role in signal transduction of invasion pattern perception and, so, their role in positively regulating resistance to pests and pathogen (Zhang et al., 2007, Cui et al., 2010, Wang et al., 2010). The solanaceous MAPKKK$_{\varepsilon}$ has previously been characterised as a positive regulator of immunity to bacterial, fungal and oomycete pathogens (Melech-Bonfil and Sessa, 2010b, King, 2013, King et al., 2014) and is a target of the *P. infestans* RXLR effector PexRD2 (King et al., 2014). It is thought that PexRD2 does so in order to suppress MAPKKK$_{\varepsilon}$ mediated cell death, a potent defence response to a hemibiotrophic pathogen like *P. infestans*.

The previous characterisation of the interaction between PexRD2 and MAPKKK$_{\varepsilon}$ and its outcomes provided the resources with which to generate a variant of this virulence target which is refractory to the effector and continues to perform its

normal cellular function despite the effector's presence. Ideally, a structurally informed, rational approach to design such a variant would be employed but MAPKKK$_\varepsilon$'s recalcitrance to soluble expression in *E. coli* necessitated the generation of a mutant library and subsequent screening in the model solanaceous crop *N. benthamiana*, in order to evaluate the ability of variants to evade PexRD2 mediated cell death suppression.

The initial screen yielded a positive hit rate of 4%, which fell to <1% of candidates exhibiting a highly reproducible cell death evasion phenotype upon re-testing of all positive candidates in lower throughput assays with less potential for error. It remains a possibility that PexRD2-insensitive MAPKKK$_\varepsilon$ variants are not actually PexRD2-insensitive but, rather, simply have enhanced activity relative to MAPKKK$_\varepsilon^{Wt}$. This could be tested through expression of MAPKKK$_\varepsilon$ $^{Wt}$ alongside MAPKKK$_\varepsilon$ variants without the presence of PexRD2 in order to assess their relative cell death eliciting activities. Subsequent Y2H experiments, however, lend weight to the idea that variants genuinely are PexRD2-insensitive.

Sequencing of the two most reproducible candidates revealed 2 polymorphisms at the amino acid level per variant. These were; Phe10Leu & Gln50Arg in one, and Ser200Thr & Asp241Asn in the other. Upon testing these polymorphisms individually, it was found that all four contributed to some extent to insensitivity to PexRD2, although Q50R and S200T seemed to produce the most robust, reproducible phenotype. This is a surprising observation. The probability of all four variants producing such a phenotype when they resulted from a population with no selective pressure for generation of effector-insensitive variants is extremely remote. Given that the positive and negative controls behaved exactly as predicted in these assays, it is likely that any potential issue does not lie in the experimental procedure itself. Furthermore, the PexRD2$^{Wt}$ construct which completely suppresses cell death associated with MAPKKK$^{Wt}$ came from the same bacterial culture as the PexRD2$^{Wt}$ which fails to suppress cell death associated with the four PexRD2-insensitive variants removing, making it unlikely that the cell-death suppressor failed during these assays. The idea that putative PexRD2-insensitive kinase variants possess enhanced activity relative to the wild type discussed in the previous chapter is a possibility but encounters the same problems as the observation that all four SNPs

appear to contribute to insensitivity: improbability. The most likely alternative lies in the vectors in which the PexRD2-insensitive MAPKKK$_\varepsilon$ variants were delivered. MAPKKK$_\varepsilon^{Wt}$ was submitted to Genscript in the bespoke pERCH vector and returned following mutagenesis. Upon return, the vector backbone itself was not checked by sequencing. Therefore, there is a possibility that some mutation has occurred within the vector's promoter region enhancing expression and, so, observable cell death symptoms. This seems the most likely explanation if all four variants are not genuinely PexRD2-insensitive.

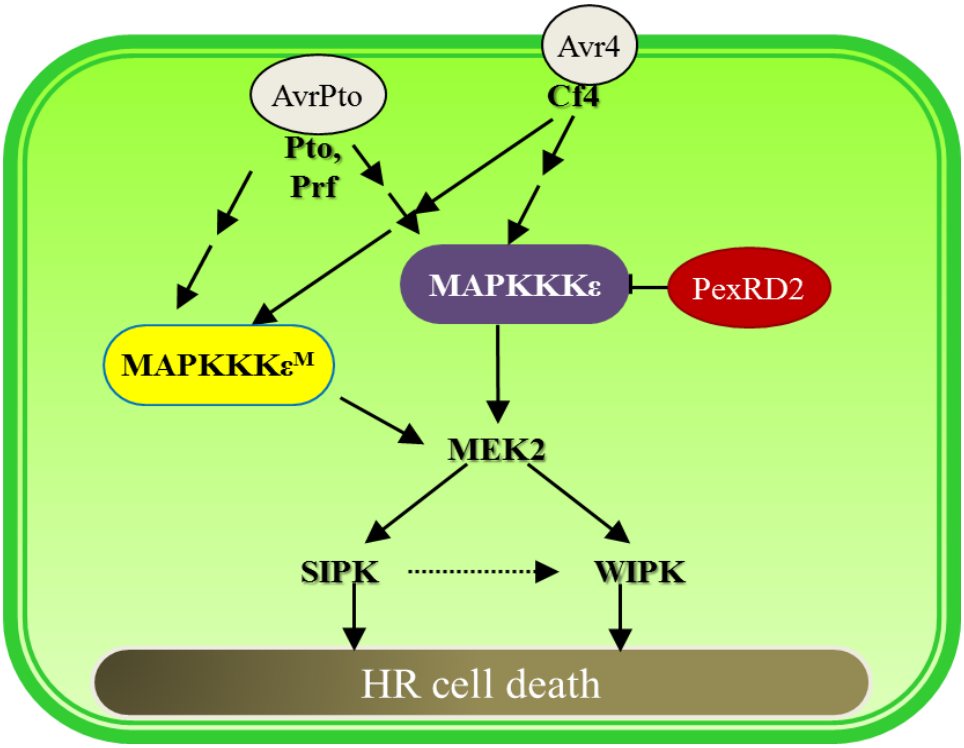



**Figure 21| Explanation for the basis of hypothetical PexRD2-insensitive MAPKKK$_\varepsilon$ variant dominant-resistant phenotype.**

MAPKKK$_\varepsilon^{Wt}$ mediated cell death response is triggered in response to perception of Avr4 by Cf4, AvrPto by Pto or Prf and – probably – by other host factors including an as yet unidentified invasion pattern from oomycetes. This is achieved by sequential phosphorylation of downstream kinases leading to transcriptional changes. The RXLR effector PexRD2 is secreted into the host cytoplasm, directly interacts with MAPKKK$_\varepsilon^{Wt}$ (purple) and suppresses this cell death response. MAPKKK$_\varepsilon$ variants which are refractory to PexRD2 (yellow) respond to the same signals in response to invasion and elicit a cell death response. PexRD2 is unable to interact with these variants or otherwise unable to interfere with their activation of downstream signalling components. Cell death, therefore, is initiated despite the presence of endogenous MAPKKK$_\varepsilon^{Wt}$. Adapted from (King et al., 2014).

Investigation of the nature of insensitivity of the variants of MAPKKK$_\varepsilon$ through qualitative protein-protein interaction studies indicated that insensitivity may arise from a weakened interaction with PexRD2, which we hypothesised would lead to a subsequent perturbation of its virulence function. In the less stringent, -LWH + 3-AT assays, we observed no differential between wild type and variants, however it is possible that investigating increased 3-AT concentrations would make a differential observable. However, in the more stringent –LWU assay a clear differential was observed indicating that the interaction between the wild type kinase and the wild type effector is stronger. This result was re-capitulated in X-GAL assays. Yeast 2-hybrid is, of course, a qualitative assay and not best-suited to measuring strength of interaction. In order to investigate this further, it may be necessary to purify both proteins and measure their binding affinities *in vitro* using a technique such as Isothermal Titration Calorimetry (ITC) or Surface Plasmon Resonance (SPR). However, the existence of a differential in activation of reporter genes in yeast does strongly suggest that PexRD2 has a reduced affinity for the PexRD2-insensitive MAPKKK$_\varepsilon$ variants developed in the present study. This reduced affinity is likely to be the basis for their insensitivity to PexRD2. Reduced affinity, as opposed to total loss of interaction, may prove useful in the long term. Total loss of interaction with PexRD2 may create selective pressure favouring promiscuity of the effector and the acquisition of new virulence targets through evolution. An insensitive kinase which retains a weakened interaction may sequester the effector and alleviate any potential selection pressure.

Efforts to generate CRISPR/Cas9 mediated knockouts of the endogenous MAPKKK$_\varepsilon$ in the $T_0$ generation were not successful, as screening revealed that primary transformants did contain large deletions but that these were not especially prevalent and that plants were likely to be chimeric. Screening of individual sgRNA sites for smaller indel mutations did reveal mutations which were present either homozygously, or heterozygously, in the tissue sampled. However, given the chimeric nature of the plants, based on PCR-RE analysis of gDNA, we decided it would be impossible to draw firm conclusions based on limited tissue sampling and that analysis of genome editing would be better assessed in the $T_1$ generation where CRISPR/Cas9 mediated editing and transgene expression would be easy to identify in lines where they are present in the germline. An alternative possibility, not

investigated, is that the aliquot of restriction enzyme used for digests had lost some activity and digests were incomplete. However, even with incomplete digestion, efficience CRISPR-mediated KO should still be detected at a higher level than observed in figure 16 if the modification was particularly prevalent. Furthermore, it remains a possibility that high levels of wild type band detection could be the result of contamination with wild type DNA, potentially of primer stocks. Primers used for the assay in Figure 14 should be tested for amplification where no tomato gDNA is present in order to rule out the presence of contaminating MAPKKK$_\varepsilon$ DNA. The lack of complete knockout of endogenous MAPKKK$_\varepsilon$ in whole transformants presents something of a limitation to the present work. However, we hypothesised that expression of an effector-insensitive kinase variant in the same cell as the endogenous protein may produce a dominant-resistant phenotype. This is based on the idea that effector-insensitivity has been characterised through cell death assays and that PexRD2 is thought to target MAPKKK$_\varepsilon$ in order to suppress its cell death induction activity in order to maintain living tissue for *P. infestans* to colonise. In a scenario where effector-sensitive kinase and effector-insensitive kinase are expressed in the same cell, cell death will be the dominant phenotype. This is outlined in Figure 21. This, of course, is predicated on the assumptions that endogenous MAPKKK$_\varepsilon$ is not expressed preferentially to effector-insensitive variants and that insensitive variants are stable in plant cells under native expression conditions and not rapidly degraded, leaving only a pool of effector-sensitive, endogenous MAPKKK$_\varepsilon$. If these assumptions proved to be correct then failure to test effector-insensitive kinases in plants where endogenous kinase expression had not been totally knocked out would be a severe limitation to the present work. The progeny of all primary transformants will be collected and germinated for analysis of stable inheritance of both the effector-insensitive kinases and CRISPR/Cas9 mediated editing events.

With this assumption in mind, we proceeded to test primary transformants expressing V1 and V2 (F10L and Q50R) effector insensitive kinase variants for resistance to *P. infestans*. In infection time courses over 6 days post infection, one variant conditioned enhanced resistance relative to the wild type control line when challenged with *P. infestans* strain 88069 (the strain in which PexRD2 was first identified). An almost twofold reduction in mean lesion size was observed 4dpi, relative to wild type. Further to this, infection lesions looked visibly different with

lesions on resistant leaves exhibiting a darker colouration, reminiscent of cell death as opposed to the spreading, grey lesions associated with successful *P. infestans* colonisation. This fits with our hypothesis that effector-insensitive kinases will deliver a dominant-resistant phenotype in spite of the presence of endogenous, effector-sensitive kinase. This is an encouraging result. A second round of pathogen assays with V3 and V4 (S200T and D241N) lines, however, showed enhanced susceptibility of tissue expressing both putative effector-insensitive variants relative to wild type. This is a confounding result which is somewhat difficult to explain. We cannot form a rational hypothesis as to why these variants would condition enhanced susceptibility and we suggest that there is likely to be an inherent level of variability in our assays related to expression levels of the transgene and the success of CRISPR/Cas9 mediated editing in primary transformants. There may be a $MAPKKK_{\varepsilon}$ dosage effect, where endogenous $MAPKKK_{\varepsilon}$ is expressed at native levels (having not been efficiently knocked out by CRISPR/Cas9 editing) and complemented with effector-insensitive variants. It is conceivable that enhanced levels of expression of a known positive regulator of immunity could produce the enhanced resistance phenotype observed in V1 plants, and also make V3 and V4 plants appear more susceptible. Another possible explanation could be that T-DNA has inserted into some key susceptibility gene in some lines, knocking it out and rendering certain lines more resistant. This is an obvious limitation on the present work and clearly collection of progeny from self-fertilised $T_0$ plants and subsequent screening for knockout and transgene expression (in particular, expression relative to endogenous $MAPKKK_{\varepsilon}$ levels) is required before firm conclusions can be made about the efficacy of any of these effector-insensitive variants in conditioning enhanced resistance – or susceptibility – to *P. infestans*.

88069 is a well-studied laboratory strain of *P. infestans* but not especially prevalent in the field. In recent years, a member of the A2 mating type of *P. infestans* (which became prevalent in the UK in 2006 at the expense of the A1 type), known as 13_A2 (blue-13) has become especially prevalent in the UK, Europe and as far afield as India (Cooke et al., 2012, Chowdappa et al., 2013). Blue-13 is a notoriously aggressive genotype of *P. infestans* and is known to be resistant to frontline fungicides such as metalaxyl-M and other phenylamide fungicides. Blue-13

is known to express PexRD2 (Cooke et al., 2012) so future efforts should include testing transgenic lines against blue-13.

Whilst we tested effector-insensitive variants against *P. infestans* strains known to express PexRD2 (Oh et al., 2009, Cooke et al., 2012), it would be important to identify a strain of *P. infestans* which does not express PexRD2 and challenge our transgenic plants with it in order to conclusively determine whether or not enhanced resistance is based on the presence/absence of PexRD2 in the invading pathogen. Enhanced resistance to strains not expressing PexRD2 may indicate an interaction between MAPKKK$_\varepsilon$ and a further effector protein.

It is known that MAPKKK$_\varepsilon$ is a positive regulator of immunity in at least two other pathosystems (Melech-Bonfil and Sessa, 2010b). It would, therefore, be intriguing to test a range of other pathogens for which tomato is a natural host for reduced virulence on plants expressing PexRD2-insensitive kinases in order to assess the ability of these variants to confer broad-spectrum resistance. As a positive regulator of immunity in numerous pathosystems, one may envisage that effector proteins from multiple pathogens converge on MAPKKK$_\varepsilon$ as a common virulence target in order to suppress its function. However, this would require both the secretion of an effector protein which directly targets MAPKKK$_\varepsilon$ and that this effector protein interacts with MAPKKK$_\varepsilon$ in a manner similar to that of PexRD2.

Another consideration for the present work is the propensity of *Phytophthora* spp. to overcome resistance rapidly. In theory, it may only take one mutation in the sequence of PexRD2 to re-establish suppression of MAPKKK$_\varepsilon$ mediated immunity. The likelihood of this could be investigated by establishing a similar random mutagenesis screen of PexRD2, and testing variants of the effector against effector-insensitive kinase variants in the cell death assays deployed in this chapter. Related to this idea, stacking of one or more effector-insensitive kinases – similar to stacking of NLRs - may provide some insurance against this possibility. Stacking multiple mutations on one sequence may result in interference with normal protein function. Multiple effector-insensitive targets should be far more durable in the field than one. We aim to investigate both of these ideas in the future.

# 4

# A solanaceous MAPKKKK may function as a ‘helper’ of the *P. infestans* virulence effector PexRD2

# Chapter 4 A solanaceous MAPKKKK may function as a helper of the *P. infestans* virulence effector PexRD2.

## 4.1. Introduction

Signal transduction from perception of an extracellular or intracellular stimulus is critical to elicit a cellular response. This role is generally carried out by a set of proteins termed Mitogen Activated Protein Kinases (MAPKs). These kinases are activated by an immune receptor, which often contains a kinase domain of its own or associates with a co-factor possessing a kinase domain upon invasion pattern perception, through phosphorylation. From there, further kinases are sequentially activated until a response is elicited, often in the form of transcriptional re-programming. Following signal perception by a cell surface or intracellular immune receptor, the first module in a given MAPK signalling cascade is a MAPKKKK, or MAP4K. MAPKKKKs act to phosphorylate MAPKKKs (MAP3Ks).

PM4K1 (henceforth referred to as MAPKKKK) is a MAP4K from *Solanum lycopersicum*. It was identified as an interactor of the *P. infestans* RXLR effector PexRD2 in a yeast 2-hybrid screen using PexRD2 as the bait in a prey library derived from cDNA generated from potato tissue infected with *P. infestans* (King, 2013). MAPKKKK shares significant levels of homology with the kinase domain of Sterile 20 protein-related protein kinases (Ste20p-related protein kinases). Ste20p is, itself, a protein kinase from yeast known to activate a signalling cascade by direct phosphorylation of Ste11p, a MAP3K. Currently there is no evidence to suggest that MAPKKKK phosphorylates the MAPKKK$_{\varepsilon}$ discussed in Chapter 3, despite evidence for a direct interaction (King, 2013). MAPKKKK may activate some other signalling cascade. The Ste20p-related protein kinases themselves can be further subdivided into two families termed the p21-activated kinase (PAK) and Germinal Centred Kinase (GCK) families which encode a catalytically active domain at their C- or N-terminus, respectively (Dan et al., 2001). It is clear from the literature that the

Ste20p-related kinases, PAKs and GCKs are involved in many different, critical cellular processes including cytoskeletal organisation and stress response (Brown et al., 1996, Chen et al., 2004), sterol homeostasis (Lin et al., 2009) as well as pheromone response in yeast and – potentially - other organisms (Leeuw et al., 1998, Leberer et al., 2000).

A homology search – using the amino acid sequence of MAPKKKK – showed that MAPKKKK contains a putative Serine/Threonine catalytic domain with homology to two members of the GCK sub-family VI; Oxidative Stress Response 1 (OSR1) and Ste20-related proline/alanine-rich kinase (SPAK) (King, 2013). OSR1 and SPAK share a number of conserved features in their regulatory regions including a putative Nuclear Localisation Signal (NLS), a Caspase cleavage site and putative, C-terminal protein-protein interaction domains (Piechotta et al., 2003, Delpire and Gagnon, 2008). By comparison, MAPKKKK shares no real conservation with the regulatory regions of OSR1 or SPAK, suggesting that the mechanisms of regulation and downstream targets of MAPKKKK are likely to be divergent from those of OSR1 and SPAK. In addition, MAPKKKK is predicted to encode a coiled-coil domain at its C-terminus which is, once more, in contrast with OSR1 and SPAK but similar to other Ste20-like kinases, such as SLK, a regulator of renal development in mammals (Delarosa et al., 2011). These coiled-coil domains have been hypothesised to mediate homodimerisation, enhancing their signalling activity. It has been demonstrated, through yeast 2-hybrid assays, that MAPKKKK can self-associate (King, 2013) although the implications of this are unclear as its *in planta* function is not currently known.

As discussed in section 1.4, MAP Kinases are of critical importance as regulators of a range of processes in plants and, as such, are a general target of effector proteins secreted by plant pathogens. Of particular importance here, the solanaceous $MAPKKK_{\varepsilon}$ was identified as a virulence target of the *P. infestans* effector protein PexRD2 (King et al., 2014). The importance of this interaction is discussed in greater detail in Chapter 3.

The *S. tuberosum* MAPKKKK identified in the initial Y2H screen has homologous proteins in *S. lycopersicum* and in the model host, *N. benthamiana* with sequence homology at the amino acid level of 99% and 92%, respectively. This

indicates that any findings based on the potato protein will likely be broadly applicable to other solanaceous crops and that *N. benthamiana* will serve as a suitable model system to study the interaction between MAPKKKK and PexRD2. We hypothesised that MAPKKKK may, like MAPKKK$_{\varepsilon}$, be a *bona fide* virulence target of the *P. infestans* RXLR effector PexRD2 and set out to explore this hypothesis through pathogenicity assays in the model, solanaceous host *Nicotiana benthamiana*. During the course of the research presented in the present chapter, it became apparent that MAPKKKK may not be a virulence target of PexRD2 in the traditional sense of the term but may serve to modify PexRD2 *in planta* and enhance its virulence function. The helper concept is discussed more thoroughly in section 4.2.3. This idea was explored using co-expression of the kinase and the effector, once more in *N. benthamiana* as well as using mass spectrometry on proteins expressed *in planta* to test for post-translational modifications, including phosphorylation.

## 4.2. Results

### 4.2.1. The kinase domain of MAPKKKK and the effector domain of PexRD2 are necessary and sufficient for interaction

The C-terminal effector domain of PexRD2 is necessary and sufficient for its biochemical function *in planta*, which is to suppress cell death signalling associated with MAPKKK$_\varepsilon$ (King, 2013, King et al., 2014). Further, the N-terminal kinase domain of MAPKKK$_\varepsilon$ was shown to be necessary and sufficient for interaction with PexRD2 and the two ARMADILLO repeat domains spanning residues in the C-terminal portion of the protein were not required, despite their previously reported role in mediating protein-protein interactions (Huber et al., 1997).

Based on this it was hypothesised that the PexRD2/MAPKKKK interaction would follow a similar pattern with the N-terminal kinase domain - spanning residues 16 to 277 – being necessary and sufficient for interaction with PexRD2. To test this, a yeast 2-hybrid approach was used testing three truncations of MAPKKKK for interaction with PexRD2 lacking the N-terminal signal peptide. Also tested for MAPKKKK interaction were two additional PexRD2 truncations, one lacking the residues to the N-terminus of the RXLR motif and one lacking the residues to the start of WY-domain. These constructs had previously been prepared by Stuart King (JIC). Constructs were transferred into yeast 2-hybrid expression vectors pDEST22™ and pDEST32™ by Gateway Cloning. These truncated constructs are outlined in Figure 22 and yeast 2-hybrid reporter assays are shown in Figure 23. Two orthologues of PexRD2 – PITG_14787 and PITG_14984 – known not to interact with full length MAPKKKK in yeast 2-hybrid assays were also transferred into pDEST32™ to act as negative controls.

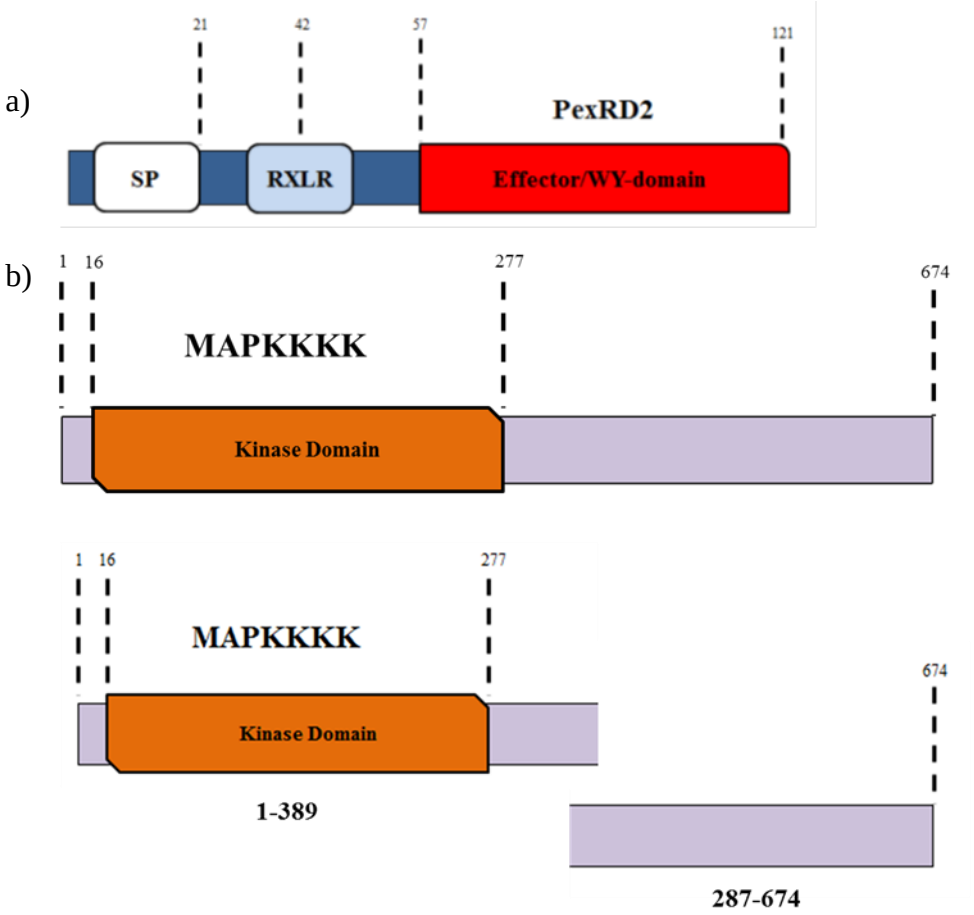



**Figure 22| Truncated constructs used in section 4.2.1.**

a) Schematic representation of PexRD2 starting from residue 21. Residues 21-121 correspond to 'full length' construct lacking its signal peptide, PexRD2$^{42\text{-}121}$ lacks the signal peptide and the RXLR motif whilst PexRD2$^{57\text{-}121}$ lacks the RXLR-dEER motif and is comprised of the WY-domain only. Constructs hereafter referred to as PexRD2$^{\Delta 20}$, PexRD2$^{\Delta 41}$ and PexRD2$^{\Delta 56}$, respectively. b) Shows a schematic representation of the 'full length' MAPKKKK construct in addition to the two truncated constructs used. The first comprises MAPKKKK$^{1\text{-}389}$ containing the kinase domain but excluding most of the C-terminus. The second comprises MAPKKKK$^{287\text{-}674}$ excluding the kinase domain but including most of the C-terminal region of MAPKKKK.

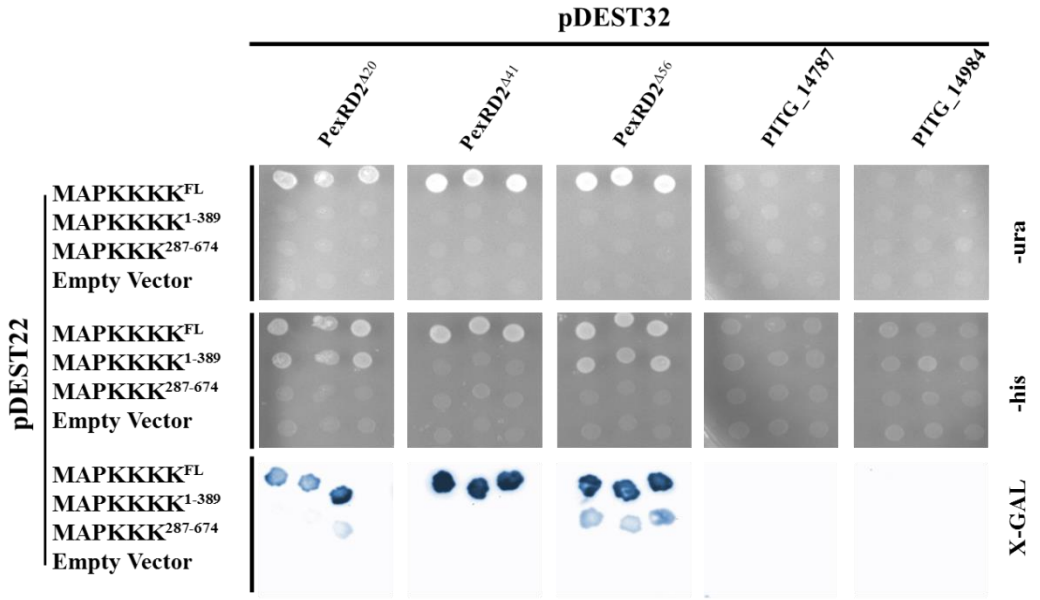



**Figure 23| Yeast 2-hybrid reporter assays for MAPKKKK truncations and PexRD2 truncations.**

Rows of 3 spots of yeast growth within individual panels derive from 3 independent colonies per bait and prey combination selected following co-transformation of bait and prey constructs. Images were recorded 48 h after transformants were spotted onto reporter assay plates. Panels, therefore, all contain the same bait construct and each row represents that bait construct tested against a different prey construct.

Figure 23 shows results of yeast 2-hybrid reporter assays. Full length PexRD2, as well as PexRD2$^{\Delta 41}$ and PexRD2$^{\Delta 56}$ interacted with full length MAPKKKK in all three reporter assays. PexRD2$^{\Delta 20}$ and PexRD2$^{\Delta 56}$ also interacted with MAPKKKK$^{1-389}$ in -*HISTIDINE* assays and, weakly, in X-Gal assays. No interaction was observed in –*URACIL* assays. This may indicate that, whilst not critical for a physical interaction, MAPKKKK's C-terminal residues may interact with PexRD2 or otherwise exert some effect on the conformation of the rest of MAPKKKK which stabilises the interaction and that removing these residues destabilises its complex with PexRD2. Interestingly, PexRD2$^{\Delta 41}$ did not interact with MAPKKKK$^{1-389}$ in the three reporter assays. This construct is only 21 amino acids shorter than the PexRD2$^{\Delta 20}$ construct which still interacted and does not lack any known domain

thought to be critical for interaction. Given that both the 'full length' and 'effector domain only' constructs still interacted with MAPKKKK's kinase domain it seems likely that this observation is a false negative or an artefact of an expressed protein construct which is not predicted to exist in the natural potato/tomato-*P. infestans* pathosystem. The C-terminal region of MAPKKKK did not interact with any PexRD2 construct in any reporter assay and neither of the PexRD2 orthologues interacted with any MAPKKKK construct in any reporter assay indicating that the kinase domain of MAPKKKK can interact with PexRD2's effector domain – although the interaction between PexRD2 and full length MAPKKKK is stronger - and that this interaction does not extend to close relatives of PexRD2.

This confirmed interaction between PexRD2 and MAPKKKK suggests that MAPKKKK may represent a genuine virulence target of PexRD2, alongside MAPKKK$_{\varepsilon}$. In addition, previous work suggested the PexRD2/MAPKKKK interaction could be observed by *in planta* co-IP and preliminary evidence suggested an interaction may be observed when the proteins are purified after co-expression in *E. coli* (work by Dr Abbas Maqbool, JIC). This was determined by co-expression of hexa-histidine tagged PexRD2 and MAPKKKK in *E. coli* and comparison of size exclusion chromatography (SEC) peaks to both MAPKKKK and PexRD2 expressed alone. A clear peak shift was observed when MAPKKKK and PexRD2 were co-expressed with no peak observed for either of the two proteins on their own (Figure 24). Analysis of fractions collected under this shifted peak by SDS-PAGE indicated the presence of both proteins and provided clear evidence of an interaction *in vitro*.

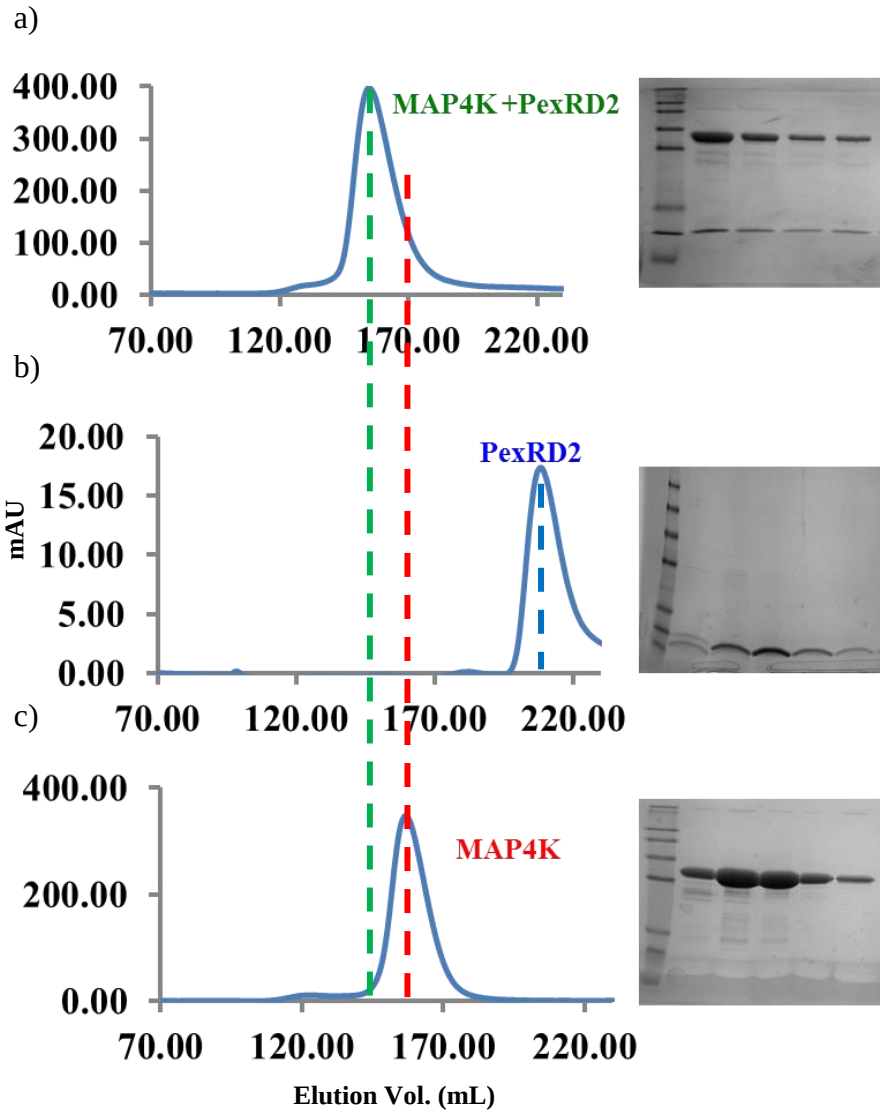



**Figure 24| PexRD2 co-elutes with MAPKKKK following SEC.**

a) SEC peak for MAPKKKK and PexRD2 co-expressed alongside SDS-PAGE gel of fractions under the peak. Gel indicates co-elution of kinase and effector. Upper bands represent MAPKKKK whilst lower bands represent PexRD2 Peak indicates earlier elution volume relative to b) and c) (PexRD2 and MAPKKKK expressed alone, respectively) indicating the formation of a complex between the two proteins. SDS-PAGE gels of fractions under these peaks indicate the presence of PexRD2 or MAPKKKK alone. Dashed lines indicate the elution volumes of each peak.

### 4.2.2. Transient overexpression of MAPKKKK *in planta* does not affect virulence of *P. infestans* on *N. benthamiana* leaves

In order to test the hypothesis that MAPKKKK is a virulence target of PexRD2, MAPKKKK was overexpressed transiently from the pB7WGF(3xHA)2 vector in one half of a ~4 week old *N. benthamiana* leaf, alongside an empty vector control in the other half of the same leaf as a direct comparison. Leaves were allowed to recover from agroinfiltation for around 48 h before harvesting. Leaves were detached and placed in trays on top of paper towel saturated with water to create an environment of 100% relative humidity. Leaves were then challenged with the *P. infestans* 88069 prepared and applied to leaves as described in section 2.9.7.

Infection was allowed to proceed for 7 days and images of leaves and developing lesions were recorded daily from appearance of visible lesions (4 days).

Over the time course from day 5 to day 8 post infection, mean lesion size on tissue overexpressing MAPKKKK did not differ significantly (student's t-test ($n$ = 20)) from mean lesion size on tissue not overexpressing MAPKKKK at any given time point, suggesting that MAPKKKK is not deleterious to *P. infestans* 88069 virulence on this variety of *N. benthamiana* (Figure 25).

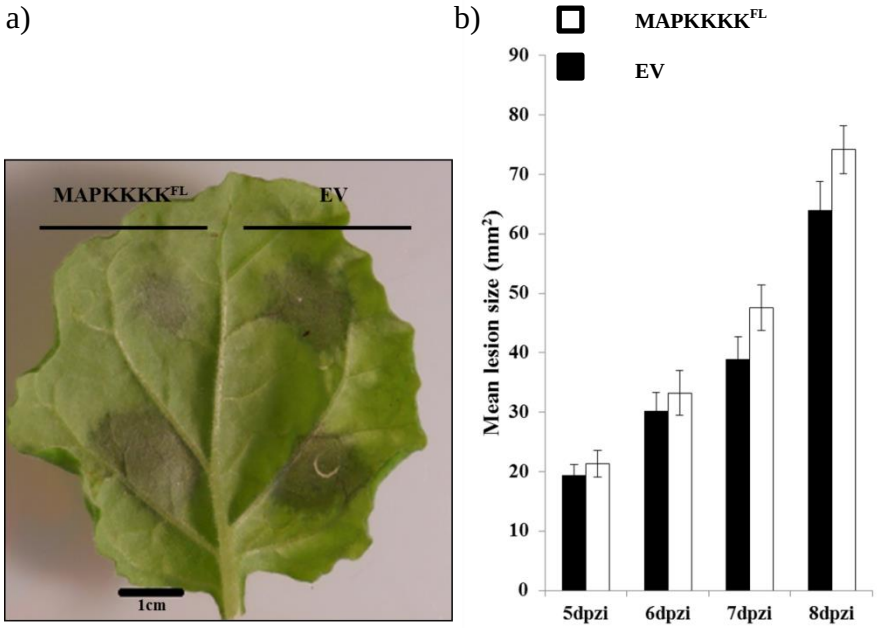



**Figure 25| Overexpression of MAPKKKK$^{FL}$ does not affect *P. infestans* virulence.**

a) Detached *N. benthamiana* leaf infected with *P. infestans*. Left hand side transiently infiltrated with pB7WGF(3xHA)2.MAPKKKK$^{FL}$ and right hand side transiently infiltrated with empty pB7WGF(3xHA)2. Characteristic spreading, grey *P. infestans* lesions visible on lower portion of leaf. b) Bar chart showing mean lesion area in mm$^2$ over a time course of 5-8 days post zoospore inoculation. Error bars represent standard error ($n$ = 20). White bars represent tissue expressing MAPKKKK$^{FL}$.

Although the interaction between MAPKKKK$^{N-319}$ and PexRD2 was weaker in yeast, we investigated the possibility that the kinase domain of MAPKKKK, lacking its predicted, disordered regulatory C-terminal region may enhance virulence of *P. infestans 88069* on *N. benthamiana* leaves. *N. benthamiana* leaves were infiltrated with the constitutively expressed pB7WGF(3xHA)2.MAPKKKK$^{N-319}$ construct. Leaves were detached ~48 h post infiltration and placed in trays on top of paper towels saturated with water, at 100% humidity. Leaves were infected with *P.*

*infestans* 88069 as described previously and infection was allowed to proceed for 7 days, with images recorded daily.

Over days 5-8 post zoospore inoculation, no significant difference (student's t-test ($n$ = 20)) was observed between lesions on leaf tissue expressing MAPKKKK$^{N-319}$ and lesions on leaf tissue infiltrated with empty vector control on any given day. This suggests that the kinase domain of MAPKKKK is not sufficient to impede the pathogenicity of *P. infestans* 88069 on this variety of *N. benthamiana* and questions the hypothesis that MAPKKKK is a *bona fide* virulence target of PexRD2. These results are shown in Figure 26.

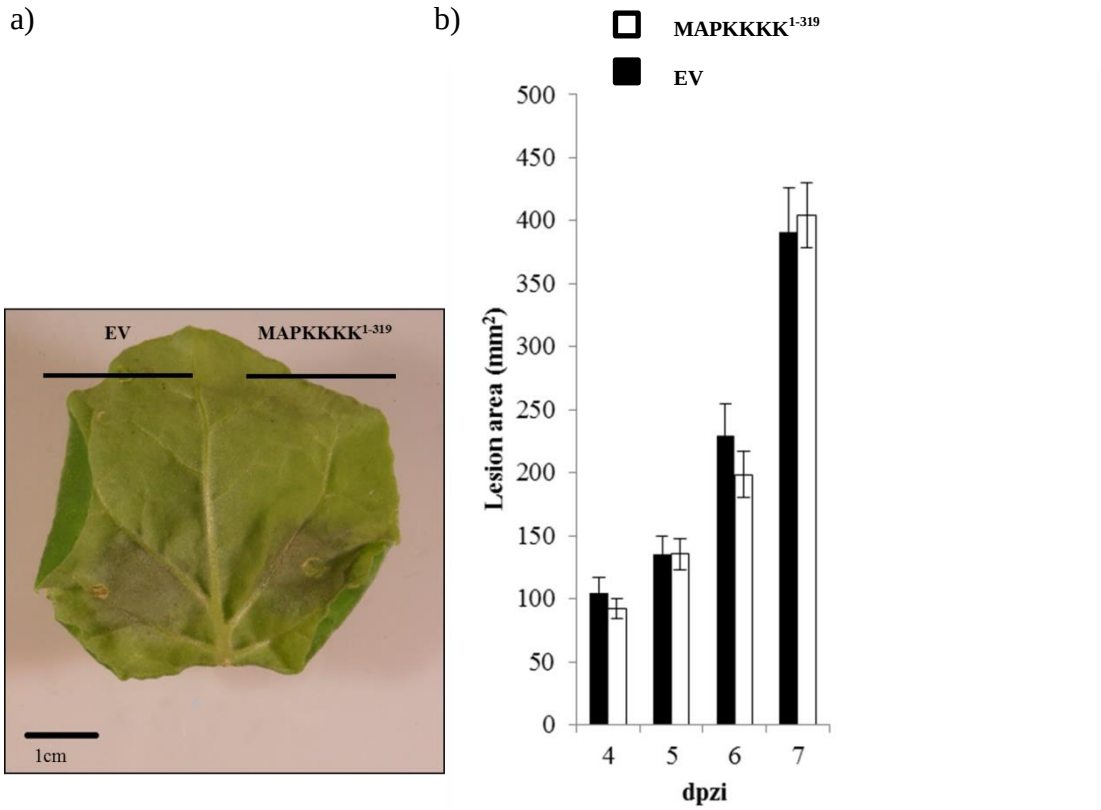



**Figure 26| Overexpression of MAPKKKK$^{N-319}$ does not affect *P. infestans* virulence.**

a) Detached *N. benthamiana* leaf infected with *P. infestans*. Right hand side transiently infiltrated with pB7WGF(3xHA)2.MAPKKKK$^{N-319}$ and left hand side transiently infiltrated with empty pB7WGF(3xHA)2. Characteristic spreading, grey *P. infestans* lesions visible on lower portion of leaf. b) Bar chart showing mean lesion area in mm$^2$ over a time course of 4-7 days post zoospore inoculation. Error bars represent standard error ($n$ = 20). White bars represent tissue expressing MAPKKKK$^{1-319}$.

### 4.2.3. Transient overexpression of MAPKKKK enhances *P. infestans* virulence on *N. benthamiana* leaves when co-expressed with PexRD2

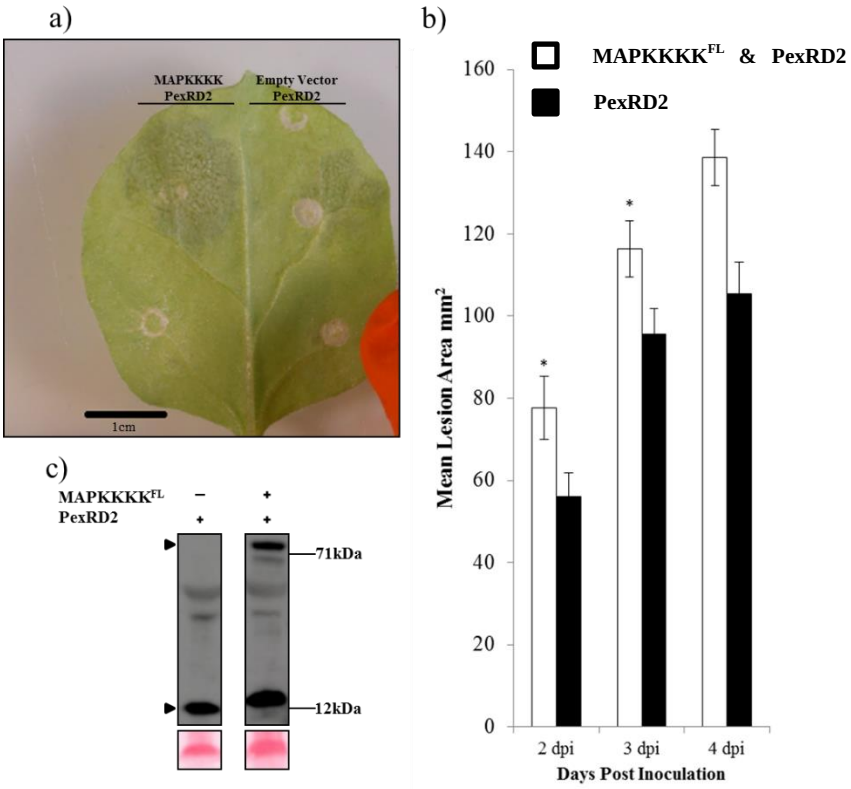



**Figure 27| Co-expression of MAPKKKK$^{FL}$ with PexRD2 appears to enhance its virulence function in planta.**

a) Detached *N. benthamiana* leaf infected with *P. infestans* 88069. Left hand side co-expressing MAPKKKK$^{FL}$ and PexRD2, right hand side expressing PexRD2 alone. White bars represent kinase & effector. Black bars represent effector only b) Bar chart showing mean lesion area in mm$^2$ over a time course of 2-4 days post zoospore inoculation. Error bars represent standard error. Significant differences at the p = <0.05 were determined by student's t-test. *n* = 40. c) Immunoblot demonstrating protein expression. Membranes were incubated with primary anti-HA antibody and a secondary antibody conjugated to HRP. Left hand lane shows HA-PexRD2 expressed alone. Right hand lane shows HA-MAPKKKK$^{FL}$ and HA-PexRD2 co-expressed. Arrowheads show expected size of each protein. Upper band is MAPKKKK$^{FL}$ and lower band is PexRD2. Ponceau S stain of the large RuBisCo sub-unit to demonstrate protein loading.

Although MAPKKKK expression alone in *N. benthamiana* did not affect growth of *P. infestans* 88069, it remains possible that the interaction between MAPKKKK and PexRD2 is relevant during infection. In order to investigate the hypothesis that MAPKKKK somehow activates PexRD2 and enhances its ability to promote *P. infestans* virulence *in planta*, PexRD2$^{\Delta 21}$ was expressed transiently in *N. benthamiana* leaves under the control of a constitutively active promoter in the pB7WGF(3xHA)2 vector. In one half of a given leaf, PexRD2 was co-expressed with full length MAPKKKK in the pB7WGF(3xHA)2 vector and in the other half of the same leaf, PexRD2 was expressed alone. Leaves were detached 48 h post infiltration and placed in trays on top of water saturated paper towel in 100% humidity. Leaves were then challenged with *P. infestans* 88069 prepared and applied as described in section 2.10.7 and infection was monitored, recorded and analysed as described previously with the exception that disease lesions appeared earlier (presumably due to overexpression of the virulence effector PexRD2) so lesions were monitored from 2-4 days post zoospore incoculation. The results of this assay are shown in Figure 27.

We also tested the kinase domain of MAPKKKK in order to see if it was sufficient to provide the same enhancement of PexRD2 virulence function observed when full length MAPKKKK was co-expressed with PexRD2. MAPKKKK$^{N\text{-}319}$ was co-expressed alongside PexRD2 in one half of a *N. benthamiana* leaf, whilst PexRD2 was expressed on its own in the other half. Leaves were infiltrated, detached and inoculated with *P. infestans* 88069 as described previously, and symptoms were monitored from 3-5 days post zoospore inoculation. The results of this assay are presented in Figure 28.

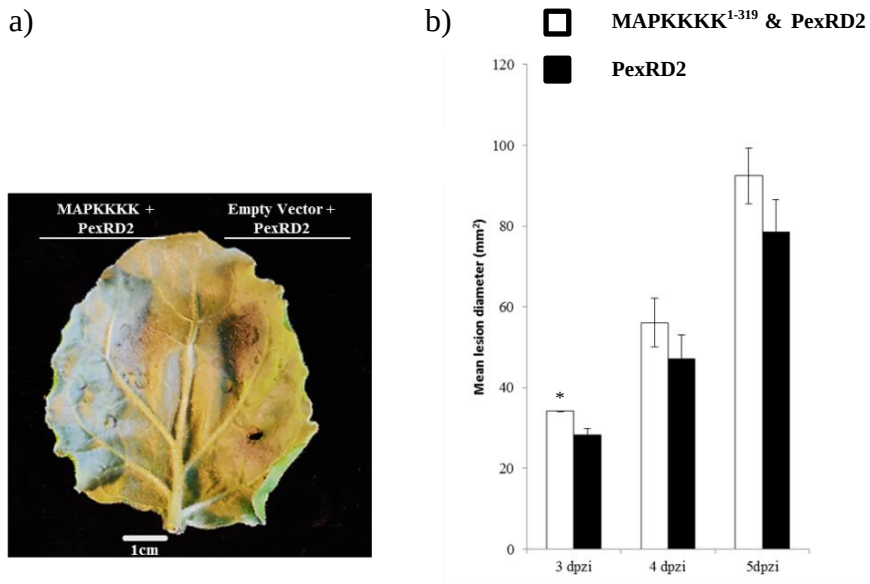



**Figure 28| Co-expression of MAPKKKK$^{N-319}$ with PexRD2 may enhance PexRD2's virulence function.**

a) Detached N. benthamiana leaf infected with P. infestans 88069. Left hand side co-expressing MAPKKKK$^{N-319}$ and PexRD2, right hand side expressing PexRD2 alone. b) Bar chart showing mean lesion area in mm$^2$ over a time course of 3-5 days post zoospore inoculation. White bars represent tissue expressing effector and kinase, black bars represent effector only. Error bars represent standard error. Significant difference at p = <0.05 differences determined by student's t-test. *n* = 24 inoculation sites.

It was observed that the kinase domain of MAPKKKK does enhance virulence function, as measured by mean lesion size, significantly 3 days post zoospore inoculation (student's t-test) although this enhancement of virulence function does not continue 4 and 5 days post inoculation. However, whilst not statistically significant, it should still be noted that mean lesion size for these two days is still measurably higher in tissue where the kinase domain of MAPKKKK and PexRD2 are co-expressed compared to where PexRD2 is overexpressed alone.

### 4.2.4. MAPKKKK mediated phosphorylation of PexRD2 cannot be detected *in planta*

We hypothesised that the enhanced virulence activity associated with co-expression of MAPKKKK and PexRD2 may be due to MAPKKKK mediated phosphorylation of PexRD2 due to the clear evidence for a direct interaction between kinase and effector.

The amino acid sequence of PexRD2$^{\Delta 21}$ was submitted to the NetPhosk 2.0 server which predicts probability of phosphorylation of serine, threonine and tyrosine residues based on amino acid sequence (Blom et al., 2004). The output predicted the potential phosphorylation of 6 residues at a probability of >0.5. This included serine residues at positions 22, 53 and 92, Threonine residues at positions 60 and 72 and a tyrosine residue at position 76 (Figure 29). As MAPKKKK is a predicted serine/threonine protein kinase, it is unlikely to act on Tyr76.

To investigate whether any of these sites were phosphorylated by MAPKKKK, pTRBO.PexRD2$^{21\text{-}121}$ was overexpressed transiently in *N. benthamiana* in the presence of pB7WGF(3xHA)2.MAPKKKK$^{1\text{-}319}$, and in the absence of MAPKKKK as a direct comparison. The CaMV 35S driven pTRBO vector, in addition to expressing the TBSV p19 silencing suppressor, encodes an N-terminal FLAG tag based on the expression vector constructed in (Lindbo, 2007). Following extraction of crude cell lysate from *N. benthamiana* leaves, lysate was incubated with anti-FLAG resin conjugated to magnetic beads in order to enrich for PexRD2 by immune-precipitation. Samples were analysed by SDS-PAGE, western blotting and immuno-detection of HA-tagged kinase and FLAG-tagged effector (Figure 30). PexRD2 appears to be moderately enriched following pulldown compared with crude lysate whilst MAPKKKK$^{1\text{-}319}$ appears to be significantly enriched compared with crude lysate. Some MAPKKKK$^{1\text{-}319}$ was detected following pulldown with anti-FLAG resin from lysate containing no FLAG-tagged PexRD2 indicating that there may be some cross-reactivity or that washing was insufficient during pulldown. As this was not a test for protein-protein interactions, this was not considered a major cause for concern.

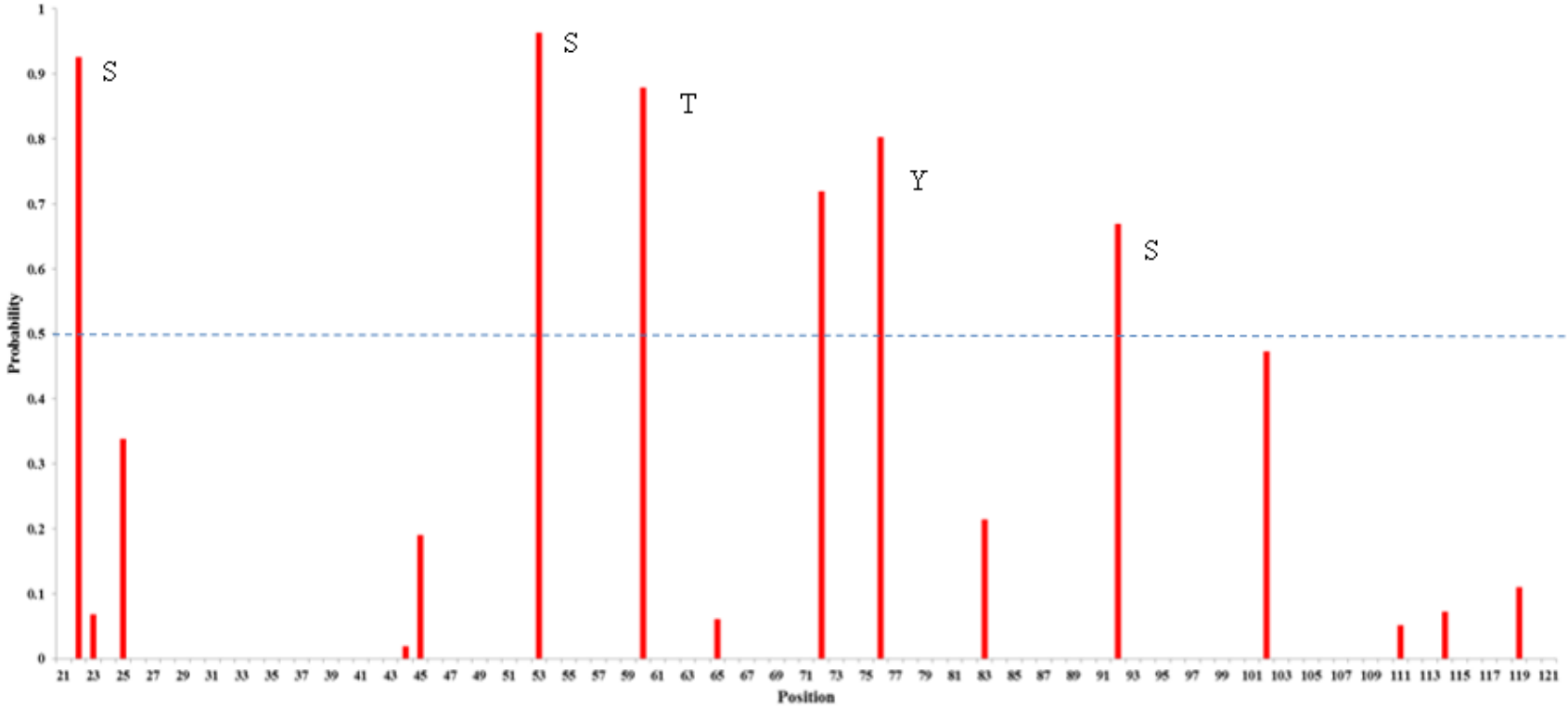



**Figure 29| Several amino acids in PexRD2Δ21 are predicted to be phosphorylated.**

NetPhosK 2.0 output presented as probability of phosphorylation against residues 21-121 of PexRD2. Blue dashed line indicates probability cut-off for residues likely to be phosphorylated according to the NetPhosK 2.0 server. Labels to the left of bars representing probability values of >0.5 indicate the identity of the amino acid predicted to be phosphorylated with high confidence.

Following confirmation of expression by immuno-detection shown in Figure 30, protein bands in SDS-PAGE gels corresponding to FLAG-tagged PexRD2 stained with InstantBlue™ were excised from the gel with a sterile scalpel and submitted for mass spectrometry analysis. Notably, when stained with instant blue, enrichment of FLAG-tagged PexRD2 became more apparent.

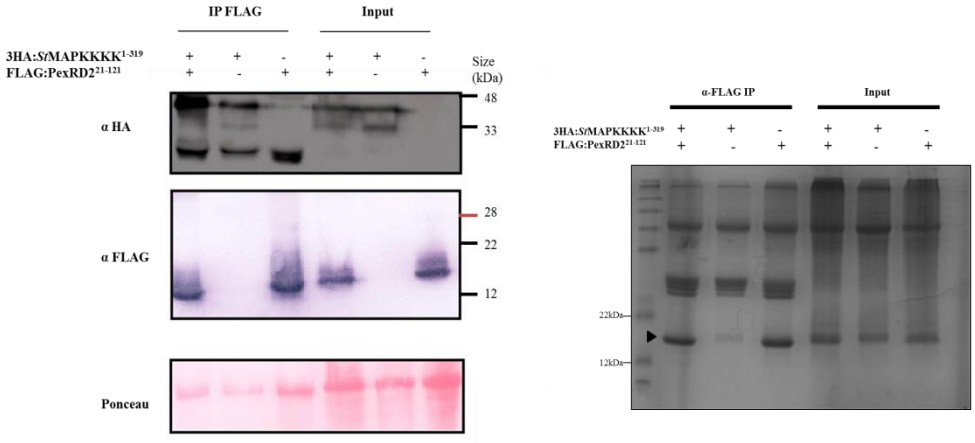



**Figure 30| PexRD2 co-expressed with MAPKKKK$^{N-319}$ and enriched by immunoprecipitation.**

(Left) Western blot showing; horseradish peroxidase assay for detection of HA tagged protein in pulldowns (IP Flag) or in crude lysate (Input), alkaline phosphatase assay for detection of FLAG tagged protein in pulldowns or crude lysate and Ponceau red staining of the large RuBisCo subunit to confirm protein loading (Note – total protein loading expected to be lower in pulldown relative to input). (Right) InstantBlue stained 17% SDS-PAGE gel ran simultaneously with western blot to the left with aliquots of the same samples loaded, demonstrating enrichment of FLAG-tagged PexRD2 in pulldown lanes relative to crude extract. Bands for PexRD2 indicated by black arrow. These bands were excised and submitted for LC-MS/MS analysis. The bands of similar size to PexRD2 in input lanes do not correspond to PexRD2.

*(Note – mass spectrometry was performed by Dr Gerhard Saalbach of the JIC proteomics platform).*

Excised gel slices containing FLAG-tagged PexRD2 were washed and digested in gel with trypsin prior to extraction according to procedures adapted from (Shevchenko et al., 2007). Detected peptides for both samples containing either PexRD2 alone (sample 1) of PexRD2 co-expressed with the MAPKKKK (sample 2)

kinase domain are presented in Table 10 alongside their observed masses, their unmodified predicted masses and their predicted masses based on the phosphorylation prediction presented in Figure 29.

In both samples, coverage was not perfect. In sample 1, 68/101 amino acids were detected (67%). Included in the missing amino acids was T60, which was predicted to be phosphorylated. However, given that this sample was serving as a negative comparison for MAPKKKK's activity, this is not necessarily an issue as we predicted that no residues here should be phosphorylated. Furthermore, given that this residue is covered in sample 2, where we would predict it may be phosphorylated yet no phosphorylation can be detected, it is a reasonable assumption that this residue is also not phosphorylated in sample 1. In sample 2, 70/101 amino acids were detected (69%) which was roughly equivalent to sample 1. Again, only one of the residues predicted to be phosphorylated based on Figure 29 was not detected: S22.

| **PexRD2 alone** | | | |
|---|---|---|---|
| Peptide Seq. | Predicted mass (No PTM) | Observed mass | Δ |
| LSTNTGVQAANLVGPAQR | 1797.0 Da | 1795.95 Da | -1.05 Da |
| KHYTAAENDDDSEAR | 1721.7 Da | 1720.73 Da | -0.97 Da |
| AGMTVDDYAAK | 1141.2 Da | 1140.51 Da | -0.69 Da |
| LTDKIAAAANSAR | 1301.4 da | 1300.71 Da | -0.69 Da |
| YLNYVAEHTAV | 1279.4 Da | 1278.62 Da | -0.78 Da |

| **MAPKKKK + PexRD2** | | | |
|---|---|---|---|
| Peptide Seq. | Predicted mass (No PTM) | Observed mass | Δ |
| KHYTAAENDDDSEAR | 1721.7 Da | 1720.73 Da | -0.97 Da |
| ALNTEKMK | 934.1 Da | 949.49 Da | +15.39 Da |
| AGMTVDDYAAKLK | 1382.5 Da | 1397.69 Da | +15.19 Da |
| LTDKIAAAANSAR | 1301.4 Da | 1300.71 Da | -0.69 Da |
| AMEKLGETLK | 1119.3 Da | 1118.6 Da | -0.7 Da |
| YLNYVAEHTAV | 1279.4 Da | 1278.62 Da | -0.78 Da |

**Table 10| Predicted and observed masses of PexRD2 derived peptides expressed alone in *N. benthamiana* tissue or co-expressed with MAPKKKK.**

First column shows peptide amino acid sequence, second column shows predicted molecular mass of peptide as determined by the ProtParam server, third column shows observed peptide mass and fourth column shows mass shift between predicted and observed peptide masses. Residues bioinformatically predicted to be phosphorylated in Figure 29 are highlighted red.

The observed masses of peptides detected in Table 10 correlate well with the masses predicted for each peptide by the ProtParam server indicating that PexRD2, overexpressed alone in *N. benthamiana* leaves, undergoes no post-translational modifications. In sample 2, mass changes in two of the peptides are observed. ALNTEKMK has a Δ of +15.39Da and AGMTVDDYAAKLK has a Δ of +15.19Da. These mass shifts likely correspond to oxidation of the sulphur moieties on methionine residues forming methionine sulphoxide. This modification is a frequent artefact observed in mass spectrometry experiments and not the result of any MAPKKKK activity on PexRD2. What is clear from the data is that MAPKKKKN-319 expression does not phosphorylate any peptide detected in the present assay. Phosphorylation would correspond to a Δ in the range of +80 Da, which was not detected for any peptide. However, it is impossible to formally rule out the possibility that PexRD2 is phosphorylated by MAPKKKK$^{N\text{-}319}$ as incomplete coverage has resulted in the peptide containing Ser22, which is predicted to be phosphorylated with a probability of ~0.9 (Figure 29), not being detected.

## 4.3. Discussion

MAP Kinases are firmly established as common virulence targets of virulence effectors secreted by a variety of plant-associated microbes (Zhang et al., 2007, Cui et al., 2010, Wang et al., 2010, King et al., 2014). This is unsurprising given their integral role in the transduction and amplification of signals triggered by a wide range of invasion patterns, including classically defined MAMPs perceived by PRRs and avirulence effectors perceived by NLRs.

We investigated the possibility that MAPKKKK is a virulence target of PexRD2 through a series of *in planta* pathogen growth assays, using *Agrobacterium* mediated overexpression of MAPKKKK in *N. benthamiana* (a model host of *P. infestans*) to test this hypothesis. If MAPKKKK is a *bona fide* virulence target of PexRD2, we would hypothesise that its overexpression would be detrimental to pathogen fitness, presenting as smaller disease lesions on leaf tissue overexpressing MAPKKKK relative to those infiltrated with a vector control. This assay was performed using a full length MAPKKKK construct as well as a construct expressing only the minimal

interacting domain of MAPKKKK - residues 1-319 – which encode its catalytically active kinase domain (Figure 25 and Figure 26, respectively). In both assays, mean lesion size was roughly the same on leaf tissue expressing a kinase construct as it was on leaf tissue transformed with an empty vector control. This strongly suggests that MAPKKKK may not be a virulence target of PexRD2. We cannot formally exclude the possibility that MAPKKKK is a susceptibility factor, however. Overexpression does not promote enhanced virulence but native expression may be sufficient to observe enhanced susceptibility. VIGS constructs targeted against endogenous MAPKKKK mRNA could reveal whether or not MAPKKKK is a susceptibility factor.

One hypothesis may be that the interaction is vestigial, and at one time MAPKKKK mediated signalling in a pathway which promoted defence against *P. infestans* and – whilst the pathway no longer functions – the interaction remains. Given the continuous arms race between host and pathogen, it is easy to envisage a scenario in which a particular signalling pathway was lost as it no longer conferred any benefit due to pathogen targeting and maintaining it was too costly. Another hypothesis is the idea that MAPKKKK may function as a 'helper' of PexRD2. The process of effector entry into a host cell and subsequent modification by a host derived factor is a well-understood idea in plant pathology known as the 'helper concept' (Win J, 2012). A helper protein can be defined as a host protein which; post-translationally modifies an effector protein, is required for maturation of an effector protein, acts as a co-factor for effector function or is required for trafficking of an effector protein within the host cell. A basic model of the helper concept, using MAPKKKK and PexRD2 as examples, is presented in Figure 31. Classical examples of helper proteins include myristoyl transferases which myristoylate the TIIISS secreted effector AvrPto from *Pseudomonas syringae*. AvrPto is recognised by the intracellular immune receptor Pto and this is dependent on myristoylation mediated localisation to the host plasma membrane (Shan et al., 2000). Another well studied example of a helper protein is the *Glycine max* (Soybean) cyclophilin GmCYP1. GmCYP1 possesses proly-pedtidyl isomerase (PPIase) activity which is required for the proper formation of intra-molecular prolyl bonds. GmCYP1 physically interacts with a GP peptide of the Avr3b effector from *P. sojae* and this PPIase activity is required for proper maturation of Avr3b. In the absence of GmCYP1 – or

homologous cyclophilins from *N. benthamiana* extract, Avr3b loses its virulence function (as a nudix hydrolase) and also can no longer be recognised by the NLR Rps3b (Kong et al., 2015).

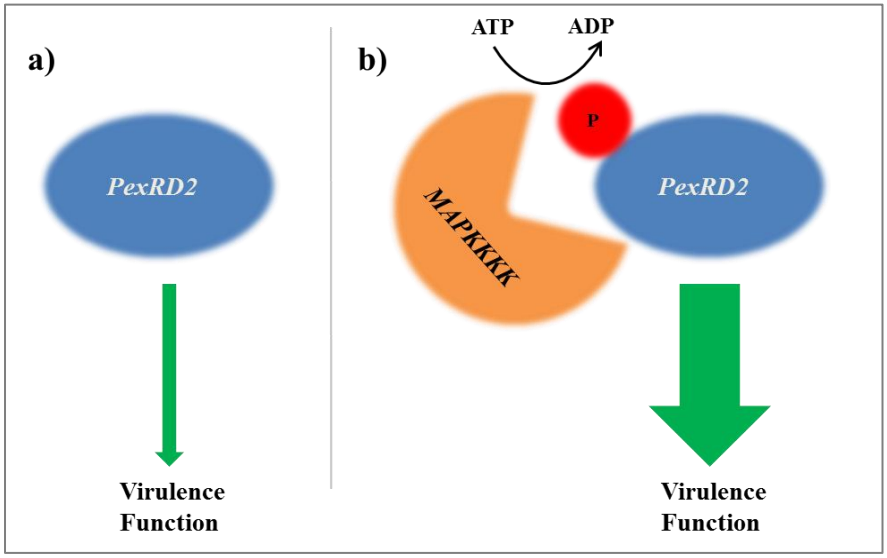



**Figure 31| Hypothetical model of MAPKKKK acting as a 'helper' for PexRD2.**

a) PexRD2 is translocated into a host cell and, in the absence of its helper protein, exerts a weak virulence function. b) PexRD2 is translocated into a host cell and interacts with – and is directly phosphorylated by – its helper protein MAPKKKK leading to enhanced virulence function.

We set out to test the hypothesis that MAPKKKK could post-translationally modify PexRD2 *in planta* through transient co-expression in *N. benthamiana* tissue and subsequent infection with live *P. infestans*. Once more, this was carried out with the two constructs discussed previously, one comprising full length MAPKKKK and the other comprising just its N-terminal kinase domain. These were co-expressed with PexRD2 and pathogen growth was compared to growth on leaf tissue expressing PexRD2 alone. In both cases, a statistically significant enhancement of *P. infestans* mean lesion size was observed where one of the MAPKKKK constructs was co-expressed with PexRD2 relative to tissue where PexRD2 was expressed alone. This strongly suggests that MAPKKKK acts as a 'helper' to PexRD2, enhancing its virulence function and, so, facilitating the growth of *P. infestans* on plant tissue. Given that there is robust evidence for a direct interaction between MAPKKKK and PexRD2, the simplest explanation for this observation is that

MAPKKKK directly phosphorylates PexRD2 and that this post-translational modification enhances its virulence function.

It remains a possibility that MAPKKKK is a *bona fide* virulence target of PexRD2 and does indeed enhance resistance to *P. infestans* when overexpressed. It is feasible that expression of MAPKKKK in one half of a leaf could elicit a systemic immune response which spreads to areas of the leaf infiltrated with an empty vector control only. In such a scenario, any positive regulation of immunity by MAPKKKK would not be detected as pathogen growth would be limited in all parts of the leaf to which such a signal had spread. It may be possible to test this by performing whole leaf pathogen assays where one whole leaf is infiltrated with MAPKKKK, another is infiltrated with a negative control, the two are inoculated with *P. infestans* and disease symptoms are compared. In order to test whether a systemic response is responsible for not being able to detect positive regulation of immunity mediated by MAPKKKK, one half of a leaf could be infiltrated with MAPKKKK whilst the other half could be assayed for classical defence responses such as ROS burst or upregulation of defence-related genes.

This hypothesis was tested using a mass spectrometry approach to identify post-translational modifications of PexRD2. No MAPKKKK-mediated phosphorylation of PexRD2 could be detected. However, it should be noted that coverage of the sample where kinase and effector had been co-expressed was not perfect and that one residue which we predicted may be phosphorylated with a high degree of probability was not detected (Ser22). This residue was predicted to be phosphorylated with a ~0.9 probability score. Therefore, the present analysis cannot formally exclude the possibility that MAPKKKK$^{N\text{-}319}$ is able to phosphorylate PexRD2 and we can only conclude that we cannot detect it with the methods available to us. Examples of other *Phytophthora* effector proteins from the literature suggest serine residues may typically be phosphorylated *in planta* (Van Damme et al., 2012). Three serine residues in PexRD2 are predicted to be phosphorylated with a high degree of confidence. These should serve as starting points for mutational analyses to see whether or not the observed MAPKKKK-mediated enhancement of virulence can be abolished. One cautionary point should be noted, however: the NetPhosK server makes predictions based on a machine learning algorithm which

incorporates data on residues known to be phosphorylated by human kinases. Any predictions made in this way should be treated purely as predictions.

If MAPKKKK does indeed phosphorylate PexRD2 and it simply has not been detected with the methods previously described, one option may be to try enriching for phosphopeptides using Titanium Dioxide coupled to replacing trypsin with a low specificity protease to enhance sequence coverage (Schlosser et al., 2005). Further to this, a phosphor-staining gel or an antibody against specific phosphorylated residues could be deployed before further attempts are made to optimise MS approaches to detect the exact nature of any MAPKKKK-mediated phosphorylation of PexRD2. In addition to this – only co-expression of the kinase domain with PexRD2 was analysed. Given that the kinase domain was sufficient for the interaction with PexRD2 we felt it may be sufficient for any potential enzymatic activity. However, it is also possible that MAPKKKKs C-terminal domain performs an important regulatory function related to MAPKKKK's biochemical activity. Indeed, the literature lends support to the idea that the C-terminal domain of MAPKKKK may be critical (Delarosa et al., 2011). This correlates with the observations made in yeast 2-hyrbid assays that whilst MAPKKKK's kinase domain can mediate an interaction with PexRD2, only the full length kinase interacts with PexRD2 in –Uracil auxotrophy assays, suggesting a role for MAPKKKK's C-terminal region in regulation/stability of the protein. Whilst the regulatory domains of PAK kinases appear to be negative regulators, recruited to prevent autoactivity (House and Kemp, 1987, Zhao et al., 1998), the GCK kinases (of which MAPKKKK is a member) tend to have C-terminal domains observed to be important for activation of downstream signalling modules (Su et al., 1997, Kyriakis, 1999). Further work should focus on assessing the potential of MAPKKKK's C-terminal domain for mediating phosphorylation of PexRD2.

Clarification on any biological relevance for the MAPKKKK-PexRD2 interaction could be provided by improved knowledge of the pathways in which it operates. Currently, its role in immune signalling is not understood. Identification of the invasion pattern which MAPKKKK is involved in signalling the perception of, as well as identification of MAPKKKK's substrate, or other downstream signalling modules, would provide the toolkit necessary to fully interrogate the relevance of the

interaction in *P. infestans*-host interactions. One clue could be provided by the substrate of MAPKKKK's homologue in yeast; Ste20p. As discussed earlier, Ste20p interacts with, and phosphorylates Ste11p, whose C-terminus shares significant homology with a predicted NPK-1-related protein kinase in *S. lycopersicum* (Banno et al., 1993, Jin et al., 2002). *N. tabacum* NPK1 is able to complement *Saccharomyces cerevisiae* strains deficient in related kinases PKC1 and BCK1 but interestingly not Ste11p. The function of NPK1 and related kinases *in planta* is well understood to be the control of cytokinesis and the formation of cell plates, and overexpression of an inactive mutant form of NPK1 in *N. tabacum* cell suspensions leads to the formation of multi-nucleated cells (Nishihama et al., 2001). However, it is also generally accepted that NPK1 probably functions in many, diverse, unrelated pathways in the plant cell, including immune signalling (Hirt, 2000, Jin et al., 2002). If these pathways are conserved between plants and yeast, then plant NPK1 homologues should be interrogated as a potential substrate of MAPKKKK and as a regulator of immunity. This could provide insights into whether or not MAPKKKK is a potential target of PexRD2 or whether it may function as a helper of PexRD2. Figure 32 shows the *S. cerevisiae* Ste11p sequence (residues 301-727) aligned with a subset of plant NPK1-like proteins which share 44-45% sequence identity with this region of Ste11p. These proteins should be investigated through protein-protein interaction studies and through assaying MAPKKKK's ability to phosphorylate them *in planta*.

Presently, the evidence points toward MAPKKKK not being a direct virulence target of PexRD2, in the sense that there is no evidence to support the hypothesis that PexRD2 interacts with MAPKKKK in order to perturb some function which either positively regulates immunity to *P. infestans* or negatively regulates *P. infestans*' ability to derive nutrients from the host cell. *In planta* assays with live *P. infestans* suggest a role for MAPKKKK in enhancing PexRD2's virulence function and we hypothesised that this is through PexRD2 hi-jacking MAPKKKK's role as a component of a MAPK signalling pathway, although conclusive evidence for or against this has not been forthcoming and the mechanism by which MAPKKKK enhances PexRD2's virulence function remains unclear.

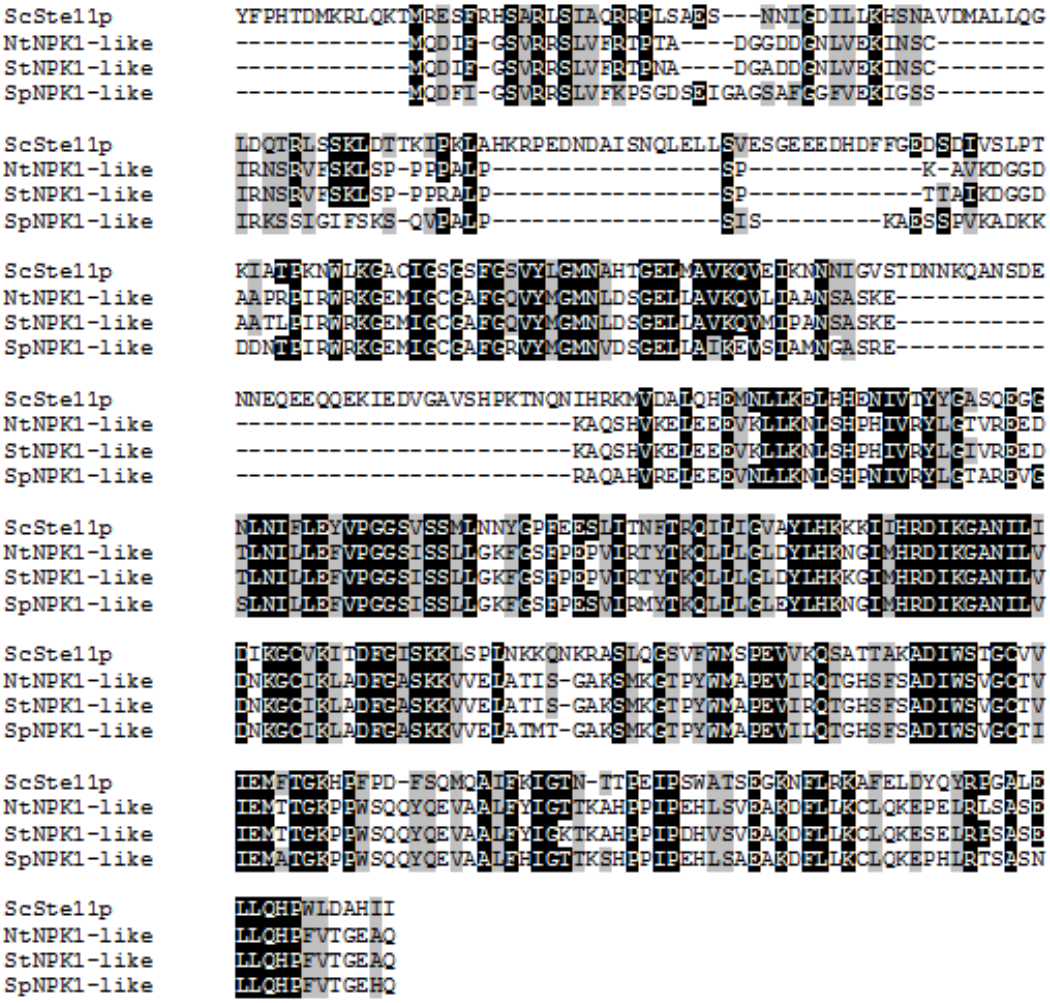



**Figure 32| Clustal Omega multiple sequence alignment of plant NPK1-like proteins against *S. cerevisiae* Ste11p.**

*N. tabacum* (44% ID), *S. tuberosum* (45% ID) and *S. penelli* (44% ID) aligned to the C-terminal region of Ste11p. Residues highlighted in black indicate identical residues whilst residues highlighted in grey highlight similar residues (e.g. hydrophobic side chains).

# 5

# Towards the crystal structure of CSEP0162 from the barley powdery mildew pathogen

# Chapter 5 Towards the crystal structure of CSEP0162 from the barley powdery mildew pathogen

## 5.1. Introduction

**C**andidate **S**ecreted **E**ffector **P**rotein_0162 (CSEP_0162) is one of the 248 CSEPs identified through analysis of the *Blumeria graminis* f. sp. *hordei* haploid genome (Spanu et al., 2010) which were predicted on the basis of the presence of a signal peptide (SP), lack of predicted transmembrane domain (TM Domain) and a lack of homology to known proteins outside the Erisyphales (Powdery Mildews). CSEP0162 is a family 4 CSEP comprising 155 amino acids, a characteristic, N-terminal W/F/YxC motif (FHC) and two predicted disulphide bonds. The N-terminal YxC motif appears to be an ancient, and highly conserved, feature of effector proteins from both powdery mildews and rust pathogens. However, any potential significance in pathogenesis, effector function and effector transport for this three amino acid motif has yet to be established (Godfrey et al., 2010).

The insights into biological function that a deep structural understanding of a protein involved in pathogenicity or resistance can offer, and the consequences of this for engineering disease resistance, were discussed in section 1.7. In the field of fungal effectors, structural biology has provided a number of key insights into the mechanisms of effector translocation and recognition. To date, the crystal structures of a number of fungal effector proteins have been described.

*Melampsora limi* infects flax and secretes a number of effector proteins, one of which is AvrM, which directly associates with – and is recognised by – the NLR protein ‘M’ (Catanzariti et al., 2010). A number of allelic variants of AvrM exist in nature and the crystal structures of two have been solved; AvrM-A (which is

recognised by M, eliciting HR) and AvrM (which is not recognised) (Catanzariti et al., 2006, Ve et al., 2013). It was observed that both variants adopted a similar, dimeric conformation as well as a structural motif reminiscent of the WY-fold from oomycete effectors. Further to this, a hydrophobic surface patch of both variants was demonstrated to mediate pathogen independent internalisation of AvrM. M-mediated recognition and ETI induction was also shown to be dependent on the C-terminal CC domain of AvrM-A – a domain previously identified as necessary for interaction in yeast and by far the most polymorphic region of the protein.

The *Leptosphaeria maculans* (Oilseed rape stem canker) AvrLm4-7 effector is recognised by two NLR proteins; Rlm4 and Rlm7, and its silencing dramatically affects the virulence of *L. maculans*. The analysis revealed a protein with no close structural similarity to any known protein and which contained 4 disulphide bridges (Blondeau et al., 2015). A conserved C-terminal domain, together with a RAWG motif located nearby combine to form a positively charged region which may be involved in pathogen independent translocation of AvrLm4-7 into plant cells. Further to this, the basis of recognition of AvrLm4-7 by Rlm4 and Rlm7 was deconvoluted thanks to structurally informed mutagenesis experiments.

Structure-based insights into effector function have also come from extracellular effectors. *C. fulvum* secretes a LysM effector named Ecp6 which interferes with either the perception of chitin, or the transduction of its perception into a cellular response (Bolton et al., 2008, de Jonge et al., 2010). The crystal structure of Ecp6 revealed the presence of three LysM domains, of which LysM1 and LysM3 closely associate and form a chitin-binding groove and that this conformation is induced by chitin binding. The chitin-binding groove displayed picomolar binding affinity for chitin, suggesting that Ecp6 outcompetes host chitin receptors for elicitor binding (Sanchez-Vallet et al., 2013).

Despite the availability of structures for several different fungal effectors, there is an overall paucity of structural information available relative to the numbers of fungal candidate effectors identified and the likely diversity of structures and functions they adopt. Homology-based secondary structure prediction based on CSEP0162's amino acid sequence using the Phyre2 server yielded a model based on an Acyl-CoA N-Acyltransferase with a low confidence value of 27.3, as well on

several other protein structures with similar, low confidence ratings. Homology modelling, therefore, is unsuitable for making predictions on CSEP0162's function and experimentally derived structural data is required for reliable inferences to be made. Here, we aimed to solve the tertiary structure of CSEP0162 by X-ray crystallography, to take the first steps toward understanding the function of this effector protein.

### 5.1.1. Principles of crystallisation.

Single molecules are inherently unsuitable for structure solution from x-ray diffraction data as the signal would be essentially unmeasurable. The arrangement of protein molecules in an ordered, three dimensional lattice produces a diffraction signal sufficient for solution of their macromolecular structure. For crystallisation to occur, the concentration of a molecule in solution needs to gradually increase until a state of super-saturation is reached, in which crystal nuclei form allowing crystals to grow. Development of protein crystals in a solution in which concentration had been increasing halts the increase in concentration and subsequent precipitation out of solution as the sequestration of protein molecules in the crystal lowers their concentration in solution. This state is referred to as the 'metastable zone'. In this state, no further nucleation points should form and crystals may continue to grow. In the present study, this was achieved by vapour diffusion from sitting drops. A protein drop mixed with a precipitant and a well containing only a precipitant are set up together in a closed system. The precipitant well contains a higher concentration of precipitant than the protein drop. Therefore, water diffuses from the protein drop as vapour allowing the concentration of protein in the drop to increase in a linear fashion. A representation of this is shown in Figure 33.

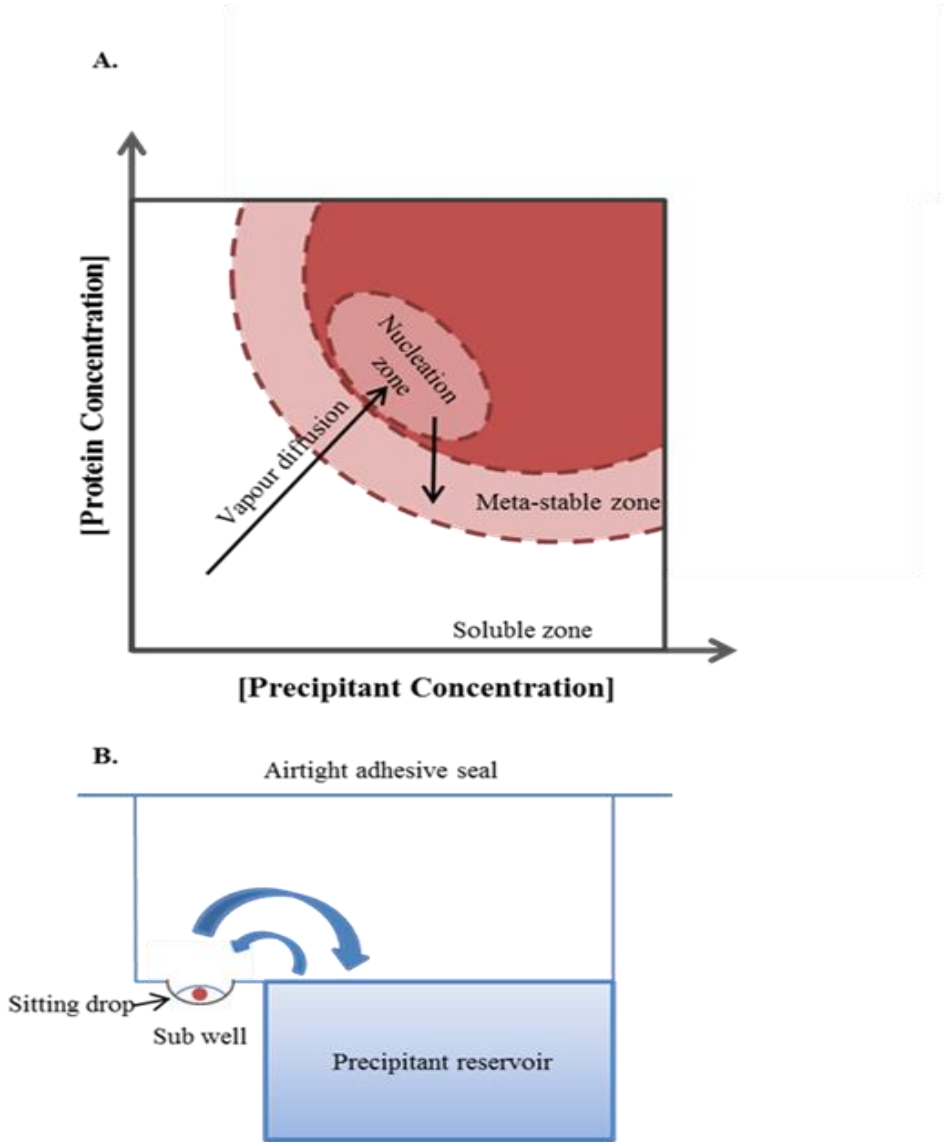



**Figure 33| Diagram demonstrating crystal nucleation and sitting drop vapour diffusion.**

(A) Diagram demonstrating crystallisation by vapour diffusion. Protein and precipitant concentration increases as vapour diffuses into precipitant well until concentration is sufficient for nucleation to occur. Sequestration of protein in crystals reduces concentration and creating the 'meta-stable zone'. (B) Diagram showing the architecture of crystallisation champers in the sitting drop vapour diffusion method.

## 5.2. Results

### 5.2.1. Expression and purification of native CSEP0162

*Note: prior to starting the work described in the present chapter, Agniezska Siwoszek (University of Copenhagen) had constructed the recombinant plasmid pOPINF:CSEP0162 for heterologous expression in E. coli.*

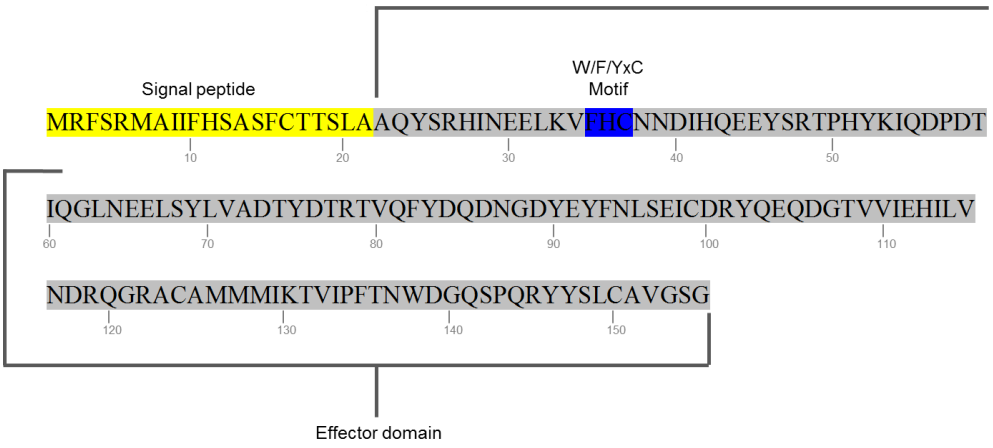



**Figure 34| Amino acid sequence of CSEP0162.**

Signal peptide is highlighted in yellow. Effector domain highlighted in grey. W/F/YxC (FHC) motif highlighted in blue.

Native *Bgh*CSEP0162 (amino acids 21-155) codon optimised for *E. coli* expression and synthesised by Genscript USA was ligated into the *E. coli* expression vector pOPINF (Berrow et al., 2007) using IN-FUSION® HD ligation independent cloning to produce the expression construct pOPINF:CSEP0162. CSEP0162 is predicted to contain an N-terminal signal peptide which ends at residue 21. This was omitted from the final expression construct on the basis that signal peptides are not thought to be part of mature, secreted effector proteins (Nielsen et al., 1997). The amino acid sequence of the final construct is shown in Figure 34**.** Disorder predictions indicated that residues 21-155 were well ordered (Figure 35) so were included in the final expression construct. The final construct was expressed with a 3C cleavable hex-histidine tag. This is shown, along with other features of the pOPINF cloning site, in Figure 36.

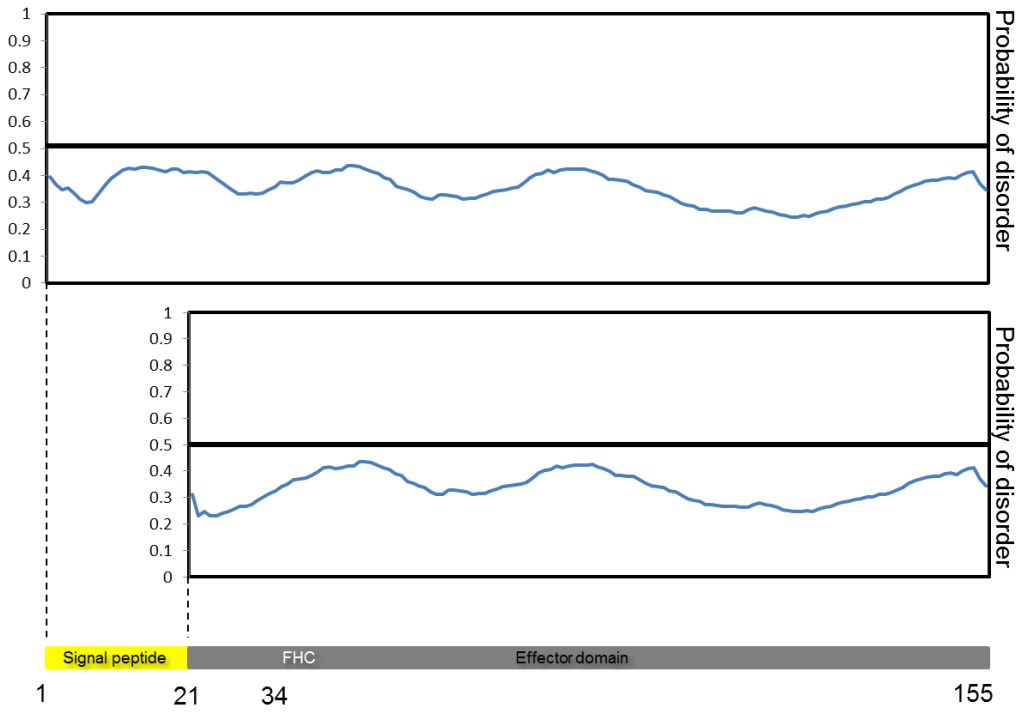



**Figure 35| Disorder prediction for full length CSEP0162 and truncated, effector domain only construct of CSEP0162 used for expression and crystallisation.**

Disorder predictions were made using the RONN protein disorder prediction server (**http://www.strubi.ox.ac.uk/RONN**). Disorder prediction traces are shown aligned with a schematic representation of CSEP0162 with signal peptide and effector domain highlighted as in figure 38. Black line at 'Probability of disorder' value of 0.5 represents a cut-off point. A trace above this line indicates regions of protein likely to be disordered.

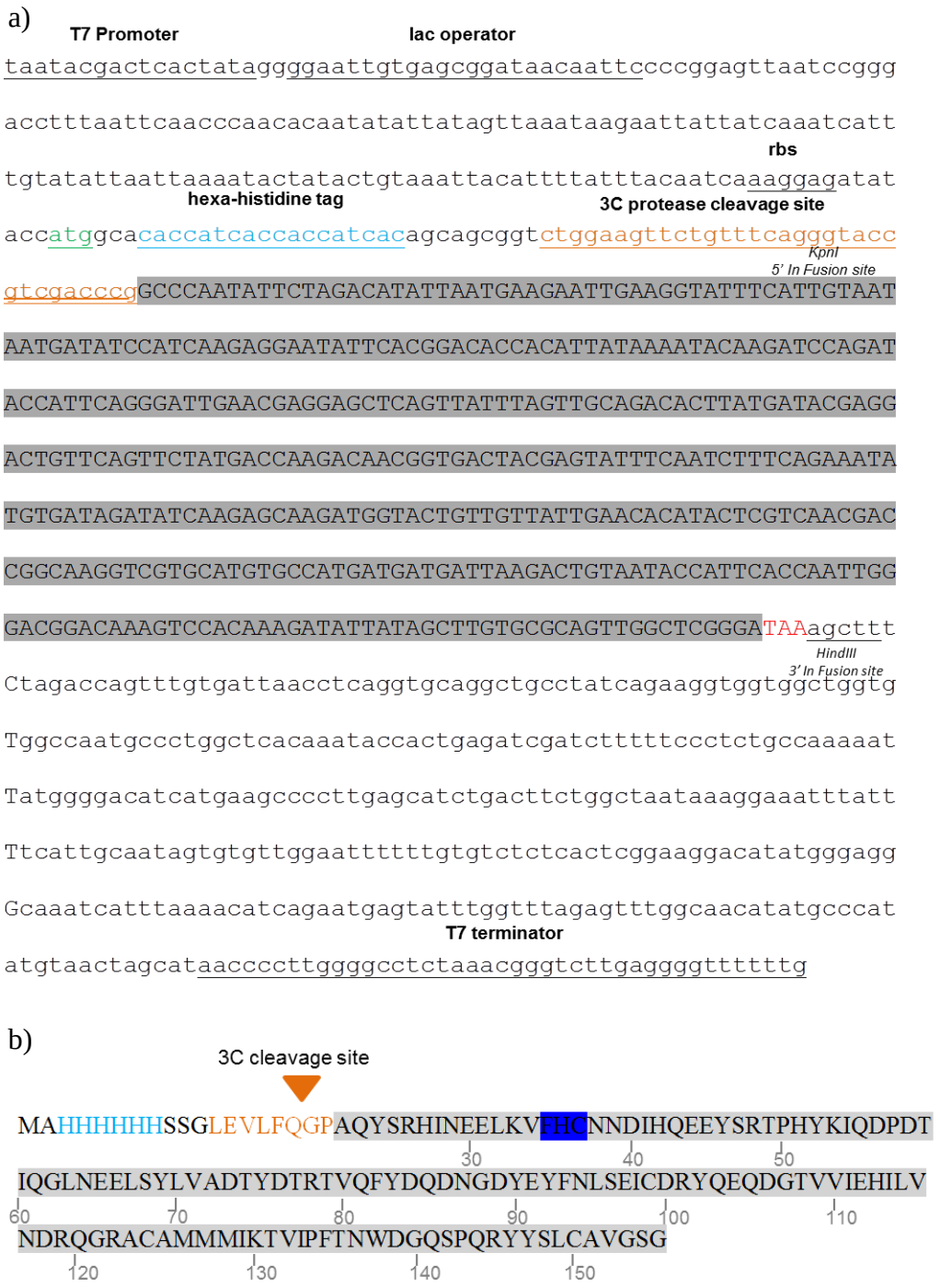



**Figure 36| pOPINF:CSEP0162.**

a) DNA sequence of CSEP0162 (shaded grey) cloned into pOPINF. Important vector features have been indicated by labelling. Hexa-histidine tag and 3C protease site are indicated in light blue, and orange type, respectively. b) Amino acid sequence of CSEP0162 cloned into pOPINF, with effector domain highlighted in grey, FHC motif highlighted in blue and hexa-histidine tag and 3C protease cleavage sites shown in light blue and orange type, respectively.

Like other YxC motif containing powdery mildew CSEPs, CSEP0162 is predicted to form at least one internal disulphide bond. Therefore, to aid soluble expression, SHuffle® Express T7 Competent *E. coli* cells (New England Biolabs) were transformed with pOPINF:CSEP0162 and selected on LB agar supplemented with Carbenicillin. SHuffle® cells promote the soluble expression of disulphide bond containing proteins through the cytoplasmic expression of the disulphide bond isomerase DsbC. Single, well separated colonies were selected and used to inoculate 10mL aliquots of LB supplemented with Carbenicillin to be used as starter cultures for scaled-up expression as described in section 2.6.2.

Cell pellets were harvested from liquid cultures and protein was purified as described in section 2.6.4. Native CSEP0162 eluted as a single, broad peak with an elution volume consistent with that previously observed by Agnieszka Siwoszek (Univ. Copenhagen). A reasonably pure protein, migrating between molecular weight markers of 12kDa and 22kDa (consistent with pOPINF:CSEP0162's predicted molecular mass of ~16kDa) was visualised through SDS-PAGE analysis of the fractions collected under the peak alongside the insoluble and soluble fractions collected earlier in the purification procedure (Figure 37).

These fractions were pooled and concentrated to a volume of ~8mL at a concentration of ~16mg/mL.

A sample of the concentrated protein was retained for SDS-PAGE analysis. Recombinant 3C protease was added to concentrated protein prior to incubation at 4˚C overnight in order to cleave the hexa-histidine tag. A further sample was retained immediately following removal from 4˚C for SDS-PAGE analysis whilst the rest of the sample was applied to a pre-equilibrated nickel column manually. Eluted protein was concentrated to <10mL total volume and manually applied to a SEC column for further, preparative purification. Once again, CSEP016 eluted as a single, broad peak with a slight shift in peak elution volume from ~200mL to ~210mL, which is expected given the smaller molecular mass of the protein lacking a hexa-histidine tag (see Figure 38 and compare peak to Figure 37). Fractions under the peak were pooled and concentrated and a sample of this final preparation was prepared for SDS-PAGE analysis alongside pre-3C protease incubation and post-3C

protease incubation samples which demonstrated almost no observable contamination and a pure protein sample suitable for crystallisation screening.

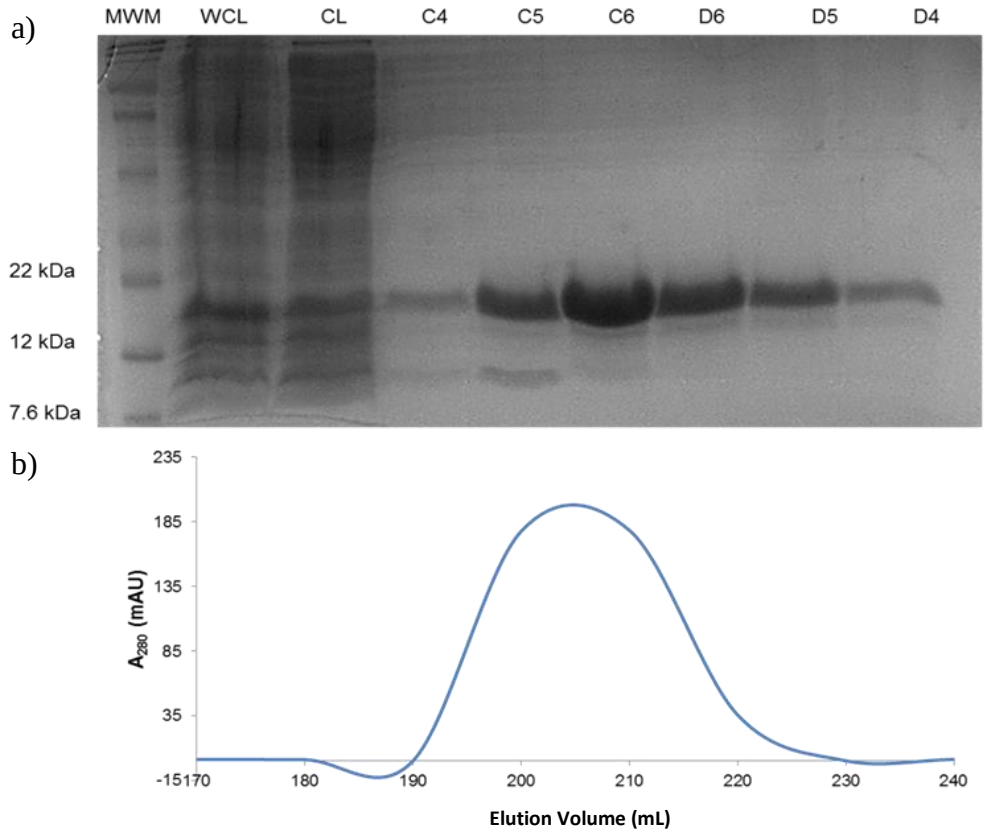



**Figure 37| IMAC purification of native CSEP0162.**

SDS-PAGE (17%) analysis of purified CSEP0162. Lane one contains molecular weight marker, lane two contains whole cell lysate, lane three contains clarified lysate and subsequent lanes contain SEC fractions collected under the $A_{280}$ peak shown (below). $A_{280}$ trace shows profile of SEC elution of $HIS^6$-CSEP0162. Peak spans all fractions shown in SDS-PAGE gel (above).

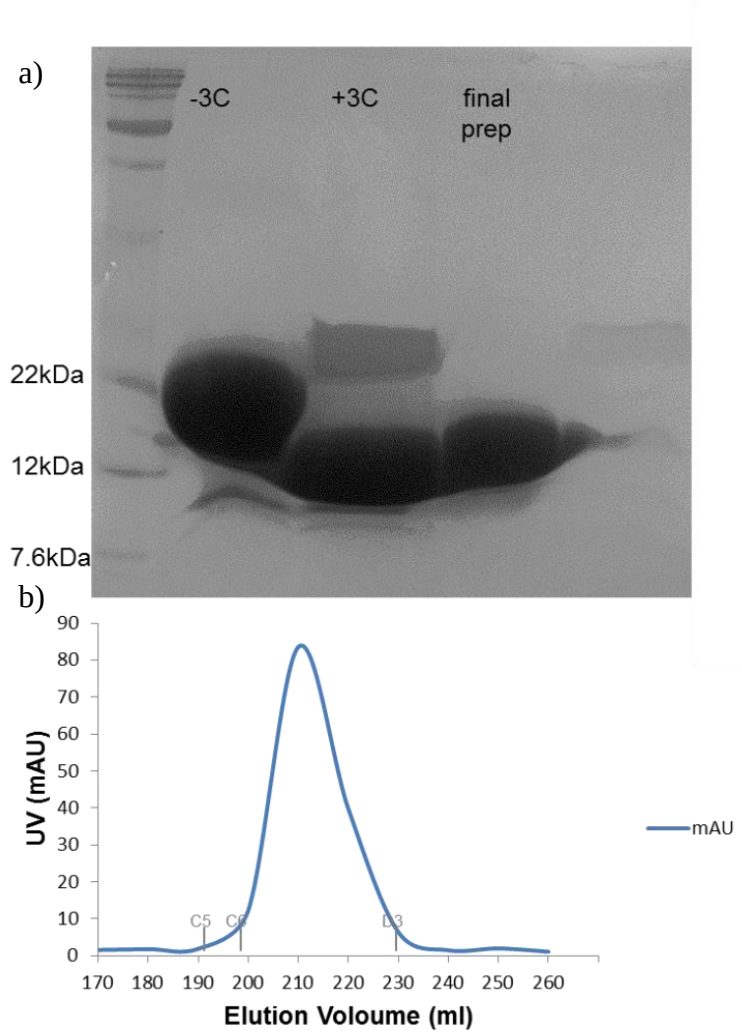



**Figure 38| Purification of untagged CSEP0162.**

Following $His^6$ tag cleavage, CSEP0162 was purified once more to remove 3C protease from the final prep. SDS-PAGE (17%) analysis showing; Lane 1 – molecular weight marker, lane 2 – CSEP0162 without 3C protease, lane 3 – CSEP0162 with 3C protease and lane 4 final prep of CSEP0162 used in crystallisation trials following IMAC purification to separate CSEP0162 and 3C protease and a second round of SEC. SEC $A_{280}$ trace shown below. Peak was collected and concentrated to give 'final prep'.

### 5.2.2. Circular Dichroism Spectroscopy of native CSEP0162

To determine whether or not purified CSEP0162 formed an ordered protein with interpretable secondary structure, we subjected the protein to Circular Dichroism (CD) Spectroscopy. CD spectroscopy is a technique which analyses the difference in absorption of Left Circularly-Polarised Light (LCPL) and Right Circularly-Polarised Light (RCPL) by a molecule containing one, or more, chiral chromophores. As different secondary structure elements absorb circularly polarised light of different wavelengths at different rates, CD spectroscopy can be used to determine the relative composition of secondary structural elements in a protein.

Native CSEP0162, prepared as described previously, was diluted to ~0.8mg/mL in phosphate buffer in order to reduce the interfering effects of A4 buffer components (NaCl and HEPES). In a 0.5mm cell, at a temperature of 20˚C, CD spectra were collected as described in (Walden et al., 2014) from 180nm to 260nm at 1nm intervals, with measurements at each wavelength being taken in triplicate and averaged. Raw output in millidegree units was corrected by subtracting the absorbance of the phosphate buffer and converted to Mean Residue Elipticity (MRE) (degrees $cm^2$ $dmol^{-1}$ $residue^{-1}$) using the equation:

$$MRE = \theta \, x \, \frac{(0.1 \, x \, MRW)}{(l \, x \, c) x \, 3298}$$

Where θ = raw millidegree units, MRW = Mean Residue Weight as determined by dividing the molecular weight of the protein in Da by the number of residues – 1, $l$ = path length of CD cell and $c$ = concentration in mg/mL. The CD spectrum for native CSEP0162 is shown in Figure 39.

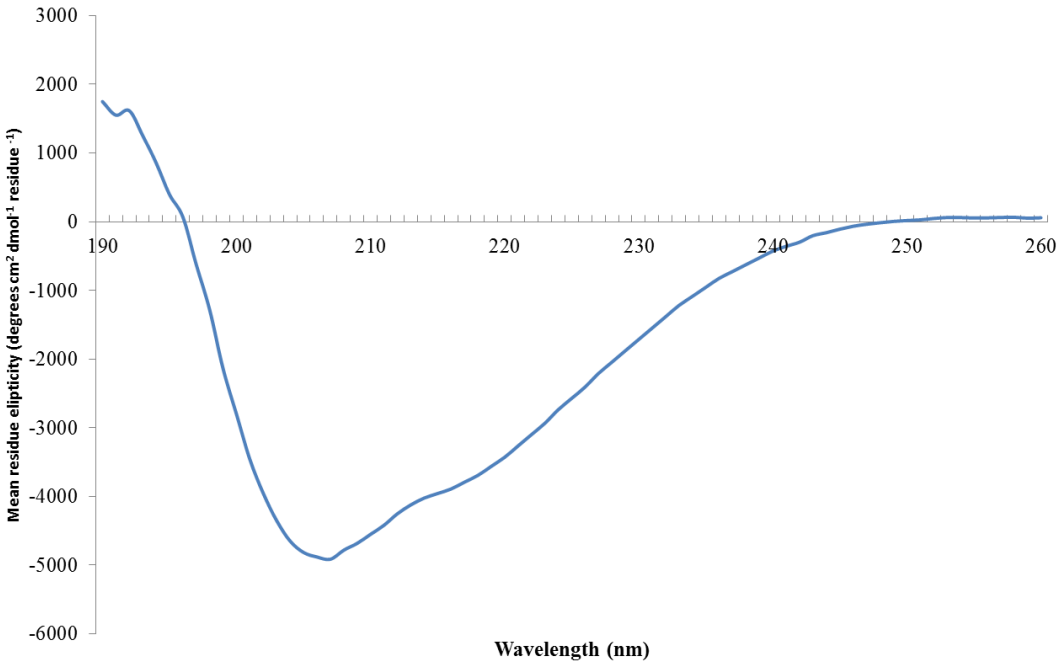



**Figure 39| Circular dichroism spectrum of native CSEP0162.**

Average (3 measurements per wavelength) and buffer subtracted CD spectrum for native CSEP0162. Data were collected from 190nm to 260nm in 1nm steps and converted from millidegree units to mean residue elipticity. Signal from averaged buffer only sample was subtracted in order to give final trace.

In order to assign secondary structure elements, the averaged CD spectrum was submitted to the DichroWeb server where the Cdsstr method – using reference set 7 – was applied to the data (Whitmore and Wallace, 2008, Sreerama et al., 1999, Sreerama and Woody, 2000). Interpretation is shown in **Error! Reference source not found.**.

| Helix 1 | Helix 2 | Total | Strand 1 | Strand 2 | Total | Turns | Unordered | Nrmsd |
|---|---|---|---|---|---|---|---|---|
| 0.00 | 0.04 | 0.04 | 0.22 | 0.09 | 0.31 | 0.17 | 0.48 | 0.029 |

**Table 11| Secondary structural assignment from CD spectra of native CSEP0162.**

Figures represent fraction of CSEP0162 secondary structure composed of each secondary structure element. Helix1 and Strand 1 refer to regular helices and strands whilst Helix 2 and Strand 2 refer to distorted helices and strands. Nrmsd = normalised root mean squared deviation and provides a measure of goodness of fit of this result to the raw data.

Analysis of secondary structure elements of native CSEP0162 reveals a predominantly β-stranded protein, with a smaller proportion of α-helical elements and nearly half of the protein made up of unordered loops. An Nrmsd value of 0.029 gives a strong indication that the model is reliable.

## 5.2.3. Crystallisation of native CSEP0162

### 5.2.3.1 **Screening**

Protein purified as outlined in section 5.2.1 was used in crystallisation trials at a final concentration of ~49mg/mL using the sitting drop vapour diffusion method (Figure 33). Crystallisation proceeded in 96 well plates capable of accommodating two protein drops per well, one of which contained CSEP0162 at 1x concentration with the other containing CSEP0162 at 0.5x concentration in order to explore the effects of concentration on crystallisation. Initially, one commercially available crystallisation trial was used – the JCSG-*plus*™ screen (Molecular Dimensions) – in order to assess a sparse matrix of crystallisation conditions exploring well known polyethylene glycol (PEG) and salt conditions and pH values ranging from 4.0-10.0.

Several conditions yielded crystalline precipitate or small crystals/needles whilst two conditions yielded substantial crystals, one of which yielded crystals considered suitable for harvesting (see Figure 40a). All crystals from this screen grew in clusters of rods around a central nucleation point. Both conditions yielding clear, rod-like crystals contained ammonium sulphate, Bis-TRIS at pH 5.5 and PEG3350 at varying concentrations. The condition yielding harvestable crystals contained 1.0M

Ammonium Sulphate, 0.1M Bis-TRIS pH 5.5 and 1% PEG3350. Crystals formed in both 1x and 0.5x concentrated protein drops after 2 days and ceased growth after 3-4 days. Three crystals were harvested from the 1x concentrated drop, one of which was suspended in cryoprotectant solution comprising the mother liquor supplemented with 25% ethylene glycol as described in section 2.7.3. A further two crystals were harvested and soaked in the aforementioned cryoprotectant solution supplemented with 0.5M Potassium Iodide (KI) for ~60s. Notably, the crystals were extremely fragile, particularly when suspended in the 0.5M KI soaking solution. Despite the presence of far more than three crystals, the Ammonium Sulphate in the drop was quick to precipitate once the drop was allowed to evaporate making it difficult to distinguish between protein crystal and salt crystal.

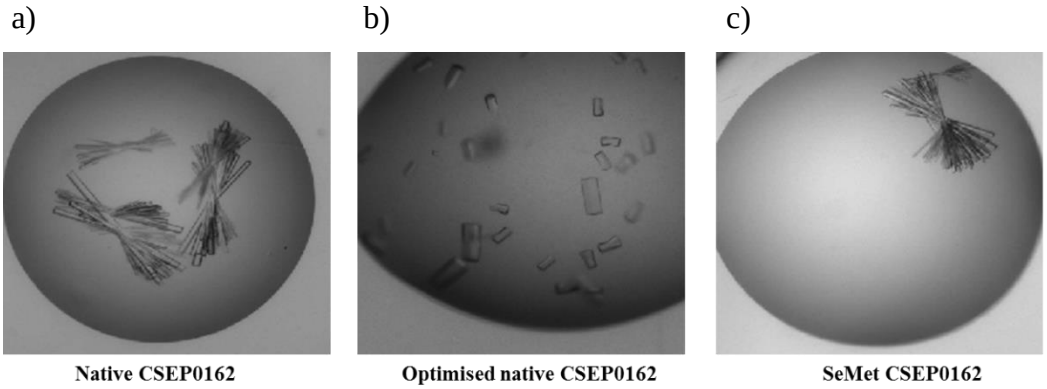



**Figure 40| Native and SeMet derivative crystals of CSEP0162.**

a) Representative image of initial hit described in section 5.2.3.1. Rod-like crystals growing in clusters around a central nucleation point. b) Optimised native CSEP0162 crystals from screens described in section 5.2.3.2. c) SeMet derivative crystals from screens described in section 5.2.4.

#### 5.2.3.2 **Optimisation**

The condition described in 5.2.2.1 from which crystals were harvested was used to design a 48 condition optimisation. Bis-TRIS concentration was kept constant and the same protein prep from the initial screen was used, again, at ~49mg/mL. Ammonium sulphate concentration was varied from 0.0M to 1.0M and PEG3350 was varied from 0% to 25%.

Crystals of a new form appeared after ~3 weeks in 1.0M Ammonium Sulphate, 0.1M Bis-Tris pH 5.5 and 5% PEG3350 (see Figure 40b). Crystals were harvested as described in 5.2.2.1 and 2.7.3. Notably, these crystals were far more stable than their counterparts from initial screening conditions.

### 5.2.4. Expression and purification of a CSEP0162 Selenomethionine derivative

Two pieces of information are required in order to derive the 3D structure of a macromolecule by X-ray crystallography; the amplitude of diffracted X-rays and their phases. A diffraction spot on a detector is produced by a diffracted X-ray wave which has both an amplitude and a phase, but the detector only records information related to the amplitude of the wave. The phase information is lost. If one were to attempt to solve a crystal structure without phase information, a map of vectors between atoms could be produced – known as a Patterson map – but in order to obtain information on the electron density and produce an accurate map of a macromolecule then phase information is required. This is known in crystallography as 'the phase problem' and solving the phase problem is often the most important step toward solving the 3D structure. In order to provide more data with which to solve the phase problem, it was necessary to produce a derivative of CSEP0162 in which native methionine residues were replaced with L-selenomethionine to produce a protein henceforth referred to as 'SeMet CSEP0162'. As outlined in Figure 34, there are three methionine residues adjacent to each other from residue 126-129 in the effector domain. In the final pOPINF:CSEP0162 construct produced by Agnieszka Siwoszek (Univ. Copenhagen) there is an additional methionine residue at residue 3 which is retained following hexa-histidine tag cleavage giving a total of 4 methionine residues in the final protein.

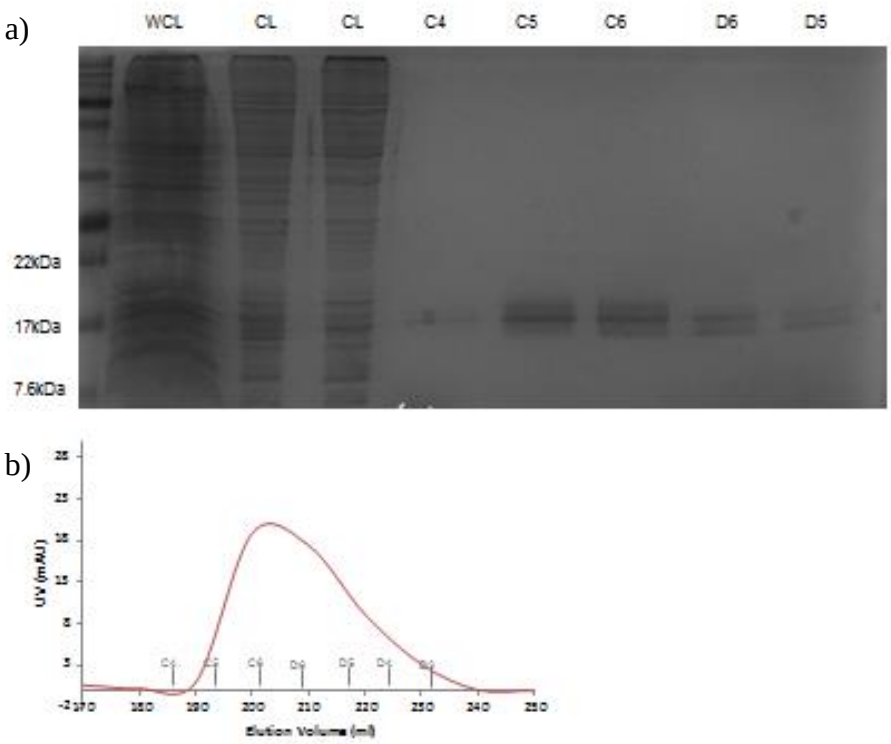



**Figure 41| Expression and purification of L-selenomethionine derivative of CSEP0162.**

SDS-PAGE (17%) analysis of L-selenomethionine derivative CSEP0162. Lane one contains molecular weight marker, lane two whole cell lysate, lanes 3 and 4 contain clarified lysate and following lanes contain size exclusion chromatography (SEC) fractions collected under the A280 peak shown below. A280 trace shows profile of SEC purification of HIS6-CSEP0162 (L-selenomethionine derivative). Peak spans all fractions shown in SDS-PAGE gel above.

Full procedure for production of SeMet CSEP0162 is outlined in section 2.6.3 but broadly speaking two methods were attempted and highlighted the importance of proper disulphide bond formation for CSEP0162 solubility. Initially, a methionine auxotrophic strain of *E. coli* (B834(DE3)) was used and protein expression was induced with 1mM IPTG in minimal media supplemented with L-selenomethionine only. CSEP0162 could be detected using this method but it was almost completely insoluble as the band corresponding to CSEP0162 was more intense in the lane containing whole cell lysate relative to the lane containing clarified lysate (data not shown). A second method, entailing metabolic inhibition of methionine biosynthesis

was employed to greater success. Briefly, the SHuffle *E. coli* cells successfully used to produce native protein were grown in minimal media until an OD600 of 0.6 was reached. At this point high concentrations of lysine, isoleucine and threonine are added to the growth medium in order to inhibit aspartate kinase activity and block synthesis of native methionine. Subsequent supplementation of growth medium with L-selenomethionine followed by induction with 1mM IPTG encourages incorporation of L-selenomethionine into nascent proteins. Cells were harvested and protein purified as described previously. Once again, protein eluted as a single, broad peak and fractions collected under the peak were analysed by SDS-PAGE (Figure 41). Yield was considerably lower than obtained from expression of native protein but sufficient levels of pure, soluble protein were obtained for crystallisation. Samples were pooled and concentrated as described previously and hexa-histidine tag was cleaved. Cleaved protein was manually applied to a nickel IMAC column and eluted protein lacking a hexa-histidine tag was concentrated with buffer exchange to replace IMAC elution buffer with gel filtration buffer. The final gel filtration step described for native CSEP0162 was not carried out to avoid sample loss and was deemed unnecessary as SDS-PAGE analysis revealed the sample was already very pure (Figure 42a).

In order to confirm the identity of expressed protein, a 25 pmole sample of protein in solution was submitted for intact mass analysis by Dr Gerhard Saalbach (JIC Proteomics Platform) which confirmed an intact mass of 16188.17 Da, this was consistent with a theoretical intact mass of 16,188.59 Da (Figure 42b).

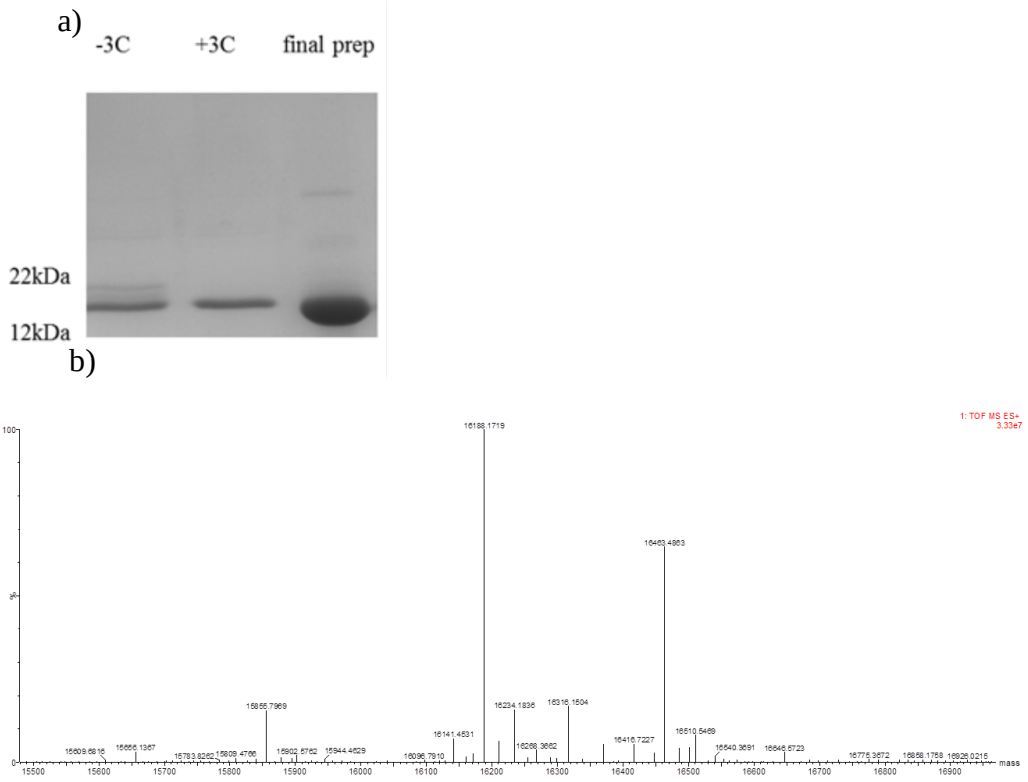



**Figure 42| His6 tag cleavage by 3C protease and subsequent intact mass analysis of selenomethionine CSEP0162 derivative.**

a) SDS-PAGE (17%) analysis of CSEP0162. Lane 1 shows protein without 3C protease addition (auto-cleavage is evident), lane 2 shows protein with 3C protease addition and lane 3 shows final prep following IMAC purification to remove 3C protease and subsequent concentration for crystallographic screening. b) Intact mass analysis confirms incorporation of selenomethionine.

## 5.2.5. Crystallisation of CSEP0162 Selenomethionine derivative

Protein was concentrated to ~6mg/mL and taken forward into the previously described JCSG-*Plus*™ screen as well as a screen devised by Dr Clare Stevenson (John Innes Centre Protein Crystallography Platform) – Keep It Simple Screen (KISS). KISS explores Ammonium Sulphate concentrations ranging from 0.0M to

3.2M and PEG3350 concentrations ranging from 0% to 40% against a range of buffers and pH values. A total of 192 unique conditions, therefore, were screened. KISS was chosen based on the knowledge that native CSEP0162 crystallises in the presence of Ammonium Sulphate and PEG3350. SeMet CSEP0162 formed no crystals in the JCSG-*Plus*™ screen. However, harvestable crystals appeared in one condition from KISS within 24 h (0.8M Ammonium Sulphate, 0.1M CHES pH 9.0). Like crystals from initial screens of native CSEP0162, crystals grew as long rod shapes in clusters around a central nucleation point (Figure 40c). The concentration at which SeMet crystals grew was markedly lower than the concentration at which native crystals grew. In most cases SeMet crystals will crystallise in similar conditions and at similar concentrations to native crystals of the same protein. However, overall solubility of SeMet crystals is often slightly lower than that of their native counterparts due to the enhanced hydrophobicity of Selenomethionine relative to native methionine. It should also be noted that we did not investigate lower concentrations of native CSEP0162.

### 5.2.6. Data collection

Crystals were harvested and cryo-protected as described in section 5.2.3.1 and 5.2.3.2 and flash frozen in liquid $N_2$ before transport to the Diamond Light Source (DLS) synchrotron radiation facility, Oxfordshire, UK for collection of X-ray diffraction data. Table 12 provides a list of crystals that best datasets were collected from at the synchrotron, their origin, details of cryoprotectant and any heavy atom soak.

| Xtal ID | Native/SeMet | Cryoprotectant | Soak |
|---|---|---|---|
| **CSEP0162_23** | Native | 1M $(NH_4)_2SO_4$, 0.1M BIS-TRIS (pH 5.5), 5% PEG3350 + 25% ethylene glycol | N/A |
| **CSEP0162_18** | Native | As above | 0.5M KI (~60s) |
| **CSEP0162_32** | SeMet | 0.1M CHES (pH 9.0), 0.8M $(NH_4)_2SO_4{}^+$ 25% (w/v) ethylene glycol | N/A |
| **CSEP0162_33** | SeMet | As above | N/A |

**Table 12| List of crystals from which data were collected in this study including details of cryoprotectant and heavy atom soaks.**

All crystals were subjected to 5 test shots ($\Omega = 0.5^\circ$) to inform the automated data collection strategy calculation software EDNA.

CSEP0162_23 provided the best native dataset without a heavy atom soak. Data was collected on the i02 beamline at an incident wavelength of 0.9795Å with a maximum resolution of 2.2Å. 2400 images in total were collected with a rotation angle of $\Omega = 0.15^\circ$. Heavy atom soak data was obtained from CSEP0162_18, which was soaked in a 0.5M Potassium Iodide solution for ~60s before being washed in cryoprotectant and flash frozen in liquid $N_2$. Two datasets were collected on this crystal, with beamline intensity at 5% for one dataset and increased to 20% for the second dataset. Both datasets were collected with an incident wavelength of 1.8Å and a maximum resolution of 2.6Å. 3000 images were collected for both datasets with a rotation angle of $\Omega = 0.2^\circ$.

Four SeMet datasets were collected, on crystals CSEP0162_32 and CSEP0162_33, using the i04-1 beamline. Three datasets were collected from CSEP0162_32 with an incident wavelength of 0.9282Å and a maximum resolution of 3.2Å. 5000 images were collected initially with a rotation angle of $\Omega = 0.1^\circ$ and a further two datasets comprising 3000 images each were collected subsequently to enhance multiplicity of data. A single dataset comprising 8000 images with a rotation angle of $\Omega = 0.1^\circ$ was collected from CSEP0162_33 with an incident wavelength of 0.9282Å and a maximum resolution of 3.0Å. All four datasets were collected at a beamline intensity of 47%.

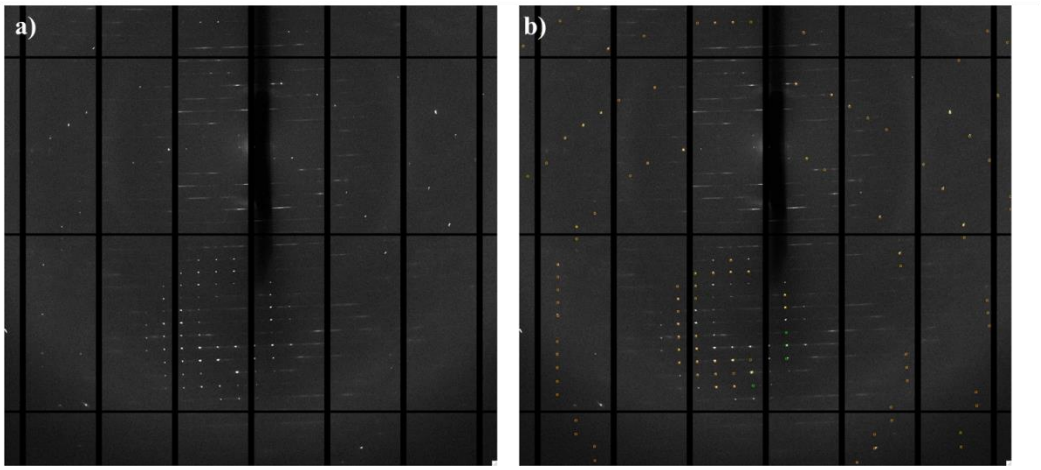



**Figure 43| Diffraction images from crystal CSEP0162_23 indicates possible OD-twinning pathology.**

(a) Shows observed reflections from one representative diffraction image from CSEP0162_23 (native) with well-defined, punctate diffraction spots interspersed with streaky reflections. (b) Shows the same representative image with predicted positions of reflections if a hexagonal or trigonal indexing solution were valid. Punctate diffraction spots correspond to these predictions but streaky spots do not. Predictions were made using iMOSFLM (Leslie and Powell, 2007).

Close analysis of diffraction images from native and KI soaked crystals revealed the presence of streaky reflections which do not correspond to where reflections are predicted to appear if hexagonal or trigonal space groups are accurate representations of crystal geometry (Figure 43). This may indicate a pathology in the data known as partially disordered ordered-disordered twinning (OD twinning). This occurs when a translation operator along a given axis for a subset of layers in a crystal lattice is offset relative to other layers in the same lattice. This type of twinning is characterised by the streaky, diffuse diffraction spots also observed in Figure 44 which are indicative of a perturbation in translational symmetry. This phenomena was first reported in (Bragg et al., 1954) and methods for solving crystal structures despite the presence of this pathology were made by (Wang et al., 2005) which also presents an example of OD twinning in Figure 1.

Tellingly, when data were re-processed in lower symmetry space groups, the twinning prediction programme ‘truncate’ indicated that the data collected is most likely the result of twinning in the crystal. Only in higher symmetry space groups

like P6 2 2 did the probability of twinning reduce. Twinning is a fairly common problem in macromolecular crystallography and occurs when multiple crystals share lattice points in a symmetrical manner forming an interface between the separate crystals often called the composition surface. Twinning convolutes the interpretation of X-ray diffraction data as data is produced from multiple crystal domains in different orientations to each other, producing difficult to interpret diffraction patterns.

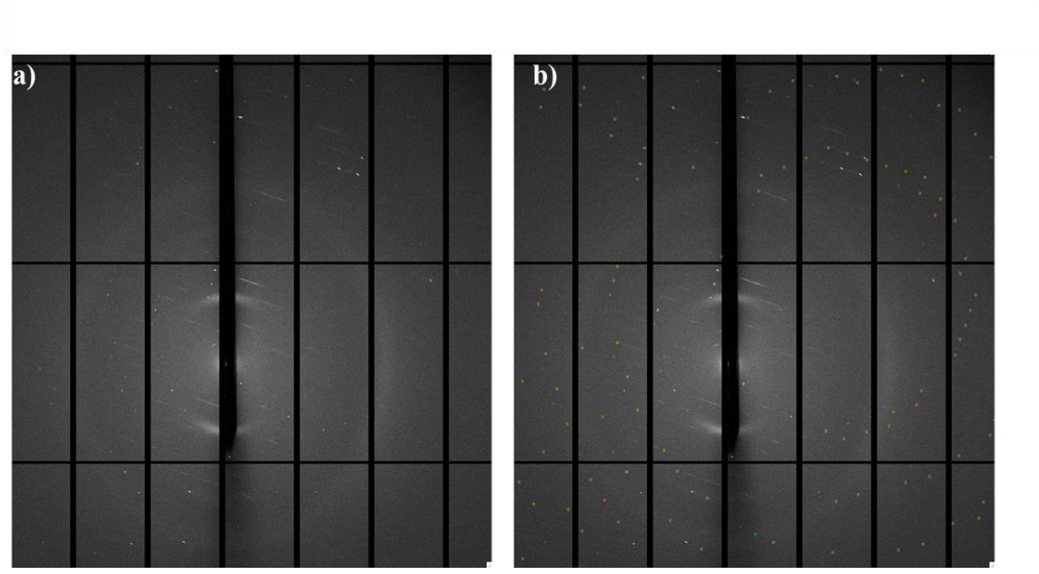



**Figure 44| Diffraction images from crystal CSEP0162_18 indicate possible OD-twinning pathology.**

As Figure 43, except data shown are representative images from crystal CSEP0162_18 (KI soaked).

### 5.2.6.1 **Fluorescence scan of SeMet CSEP0162**

In order to accurately determine appropriate wavelengths for Multi-wavelength Anomalous Diffraction (MAD) experiments, and to further confirm the presence of Selenomethionine in SeMet CSEP0162 crystals, we performed a fluorescence scan at the K edge of selenium. Tabulated, theoretical values for the wavelength an element's absorption edge corresponds to are insufficiently accurate as the absorption properties of the heavy atom in a given macromolecule will almost certainly be affected by its immediate environment. When an atom absorbs X-ray radiation, one of its electrons may be excited to a higher energy level or, indeed,

ejected from the atom altogether. Energy is lost during this process in the form of fluorescence. The likelihood of a package of fluorescence being produced at an absorption edge is known as fluorescence yield. The absorption coefficient is directly related to fluorescence by this value and can be experimentally determined by a fluorescence scan. From this information, the anomalous scattering factors – *f''* and *f'* – can be derived and appropriate wavelengths for MAD experiments determined. In X-ray crystallography, the absorption edge of the innermost electron shell – the K shell – is determined.

Selenium K edge scan of SeMet CSEP0162 was carried out on the i03 beamline on crystal CSEP0162_36, between X-ray energies 12628eV and 1282eV. Data was analysed automatically on the DLS pipeline using the CHOOCH programme (Evans and Pettifer, 2001) which normalises the recorded fluorescence spectrum and automatically calculates experimentally determined *f*' and *f*' values. Output from CHOOCH is shown in Figure 45.

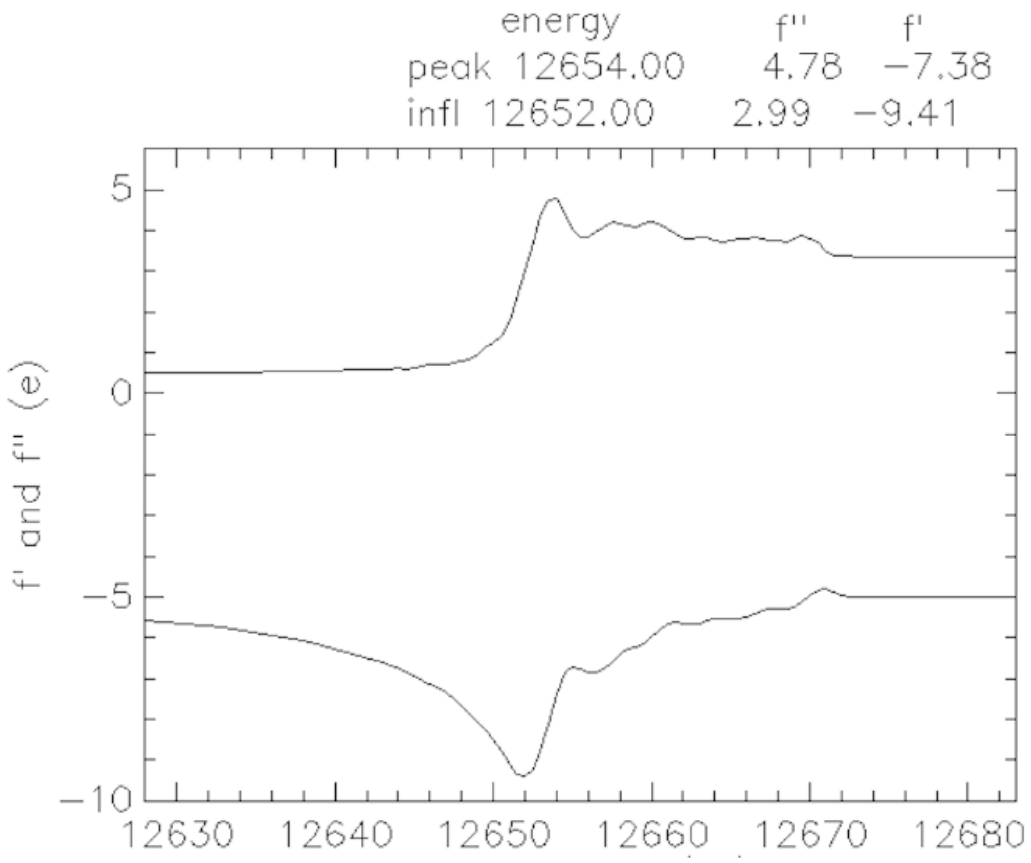



**Figure 45| CHOOCH output showing normalised fluorescence spectrum for SeMet CSEP0162.**

Crystal was scanned at wavelengths along the theoretical K edge for Selenium and values for anomalous scattering factors *f''* and *f'* were experimentally determined by CHOOCH as; *f''* = 4.78e and *f'* = -7.38e where energy at the peak was 12.654KeV.

## 5.2.7. Data processing and structure solution

### 5.2.7.1 **SAD approaches to experimental phasing**

All data were processed with the XIA2 pipeline with XDS and XSCALE (Evans, 2006, Kabsch, 2010, Winter et al., 2013). XIA2 indexes and integrates all data, reducing its complexity. A summary of data used for attempts at initial structure solution is presented in Table 13.

| | Native | KI | SeMet |
|---|---|---|---|
| **Data collection** | | | |
| **Instrumentation** | DLS-i02 | DLS-i02 | DLS-i04-1 |
| **Wavelength (Å)** | 0.9795Å | 1.8Å | 0.9282Å |
| **Resolution range (Å)** | 51.64-2.40 | 42.32-2.8 | 52.19-3.20 |
| **Space group** | $P6_2\,2\,2$ | $P6_2\,2\,2$ | $P6_2\,2\,2$ |
| **Unit cell parameters** | A=B=59.62, C=74.50, α=β=90.0, γ=120.0 | A=B= 59.41, C=74.39, α=β=90.0, γ=120.0 | A=B=60.26, C=75.10, α=β=90.0, γ=120.0 |
| **Completeness (%)** | 99.9 (100) | 98.5 (95.1) | 99.0 (98.9) |
| **Unique reflections** | 3373 | 2127 | 1537 |
| **Redundancy** | 34.5 (35.2) | 57.4 (57.2) | 149.5 (150.9) |
| **Anomalous redundancy** | - | 34.1 (32.4) | 91.0 (87.2) |
| **Anomalous slope** | - | 1.298 (0.000) | 1.443 (0.000) |
| **Anomalous correlation** | - | 0.853 (0.108) | 0.908 (0.186) |
| **$R_{merge}$** | 0.07 (2.220) | 0.145 (3.067) | 0.156 (1.115) |
| **I/σ(I)** | 29.6 (2.4) | 29.7 (1.9) | 42.2 (9.7) |
| **No. of SAD sites** | - | 7 | 4 |

**Table 13| Summary statistics of data used for experimental phasing.**

Data from XIA2 automated pipeline. Data in parentheses corresponds to data from high-resolution shells. SAD sites were determined using the AutoSol programme in the PHENIX pipeline. Note – $P6_2\,2\,2$ space group assignment is an estimate.

AutoSol, in the PHENIX (Adams et al., 2010) suite of crystallography software, was used for initial experimental phasing. Heavy atom soak data was combined with high-resolution native data, heavy atom soak data from two different beamline intensities was combined with high-resolution native data and SeMet data was combined with high-resolution native data. However, no combination of datasets yielded an interpretable electron density map, with $R_{free}$ values (a measure of how confident one can be of a solution) ranging from 0.4-0.59 for all strategies employed, indicating that no solutions were reliable. Further to this, all Figure of Merit (FOM) values were lower than 0.4 (FOM $\leq$ 0.45 means that initial phases are 'marginally accurate'). This, coupled with the electron density map itself not suggesting the interpretation of any secondary structural elements and not showing many regions of

continuous electron density or clear solvent channels, indicated that attempts to derive initial phases from SAD experiments had been unsuccessful.

The space group of a crystal provides a description of its symmetry within the unit cell by giving information on the point of a molecule (its origin of symmetry), and how it is rotated, translated and inverted in a crystal lattice. All raw datasets were assigned to space group $P6_2$ 2 2 by automated XIA2 processing. However, all space groups assigned to a crystal during the process of solving its structure are hypothetical and applied as a best fit of the data available. Data were re-indexed in alternative hexagonal space groups; P6, $P6_2$ and p6 2 2. In addition to this, the trigonal space groups (starting at P3) were also explored as the trigonal and hexagonal space groups are related. As were the C-centred orthorhombic space groups and monoclinic space groups. The lowest symmetry space group – P1 – was also investigated. A solution remained unobtainable with data re-indexed in the aforementioned space groups.

*(Note – Subsequent independent attempts by Professor Mark Banfield (JIC) and Dr David Lawson (JIC protein crystallography platform) using the CCP4 suite of crystallography programmes also proved unsuccessful in obtaining initial phase information.)*

#### 5.2.7.2 **Molecular replacement approach to experimental phasing**

Where structural information is available for a protein which may have a similar 3D structure to the protein of interest, a technique for obtaining phase information known as molecular replacement (MR) may be used (Tickle and Driessen, 1996). The technique requires an experimentally determined diffraction pattern for the protein of unknown structure and the atomic coordinates for the protein of known structure (termed the search model). An attempt is then made, computationally, to place the search model in the unit cell of the protein of unknown structure so as to match – as closely as possible – the theoretical diffraction pattern of the search model to the experimentally determined diffraction pattern for the protein of unknown structure.

With many potential search models deposited in the PDB, the chance of finding a suitable structure with which to attempt MR is often high. Typically, 30-40% sequence identity is a minimum requirement and the closer the identity, the greater the chances of success. At the time of writing, there are no published structures for any *B. graminis* CSEPs and there are no crystal structures available for any close homologs of CSEP0162. One crystal structure exists for a CSEP from *B. graminis f. sp. hordei* (unpublished data). However, there is limited sequence similarity between this potential search model and CSEP0162. Therefore, attempts at a MR solution for the structure of CSEP0162 using this search model in the Phenix Phaser-MR pipeline failed.

## 5.3. Discussion

Our understanding of the roles *B. graminis* CSEPs play in virulence is currently limited but developing, with studies utilising HIGS against putative CSEPs unravelling their importance in pathogen fitness (Ahmed et al., 2015, Ahmed et al., 2016). Some bioinformatically identified CSEPs are predicted to adopt a ribonuclease-like core fold which is hypothesised to be the ancestral state of this gene family (Pedersen et al., 2012). It has even been suggested that the CSEPs diverged from ribonuclease proteins and subsequently lost ribonuclease activity, but retained the core fold. It is currently unknown whether or not all of these predicted ribonuclease-like proteins adopt this fold in reality. Furthermore, only 72 of the >400 CSEPs identified are predicted to adopt this structure. CSEP0162 is not predicted to adopt the ribonuclease-like fold based on its primary amino acid sequence and solving the crystal structure of this protein would provide a unique insight into whether or not sequence divergent CSEP effectors adopt the same core-structure as one another, as a template around which functional diversity can evolve in a manner similar to the WY-fold identified in oomycete RXLR effectors (Boutemy et al., 2011, Win et al., 2012). This may point to a general strategy among plant-associated microbes of adopting a stable, core structural scaffold around which functional diversity can evolve.

We aimed to use protein expressed heterologously in *E. coli* in order to solve the crystal structure of CSEP0162 from *B. graminis f. sp. Hordeii* and to provide novel insights into the type of virulence functions barley powdery mildew CSEPs might perform, as well as into their evolution. One potential bottleneck in protein crystallography studies is the production of sufficient soluble protein from heterologous expression systems. This can be particularly problematic when attempting to express eukaryotic proteins in prokaryotes which do not have the capacity to perform all of the post-translational modifications which may be required for proper folding of eukaryotic proteins. However, soluble expression of native CSEP0162 in *E. coli* under standard growth and induction conditions, using a strain of *E. coli* well suited to expression of disulphide bond containing proteins, proved successful.

To date diffraction data from native CSEP0162 and native CSEP0162 soaked in a solution containing heavy atoms proved insufficient to solve the phase problem and produce an interpretable electron density map despite the presence of a clear anomalous signal from heavy atom-soaked crystals. We, therefore, attempted to incorporate anomalous scatterers into the protein directly – by replacing methionine residues with Selenomethionine. Two methods for incorporation of Selenomethionine into proteins are commonly adopted. Expression of SeMet CSEP0162 by the first method – which uses a methionine auxotrophic strain of *E. coli* - yielded almost no soluble protein, although the protein was expressed. However, the second method permits the use of *E. coli* strains which specialise in the proper folding of disulphide bond containing proteins. Tellingly, when this method was used and expression was induced in Shuffle cells, soluble protein was produced in quantities sufficient for crystallisation, indicating the presence of at least one disulphide bond in CSEP0162 (as predicted) and highlighting the importance of choosing suitable expression hosts in protein purification studies.

Unfortunately, despite the presence of a clear anomalous signal in data collected from SeMet crystals, confirmation of SeMet incorporation from intact mass spectrometry, a fluorescence scan of selenium's K-edge and what appeared to be reasonable data based on summary statistics, SAD phasing again proved unsuccessful in yielding an interpretable electron density map. Automated Xia2 data

processing on the DLS pipeline classified the crystals in the hexagonal space group $P6_2$ 2 2. Although we investigated re-indexing the data to a variety of other space groups including lower symmetry hexagonal space groups, the trigonal space groups (which are related to the hexagonal space groups), C-centred orthorhombic space groups, monoclinic groups and back to the lowest symmetry space group: P1 with no success.

One method for obtaining phase information which was not attempted in the present work is multi-wavelength anomalous diffraction (MAD), which became a possibility with the advent of beamlines using tuneable synchrotron radiation. MAD experiments involve the exposure of a single crystal to multiple incident X-ray wavelengths. Typically, data is collected at three different wavelengths; the K-edge (absorption edge) of the anomalous scatterer (e.g. selenium), at the *f*' maximum (where differences between Bijoevet pairs are greatest) and at a wavelength distant from the K-edge where $\Delta f$ and *f*' are relatively small. Theoretical values for these wavelengths do exist (Cromer and Liberman, 1981) but are unsuitable for the design of MAD experiments as they fail to take into account the context in which the anomalous scatterer exists within a macromolecule. The inaccuracies with these experimental values are most pronounced at wavelengths near the absorption edge of a given element. Wavelengths for a MAD experiment, therefore, must be experimentally determined by a fluorescence scan of the crystal, as in Figure 45. The fluorescence scan performed here should inform future MAD experiments on SeMet derivative CSEP0162 crystals.

Despite data which was processed and indexed with acceptable statistics and the identification of a clear anomalous signal from both KI soaked and SeMet derivative crystals, indicating that a structure solution was possible, it has not been possible to obtain a realistic electron density map for CSEP0162. This may indicate a pathology in the data that is not immediately obvious. Despite a clear preference for the space group $P6_2$ 2 2 by auto-processing software, this space group is not necessarily an accurate description of the symmetry of the crystals in the lattice and it remains a prediction until the crystal structure is solved. When data is re-processed in lower symmetry space groups, twinning fractions of almost 40% are predicted by truncate. Given that a solution is not forthcoming when data are assigned the $P6_2$ 2 2 space

group, it seems likely that one of these lower symmetry space groups are a better description of the crystal and that the data may be twinned. Whilst twinned-data is not always useless, as modern computational methods are able to deconvolute diffraction patterns produced by twinned crystals, the data collected here has not provided a solution for the 3D structure of CSEP0162. Crystallisation at higher temperatures may enhance the risk of twinning in crystals. Growing crystals at a lower temperature may produce un-twinned crystals a solution for the 3D structure of CSEP0162.

Whilst we have been unable to solve the crystal structure of CSEP0162, we have developed reliable methods for expressing, purifying and crystallising both native CSEP0162 and a SeMet derivative. From this, we have been able to gain limited information on the composition of secondary structural elements in CSEP0162 through CD spectroscopy, which reveals a primarily β-stranded protein. Given that we can reliably produce SeMet protein, it should be possible to attempt MAD experiments in the near future if crystals free from pathologies like twinning and OD-twinning can be produced.

# 6

# General Discussion

# Chapter 6 General Discussion

In the post-genomic era, unprecedented levels of data are available with which to investigate the factors governing the outcomes of plant-microbe interactions at the molecular level. Whole genomes – and transcriptomes - of plant pathogens, pests and symbionts can be readily sequenced and assembled more cheaply, and in a more automated fashion, than ever before. Effector proteins can be predicted bioinformatically from these datasets through conserved sequence motifs and characterised through investigating their virulence functions, avirulence function and through assessing the effects on pathogen fitness of knocking them down through techniques such as HIGS or through the exogenous application of siRNA. Predicted effectors validated as *bona fide* effectors which contribute to the virulence activity of the pathogen and/or recognition by host NLRs, can then be functionally characterised. In the case of avirulence function, this may take the form of identifying a cognate NLR(s). In the case of virulence function, this will often take the form of searching for a host target. It is increasingly apparent, again through improved bioinformatics techniques, that the former may inform the latter. Many NLR genes have now been identified which contain non-canonical domains – termed 'integrated domains' – which are thought to be the result of integration of a host target into the NLR in order to act as a decoy, promoting recognition of the effector by NLRs (Kroj et al., 2016, Sarris et al., 2016). Methods for this include pulling down an epitope-tagged construct of the effector expressed *in planta* and analysing – by mass spectrometry – its interactors. Also, large cDNA libraries derived from infected plant tissues can be maintained and used to screen bait constructs derived from an effector of interest.

The process of functional characterisation of *bona fide* effector proteins may be rapidly accelerated by knowledge of their structure. Where clear structural homology to a protein of known function is revealed, mutagenesis experiments to investigate whether or not that homology is artefactual or contributing to real effector function can be designed rapidly. This allows functional characterisation both *in planta* and *in vitro*, as a protein which has been successfully crystallised is by default amenable to heterologous expression and purification to high levels of homogeneity. Advances in

eukaryotic heterologous expression systems will serve to further accelerate this process for effector proteins secreted by eukaryotic pathogens like oomycetes and fungi, as well as the expression and purification of host targets, susceptibility factors and NLRs.

This increasingly well-developed level of understanding of the molecular interplay between plants and the microbes with which they interact has coincided with a revolution in our ability to manipulate genomes. Thanks to targeted genome editing technologies like TALENs, ZFNs and CRISPR/Cas9, we are in a unique position to take the functional knowledge gained through *in planta* and structural studies of plant pathogen effectors and their host targets and rapidly translate this into crop improvement, in order to further safeguard the world's food supply.

The present project aimed to investigate all three themes discussed above; the dissection of an interaction identified between an oomycete effector protein and a host MAPK, the possibility of engineering a well-characterised virulence target of an oomycete effector protein for enhanced resistance and to investigate the 3D structure of an as yet uncharacterised effector protein from a fungal pathogen of barley.

## 6.1. Development of variants of a host target which exhibit insensitivity to its cognate effector in a model system

At the time of writing, efforts to manipulate host-targets of plant pathogenic effector proteins were scarce. This may be due to a reluctance to interfere with what are – in general – proteins which are targeted due to their important roles in immunity, development, hormone signalling, intracellular trafficking and other processes.

We deployed random mutagenesis PCR in order to generate a library of variants of a well characterised effector target and screened them for insensitivity to PexRD2. A small number of these variants proved reproducible and were investigated further. In total, four amino acid substitutions were found to contribute to insensitivity to PexRD2 and protein-protein interaction studies revealed that this is likely due to a weakened interaction between the effector and the insensitive variants of its target.

The observation that all four variants contribute, to some degree, to PexRD2-insensitivity is surprising and may signal the need for further investigation. The result seems improbable given the procedure used for generating the variants and if they are not *bona fide* PexRD2-insensitive variants the most likely explanation is a mutation in the promoter region of the pERCH vector leading to higher levels of kinase expression overcoming the cell death suppression activity of PexRD2, possibly through enhanced response to β-estradiok. Sequencing of the vectors would indicate whether or not there is a mutation in the promoter region.

The strategy used to develop effector-insensitive targets here should be employed by other researchers where there is a well characterised effector-target interaction, and little to no knowledge of the protein's structure with which to inform rationally designed mutations which one would predict to condition insensitivity.

Future experiments on this theme might include *in planta* protein-protein interaction studies – in particular Co-IP, to further confirm the loss of interaction between effector and variant which correlates with insensitivity. In addition to this, developing a reliable strategy for expression of MAPKKK$_\varepsilon$ and the PexRD2-insensitive variants described here would enable actual binding affinities – and any difference between the wild type and the variants – to be accurately determined through *in vitro* techniques such as surface plasmon resonance. Further to this, the ability to express and purify wild type MAPKKK$_\varepsilon$ would be a step toward a long-standing goal of co-crystallising the effector in complex with its target and deriving a 3D structure of the complex. This would provide insights into how the PexRD2 dimer interacts with its target and may help explain why the insensitive mutants identified in this screen interfere with the interaction between PexRD2 and MAPKKK$_\varepsilon$.

One concern may be the ability of an effector – which exists in the context of a rapidly evolving genome – to evolve in such a way that makes the variants of MAPKKK$_\varepsilon$ described here sensitive to its cell death suppression effects once more. One way to tackle this eventuality may be to 'stack' the variants in a manner similar to the strategy employed by those looking to introduce multiple NLRs into crop cultivars. A method to investigate the possibility or PexRD2 overcoming insensitivity could involve an experiment similar to the random mutagenesis

approach used to generate kinase variants. Natural evolutionary processes could be simulated by subjecting PexRD2's effector domain to random mutagenesis, and variants could be screened against the four kinase variants in the transient expression, cell death suppression assays described in the present thesis – any PexRD2 variants identified as able to re-establish cell death suppression activity may be indicative that any resistance derived from these MAPKKK$_{\varepsilon}$ mutants may be relatively easy to overcome.

It is worth noting that of the integrated domains in NLRs mentioned earlier (Kroj et al., 2016, Sarris et al., 2016), domains encoding putative kinase domain are the best represented. As well as pointing to kinase signalling cascades as a highly conserved, general target of plant pathogen effector proteins, this also opens up the possibility that gaining a deep understanding of the interaction between kinases and effectors will also inform our understanding of recognition specificity by NLRs. Such understanding could, in the future, lead to the development of a synthetic, universal kinase integrated domain capable of conditioning immunity to a wide range of pathogens, in multiple hosts.

## 6.2. Deploying the CRISPR/Cas9 system in tomato

Few examples of CRISPR/Cas9 genome editing for disease resistance in tomato existed at the time of writing as most efforts have, thus far, been directed at establishing the technique in model systems such as *A. thaliana* and *N. benthamiana*. Here, we used a transient method with which to confirm the activity of sgRNAs prior to stable transformation using a PCR-RE assay.

Unfortunately, stable transformation of tomato plants with our confirmed active sgRNAs and Cas9 did not result in the generation of any plants homozygous, or even heterozygous, for deletion of endogenous MAPKKK$_{\varepsilon}$, as determined by PCR-RE assay. Screening of plants for smaller INDELs suggested the presence of a small number, although the only INDEL detected homozygously resulted in deletion of one amino acid and no frameshift in the rest of the protein resulting in aberrant termination of translation. We could detect the deletion in EcoRV treated gDNA samples subjected to PCR using primers bounding the targeted genomic region, but a

strong band corresponding to endogenous MAPKKK$_\varepsilon$ gDNA was present in all samples tested, suggesting that induction of large deletions was a rare event.

Nevertheless, we were able to sequence the large deletion, through cloning, in order to determine its exact nature and revealed two cuts precisely 3bp upstream of the sgRNA2 PAM site and 4bp upstream of the sgRNA4 PAM site. Unfortunately, due to low yields from PCR we could not sequence the deletion directly from gel extractions in order to determine its zygosity in the tissue sampled.

CRISPR/Cas9 mediated genome editing is clearly a technology of the future, however it is still in its infancy. The present work demonstrates, to the author's knowledge, the first instance of CRISPR/Cas9 mediated knockout of an endogenous gene with simultaneous complementation of a transgene. This should demonstrate that additional genes can be studied, alongside any gene presented as a HDR cassette in CRISPR/Cas9 experiments.

## 6.3. Transgene expression indicates that one insensitive kinase variant may contribute to enhanced resistance to *P. infestans*

Despite a failure to knock out endogenous MAPKKK$_\varepsilon$, primary transformants were assayed for enhanced resistance to two strains of *P. infestans*. The Q50R mutation appeared to contribute to significantly enhanced resistance to two strains of *P. infestans*. However, given that we have failed to knock out endogenous MAPKKK$_\varepsilon$ it is impossible to rule out the possibility that enhanced resistance is due to a dosage effect related to greater than normal expression levels of MAPKKK$_\varepsilon$, a known positive regulator of immunity. Given that V2 (Q50R) did exhibit some developmental abnormalities relative to other transformants, which may be associated with aberrant expression levels of MAPKKK$_\varepsilon$, it would be wise to remain cautious until the $T_1$ generation is investigated and, hopefully, homozygous MAPKKK$_\varepsilon$ knockouts are obtained. An alternative possibility explaining developmental abnormalities is that T-DNA integration, in certain plants, has

occurred in a genomic context which has disrupted some gene important for normal development. However, this does not account for enhanced resistance to *P. infestans*.

V3 and V4 (S200T and D241N, respectively) appeared to condition enhanced susceptibility when stably expressed in *S. lycopersicum*. This is a confusing result as there is no obvious, logical basis for this phenotype.

Taken together, we feel that the work presented on manipulation of MAPKKK$_{\varepsilon}$ in the present thesis provides a valuable proof of concept that host targets can certainly be manipulated for insensitivity to their cognate effector proteins. We also tentatively suggest that this can confer enhanced resistance *in planta*.

## 6.4. STE20-Like MAPKKKK is not a virulence target of PexRD2

Potato MAPKKKK was investigated as a potential virulence target of PexRD2 through a series of overexpression assays in the model host *N. benthamiana*. This revealed that it is unlikely to be a virulence target as overexpression does not reduce pathogen fitness, a hallmark of virulence targets. Although it remains a possibility that it is, indeed, a positive regulator of immunity to *P. infestans* and a likely virulence target but that it produces a systemic effect which obfuscated any detection of suppression of *P. infestans* virulence in our assays.

Investigations into whether or not it may function as a helper protein of PexRD2 remain inconclusive although pathogenicity assays suggest that it is. The discovery of a novel helper of an effector protein from *P. infestans* would be of particular interest as this has not been reported previously. The importance of helper proteins in plant-microbe interactions for virulence – or avirulence – outcomes is becoming increasingly well-understood (Win J, 2012) and serves to highlight both the sophistication of microbial infection strategies as well as the internal surveillance mechanisms of plants.

## 6.5. Expression of CSEP0162 yields protein crystals which diffract X-ray radiation but 3D structure remains unsolved

A reliable system for producing SeMet-tagged CSEP0162 was established which should serve to produce data for solving the 3D structure of CSEP0162 by MAD techniques, as SAD, IR and MR phasing techniques have, so far, failed to yield an interpretable electron density map suitable for refinement to a high resolution 3D structure.

It is hoped that understanding the 3D structure of CSEP0162 will provide deeper insights into not only the function of CSEP0162 itself, but into the evolution of the CSEPs themselves.

Of particular interest in the emerging *B. graminis* CSEP story is the functional significance of the YxC motif. Our expressed and crystallised construct contains the YxC motif, so a 3D structure of this protein should provide some clue as to its function dependent on where it is presented on, or inside any fold/pocket of, the protein.

## 6.6. Outlook

The future of genetic manipulation of crops in order to generate enhanced disease resistance looks promising. With better understanding of the interactions, at a molecular level, between microbes and their hosts and greater control than ever before over the genomes of organisms, we've never been in a better place to safeguard food security for a rapidly expanding population. Indeed, the present work demonstrates the first instance of simultaneous CRISPR/Cas9-mediated gene knockout and expression of an additional transgene. This, in and of itself, should prove useful to the community in addition to the preliminary results suggesting that the effector-insensitive variants in the present thesis may confer some resistance to *P. infestans*.

We have attempted to make strides into gaining fundamental insights into how a host MAP Kinase may act as a helper of a *P. infestans* secreted effector protein. Whilst this remains very much an open question, the present thesis has opened up several avenues for study. Efforts should be made into ascertaining, for certain, whether or not MAPKKKK phosphorylates PexRD2 and, if so, where phosphorylation occurs. Further to this, there remains a possibility that MAPKKKK does indeed still function as a positive regulator of immunity, making it a genuine candidate for a *bona fide* virulence target of the effector. It seems possible that a systemic effect of the kinases immune signalling may have prevented detection of an effect in side-by-side leaf pathogenesis assays. Further experiments are suggested which should shed light on this.

Progress has been made into solving the 3D structure of a candidate secreted effector protein (with confirmed importance for pathogen fitness) from *B. graminis*.

This work also offers proof of concept that effector-insensitive variants of effector targets can be generated using a random mutagenesis approach and will hopefully encourage further work on other well-characterised effector targets. We also tentatively suggest that one of these variants can confer enhanced resistance to the pathogen, although clearly further work is needed to confirm this. Whilst we have not fully solved the questions we posed, the experiments required to answer these questions are now clear and well-defined based on the work done in the present thesis.

References

ABRAMOVITCH, R. B., JANJUSEVIC, R., STEBBINS, C. E. & MARTIN, G. B. 2006. Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 2851-2856.

ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J. & GROSSE-KUNSTLEVE, R. W. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 213-221.

ADE, J., DEYOUNG, B. J., GOLSTEIN, C. & 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 Sciences,* 104**,** 2531-2536.

AHMED, A. A., PEDERSEN, C., SCHULTZ-LARSEN, T., KWAAITAAL, M., JØRGENSEN, H. J. L. & THORDAL-CHRISTENSEN, H. 2015. The Barley Powdery Mildew Candidate Secreted Effector Protein CSEP0105 Inhibits the Chaperone Activity of a Small Heat Shock Protein. *Plant Physiology,* 168**,** 321-333.

AHMED, A. A., PEDERSEN, C. & THORDAL-CHRISTENSEN, H. 2016. The Barley Powdery Mildew Effector Candidates CSEP0081 and CSEP0254 Promote Fungal Infection Success. *PLoS ONE,* 11**,** e0157586.

ANTONY, G., ZHOU, J., HUANG, S., LI, T., LIU, B., WHITE, F. & YANG, B. 2010. Rice xa13 Recessive Resistance to Bacterial Blight Is Defeated by Induction of the Disease Susceptibility Gene Os-11N3. *The Plant Cell,* 22**,** 3864-3876.

ARMSTRONG, M. R., WHISSON, S. C., PRITCHARD, L., BOS, J. I. B., VENTER, E., AVROVA, A. O., REHMANY, A. P., BÖHME, U., BROOKS, K., CHEREVACH, I., HAMLIN, N., WHITE, B., FRASER, A., LORD, A., QUAIL, M. A., CHURCHER, C., HALL, N., BERRIMAN, M., HUANG, S., KAMOUN, S., BEYNON, J. L. & BIRCH, P. R. J. 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 Sciences of the United States of America,* 102**,** 7766-7771.

AXTELL, M. J. & STASKAWICZ, B. J. 2003. Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell,* 112**,** 369-77.

AYLOR, D. E., FRY, W. E., MAYTON, H. & ANDRADE-PIEDRA, J. L. 2001. Quantifying the Rate of Release and Escape of Phytophthora infestans Sporangia from a Potato Canopy. *Phytopathology,* 91**,** 1189-1196.

BALDAUF, S. L., ROGER, A. J., WENK-SIEFERT, I. & DOOLITTLE, W. F. 2000. A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data. *Science,* 290**,** 972-977.

BANNO, H., HIRANO, K., NAKAMURA, T., IRIE, K., NOMOTO, S., MATSUMOTO, K. & MACHIDA, Y. 1993. NPK1, a tobacco gene that encodes a protein with a domain homologous to yeast BCK1, STE11, and Byr2 protein kinases. *Molecular and Cellular Biology,* 13**,** 4745-4752.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BEAKES, G., GLOCKLING, S. & SEKIMOTO, S. 2012. The evolutionary phylogeny of the oomycete "fungi". *Protoplasma,* 249**,** 3-19.

BERROW, N. S., ALDERTON, D., SAINSBURY, S., NETTLESHIP, J., ASSENBERG, R., RAHMAN, N., STUART, D. I. & OWENS, R. J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research,* 35**,** e45-e45.

BERTANI, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol,* 62**,** 293-300.

BHATTACHARJEE, S., HILLER, N. L., LIOLIOS, K., WIN, J., KANNEGANTI, T.-D., YOUNG, C., KAMOUN, S. & HALDAR, K. 2006. The Malarial Host-Targeting Signal Is Conserved in the Irish Potato Famine Pathogen. *PLoS Pathog,* 2**,** e50.

BHATTACHARJEE, S., STAHELIN, R. V., SPEICHER, K. D., SPEICHER, D. W. & HALDAR, K. 2012. Endoplasmic reticulum PI(3)P lipid binding targets malaria proteins to the host cell. *Cell,* 148**,** 201-12.

BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

BIRNBOIM, H. C. & DOLY, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res,* 7**,** 1513-23.

BJERRUM, O. J., SELMER, J. C. & LIHME, A. 1987. Native immunoblotting: Transfer of membrane proteins in the presence of non-ionic detergent. *ELECTROPHORESIS,* 8**,** 388-397.

BLONDEAU, K., BLAISE, F., GRAILLE, M., KALE, S. D., LINGLIN, J., OLLIVIER, B., LABARDE, A., LAZAR, N., DAVERDIN, G., BALESDENT, M.-H., CHOI, D. H. Y., TYLER, B. M., ROUXEL, T., VAN TILBEURGH, H. & FUDAL, I. 2015. Crystal structure of the effector AvrLm4–7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. *The Plant Journal,* 83**,** 610-624.

BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science,* 326**,** 1509-1512.

BOEVINK, P. C., WANG, X., MCLELLAN, H., HE, Q., NAQVI, S., ARMSTRONG, M. R., ZHANG, W., HEIN, I., GILROY, E. M., TIAN, Z. & BIRCH, P. R. 2016. A Phytophthora infestans RXLR effector targets plant PP1c isoforms that promote late blight disease. *Nat Commun,* 7**,** 10311.

BÖHM, H., ALBERT, I., OOME, S., RAAYMAKERS, T. M., VAN DEN ACKERVEKEN, G. & NÜRNBERGER, T. 2014. A Conserved Peptide Pattern from a Widespread Microbial Virulence Factor Triggers Pattern-Induced Immunity in <italic>Arabidopsis</italic>. *PLoS Pathog,* 10**,** e1004491.

BOLOTIN, A., QUINQUIS, B., RENAULT, P., SOROKIN, A., EHRLICH, S. D., KULAKAUSKAS, S., LAPIDUS, A., GOLTSMAN, E., MAZUR, M., PUSCH, G. D., FONSTEIN, M., OVERBEEK, R., KYPRIDES, N., PURNELLE, B., PROZZI, D., NGUI, K., MASUY, D., HANCY, F., BURTEAU, S., BOUTRY, M., DELCOUR, J., GOFFEAU, A. & HOLS, P. 2004. Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nat Biotech,* 22**,** 1554-1558.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-61.

BOLTON, M. D., VAN ESSE, H. P., VOSSEN, J. H., DE JONGE, R., STERGIOPOULOS, I., STULEMEIJER, I. J., VAN DEN BERG, G. C., BORRAS-HIDALGO, O., DEKKER, H. L., DE KOSTER, C. G., DE WIT, P. J., JOOSTEN, M. H. & THOMMA, B. P. 2008. The novel Cladosporium fulvum lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Mol Microbiol,* 69**,** 119-36.

BOS, J. I., KANNEGANTI, T. D., YOUNG, C., CAKIR, C., HUITEMA, E., WIN, J., ARMSTRONG, M. R., BIRCH, P. R. & 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. *The Plant Journal,* 48**,** 165-176.

BOS, J. I. B., ARMSTRONG, M. R., GILROY, E. M., BOEVINK, P. C., HEIN, I., TAYLOR, R. M., TIAN, Z. D., ENGELHARDT, S., VETUKURI, R. R., HARROWER, B., DIXELIUS, C., BRYAN, G., SADANANDOM, A., WHISSON, S. C., KAMOUN, S. & BIRCH, P. R. J. 2010. Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 9909-9914.

BOS, J. I. B., CHAPARRO-GARCIA, A., QUESADA-OCAMPO, L. M., GARDENER, B. B. M. & KAMOUN, S. 2009. Distinct Amino Acids of the Phytophthora infestans Effector AVR3a Condition Activation of R3a Hypersensitivity and Suppression of Cell Death. *Molecular Plant-Microbe Interactions,* 22**,** 269-281.

BOTH, M., CSUKAI, M., STUMPF, M. P. H. & SPANU, P. D. 2005. Gene Expression Profiles of Blumeria graminis Indicate Dynamic Changes to Primary Metabolism during Development of an Obligate Biotrophic Pathogen. *The Plant Cell,* 17**,** 2107-2122.

BOUTEMY, L. S., KING, S. R., WIN, J., HUGHES, R. K., CLARKE, T. A., BLUMENSCHEIN, T. M., KAMOUN, S. & BANFIELD, M. J. 2011. Structures of Phytophthora RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J Biol Chem,* 286**,** 35834-42.

BOUWMEESTER, K., DE SAIN, M., WEIDE, R., GOUGET, A., KLAMER, S., CANUT, H. & GOVERS, F. 2011. The Lectin Receptor Kinase LecRK-I.9 Is a Novel <italic>Phytophthora</italic> Resistance Component and a Potential Host Target for a RXLR Effector. *PLoS Pathog,* 7**,** e1001327.

BOZKURT, T. O., SCHORNACK, S., WIN, J., SHINDO, T., ILYAS, M., OLIVA, R., CANO, L. M., JONES, A. M., HUITEMA, E., VAN DER HOORN, R. A. & KAMOUN, S. 2011. Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc Natl Acad Sci U S A,* 108**,** 20832-7.

BRAGG, L., HOWELLS, E. & PERUTZ, M. The structure of haemoglobin. II. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1954. The Royal Society, 33-44.

BROWN, J. L., STOWERS, L., BAER, M., TREJO, J., COUGHLIN, S. & CHANT, J. 1996. Human Ste20 homologue hPAK1 links GTPases to the JNK MAP kinase pathway. *Curr Biol,* 6**,** 598-605.

CATANZARITI, A.-M., DODDS, P. N., LAWRENCE, G. J., AYLIFFE, M. A. & ELLIS, J. G. 2006. Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. *The Plant Cell,* 18**,** 243-256.

CATANZARITI, A.-M., DODDS, P. N., VE, T., KOBE, B., ELLIS, J. G. & STASKAWICZ, B. J. 2010. The AvrM Effector from Flax Rust Has a Structured C-Terminal Domain and Interacts Directly with the M Resistance Protein. *Molecular plant-microbe interactions : MPMI,* 23**,** 49-57.

CAVALIER-SMITH, T. & CHAO, E. E. Y. 2006. Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *Journal of Molecular Evolution,* 62**,** 388-420.

CESARI, S., BERNOUX, M., MONCUQUET, P., KROJ, T. & DODDS, P. N. 2014. A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. *Frontiers in Plant Science,* 5.

CESARI, S., THILLIEZ, G., RIBOT, C., CHALVON, V., MICHEL, C., JAUNEAU, A., RIVAS, S., ALAUX, L., KANZAKI, H., OKUYAMA, Y., MOREL, J.-B., FOURNIER, E., THARREAU, D., TERAUCHI, R. & KROJ, T. 2013. The Rice Resistance Protein Pair RGA4/RGA5 Recognizes the Magnaporthe oryzae Effectors AVR-Pia and AVR1-CO39 by Direct Binding. *The Plant Cell Online*.

CHEN, H., KOVALCHUK, A., KERIO, S. & ASIEGBU, F. O. 2013. Distribution and bioinformatic analysis of the cerato-platanin protein family in Dikarya. *Mycologia,* 105**,** 1479-88.

CHEN, W., YAZICIOGLU, M. & COBB, M. H. 2004. Characterization of OSR1, a member of the mammalian Ste20p/germinal center kinase subfamily. *J Biol Chem,* 279**,** 11129-36.

CHENG, W., MUNKVOLD, KATHY R., GAO, H., MATHIEU, J., SCHWIZER, S., WANG, S., YAN, Y.-B., WANG, J., MARTIN, GREGORY B. & CHAI, J. 2011. Structural Analysis of Pseudomonas syringae AvrPtoB Bound to Host BAK1 Reveals Two Similar Kinase-Interacting Domains in a Type III Effector. *Cell Host & Microbe,* 10**,** 616-626.

CHOWDAPPA, P., KUMAR, N. B. J., MADHURA, S., KUMAR, M. S. P., MYERS, K. L., FRY, W. E., SQUIRES, J. N. & COOKE, D. E. L. 2013. Emergence of 13_A2 Blue Lineage of Phytophthora infestans was Responsible for Severe Outbreaks of Late Blight on Tomato in South-West India. *Journal of Phytopathology,* 161**,** 49-58.

CHYLINSKI, K., MAKAROVA, K. S., CHARPENTIER, E. & KOONIN, E. V. 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research,* 42**,** 6091-6105.

CLARKE, C. R., CHINCHILLA, D., HIND, S. R., TAGUCHI, F., MIKI, R., ICHINOSE, Y., MARTIN, G. B., LEMAN, S., FELIX, G. & VINATZER, B. A. 2013. Allelic variation in two distinct Pseudomonas syringae flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. *New Phytologist,* 200**,** 847-860.

COLCOMBET, J. & HIRT, H. 2008. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochemical Journal,* 413**,** 217-226.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339**,** 819-823.

COOK, D. E., MESARICH, C. H. & THOMMA, B. P. 2015. Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol,* 53**,** 541-63.

COOKE, D. E. L., CANO, L. M., RAFFAELE, S., BAIN, R. A., COOKE, L. R., ETHERINGTON, G. J., DEAHL, K. L., FARRER, R. A., GILROY, E. M., GOSS, E. M., GRÜNWALD, N. J., HEIN, I., MACLEAN, D., MCNICOL, J. W., RANDALL, E., OLIVA, R. F., PEL, M. A., SHAW, D. S., SQUIRES, J. N., TAYLOR, M. C., VLEESHOUWERS, V. G. A. A., BIRCH, P. R. J., LEES, A. K. & KAMOUN, S. 2012. Genome Analyses of an Aggressive and Invasive Lineage of the Irish Potato Famine Pathogen. *PLoS Pathog,* 8**,** e1002940.

CROMER, D. T. & LIBERMAN, D. A. 1981. Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* 37**,** 267-268.

CUI, H., WANG, Y., XUE, L., CHU, J., YAN, C., FU, J., CHEN, M., INNES, R. W. & ZHOU, J. M. 2010. Pseudomonas syringae effector protein AvrB perturbs Arabidopsis hormone signaling by activating MAP kinase 4. *Cell Host Microbe,* 7**,** 164-75.

CURTIN, S. J., ZHANG, F., SANDER, J. D., HAUN, W. J., STARKER, C., BALTES, N. J., REYON, D., DAHLBORG, E. J., GOODWIN, M. J., COFFMAN, A. P., DOBBS, D., JOUNG, J. K., VOYTAS, D. F. & STUPAR, R. M. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol,* 156**,** 466-73.

DAGDAS, Y. F., BELHAJ, K., MAQBOOL, A., CHAPARRO-GARCIA, A., PANDEY, P., PETRE, B., TABASSUM, N., CRUZ-MIRELES, N., HUGHES, R. K., SKLENAR, J., WIN, J., MENKE, F., FINDLAY, K., BANFIELD, M. J., KAMOUN, S. & BOZKURT, T. O. 2016. An effector of

the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. *eLife,* 5**,** e10856.

DALE, D. C., BOXER, L. & LILES, W. C. 2008. *The phagocytes: neutrophils and monocytes*.

DAN, I., WATANABE, N. M. & KUSUMI, A. 2001. The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol,* 11**,** 220-30.

DE JONGE, R., VAN ESSE, H. P., KOMBRINK, A., SHINYA, T., DESAKI, Y., BOURS, R., VAN DER KROL, S., SHIBUYA, N., JOOSTEN, M. H. & THOMMA, B. P. 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science,* 329**,** 953-5.

DELAROSA, S., GUILLEMETTE, J., PAPILLON, J., HAN, Y. S., KRISTOF, A. S. & CYBULSKY, A. V. 2011. Activity of the Ste20-like kinase, SLK, is enhanced by homodimerization. *Am J Physiol Renal Physiol,* 301**,** F554-64.

DELPIRE, E. & GAGNON, K. B. 2008. SPAK and OSR1: STE20 kinases involved in the regulation of ion homoeostasis and volume control in mammalian cells. *Biochem J,* 409**,** 321-31.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-607.

DEYOUNG, B. J., QI, D., KIM, S. H., BURKE, T. P. & INNES, R. W. 2012. Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cell Microbiol,* 14**,** 1071-84.

DODDS, P. N. & RATHJEN, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet,* 11**,** 539-48.

DOU, D. L., KALE, S. D., WANG, X., JIANG, R. H. Y., BRUCE, N. A., ARREDONDO, F. D., ZHANG, X. M. & TYLER, B. M. 2008. RXLR-mediated entry of Phytophthora sojae effector Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell,* 20**,** 1930-1947.

DU, Y., MPINA, M. H., BIRCH, P. R., BOUWMEESTER, K. & GOVERS, F. 2015. Phytophthora infestans RXLR Effector AVR1 Interacts with Exocyst Component Sec5 to Manipulate Plant Immunity. *Plant Physiol,* 169**,** 1975-90.

EVANS, G. & PETTIFER, R. F. 2001. CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of applied crystallography,* 34**,** 82-86.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 72-82.

FISHER, M. C., HENK, D. A., BRIGGS, C. J., BROWNSTEIN, J. S., MADOFF, L. C., MCCRAW, S. L. & GURR, S. J. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature,* 484**,** 10.1038/nature10947.

FLOR, H. H. 1942. Inheritance of pathogenicity in Melampsora lini. *Phytopathology,* 32**,** 653-669.

FLOR, H. H. 1947. Inheritance of reaction to rust in flax. *J. Agric. Res. ,* 74**,** 241-262.

FLOR, H. H. 1955. Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology,* 45**,** 680-685.

FOSTER, S. J., PARK, T.-H., PEL, M., BRIGNETI, G., ŚLIWKA, J., JAGGER, L., VAN DER VOSSEN, E. & JONES, J. D. G. 2009. Rpi-vnt1.1, a Tm-22 Homolog from Solanum venturii, Confers Resistance to Potato Late Blight. *Molecular Plant-Microbe Interactions,* 22**,** 589-600.

FURUKAWA, T., INAGAKI, H., TAKAI, R., HIRAI, H. & CHE, F. S. 2014. Two distinct EF-Tu epitopes induce immune responses in rice and Arabidopsis. *Mol Plant Microbe Interact,* 27**,** 113-24.

GARNEAU, J. E., DUPUIS, M. E., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67-71.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences,* 109**,** E2579–E2586.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S. E., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In:* WALKER, J. M. (ed.) *The Proteomics Protocols Handbook.* Totowa, NJ: Humana Press.

GIANNAKOPOULOU, A., STEELE, J. F. C., SEGRETIN, M. E., BOZKURT, T., ZHOU, J., ROBATZEK, S., BANFIELD, M. J., PAIS, M. & KAMOUN, S. 2015. Tomato I2 immune receptor can be engineered to confer partial resistance to the oomycete Phytophthora infestans in addition to the fungus Fusarium oxysporum. *Molecular Plant-Microbe Interactions*.

GOBLER, C. J., BERRY, D. L., DYHRMAN, S. T., WILHELM, S. W., SALAMOV, A., LOBANOV, A. V., ZHANG, Y., COLLIER, J. L., WURCH, L. L., KUSTKA, A. B., DILL, B. D., SHAH, M., VERBERKMOES, N. C., KUO, A., TERRY, A., PANGILINAN, J., LINDQUIST, E. A., LUCAS, S., PAULSEN, I. T., HATTENRATH-LEHMANN, T. K., TALMAGE, S. C., WALKER, E. A., KOCH, F., BURSON, A. M., MARCOVAL, M. A., TANG, Y.-Z., LECLEIR, G. R., COYNE, K. J., BERG, G. M., BERTRAND, E. M., SAITO, M. A., GLADYSHEV, V. N. & GRIGORIEV, I. V. 2011. Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. *Proceedings of the National Academy of Sciences,* 108**,** 4352-4357.

GODFREY, D., BOHLENIUS, H., PEDERSON, C., ZHANG, Z., EMMERSON, J. & THORDAL-CHRISTENSEN, H. 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC motif. *BMC Genomics,* 11.

GREENBERG, J. T. & AUSUBEL, F. M. 1993. Arabidopsis mutants compromised for the control of cellular damage during pathogenesis and aging. *Plant J,* 4**,** 327-41.

GRISSA, I., VERGNAUD, G. & POURCEL, C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics,* 8**,** 172-172.

HAAS, B. J., KAMOUN, S., ZODY, M. C., JIANG, R. H. Y., HANDSAKER, R. E., CANO, L. M., GRABHERR, M., KODIRA, C. D., RAFFAELE, S., TORTO-ALALIBO, T., BOZKURT, T. O., AH-FONG, A. M. V., ALVARADO, L., ANDERSON, V. L., ARMSTRONG, M. R., AVROVA, A., BAXTER, L., BEYNON, J., BOEVINK, P. C., BOLLMANN, S. R., BOS, J. I. B., BULONE, V., CAI, G. H., CAKIR, C., CARRINGTON, J. C., CHAWNER, M., CONTI, L., COSTANZO, S., EWAN, R., FAHLGREN, N., FISCHBACH, M. A., FUGELSTAD, J., GILROY, E. M., GNERRE, S., GREEN, P. J., GRENVILLE-BRIGGS, L. J., GRIFFITH, J., GRUNWALD, N. J., HORN, K., HORNER, N. R., HU, C. H., HUITEMA, E., JEONG, D. H., JONES, A. M. E., JONES, J. D. G., JONES, R. W., KARLSSON, E. K., KUNJETI, S. G., LAMOUR, K., LIU, Z. Y., MA, L. J., MACLEAN, D., CHIBUCOS, M. C., MCDONALD, H., MCWALTERS, J., MEIJER, H. J. G., MORGAN, W., MORRIS, P. F., MUNRO, C. A., O'NEILL, K., OSPINA-GIRALDO, M., PINZON, A., PRITCHARD, L., RAMSAHOYE, B., REN, Q. H., RESTREPO, S., ROY, S., SADANANDOM, A., SAVIDOR, A., SCHORNACK, S., SCHWARTZ, D. C., SCHUMANN, U. D., SCHWESSINGER, B., SEYER, L., SHARPE, T., SILVAR, C., SONG, J., STUDHOLME, D. J., SYKES, S., THINES, M., VAN DE VONDERVOORT, P. J. I., PHUNTUMART, V., WAWRA, S., WEIDE, R., WIN, J., YOUNG, C., ZHOU, S. G., FRY, W., MEYERS, B. C., VAN WEST, P., RISTAINO, J., GOVERS, F., BIRCH, P. R. J., WHISSON, S. C., JUDELSON, H. S. & NUSBAUM, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature,* 461**,** 393-398.

HANAHAN, D. 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology,* 166**,** 557-580.

HARPER, J. T., WAANDERS, E. & KEELING, P. J. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *International Journal of Systematic and Evolutionary Microbiology,* 55**,** 487-496.

HAVERKORT, A. J., BOONEKAMP, P. M., HUTTEN, R., JACOBSEN, E., LOTZ, L. A. P., KESSEL, G. J. T., VISSER, R. G. F. & VOSSEN, E. A. G. 2008. Societal Costs of Late Blight in Potato and Prospects of Durable Resistance Through Cisgenic Modification. *Potato Research,* 51**,** 47-57.

HERRMANN, C., HORN, G., SPAARGAREN, M. & WITTINGHOFER, A. 1996. Differential interaction of the ras family GTP-binding proteins H-Ras, Rap1A, and R-Ras with the putative effector molecules Raf kinase and Ral-guanine nucleotide exchange factor. *Journal of Biological Chemistry,* 271**,** 6794-6800.

HIRT, H. 2000. Connecting oxidative stress, auxin, and cell cycle regulation through a plant mitogen-activated protein kinase pathway. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 2405-2407.

HOMAN, K. T., WALDSCHMIDT, H. V., GLUKHOVA, A., CANNAVO, A., SONG, J., CHEUNG, J. Y., KOCH, W. J., LARSEN, S. D. & TESMER, J. J. G. 2015. Crystal Structure of G Protein-coupled Receptor Kinase 5 in Complex with a Rationally Designed Inhibitor. *Journal of Biological Chemistry,* 290**,** 20649-20659.

HOUSE, C. & KEMP, B. E. 1987. Protein kinase C contains a pseudosubstrate prototope in its regulatory domain. *Science,* 238**,** 1726-1728.

HUBER, A. H., NELSON, W. J. & WEIS, W. I. 1997. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell,* 90**,** 871-82.

IANNACONE, M., SITIA, G. & GUIDOTTI, L. G. 2006. Pathogenetic and antiviral immune responses against hepatitis B virus. *Future Virology,* 1**,** 189-196.

ICHIMURA, K., SHINOZAKI, K., TENA, G., SHEEN, J., HENRY, Y., CHAMPION, A., KREIS, M., ZHANG, S., HIRT, H. & WILSON, C. 2002. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in plant science,* 7**,** 301-308.

JAUCH, R., JÄKEL, S., NETTER, C., SCHREITER, K., AICHER, B., JÄCKLE, H. & WAHL, M. C. 2005. Crystal Structures of the Mnk2 Kinase Domain Reveal an Inhibitory Conformation and a Zinc Binding Site. *Structure,* 13**,** 1559-1568.

JIA, Y., MCADAMS, S. A., BRYAN, G. T., HERSHEY, H. P. & VALENT, B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO Journal,* 19**,** 4004.

JIANG, R. H., TYLER, B. M. & GOVERS, F. 2006. Comparative analysis of Phytophthora genes encoding secreted proteins reveals conserved synteny and lineage-specific gene duplications and deletions. *Mol Plant Microbe Interact,* 19**,** 1311-21.

JIN, H., AXTELL, M. J., DAHLBECK, D., EKWENNA, O., ZHANG, S., STASKAWICZ, B. & BAKER, B. 2002. NPK1, an MEKK1-like mitogen-activated protein kinase kinase kinase, regulates innate immunity and development in plants. *Dev Cell,* 3**,** 291-7.

JIN, S. G., KOMARI, T., GORDON, M. P. & NESTER, E. W. 1987. Genes responsible for the supervirulence phenotype of Agrobacterium tumefaciens A281. *J Bacteriol,* 169**,** 4417-25.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JONES, J. D. G. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.

KABSCH, W. 2010. Xds. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 125-132.

KALE, S. D., GU, B. A., CAPELLUTO, D. G. S., DOU, D. L., FELDMAN, E., RUMORE, A., ARREDONDO, F. D., HANLON, R., FUDAL, I., ROUXEL, T., LAWRENCE, C. B., SHAN, W. X. & TYLER, B. M. 2010. External Lipid PI3P Mediates Entry of Eukaryotic Pathogen Effectors into Plant and Animal Host Cells. *Cell,* 142**,** 284-295.

KAMOUN, S., VAN WEST, P., VLEESHOUWERS, V. G. A. A., DE GROOT, K. E. & GOVERS, F. 1998. Resistance of Nicotiana benthamiana to Phytophthora infestans Is Mediated by the Recognition of the Elicitor Protein INF1. *The Plant Cell Online,* 10**,** 1413-1425.

KAMPER, J., KAHMANN, R., BOLKER, M., MA, L. J., BREFORT, T., SAVILLE, B. J., BANUETT, F., KRONSTAD, J. W., GOLD, S. E., MULLER, O., PERLIN, M. H., WOSTEN, H. A., DE VRIES, R., RUIZ-HERRERA, J., REYNAGA-PENA, C. G., SNETSELAAR, K., MCCANN, M., PEREZ-MARTIN, J., FELDBRUGGE, M., BASSE, C. W., STEINBERG, G., IBEAS, J. I., HOLLOMAN, W., GUZMAN, P., FARMAN, M., STAJICH, J. E., SENTANDREU, R., GONZALEZ-PRIETO, J. M., KENNELL, J. C., MOLINA, L., SCHIRAWSKI, J.,

MENDOZA-MENDOZA, A., GREILINGER, D., MUNCH, K., ROSSEL, N., SCHERER, M., VRANES, M., LADENDORF, O., VINCON, V., FUCHS, U., SANDROCK, B., MENG, S., HO, E. C., CAHILL, M. J., BOYCE, K. J., KLOSE, J., KLOSTERMAN, S. J., DEELSTRA, H. J., ORTIZ-CASTELLANOS, L., LI, W., SANCHEZ-ALONSO, P., SCHREIER, P. H., HAUSER-HAHN, I., VAUPEL, M., KOOPMANN, E., FRIEDRICH, G., VOSS, H., SCHLUTER, T., MARGOLIS, J., PLATT, D., SWIMMER, C., GNIRKE, A., CHEN, F., VYSOTSKAIA, V., MANNHAUPT, G., GULDENER, U., MUNSTERKOTTER, M., HAASE, D., OESTERHELD, M., MEWES, H. W., MAUCELI, E. W., DECAPRIO, D., WADE, C. M., BUTLER, J., YOUNG, S., JAFFE, D. B., CALVO, S., NUSBAUM, C., GALAGAN, J. & BIRREN, B. W. 2006. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. *Nature,* 444**,** 97-101.

KANNAN, N. & NEUWALD, A. F. 2005. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J Mol Biol,* 351**,** 956-72.

KANZAKI, H., YOSHIDA, K., SAITOH, H., FUJISAKI, K., HIRABUCHI, A., ALAUX, L., FOURNIER, E., THARREAU, D. & TERAUCHI, R. 2012. Arms race co-evolution of Magnaporthe oryzae AVR-Pik and rice Pik genes driven by their physical interactions. *The Plant Journal,* 72**,** 894-907.

KAPILA, J., DE RYCKE, R., VAN MONTAGU, M. & ANGENON, G. 1997. An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant science,* 122**,** 101-108.

KASCHANI, F., SHABAB, M., BOZKURT, T., SHINDO, T., SCHORNACK, S., GU, C., ILYAS, M., WIN, J., KAMOUN, S. & VAN DER HOORN, R. A. 2010. An effector-targeted protease contributes to defense against Phytophthora infestans and is under diversifying selection in natural hosts. *Plant physiology,* 154**,** 1794-1804.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols,* 10**,** 845-858.

KIM, S. H., QI, D., ASHFIELD, T., HELM, M. & INNES, R. W. 2016. Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science,* 351**,** 684-687.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 1156-1160.

KING, S. R. F. 2013. *STRUCTURE-FUNCTION STUDIES OF TRANSLOCATED EFFECTORS FROM THE LATE BLIGHT PATHOGEN.* PhD, University of East Anglia.

KING, S. R. F., MCLELLAN, H., BOEVINK, P. C., ARMSTRONG, M. R., BUKHAROVA, T., SUKARTA, O., WIN, J., KAMOUN, S., BIRCH, P. R. J. & BANFIELD, M. J. 2014. Phytophthora infestans RXLR Effector PexRD2 Interacts with Host MAPKKKε to Suppress Plant Immune Signaling. *The Plant Cell Online,* 26**,** 1345-1359.

KONCZ, C. & SCHELL, J. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Molecular and General Genetics MGG,* 204**,** 383-396.

KORDES, M., RÖRING, M., HEINING, C., BRAUN, S., HUTTER, B., RICHTER, D., GEÖRG, C., SCHOLL, C., GRÖSCHEL, S. & ROTH, W. 2015. Cooperation of BRAFF595L and mutant HRAS in histiocytic sarcoma provides new insights into oncogenic BRAF signaling. *Leukemia*.

KROJ, T., CHANCLUD, E., MICHEL-ROMITI, C., GRAND, X. & MOREL, J.-B. 2016. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist,* 210**,** 618-626.

KYRIAKIS, J. M. 1999. Signaling by the germinal center kinase family of protein kinases. *J Biol Chem,* 274**,** 5259-62.

LAEMMLI, U. K. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature,* 227**,** 680-685.

LAREBEKE, N. V., ENGLER, G., HOLSTERS, M., DEN ELSACKER, S. V., ZAENEN, I., SCHILPEROORT, R. A. & SCHELL, J. 1974. Large plasmid in Agrobacterium tumefaciens essential for crown gall-inducing ability. *Nature,* 252**,** 169-170.

LEBERER, E., DIGNARD, D., THOMAS, D. Y. & LEEUW, T. 2000. A conserved Gbeta binding (GBB) sequence motif in Ste20p/PAK family protein kinases. *Biol Chem,* 381**,** 427-31.

LEEUW, T., WU, C., SCHRAG, J. D., WHITEWAY, M., THOMAS, D. Y. & LEBERER, E. 1998. Interaction of a G-protein beta-subunit with a conserved sequence in Ste20/PAK family protein kinases. *Nature,* 391**,** 191-5.

LESLIE, A. G. W. & POWELL, H. R. 2007. Processing diffraction data with mosflm. *In:* READ, R. J. & SUSSMAN, J. L. (eds.) *Evolving Methods for Macromolecular Crystallography: The Structural Path to the Understanding*

*of the Mechanismof Action of CBRN Agents.* Dordrecht: Springer Netherlands.

LI, T., LIU, B., SPALDING, M. H., WEEKS, D. P. & YANG, B. 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotech,* 30**,** 390-392.

LIN, M., UNDEN, H., JACQUIER, N., SCHNEITER, R., JUST, U. & HOFKEN, T. 2009. The Cdc42 effectors Ste20, Cla4, and Skm1 down-regulate the expression of genes involved in sterol uptake by a mitogen-activated protein kinase-independent pathway. *Mol Biol Cell,* 20**,** 4826-37.

LINDBO, J. A. 2007. High-efficiency protein expression in plants from agroinfection-compatible Tobacco mosaic virus expression vectors. *BMC Biotechnology,* 7**,** 52-52.

LU, Y.-J., SCHORNACK, S., SPALLEK, T., GELDNER, N., CHORY, J., SCHELLMANN, S., SCHUMACHER, K., KAMOUN, S. & ROBATZEK, S. 2012. Patterns of plant subcellular responses to successful oomycete infections reveal differences in host cell reprogramming and endocytic trafficking. *Cellular Microbiology,* 14**,** 682-697.

LUO, Y., CALDWELL, K. S., WROBLEWSKI, T., WRIGHT, M. E. & MICHELMORE, R. W. 2009. Proteolysis of a negative regulator of innate immunity is dependent on resistance genes in tomato and Nicotiana benthamiana and induced by multiple bacterial effectors. *Plant Cell,* 21**,** 2458-72.

MACKEY, D., BELKHADIR, Y., ALONSO, J. M., ECKER, J. R. & DANGL, J. L. 2003. Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell,* 112**,** 379-89.

MACKEY, D., HOLT, B. F., 3RD, WIIG, A. & DANGL, J. L. 2002. RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell,* 108**,** 743-54.

MALCOLMSON, J. F. 1969. Races of Phytophthora infestans occurring in Great Britain. *Transactions of the British Mycological Society,* 53**,** 417-IN2.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-Guided Human Genome Engineering via Cas9. *Science,* 339**,** 823-826.

MAQBOOL, A., SAITOH, H., FRANCESCHETTI, M., STEVENSON, C. E. M., UEMURA, A., KANZAKI, H., KAMOUN, S., TERAUCHI, R. & BANFIELD, M. J. 2015. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife,* 4**,** e08709.

MARRAFFINI, L. A. & SONTHEIMER, E. J. 2008. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science,* 322**,** 1843-1845.

MARTI, M., GOOD, R. T., RUG, M., KNUEPFER, E. & COWMAN, A. F. 2004. Targeting Malaria Virulence and Remodeling Proteins to the Host Erythrocyte. *Science,* 306**,** 1930-1933.

MATHIEU, J., SCHWIZER, S. & MARTIN, G. B. 2014. Pto Kinase Binds Two Domains of AvrPtoB and Its Proximity to the Effector E3 Ligase Determines if It Evades Degradation and Activates Plant Immunity. *PLoS Pathog,* 10**,** e1004227.

MAYTON, H., SMART, C., MORAVEC, B., MIZUBUTI, E., MULDOON, A. & FRY, W. 2000. Oospore survival and pathogenicity of single oospore recombinant progeny from a cross involving US-17 and US-8 genotypes of Phytophthora infestans. *Plant Disease,* 84**,** 1190-1196.

MCLAUGHLIN, P. D., BOBAY, B. G., REGEL, E. J., THOMPSON, R. J., HOCH, J. A. & CAVANAGH, J. 2007. Predominantly buried residues in the response regulator Spo0F influence specific sensor kinase recognition. *FEBS letters,* 581**,** 1425-1429.

MELECH-BONFIL, S. & SESSA, G. 2010a. Tomato MAPKKK epsilon is a positive regulator of cell-death signaling networks associated with plant immunity. *Plant Journal,* 64**,** 379-391.

MELECH-BONFIL, S. & SESSA, G. 2010b. Tomato MAPKKKepsilon is a positive regulator of cell-death signaling networks associated with plant immunity. *Plant Journal,* 64**,** 379-91.

MILLER, J. C., TAN, S., QIAO, G., BARLOW, K. A., WANG, J., XIA, D. F., MENG, X., PASCHON, D. E., LEUNG, E., HINKLEY, S. J., DULAY, G. P., HUA, K. L., ANKOUDINOVA, I., COST, G. J., URNOV, F. D., ZHANG, H. S., HOLMES, M. C., ZHANG, L., GREGORY, P. D. & REBAR, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotech,* 29**,** 143-148.

MOJICA, F. J., FERRER, C., JUEZ, G. & RODRIGUEZ-VALERA, F. 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol Microbiol,* 17**,** 85-93.

MOJICA, F. J., JUEZ, G. & RODRIGUEZ-VALERA, F. 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol,* 9**,** 613-21.

MOJICA, F. M., DÍEZ-VILLASEÑOR, C. S., GARCÍA-MARTÍNEZ, J. & SORIA, E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution,* 60**,** 174-182.

MOORE, J. K. & HABER, J. E. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology,* 16**,** 2164-2173.

MOSCOU, M. J. & BOGDANOVE, A. J. 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science,* 326**,** 1501-1501.

MURRAY, M. & THOMPSON, W. F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic acids research,* 8**,** 4321-4326.

NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & VON HEIJNE, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng,* 10**,** 1-6.

NISHIHAMA, R., ISHIKAWA, M., ARAKI, S., SOYANO, T., ASADA, T. & MACHIDA, Y. 2001. The NPK1 mitogen-activated protein kinase kinase kinase is a regulator of cell-plate formation in plant cytokinesis. *Genes & Development,* 15**,** 352-363.

NOWARA, D., GAY, A., LACOMME, C., SHAW, J., RIDOUT, C., DOUCHKOV, D., HENSEL, G., KUMLEHN, J. & SCHWEIZER, P. 2010. HIGS: Host-Induced Gene Silencing in the Obligate Biotrophic Fungal Pathogen Blumeria graminis. *The Plant Cell,* 22**,** 3130-3141.

OH, S. K., YOUNG, C., LEE, M., OLIVA, R., BOZKURT, T. O., CANO, L. M., WIN, J., BOS, J. I. B., LIU, H. Y., VAN DAMME, M., MORGAN, W., CHOI, D., VAN DER VOSSEN, E. A. G., VLEESHOUWERS, V. G. A. A. & KAMOUN, S. 2009. In Planta Expression Screens of Phytophthora infestans RXLR Effectors Reveal Diverse Phenotypes, Including Activation of the Solanum bulbocastanum Disease Resistance Protein Rpi-blb2. *Plant Cell,* 21**,** 2928-2947.

OOME, S., RAAYMAKERS, T. M., CABRAL, A., SAMWEL, S., BÖHM, H., ALBERT, I., NÜRNBERGER, T. & VAN DEN ACKERVEKEN, G. 2014. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in Arabidopsis. *Proceedings of the National Academy of Sciences,* 111**,** 16955-16960.

OTTMANN, C., LUBERACKI, B., KÜFNER, I., KOCH, W., BRUNNER, F., WEYAND, M., MATTINEN, L., PIRHONEN, M., ANDERLUH, G., SEITZ, H. U., NÜRNBERGER, T. & OECKING, C. 2009. A common toxin

fold mediates microbial attack and plant defense. *Proceedings of the National Academy of Sciences,* 106**,** 10359-10364.

PATRON, N. J., ROGERS, M. B. & KEELING, P. J. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryotic Cell,* 3**,** 1169-1175.

PEDERSEN, C., VAN THEMAAT, E. V. L., MCGUFFIN, L. J., ABBOTT, J. C., BURGIS, T. A., BARTON, G., BINDSCHEDLER, L. V., LU, X., MAEKAWA, T., WEßLING, R., CRAMER, R., THORDAL-CHRISTENSEN, H., PANSTRUGA, R. & SPANU, P. D. 2012. Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics,* 13**,** 1-21.

PETERSEN, M., BRODERSEN, P., NAESTED, H., ANDREASSON, E., LINDHART, U., JOHANSEN, B., NIELSEN, H. B., LACY, M., AUSTIN, M. J., PARKER, J. E., SHARMA, S. B., KLESSIG, D. F., MARTIENSSEN, R., MATTSSON, O., JENSEN, A. B. & MUNDY, J. 2000. Arabidopsis map kinase 4 negatively regulates systemic acquired resistance. *Cell,* 103**,** 1111-20.

PFUND, C., TANS-KERSTEN, J., DUNNING, F. M., ALONSO, J. M., ECKER, J. R., ALLEN, C. & BENT, A. F. 2004. Flagellin Is Not a Major Defense Elicitor in Ralstonia solanacearum Cells or Extracts Applied to Arabidopsis thaliana. *Molecular Plant-Microbe Interactions,* 17**,** 696-706.

PIECHOTTA, K., GARBARINI, N., ENGLAND, R. & DELPIRE, E. 2003. Characterization of the Interaction of the Stress Kinase SPAK with the Na+-K+-2Cl– Cotransporter in the Nervous System: EVIDENCE FOR A SCAFFOLDING ROLE OF THE KINASE. *Journal of Biological Chemistry,* 278**,** 52848-52856.

POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology,* 151**,** 653-63.

POZO, O. D., PEDLEY, K. F. & MARTIN, G. B. 2004. MAPKKKα is a positive regulator of cell death associated with both plant immunity and disease. *The EMBO Journal,* 23**,** 3072-3082.

QI, J., ASANO, T., JINNO, M., MATSUI, K., ATSUMI, K., SAKAGAMI, Y. & OJIKA, M. 2005. Characterization of a Phytophthora Mating Hormone. *Science,* 309**,** 1828.

QUTOB, D., KAMOUN, S. & GIJZEN, M. 2002. Expression of a Phytophthora sojae necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *Plant J,* 32**,** 361-73.

RAFFAELE, S., WIN, J., CANO, L. M. & KAMOUN, S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. *BMC Genomics,* 11**,** 637.

RAMIREZ, C. L., FOLEY, J. E., WRIGHT, D. A., MULLER-LERCH, F., RAHMAN, S. H., CORNU, T. I., WINFREY, R. J., SANDER, J. D., FU, F., TOWNSEND, J. A., CATHOMEN, T., VOYTAS, D. F. & JOUNG, J. K. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Meth,* 5**,** 374-375.

ROUXEL, T., PENAUD, A., PINOCHET, X., BRUN, H., GOUT, L., DELOURME, R., SCHMIT, J. & BALESDENT, M.-H. 2003. A 10-year Survey of Populations of Leptosphaeria maculans in France Indicates a Rapid Adaptation Towards the Rlm1 Resistance Gene of Oilseed Rape. *European Journal of Plant Pathology,* 109**,** 871-881.

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

SAPRANAUSKAS, R., GASIUNAS, G., FREMAUX, C., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research,* 39**,** 9275-9282.

SARRIS, P. F., CEVIK, V., DAGDAS, G., JONES, J. D. & KRASILEVA, K. V. 2016. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC biology,* 14**,** 1.

SAUNDERS, D. G., BREEN, S., WIN, J., SCHORNACK, S., HEIN, I., BOZKURT, T. O., CHAMPOURET, N., VLEESHOUWERS, V. G., BIRCH, P. R. & GILROY, E. M. 2012. Host protein BSL1 associates with Phytophthora infestans RXLR effector AVR2 and the Solanum demissum immune receptor R2 to mediate disease resistance. *The Plant Cell,* 24**,** 3420-3434.

SCHIRAWSKI, J., MANNHAUPT, G., MÜNCH, K., BREFORT, T., SCHIPPER, K., DOEHLEMANN, G., DI STASIO, M., RÖSSEL, N., MENDOZA-MENDOZA, A., PESTER, D., MÜLLER, O., WINTERBERG, B., MEYER, E., GHAREEB, H., WOLLENBERG, T., MÜNSTERKÖTTER, M., WONG, P., WALTER, M., STUKENBROCK, E., GÜLDENER, U. & KAHMANN, R. 2010. Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison. *Science,* 330**,** 1546-1548.

SCHLOSSER, A., VANSELOW, J. T. & KRAMER, A. 2005. Mapping of Phosphorylation Sites by a Multi-Protease Approach with Specific Phosphopeptide Enrichment and NanoLC−MS/MS Analysis. *Analytical Chemistry,* 77**,** 5243-5250.

SCHORNACK, S., VAN DAMME, M., BOZKURT, T. O., CANO, L. M., SMOKER, M., THINES, M., GAULIN, E., KAMOUN, S. & HUITEMA, E. 2010. Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings of the National Academy of Sciences,* 107**,** 17421-17426.

SEGRETIN, M. E., PAIS, M., FRANCESCHETTI, M., CHAPARRO-GARCIA, A., BOS, J. I., BANFIELD, M. J. & KAMOUN, S. 2014. Single amino acid mutations in the potato immune receptor R3a expand response to Phytophthora effectors. *Mol Plant Microbe Interact,* 27**,** 624-37.

SEREBRIISKII, I. G. & KOTOVA, E. 2004. Analysis of protein-protein interactions utilizing dual bait yeast two-hybrid system. *Protein-Protein Interactions: Methods and Applications***,** 263-296.

SHAH, J. 2009. Plants under attack: systemic signals in defence. *Current opinion in plant biology,* 12**,** 459-464.

SHEVCHENKO, A., TOMAS, H., HAVLIS, J., OLSEN, J. V. & MANN, M. 2007. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols,* 1**,** 2856-2860.

SPANU, P. D., ABBOTT, J. C., AMSELEM, J., BURGIS, T. A., SOANES, D. M., STÜBER, K., LOREN VAN THEMAAT, E. V., BROWN, J. K. M., BUTCHER, S. A., GURR, S. J., LEBRUN, M.-H., RIDOUT, C. J., SCHULZE-LEFERT, P., TALBOT, N. J., AHMADINEJAD, N., AMETZ, C., BARTON, G. R., BENJDIA, M., BIDZINSKI, P., BINDSCHEDLER, L. V., BOTH, M., BREWER, M. T., CADLE-DAVIDSON, L., CADLE-DAVIDSON, M. M., COLLEMARE, J., CRAMER, R., FRENKEL, O., GODFREY, D., HARRIMAN, J., HOEDE, C., KING, B. C., KLAGES, S., KLEEMANN, J., KNOLL, D., KOTI, P. S., KREPLAK, J., LÓPEZ-RUIZ, F. J., LU, X., MAEKAWA, T., MAHANIL, S., MICALI, C., MILGROOM, M. G., MONTANA, G., NOIR, S., O'CONNELL, R. J., OBERHAENSLI, S., PARLANGE, F., PEDERSEN, C., QUESNEVILLE, H., REINHARDT, R., ROTT, M., SACRISTÁN, S., SCHMIDT, S. M., SCHÖN, M., SKAMNIOTI, P., SOMMER, H., STEPHENS, A., TAKAHARA, H., THORDAL-CHRISTENSEN, H., VIGOUROUX, M., WEßLING, R., WICKER, T. & PANSTRUGA, R. 2010. Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. *Science,* 330**,** 1543-1546.

SREERAMA, N., VENYAMINOV, S. Y. U. & WOODY, R. W. 1999. Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Science,* 8**,** 370-380.

SREERAMA, N. & WOODY, R. W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem,* 287**,** 252-60.

STADHOUDERS, R., PAS, S. D., ANBER, J., VOERMANS, J., MES, T. H. M. & SCHUTTEN, M. 2010. The Effect of Primer-Template Mismatches on the Detection and Quantification of Nucleic Acids Using the 5′ Nuclease Assay. *The Journal of Molecular Diagnostics : JMD,* 12**,** 109-117.

STAM, R., HOWDEN, A. J. M., DELGADO-CEREZO, M., M. M. AMARO, T. M., MOTION, G. B., PHAM, J. & HUITEMA, E. 2013. Characterization of cell death inducing Phytophthora capsici CRN effectors suggests diverse activities in the host nucleus. *Frontiers in Plant Science,* 4**,** 387.

STASKAWICZ, B., AUSUBEL, F., BAKER, B., ELLIS, J. & JONES, J. 1995. Molecular genetics of plant disease resistance. *Science,* 268**,** 661-667.

STRONG, T. C., KAUR, G. & THOMAS, J. H. 2011. Mutations in the Catalytic Loop HRD Motif Alter the Activity and Function of Drosophila Src64. *PLoS ONE,* 6**,** e28100.

SU, Y. C., HAN, J., XU, S., COBB, M. & SKOLNIK, E. Y. 1997. NIK is a new Ste20-related kinase that binds NCK and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain. *The EMBO Journal,* 16**,** 1279-1290.

SUN, F., KALE, S. D., AZURMENDI, H. F., LI, D., TYLER, B. M. & CAPELLUTO, D. G. S. 2012. Structural Basis for Interactions of the Phytophthora sojae RxLR Effector Avh5 with Phosphatidylinositol 3-Phosphate and for Host Cell Entry. *Molecular Plant-Microbe Interactions,* 26**,** 330-344.

SUN, W., DUNNING, F. M., PFUND, C., WEINGARTEN, R. & BENT, A. F. 2006. Within-Species Flagellin Polymorphism in Xanthomonas campestris pv campestris and Its Impact on Elicitation of Arabidopsis FLAGELLIN SENSING2–Dependent Defenses. *The Plant Cell,* 18**,** 764-779.

SUNG, P. & KLEIN, H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol,* 7**,** 739-50.

THOMMA, B. P., CAMMUE, B. P. & THEVISSEN, K. 2002. Plant defensins. *Planta,* 216**,** 193-202.

TICKLE, I. J. & DRIESSEN, H. P. 1996. Molecular replacement using known structural information. *Crystallographic Methods and Protocols***,** 173-203.

TORTO, T. A., LI, S., STYER, A., HUITEMA, E., TESTA, A., GOW, N. A., VAN WEST, P. & KAMOUN, S. 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. *Genome Res,* 13**,** 1675-85.

TYLER, B. M., TRIPATHY, S., ZHANG, X., DEHAL, P., JIANG, R. H., AERTS, A., ARREDONDO, F. D., BAXTER, L., BENSASSON, D., BEYNON, J. L., CHAPMAN, J., DAMASCENO, C. M., DORRANCE, A. E., DOU, D., DICKERMAN, A. W., DUBCHAK, I. L., GARBELOTTO, M., GIJZEN, M., GORDON, S. G., GOVERS, F., GRUNWALD, N. J., HUANG, W., IVORS, K. L., JONES, R. W., KAMOUN, S., KRAMPIS, K., LAMOUR, K. H., LEE, M. K., MCDONALD, W. H., MEDINA, M., MEIJER, H. J., NORDBERG, E. K., MACLEAN, D. J., OSPINA-GIRALDO, M. D., MORRIS, P. F., PHUNTUMART, V., PUTNAM, N. H., RASH, S., ROSE, J. K., SAKIHAMA, Y., SALAMOV, A. A., SAVIDOR, A., SCHEURING, C. F., SMITH, B. M., SOBRAL, B. W., TERRY, A., TORTO-ALALIBO, T. A., WIN, J., XU, Z., ZHANG, H., GRIGORIEV, I. V., ROKHSAR, D. S. & BOORE, J. L. 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science,* 313**,** 1261-6.

UNDERWOOD, W. 2012. The plant cell wall: A dynamic barrier against pathogen invasion. *Frontiers in Plant Science,* 3.

URNOV, F. D., MILLER, J. C., LEE, Y.-L., BEAUSEJOUR, C. M., ROCK, J. M., AUGUSTUS, S., JAMIESON, A. C., PORTEUS, M. H., GREGORY, P. D. & HOLMES, M. C. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**,** 646-651.

VAN DAMME, M., BOZKURT, T. O., CAKIR, C., SCHORNACK, S., SKLENAR, J., JONES, A. M. & KAMOUN, S. 2012. The Irish potato famine pathogen Phytophthora infestans translocates the CRN8 kinase into host plant cells. *PLoS Pathog,* 8**,** e1002875.

VAN DER HOORN, R. A., LAURENT, F., ROTH, R. & DE WIT, P. J. 2000. Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr 9/Cf-9-induced and Avr 4/Cf-4-induced necrosis. *Molecular Plant-Microbe Interactions,* 13**,** 439-446.

VANETTEN, H. D., MANSFIELD, J. W., BAILEY, J. A. & FARMER, E. E. 1994. Two Classes of Plant Antibiotics: Phytoalexins versus "Phytoanticipins". *Plant Cell,* 6**,** 1191-1192.

VE, T., WILLIAMS, S. J., CATANZARITI, A.-M., RAFIQI, M., RAHMAN, M., ELLIS, J. G., HARDHAM, A. R., JONES, D. A., ANDERSON, P. A., DODDS, P. N. & KOBE, B. 2013. Structures of the flax-rust effector AvrM

reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 17594-17599.

VOINNET, O., RIVAS, S., MESTRE, P. & BAULCOMBE, D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal,* 33**,** 949-956.

WALDEN, M., CROW, A., NELSON, M. D. & BANFIELD, M. J. 2014. Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the S. pyogenes pilus adhesin Spy0125. *Proteins: Structure, Function, and Bioinformatics,* 82**,** 517-527.

WANG, J., KAMTEKAR, S., BERMAN, A. J. & STEITZ, T. A. 2005. Correction of X-ray intensities from single crystals containing lattice-translocation defects. *Acta Crystallographica Section D: Biological Crystallography,* 61**,** 67-74.

WANG, X., BOEVINK, P., MCLELLAN, H., ARMSTRONG, M., BUKHAROVA, T., QIN, Z. & BIRCH, PAUL R. 2015. A Host KH RNA-Binding Protein Is a Susceptibility Factor Targeted by an RXLR Effector to Promote Late Blight Disease(). *Molecular Plant,* 8**,** 1385-1395.

WANG, Y., LI, J., HOU, S., WANG, X., LI, Y., REN, D., CHEN, S., TANG, X. & ZHOU, J.-M. 2010. A Pseudomonas syringae ADP-Ribosyltransferase Inhibits Arabidopsis Mitogen-Activated Protein Kinase Kinases. *The Plant Cell,* 22**,** 2033-2044.

WAWRA, S., DJAMEI, A., ALBERT, I., NÜRNBERGER, T., KAHMANN, R. & VAN WEST, P. 2013. In Vitro Translocation Experiments with RxLR-Reporter Fusion Proteins of Avr1b from Phytophthora sojae and AVR3a from Phytophthora infestans Fail to Demonstrate Specific Autonomous Uptake in Plant and Animal Cells. *Molecular Plant-Microbe Interactions,* 26**,** 528-536.

WHISSON, S. C., BOEVINK, P. C., MOLELEKI, L., AVROVA, A. O., MORALES, J. G., GILROY, E. M., ARMSTRONG, M. R., GROUFFAUD, S., VAN WEST, P., CHAPMAN, S., HEIN, I., TOTH, I. K., PRITCHARD, L. & BIRCH, P. R. J. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature,* 450**,** 115-118.

WHITMORE, L. & WALLACE, B. A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers,* 89**,** 392-400.

WILLIAMS, S. J., SOHN, K. H., WAN, L., BERNOUX, M., SARRIS, P. F., SEGONZAC, C., VE, T., MA, Y., SAUCET, S. B., ERICSSON, D. J., CASEY, L. W., LONHIENNE, T., WINZOR, D. J., ZHANG, X., COERDT, A., PARKER, J. E., DODDS, P. N., KOBE, B. & JONES, J. D. 2014. Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science,* 344**,** 299-303.

WIN J, C.-G. A., BELHAJ K, SAUNDERS DGO, YOSHIDA K, DONG S, SCHORNAK S, ZIPFEL C, ROBATZEK S, HOGENHOUT SA, KAMOUN S 2012. Effector Biology of Plant-Associated Organisms: Concepts and Perspectives. *Cold Spring Harb Symp Quant Biol,* 77**,** 235-247.

WIN, J., KANNEGANTI, T. D., TORTO-ALALIBO, T. & KAMOUN, S. 2006. Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen Phytophthora infestans. *Fungal Genet Biol,* 43**,** 20-33.

WIN, J., KRASILEVA, K. V., KAMOUN, S., SHIRASU, K., STASKAWICZ, B. J. & BANFIELD, M. J. 2012. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLoS Pathog,* 8**,** e1002400.

WINTER, G., LOBLEY, C. M. & PRINCE, S. M. 2013. Decision making in xia2. *Acta Crystallographica Section D: Biological Crystallography,* 69**,** 1260-1273.

WIRTHMUELLER, L., MAQBOOL, A. & BANFIELD, M. J. 2013. On the front line: structural insights into plant-pathogen interactions. *Nat Rev Micro,* 11**,** 761-776.

YAMAGUCHI, K., YAMADA, K., ISHIKAWA, K., YOSHIMURA, S., HAYASHI, N., UCHIHASHI, K., ISHIHAMA, N., KISHI-KABOSHI, M., TAKAHASHI, A., TSUGE, S., OCHIAI, H., TADA, Y., SHIMAMOTO, K., YOSHIOKA, H. & KAWASAKI, T. 2013. A receptor-like cytoplasmic kinase targeted by a plant pathogen effector is directly phosphorylated by the chitin receptor and mediates rice immunity. *Cell Host Microbe,* 13**,** 347-57.

YANG, B., SUGIO, A. & WHITE, F. F. 2006. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proceedings of the National Academy of Sciences,* 103**,** 10503-10508.

ZHANG, F., CONG, L., LODATO, S., KOSURI, S., CHURCH, G. M. & ARLOTTA, P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol,* 29**,** 149-53.

ZHANG, J., SHAO, F., LI, Y., CUI, H., CHEN, L., LI, H., ZOU, Y., LONG, C., LAN, L., CHAI, J., CHEN, S., TANG, X. & ZHOU, J.-M. 2007. A *Pseudomonas syringae* Effector Inactivates MAPKs to Suppress PAMP-Induced Immunity in Plants. *Cell Host & Microbe,* 1**,** 175-185.

ZHANG, L., WANG, J.-C., HOU, L., CAO, P.-R., WU, L., ZHANG, Q.-S., YANG, H.-Y., ZANG, Y., DING, J.-P. & LI, J. 2015. Functional Role of Histidine in the Conserved His-x-Asp Motif in the Catalytic Core of Protein Kinases. *Scientific Reports,* 5**,** 10115.

ZHANG, W. J., PEDERSEN, C., KWAAITAAL, M., GREGERSEN, P. L., MORCH, S. M., HANISCH, S., KRISTENSEN, A., FUGLSANG, A. T., COLLINGE, D. B. & THORDAL-CHRISTENSEN, H. 2012a. Interaction of barley powdery mildew effector candidate CSEP0055 with the defence protein PR17c. *Mol Plant Pathol,* 13**,** 1110-9.

ZHANG, Z., WU, Y., GAO, M., ZHANG, J., KONG, Q., LIU, Y., BA, H., ZHOU, J. & ZHANG, Y. 2012b. Disruption of PAMP-induced MAP kinase cascade by a Pseudomonas syringae effector activates plant immunity mediated by the NB-LRR protein SUMM2. *Cell Host Microbe,* 11**,** 253-63.

ZHAO, Z. S., MANSER, E., CHEN, X. Q., CHONG, C., LEUNG, T. & LIM, L. 1998. A conserved negative regulatory region in alphaPAK: inhibition of PAK kinases reveals their morphological roles downstream of Cdc42 and Rac1. *Mol Cell Biol,* 18**,** 2153-63.

ZHENG, X., MCLELLAN, H., FRAITURE, M., LIU, X., BOEVINK, P. C., GILROY, E. M., CHEN, Y., KANDEL, K., SESSA, G., BIRCH, P. R. J. & BRUNNER, F. 2014. Functionally Redundant RXLR Effectors from <italic>Phytophthora infestans</italic> Act at Different Steps to Suppress Early flg22-Triggered Immunity. *PLoS Pathog,* 10**,** e1004057.

ZUO, J., NIU, Q.-W. & CHUA, N.-H. 2000. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant Journal,* 24**,** 265-273.

# Appendix

## Supplementary figure

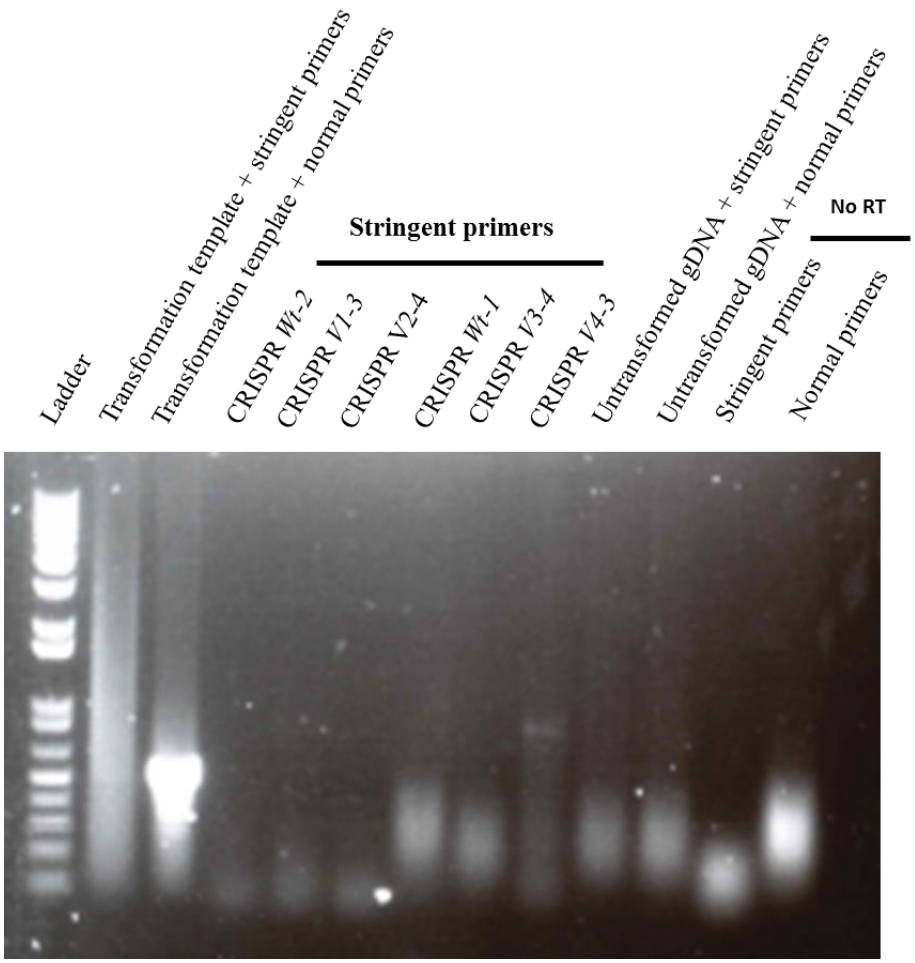



**Figure S1| Transgene screening primer check.**

1% agarose DNA gel demonstrating inability of phosphorothioate bonded oligos to amplify any target sequence and the inability of mutant specific oligos to amplify wild type sequence from untransformed tomatoes, demonstrating their specificity in amplifying transgene targets. Mutant specific oligos amplify target sequence from transformation template positive control demonstrating their ability to amplify target sequence.

**List of primers used in the present study.**

| Primer ID[13] | Sequence (5' → 3') | Purpose |
|---|---|---|
| **440F** | **CACC**GAACAGCTAACAGAGAAAAAGTTC | Cloning of *St*MAPKKKK into pENTR™/D-TOPO®. Starts at residue 2. |
| **441R** | TCAGCGATTCAATTTCTTTTCCAACTG | Cloning *St*MAPKKKK into pENTR™/D-TOPO®. Ends at C-terminus |
| **482R** | TTATCAATACATTGCCTGGTTCTGTAAAAG | Cloning *St*MAPKKKK into pENTR™/D-TOPO®. Ends at residue 319. |
| **464F** | **CACC**ATGGCACGTACAATCCTTGATGG | Cloning *St*MAPKKKK into pENTR™/D-TOPO®. Starts at residue 287. |
| | | Random mutagenesis |

[13] Primer ID number refers to an identifying number in the Banfield group's store of oligonucleotides. 'F' or 'R' refer to forward or reverse primers, respectively, with respect to the orientation of the coding sequence.

| | | |
|---|---|---|
| **614F** | CAGCTCGAGCTATGTCTAGGCAAATGG...<br>...CAAATGC | PCR of MAPKKK$_{\varepsilon}$ 1-332 and cloning into pERCH. Contains *XhoI* site. Starts at N-terminus. |
| **682R** | GGATTAATTAATTCAGGTGGTGCCAAT | Random mutagenesis PCR of MAPKKK$_{\varepsilon}$ 1-332 and cloning into pERCH. Contains *PacI* site. |
| **438F** | **CACC**TCTAGGCAAATGGCAAATGCTG | Cloning of MAPKKK$_{\varepsilon}$ into pENTR™/D-TOPO®. Starts at residue 2. |
| **1051R** | TCATTCAGGTGGTGCCAATGTTGT | Cloning MAPKKK$_{\varepsilon}$ 1-332 into pENTR™/D-TOPO®. Ends at residue 332. |
| **1147F** | TGCTATATTTTCGAAGGAAAC | PCR-RE assay. 5' of sgRNA1 target sequence. |
| **1148R** | GATCAGAATACGAGGGTCAAG | PCR-RE assay. 3' of sgRNA3 target sequence. |

| | | |
|---|---|---|
| **1149F** | AGTTAGATGTAGAAGTTCCAG | PCR-RE assay. 5’ of sgRNA2 target sequence. Also used for sequencing of ~850bp PCR products in PCR-RE assay. |
| **1149R** | GGTGTTCCAACAACTGAATGC | PCR-RE assay. 3’ of sgRNA4 binding sequence. Also used for cloning large deletion band. |
| **1212F** | ATGTCTAGGCAAATGGCA | RT-PCR of MAPKKK$_{\varepsilon}$ complementation transgene. Forward from start codon. |
| **1213R** | TTAACATCTGCtTCaGTc | RT-PCR of MAPKKK$_{\varepsilon}$ complementation transgene. Mutant specific. |
| **1227F** | **CACC**TGCTATATTTTCGAAGGAAAC | Cloning of large deletion band in edited MAPKKK$_{\varepsilon}$ pENTR™/D-TOPO®. |

**Table A Entry clones used in the present study.**

| Construct Name | Vector Backbone | Insert/Description | Start Codon | Stop Codon | Source/Reference |
|---|---|---|---|---|---|
| **pENTR:PexRD2$^{\Delta 21}$** | **pENTR™/D-TOPO®** | **PITG_21422, residues 21 - 121** | • | • | King *et al.* (2013). |
| **pENTR:PexRD2$^{\Delta 42}$** | **pENTR™/D-TOPO®** | **PITG_21422, residues 42 - 121** | • | • | King *et al.* (2013). |
| **pENTR:PexRD2$^{\Delta 57}$** | **pENTR™/D-TOPO®** | **PITG_21422, residues 57 - 121** | • | • | King *et al.* (2013). |
| **pENTR:PITG_14984** | **pENTR™/D-TOPO®** | **PITG_14984 (PexRD2 homolog), Residues 21-121** | • | • | King *et al.* (2013). |
| **pENTR:PITG_14787** | **pENTR™/D-TOPO®** | **PITG_14787 (PexRD2 homolog), Residues 21-121** | • | • | King *et al.* (2013). |
| **pENTR:MAPKKK$_{\varepsilon}^{1-300}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\varepsilon}$ residues 1-300** | • | • | King *et al.* (2013). |
| **pENTR:MAPKKK$_{\varepsilon}^{1-332\ F10L}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\varepsilon}$ residues 1-332, Phenylanine at position 10 mutated to Leucine.** | • | • | King *et al.* (2013) |
| **pENTR:MAPKKK$_{\varepsilon}^{1-332\ Q50R}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\varepsilon}$ residues 1-332, Glutamine at position 50 mutated to Arginine.** | • | • | This study. Sub-cloned from constructs synthesised by Genscript. |
| **pENTR:MAPKKK$_{\varepsilon}^{1-332\ S200T}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\varepsilon}$ residues 1-332, Serine at position 200 mutated to Threonine.** | • | • | This study. Sub-cloned from constructs synthesised by Genscript. |

| | | | | | |
|---|---|---|---|---|---|
| **pENTR:MAPKKK$_{\varepsilon}^{1\text{-}332\ D241N}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\varepsilon}$ residues 1-332, Aspartate at position 241 mutated to Asparagine.** | • | • | **This study. Sub-cloned from constructs synthesised by Genscript.** |
| **pENTR:MAPKKK$_{\alpha}^{FL}$** | **pENTR™/D-TOPO®** | ***St*MAPKKK$_{\alpha}$ residues 2-1401, Full length protein.** | | • | King *et al.* (2013). |
| **pENTR:MAPKKKK$^{FL}$** | **pENTR™/D-TOPO®** | ***St*MAPKKKK residues 1-674.** | • | • | King *et al.* (2013). |
| **pENTR:MAPKKKK$^{1\text{-}389}$** | **pENTR™/D-TOPO®** | ***St*MAPKKKK residues 1-389.** | • | • | King *et al.* (2013). |
| **pENTR:MAPKKKK$^{287\text{-}674}$** | **pENTR™/D-TOPO®** | ***St*MAPKKKK residues 287-674.** | | • | King *et al.* (2013). |

**Table B Yeast 2-Hybrid clones used in the present study**

| Construct Name | Vector Backbone | Insert/Description | Source/Reference |
|---|---|---|---|
| **pDEST32 (Bait vector control)** | pEXP™32 | *attR*1-ccdB-CmR-*attR*2 | Invitrogen™ |
| **pDEST22 (Prey vector control)** | pEXP™22 | *attR*1-ccdB-CmR-*attR*2 | Invitrogen™ |
| **pEXP32:PexRD2$^{\Delta 20}$** | pEXP™32 | PexRD2 (PITG_21422), residues 21 – 121. | Sub-cloned by LR reaction from entry clone. |
| **pEXP32:PexRD2$^{\Delta 41}$** | pEXP™32 | PexRD2 (PITG_21422), residues 42 – 121. | Sub-cloned by LR reaction from entry clone. |
| **pEXP32:PexRD2$^{\Delta 56}$** | pEXP™32 | PexRD2 (PITG_21422), residues 42 – 121. | Sub-cloned by LR reaction from entry clone. |
| **pEXP32:PexRD2$^{L109D}$** | pEXP™32 | PexRD2 (PITG_21422), residues 21 – 121, Leucine at position 109 mutated to Aspartate. | Sub-cloned by LR reaction from entry clone. |
| **pEXP32:PITG_14984** | pEXP™32 | PITG_14984 (PexRD2 homolog), residues 21-121. | Sub-cloned by LR reaction from entry clone. |
| **pEXP32:PITG_14787** | pEXP™32 | PITG_14787 (PexRD2 homolog), Residues 21-121 | Sub-cloned by LR reaction from entry clone. |

| **pEXP22:MAPKKK$_{\varepsilon}^{1-300}$** | **pEXP™22** | ***St*MAPKKK$_{\varepsilon}$ residues 1-300** | **Sub-cloned by LR reaction from entry clone.** |
|---|---|---|---|
| **pEXP22:MAPKKK$_{\varepsilon}^{1-332\ F10L}$** | pEXP™22 | *St*MAPKKK$_{\varepsilon}^{F10L}$ residues 1-332 | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKK$_{\varepsilon}^{1-332\ Q50R}$** | pEXP™22 | *St*MAPKKK$_{\varepsilon}^{Q50R}$ residues 1-332 | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKK$_{\varepsilon}^{1-332\ S200T}$** | pEXP™22 | *St*MAPKKK$_{\varepsilon}^{S200T}$ residues 1-332 | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKK$_{\varepsilon}^{1-332\ D241N}$** | pEXP™22 | *St*MAPKKK$_{\varepsilon}^{D241N}$ residues 1-332 | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKK$_{\alpha}^{FL}$** | pEXP™22 | StMAPKKKα residues 2-1401, Full length protein. | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKKK$^{FL}$** | pEXP™22 | *St*MAPKKKK residues 1-674. | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKKK$^{1-389}$** | pEXP™22 | *St*MAPKKKK residues 1-389. | Sub-cloned by LR reaction from entry clone. |
| **pEXP22:MAPKKKK$^{287-674}$** | pEXP™22 | *St*MAPKKKK residues 297-674. | Sub-cloned by LR reaction from entry clone. |

**Table C Binary constructs used in the present study**

| Construct Name | Vector Backbone | Insert/Description | Antibiotic Selection | | | Strain | Source/Reference |
|---|---|---|---|---|---|---|---|
| | | | Kan$^r$ | Spec$^r$ | Gent$^r$ | | |
| **pJL3-P19** | pCB301 derivative | p19 cloned into *PacI*/*XbaI* sites. | • | | | GV3101 | Kamoun Lab (TSL), (Lindbo, 2000) |
| **pB7WGF(3xHA)2** | pB7WGF2 derivative | | | • | | GV3101 | Generated by Dr Marina Franceschetti (Banfield Lab) |
| **pB7WGF(3xHA)2:MAPKKKK$^{FL}$** | pB7WGF2 derivative | *St*MAPKKKK (residues 1 - 674). | | • | | GV3101 | Sub-cloned by LR reaction from entry clone |
| **pB7WGF(3xHA)2:MAPKKKK$^{1-389}$** | pB7WGF2 derivative | *St*MAPKKKK (residues 1-389). | | • | | GV3101 | Sub-cloned by LR reaction from entry clone |
| **pERCH: MAPKKK$_\varepsilon^{1-332\ Wt}$** | pERCH | *Sl*MAPKKK$_\varepsilon^{1-332\ Wt}$ | | • | | GV3101 | Synthesised by Genscript USA |
| **pERCH: MAPKKK$_\varepsilon^{1-332\ F10L}$** | pERCH | *Sl*MAPKKK$_\varepsilon^{1-332\ F10L}$ | | • | | GV3101 | Synthesised by Genscript USA |
| **pERCH: MAPKKK$_\varepsilon^{1-332\ Q50R}$** | pERCH | *Sl*MAPKKK$_\varepsilon^{1-332\ Q50R}$ | | • | | GV3101 | Synthesised by Genscript USA |
| **pERCH: MAPKKK$_\varepsilon^{1-332\ S200T}$** | pERCH | *Sl*MAPKKK$_\varepsilon^{1-332\ S200T}$ | | • | | GV3101 | Synthesised by Genscript USA |
| **pERCH: MAPKKK$_\varepsilon^{1-332\ D241N}$** | pERCH | *Sl*MAPKKK$_\varepsilon^{1-332\ D241N}$ | | | | GV3101 | Synthesised by Genscript USA |

| | | | | | | |
|---|---|---|---|---|---|---|
| **pB7WGF(3xHA)2:PexRD2$^{\Delta 20}$** | **pB7WGF2** | **PexRD2 (PITG_21422), residues 21 – 121.** | | • | **GV3101** | **Sub-cloned by LR reaction from entry clone.** |
| **pB7WGF(3xHA)2:PexRD2$^{\Delta 20\ L109D}$** | pB7WGF2 | PexRD2$^{L109D}$ (PITG_21422), residues 21 – 121. | | • | GV3101 | Sub-cloned by LR reaction from entry clone. |
| **pTRBO:PexRD2 $^{\Delta 20}$** | pTRBO | | • | | GV3101 | Prepared by Stuart King |
| **Sg1-3** | TSL pICSL | Cas9 driven by 35S promoter and paired sgRNAs. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **Sg2-4** | TSL pICSL | Cas9 driven by 35S promoter and paired sgRNAs. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **pICSL11065** | TSL pICSL | sgRNA2-4, Cas9 and Wild type complementation cassette. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **pICSL11066** | TSL pICSL | sgRNA2-4, Cas9 and V1 complementation cassette. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **pICSL11067** | TSL pICSL | sgRNA2-4, Cas9 and V2 complementation cassette. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **pICSL11068** | TSL pICSL | sgRNA2-4, Cas9 and V3 complementation cassette. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |
| **pICSL11069** | TSL pICSL | sgRNA2-4, Cas9 and V4 complementation cassette. | • | • | AGL1 | Generated by Mr Mark Youles at TSL. |

**Table D pOPIN constructs used in the present study**

| Construct Name | Vector Backbone | Insert description | Source/reference |
|---|---|---|---|
| **pOPINF:CSEP0162** | pOPINF | CSEP0162 (residues 21-155) | Prepared by Miss Agniezska Siwoszek (Univ. of Copenhagen) whilst a visiting student in the Banfield lab. |